



HAL
open science

Caractérisation d'axes drivers de l'oncogénèse des léiomyosarcomes : validations fonctionnelles et cibles thérapeutiques

Lucie Darmusey

► **To cite this version:**

Lucie Darmusey. Caractérisation d'axes drivers de l'oncogénèse des léiomyosarcomes : validations fonctionnelles et cibles thérapeutiques. Cancer. Université Paul Sabatier - Toulouse III, 2020. Français. NNT : 2020TOU30155 . tel-03152462

HAL Id: tel-03152462

<https://theses.hal.science/tel-03152462>

Submitted on 25 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE
Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par
Lucie DARMUSEY

Le 10 décembre 2020

**Caractérisation d'axes drivers de l'oncogenèse des
léiomyosarcomes : validations fonctionnelles et cibles
thérapeutiques**

Ecole doctorale : **BSB - Biologie, Santé, Biotechnologies**

Spécialité : **CANCEROLOGIE**

Unité de recherche :

CRCT - Centre de Recherche en Cancérologie de Toulouse

Thèse dirigée par
Frederic CHIBON

Jury

Mme Marie CASTETS, Rapporteur
M. Mathias MERISCKAY, Rapporteur
M. Jérôme DEJARDIN, Rapporteur
Mme Estelle ESPINOS, Examinatrice
Mme Coralie SENGES, Examinatrice
M. Frédéric CHIBON, Directeur de thèse

THÈSE

En vue de l'obtention du
**DOCTORAT DE L'UNIVERSITÉ
DE TOULOUSE**

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Lucie DARMUSEY

Le 10 Décembre 2020

**Caractérisation d'axes drivers de l'oncogénèse des
léiomyosarcomes : validations fonctionnelles et cibles
thérapeutiques**

Ecole doctorale : **BSB - Biologie, Santé, Biotechnologies**

Spécialité : **CANCEROLOGIE**

Unité de recherche :

CRCT - Centre de Recherche en Cancérologie de Toulouse

Thèse dirigée par

Frédéric CHIBON

Jury

Mme Marie CASTETS, Rapporteuse

M. Mathias MERICKSKAY, Rapporteur

M. Jérôme DEJARDIN, Rapporteur

Mme Estelle ESPINOS, Examinatrice

Mme Coralie SANGENES, Examinatrice

M. Frédéric CHIBON, Directeur de thèse

« Je suis de ceux qui pensent que la science est d'une grande beauté. Un scientifique dans son laboratoire est non seulement un technicien : il est aussi un enfant placé devant des phénomènes naturels qui l'impressionnent comme des contes de fées. »

Marie Curie, Madame Curie, 1938

Remerciements

Je tiens tout d'abord à remercier les membres du jury, Estelle Espinos, Marie Castets, Coralie Sangenes, Mathias Mericskay et Jérôme Déjardin pour avoir accepté d'évaluer mon travail.

Cette thèse fut un travail intense rempli de surprises, bonnes et moins bonnes, mais surtout de rencontres scientifiques et personnelles !

Fred, merci de m'avoir accueillie dans ton équipe, il y a maintenant 4 ans et demi. Après quatre mois de master 1, six mois de master 2, un changement de décor et trois ans de thèse, je tiens à te remercier pour ta confiance et ton enseignement. Tu m'auras toujours poussée à faire mieux et à aller plus loin dans ma réflexion scientifique. Même si cela ne s'est pas toujours fait paisiblement, je te dois le chemin scientifique que j'ai parcouru. Je retiendrais ta passion pour comprendre les sarcomes et l'envie de lier recherche et clinique au maximum. J'ai adoré travailler sur ces sujets variés (presque tout), et je te remercie de la liberté que tu m'as laissée ainsi que pour les opportunités d'apprendre de nouvelles choses !

Mon trio préféré Candice, Noémie et Laura, sans vous cette thèse aurait été beaucoup moins fun !

Ma Cancan, je ne sais pas comment te remercier pour tout ce que tu as fait durant cette thèse... Nos discussions pour comprendre cette fameuse oncogenèse des sarcomes, ta présence quand la biologie c'est de la merde (merci de m'avoir amenée « boire des bières et manger gras » !!), ton aide au labo, nos pauses café du matin, nos matinées au L2 avec la musique à fond, nos soirées, les week-ends ski, nos thés et gâteaux pour le goûter le week-end...tout ça me manquera beaucoup ! Tu as toujours été présente dans les bons comme dans les mauvais moments. Merci aussi de m'avoir nourri pendant mon COVID !! Ne change rien tu es parfaite telle que tu es.

Nono le petit asticot, toujours là avec un grand sourire et des supers expressions ! Merci d'avoir fait ce joli travail sur ATRX (ces couleurs sont vraiment

magnifiques) et surtout surtout d'avoir corrigé mes (nombreuses) fautes d'orthographe dans cette thèse et tous mes mails... Je repenserai avec plaisir à toutes nos discussions au labo, dans la voiture, autour d'un thé ou d'une sangria, en gros dès qu'on le pouvait. Prends bien soin de toi.

Laura, merci d'avoir rendu ces 2 ans et demi à mourir de rire ! Plus que ça, merci pour ton aide avec ma culture pendant mes quelques vacances et pour tous les services que tu m'as rendu au labo. Ta voix si belle me manquera lorsque je réécouterai notre playlist préférée. Tes expressions du Nord et ce petit accent me manqueront (et me manque déjà un peu depuis que tu nous as lâchement abandonnés pour faire ce très beau bébé). J'espère que tout ira pour le mieux pour toi et ta petite famille.

Vous quittez toutes les trois, pour de vrai, sera un crève-cœur même si je ne vous le montrerai peut-être pas sur le coup... Sachez que vous avez pris une place énorme dans mon petit cœur, et je sais qu'on se retrouvera un jour !

Gaëlle, j'ai appris à te connaître à ton arrivée à Toulouse et ce fut un vrai plaisir de travailler avec toi ! Merci pour ta relecture, tes conseils, ton aide avec la fameuse myocardine et tous les petits déj ! Tu auras aussi été une très bonne élève et tu as de très beaux culots 😊.

Ariadna, I know how a thesis can be difficult but I'm sure everything will be okay. Thanks for your smile and the good mood you bring to the lab! Keep going, you've got it !!

Elodie, même si on ne se sera pas beaucoup vu « en vrai » durant ces trois ans, travailler avec toi à distance fut un vrai plaisir. Merci pour ton travail incroyable de bioinfo qui est à l'origine d'une grande partie de mon travail de thèse. Merci aussi pour ta gentillesse et ton sourire durant tous nos Skype !

Lucile et Tom, je vous ai découvert en arrivant en thèse et quelle découverte ! En vrac, merci pour les rires, les cours de R, les bières tard le soir au labo, les explications sur les *primers*, l'aide pour les stat, les bières en soirée, les discussions scientifiques ou pas et tout le reste. Je vous souhaite tout le bonheur du monde

ensemble et j'espère que nos chemins se recroiseront un jour.

Un grand merci à l'équipe de médecins du Sarcome : Thibault, Philippe, Gwen Christine, Anne B, Anne D, Bérénice, Dimitri, de m'avoir tant appris sur la clinique du sarcome et de m'avoir, de temps en temps, rappelé que ce qu'on fait c'est d'abord pour les patients. Mais surtout une très grande pensée pour Sophie, mon anapath' favorite, merci pour toutes ces heures au microscope et d'avoir pris le temps de tout bien m'expliquer !

Lydia et Jess, merci pour votre douceur et toutes les techniques que vous m'avez apprises durant mes stages à Bordeaux.

Aux voisins de l'étage, les équipes 13 et 21, merci de votre aide à l'aménagement et pour la bonne humeur qui règne au 4^{ème} ! Un merci à Cyp et Edouard pour votre aide avec les masto (oui oui les meilleures cellules si vous voulez). Marie V. tu auras toujours été l'oreille attentive aux moments où j'en avais besoin, merci beaucoup d'avoir été si présente.

Ma Thildou, ma copine de puis tant de temps. Merci pour toutes ces années, ta gentillesse, ton sourire et ton rire qui me rendent toujours heureuse. Même si on ne se voit plus autant, à chaque fois qu'on se retrouve c'est toujours autant de joie !

Mes vieux copains, Loïse, Flo, Inès et Marie merci de m'avoir fait penser à autre chose qu'à ma thèse. A tous nos moments passés ensemble depuis toujours ! Merci aussi à Michmich, Zabou, Franck, Nath, Michou, Didier, Pierre et Sophie, pour avoir toujours été là et m'avoir soutenue pendant quelques jours durant mon écriture !

Mes mamies, mon papi, mes taties et tontons chéris, merci d'être présents dans ma vie et de m'avoir accueillie pour un (enfin plusieurs) bons repas durant ma thèse, toujours un sourire au visage et une bouteille à la main ! Tous mes cousins et les « pièces rapportées », merci pour vos bêtises sur les conversations de familles et pour m'avoir soutenue jusqu'au bout. Mentions spéciales à Marine et Allan pour les week-ends long au ski et à la mer c'était toujours dépaysant, à mes trois petites puces pour les week-ends parc d'attraction et tout le reste, à Sosso pour les (très) nombreuses

heures au téléphone, les week-ends ski ou à Toulouse et les restos !

Unaï, Aïna, Mattin et Robin, vous ne le savez pas mais vous m'avez bien fait rire au travers des photos/vidéos que vos mamans respectives ont envoyées.

Papa et Maman comment dire suffisamment merci pour tout ce que vous avez fait. Merci pour votre soutien au cours de toutes ces années d'études, d'avoir cru en moi même quand moi je n'y croyais pas, et de m'avoir obligé à toujours me dépasser ! Merci aussi de m'avoir nourri et logé pendant l'écriture de cette thèse c'était parfait. Même si je sais que vous n'allez pas tout comprendre, cette thèse est aussi un peu la vôtre.

Le meilleur (quand t'es tout seul) pour la fin, mon frerot préféré ! Merci d'avoir relu cette thèse même sans rien comprendre à la science, ça m'aura bien aidé ! Merci aussi et surtout de m'envoyer des blagues et des trucs plus sérieux, tu m'auras appris à toujours rester attentive à ce qu'il se passe dans le monde. Merci d'être là tout simplement (sauf quand t'es relou).

Table des matières

Remerciements.....	3
Table des matières	7
Index Figures.....	10
Index des tableaux.....	12
Abréviations	13
Partie 1 : Introduction	17
A. Les apports des « OMICS » dans la compréhension des cancers	18
I. La génomique : déchiffrer les génomes tumoraux	20
1. Comprendre les génomes cancéreux	20
a. Mutation au niveau nucléotidique.....	20
b. Altération de structure.....	21
2. Analyses pan-cancers des génomes.....	24
a. Des profils d'altérations différents selon les cancers.....	24
b. Les signatures génomiques	25
II. La transcriptomique : appréhender l'oncogenèse par l'expression des gènes.....	26
1. Détection de l'expression des mutations génomiques.....	26
a. Les altérations des transcrits et leurs conséquences.....	26
b. Cas particulier du chromosome X.....	27
2. Les signatures transcriptomiques	28
3. Déceler les cellules immunitaires au sein des tumeurs	30
III. Pour une compréhension plus globale : l'analyse de l'épigénome et des ARN non codants	33
1. Le méthylome	33
2. Les micro ARN.....	35
B. Les sarcomes : un groupe de cancers hétérogènes	38
I. Des histologies très diversifiées	38
II. La génomique hétérogène des sarcomes	39
1. Des <i>drivers</i> forts et leurs conséquences.....	39
a. Qu'est-ce qu'un <i>driver</i> fort ?	39
b. Activation d'un oncogène	40
c. Création de protéines chimériques	41
d. Un transcriptome spécifique dépendant de chaque <i>driver</i> fort	43
2. Des génomes complexes sans <i>driver</i> connu	44

a.	Hétérogénéité des génomes et des transcriptomes.....	44
b.	Instabilité des génomes.....	46
c.	Des altérations communes malgré le chaos.....	49
III.	Prises en charge thérapeutiques variées des sarcomes.....	52
1.	La chirurgie.....	53
2.	La chimiothérapie.....	54
3.	La radiothérapie.....	54
4.	Les thérapies ciblées.....	55
a.	Les inhibiteurs.....	55
b.	L'immunothérapie.....	55
C.	Zoom : les léiomyosarcomes.....	58
I.	Descriptions anatomopathologique et épidémiologique.....	58
II.	Des génomes complexes avec des altérations communes.....	59
1.	Des génomes très remaniés.....	59
2.	La perte de PTEN.....	60
3.	La perte de DMD.....	61
4.	L'amplification de <i>MYOCD</i>	64
5.	Transcriptomes des léiomyosarcomes.....	65
III.	Physiologie de la cellule musculaire lisse.....	67
1.	Les muscles lisses et les cellules musculaires lisses.....	67
2.	Des origines multiples.....	68
a.	Le mésoderme.....	68
b.	L'ectoderme.....	70
3.	Spécificités des différentes cellules musculaires lisses.....	70
4.	Plasticité de la différenciation.....	71
a.	Capacité de dédifférenciation.....	72
b.	Des capacités plastiques au-delà du phénotype musculaire lisse	74
5.	Les voies de régulation de la différenciation.....	76
a.	La voie de la différenciation SRF/ <i>MYOCD</i>	76
b.	La dédifférenciation : l'inhibition de SRF/ <i>MYOCD</i>	77
c.	Les miRNA de la régulation du <i>switch</i> phénotypique.....	78
D.	Objectifs.....	80
	Partie 2 : Résultats.....	81
A.	Comprendre l'oncogenèse des LMS grâce aux « OMICS ».....	82
I.	Contexte.....	82
II.	Article 1 : <i>SRF-MYOCD axis is the targetable driver of a well differentiated and aggressive subtype of Leiomyosarcomas</i>	83
III.	Conclusions.....	109

B.	ATRX : son altération est-elle un <i>driver</i> ?	111
I.	Contexte	111
II.	Article 2 : <i>ATRX alteration contributes to tumor growth and immune escape in pleomorphic sarcomas</i>	112
III.	Conclusions	144
C.	Comment la sur-activation de la voie SRF/MYOCD est-elle le <i>driver</i> des hLMS ?	146
I.	Contexte	146
II.	Résultats.....	147
1.	Validation des résultats protéiques précédents dans l'ensemble de la cohorte de LMS.....	147
2.	Analyse de l'effet de la suractivation de SRF/MYOCD dans les hLMS 149	
a.	Stratégie d'analyse.....	149
b.	Recherche d'anticorps pour MYOCD.....	149
c.	Mise au point d'un TAG de MYOCD par CRISPR.....	151
d.	Mise au point de ChIP dans des lignées de sarcomes.....	155
e.	Comparaison des profils qPCR de ChIP par rapport aux SMC 157	
3.	Cibler la voie SRF/MYOCD	158
III.	Conclusions	163
Partie 3 : Discussion et Perspectives		165
A.	LMS et sarcomes pléomorphes : le remaniement du génome comme <i>driver</i> ? 166	
I.	Des altérations permissives de modifications génomiques importantes	166
II.	Sélection des clones ayant les altérations les plus oncogéniques.....	167
1.	oLMS, LMS utérins et sarcomes pléomorphes : la perte d'ATRX, un <i>driver</i> important	168
2.	hLMS : la sur-activation de SRF/MYOCD comme <i>driver</i> principal.....	170
B.	Comment la différenciation de la cellule d'origine détermine-t-elle l'oncogenèse ?	174
C.	Doit-on proposer des thérapies spécifiques aux altérations des tumeurs ?	179
I.	Vers une utilisation clinique des statuts h/oLMS.....	179
II.	Immunothérapie lors de la perte d'ATRX	180
III.	Une thérapie prometteuse des hLMS : inhibition de la voie SRF/MYOCD 181	
D.	Perspectives.....	183
Partie 4 : Matériels et Méthodes		188
	Bibliographie.....	198

Index Figures

Figure 1 : Les 8 capacités des cancers et leurs 2 caractéristiques leur permettant d’y accéder (Hanahan and Weinberg, 2011)	18
Figure 2 : Représentation des différents types de variants structuraux (Inaki and Liu, 2012).....	22
Figure 3 : Le chromothripsis (Marcozzi <i>et al.</i> , 2018)	23
Figure 4 : Nombre médian de mutations somatiques non-synonymes par rapport au nombre médian de variants structuraux déséquilibrés (TCGA, 2017)	24
Figure 5 : Les fonctions nucléaires des micro ARN (adapté de Catalanotto <i>et al.</i> , 2016)	36
Figure 6 : Taxonomie des sarcomes (Taylor <i>et al.</i> , 2011)	38
Figure 7 : Les 4 altérations chromosomiques pouvant amener à la formation de gènes de fusion (Annala <i>et al.</i> , 2013)	41
Figure 8 : Représentation des principales translocations dans les sarcomes à <i>drivers</i> forts les plus fréquents (Nakano and Takahashi, 2018)	42
Figure 9 : Classification non supervisée de 184 sarcomes à petites cellules rondes selon l’expression de leurs gènes (Watson <i>et al.</i> , 2018).....	43
Figure 10 : Nombre de mutations somatiques non-synonymes par rapport au nombre d’altérations du nombre de copies (TCGA, 2017).....	45
Figure 11 : Analyse non supervisée de l’expression des ARN messagers de 206 sarcomes à génomique complexe (TCGA, 2017).....	46
Figure 12 : Modèles théoriques du mécanisme ALT (Amorim <i>et al.</i> , 2016)	48
Figure 13 : Insertion d’ADN télomérique par le système ALT amenant une instabilité chromosomique (Marzec <i>et al.</i> , 2015)	49
Figure 14 : La perte d’ATRX entraîne l’activation du mécanisme ALT (Strum <i>et al.</i> , 2014).....	51
Figure 15 : Gestion des sarcomes des tissus mous localisés (Casali <i>et al.</i> , 2018c).....	53
Figure 16 : Léiomyosarcome différencié (Robin <i>et al.</i> , 2013)	59
Figure 17 : Fréquence de la délétion profonde de <i>PTEN</i> dans différents types de sarcome (Stefano and Giovanni, 2019)	60
Figure 18 : Représentation du nombre de copies du chromosome X (Wang <i>et al.</i> , 2014)	62
Figure 19 : Complexe Dystrophine/Glycoprotéine (Gao et McNally, 2015)	63
Figure 20 : Amplifications récurrentes dans les 80 LMS de la cohorte du TCGA (TCGA, 2017).....	64

Figure 21 : Regroupement non supervisé de l'expression des gènes venant de différents types de sarcome à génomique complexe (Gibault <i>et al.</i> , 2011).....	66
Figure 22 : Contraction musculaire des cellules musculaires lisses (Heyden <i>et al.</i> , 2013)	68
Figure 23 : Origine cellulaire des différentes parties du système vasculaire (Wang <i>et al.</i> , 2015).....	69
Figure 24 : Ultrastructure des SMC contractiles et synthétiques (Rensen <i>et al.</i> ,2007) .	73
Figure 25 : Trans-différenciation des SMC dans la plaque artériosclérotique (Liu and Gomez, 2019).....	75
Figure 26 : Représentation du complexe MYOCD/SRF/PTEN dans les SMC contractiles et lors du <i>switch</i> phénotypique (adapté de Horita <i>et al.</i> , 2016)	77
Figure 27 : Rôle des miRNA dans la régulation de la prolifération des SMC vasculaires (Wang and Atanasov, 2018)	79
Figure 28 : Caractérisation de la voie SRF/MYOCD dans l'ensemble de la cohorte de LMS.....	148
Figure 29 : Principe d'immunoprécipitation de la chromatine (ChIP)	149
Figure 30 : Recherche d'anticorps spécifiques de MYOCD.....	150
Figure 31 : Représentation schématique de la technique d'étiquetage de MYOCD endogène.....	151
Figure 32 : Représentation schématique d'une insertion dans une mauvaise cible ..	152
Figure 33 : PCR, Western blot et immunofluorescence pour la validation de l'étiquetage de MYOCD	154
Figure 34 : Mise au point de la sonication pour les ChIP	155
Figure 35 : qPCR de validation des ChIP testant les anticorps dirigés contre SRF....	156
Figure 36 : Profils qPCR des ChIP sur l'ensemble des lignées cellulaires modèles...	158
Figure 37 : Prolifération cellulaire selon une gamme de CCG-100602	160
Figure 38 : Pourcentage de cellules caspase 3 positives selon une gamme de CCG-100602.....	161
Figure 39 : qPCR de validation des ChIP avec un anti-SRF sur les lignées après traitement à l'IC50 du CCG-100602 ou au DMSO	162
Figure 40 : Modèle de travail des voies de signalisation à l'origine de l'oncogenèse des hLMS	173
Figure 41 : Modèle d'oncogenèse des hLMS et des oLMS	178
Figure 42 : Vecteurs et séquences ayant servis à la préparation de l'étiquetage de MYOCD	193

Index des tableaux

Tableau 1 : Conséquences fonctionnelles des mutations de structures (Inaki and Liu, 2012).....	27
Tableau 3 : Sélection d'études complètes utilisant de l'immunothérapie dans les sarcomes des tissus mous (Ayodele and Abdul Razak, 2020)	56
Tableau 4 : Séquences des primers pour les PCR de validation de l'insertion du Tag-FLAG et des séquences de sélections	194
Tableau 5 : Séquences des <i>primers</i> utilisés pour les qPCR de validation des CHIP....	196

Abréviations

~A~

ACTA2 : *alpha actin 2*
ACTB : *beta actin*
AKT : *protein kinase B*
ALT : *Allongement alternatif des télomères*
ASK1 : *Mitogen-Activated Protein Kinase Kinase Kinase 5*
ATRX : *Alpha Thalassemia/Mental Retardation Syndrome X-Linked*
 α -SM actine : *Smooth muscle alpha actin*

~B~

BAFFR : *Tumor Necrosis Factor Receptor Superfamily Member 13C*
BRAF : *B-Raf Proto-Oncogene, Serine/Threonine Kinase*

~C~

CALD1 : *Caldesmon 1*
CCL24 : *C-C Motif Chemokine Ligand 24*
CCL3 : *C-C Motif Chemokine Ligand 3*
CCL5 : *C-C Motif Chemokine Ligand 5*
CCR3 : *C-C Motif Chemokine Receptor 3*
CCR5 : *C-C Motif Chemokine Receptor 5*
CD4 : *T-Cell Surface Glycoprotein CD4*
CD8 : *T-Cell Surface Glycoprotein CD8*
CD19 : *Differentiation Antigen CD19*
CD20 : *Membrane Spanning 4-Domains A1*
CD68 : *Macrophage Antigen CD68*
CD79 : *B-Cell Antigen Receptor Complex-Associated Protein*
CD117 : *KIT Proto-Oncogene, Receptor Tyrosine Kinase*
CDK4 : *Cyclin Dependent Kinase 4*
CDKN2A : *Cyclin Dependent Kinase Inhibitor 2A*
CENP-C : *Centromere Protein C*
ChIP : *Immunoprécipitation de la chromatine*
Chr22 : *Chromosome 22*
CINSARC : *Complexity INdex in SARComas*
CNN1 : *Calponin 1*
COL1A1 : *Collagen Type I Alpha 1 Chain*
COL1A1/2 : *Collagen Type I Alpha 1/2 Chain*
Cox2 : *Prostaglandine synthase 2*
CRISPR : *Clustered regularly interspaced short palindromic repeats*
CXCL1 : *C-X-C Motif Chemokine Ligand 1*
CXCL3 : *C-X-C Motif Chemokine Ligand 3*
CXCL8 : *C-X-C Motif Chemokine Ligand 8*
CXCL10 : *C-X-C Motif Chemokine Ligand 10*

~D~

DAXX : *Death Domain Associated Protein*
DDLPS : *Dedifferentiated Liposarcoma*
DES : *Desmin*
DMD : *Dystrophin*
DNMT : *DNA Methyltransferase 1*

~E~

E2F1 : *E2F Transcription Factor 1*
Elk-1 : *ETS Transcription Factor*
ENCODE : *Encyclopedia of DNA Elements*
ERK : *Mitogen-Activated Protein Kinase 1*
ESMO : *European Society for Medical Oncology*
EWSR1 : *Ewing Sarcoma Breakpoint Region 1*

~F~

FCGR3 : *Fc Fragment Of IgG Receptor IIa*
FceRI : *Fc Fragment Of IgE Receptor Ia*
FGF2 : *Fibroblast Growth Factor 2*
FLI1 : *Friend Leukemia Integration 1 Transcription Factor*
FN1 : *Fibronectin 1*
FNCLCC : *Fédération nationale des centres de lutte contre le cancer*
FOLFOX : *Folinic acid, Fluorouracil, and Oxaliplatin*
FOXO : *Forkhead Box O*
FoxO3a : *Forkhead Box O3a*
FoxO4 : *Forkhead Box O4*

~G~

GIST : *Gastrointestinal stromal tumor*

~H~

H3K36me3 : *Triméthylation de la lysine 36 de l'histone 3*
H3K4dMe : *Diméthylation de la lysine 4 de l'histone 3*
H3K9me3 : *Triméthylation de la lysine 9 de l'histone 3*
H4Ac : *Acétylation de l'histone 4*
H4K20me3 : *Triméthylation de la lysine 20 de l'histone 4*
HIF : *Hypoxia Inducible Factor*
hLMS : *Homogenous leiomyosarcoma*
HNMT : *Histamine N-Methyltransferase*
HOXB13 : *Homeobox B13*
hTERT : *Human Telomerase Reverse Transcriptase*

~I~

ICGC : *International Cancer Genome Consortium*
IFN- γ : *Interferon gamma*
IgE : *Immunoglobulin E*
IGF-1 : *Insulin-like growth factor 1*
IgG : *Immunoglobulin G*
IgH : *Immunoglobulin H*
IHC : *Immunohistochimie*
IL1 : *Interleukin 1*
IL4 : *Interleukin 4*
IL6 : *Interleukin 6*
IL8 : *Interleukin 8*
IL9 : *Interleukin 9*
IL10 : *Interleukin 10*
IL1R : *Interleukin-1 receptor*
IL17BR : *Interleukin 17 Receptor B*
IL-5R α : *Interleukin 5 Receptor Subunit Alpha*
iNOS : *Nitric oxide synthases calcium-insensitive*

~J~

JUN : *Jun Proto-Oncogene, AP-1 Transcription Factor Subunit*

~K~

kb : *kilo base*
KIT : *Receptor Tyrosine Kinase*
KLF4 : *Kruppel Like Factor 4*
KLF1 : *Kruppel Like Factor receptor &*
KRAB : *Krüppel associated box*
KRAS : *Kirsten Rat Sarcoma Viral Oncogene Homolog*

~L~

LEF1 : *Lymphoid Enhancer Binding Factor 1*
LGALS3 : *Galectin 3*
LMS : *Léiomyosarcomes*
LTB4 : *Leukotriene B4*

~M~

MAPK : *Mitogen-Activated Protein Kinase 1*
Maspin : *Serpin Family B Member 5*
MCL1 : *MCL1 Apoptosis Regulator, BCL2 Family Member*
MCP3 : ou CCL7 : *C-C Motif Chemokine Ligand 7*
MCP4 : ou CCL13 : *C-C Motif Chemokine Ligand 13*
MDM2 : *Double Minute 2*
MEF2C : *Myocyte Enhancer Factor 2C*
MEK : *Mitogen-Activated Protein Kinase Kinase 1*
MFAP4 : *Micofibril Associated Protein 4*
MFS : *Mixofibrosarcomes*
MITF : *Melanocyte Inducing Transcription Factor*
MMP2 : *Matrix Metalloproteinase 2*

MMP9 : *Matrix Metalloproteinase 9*
MRTF-A : *Myocardin Related Transcription Factor A*
Ms4a2 : *Membrane Spanning 4-Domains A2*
MSN : *Moesin*
mTOR : *Mechanistic Target Of Rapamycin Kinase*
MYC : *MYC Proto-Oncogene, BHLH Transcription Factor*
MYF5 : *Myogenic Factor 5*
MYF6 : *Myogenic Factor 6*
MYH11 : *Myosin Heavy Chain 11*
MYL6 : *Myosin Light Chain 6*
MYOCD : *Myocardin*

~N~

NCF4 : *Neutrophil Cytosolic Factor 4*
NF1 : *Neurofibromin 1*
NF κ B : *Nuclear Factor Kappa B*
NK : *Natural killer*
Nm : *nanomètre*
NR2C/F : *Nuclear Receptor Subfamily 2 Group C Member 2/Group F Member 2*
NRAS : *Transforming Protein N-Ras*
NY-ESO-1 : *Cancer/Testis Antigen 1B*

~O~

oLMS : *other léiomyosarcoma*
OMICS : *Différentes disciplines de biologie comprenant : la génomique, la transcriptomique, la protéomique...*
OS : *Ostéosarcomes*

~P~

P14 : ou CDKN2A : *Cyclin Dependent Kinase Inhibitor 2A*
p53 : *Tumor Protein P53*
PAX5 : *Paired Box 5*
pb : *paires de bases*
PBS : *Phosphate-buffered saline*
PCR : *Polymerase chain reaction*
PD-1 : *Programmed Cell Death 1*
PDGF : *Platelet Derived Growth Factor*
PDGFB : *Platelet Derived Growth Factor Subunit B*
PDGFR α : *Platelet Derived Growth Factor Receptor Alpha*
PDGFR β : *Platelet Derived Growth Factor Receptor Beta*
PDX : *Patient-derived xenograft*
PI3K : *Phosphatidylinositol-4,5-Bisphosphate 3-Kinase*
Pim-1 : *Pim-1 Proto-Oncogene, Serine/Threonine Kinase*
PIP3 : *Phosphatidylinositol (3,4,5)-trisphosphate*

PoIII : RNA polymerase II
PPP2CA : Protein Phosphatase 2 Catalytic Subunit Alpha
pRb : phoso-retinoblastoma
PTEN : Phosphatase And Tensin Homolog

~Q~

qPCR : Quantitative Polymerase chain reaction

~R~

Rad51 : RAD51 Recombinase
RAS : Proto-Oncogene
Rb : Retinoblastoma
RB1 : Retinoblastoma 1
REO : Relative expression ordering
RICTOR : Rapamycin-Insensitive Companion Of mTOR
RIPA: Radio-immunoprécipitation
RISC : RNA-induced silencing complex
RMS : Rhabdomyosarcomes

~S~

SDHB : Succinate Dehydrogenase Complex Iron Sulfur Subunit B
SDHC : Succinate Dehydrogenase Complex Subunit C
SCF : Stem cell Factor
SM22 α : Transgelin
SMAD2 : Small worm phenotype/ Mothers Against Decapentaplegic Family Member 2
SMAD4 : Small worm phenotype/ Mothers Against Decapentaplegic Family Member 4
SMC : Smooth Muscle Cell ou Cellule musculaire lisse
SM-MHC : Smooth muscle myosin heavy chain
SRF : Serum Response Factor
SS18 : SS18 Subunit Of BAF Chromatin

Remodeling Complex
SSX : Synovial Sarcoma, X Breakpoint
ST2 : Suppression Of Tumorigenicity 2
SV40 : virus simien 40
SWI/SNF : SWItch/Sucrose Non-Fermentable

~T~

TAGLN : Transgelin
TCF : Transcription Factor
TCGA : The Cancer Genome Atlas
TCRA : T Cell Receptor Alpha Locus
TGF- β : Transforming Growth Factor Beta
TMA : Tissue Microarray
TNC : Tenascin C
TNF : Tumor Necrosis Factor
TP53 : Tumor Protein P53
Treg : Regulatory T cells

~U~

UPS : Sarcomes pléomorphiques indifférenciées ou undifferentiated pleomorphic sarcoma

~V~

VEGF : Vascular Endothelial Growth Factor
VNN2 : Vanin 2

~X~

Xist : X-inactive specific transcript

~Z~

ZNF : Zinc Finger Protein

~ α ~

α -SM actine : Smooth muscle α -actin

Partie 1 : Introduction

A. Les apports des « OMICS » dans la compréhension des cancers

En 2000, Hanahan et Weinberg proposent six capacités que chaque cellule précancéreuse doit acquérir afin de pouvoir se transformer en cancer capable de produire des métastases ; ils ont par la suite, en 2011, ajouté deux autres capacités ainsi que deux caractéristiques permettant d'accéder à ces capacités (Hanahan and Weinberg, 2011). Ces capacités sont l'augmentation de la prolifération, la résistance à la mort cellulaire, l'immortalité répliquative, l'activation de l'invasion, l'induction de l'angiogenèse, la dérégulation du métabolisme, l'échappement aux suppresseurs de croissance et l'échappement au système immunitaire (**Figure 1**). Les deux caractéristiques d'accession à ces capacités sont d'une part l'inflammation tumorale qui aide la tumeur à proliférer et à survivre en apportant, par exemple, des facteurs de croissance (Hanahan and Weinberg, 2011). Et d'autre part, l'acquisition d'instabilité génomique et de mutations qui permettent l'acquisition de la plupart des capacités décrites. Bien que les altérations du génome soient très diversifiées entre les différents cancers, le nombre important de défaut de la maintenance et de la réparation du génome que l'on retrouve dans quasiment toutes les tumeurs humaines ainsi que le haut niveau de déstabilisation du nombre de copies dans ces tumeurs indiquent que l'instabilité du génome est inhérente à une grande majorité des cellules cancéreuses (Hanahan and Weinberg, 2011).

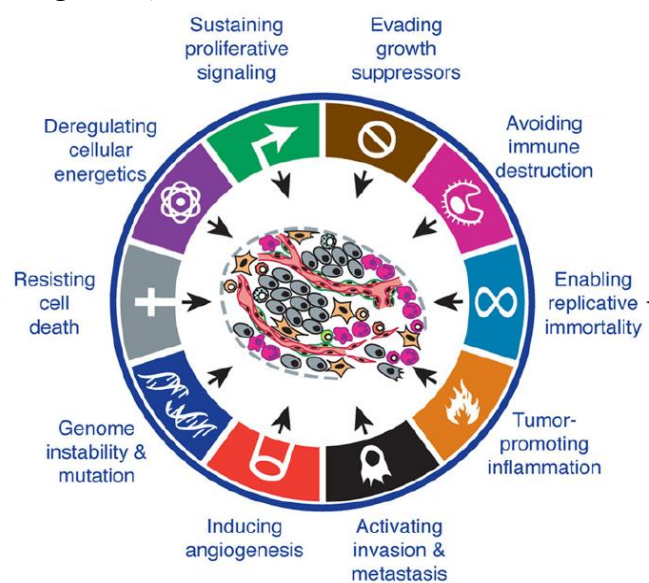


Figure 1 : Les 8 capacités des cancers et leurs 2 caractéristiques leur permettant d'y accéder (Hanahan and Weinberg, 2011)

La connaissance des génomes tumoraux et de leurs modifications que ce soit de structure, de nombre, ou leurs mutations est devenue nécessaire à la compréhension de la biologie de chaque type de cancer. Deux techniques de séquençage de l'ADN ont été développées en 1977 par Gilbert (Maxam and Gilbert, 1977) et Sanger (Sanger, Nicklen and Coulson, 1977) mais ne permettaient de séquencer que 250 nucléotides pour la méthode de Gilbert, qui par la suite n'a quasiment plus été utilisée, et jusqu'à 200 nucléotides pour celle de Sanger. L'amélioration de la technique de Sanger et l'avènement de la bio-informatique ont permis le séquençage complet du génome humain (Lander *et al.*, 2001; Venter *et al.*, 2001). Depuis, les techniques de séquençage *next generation* à haut débit ont permis, depuis 2005, un séquençage massif et à faible coût de n'importe quel génome.

De grands consortiums internationaux se sont formés dans l'optique d'appliquer ces séquençages aux tumeurs. On retrouve les programmes du TCGA pour *The Cancer Genome Atlas* et de l'ICGC pour *International Cancer Genome Consortium*. Le TCGA a démarré en 2006 et avait pour ambition de séquencer 33 types de cancers au niveau génomique, épigénomique, transcriptomique et protéomique (*The Cancer Genome Atlas Program - National Cancer Institute, 2018*). L'ICGC a débuté un an plus tard et avait pour but de réaliser une étude génomique, transcriptomique et épidémiologique de 250 000 tumeurs venant de 50 types de cancer afin de trouver des cibles thérapeutiques potentielles (*International Cancer Genome Consortium, 2020*). Dans les deux cas, ces techniques à grande échelle ont permis de mettre à jour des points communs avec des gènes altérés dans de nombreux types de cancers dont les deux plus fréquents sont *TP53* (altéré dans 25 types de cancer sur 26 étudiés) et *CDKN2A* (altéré dans 20 types de cancer sur 26) (*The ICGC/TCGA Pan-Cancer, 2020*). Ces techniques ont aussi révélé des altérations spécifiques à certains types de cancer telle que l'altération d'*ERG* dans 50% des adénocarcinomes de la prostate ou encore celle de *VHL* dans 60% des cancers du rein à cellules claires (*The ICGC/TCGA Pan-Cancer, 2020*).

I. La génomique : déchiffrer les génomes tumoraux

1. Comprendre les génomes cancéreux

Le séquençage de génomes tumoraux est d'abord comparé avec le séquençage de l'ADN normal du patient ou, si ce dernier n'est pas disponible ou faisable, avec une base de données d'ADN normaux comme celle du projet 1000 génomes. La soustraction du génome normal à celui du génome tumoral permet de mettre en évidence l'ensemble des altérations génétiques et génomiques somatiques, de chaque tumeur.

a. Mutation au niveau nucléotidique

Les mutations au niveau nucléotidique peuvent être de deux types. Soit une substitution d'une paire de bases, soit une insertion ou une délétion d'une ou plusieurs bases.

Une substitution est un échange d'une paire de bases pour une autre. Cela peut avoir lieu n'importe où dans le génome. Dans un exon, cet échange modifie un codon pouvant entraîner différentes sortes de modifications dans la séquence protéique : pas de modification d'acide aminé (mutation synonyme), la modification d'un acide aminé (mutation faux-sens) ou l'apparition d'un codon stop (mutation non-sens). Ces deux dernières altérations auront, dans la majorité des cas, un impact direct sur la protéine entraînant un gain ou une perte de fonction. L'apparition d'une substitution dans une région non-codante peut avoir un effet tout aussi important pour la cellule puisqu'une modification dans une région promotrice, dans un site d'épissage ou dans un site impliqué dans la stabilité de l'ARN messager peut conduire à la transformation tumorale (Diederichs *et al.*, 2016).

Les insertions et délétions d'une ou plusieurs paire(s) de bases, si elles ne sont pas un multiple de trois et sont dans une région transcrite, entraînent un décalage dans le cadre de lecture du gène amenant à une protéine tronquée ou inactive. Lorsque l'insertion ou la délétion concerne un multiple de trois dans une région transcrite, cela aura pour conséquence la suppression ou l'ajout d'acide(s) aminé(s) dans la séquence

sauf si l'insertion est un codon stop, dans ce cas la protéine sera tronquée. En dehors des régions codantes une telle mutation n'aura soit aucun effet soit les mêmes effets qu'une substitution dans une région non codante.

Ces mutations sont les conséquences de plusieurs mécanismes endogènes ou exogènes. Les causes endogènes comprennent les erreurs de réplication, les espèces réactives de l'oxygène, les aldéhydes ou encore la déamination des nucléotides. Les causes exogènes proviennent de sources chimiques ou physiques comme les produits chimiques, les radiations ionisantes ou les ultra-violets (Martincorena and Campbell, 2015). La localisation chromosomique et le type de mutation associé sont influencés par la topologie de l'ADN, par le niveau de la transcription et la réplication (Maki, 2002). En effet, le taux de mutation est variable selon l'état de la chromatine, il sera alors plus élevé dans les régions à hétérochromatine par rapport aux régions d'euchromatine (Schuster-Böckler and Lehner, 2012). Cet effet est d'autant plus important sur le chromosome X inactif, composé quasiment que d'hétérochromatine et qui est hyper-muté dans différents cancers (Jäger *et al.*, 2013).

b. Altération de structure

Les premières altérations de structure ont été découvertes par l'analyse de la structure des chromosomes en métaphase avec des gains ou des pertes de chromosomes et des translocations entre 2 chromosomes comme la première translocation décrite dans les cancers : la translocation de Philadelphie dans la leucémie myéloïde chronique (Benson, 1961). Mais les tailles des anomalies devaient être suffisantes pour être vue au microscope. Le séquençage des génomes permet une analyse plus fine de ces altérations qui peuvent être intra ou inter chromosomiques.

Il existe 5 types d'altération structurale retrouvés dans les génomes des cancers (**Figure 2** ; Inaki and Liu, 2012) :

- Les délétions : perte d'une partie d'un chromosome avec les deux extrémités de la perte qui fusionnent.
- Les duplications en tandem : duplication d'un segment d'un

chromosome qui se retrouve en 5' ou en 3' du segment dupliqué.

- Les amplifications : duplication d'un segment chromosomique qui peut rester sur le même chromosome ou être distribué sur d'autres.
- Les inversions : un segment de chromosome retourné entre 2 points de cassure.
- Les translocations : échange d'un segment entre 2 chromosomes.

Comparées aux variations structurales des cellules normales dans lesquelles on retrouve plus de délétions, inversions et insertions, les cellules cancéreuses comportent plus des duplications en tandem, des inversions non appariées et des translocations (Hillmer *et al.*, 2011).

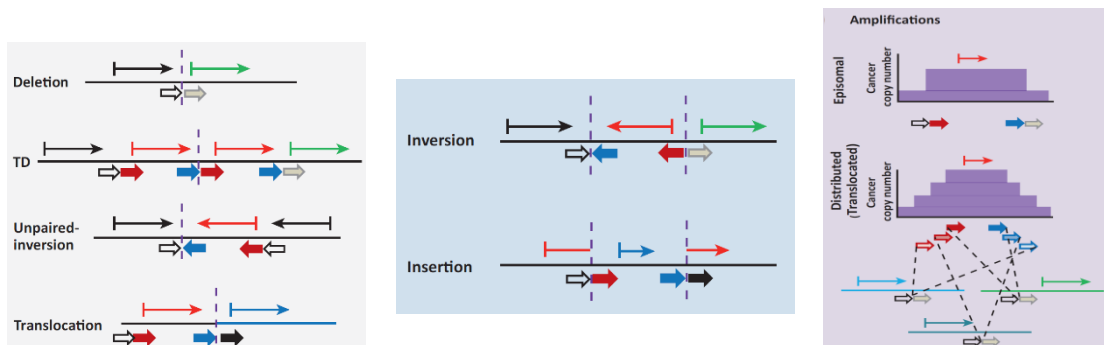


Figure 2 : Représentation des différents types de variants structuraux (Inaki and Liu, 2012)

Les variants structuraux ont de multiples conséquences dans une cellule cancéreuse. Cela peut créer directement un évènement oncogénique par la création d'un transcrit de fusion ou la délétion d'un gène suppresseur de tumeur (Inaki and Liu, 2012). Environ la moitié des réarrangements somatiques touche un gène (Hillmer *et al.*, 2011), ces réarrangements peuvent tronquer un gène, fusionner deux gènes ou réarranger les exons. Il peut aussi y avoir des changements dans les zones de régulation d'un gène avec des translocations qui déplacent un *enhancer* ou un *silencer* devant un oncogène (Mitelman, Johansson and Mertens, 2007). Il est aussi retrouvé des amplifications qui ont, majoritairement, pour conséquence la surexpression d'oncogènes clés (Beroukhim *et al.*, 2010). Le séquençage de l'ADN a aussi permis d'identifier les évènements à l'origine de réarrangements plus complexes qui rendaient jusqu'alors les chromosomes « de base » non reconnaissables. Un de ces

réarrangements est appelé le phénotype de duplication en tandem, composé par de nombreuses duplications en tandem à travers l'ensemble du génome (Hillmer *et al.*, 2011; Ng *et al.*, 2012). Le second type de réarrangement complexe est le chromothripsis. Il est caractérisé par des dizaines de réarrangements chromosomiques regroupés sur quelques chromosomes ou bras de chromosomes qui semblent venir d'un même évènement catastrophique (Marcozzi, Pellestor and Kloosterman, 2018). Cet évènement peut être lié soit l'internement dans un micronoyau d'un chromosome retardataire lors de la mitose soit à la fusion de télomères formant un pont chromosomique. Le résultat du réarrangement amène à la délétion de segments chromosomiques ainsi qu'à des segments inversés et des segments droits mais désordonnés (Figure 3 ; Marcozzi, Pellestor and Kloosterman, 2018).

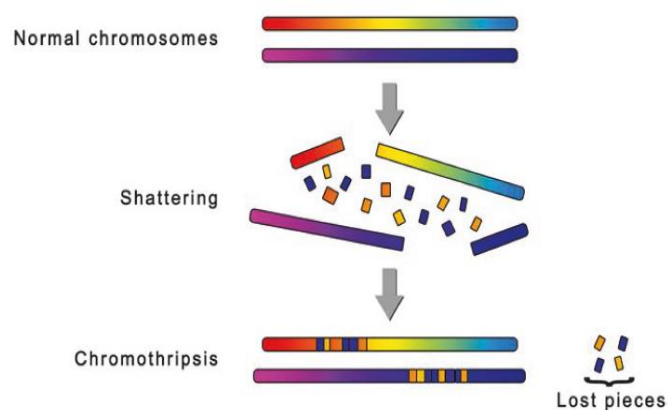


Figure 3 : Le chromothripsis (Marcozzi *et al.*, 2018)

aneuploïde amenant la possibilité que certaines combinaisons de pertes et de gains puissent être suffisantes pour initier l'oncogenèse (Davoli *et al.*, 2013).

Les réarrangements complexes tels que le chromothripsis sont retrouvés dans 22,3% des tumeurs et plus fréquemment dans les sarcomes, les glioblastomes, les carcinomes squameux du poumon, les mélanomes et les adénocarcinomes du sein (Cortés-Ciriano *et al.*, 2020). Les mutations *drivers* les plus souvent associées au chromothripsis sont celles de *TP53* (Rausch *et al.*, 2012). Le chromothripsis est un événement pouvant participer à l'oncogenèse puisque les régions comprises dans celles du chromothripsis coïncident avec 3,6% des mutations *drivers* identifiées, notamment des amplifications, des délétions homozygotes et des cassures dans un gène ou un promoteur (The ICGC/TCGA Pan-Cancer, 2020).

b. Les signatures génomiques

Les signatures mutationnelles correspondent à des mutations spécifiques dans des séquences déterminées dont l'origine peut être liée à un même facteur endogène ou exogène. La première caractérisation d'un motif de mutation est le passage d'une ou deux cytosine(s) à une ou deux thymine(s) par l'effet des UV (Howard and Tessman, 1964; Setlow and Carrier, 1966).

Grâce au séquençage à haut débit, 22 signatures mutationnelles ont été mises à jour (Alexandrov and Stratton, 2014), ce nombre a été augmenté à 30 après l'analyse de 23 000 tumeurs (Phillips, 2018). Chaque type de cancer comporte au moins 2 signatures mutationnelles actives et jusqu'à 6 pour le cancer du foie et de l'utérus (Alexandrov and Stratton, 2014). Pour 14 de ces signatures le facteur responsable n'est pas connu. Pour les autres signatures, il y a soit des facteurs exogènes tels que les UV, le tabac et les chimiothérapies alkylantes. Soit dans un tiers des cas, il y a des facteurs endogènes tels que l'âge, une déficience en réparation de l'ADN ou un défaut en ADN polymérase (Phillips, 2018). L'ensemble de ces signatures est disponible dans la base de données COSMIC (<https://cancer.sanger.ac.uk/cosmic>).

II. La transcriptomique : appréhender l'oncogenèse par l'expression des gènes

1. Détection de l'expression des mutations génomiques

La transcriptomique est l'étude de l'expression des gènes par le séquençage des ARN. Cela permet une étude de l'expression génomique qualitative, par l'analyse des mutations, et également quantitative, par l'analyse différentielle de l'expression des gènes.

a. Les altérations des transcrits et leurs conséquences

Une des conséquences possibles des modifications du génome est la modification d'ARN. Ces modifications peuvent soit impacter la quantité d'ARN transcrits soit les transcrits eux-mêmes. Dans la plupart des cas, la modification d'un transcrit altère le fonctionnement physiologique de la protéine codée.

Le déplacement d'un *enhancer* devant un gène ou une amplification de ce gène entraîne une surexpression de celui-ci comme la translocation dans les lymphomes de Burkitt qui associe l'*enhancer* de *IgH* à l'oncogène *MYC* (Mitelman, Johansson and Mertens, 2007). L'effet inverse existe avec la délétion totale d'un gène, d'un promoteur ou le déplacement d'un *silencer* devant un gène qui entraînera une sous-expression ou la suppression d'un transcrit (Inaki and Liu, 2012).

Certaines des altérations génomiques modifient directement le transcrit d'origine. Une translocation peut être à l'origine d'un transcrit de fusion fonctionnel. Deux morceaux de deux gènes se retrouvent donc à former un transcrit unique qui donnera une protéine chimérique généralement oncogène (**Tableau 1** ; Mitelman, Johansson and Mertens, 2007). Par exemple, la première translocation mise à jour, est le chromosome de Philadelphie dans la leucémie myéloïde chronique, cette translocation se produit entre les chromosomes 9 et 22 liant les gènes *BRC* et *ABL1* (Benson, 1961). Certains de ces gènes de fusion peuvent devenir la cible de thérapies ciblées. Bien que des fusions spécifiques entre deux gènes invariants arrivent, dans la plupart des cas un gène peut fusionner avec plusieurs autres gènes. De plus, à

l'exception de quelques gènes de fusion récurrents, la grande majorité est spécifique d'un type de tumeur (**Tableau 1** ; Stephens *et al.*, 2009; Totoki *et al.*, 2011; Delepaul *et al.*, 2017). Une délétion comportant une partie codante d'un gène amène à un transcrit tronqué et donc à une protéine tronquée. Cette nouvelle conformation peut soit inactiver la protéine soit lui conférer des nouvelles propriétés oncogéniques (**Tableau 1** ; Inaki and Liu, 2012). Par exemple, dans le cancer du sein, la présence d'altérations tronquantes de *BRCA2* entraîne une délocalisation de la protéine dans le cytoplasme qui ne pourra plus jouer son rôle physiologique dans le noyau et est à l'origine de l'oncogénèse (Spain *et al.*, 1999). Une mutation ponctuelle sur un transcrit d'ARN, selon le type, peut avoir un effet soit activateur soit inactivateur de fonction de la protéine. Par exemple, certaines mutations de *TP53* induisent un gain de fonction de la protéine qui devient capable de limiter l'apoptose et d'augmenter la migration (Oren and Rotter, 2010).

Break points	Types	Functional Consequences
Intergenic regions	Amplification (focal)	Overexpression
	Deletion (focal)	Silencing
		Activating target genes (miRNA)
Gene regulatory regions	Enhancer juxtaposition	Overexpression
	Silencer juxtaposition	Silencing
	Truncating promoter	Silencing
Intragenic	Gene truncation	Silencing and activating
	Fusion gene	Activating protein function and silencing
	Exon rearrangements	
	Fusion gene, untranslated region	Overexpression and silencing

Tableau 1 : Conséquences fonctionnelles des mutations de structure (Inaki and Liu, 2012)

b. Cas particulier du chromosome X

Contrairement aux autosomes, les altérations touchant le chromosome X auront des conséquences différentes selon le sexe de l'individu. En effet, chez un homme, l'X est systématiquement à l'état hémizygote (à l'exclusion des zones homologues du chromosome Y) hormis en cas de tétraploïdie, dans ce cas les deux copies du X s'exprimeront (Chaligné and Heard, 2014). En général, les altérations qui modifient les zones codantes du chromosome X ont des conséquences qui seront toujours exprimées dans les cellules d'un homme. Chez une femme, un seul des chromosomes X s'exprime du fait de l'inactivation d'un X dès l'embryogenèse par l'ARN non-codant *Xist*, bien

qu'environ 15% des gènes du X échappent à cette inactivation (Chaligné and Heard, 2014). Dans les cellules cancéreuses féminines, le X inactif est retrouvé hyper muté dans environ 30% des cas, ce qui serait le résultat d'un manque de réparation du X inactif durant l'oncogenèse (Jäger *et al.*, 2013). Les mutations somatiques du chromosome X d'une femme ne sont donc exprimées que lorsqu'elles apparaissent sur le chromosome X actif. Le séquençage des ARN messenger est donc nécessaire à l'analyse des altérations du chromosome X et de leurs conséquences.

2. Les signatures transcriptomiques

L'analyse de l'expression des gènes dans une tumeur peut faire apparaître des signatures d'expression spécifiques. Cela signifie que les changements d'expression d'un ou plusieurs gènes entre différentes tumeurs ou par rapport au tissu normal permettent de poser un diagnostic, de proposer un pronostic en déterminant la capacité de la tumeur à produire des métastases ou de prédire la réponse d'une tumeur à un type de traitement. L'une des premières signatures mise à jour est une signature au départ définie comme pronostic qui est au final plus diagnostique dans le cancer du sein et permettant de différencier les sous-groupes de ce cancer. Pour ce faire, les chercheurs ont réalisé une analyse non supervisée des gènes corrélant le plus avec la survie du patient et les plus différentiellement exprimés dans leur cohorte de 78 tumeurs du sein. Parmi les différents sous-groupes, il y a ceux exprimant les récepteurs des œstrogènes et les facteurs de transcription retrouvés dans l'épithélium luminal. Ainsi qu'un groupe présentant une surexpression des gènes présents dans l'amplicon du chromosome 17q22.24, contenant *ERBB2* et *GRB7* ; ce groupe est nommé ERBB2+. Ce gène, *ERBB2*, code pour HER2, un récepteur transmembranaire de facteur de croissance qui, lors de sa surexpression, induit un avantage prolifératif, des capacités anti-apoptotiques et de transformation oncogénique (Moasser, 2007). Enfin, le dernier groupe englobe les tumeurs exprimant des gènes du tissu adipeux et des cellules non épithéliales (Sørli *et al.*, 2001). En plus de permettre un diagnostic plus précis, cela aide à la compréhension de l'oncogenèse de ces différents sous-groupes. Séparer les différents groupes, durant les études, dans les années qui ont suivi, a permis de développer des traitements spécifiques de chaque type de tumeur du sein. En effet,

pour les patients du premier groupe une thérapie hormonale à base de Tamoxifen a été développée. Pour ceux du groupe ERBB2+, le développement de la thérapie ciblée à base d'anticorps monoclonaux ciblant HER2 puisque le taux de survie sans maladie à 5 ans est de 99%. Enfin pour le troisième groupe, le pronostic est plus défavorable car le taux de survie à 5 ans est de 85%. Seules les chimiothérapies créant des dommages à l'ADN sont utilisées au vu de la déficience en gène réparateur de l'ADN de ces tumeurs (Waks and Winer, 2019).

Depuis le début des années 2000, plusieurs signatures permettant de déterminer comment une tumeur réagira à un traitement donné ont été établies. Un des cancers qui a bénéficié de nombreuses études à ce sujet est le cancer du sein, pour lequel plusieurs signatures prédictives de la réponse des tumeurs traitées au Tamoxifen ont été développées utilisant soit un ratio entre deux gènes *HOXB13* et *IL17BR* soit l'expression d'un set de 36 gènes choisi pour sa capacité de classification correct des patient avec ou sans métastase (Ma *et al.*, 2004; Jansen *et al.*, 2007; Chanrion *et al.*, 2008). Cela permet aussi de comprendre que l'expression de *HOXB13* dans ce contexte stimule la migration et l'invasion (Ma *et al.*, 2004) et que l'expression de gènes associés à la réplication et à la réparation est de mauvais pronostic (Chanrion *et al.*, 2008). Mais elles se sont révélées être plus un marqueur pronostique de la récurrence des tumeurs une fois sous traitement que réellement prédictives de la réponse au Tamoxifen. Par la suite, une autre signature a été développée permettant de déterminer si un cancer du sein sera sensible à la radiothérapie. Dans cette étude, les 192 gènes formant la signature sont établis avec un différentiel d'expression entre des tumeurs répondantes ou non à la radiothérapie. Ces gènes sont liés à la régulation du cycle cellulaire, à la réplication et au métabolisme du glucose qui sont surexprimés dans les tumeurs non répondantes à la radiothérapie (Tanić *et al.*, 2018). Dans les autres cancers, il existe aussi de très nombreuses signatures dépendantes du type de tumeur et qui prédisent la réponse à un traitement donné. Dans la leucémie lymphoblastique aiguë, la résistance aux traitements à base de glucocorticoïdes est dépendante de l'expression de *MCL1*, un anti-apoptotique. Cette résistance peut être reversée avec de la rapamycine qui sous-régule *MCL1* et permet aux cellules de leucémie lymphoblastique aiguë de devenir sensible à l'apoptose induite par les glucocorticoïdes (Wei *et al.*, 2006). Pour le

cancer du côlon, il existe la signature REO permettant de déterminer la sensibilité des tumeurs à la chimiothérapie FOLFOX. Les cinq gènes de la signature sont en lien direct des gènes associés à la métabolisation de la chimiothérapie (He *et al.*, 2020).

D'autres signatures sont aussi venues enrichir les possibilités de déterminer le pronostic d'une tumeur. Une des plus connues est la PAM50 qui est une signature composée de 50 gènes prédisant le risque de rechute dans le cancer du sein. Cette signature a été déterminée par une analyse hiérarchisée selon quatre analyses transcriptomiques précédemment réalisées par différentes équipes. Les gènes de cette signature sont intrinsèquement associés aux différents groupes et à leur oncogénèse. Il y est présent, par exemple, les gènes *ERBB2* et *GRB7* amplifiés dans l'amplicon du groupe ERBB2+ (Parker *et al.*, 2009). Depuis, un grand nombre de signatures pronostiques ont été développées pour une majorité de cancers. Il a, notamment, été développé une signature CINSARC dans les sarcomes pouvant déterminer de manière significative la capacité d'un sarcome à développer des métastases. Cette signature est plus prédictive que le grade actuellement utilisé. Elle permet de séparer les tumeurs en deux catégories, à faible et haut risque métastatique, contre trois groupes pour le grade, dont un groupe intermédiaire présentant un avenir métastatique incertain (Chibon *et al.*, 2010). Cette signature est capable de classer correctement en faible ou haut risque les tumeurs de 39 types de cancer (Lesluyes *et al.*, 2017). Les 67 gènes de cette signature sont co-régulés et présentent des profils d'expression spécifiques de chacun des deux groupes, à faible ou haut risque métastatique. Ces gènes sont associés au contrôle de la mitose, à la biogénèse et à la ségrégation des chromosomes. L'analyse de l'expression de ces gènes et de la biologie des tumeurs avec un mauvais pronostic a permis d'associer les facteurs de transcription NFY et E2F ainsi que l'instabilité génétique à la surexpression de la signature (Chibon *et al.*, 2019). Cette signature en plus de son rôle dans le pronostic, permet de mieux comprendre les modifications transcriptomiques associées la formation de métastases.

3. Déceler les cellules immunitaires au sein des tumeurs

L'analyse du séquençage de l'ARN d'une tumeur permet de déceler quel microenvironnement est présent dans la tumeur et notamment quelles sont les

populations de cellules immunitaires présentes. Chaque type de cellule immunitaire exprime des marqueurs spécifiques, leur présence et leur quantité permettent donc de déterminer quelles populations sont présentes et d'en déduire leur quantité relative. Le microenvironnement tumoral peut donc être composé de quantités différentes de lymphocytes T CD4 et CD8, cellules B, éosinophiles, neutrophiles, *natural killer* (NK), macrophages et mastocytes (Thorsson *et al.*, 2018). L'expression des gènes de chacun des sous types, en plus de leurs marqueurs spécifiques, a été analysée en lignée puis à l'intérieur de tumeurs pour créer une méthode informatique appelée CIBERSORT qui permet de déterminer quelles populations sont présentes au sein des tumeurs ainsi que leur proportion (Newman *et al.*, 2015).

Les lymphocytes T sont reconnus principalement par l'expression de TCRA, le récepteur alpha des cellules T et de LEF1, un facteur de transcription de la différenciation des lymphocytes. Chaque sous-population de lymphocytes T exprime en plus des marqueurs spécifiques comme CD4 ou CD8 mais aussi la cytokine IL9 pour les lymphocytes mémoires (Abbas *et al.*, 2005). Dans une tumeur, les lymphocytes T sont largement reconnus comme protecteurs puisqu'une proportion haute de lymphocytes T dans les tumeurs est corrélée à un bon pronostic (Fridman *et al.*, 2017). Cependant, il est aussi souvent retrouvé, dans le microenvironnement, des cellules T régulatrices ou Treg dont le rôle physiologique est de limiter l'auto-immunité, rôle détourné dans les tumeurs vers un effet immunosuppresseur (Ohue and Nishikawa, 2019).

Les cellules B sont représentées par plusieurs marqueurs comme CD20, CD19 et BAFFR, le principal récepteur de la voie de signalisation qui maintient la survie des cellules B, PAX5, un facteur de transcription spécifique du lignage des cellules B et CD79 A et B, deux composants du récepteur antigénique des cellules B (Abbas *et al.*, 2005). Dans la plupart des cancers les cellules B infiltrées sont associées à un bon pronostic (Wouters and Nelson, 2018).

Les éosinophiles expriment un ensemble de marqueurs de surface tels que IL-5R α , ST2, CCR3 et CCR5 (Grisaru-Tal *et al.*, 2020) qui permettent de les reconnaître. Leur présence est liée à un bon pronostic dans certains cancers comme le cancer

colorectal, du sein et de la prostate mais aussi à un mauvais pronostic dans le cas du lymphome de Hodgkin (Sakkal, Miller and Nurgali, 2016). Parmi les effets anti-oncogéniques des éosinophiles, il y a leur capacité cytotoxique et de recrutement d'autres cellules immunitaires. Leur fonctionnalité pro-tumorale est liée à la sécrétion de facteurs de croissance et à l'activation de l'angiogenèse (Grisaru-Tal *et al.*, 2020).

Les neutrophiles ont des marqueurs bien établis : FCGR3, un marqueur membranaire ; NCF4, intervenant dans la mort intracellulaire des microorganismes internalisés ; VNN2, permettant la migration des neutrophiles (Abbas *et al.*, 2005). Les neutrophiles sont retrouvés en nombre important dans le sang des patients avec un cancer mais aussi infiltrant la tumeur (Shaul and Fridlender, 2019). Il y a une corrélation, dans les tumeurs solides, entre survie et la présence de neutrophiles circulants : plus il y a de neutrophiles circulants, plus la survie est augmentée dans la plupart des tumeurs solides (Templeton *et al.*, 2014). Les neutrophiles associés aux tumeurs ont des rôles différents selon le type de cancer et leur localisation dans celui-ci. Ils peuvent être anti-tumoraux avec une action cytotoxique directe ou pro-tumoraux en stimulant l'angiogenèse, la motilité cellulaire, la migration, l'invasion et en modulant les autres cellules immunitaires du *switch* immunosuppresseur (Shaul and Fridlender, 2019).

Le gène spécifique le plus exprimé par les *natural killers* est *KLFR1*, une lectine qui stimule la cytolysse (Abbas *et al.*, 2005). Les NK ont un rôle cytotoxique sur les cellules tumorales et un rôle important de recrutement des autres cellules immunitaires. Ces cellules peuvent aussi être à l'origine de l'échappement immunitaire lorsque les tumeurs empêchent leur reconnaissance en exprimant les molécules du complexe d'histocompatibilité de classe I (Choucair *et al.*, 2019).

Les macrophages expriment fortement MMP9, une métalloprotéase de la matrice, ainsi que plusieurs chimiokines, telles que CCL24, CXCL1 et CXCL3, dont l'expression sera perdue une fois le macrophage complètement différencié (Abbas *et al.*, 2005). Dans la grande majorité des cas, les macrophages sont pro-tumoraux dans les cancers solides. Ils permettent la croissance tumorale en produisant des facteurs pro-angiogéniques, mais surtout en supprimant la réponse immunitaire. Pour ce faire,

ils produisent des facteurs immunosuppresseurs comme IL-10, TGF- β et la prostaglandine E₂ pour recruter des cellules T régulatrices immunosuppressives (Ostuni *et al.*, 2015).

Les mastocytes sont caractérisés par l'expression à la membrane de CD117 (KIT) et Fc ϵ RI, un récepteur d'IgE. Ces cellules produisent aussi des composants spécifiques comme l'histamine, la tryptase, la sérotonine, la chymase et la carboxypeptidase A3 (Derakhshani *et al.*, 2019). Les mastocytes, comme les éosinophiles et les neutrophiles ont des rôles à la fois pro et anti-tumoraux. Ces cellules peuvent produire des molécules pro-angiogéniques et sont capables de sécréter des cytokines immunosuppressives qui activent les cellules T régulatrices (Derakhshani *et al.*, 2019). Les mastocytes recrutent aussi d'autres cellules immunitaires au sein des tumeurs en exprimant les cytokines CCL3, CCL5 et CXCL10 qui mobilisent les cellules T cytotoxiques (Ott *et al.*, 2003). Après interaction avec les cellules T, les mastocytes induisent la migration des neutrophiles par CXCL8 (Salamon *et al.*, 2005). L'effet des mastocytes sur une tumeur dépend du type de tumeur, de la localisation des mastocytes dans la tumeur et du grade de celle-ci (Komi and Redegeld, 2019).

III. Pour une compréhension plus globale : l'analyse de l'épigénome et des ARN non codants

Plus la compréhension de la régulation des gènes et du dysfonctionnement de cette régulation sera développée plus la connaissance de la biologie des cancers permettra de trouver des traitements plus adaptés. L'épigénome et les ARN non-codants ont un rôle primordial dans cette régulation que ce soit en pré ou en post-transcription.

1. Le méthylome

La méthylation du carbone 5 des cytosines constitue une importante part de la régulation épigénétique de la transcription. Bien que la méthylation des cytosines soit une modification stable de l'ADN puisqu'elle peut être héritée, c'est aussi un évènement dynamique qui peut changer au cours de la vie de certaines cellules et tissus et qui est influencé par l'environnement comme le régime alimentaire ou le

stress. Les motifs de méthylation sont mis en place par les méthyltransférases de l'ADN (DNMT) durant le développement précoce de l'embryon (Okano *et al.*, 1999). La méthylation est nécessaire pour avoir une différenciation correcte et pour définir les programmes transcriptomiques spécifiques des tissus (Pelizzola and Ecker, 2011). Le méthylome de différents tissus cellulaires humains normaux ou cancéreux est disponible sur différentes databases comme la base du Consortium international de l'épigénome humain, de l'encyclopédie des éléments de l'ADN ou la base NSGmethDB que l'on peut retrouver sur le site <https://epigenie.com/epigenetic-tools-and-databases/>.

Des changements épigénétiques sont présents dans l'ensemble des cancers et sont connus pour coopérer avec des altérations génomiques ou génétiques pour être à l'origine des phénotypes du cancer. Les modifications épigénétiques concernent : la méthylation de l'ADN, les modificateurs des histones, les remodeleurs de la chromatine et tout autre élément de la chromatine (Baylin and Jones, 2016). Les modifications génétiques et épigénétiques coopèrent dans le développement d'un cancer. Des altérations génétiques peuvent avoir lieu dans des gènes régulant l'épigénome (Baylin and Jones, 2016). Par exemple, dans le cancer colorectal il est retrouvé des mutations de DNMT1, une DNA méthyltransférase, (Kanai Hiroyoshi *et al.*, 2001). Dans les cancers du sein, de la prostate, de la vessie, du colon, du pancréas, du foie, utérins, gastrique, le mélanome, le lymphome, le myélome et le sarcome d'Ewing, il est présent une mutation qui induit une expression aberrante de EZH2, une histone méthyltransférase H3K27, (Chase and Cross, 2011; Tsang and Cheng, 2011). D'un autre côté, les changements épigénétiques peuvent avoir un effet sur l'expression des gènes, en changeant la distribution des méthylations de l'ADN (Sharma, Kelly and Jones, 2010). Ces changements contribuent directement à l'oncogenèse en inhibant la transcription de gènes suppresseurs de tumeur par hyper-mutation des îlots CpG au niveau des promoteurs (Berman *et al.*, 2012). L'analyse du méthylome à la base près montre que les *loci* ciblés par les groupes de polycombs répresseurs dans les cellules souches embryonnaires, impliqués dans la différenciation cellulaire, sont prédisposés à une méthylation aberrante dans les cellules cancéreuses (Berman *et al.*, 2012). De plus, une hypo-méthylation globale induit de l'instabilité génomique et contribue à la

transformation cellulaire (Kulis and Esteller, 2010).

Les modifications épigénétiques permettent aussi à la cellule cancéreuse de résister à la chimiothérapie ou d'échapper au système immunitaire (Jones, Issa and Baylin, 2016). Les utiliser comme cible thérapeutique est une approche de plus en plus développée. Des drogues ayant un effet de reprogrammation transcriptomique important avec des inhibiteurs des DNMT ou des déacétylases des histones ont été développés. Mais, malgré des effets anti-tumoraux importants sur le long terme (Issa *et al.*, 2004), des résistances se créent dans plusieurs types de cancer comme les leucémies, les lymphomes, les cancers du colons, du sein et de la prostate par insuffisance de triphosphate intracellulaire menant à une expression aberrante du gène de la déoxycytidine kinase (Qin *et al.*, 2009, 2011). D'un autre côté, des inhibiteurs des bromodomaines des BET, qui sont eux aussi des reprogrammeurs larges, permettent notamment d'inhiber BRD4 qui est nécessaire à l'expression forte d'oncogène comme MYC, sont en cours de développement clinique (Jones, Issa and Baylin, 2016). Des thérapies ciblant des défauts génétiques dans les gènes régulant l'épigénome ont aussi été mises au point notamment dans les lymphomes. Les lignées cellulaires venant de lymphomes sont, en effet, sélectivement tuées par un inhibiteur du gène EZH2 si les cellules portent une mutation sur ce gène (Morin *et al.*, 2010). L'utilisation de ces drogues changeant l'épigénome a aussi démontré un effet important sur l'activation de gènes, principalement sur l'expression des antigènes cancers/testicules (Karpf and Jones, 2002). Cette réactivation entraîne alors une visibilité de la tumeur par le système immunitaire du patient augmentant ainsi les effets anti-tumoraux de ces drogues (Jones, Issa and Baylin, 2016).

2. Les micro ARN

Les micro ARN sont une classe d'ARN non-codants qui agissent pour réguler l'expression des protéines en déstabilisant les ARN messagers et en inhibant la traduction. Les micro ARN matures s'associent avec les protéines se liant aux ARN formant ainsi le complexe RISC (*RNA-induced silencing complex*) et s'hybrident aux ARN messagers cibles, généralement dans la partie 3' non traduite (Nikolic *et al.*, 2017). Cette déstabilisation induit une répression partielle de la traduction puisque

l'expression des protéines est diminuée de 6 à 25% (Guo *et al.*, 2010; Eichhorn *et al.*, 2014). Les micro ARN sont aussi présents dans le noyau, environ 75% des micro ARN de la cellule se trouvent à la fois dans le cytoplasme et dans le noyau (Gagnon *et al.*, 2014). Dans le noyau les micro ARN ont plusieurs capacités : celle de réguler l'épissage des ARN messagers ainsi que celle de réguler la transcription des ARN messagers en l'activant ou en l'inhibant (**Figure 5** ; Catalanotto, Cogoni and Zardo, 2016).

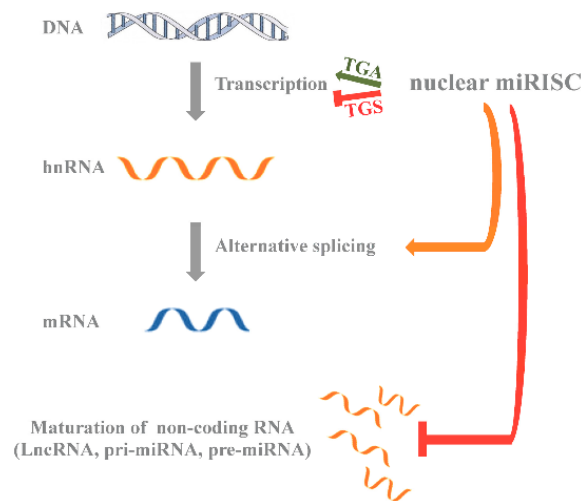


Figure 5 : Les fonctions nucléaires des micro ARN (adapté de Catalanotto *et al.*, 2016)

L'éventail des capacités des micro ARN est important, ils sont donc impliqués dans de nombreux processus physiologiques comme la différenciation, la prolifération, l'apoptose et le développement (Catalanotto, Cogoni and Zardo, 2016). La première preuve la plus directe du lien entre micro ARN et cancer est la délétion du *cluster* de micro ARN mir-15a/16-1 du chromosome 13q14 dans les cellules de leucémie lymphatique (Calin *et al.*, 2002). Les micro ARN surexprimés dans certains cancers sont appelés « oncomiR » alors que ceux sous-exprimés ont souvent un rôle de suppresseur de tumeur et sont connus sous le nom de « miR suppresseurs de tumeur » (Hata and Lieberman, 2015). Parmi les micro ARN sous-exprimés dans la plupart des cancers, il y a let-7 qui permet physiologiquement la différenciation des cellules souches et n'est pratiquement pas détectable dans ces dernières ainsi que dans les tumeurs agressives et peu différenciées (Yu *et al.*, 2007). Mir-200 fait aussi partie de ces micro ARN majoritairement sous-exprimés dans les cancers. Cette sous-expression augmente la transition épithéliomésenchymateuse, la migration et l'invasion des cellules cancéreuses (Mongroo and Rustgi, 2010). Le cluster de micro ARN mir-17-92

se compose de 6 micro ARN qui font partie des deux groupes. En effet, l'oncogène Myc active l'expression du cluster et l'inactivation de p53 entraîne la surexpression d'une partie du cluster (He *et al.*, 2005; Li *et al.*, 2012). L'inactivation de ce groupe de micro ARN supprime la formation de rétinoblastome dans des modèles murins (Nittner *et al.*, 2012). Connaître leur expression et celle de tous les autres micro ARN par rapport aux tissus normaux permet une compréhension plus profonde de la biologie des cancers et ainsi de proposer de nouveaux traitements (Rupaimoole and Slack, 2017). De plus, les programmes de transcription des micro ARN sont aussi utilisés pour identifier les différents sous-types d'une même sorte de cancer ainsi que pour identifier la cellule d'origine des tumeurs (Hata and Lieberman, 2015).

B. Les sarcomes : un groupe de cancers hétérogènes

I. Des histologies très diversifiées

Les sarcomes forment un groupe de tumeurs d'origine mésenchymateuse hétérogène et rare puisque ces tumeurs représentent environ 1% des cancers (Fletcher *et al.*, 2013). Les sarcomes sont classés en fonction de la morphologie et de la différenciation cellulaire, ils sont ainsi divisés en 12 grandes lignes de différenciation, elles-mêmes divisées en 113 sous-groupes histologiques (Figure 6 ; Taylor *et al.*, 2011).

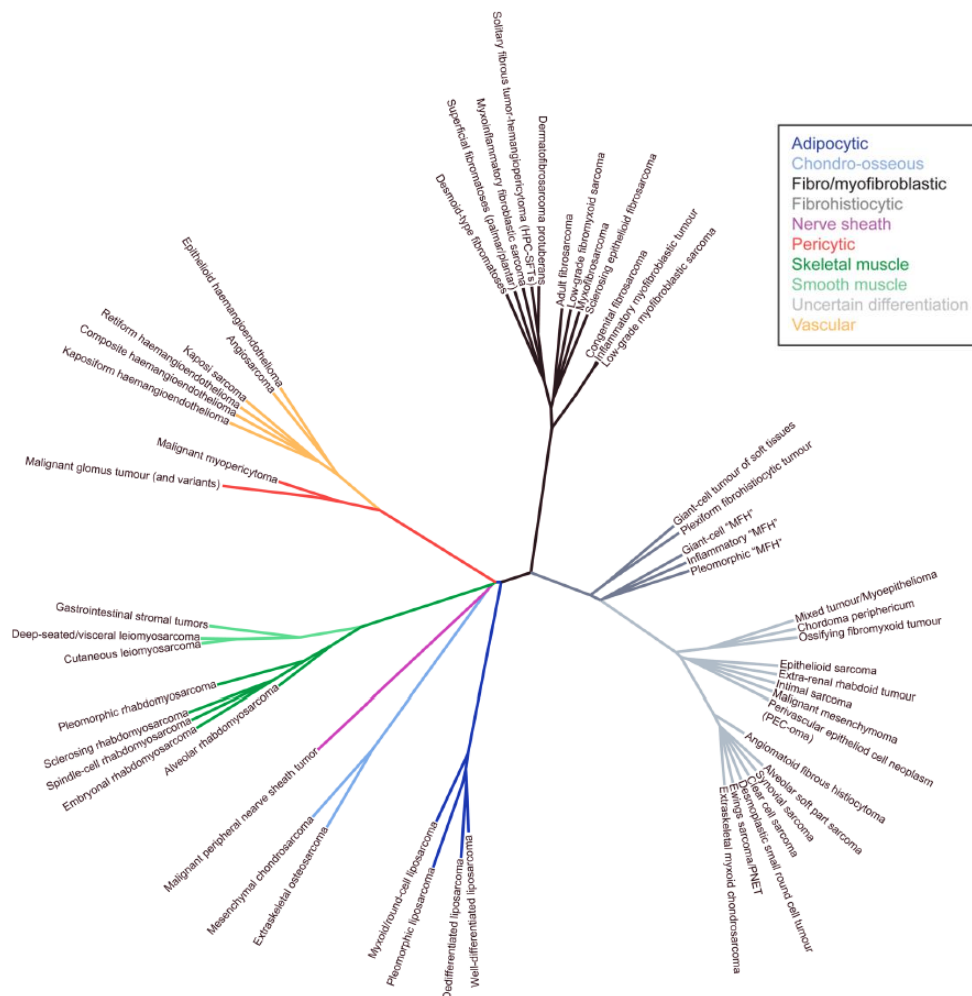


Figure 6 : Taxonomie des sarcomes (Taylor *et al.*, 2011)

Représentation de la phylogénie d'environ 60 sous-types de sarcomes (exclusion de 52 sous-types bénins). MFH : sarcome pléomorphe indifférencié

Les grandes lignes de différenciation correspondent aux différents types de tissu dont les tumeurs présentent des caractéristiques : adipeux, fibroblastique, musculaire lisse, péri-vasculaire, musculaire squelettique, vasculaire, osseux et cartilagineux, gaine nerveuse, gastro-intestinal. La classification de chaque sarcome

dépendra des analyses anatomopathologiques macroscopique et microscopique. Puis, des analyses immunohistochimiques et moléculaires permettent de définir le sous-groupe et d'adapter le traitement. Certaines tumeurs ne montrent pas de différenciation particulière ni de marqueurs moléculaires spécifiques et sont rassemblées sous le nom de sarcomes indifférenciés.

Dans une grande majorité des cas diagnostiqués (99%) les tumeurs sont bénignes et une résection chirurgicale seule permet quasi systématiquement une rémission complète (P.G. Casali *et al.*, 2018). Les autres cas sont des tumeurs agressives puisque 40 à 50% de ces cas auront une dissémination métastatique (Fletcher *et al.*, 2013). Il ne sera présenté par la suite que les tumeurs les plus agressives.

Les sarcomes sont principalement retrouvés chez des patients dont l'âge médian est de 60 ans (Ducimetière *et al.*, 2011) et localisés au niveau des membres dans 60% des cas, du tronc chez 30% des patients ou de la région de la tête et du cou pour 10% des tumeurs (Guillou and Aurias, 2010). Le ratio homme-femme est de 1,1/1 avec des différences selon la localisation de la tumeur. Les sarcomes viscéraux touchent 59% de femmes et ceux des membres 57% d'hommes (Ducimetière *et al.*, 2011).

II. La génomique hétérogène des sarcomes

Les anomalies moléculaires sont au cœur du diagnostic des sarcomes puisqu'ils présentent des caractéristiques génétiques et/ou génomiques propres (Fletcher *et al.*, 2013). Ces anomalies peuvent être associées à des caractéristiques et des comportements cliniques spécifiques. Dans certains cas, il est possible d'utiliser une anomalie spécifique pour développer un traitement, il est donc primordial de décrire le plus précisément les anomalies génétiques et/ou génomiques de chaque sous-groupe de sarcome.

1. Des *drivers* forts et leurs conséquences

a. Qu'est-ce qu'un *driver* fort ?

Un *driver* fort est un gène modifié par une altération génétique ou génomique de façon récurrente dans un sous-type de sarcome et qui est capable de diriger

l'ensemble de l'oncogénèse. L'expression de cette altération *driver* est suffisante pour déclencher l'oncogénèse et pour diagnostiquer la tumeur, comme c'est le cas pour les GIST (tumeur stromale gastro-intestinale)(Heinrich *et al.*, 2019). Dans certains cas, ces altérations *drivers* se produisent malgré un génome avec de nombreux remaniements. Ces altérations *drivers* vont soit activer un oncogène soit créer une protéine chimérique.

b. Activation d'un oncogène

Pour activer un oncogène *driver* il y a trois possibilités : une mutation ponctuelle, une translocation qui modifie son expression ou une amplification.

Les sarcomes avec des mutations activatrices sont très majoritairement représentés par les GIST qui ont des mutations activatrices des gènes *KIT* ou *PDGFRa*. Dans les deux cas, la mutation entraîne une activation constitutive du récepteur tyrosine kinase amenant une augmentation de la prolifération et de la survie cellulaire (Rossi *et al.*, 2006). Dans 70 à 80% des cas c'est le gène *KIT* qui est muté au niveau principalement des exons 9 ou 11 et dans 10% des cas c'est le gène *PDGFRa* qui est muté au niveau de l'exon 18 majoritairement (Coindre, 2010). Les autres 10% ne montrent pas de mutation de *KIT* ou *PDGFRa* mais des altérations de *NF1*, *BRAF*, *KRAS*, *NRAS* et *SDHB/SDHC* (Blay *et al.*, 2012).

Les dermatofibrosarcomes de Darier-Ferrand sont des tumeurs superficielles localement récurrentes mais rarement invasives (Saiag *et al.*, 2015) présentant une translocation récurrente $t(17;22)(q22;q13)$ (Pedeutour *et al.*, 1996). Cette dernière permet la surexpression du gène *PDGFB* grâce au déplacement du promoteur fort de *COL1A1* en amont. Sa surexpression stimule *PDGFRβ* qui est un récepteur à tyrosine kinase et augmente la prolifération (Shimizu *et al.*, 1999).

Les sarcomes présentant une amplification spécifique en 12q14-15 sont majoritairement les liposarcomes bien différenciés ou dédifférenciés et plus rarement les sarcomes intimaux et ostéosarcomes parostéaux (Pedeutour *et al.*, 1994; Coindre, 2010). Ces tumeurs sont caractérisées par la présence de chromosomes géants ou en anneaux surnuméraires contenant les séquences de la région 12q14-15. La région

amplifiée implique toujours le gène *MDM2* et dans 90% des cas également *CDK4* (Coindre, 2010). Ces gènes inactivent p53 et pRb permettant aux cellules d'augmenter leur survie et leur prolifération. Dans le cas des liposarcomes dédifférenciés, le reste du génome peut être également réarrangé avec plus particulièrement une co-amplification exclusive des régions 1p32 ou 6q23 amplifiant respectivement *JUN* et *ASK1*, entraînant un arrêt de la différenciation adipocytaire (Chibon *et al.*, 2004; Mariani *et al.*, 2007).

c. Création de protéines chimériques

Les protéines chimériques sont le résultat d'altérations chromosomiques induisant la création d'un ARN messager composé de parties d'ARN messagers venant de deux gènes différents. Cela se traduit donc par une protéine avec de nouvelles capacités qui peuvent être oncogènes. Les altérations à l'origine de ces protéines sont majoritairement des translocations mais aussi des délétions, des duplications et des inversions (Figure 7 ; Annala *et al.*, 2013).

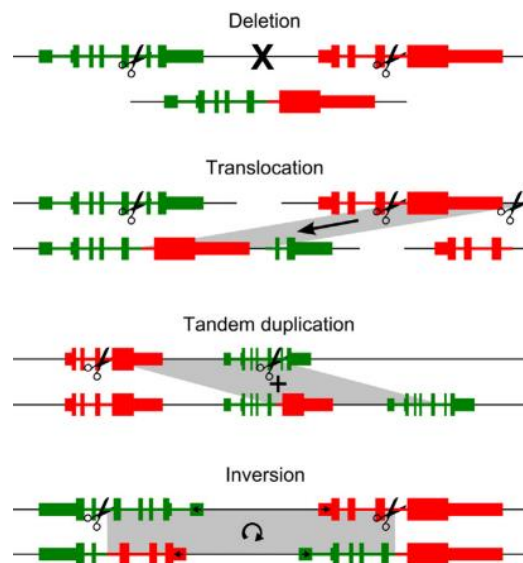


Figure 7 : Les 4 altérations chromosomiques pouvant amener à la formation de gènes de fusion (Annala *et al.*, 2013)

Dans les sarcomes, il existe plus de 300 gènes de fusion (Nakano and Takahashi, 2018b) bien que la majorité ne soit pas ou peu récurrent (moins de 1% des cas)(Mertens *et al.*, 2009), certains sont très spécifiques et leur présence permet un diagnostic clair (Coindre, 2010). Les sarcomes avec translocation, constituées de petites cellules rondes ou fusiformes et souvent agressives, représentent environ 20% des cas et surviennent

chez des patients jeunes (Coindre, 2010).

La première translocation identifiée dans une tumeur solide se trouve dans le sarcome d'Ewing entre le chromosome 22 et le chromosome 11 (Aurias *et al.*, 1983). Cette translocation amène à la fusion des gènes *EWSR1* et *FLI1* formant ainsi un facteur de transcription chimérique qui modifie l'expression de nombreux gènes (Delattre *et al.*, 1992). Parmi les sarcomes à protéine chimérique les plus fréquents, on retrouve les synoviosarcomes qui présentent de façon exclusive l'un de ces trois gènes de fusion : *SS18-SSX1* dans 65% des cas (Clark *et al.*, 1994), *SS18-SSX2* dans 35% des cas (Fligman *et al.*, 1995) ou *SS18-SSX4* de façon rare (Skytting *et al.*, 1999). Cette protéine chimérique *SS18-SSX* s'intègre au complexe SWI/SNF de remodelage de la chromatine, dont fait partie physiologiquement *SS18*, modifiant ainsi l'accessibilité à la chromatine (Banito *et al.*, 2018).

Il existe de nombreux autres sarcomes ayant une translocation produisant un gène de fusion dont la plupart code pour un facteur de transcription aberrant (Mertens *et al.*, 2009). Les principales translocations dans les sarcomes à *drivers* forts les plus fréquents sont représentées dans la **Figure 8** (Nakano and Takahashi, 2018b).

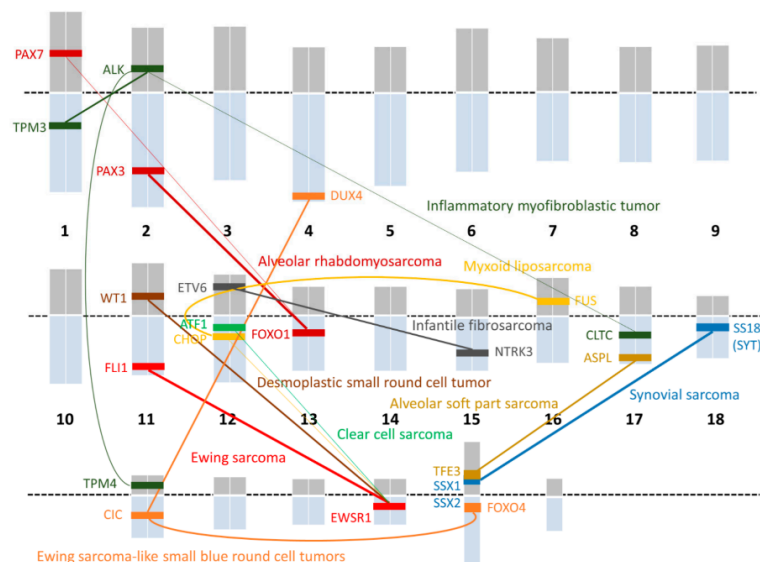


Figure 8 : Représentation des principales translocations dans les sarcomes à *drivers* forts les plus fréquents (Nakano and Takahashi, 2018)

d. Un transcriptome spécifique dépendant de chaque *driver* fort

Les altérations génétiques à l'origine de la création de *drivers* forts de l'oncogenèse définissent des groupes homogènes au niveau transcriptomique. Cela signifie que les tumeurs présentant un même *driver* ont une transcriptomique similaire et se regroupent entre elles lors d'analyses non supervisées de l'expression des gènes de différents sous-groupes de sarcomes (Figure 9 ; Segal *et al.*, 2003; Baird *et al.*, 2005; Watson *et al.*, 2018).

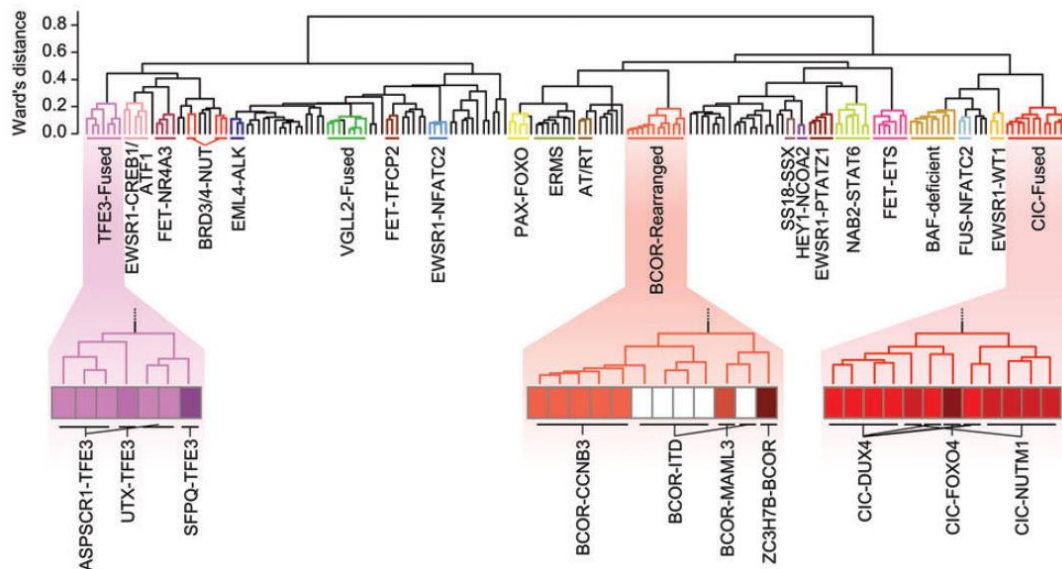


Figure 9 : Classification non supervisée de 184 sarcomes à petites cellules rondes selon l'expression de leurs gènes (Watson *et al.*, 2018)

Dans un hôte murin, des myoblastes exprimant *MYF5*, peu différenciés et induits avec SS18-SSX2 produisent des tumeurs proches des synoviosarcomes. Alors que l'induction de cette même protéine de fusion dans des myocytes ou des myofibres exprimant *MYF6*, plus avancés dans la différenciation musculaire squelettique, aura pour effet de générer une myopathie mais pas de tumeur (Haldar *et al.*, 2007). Quand le transcrit des tumeurs d'Ewing, *EWS-FLI*, est inhibé les cellules qui sont très dédifférenciées à la base expriment des marqueurs spécifiques des cellules souches mésenchymateuses (Tirode *et al.*, 2007). La capacité oncogénique des altérations à l'origine de ces transcriptomes homogènes dépend donc du fond génétique dans lequel elle a lieu. C'est-à-dire que l'altération ne sera *driver* d'une oncogenèse particulière que si elle apparait dans un type de cellule et dans une différenciation spécifique.

2. Des génomes complexes sans *driver* connu

Environ la moitié des sarcomes ne présentent pas d'altération récurrente et spécifique et sont nommés sarcomes à génomique complexe. Il en existe plusieurs sous-types ayant des différenciations différentes et dont les plus fréquents sont :

- Les léiomyosarcomes (LMS) à différenciation musculaire lisse qui font l'objet de la présente thèse et sont introduits en détail dans le paragraphe « **Zoom : les léiomyosarcomes** »
- Les sarcomes pléomorphes indifférenciés (UPS) sans différenciation spécifique
- Les myxofibrosarcomes (MFS) à différenciation fibreuse avec une matrice extracellulaire myxoïde
- Les liposarcomes dédifférenciés (DDLPS) à faible différenciation adipocytaire
- Les ostéosarcomes (OS) à différenciation ostéogénique et présentant des calcifications
- Les rhabdomyosarcomes pléomorphes (RMS) à différenciation musculaire squelettique

a. Hétérogénéité des génomes et des transcriptomes

Ces génomes complexes comportent un grand nombre de modifications de la structure des chromosomes caractérisant une forte instabilité génomique. Au niveau génomique, ils comportent un faible taux de mutation et un taux important d'altération du nombre de copies associées à de nombreux réarrangements structuraux (**Figure 10**; The Cancer Genome Atlas Research Network *et al.*, 2017). Cela signifie que ces génomes présentent de nombreuses pertes, gains et amplifications au niveau de chromosomes entiers ou de bras de chromosomes. La grande majorité de ces altérations sont spécifiques de chaque tumeur en dehors des altérations récurrentes des voies TP53 et RB1 (Pérot *et al.*, 2010; Gibault *et al.*, 2011; The Cancer Genome Atlas Research Network *et al.*, 2017).

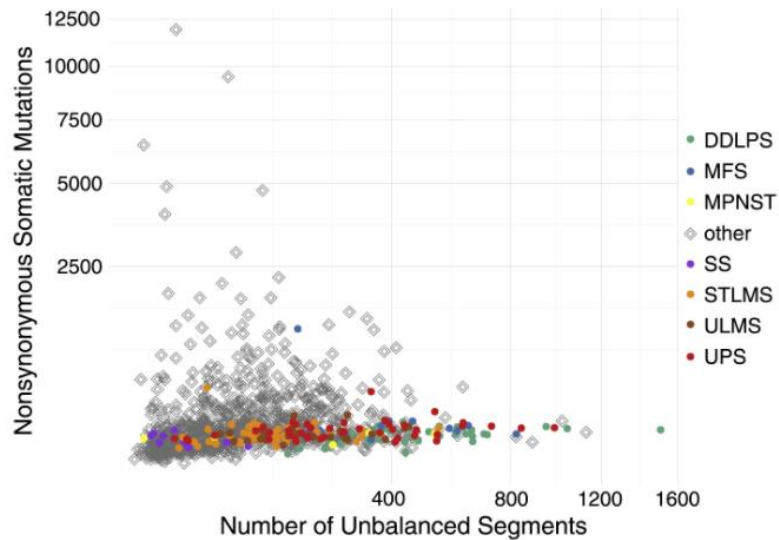


Figure 10 : Nombre de mutations somatiques non-synonymes par rapport au nombre d'altérations du nombre de copies (TCGA, 2017)

Représentation des 206 sarcomes en couleur selon leur histotype et de 1130 tumeurs prise au hasard dans la cohorte du TCGA.

Au niveau transcriptomique, au contraire des sarcomes avec un *driver* fort, les différents types de tumeurs sont associés à peu d'altérations spécifiques et récurrentes. Il est retrouvé un seul transcrit de fusion récurrent impliquant TRIO et différents partenaires dans 6% des sarcomes à génomique complexe (Delespaul *et al.*, 2017). Il est aussi retrouvé trois gènes mutés de façon récurrente : *TP53*, *RB1* et *ATRX* avec une prévalence dans les LMS de ce dernier (The Cancer Genome Atlas Research Network *et al.*, 2017). Actuellement, les analyses transcriptomiques non supervisées forment plusieurs groupes de sarcomes à génomique complexe sans lien avec leur histologie (Gibault *et al.*, 2011; The Cancer Genome Atlas Research Network *et al.*, 2017), hormis une partie des LMS qui se groupe entre eux. Au sein de ces groupes de sarcomes à génomique complexe, les profils transcriptomiques sont hétérogènes (Figure 11).

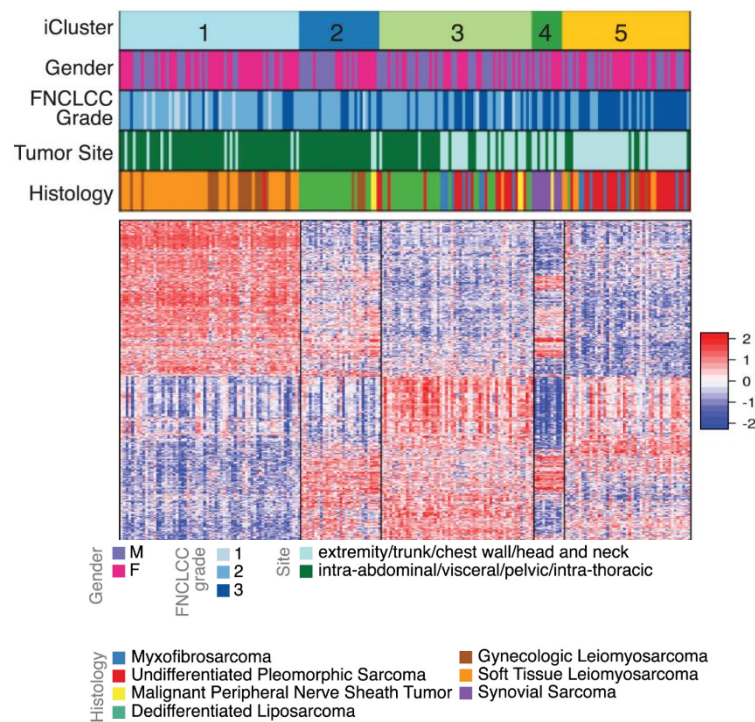


Figure 11 : Analyse non supervisée de l'expression des ARN messagers de 206 sarcomes à génomique complexe (TCGA, 2017)

b. Instabilité des génomes

L'instabilité génomique est la capacité des cellules tumorales à acquérir des altérations chromosomiques au cours du temps. Cette instabilité est observée dans 90% des tumeurs (Sansregret, Vanhaesebroeck and Swanton, 2018) et est la conséquence de différents mécanismes : soit la mutation de gènes impliqués dans la maintenance de l'intégrité chromosomique, soit des événements touchant l'ensemble du génome (Bakhoun and Cantley, 2018). Ces derniers sont les plus fréquents dans les sarcomes à génomique complexe.

i. Doublement du génome

Les sarcomes à génomique complexe sont significativement associés à un doublement du génome (The Cancer Genome Atlas Research Network *et al.*, 2017). Ces tumeurs sont majoritairement tétraploïdes comme 55% des LMS (Chudasama *et al.*, 2018). Cette tétraploïdisation serait un événement favorable à l'évolution de la tumeur car, dans un des LMS de l'étude de Chudasama, une métastase est tétraploïde mais pas sa tumeur d'origine (Chudasama *et al.*, 2018). De plus, la tétraploïdisation de

lignées cellulaires de sarcome favorise la migration cellulaire sans impacter la prolifération (Jemaà *et al.*, 2016). Les génomes tétraploïdes sont instables, favorisant la restructuration du génome puisqu'ils sont permissifs aux altérations génomiques et plus tolérants à l'aneuploïdie (Dewhurst *et al.*, 2014). La tétraploïdisation serait donc un état intermédiaire dans l'évolution tumorale amenant à ces génomes complexes.

A l'origine de ce doublement, on retrouve plusieurs mécanismes possibles : l'endoréplication, qui est un doublement du génome sans division cellulaire, une cytokinèse abortive, qui est une mitose sans la séparation des deux cellules et est induite par une altération des protéines impliquées dans ce processus, ou encore la fusion cellulaire.

Récemment plusieurs études sur le rôle de la fusion cellulaire dans les sarcomes montrent que ce mécanisme peut être à l'origine de l'oncogenèse de ces tumeurs. La fusion *in vitro* de différents types de cellule reproduit les génomes très remaniés caractéristiques des sarcomes à génomique complexe (Delespaul *et al.*, 2019; Merle *et al.*, 2020). De plus, selon les types de cellules fusionnées *in vitro*, les tumeurs développées *in vivo* comportent les altérations spécifiques du type histologique des tumeurs : la fusion de fibroblastes donne des UPS avec gain du chromosome 1 et perte des chromosomes 13, 10 et 16 (Delespaul *et al.*, 2019) et celle de myoblastes donne des sarcomes myogéniques avec délétion de *DMD* (Merle *et al.*, 2020). Enfin, ce même évènement de fusion peut être à l'origine de la capacité des cellules de sarcome à former des métastases (Lartigue *et al.*, 2020).

ii. Le mécanisme ALT

Un autre processus à l'origine d'instabilité chromosomique est le mécanisme ALT pour *Alternative Lengthening of Telomeres* (allongement alternatif des télomères). Ce mécanisme est retrouvé dans environ 41% des sarcomes des tissus mous (Lawlor *et al.*, 2019) avec une prédominance dans les LMS et les UPS/MFS (The Cancer Genome Atlas Research Network *et al.*, 2017). C'est un moyen pour la cellule tumorale de devenir immortelle ; cela permet aux télomères « frères » (venant de chromatides sœurs) d'être allongés sans intervention de la télomérase (Henson *et al.*, 2002). Mais le modèle utilisé pour répliquer les télomères peut aussi être son propre modèle, en créant une boucle de réplication ou même le télomère d'un autre chromosome, ce qui engendre des translocations de la fin des chromosomes (Figure 12 ; Henson *et al.*, 2002; Amorim *et al.*, 2016).

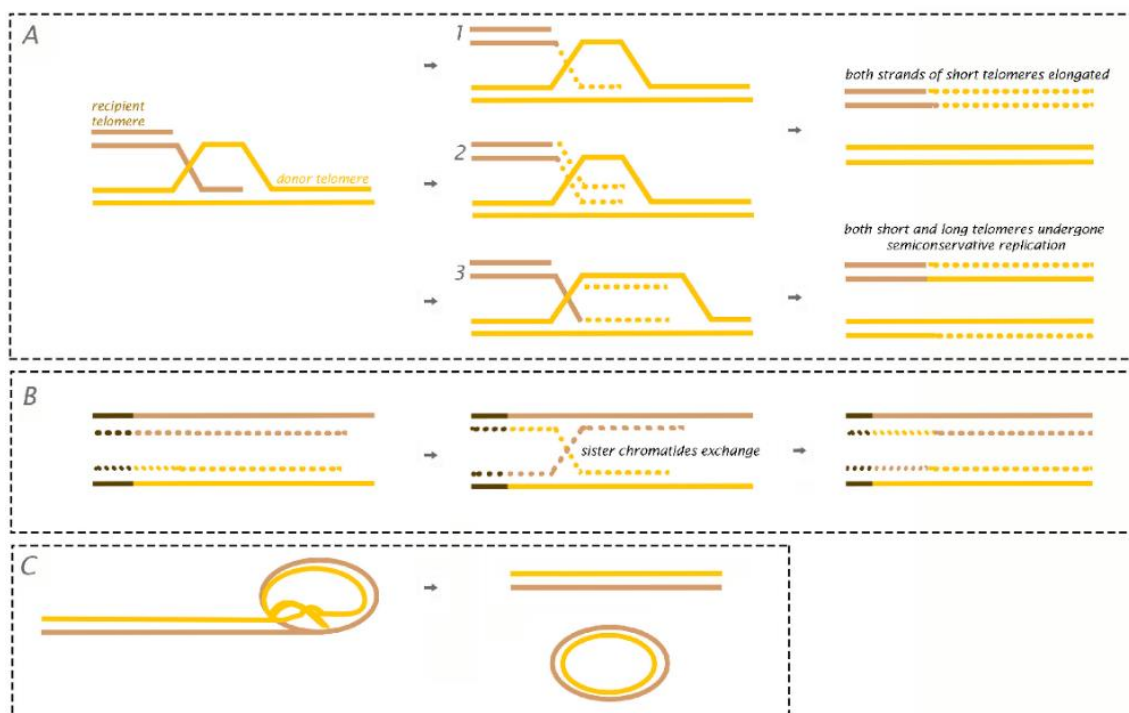


Figure 12 : Modèles théoriques du mécanisme ALT (Amorim *et al.*, 2016)

A - Modèles de réplication induit par des cassures B - Echange entre chromatides sœurs C - Formation de Cercle T (télomérique) après résolution de la boucle T

Le mécanisme ALT est permis par la présence des protéines NR2C/F aux télomères qui ont un rôle architectural, en créant un pont entre elles, et favorisent la proximité spatiale et les recombinaisons télomère-télomère. Mais la présence de NR2C/F aux télomères crée aussi des rapprochements entre les télomères et les autres séquences régularées par NR2C/F dans l'ensemble du génome. Cette organisation entraîne des insertions d'ADN télomérique créant des sites fragiles connus pour se casser et transloquer facilement. ALT, *via* les protéines NR2C/F, est donc à l'origine de nombreux remaniements et de l'instabilité chromosomique (Figure 13 ; Marzec *et al.*, 2015).

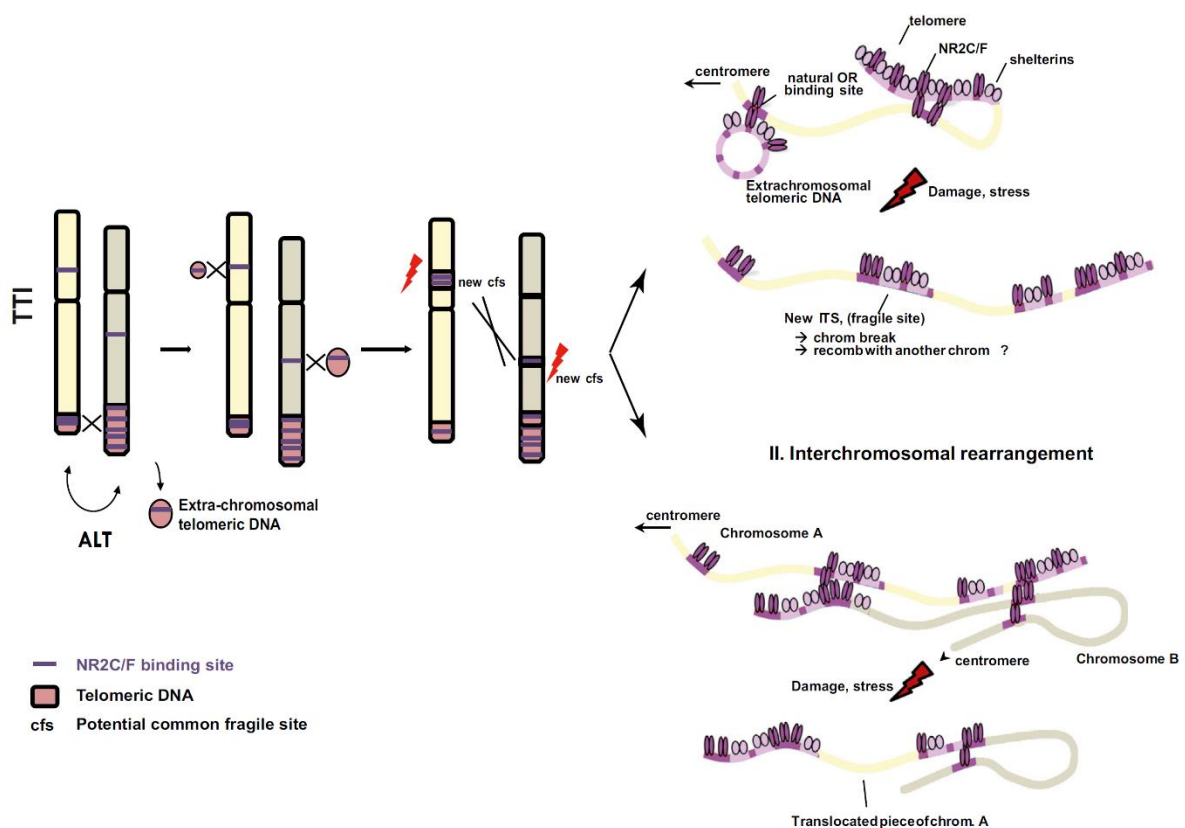


Figure 13 : Insertion d'ADN télomérique par le système ALT amenant une instabilité chromosomique (Marzec *et al.*, 2015)

c. Des altérations communes malgré le chaos

i. Les gènes du cycle cellulaire

Deux voies du cycle cellulaire sont fréquemment inactivées dans les sarcomes à génomique complexe : celles de p53 et de Rb.

TP53, qui se situe en 17p13.1, est délété de manière homozygote dans 9% des LMS, 16% des UPS et 12% des MFS (The Cancer Genome Atlas Research Network *et al.*, 2017). Plus précisément, dans les UPS et LMS on retrouve plusieurs combinaisons d'altérations en plus des délétions homozygotes : la mutation des deux allèles, la perte d'un allèle et la mutation du second, la mutation d'un seul des deux allèles uniquement, et la perte d'un seul des deux allèles. Cela ne laisse que 20% des LMS et 29% des UPS sans altération de *TP53*. Dans ces cas, il est fréquemment retrouvé une altération de *CDKN2A* (Pérot *et al.*, 2010). Plusieurs modèles murins de sarcomes pléomorphes sont basés sur l'inactivation de la voie p53 et l'altération d'un second gène (activation de *Kras* ou inactivation de *Pten*) (Tsumura *et al.*, 2006; Hernando *et al.*, 2007; Kirsch *et al.*, 2007). p53 a un rôle majeur dans le respect de l'intégrité du génome lors de la réplication puisqu'il est activé en réponse à différents stress tels que les dommages à l'ADN, l'hypoxie ou les irradiations UV. L'activation de p53 induit l'arrêt du cycle cellulaire ou l'apoptose de la cellule (Lavin and Gueven, 2006). p14 inhibe l'ubiquitine ligase MDM2 dont l'activation permet la dégradation par le protéasome de p53 (Matheu, Maraver and Serrano, 2008).

RB1 se situant en 13q14.2 est délété de manière homozygote dans 14% des LMS, 16% des UPS et 24% des MFS. (The Cancer Genome Atlas Research Network *et al.*, 2017). Si les délétions hétérozygotes de la région sont ajoutées, *RB1* est perdu dans 78% des UPS (Chibon *et al.*, 2000) et dans 82% des LMS (The Cancer Genome Atlas Research Network *et al.*, 2017). Il est aussi muté dans 15% des LMS (The Cancer Genome Atlas Research Network *et al.*, 2017). La protéine Rb doit être hypo-phosphorylée pour être activée et se lier à E2F, empêchant la liaison de ce facteur de transcription à des gènes de la transition G1/S. Quand Rb est phosphorylée par la cycline D, il se dissocie de E2F1 (Dick and Rubin, 2013). Lors de dommages à l'ADN, Rb est activée pour permettre la répression des gènes du cycle cellulaire.

ii. ATRX : le remodeleur de la chromatine

ATRX fait partie des 3 gènes les plus fréquemment mutés avec une prévalence plus importante dans les LMS, ces mutations sont dans la majorité des cas des mutations tronquantes (The Cancer Genome Atlas Research Network *et al.*, 2017). La

protéine est retrouvée perdue dans environ 16% des sarcomes à génomique complexe (Liau, Lee, *et al.*, 2015; Yang *et al.*, 2015; Lee *et al.*, 2016). La perte d'ATRAX est directement liée au mécanisme ALT puisque l'expression d'ATRAX supprime ALT (Clynes *et al.*, 2015) et que les sarcomes présentant le phénotype ALT sont associés à la perte d'ATRAX (Heaphy *et al.*, 2011; Lovejoy *et al.*, 2012; Chen *et al.*, 2014; Lee *et al.*, 2015; Liau, Lee, *et al.*, 2015).

Physiologiquement ATRX joue un rôle important dans le maintien des boucles T à la fin des chromosomes qui permettent de garder des télomères intègres. De par son domaine ADD (Aapola *et al.*, 2000), ATRX reconnaît des méthylations spécifiques des histones : la triméthylation de la lysine 9 de l'histone 3 (H3K9me3) et la triméthylation de la lysine 20 de l'histone 4 (H4K20me3) (Schotta *et al.*, 2004). Après s'être lié aux histones comportant ces marques, ATRX et son partenaire DAXX placent le variant histone 3.3 dans le nucléosome suivant, contribuant ainsi à la stabilisation de structures de chromatine répétée (Goldberg *et al.*, 2010). L'absence d'ATRAX ou de DAXX entraînera l'absence du variant histone 3.3 et donc une déstabilisation de la boucle T laissant ainsi le télomère libre aux recombinaisons (Figure 14).

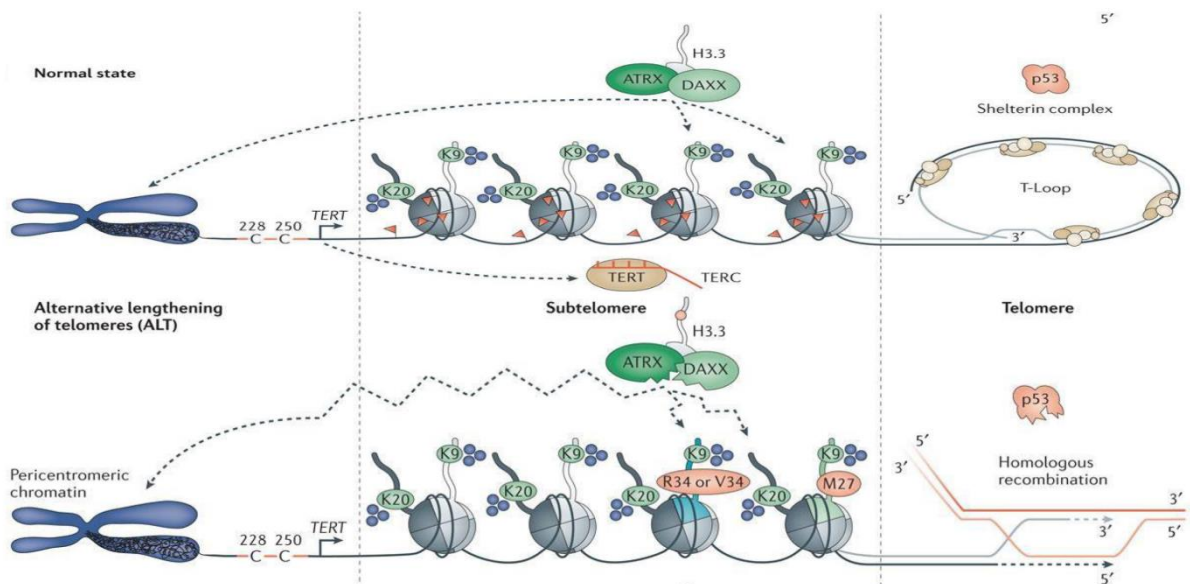


Figure 14 : La perte d'ATRAX entraîne l'activation du mécanisme ALT (Strum *et al.*, 2014)

Des enrichissements d'ATRAX sont aussi retrouvés au niveau des sites hétérochromatiques constitutifs tels que l'hétérochromatine péricentrique, l'X inactif (Baumann and Fuente, 2009) et des familles de rétrotransposons (Goldberg *et al.*, 2010).

En plus de son rôle sentinelle du maintien de l'hétérochromatine constitutive, ATRX se lie à l'ADN réprimé dans les différents types cellulaires ainsi qu'aux exons 3' d'un sous-groupe des gènes *ZNF* transcrits, la plus grande famille de facteurs de transcription. Les ZNF de ce sous-groupe contiennent les marques histones H3K9me3 et H3K36me3 ainsi que le domaine de répression KRAB (Valle-García *et al.*, 2016). En plus de son rôle de répresseur de la chromatine, ATRX aurait donc un rôle de maintien des marques d'histone permettant l'expression de certains facteurs de transcription.

III. Prises en charge thérapeutiques variées des sarcomes

Les sarcomes étant un groupe très hétérogène de tumeurs, les traitements le sont tout autant. La prise en charge thérapeutique est donc discutée au cas par cas lors de réunions de concertation pluridisciplinaire composées de l'ensemble des médecins pouvant prendre en charge les sarcomes du diagnostic au traitement (oncologues, anatomopathologistes, chirurgiens, radiologistes interventionnels, radiothérapeutes, *etc.*). Pour les aider dans leurs décisions les principales recommandations sont publiées sous formes de rapports réguliers de l'*European Society for Medical Oncology* (ESMO). Les caractéristiques cliniques des sarcomes étant variées un grand nombre de traitements peuvent être envisagés. Cependant, ce cancer n'a bénéficié que de peu de grandes avancées thérapeutiques depuis plusieurs décennies (Demicco *et al.*, 2012).

La première étape de tout traitement est le diagnostic réalisé par un anatomopathologiste spécialisé dans les sarcomes. Le diagnostic repose sur :

- L'analyse microscopique d'une micro-biopsie qui permet de décrire la forme des cellules et de leur noyau, l'infiltrat immunitaire, la composition de la matrice extracellulaire et le nombre moyen de mitoses
- L'analyse d'immunohistochimies dont les différents marquages permettront de déterminer la ligne de différenciation de la tumeur et dans certains cas de déterminer le type de sarcome
- L'analyse moléculaire qui permet de déterminer le type de sarcome notamment dans le cas des sarcomes à *driver* fort
- Et la détermination du grade qui se fait selon les critères FNCLCC

(Coindre, 1998) donnant 3 grades selon la propension d'une tumeur à former une métastase sauf pour les GIST qui sont évalués par le système de l'Armed Forces Institute of Pathology (Miettinen and Lasota, 2006).

A noter que plusieurs études cliniques sont en cours afin de remplacer le grade par la signature CINSARC dans le pronostic des sarcomes.

1. La chirurgie

Dans les sarcomes des tissus mous et les GIST localisées, la résection chirurgicale avec, si possible, des marges saines à l'examen microscopique (dite chirurgie R0) est le traitement de référence (P. G. Casali, Abecassis, *et al.*, 2018a, 2018b). Selon la localisation et/ou le volume de la tumeur, une résection incomplète peut être planifiée mais sera de mauvais pronostic (Figure 15).

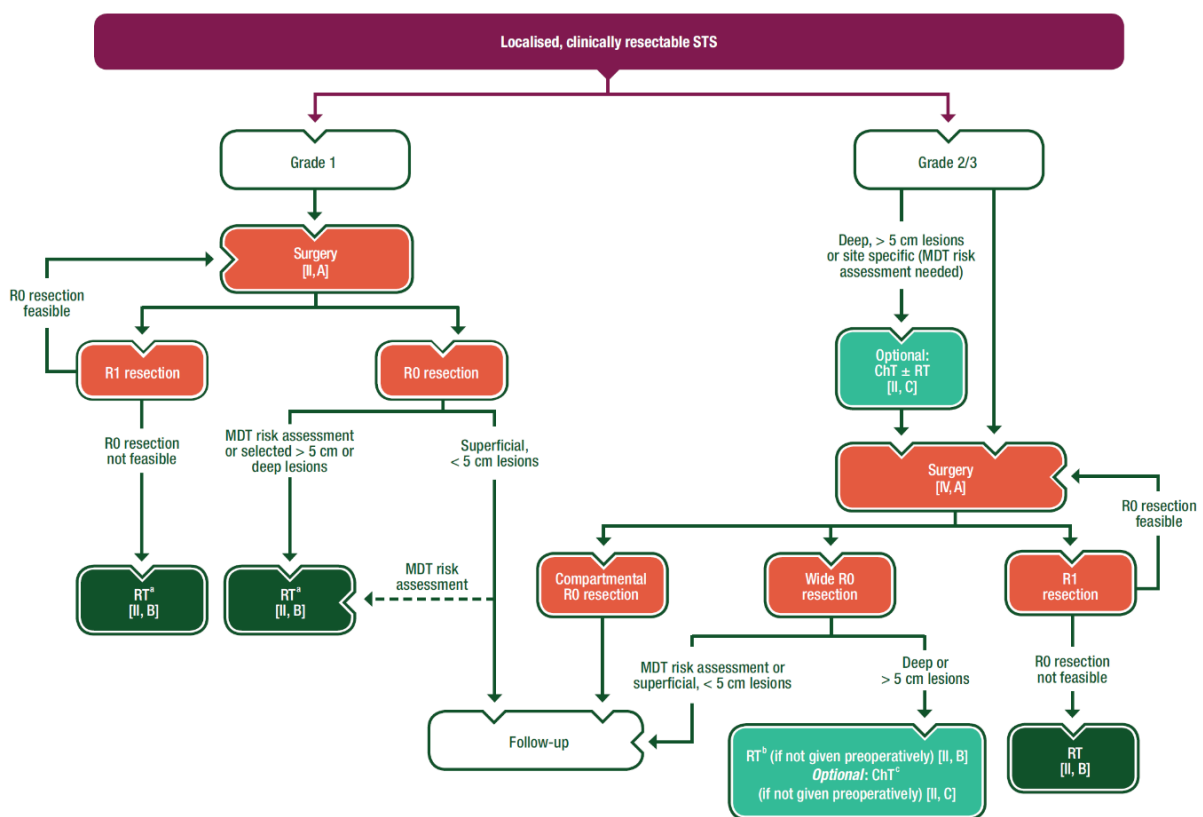


Figure 15 : Gestion des sarcomes des tissus mous localisés (Casali *et al.*, 2018c)

MDT : équipe pluridisciplinaire, RT : radiothérapie, ChT : chimiothérapie

La non-spécialisation du chirurgien dans les sarcomes et une chirurgie incomplète (dite R1) voir une chirurgie dite « oups », c'est-à-dire non préparée et avec effraction tumorale, sont des facteurs pronostiques péjoratifs quant à l'évolution tumorale (Toulmonde *et al.*, 2014). Dans la plupart des cas, la non prise en charge dans un centre expert se traduit par une seconde chirurgie dans un centre spécialisé (Koulaxouzidis *et al.*, 2015) mais le pronostic ne sera pas rattrapé par un quelconque traitement ultérieur (Honoré *et al.*, 2015).

2. La chimiothérapie

Dans le cas des sarcomes métastatiques des tissus mous ou de sarcomes des os (ostéosarcomes et sarcomes d'Ewing), à haut risque ou non résécable, la chimiothérapie est recommandée en néoadjuvant afin de réduire le volume tumoral initial facilitant ainsi la chirurgie (P. G. Casali, Abecassis, *et al.*, 2018b; P. G. Casali, Bielack, *et al.*, 2018). Bien qu'il n'y ait pas de bénéfice démontré, sur la survie globale ou la survie sans rechute, de la chimiothérapie adjuvante dans les cas de sarcomes à haut risque, ce traitement reste discuté au cas par cas et avec le patient (P. G. Casali, Abecassis, *et al.*, 2018b).

La première ligne de traitement par chimiothérapie pour les patients métastatiques est l'utilisation d'anthracyclines comme la doxorubicine. Celle-ci peut être additionnée d'ifosfamide ou de cisplatine et de methotrexate dans les OS ou de dacarbbazine dans les LMS (P. G. Casali, Abecassis, *et al.*, 2018b; P. G. Casali, Bielack, *et al.*, 2018).

3. La radiothérapie

La radiothérapie est proposée généralement en adjuvant pour les sarcomes de haut grade pour améliorer le contrôle local (Toulmonde *et al.*, 2014). Elle est recommandée en néoadjuvant dans deux cas : les complications seraient trop importantes lors d'une administration post-opératoire ou lors d'une chirurgie incomplète qui ne sera pas réopérée (P. G. Casali, Abecassis, *et al.*, 2018b).

L'efficacité de la radiothérapie dépend de la localisation de la tumeur mais

surtout du sous-type histologique du sarcome. Les synoviosarcomes et les liposarcomes sont reconnus comme généralement radiosensibles. Alors que la plupart des sarcomes pléomorphes ne répondent pas bien à la radiothérapie (P. G. Casali, Abecassis, *et al.*, 2018b; P. G. Casali, Bielack, *et al.*, 2018).

4. Les thérapies ciblées

a. Les inhibiteurs

Actuellement les thérapies ciblées sont très limitées pour les sarcomes même pour les sarcomes ayant un *driver* fort qui pour la très grande majorité n'ont pas de traitement spécifique. Seul les GIST sont traités par un inhibiteur des tyrosines kinases appelé imatinib. L'utilisation de cet inhibiteur a permis une augmentation de la survie globale des patients métastatiques de moins d'1 an à pratiquement 5 ans. Il existe cependant une limitation à ce traitement puisque les patients avec une mutation dans l'exon 9 de *KIT* (environ 10 à 20% des cas) ne répondent pas à cet inhibiteur. De nouveaux inhibiteurs des tyrosines kinases sont développés et permettent d'augmenter la survie sans progression des patients avec un GIST résistant à l'imatinib (Nakano and Takahashi, 2018a).

Le pazopanib, un inhibiteur de multiples tyrosines kinases, a aussi été testé sur des léiomyosarcomes, des synoviosarcomes, des liposarcomes et d'autres histologies de sarcomes des tissus mous métastatiques. Les liposarcomes ont été sortis de l'étude par manque de réponse. Pour les autres histologies, les résultats montrent une augmentation de la médiane de survie sans progression de 3 mois par rapport au placebo ainsi qu'une augmentation de la survie globale de 1,8 mois par rapport au placebo (van der Graaf *et al.*, 2012). Bien que ce traitement soit proposé comme une nouvelle option pour les patients avec un sarcome métastatique non-adipocytaire, les bénéfices restent très limités.

b. L'immunothérapie

Pour le moment aucune immunothérapie n'est proposée comme traitement de référence des sarcomes, hormis l'olatumab, ciblant PDGFR α , en combinaison avec

de la doxorubicine dans les sarcomes des tissus mous métastatiques et non résecables (Tap *et al.*, 2016; P. G. Casali, Abecassis, *et al.*, 2018b).

Aux vues des réussites des immunothérapies dans les autres cancers, de nombreux essais cliniques ont été réalisés dans les sarcomes, ciblant différents niveaux du système immunitaire. Cependant dans l'ensemble des études publiées à ce jour, le niveau de réponse reste très limité avec une survie globale d'un maximum de 14 mois (Tableau 2 ; Ayodele and Abdul Razak, 2020).

Reference	Agent	Phase	Pts (n)	Indication	Response rate	Survival
<i>Checkpoint inhibitors</i>						
Maki <i>et al.</i> , 2013 ¹⁹	Ipilimumab	I	6	Advanced SyS	0 of 6	mOS: 8.75 months
Tawbi <i>et al.</i> , 2017 ¹⁶	Pembrolizumab	II		Selected STSs and bone sarcomas	18% in STS, 40% in UPS, 20% in LPS, 10% in SyS	mPFS: 18 weeks; OS: 49 weeks
D'Angelo <i>et al.</i> , 2018 ²¹	Nivolumab with or without ipilimumab	II	96	Metastatic STS	Nivolumab: 5%; Ipilimumab–nivolumab: 16%	mPFS: 4.1 months; OS: 14.3 months
Toulmonde <i>et al.</i> , 2018 ²⁶	Pembrolizumab, cyclophosphamide	II	57	Advanced STS	Solitary fibrous tumour in 1 patient	NA
Wilky <i>et al.</i> , 2019 ²²	Axitinib, pembrolizumab	II		ASPS and other STSs	25%, all STS patients; 50.4%, ASPs patients	3-Month PFS: 66%; in ASPs patients: 73%
<i>Adoptive cell therapy</i>						
Robbins <i>et al.</i> , 2011 ²⁷	Adoptively transferred autologous T cells transduced with a T cell receptor directed against NY-ESO-1	I	6	Metastatic SyS expressing NY-ESO-1	4 of 6	NA
<i>Vaccines</i>						
Mahvi <i>et al.</i> , 2002 ²⁸	Tumour cells treated with granulocyte macrophage colony-stimulating factor	I	16	Melanoma and sarcoma	1 of 16	NA
Dillman <i>et al.</i> , 2004 ²⁹	Autologous tumour cell-line-derived vaccines	I/II	23	Recurrent or metastatic sarcoma	No objective response	10 Patients lived more than 1 year
Kawaguchi <i>et al.</i> , 2005 ³⁰	Vaccination with SYT–SSX junction peptide	I	6	Metastatic SyS	0 of 6	NA
Finkelstein <i>et al.</i> , 2012 ³¹	Radiotherapy with intratumoural injection of dendritic cells	I/II	17	Neoadjuvant treatment in high-risk STS	9 of 17	1-Year PFS: 70.6%
Kawaguchi <i>et al.</i> , 2012 ³²	SYT–SSX breakpoint peptide vaccines	I/II	21	Metastatic SyS	1 of 21 (stable disease: 6 of 21)	NA

Pts = patients; SyS = synovial sarcoma; mOS = median overall survival; UPS = undifferentiated pleomorphic sarcoma; LPS = liposarcoma; mPFS = median progression-free survival; OS = overall survival; NA = not applicable; ASPs = alveolar soft-part sarcoma; PFS = progression-free survival.

Tableau 2 : Sélection d'études complètes utilisant de l'immunothérapie dans les sarcomes des tissus mous (Ayodele and Abdul Razak, 2020)

Les inhibiteurs des points de contrôle du système immunitaire sont des drogues qui bloquent les protéines permettant de limiter la réponse du système immunitaire. Cette inhibition doit permettre aux cellules immunitaires comme les cellules T de mieux reconnaître et mieux tuer les cellules cancéreuses. Pour que ces drogues fonctionnent, il est nécessaire d'avoir un infiltrat immunitaire important déjà présent

dans la tumeur. Or dans les sarcomes plus de la moitié (54%), notamment les LMS, les synoviosarcomes et les chondrosarcomes, ont une faible densité d'infiltrat immunitaire en lymphocytes et macrophages (D'Angelo *et al.*, 2018). Les inhibiteurs des points de contrôle du système immunitaire ne seraient donc efficaces que dans certains sous-types histologiques. Cela a été validé par l'étude SARC08 de Tawbi *et al.* dans laquelle il y a eu des réponses avec un inhibiteur de PD-1 de 40% des UPS mais aucune avec les LMS. Malgré un taux de réponse correct, la médiane de survie sans progression était de 18 semaines (Tawbi *et al.*, 2017).

Il existe d'autres thérapies basées sur la modulation, la manipulation et la sélection de cellules T autologues *in vitro*. Les cellules T sont récupérées sur le patient et réinjectées après expansion de la population cellulaire choisie ou modifiée. Ce genre de stratégie n'a été testé que sur une petite cohorte de synoviosarcomes. Dans cette étude les cellules T autologues étaient transduites avec un récepteur dirigé contre NY-ESO-1, un antigène exprimé par 70 à 80% des synoviosarcomes. Quatre des six synoviosarcomes, sélectionnés pour leur expression de NY-ESO-1, ont une réponse objective au traitement (Robbins *et al.*, 2011). Ce traitement semble prometteur mais son efficacité sur une cohorte plus importante et en comparaison avec les standards actuels reste à démontrer.

Il est aussi possible d'utiliser des vaccins thérapeutiques qui sont basés sur l'activation du système immunitaire inné en activant les cellules dendritiques par la présence d'antigènes immunogéniques prédéterminés. La plupart des essais sur les sarcomes, utilisant soit des antigènes non spécifiques soit la cellule cancéreuse entière après irradiation, n'ont pas eu d'efficacité démontrée (Ayodele and Abdul Razak, 2020). Une étude, ayant utilisé des vaccins constitués par des peptides personnalisés sur 20 sarcomes des tissus mous et des os réfractaires à tout traitement, a montré une réduction des métastases pulmonaires et une stabilisation de la maladie dans tous les cas mais avec une médiane de survie globale à 9,6 mois (Takahashi *et al.*, 2013).

Pour l'ensemble de ces techniques, des essais sont encore en cours sur de nouvelles molécules ou des combinaisons de molécules afin d'améliorer la survie des patients.

C.Zoom : les léiomyosarcomes

I. Descriptions anatomopathologique et épidémiologique

Les léiomyosarcomes ou LMS représentent environ 11% des nouveaux cas de sarcomes des tissus mous diagnostiqués (WHO Classification of Tumours Editorial Board, 2020). Ce qui représente, en France, environ 600 nouveaux cas chaque année (Blay *et al.*, 2017). Les LMS sont retrouvés dans tout le corps avec 3 localisations majoritaires : le rétropéritoine, l'utérus et les membres. Au niveau du rétropéritoine et de la cavité abdominale, les LMS sont les sarcomes prédominants parmi ceux ayant une origine dans les vaisseaux sanguins et notamment venant de la veine cave (Serrano and George, 2013). L'incidence des LMS augmente avec l'âge avec un pic autour des 70-80 ans, hormis les LMS utérins dont le pic d'incidence est vers la péri-ménopause des femmes, aux alentours de 50 ans (Miettinen, 2016). Enfin, le sexe a une incidence sur la localisation du LMS, on retrouve plus de femmes chez les patients avec une tumeur rétro-péritonéale (Hashimoto, Tsuneyoshi and Enjoji, 1985) et plus d'hommes ayant une tumeur des membres non-cutanée (Gustafson, 1994) ou cutanée (Fields and Helwig, 1981).

En histologie, les LMS sont caractérisés par des cellules fusiformes entrecroisées. Ce sont des tumeurs généralement cellulièrement compactes mais pouvant présenter des remaniements locaux fibreux ou myxoïdes (Fletcher *et al.*, 2013). Leur caractéristique principale étant l'expression de marqueurs des muscles lisses ; au moins cinq d'entre eux peuvent être utilisés en immunohistochimie pour aider au diagnostic. On retrouve l' α -actine musculaire lisse (*ACTA2*), la desmine (*DES*), la h-caldesmone (*CALD1*), la calponine (*CNN1*) et la transgeline (*TAGLN* ou *SM22a*) (**Figure 16**). Parmi eux, la transgeline a les meilleures sensibilité et spécificité, ce qui en fait un marqueur diagnostique robuste (Robin *et al.*, 2013).

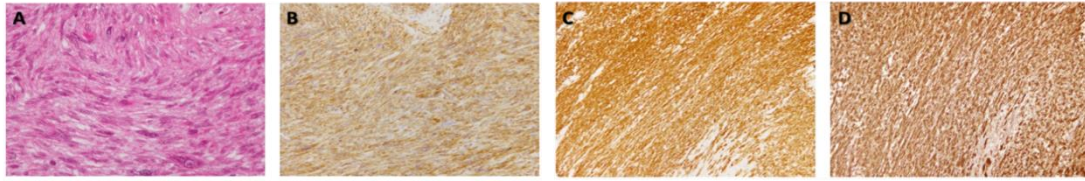


Figure 16 : Léiomyosarcome différencié (Robin *et al.*, 2013)

Marquage de LMS (a) Hématoxyline et éosine, (b) actine musculaire lisse, (c) h-caldesmone et (d) transgéline.

Il a été aussi proposé de réaliser un score composé de 4 marqueurs : l' α -actine musculaire lisse, la caldesmone, la calponine et la transgéline permettant un diagnostic fiable des LMS ainsi qu'une classification selon leur différenciation (Pérot *et al.*, 2014).

II. Des génomes complexes avec des altérations communes

1. Des génomes très remaniés

Les LMS font partie des sarcomes ayant un génome très remanié contenant des variations importantes du nombre de copies. En effet, comme les autres sarcomes ayant une génomique complexe, leur génome se compose d'un faible taux de mutation mais de nombreux gains et délétions avec une part plus importante de délétions (The Cancer Genome Atlas Research Network *et al.*, 2017). Certaines régions chromosomiques sont fréquemment perdues comme les régions 13q14, 17p12, 16q et 10q (Gibault *et al.*, 2011). On retrouve à l'intérieur de ces régions perdues des gènes plus altérés que les autres comme *RB1*, en 13q14, perdu dans 82% des LMS et *TP53*, en 17p12, perdu dans 69% des LMS et muté dans 50% des cas (The Cancer Genome Atlas Research Network *et al.*, 2017). En plus des altérations communes à de nombreux sarcomes à génomique complexe, les LMS présentent des altérations récurrentes moins communes comme la perte du 10q contenant *PTEN* ou la délétion de *DMD* sur le chromosome X. Enfin les LMS ont une particularité génomique puisque entre 20 et 50% ont une amplification de *MYOCD* (Pérot *et al.*, 2009; The Cancer Genome Atlas Research Network *et al.*, 2017). En plus de ces altérations de structure, *ATRX* est l'un des trois gènes, avec *TP53* et *RB1*, les plus mutés des LMS (The Cancer Genome Atlas Research Network *et al.*, 2017). Ces mutations sont associées au phénotype ALT (Liau,

Tsai, *et al.*, 2015) comme présenté dans le paragraphe « **ATRX : le remodelleur de la chromatine** ». Cette caractéristique est commune aux sarcomes à génomique complexe mais les LMS ont une plus grande probabilité de présenter le phénotype ALT (The Cancer Genome Atlas Research Network *et al.*, 2017).

Les altérations de *RB1* et *TP53* et leurs conséquences ayant été présentées dans le paragraphe « **Les gènes du cycle cellulaire** », dans les prochains paragraphes ne seront décrites que les autres altérations les plus récurrentes des LMS : la perte de *PTEN*, la délétion de *DMD* et l'amplification de *MYOCD*.

2. La perte de PTEN

Dans les LMS, la perte de *PTEN* est retrouvée dans 13% des cas de manière homozygote et dans 68% de manière hétérozygote (The Cancer Genome Atlas Research Network *et al.*, 2017). Sa perte génomique est retrouvée dans différents sous-types de sarcome à génomique complexe mais est plus prévalente dans les LMS (**Figure 17** ; Stefano and Giovanni, 2019).

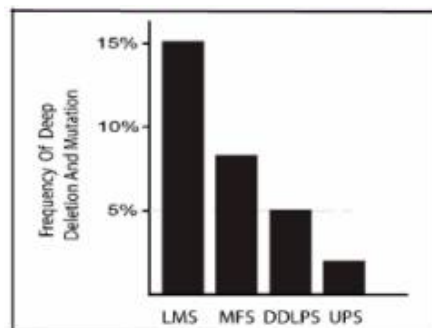


Figure 17 : Fréquence de la délétion profonde de *PTEN* dans différents types de sarcome (Stefano and Giovanni, 2019)

Dans un modèle murin ayant un *knockout* conditionnel de *PTEN* dans les cellules musculaires lisses, il y a une hyper-activation de la voie de signalisation d'AKT causant une progression maligne vers des LMS dans 80% des cas. Dans ces cas, la perte de *PTEN* était nécessaire mais non suffisante car la perte de p53 ou sa répression par MDM2 était présente dans l'ensemble des tumeurs contrairement aux lésions bénignes des muscles lisses (Hernando *et al.*, 2007).

Chez l'humain, dans la plupart des cas de LMS bien différenciés, la perte de

PTEN est aussi associée à l'activation constante d'AKT et donc de la voie PI3K/mTOR (Hernando *et al.*, 2007; Gibault *et al.*, 2012; Cuppens *et al.*, 2017). Dans ces LMS bien différenciés, il est aussi retrouvé une surexpression de *RICTOR*, ayant un rôle important dans la différenciation musculaire lisse par sa fonction d'organisateur du cytosquelette (Sarbasov *et al.*, 2006) et qui est un composant majeur du complexe mTOR capable de phosphoryler AKT (Gibault *et al.*, 2012). La voie PI3K/AKT/mTOR, lorsqu'elle est activée, est connue dans de nombreux cancers pour contrôler des voies cellulaires ayant un rôle majeur dans l'oncogenèse : le cycle cellulaire, la prolifération, la survie, la synthèse protéique, le métabolisme, la migration et la différenciation (Stiles *et al.*, 2002; O'Donnell *et al.*, 2018). Dans des modèles murins de LMS, l'activation de cette voie par la délétion génétique de *PTEN* est responsable de l'initiation de l'oncogenèse (Hernando *et al.*, 2007).

PTEN est un suppresseur de tumeur très étudié, se trouvant sur le chromosome 10 en 10q23. Il a un rôle prépondérant, au niveau cytoplasmique, dans la voie PI3K/AKT/mTOR de par sa fonction d'inhibition par déphosphorylation de PIP3, un activateur fort d'AKT. *PTEN* est aussi nucléaire de manière cycle-dépendant durant les phases G0-G1 (Ginn-Pease and Eng, 2003). La présence de *PTEN* au noyau diminue la phosphorylation de MAPK et le niveau de cycline D1, permettant ainsi l'arrêt du cycle en G0-G1 (Chung *et al.*, 2006). *PTEN* se lie aussi à CENP-C stabilisant les centromères et limitant ainsi les cassures centromériques. De plus, *PTEN* peut réguler, en se liant à la chromatine, le taux de Rad51 réduisant ainsi la quantité finale de cassures double brins (Shen *et al.*, 2007).

3. La perte de DMD

DMD ou dystrophine est un long gène de 2,5 Mb se situant sur le chromosome X en Xq21, il est composé de 79 exons et s'exprime sous la forme de 6 isoformes (Blake *et al.*, 2002). Dans les sarcomes à différenciation myogénique tels que les LMS, RMS et GIST, la perte de la plus grande isoforme de *DMD* est retrouvée dans 63% des cas de haut grade et plus spécifiquement dans 3 sur 7 LMS (**Figure 18** ; Wang *et al.*, 2014). De plus, la perte protéique de la grande isoforme est quasi-totale dans les GIST et les RMS

métastatiques mais de 62% dans les LMS métastatiques (Wang *et al.*, 2014).

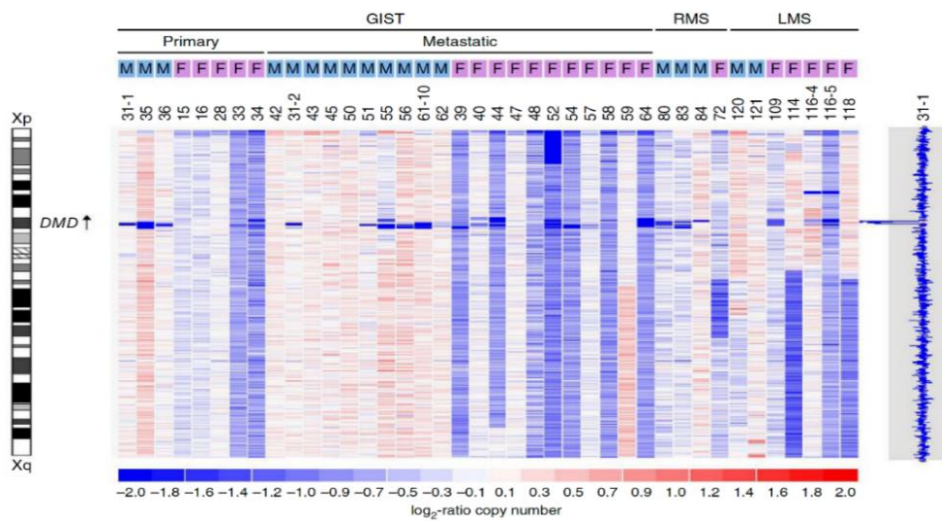


Figure 18 : Représentation du nombre de copies du chromosome X (Wang *et al.*, 2014)

Log₂ du ratio du nombre de copies, gain en rouge et perte en bleu, du chromosome X dans une cohorte de GIST, RMS et LMS montrant la perte spécifique de DMD

L'expression de la grande isoforme Dp427 est régulée par trois promoteurs tissu dépendant : Dp427b dans le cerveau, plus particulièrement dans les neurones corticaux et l'hippocampe, Dp427m dans les muscles et Dp427p dans les cellules de Purkinje (Barnea *et al.*, 1990; Chelly *et al.*, 1990). L'expression des autres isoformes de DMD, est dépendante de plusieurs promoteurs qui se situent à partir de l'intron 29 de Dp427. Ces isoformes sont Dp260, Dp140, Dp116, Dp71 et Dp40 ; chaque isoforme est nommée selon sa taille relative (Blake *et al.*, 2002). Dp71, qui n'est physiologiquement exprimée que dans les tissus non musculaires squelettiques (Hugnot *et al.*, 1992), a un rôle pro-prolifératif puisque son inhibition entraîne une diminution importante de la prolifération de lignées cellulaires (cancers de la glande surrénale du rat, carcinome pulmonaire, RMS, UPS, LPS et LMS) (Villarreal-Silva *et al.*, 2011; Wang *et al.*, 2014; Tan *et al.*, 2016; Mauduit *et al.*, 2019). Dans les sarcomes myogéniques, la perte de la grande isoforme de DMD, est liée à une délétion dans la partie 5' du gène qui ne touche jamais Dp71. Cette délétion entraîne la perte à la membrane de Dp427 et une relocalisation des autres isoformes, incluant Dp71, au noyau (Merle *et al.*, 2020).

La dystrophine, qui vient de la plus grande isoforme Dp427, fait partie du complexe dystrophine/glycoprotéine (DGC). C'est un grand complexe agissant comme un cadre afin de connecter le cytosquelette intracellulaire à la matrice extracellulaire environnante (Bhat *et al.*, 2018). Pour ce faire, la dystrophine est composée de quatre domaines fonctionnels : un domaine amino-terminal de liaison à l'actine, un domaine central rod, un domaine riche en cystéines et un domaine carboxy-terminal (Gao and McNally, 2015). Le domaine de liaison à l'actine se lie à l'actine F, liant la dystrophine au réseau subsarcolemmique de l'actine (Way *et al.*, 1992). Le domaine rod permet l'interaction avec les microtubules et est nécessaire à l'organisation du maillage des microtubules dans les cellules musculaires (Prins *et al.*, 2009). Le domaine riche en cystéines permet à la dystrophine d'interagir avec le β -dystroglycan en crant la dystrophine au sarcolemme (Rentschler *et al.*, 1999). Le reste du complexe est formé par de nombreuses protéines ayant des localisations cellulaires différentes : extracellulaire (α -dystroglycan), transmembranaire (β -dystroglycan, sarcoglycans, sarcospan) et cytoplasmique (dystrobrevin, syntrophins, synthase oxide nitrique neuronale)(Figure 19 ; Gao and McNally, 2015).

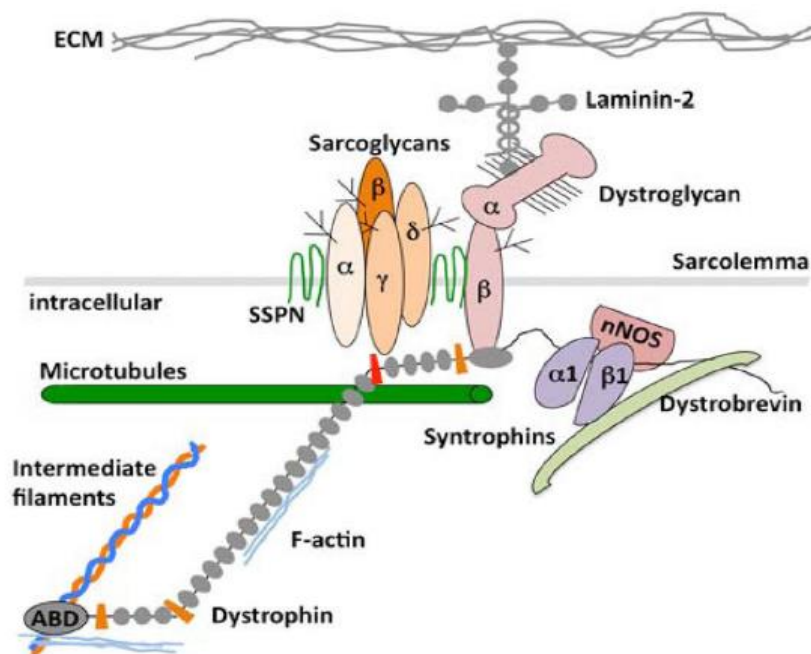


Figure 19 : Complexe Dystrophine/Glycoprotéine (Gao et McNally, 2015)

Ce complexe est une spécificité des cellules musculaires mais il est différent dans les muscles lisses comparé aux muscles squelettiques et cardiaques puisque dans ces cellules, α -dystroglycan est remplacé par ε -dystroglycan (Straub *et al.*, 1999). De plus, dans les cellules musculaires lisses vasculaires, *DMD* est plus fortement exprimé que dans les muscles striés et son expression dépend de la différenciation de la cellule : plus elle est différenciée plus *DMD* est exprimé. Dans ce contexte cellulaire, l'expression de *Dp427m* de *DMD* est régulée par deux facteurs de transcription spécifiques des cellules musculaires lisses : *MYOCD* et *MRTF-A* (Turczyńska *et al.*, 2015).

4. L'amplification de *MYOCD*

MYOCD ou *myocardin* se situe sur le chromosome 17 en 17p12 et est retrouvé fortement amplifié dans 20 à 50% des LMS (Figure 20 ; Pérot *et al.*, 2009; The Cancer Genome Atlas Research Network *et al.*, 2017). Ces LMS présentent une différenciation importante (Pérot *et al.*, 2009) et sont associés à une perte de *PTEN* plus fréquente ainsi qu'une activation constante de la voie *AKT* (Gibault *et al.*, 2012; The Cancer Genome Atlas Research Network *et al.*, 2017). Malgré cette association significative, actuellement aucune étude publiée ne montre de lien fonctionnel entre l'amplification de *MYOCD* et la perte de *PTEN* dans les LMS.

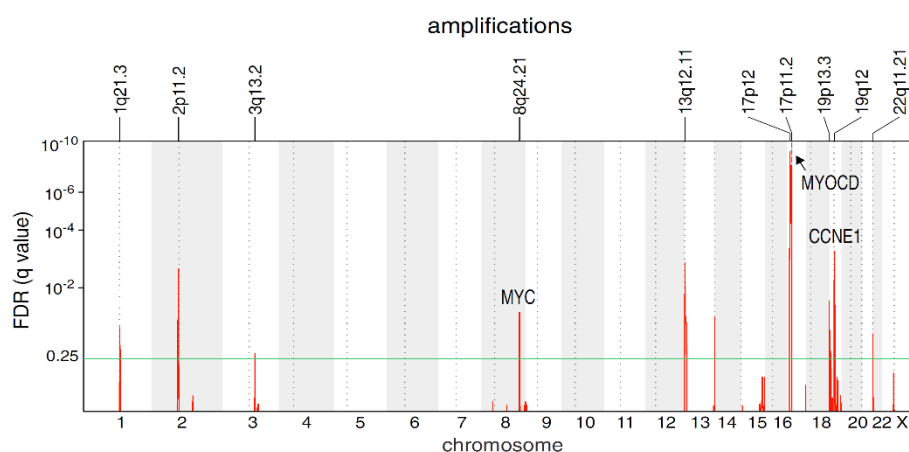


Figure 20 : Amplifications récurrentes dans les 80 LMS de la cohorte du TCGA (TCGA, 2017)

Dans la plupart des autres cancers, MYOCD est décrit comme un suppresseur de tumeur. En effet, il est retrouvé sous exprimé dans de nombreux cas comme dans les carcinomes naso-pharyngés (Chen *et al.*, 2011). De plus, il démontre des capacités indirectes d'induction de l'apoptose en sur-exprimant *Maspin*, un suppresseur de tumeur pouvant déclencher l'apoptose (Mahajan *et al.*, 2019) dans des cellules de cancer du sein (Liao *et al.*, 2014). Enfin, il a été démontré que MYOCD peut bloquer la prolifération cellulaire par l'inhibition de NF κ B (Tang *et al.*, 2008), la réduction de l'expression de la cycline D1 (Long *et al.*, 2008) et l'induction de MIR-1 (Chen Jie *et al.*, 2011).

MYOCD est le principal facteur de transcription du muscle cardiaque ainsi que des cellules musculaires lisses. En effet, la surexpression seule de cette protéine est suffisante pour trans-différencier des cellules déjà engagées dans un lignage défini comme des hépatocytes en cellules musculaires lisses (McDonald *et al.*, 2006). Cependant, MYOCD ne semble pas capable de se lier directement à l'ADN (Norman *et al.*, 1988), il joue donc son rôle en se fixant à des cofacteurs qui pourront se lier à des séquences spécifiques de l'ADN. Un des cofacteurs les plus importants de MYOCD est SRF, ensemble ils pourront former un complexe tertiaire se fixant aux CARG boxes et moduler l'expression de gènes spécifiques du muscle cardiaque ou des muscles lisses (Miano, 2015). MYOCD a aussi des rôles indépendants de SRF en se fixant notamment à MEF2C pour s'autoréguler (Creemers *et al.*, 2006).

5. Transcriptomes des léiomyosarcomes

L'analyse de l'expression des gènes touchés par ces altérations ainsi que de l'ensemble du génome permet une compréhension de la biologie des LMS plus approfondie. Lors d'analyses transcriptomiques de différents types de sarcome à génomique complexe des tissus mous, une partie des LMS se regroupent entre eux et l'autre partie reste mélangée aux autres sarcomes (**Figure 21** ; Gibault *et al.*, 2011).

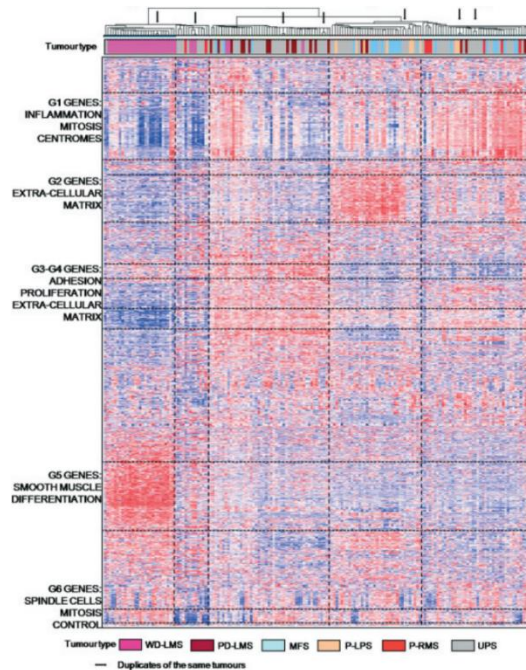


Figure 21 : Regroupement non supervisé de l'expression des gènes venant de différents types de sarcome à génomique complexe (Gibault *et al.*, 2011)

Ces deux comportements sont retrouvés lors de plusieurs analyses d'expression des gènes des LMS non utérins qui montrent deux sous-groupes (Ren *et al.*, 2003; Gibault *et al.*, 2011; Italiano *et al.*, 2013; Lee *et al.*, 2016; Chudasama *et al.*, 2018; Hemming *et al.*, 2020).

Le premier groupe exprime majoritairement des gènes liés aux muscles lisses, tels que *MYOCD*, *DMD*, *CALD1* et *MYL6* (Chudasama *et al.*, 2018). Dans ce groupe les LMS ont des altérations génomiques dites de bras, c'est-à-dire que la majorité des altérations touche l'ensemble du bras d'un chromosome. Les tumeurs présentent dans environ 40% des cas une amplification de *MYOCD* (Italiano *et al.*, 2013), et perdent aussi dans 80% des LMS le 10q (comportant *PTEN*) et le 13q (comportant *RB1*) (Gibault *et al.*, 2011). Les tumeurs de ce groupe sont plutôt de bas grade (grade 1 ou 2) (Lee *et al.*, 2016). Ils sont aussi plus différenciés, localisés dans le rétropéritoine à 64% et 78% sont chez des femmes (Gibault *et al.*, 2011; Italiano *et al.*, 2013).

Le second groupe surexprime des gènes du métabolisme, de la prolifération, de la matrice extracellulaire et de l'immunité (Gibault *et al.*, 2011; Hemming *et al.*, 2020).

Ces tumeurs présentent un profil génomique à haut niveau de complexité chromosomique, avec plus de 30 altérations par génome (Italiano *et al.*, 2013). Le second groupe est associé à des LMS peu différenciés, de grade III et situés dans les membres (Gibault *et al.*, 2011; Italiano *et al.*, 2013; Hemming *et al.*, 2020).

Aucun des deux groupes ne sont liés à une capacité plus importante que l'autre à produire des métastases (Italiano *et al.*, 2013; Chudasama *et al.*, 2018). Lorsque les LMS utérins sont ajoutés aux analyses, une troisième catégorie apparaît et exprime un mix des gènes entre les deux groupes précédents (Chudasama *et al.*, 2018; Hemming *et al.*, 2020).

Ces différents groupes avec des transcriptomes différents et des caractéristiques génomiques et cliniques spécifiques à chaque groupe laissent penser qu'il doit y avoir au moins deux oncogenèses différentes selon la localisation tumorale.

III. Physiologie de la cellule musculaire lisse

Les LMS sont des tumeurs dont le tissu d'origine n'est pas toujours déterminé. Ceux dont l'origine est connue proviennent de l'utérus ou des gros vaisseaux sanguins tels que la veine cave. Bien que ces tumeurs expriment des marqueurs des muscles lisses qui ne sont physiologiquement exprimés que par ces derniers, la cellule d'origine n'est actuellement pas déterminée. Compte tenu des marqueurs et de la composition des tissus de provenance connue, l'hypothèse principale est que les LMS proviennent de cellules musculaires lisses (SMC)(Fletcher *et al.*, 2013). La connaissance de la biologie physiologique des SMC devient donc nécessaire pour comprendre comment les LMS utilisent leur machinerie biologique au profit de l'oncogenèse.

1. Les muscles lisses et les cellules musculaires lisses

Les muscles lisses sont composés par un ensemble de cellules uniques, appelées cellules musculaires lisses ou léiomyocytes, et sont retrouvés dans l'ensemble du corps humain puisqu'ils sont localisés au niveau des organes creux, de la peau et des yeux. Plus spécifiquement, ils sont soit organisés en une grande unité unique dans les parois

de la vessie, de l'utérus et des voies digestives, soit en multi-unités dans les vaisseaux sanguins, les voies respiratoires, l'iris et sous la peau au niveau des racines des poils. Dans les deux cas, la contraction est involontaire, c'est-à-dire qu'elle est dirigée par le système nerveux autonome ou des hormones. Dans l'unité unique, les cellules sont attachées par des jonctions serrées dites gap permettant à toutes les cellules de se contracter en même temps puisque le signal se transmet de cellule en cellule. A l'inverse, dans les multi-unités le signal n'est pas transmis car les cellules le composant ne forment pas de jonction gap (Heyden *et al.*, 2013).

Toutes les cellules musculaires lisses ont une forme allongée avec un noyau unique et sont capables de produire leur matrice extracellulaire, l'endomysium. Leur cytoplasme contient de l'actine et de la myosine en filaments qui sont ancrés à la membrane, le sarcolemme, par des corps denses (Figure 22). La contraction est médiée par un afflux de Ca^{2+} qui se lie à la calmoduline déclenchant ainsi la contraction musculaire (Heyden *et al.*, 2013).

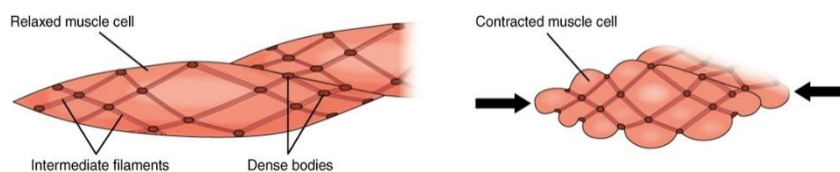


Figure 22 : Contraction musculaire des cellules musculaires lisses (Heyden *et al.*, 2013)

2. Des origines multiples

a. Le mésoderme

Les muscles lisses vasculaires viennent majoritairement du mésoderme mais les différentes sections du système vasculaire proviennent de plusieurs groupes de cellules (Figure 23 ; G. Wang *et al.*, 2015). On retrouve plus de cinq origines différentes pour les SMC vasculaires incluant la crête neurale (ectoderme), le pro-épicaarde, le mésothélium sérosal, le champ cardiaque secondaire et les somites (Xie Changqing *et al.*, 2011).

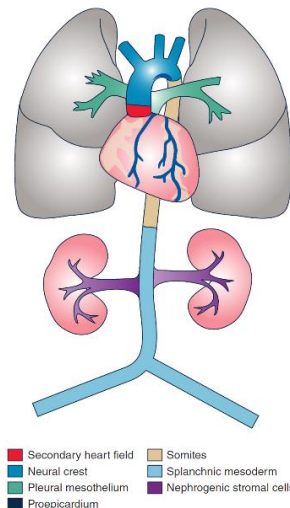


Figure 23 : Origine cellulaire des différentes parties du système vasculaire (Wang *et al.*, 2015)

L'aorte dorsale est composée de SMC provenant du mésoderme et plus précisément majoritairement du sclérotome (Pouget, Pottin and Jaffredo, 2008) mais aussi de quelques SMC venant des somites et partageant une même origine que des muscles striés (Esner *et al.*, 2006).

Les muscles lisses viscéraux dérivent du mésoderme viscéral et représentent une grande partie du système digestif. Les progéniteurs de cette couche se différencient à partir du mésenchyme entourant l'intestin primitif et tapissent l'ensemble du tube digestif de l'estomac au colon (McLin, Henning and Jamrich, 2009). Les léiomyoblastes composant ce feuillet se différencient rapidement en muscles lisses immatures durant l'embryogenèse mais leur différenciation finale n'aura lieu qu'après la naissance (McHugh, 1996).

Les muscles lisses des voies respiratoires se développent à partir de cellules mésenchymateuses localisées au bout des bourgeons épithéliaux en croissance. Ces cellules se déplacent graduellement d'une position distale à une position proximale le long de l'arbre bronchial, puis elles s'allongent et se différencient (Badri, Zhou and Schuger, 2008).

Le *detrusor*, le muscle lisse de la vessie, est formé à partir des cellules mésenchymateuses du mésoderme splanchnopleure (Rehman and Ahmed, 2020).

Les muscles *arrector pili*, les muscles lisses permettant l'érection des poils et se situant à la racine de ces derniers, se différencient à partir de cellules mésenchymateuses se situant au contact du renflement du futur bulbe du poil. Dans ce renflement se trouve des cellules souches qui déposent le long de la membrane basale de la nephronectine qui permettra aux cellules mésenchymateuses de se différencier en cellules musculaires lisses (Fujiwara *et al.*, 2011).

b. L'ectoderme

L'iris est composé des deux types de muscle lisse : dilatateur et sphincter. En 1903, Lewis décrit, chez le poulet, ces SMC comme provenant de la coupe optique neuroépithéliale et donc ayant une origine ectodermique (Lewis, 1903). Depuis, il a été démontré que ces cellules, durant leur développement, expriment à la fois des marqueurs neuro-épithéliaux et des marqueurs des muscles lisses (Jensen, 2005) validant ainsi la provenance ectodermique de ces SMC.

Une partie des SMC vasculaires viennent de la crête neurale pour donner l'arc branchial et se différencient grâce notamment au facteur SMAD2. Ces SMC vasculaires de l'arc branchial se différencient donc à partir de l'ectoderme contrairement à la majorité des cellules musculaires lisses vasculaires (Xie Wei-Bing *et al.*, 2013).

3. Spécificités des différentes cellules musculaires lisses

Malgré des points communs entre les différentes SMC tels que leur capacité de contraction et l'expression de marqueurs spécifiques de leur lignage, dès 1996 l'étude de l'expression tissu dépendant du gène SM22 α (ou transgéline) montre qu'il existe des programmes de régulation transcriptionnelle différents entre les différents types de SMC (Li *et al.*, 1996). De plus, l' α -actine du muscle lisse (α -SM actine) est exprimée dans les SMC viscérales dédifférenciées ou non-différenciées mais pas dans les différenciées, alors que cette protéine est exprimée à l'opposé dans les SMC vasculaires du poulet (Saga *et al.*, 1999). Il existe aussi une isoforme spécifique des SMC viscérales de la chaîne lourde de la myosine des muscles lisses (Babij, 1993).

Avec l'avènement des techniques d'analyse de l'expression des gènes à grande

échelle, 36 lignées de SMC purifiées selon leur tissu d'origine ont été analysées par puce d'expression et les SMC avaient effectivement un programme transcriptionnel différent selon leur tissu d'origine (Chi *et al.*, 2007). L'analyse globale de l'expression de 16 352 gènes donne deux grands groupes : les SMC viscérales, comprenant trois sous-groupes : le tractus urinaire, les organes reproducteurs féminins et le colon, et les SMC vasculaires. Les SMC viscérales expriment préférentiellement les gènes *MITF* (facteur associé à la microphthalmie), inhibant *TGF- β* , et *HNMT* (Histamine N-méthyltransférase, régulateur de l'histamine), un agoniste de la contraction des muscles lisses, ainsi que les voies de l'insuline et du glutathione. Les gènes préférentiellement exprimés par les SMC vasculaires sont dans la voie *TGF- β* , affectant la prolifération, la migration et la production de la matrice extracellulaire (Kanai Hiroyoshi *et al.*, 2001), mais aussi des gènes impliqués dans la réponse inflammatoire, dans les signaux de communication avec les cellules endothéliales, la voie *IL1R* (récepteur interleukine 1), la voie *TNF* (facteur de nécrose tumorale) et la voie *HIF* (facteur induit par l'hypoxie). Les différents types de SMC vasculaires : artères, veines et bronchiques expriment elles aussi préférentiellement certains gènes selon leur type : les artères expriment des gènes régulateurs du système immunitaire et les bronchiques des gènes régulateurs des poumons (Chi *et al.*, 2007).

4. Plasticité de la différenciation

Les cellules musculaires lisses font partie des cellules du corps humain ayant une très grande plasticité (Campbell Julie H. and Campbell Gordon R., 2012). Cela signifie qu'elles peuvent passer d'un stade différencié avec la capacité de se contracter, état appelé SMC contractile, à un stade moins différencié avec des capacités de prolifération et de migration, état appelé SMC synthétique. Dans les conditions physiologiques normales, les SMC ont un phénotype contractile mais sous l'effet de différents *stimuli* comme des facteurs de croissance, des médiateurs de l'inflammation ou des contraintes mécaniques elles perdent leur capacité de contraction et peuvent migrer et proliférer (House *et al.*, 2008). Cette capacité est appelé le *switch* phénotypique. Il a lieu dans des conditions physiologiques pour la réparation des vaisseaux sanguins mais aussi la création de nouveaux vaisseaux lors d'une grossesse

ou d'exercices sur le long terme et dans des conditions pathologiques comme dans l'angiogenèse tumorale (Rensen, Doevendans and van Eys, 2007). Ce *switch* a été extensivement étudié dans les SMC vasculaires mais l'ensemble des SMC en sont capables bien qu'il semblerait être moins présent dans les autres SMC (il est possible que cet effet soit dû à un manque d'étude du *switch* et de ses conséquences dans les SMC non vasculaires)(Scirocco *et al.*, 2016). Les LMS provenant majoritairement de vaisseaux sanguins, les prochains paragraphes décriront le *switch* seulement dans les SMC vasculaires.

a. Capacité de dédifférenciation

Sur le plan morphologique, les SMC contractiles ont la forme caractéristique des cellules de muscle lisse, c'est-à-dire allongées en fuseau et présentant des fibres de stress d'actine ainsi que des faisceaux de myofilaments leur conférant leur capacité contractile. A l'inverse, les SMC synthétiques ont une morphologie plus ronde mentionnées comme étant épithélioïdes ou rhabdoïdes (**Figure 24** ; Campbell and Campbell, 1994). Les SMC synthétiques contiennent un grand nombre d'organelles impliquées dans la synthèse protéique et expriment des gènes impliqués dans la formation de matrice indiquant ainsi que c'est sous cette forme qu'elles produisent de la matrice extracellulaire (Hao Hiroyuki, Gabbiani Giulio and Bochaton-Piallat Marie-Luce, 2003; Rensen, Doevendans and van Eys, 2007; Shinohara *et al.*, 2012).

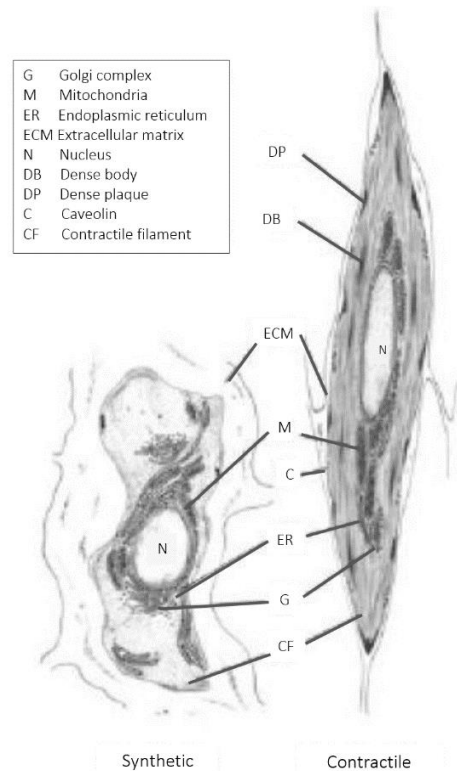


Figure 24 : Ultrastructure des SMC contractiles et synthétiques (Rensen *et al.*,2007)

L'expression de marqueurs des SMC dans ces deux états a été caractérisée : actine musculaire lisse α et β , la chaîne lourde de la myosine des muscles lisses (SM-MHC), la calponine, la h-caldesmon. Ces différents marqueurs sont exprimés fortement dans le phénotype contractile et diminuent lors du *switch* (Miano, 2003). De plus, parmi eux, SM-MHC est le marqueur le plus spécifique des muscles lisses contractiles. SM-MHC n'a jamais été détectée dans des cellules qui ne sont pas des SMC et est le seul marqueur spécifique des SMC durant l'embryogenèse (Miano J M *et al.*, 1994). Le *switch* phénotypique correspond à une perte de l'expression des gènes de la contraction musculaire et d'une augmentation des gènes liés à la prolifération, la migration et la synthèse des composants de la matrice extracellulaire (Rzucidlo, Martin and Powell, 2007).

Il existe aussi certains facteurs qui peuvent être exprimés dans les deux états des SMC et avoir des rôles différents selon l'état de la cellule. Par exemple, IGF-1 permet la maintenance du phénotype contractile des SMC allongées mais aussi celle du phénotype synthétique des SMC rondes (Hayashi *et al.*, 2004).

Les deux phénotypes se retrouvent mélangés dans les vaisseaux sanguins.

Même lors d'une blessure les réponses des SMC du même vaisseau sont hétérogènes et la réexpression des marqueurs des SMC contractiles n'est pas uniforme mais vient de sous-populations de cellule (Regan *et al.*, 2000). En effet, il existe des populations distinctes de SMC dans une même artère (Li *et al.*, 2001).

Ce changement est régulé par de nombreux facteurs dont les plus importants sont les cofacteurs SRF et MYOCD qui sont à l'origine de la régulation de l'expression des gènes de la différenciation en muscle lisse contractile et dont le fonctionnement est détaillé dans le paragraphe « **La voie de la différenciation SRF/MYOCD** ». Les autres facteurs peuvent être biochimiques, extracellulaires ou physiques (Rensen, Doevendans and van Eys, 2007). Les facteurs biochimiques les plus étudiés et dont les rôles dans le *switch* sont bien connus sont PDGF, qui induit le phénotype synthétique (X. Li *et al.*, 1997; Hao *et al.*, 2002) et TGF- β , essentiel à l'induction du phénotype contractile (Hautmann, Madsen and Owens, 1997). La matrice extracellulaire peut promouvoir l'un ou l'autre des phénotypes selon sa composition et les intégrines auxquelles se lient les composants. Par exemple, le collagène I fibrillaire induit le phénotype contractile alors que le collagène I monomérique active la prolifération (Ichii *et al.*, 2001). Enfin les facteurs physiques à l'origine du *switch* sont les contraintes de cisaillement retrouvées dans les vaisseaux et activent l'un ou l'autre des phénotypes (Rensen, Doevendans and van Eys, 2007).

b. Des capacités plastiques au-delà du phénotype musculaire lisse

L'étude des plaques de l'artériosclérose, une maladie inflammatoire chronique des vaisseaux sanguins, a démontré l'étendue des capacités plastiques des SMC. En effet, il a longtemps été pensé que la plaque se composait d'une fine capsule fibreuse comprenant des cellules exprimant ACTA2 présumées être des muscles lisses vasculaires, d'un grand nombre de cellules positives au CD68 ou au LGALS3 supposées être des macrophages et d'un corps nécrotique contenant des cellules pleines de lipides supposées également être des macrophages (Bennett, Sinha and Owens, 2016). Les travaux de lignage cellulaire et de RNAseq en cellule unique au sein de ces plaques ont démontré que 40 à 70% des cellules de la plaque sont des SMC. Pour autant jusqu'à 80% des cellules dérivées des SMC à l'intérieur de la plaque n'expriment

pas des niveaux détectables de marqueurs des SMC comme MYH11 et ACTA2 (Liu Mingjun and Gomez Delphine, 2019). En plus de perdre leurs marqueurs des muscles lisses, ces SMC subissent une transition phénotypique vers différents types cellulaires (**Figure 25**) : des ostéochondrocytes-like (Speer Mei Y. *et al.*, 2009), des macrophages-like (Feil Susanne *et al.*, 2014), des cellules souches mésenchymateuses-like (Shankman *et al.*, 2015), des myofibroblastes (Majesky *et al.*, 2017) ou encore des cellules adipocytaires beige-like (Long *et al.*, 2014).

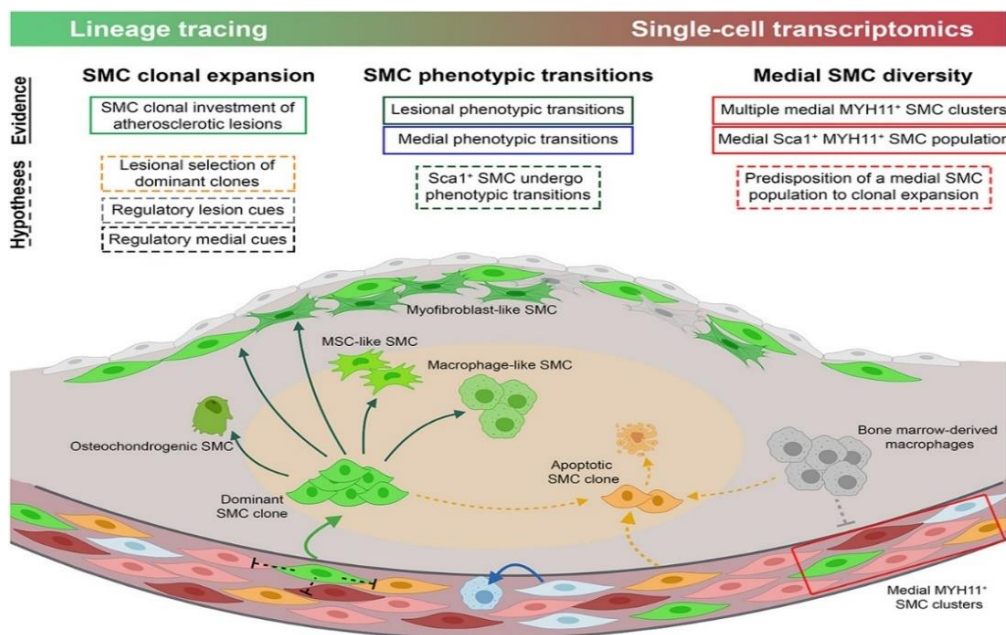


Figure 25 : Trans-différenciation des SMC dans la plaque artériosclérotique (Liu and Gomez, 2019)

Pour déterminer quelles voies sont impliquées dans ces différentes trans-différenciations, ces études de lignage cellulaire ont aussi été couplées à des modèles de suppression de gènes d'intérêt des SMC. Il a été identifié comme déterminant dans ces changements phénotypiques des facteurs ayant des rôles dans la régulation phénotypique des SMC : KLF4, inhibiteur de l'expression des marqueurs des SMC (Shankman *et al.*, 2015), TAGLN, un composant du cytosquelette des SMC contractiles (Feil Susanne, Hofmann Franz and Feil Robert, 2004) ou PTEN, un inhibiteur des SMC synthétiques (Nemenoff Raphael A. *et al.*, 2011).

5. Les voies de régulation de la différenciation

Le *switch* phénotypique permet physiologiquement notamment le maintien du tonus vasculaire et la réparation vasculaire (Rensen, Doevendans and van Eys, 2007) mais un dysfonctionnement dans celui-ci entraîne de nombreuses maladies comme l'artériosclérose, l'hypertension, l'asthme ou l'oncogenèse en participant à l'angiogenèse (Owens, Kumar and Wamhoff, 2004). La régulation de ce phénomène est donc essentielle et complexe.

a. La voie de la différenciation SRF/MYOCD

Lors de la découverte de gènes exclusifs aux SMC, la présence dans les promoteurs d'un élément de 10 pb appelé boîte CArG a été mise en évidence. Ces boîtes, composées de la séquence [CC(A/T)₆GG], servent de points de liaison à l'ADN pour le facteur de réponse au sérum (SRF), un facteur de transcription ubiquitaire, qui active la transcription de gènes impliqués dans la différenciation et la prolifération musculaire. L'association de SRF aux boîtes CArG est nécessaire à l'activation des gènes des SMC (Miano, 2003). Parmi les gènes avec des boîtes CArG dont l'expression est dépendante de SRF on retrouve : *MYH11* (SM-MHC)(Manabe and Owens, 2001), *ACTA2* (Mack Christopher P. and Owens Gary K., 1999), *TAGLN* (L. Li *et al.*, 1997), *DMD* (Gilgenkrantz *et al.*, 1992), *CALD1* (Momiyama *et al.*, 1998). Afin que cette activation ait lieu la chromatine doit être intacte et la méthylation/acétylation des sites est nécessaire (Manabe Ichiro and Owens Gary K., 2001; Qiu Ping and Li Li, 2002). Les histones autour des boîtes CArG des SMC sont spécifiquement marquées par H3K4dMe et H4Ac, or ce n'est pas SRF qui interagit avec ces marques mais MYOCD va interagir avec la marque H3K4dMe et sélectivement augmente ainsi la liaison de SRF aux boîtes CarG des gènes des SMC (McDonald *et al.*, 2006). MYOCD est le cofacteur de SRF, restreint aux muscles lisses, qui forme un complexe tertiaire avec SRF afin d'activer l'expression des gènes nécessaires aux SMC contractiles (Miano, 2003). Récemment, un troisième acteur de la spécificité de SRF dans la liaison aux boîtes CArG spécifiques des muscles lisses a été découvert. Il s'agit de PTEN (**Figure 26** ; Horita *et al.*, 2016), qui est capable de se lier, dans le noyau, à SRF et qui stabilise ce dernier uniquement sur les boîtes CArG des gènes des muscles lisses. De plus,

PTEN est transloqué avec SRF, dans le cytoplasme lors d'un *switch* phénotypique séquestrant ainsi SRF hors du noyau (Figure 26 ; Horita *et al.*, 2016).

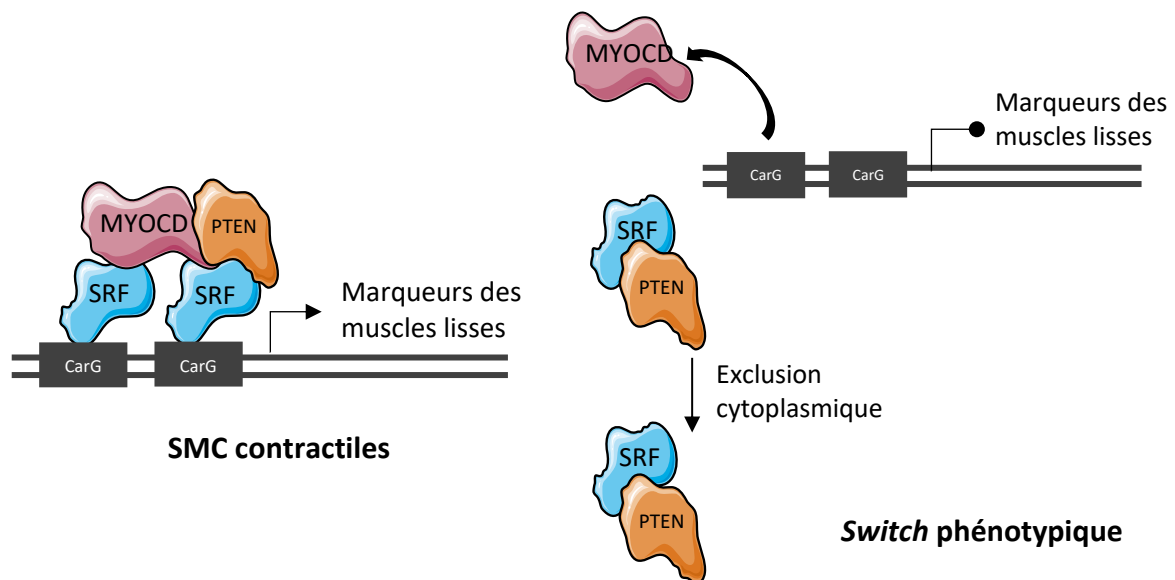


Figure 26 : Représentation du complexe MYOCD/SRF/PTEN dans les SMC contractiles et lors du *switch* phénotypique (adapté de Horita *et al.*, 2016)

La modulation phénotypique des SMC par l'axe MYOCD/SRF, quand elle est non régulée, peut devenir pathologique et peut entraîner l'augmentation de la rigidité des vaisseaux sanguins (Saphirstein and Morgan, 2014; Zhou *et al.*, 2017). Cette rigidité est une cause de l'hypertension (Townsend Raymond R. *et al.*, 2015) et peut être limitée par l'inhibition de l'interaction entre SRF et MYOCD. Pour ce faire, un inhibiteur de cette interaction, appelé CCG-100602, a été développé et testé chez le rat qui, après traitement, montre une forte diminution de la rigidité de l'aorte (Zhou *et al.*, 2017).

b. La dédifférenciation : l'inhibition de SRF/MYOCD

Le *switch* phénotypique nécessite des voies antagonistes à celle de MYOCD. Tout d'abord, FOXO4 et FOXO3A se lient à MYOCD et forment un complexe inhibant la transcription des gènes dépendants de CARG (Liu *et al.*, 2005; Yang *et al.*, 2013). Ces liaisons sont, elles-mêmes, régulées par AKT qui, lorsqu'il est activé, phosphoryle les facteurs de transcription FOXO les excluant du noyau (Yang *et al.*, 2013). En amont du promoteur de *MYOCD* il y a une région répresseur contenant 3 sites de liaison de KLF4 qui s'y lie pour réprimer *MYOCD* (Turner *et al.*, 2013). Cela aura pour effet de diminuer l'expression des marqueurs des muscles lisses (Liu, Sinha and Owens, 2003). p53 peut

aussi réprimer de façon dose dépendante un promoteur humain de *MYOCD* de 1 kb (Molchadsky *et al.*, 2008).

Il existe aussi une famille de cofacteurs du SRF qui sont en compétition directe avec *MYOCD* pour diriger SRF sur d'autres promoteurs : c'est la famille des TCF (complexe facteur tertiaire) parmi laquelle on retrouve Elk-1 (Gualdrini *et al.*, 2016). Quand SRF est lié à un TCF et plus particulièrement à Elk-1, cela promeut l'expression de gènes liés à la prolifération cellulaire, au contrôle du cycle cellulaire, à la réplication des chromosomes et à la ségrégation (Gualdrini *et al.*, 2016). Elk-1 est à la fois en compétition avec *MYOCD* pour la liaison à SRF puisqu'ils partagent le même site d'ancrage (Wang *et al.*, 2004) et permet aussi à SRF de se placer sur les sites régulateurs des gènes de la prolifération comme *c-fos* (Marais, Wynne and Treisman, 1993).

c. Les miRNA de la régulation du *switch* phénotypique

Il existe un ensemble de miRNA capables de réguler l'expression de plusieurs gènes inclus dans la régulation du *switch* phénotypique, généralement impliqués dans les mêmes voies cellulaires (Parmacek, 2009). Certains miRNA inhibent les SMC prolifératives, d'autres activent les SMC prolifératives (Wang and Atanasov, 2019).

Les miRNA qui permettent l'inhibition des SMC prolifératives et donc autorisent la différenciation en SMC contractiles sont nombreux (**Figure 27**) mais les trois plus connus et plus étudiés sont mirRNA-143, 145 et 1. miRNA-143 et 145 sont des cibles directes de SRF/*MYOCD* et coopèrent pour cibler un réseau de facteurs de transcription incluant KLF4, Elk-1 et *MYOCD* inhibant ainsi la prolifération et activant la différenciation (Cordes *et al.*, 2009). Plus spécifiquement miRNA-145-5p réduit l'expression de SMAD4, un transducteur de signal pro-prolifératif, et augmente celle de TGF- β (Li *et al.*, 2018) qui ont tous les deux un rôle déterminant dans la différenciation des SMC au niveau embryologique (Mao *et al.*, 2012). miRNA-1 qui est directement régulé par *MYOCD*, peut cibler Pim-1, une sérine/thréonine kinase activant les SMC prolifératives, permettant ainsi l'inhibition de la prolifération des SMC (Chen Jie *et al.*, 2011).

D'un autre côté, de nombreux miRNA inhibent l'expression des gènes contractiles et influencent le *switch* vers la prolifération (Figure 27). Parmi les plus étudiés, miRNA146b-5p est nécessaire au *switch* phénotypique des SMC vasculaires vers la prolifération probablement par activation de KLF4 qui comporte une cible de miRNA-146 (Sun *et al.*, 2011; H. Wang *et al.*, 2015). Il y a aussi miRNA675 qui cible directement PTEN et en réduire l'expression, augmentant ainsi la prolifération de SMC (Lv *et al.*, 2018).

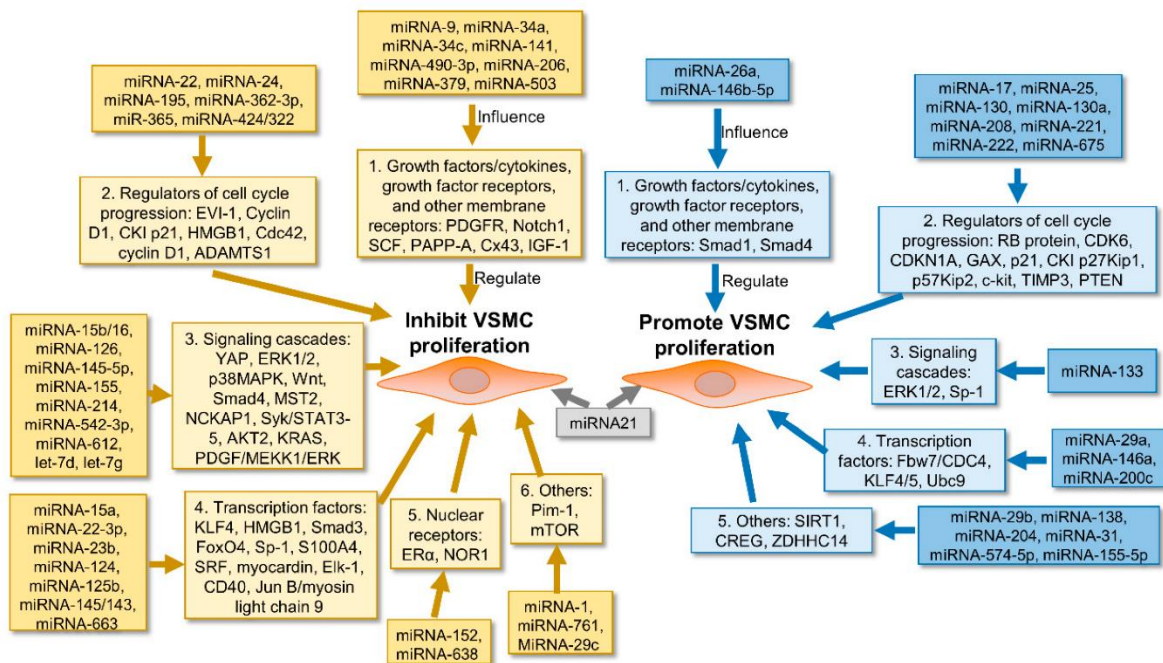


Figure 27 : Rôle des miRNA dans la régulation de la prolifération des SMC vasculaires (Wang and Atanasov, 2018)

D. Objectifs

La compréhension de la biologie des LMS se constitue actuellement par des analyses bio-informatiques de séquençage génomique ou transcriptomique, par des validations immunohistochimiques de l'expression de protéines codées par les gènes ayant des altérations récurrentes comme *TP53*, *RB1*, *ATRX* et *PTEN*, ainsi que par des analyses fonctionnelles spécifiques au rôle de chacun de ces gènes et de *MYOCD* et *DMD*. Ces différentes études permettent d'avancer dans la connaissance de l'oncogenèse des LMS mais il n'existe pas, à ce jour, d'études proposant une modélisation de l'oncogenèse des LMS ou proposant une cellule d'origine pour les LMS. En prenant en compte les différents OMICS, les données cliniques et des validations fonctionnelles, cela permettrait une compréhension globale de l'oncogenèse des LMS et le développement de traitement(s) efficace(s). Afin de séquencer au niveau génomique et transcriptomique des LMS, mon équipe est entrée dans l'*International Cancer Genome Consortium* (ICGC), l'un des deux grands programmes de séquençage massif de tumeur. Grâce à l'ICGC et avec l'aide du Groupe Sarcome Français, nous avons pu constituer une cohorte de 120 patients ayant un LMS. Pour chaque patient, l'ADN tumoral, l'ADN constitutionnel (venant du sang), l'ARN tumoral et les micro ARN tumoraux ont été séquencés. Après un travail bio-informatique d'alignement et de nettoyage des données, nous avons analysé ces résultats.

L'analyse de ces *data* a demandé en amont de mon travail de thèse une importante analyse bio-informatique à laquelle j'ai participé et qui fait l'objet de la première partie de mes résultats. En effet, cette analyse et les validations techniques que nous avons menées nous ont permis de définir solidement les deux objectifs majeurs de ma thèse :

Le premier objectif, était de comprendre les rôles des altérations du troisième gène le plus muté dans les LMS, *ATRX*, dans leur oncogenèse en plus de son intervention dans le système ALT.

Le deuxième objectif était de caractériser les différents acteurs de la voie principale de l'oncogenèse du groupe transcriptomique le plus homogène et de cibler cette voie avec un inhibiteur spécifique afin d'ouvrir la voie au développement d'un traitement efficace des LMS.

Partie 2 : Résultats

A. Comprendre l'oncogenèse des LMS grâce aux « OMICS »

I. Contexte

Depuis 2002 et la première étude non supervisée de l'expression des gènes dans les sarcomes des tissus mous (Nielsen *et al.*, 2002), il est connu que les LMS se séparent en deux groupes aux comportements transcriptomiques différents. Un premier groupe se mélange aux autres sarcomes à génomique complexe et un second groupe qui présente une transcriptomique homogène se regroupant entre eux comme les sarcomes à *driver* fort. Par la suite des caractéristiques cliniques et des types d'altérations chromosomiques ont été associés à chaque groupe. Enfin le premier groupe a été lié à l'expression de gènes de l'immunité et le second à l'expression de gènes associés aux muscles lisses (Gibault *et al.*, 2011; Italiano *et al.*, 2013; Chudasama *et al.*, 2018; Hemming *et al.*, 2020). Ces différentes études aident à la compréhension de l'oncogenèse des LMS mais aucune cartographie globale du fonctionnement des LMS n'est disponible à ce jour.

Afin de caractériser les spécificités de chacun pour mieux comprendre l'oncogenèse de ces deux groupes de LMS, nous proposons un travail d'intégration bio-informatique de données transcriptomiques, génomiques, de micro ARN, de méthylome et de données cliniques avec des validations fonctionnelles afin de comprendre l'oncogenèse des deux groupes de LMS et d'en extraire des cibles thérapeutiques.

II. Article 1 : *SRF-MYOCD axis is the targetable driver of a well differentiated and aggressive subtype of Leiomyosarcomas*

Elodie Darbo, Gaëlle Pérot, **Lucie Darmusey**, Sophie Le Guellec, Laura Leroy, Laëtitia Gaston, Nelly Desplat, Noémie Thébault, Candice Merle, Philippe Rochaix, Thibaud Valentin, Gwenaël Ferron, Christine Chevreau, Binh Bui, Eberhard Stoeckle, Dominique Ranchere-Vince, Pierre Mééus, Philippe Terrier, Sophie Piperno-Neumann, Françoise Colin, Gonzague De Pinieux, Florence Duffaud, Jean-Michel Coindre, Jean-Yves Blay, Frédéric Chibon

Soumis à Cancer Discovery, Octobre 2020

Résumé

Dans cet article, nous avons analysé l'expression des gènes de 555 sarcomes dont 98 LMS. Parmi ces LMS, 41 avaient un transcriptome homogène et ont été appelés hLMS pour *homogenous* LMS. Le différentiel d'expression entre les hLMS et les autres LMS nommées oLMS pour *other* LMS a permis de créer une signature de 1672 gènes discriminant les deux groupes. Cette signature a ensuite été appliquée sur la première phase de notre cohorte de LMS comportant 69 cas et sur la cohorte du TCGA comportant 102 LMS. L'analyse des caractéristiques cliniques montre que les hLMS sont associés à une localisation intra-abdominale, aux femmes, à une différenciation importante et à un grade faible 1 ou 2, mais ont un plus mauvais pronostic que les oLMS. L'expression des gènes et des micro ARN ainsi que les marques de méthylation des histones indiquent que les hLMS proviennent très probablement des SMC vasculaires, tout en exprimant en même temps les gènes des phénotypes contractiles et synthétiques. De plus, l'analyse des génomes et les analyses de l'expression des gènes au niveau ARN et/ou protéique montrent des récurrences dans les hLMS, ce qui a permis de démontrer que la voie SRF/MYOCD était au cœur de l'oncogenèse des hLMS. De plus, cette voie est validée par la sensibilité plus importante d'une lignée cellulaire de hLMS à un inhibiteur spécifique de l'interaction SRF/MYOCD par rapport à des lignées cellulaires de oLMS et d'UPS.

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

Elodie Darbo^{1,2,3}, Gaëlle Pérot^{4,5}, Lucie Darmusey^{4,6,7}, Sophie Le Guellec^{4,6}, Laura Leroy^{4,6}, Laëtitia Gaston⁸, Nelly Desplat¹, Noémie Thébault^{4,6}, Candice Merle^{4,6,7}, Philippe Rochaix^{4,6}, Thibaud Valentin^{4,9}, Gwenaël Ferron^{4,10}, Christine Chevreau⁹, Binh Bui¹¹, Eberhard Stoeckle¹², Dominique Ranchere-Vince¹³, Pierre Mééus¹⁴, Philippe Terrier¹⁵, Sophie Piperno-Neumann¹⁶, Françoise Colin¹⁷, Gonzague De Pinieux¹⁸, Florence Duffaud¹⁹, Jean-Michel Coindre^{1,20}, Jean-Yves Blay^{21,22}, Frédéric Chibon^{4,6}

ABSTRACT

Based on the transcriptome analysis of 555 sarcomas, we identified a group of tightly clustered leiomyosarcomas (LMS) due to their gene expression homogeneity. We named them “hLMS” and the other LMS “oLMS”. We derived a transcriptional signature able to identify each group and used it to classify 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. Based on integrative bioinformatic analysis, we show that hLMS originate from vascular smooth muscle cells presenting both contractile and synthetic characteristics, while oLMS could derive from fibroblasts. We found strong *MYOCD* expression to be an hLMS-specific driver and show that the *MYOCD*/SRF axis is essential only for hLMS survival. Identification of hLMS could become standard clinical practice, leading to the development of specific effective treatments with *MYOCD*/SRF inhibitors.

SIGNIFICANCE: Leiomyosarcomas (LMS) are currently treated as a single entity. However, we have now identified a transcriptionally, genomically and clinically homogeneous subgroup of LMS. Their oncogenesis is driven by the acquisition of high differentiation through *MYOCD* over-expression. This confers them sensitivity to *MYOCD*/SRF inhibitors, which could thus become a potential therapeutic target.

1. INSERM UMR1218, ACTION, Institut Bergonié, Bordeaux, France. 2. CNRS UMR5800, LaBRI, Talence, France. 3. Université de Bordeaux, Bordeaux, France. 4. OncoSarc, INSERM U1037, Cancer Research Center in Toulouse (CRCT), Toulouse, France. 5. Centre Hospitalier Universitaire (CHU) de Toulouse, IUCT-Oncopole, Toulouse, France. 6. Department of Pathology, Institut Claudius Régaud, IUCT-Oncopole, Toulouse, France. 7. University of Toulouse 3, Toulouse, France. 8. Department of Medical Genetics, CHU de Bordeaux, Bordeaux, France. 9. Department of Oncology, Institut Claudius Régaud, IUCT-Oncopole, Toulouse, France. 10. Department of Surgical Oncology, Institut Claudius Régaud, IUCT-Oncopole, Toulouse, France. 11. Department of Oncology, Institut Bergonié, Bordeaux, France. 12. Department of Surgery, Institut Bergonié, Bordeaux, France. 13. Department of Pathology, Centre Léon Bérard, Lyon, France. 14. Department of Surgery,

Centre Léon Bérard, Lyon, France. 15. Department of Pathology, Institut Gustave Roussy, Villejuif, France. 16. Department of Medical Oncology, Institut Curie, Paris, France. 17. Department of Pathology, Centre Georges-François Leclerc, Dijon, France. 18. Department of Pathology, Hôpital Universitaire Trousseau, Tours, France. 19. Medical Oncology Unit, Aix Marseille University, APHM Hôpital La Timone, Marseille, France. 20. Department of Pathology, Institut Bergonié, Bordeaux, France. 21. Department of Medical Oncology, Centre Léon Bérard, Lyon, France. 22. Université Claude Bernard Lyon 1, INSERM U1052, CNRS 5286, Centre Léon Bérard, Lyon, France.

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

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) (Supplementary Figure S1A, total = 555 samples). Using the strategy described in the methods and illustrated in Supplementary Figure S1C, we selected 9066 genes (out of 17854 genes common to both platforms) showing enough consistency to enable merging and normalization of all datasets. 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

Feature	Test	Affymetrix			ICGC + TCGA		
		hLMS	oLMS	p-value	hLMS	oLMS	p-value
Differentiation (%)	Well (vs Poor)	88	24	3.87x10 ⁻⁹ (+)	84	41	1.83x10 ⁻⁵ (+)
Grade (%)	Low (vs High)	58	24	6.51x10 ⁻⁴ (+)			
Sex (%)	F (vs M)	76	48	0.003 (+)	68	41	0.007 (+)
Location (%)	Internal trunk (vs other)	60	7	8.48x10 ⁻⁹ (+)	82	27	1.53x10 ⁻⁷ (+)
Mitotic counts (median)	Ranks	17	24.5	0.009 (-)	11	35	0.0006 (-)
Gene expression variance (median)	Ranks	0.7	1	2.90x10 ⁻¹³ (-)	ICGC		
					0.25	0.36	<2.2x10 ⁻¹⁶ (-)
					TCGA		
					0.45	0.9	<2.2x10 ⁻¹⁶ (-)

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.

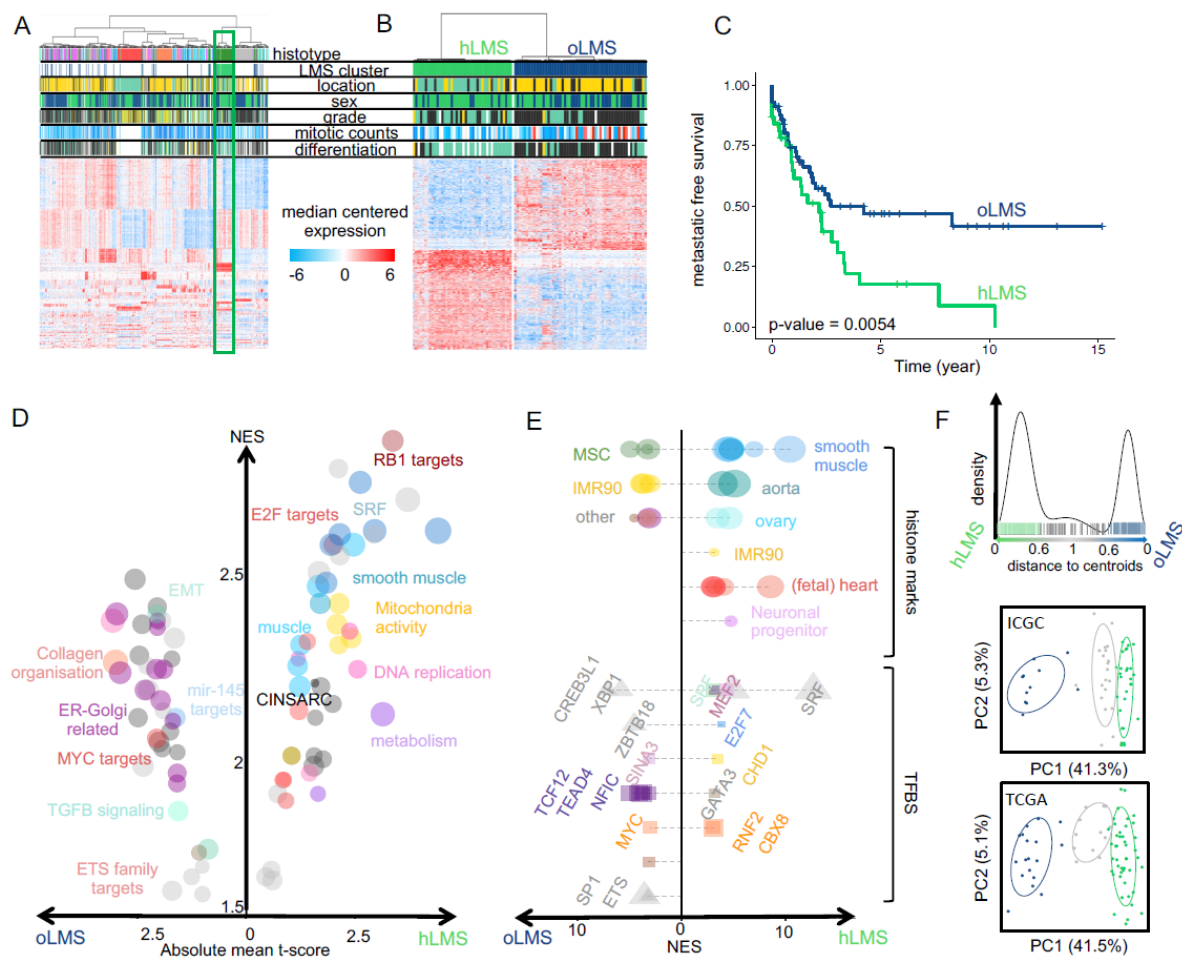


Figure 1: Transcriptional analysis and patient classification **A-B** Heatmap showing clustering of 555 sarcoma patients and 455 genes (A) and 98 LMS patients and 1672 differentially expressed genes between hLMS and oLMS (B). Patients were clustered using HCPC method and genes are grouped by co-expression 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. **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 z-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 Supplementary 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.

gene networks. 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, Supplementary 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 (Supplementary 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 Supplementary 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 *RBI* 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 (Supplementary Table S1D, Figure 1E). On the other hand, oLMS were associated with ER-Golgi related terms, epithelial-mesenchymal transition and the TGF β 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*, *ETSI* 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 (Supplementary 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 (Supplementary Figure S2B) using GSA_n (19) (Supplementary 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 (*FNI*, *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 Supplementary 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

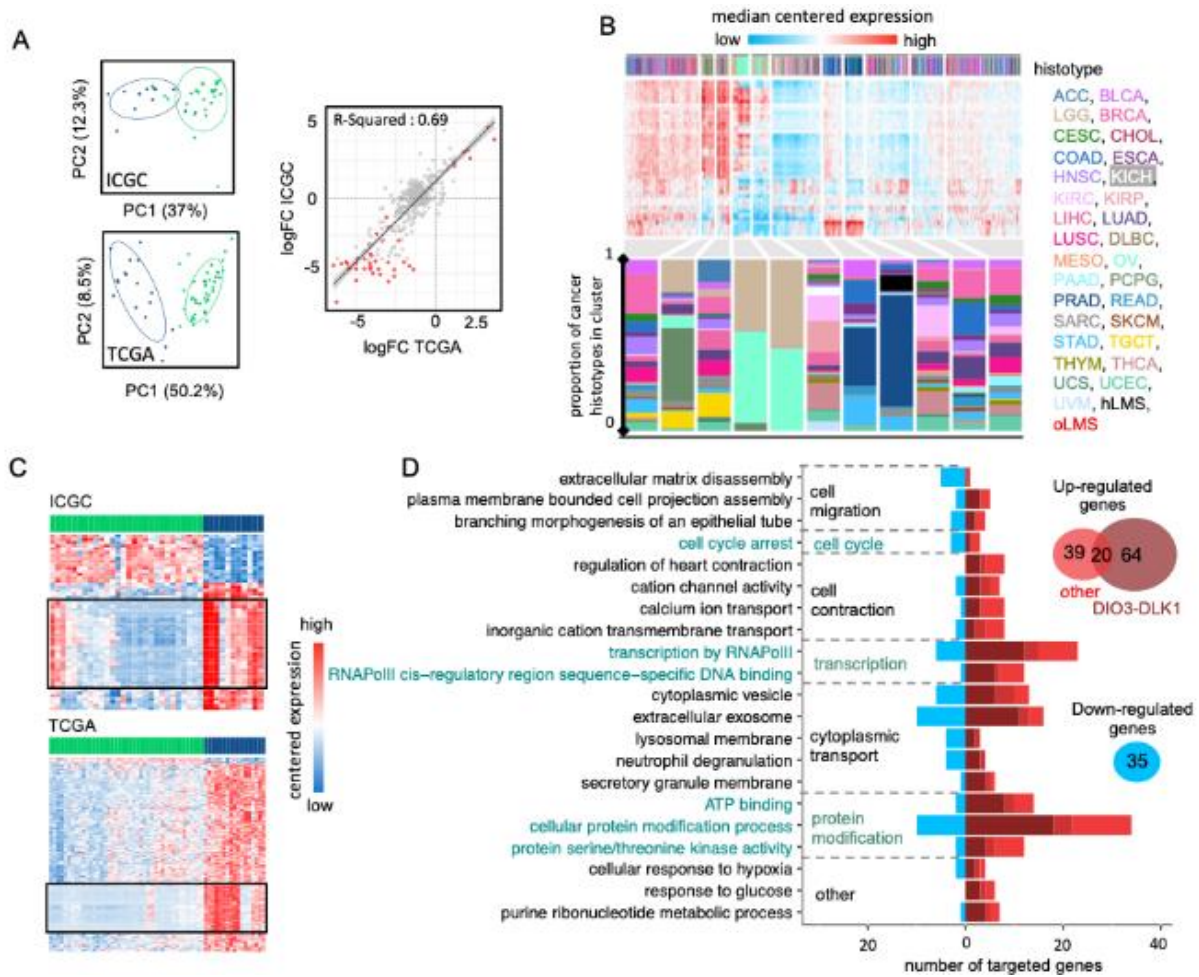


Figure 2: Analysis of miRNAs expression **A** Left panel: PCA obtained from expression of 484 mature miRNAs in 39 ICGC patients (top) and of 475 mature miRNAs in 60 TCGA patients (bottom). Colors correspond to hLMS (green, ICGC: 28 patients, TCGA: 43 patients) and oLMS (blue, ICGC: 11 patients, TCGA: 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).

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 (Supplementary Figure S3A) with very low expression in hLMS (Supplementary 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) (Supplementary 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 (Supplementary Table S2C). We annotated the 158 corresponding target genes (35 down- and 123 up-regulated in hLMS) with GSA_n (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 Supplementary 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* (23)(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

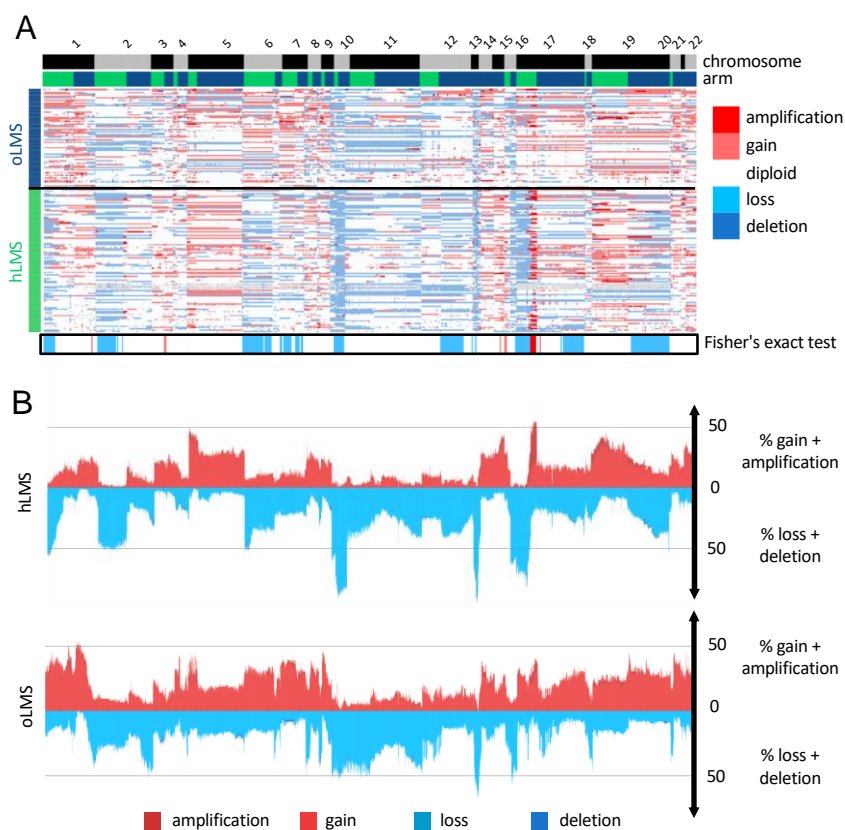


Figure 3: Copy number analysis of 7479 genes in LMS merged cohort (84 Affymetrix (CGH), 39 ICGC (WGS) and 62 TCGA (CGH) patients). **A** Heatmap showing copy number of genes (columns) in each patient (rows). Patients grouped according to LMS type (hLMS: green, oLMS: blue). Annotations above heatmap show chromosomes from 1 to 22 alternating grey and black and arms (p: green, q: blue). Annotation below shows significantly enriched events in hLMS. 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.

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 (Supplementary 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, Supplementary 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

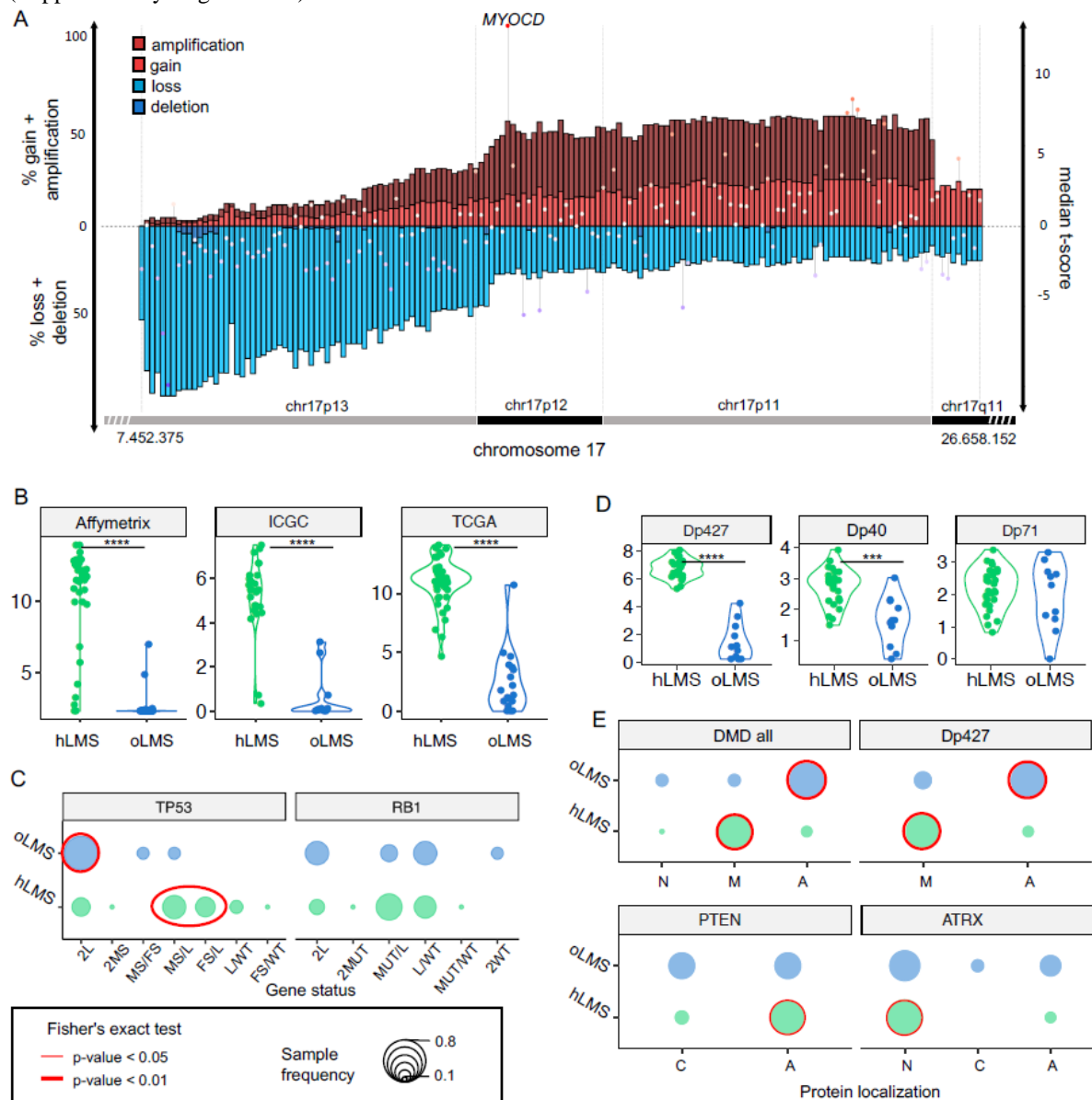


Figure 4: Zoom on genes of interest. **A.** Zoom on chr17p13-11/q11 genomic region (x-axis) penetrance profile containing *MYOCD*. Left y-axis indicates percentage of loss (light blue), deletion (dark blue), gain (light red) and amplification (dark red); right y-axis shows hLMS/oLMS median t-scores (from 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) 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 test p-value < 0.01 (bold) and < 0.05 (thin).

hLMS compared to oLMS ($P < 10^{-7}$ all cohorts considered, Figures 4A, 4B, Supplementary Table S3B). *MYOCD* expression was very high in 84% of hLMS (97/115, detailed per cohort in Supplementary 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 S4C, Supplementary Table S3B, Figure 4B).

Interestingly, the well-known tumor suppressors *RBI*, *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, Supplementary 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 ($P < 0.05$, Supplementary Figure S4D) than hLMS but that the mutational burden was similar between them ($P = 0.5$, Supplementary 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. (Supplementary 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, *RBI* in 97.4%, *PTEN* in 82%, *ATRX* in 28.2% and *DMD* in 25.6%. (Table 2, Supplementary Tables S4).

We found no significant difference in *RBI* 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 *RBI*, without reaching significance (Figure 4C, Supplementary Table S4A).

PTEN was almost exclusively altered by complete gene deletion, regardless of the LMS type (Table 2, Supplementary 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 (Darmusey *et al.*, 2020). Accordingly, *ATRX* nuclear localization was significantly enriched in hLMS (Supplementary 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, Supplementary 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 ($P = 0.63$) (Supplementary 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 *RBI*, 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*,

Group	Alterations	<i>TP53</i>	<i>RBI</i>	<i>PTEN</i>	<i>ATRX</i>	<i>DMD</i>
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)

Table 2: Summary of genetic alterations in 39 ICGC patients for *TP53*, *RBI*, *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 number are reported between brackets.

the main drivers of smooth muscle cell differentiation, given that PTEN also interacts with SRF (27). 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 (28), and CCG-100602, an inhibitor of the SRF/MYOCD interaction (29), were tested on 3 LMS (OC80: hLMS with a *MYOCD* amplification, OC88: oLMS and OC48: oLMS with a *MYOCD* gain) 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 5 B 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,

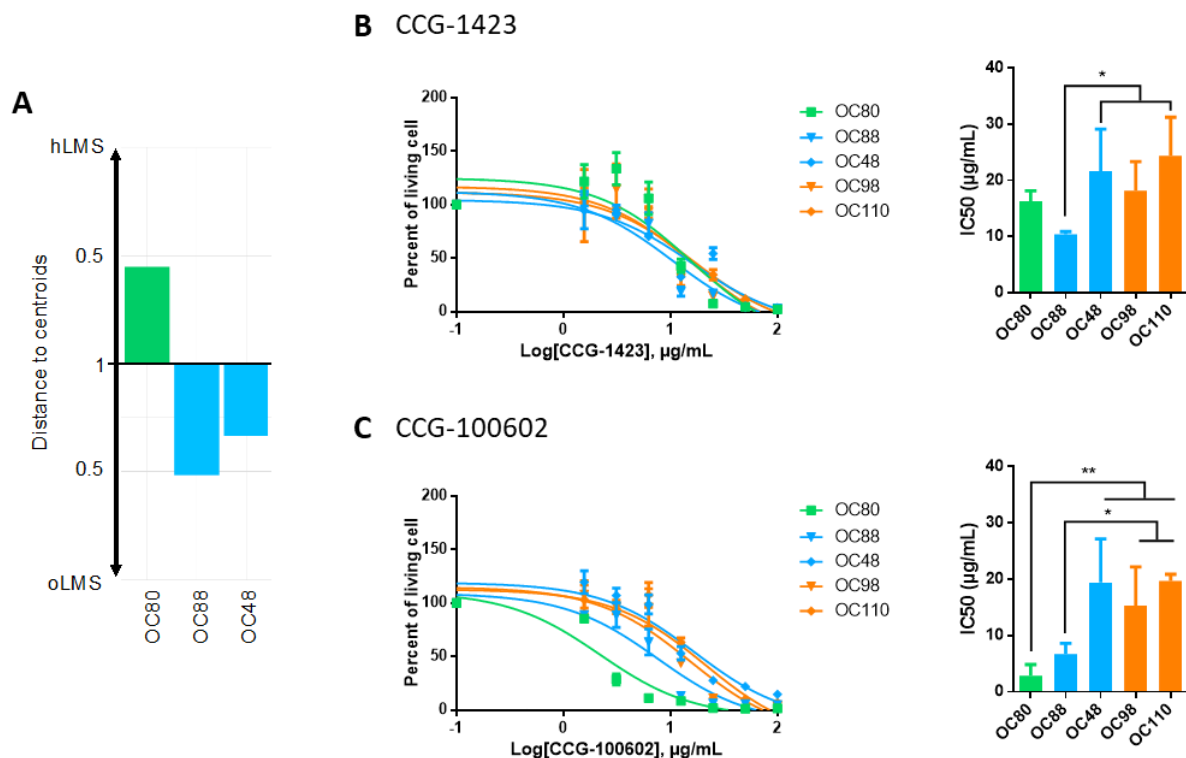


Figure 5: SRF/MYOCD inhibitor can specifically target hLMS. **A** Distance to centroids determining h/oLMS status on 1672 genes of 3 LMS cell lines. **B** Cytotoxicity curves of CCG-1423, inhibitor of SRF/MRTF axis, on 3 LMS and 2 UPS cell lines, using MTT assay after 72h of treatment at increasing concentrations (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-100602, 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**.

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 (30), 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 a ubiquitous transcription factor (TF), SRF (31), 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 (32,33). 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 (34), 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 (35), 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 (36). 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 (37). 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 (38). 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 (27). 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)(32), 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 (27), 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 underestimated.

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 (22,32) 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* (39) 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 (40), so it is probably involved in this dual phenotype. However, this cluster involves around 50 miRNAs with different functions in different cellular contexts (41), 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 (38,41) 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 (43). Subsequently, however, it was demonstrated to induce resistance (44). 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

Tumor samples and histological classification

Two hundred seventy-eight out of the 387 complex genetics sarcomas (5,45), the 60 GIST (46), and the 58 synovial sarcomas (47) are part of cohorts previously described (Supplementary Table S5A).

Fifty-nine leiomyosarcomas are part of the ICGC program (International Cancer Genome Consortium), for which constitutional DNA and tumor DNA/RNA were available.

All cases were systematically reviewed by expert pathologists from the French Sarcoma Group according to the World Health Organization recommendations (48).

LMS clinico-pathological data and patient information are summarized in Supplementary Tables S6.

Access to data

Genomic and expression arrays are available on Gene Expression Omnibus (GEO). Raw files will be available on sequence read archive. ICGC Whole-Genome sequencing and RNA sequencing data for the 67 LMS are available at <https://dcc.icgc.org/projects/LMS-FR>. ICGC miRNA-sequencing data are available on Gene Expression Omnibus (GEO).

Database accession number for data relative to samples previously published are presented in Supplementary Tables S5.

The code is available on Oceancode.

TCGA data

Clinical, gene expression and copy number data from 75 leiomyosarcomas were obtained from UCSC cancer genome browser (<https://genome-cancer.ucsc.edu/>) (49) and XenaBrowser (50). We used RNA-seq gene-level transcription estimates, as in log2(x+1) transformed RSEM normalized count from

SARC project (version 2015-02-24) and GTEx+TCGA combined data (version 2016-04-12), mature miRNA strand expression (version 2017-09-08) and GISTIC2 thresholded gene-level copy number variation (CNV, version 2017-09-08). For PANCAN miRNA-seq data, we used the batch corrected (version 2016-12-29).

Normalization of Affymetrix and Agilent micro-arrays and gene selection

We used 87 samples analyzed on Agilent and Affymetrix platforms (Supplementary Table S5A). 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 quantile normalization (preprocessCore R package (51)). We harmonized the expression between the platforms by gene expression median centering in each experiment and then adding the mean of the experiment medians. Details are presented 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 (52).

Patient classification and clinical associations

We computed centroids (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 (53). We selected the Gaussian mixture distribution estimation that best fitted the hLMS centroid distance distribution (maximization of Bayesian Information Criterion).

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

Survival analysis was performed using the survival R package (54) 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* (55).

Sample clustering and PCA analysis

In all cases concerning unsupervised clustering, samples were clustered using *PCA* and *HCPC*

functions from the FactorMiner R package (56) and were visualized using the *pheatmap* R package (57). PCA analysis on ICGC and TCGA miRNA transcriptome were computed using the *prcomp* R function and visualized using the *ggbiplot* R package (58). We used the R package *Rtsne* (59) to visualize GTEx data with parameters *dims=2* *perplexity=100* and *max_iter=1000*.

Differential expression analysis

mRNA and miRNA differential expression analyses were performed with classical methods as detailed in Supplementary Methods.

miRNA-mRNA interaction analysis

miRNA data acquisition and pre-treatments are detailed in Supplementary Methods. The *MiRComb* R package (60) was used to integrate miRNA, mRNA expression data and experimentally validated miRNA-mRNA interactions from *miRecords v4* (61) and *miRTarBase v7.0* (62). 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 ($\log_{2}FC$) > 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.

Functional enrichment and mapping

Modules of co-expressed genes were analyzed using the *enricher* function from the ClusterProfiler R package (63) with the Molecular Signatures Database version v6 (MSigDB)(64). 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)(64) 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/>) (65). 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 GSA webserver (<https://gsan.labri.fr/>) (19) with default parameters to exhaustively annotate genes with the most precise Gene Ontology term.

CNV analysis

Acquisition of copy number data for 84 Affymetrix cohort patients (Supplementary Table S5B) and 39

ICGC cohort patients is detailed in Supplementary methods.

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 (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 (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.

Mutation analysis

Somatic variants were detected from 59 whole genome-sequenced ICGC tumor/normal paired frozen samples (HiSeq2000 Technology, Illumina Inc., San Diego, CA, USA). Read cleaning, mapping (Human Genome version hg19) and detection of variants are detailed in Supplementary Methods.

We analyzed somatic mutations patterns using the MutationalPatterns R package (66). We first generated a 96 tri-nucleotide mutation count matrix per patient which we compared to the 30 COSMIC signatures v3.1 (67). 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).

Alterations verification strategy

For the ICGC cohort, *ATRX*, *TP53*, *RBI*, *PTEN* and *DMD* sequences for each case obtained by whole genome sequencing were entirely screened using the Integrative Genomics Viewer (IGV_2.6.3 (68)) 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.

Inhibitors

To assess inhibition of interaction between SRF and MRTF or MYOCD upon cell viability, we performed an MTT assay using either CCG-1423 (10010350) or CCG-100602 (10787) (Bertin Bioreagent, Montigny

le Bretonneux, France) on LMS and UPS cells.

Additional information about sample selection, DNA and RNA extraction, expression and genomic arrays, PCR and RT-PCR, sequencing, FISH, immunohistochemistry, immunofluorescence, cell culture, cytotoxicity assay and analyses can be found in the Supplementary material and methods section.

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.

Authors' Contributions:

Conception and design: E. Darbo, F. Chibon, G. Pérot, L. Darmusey

Development of methodology: E. Darbo, F. Chibon

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): E. Darbo, G. Pérot, L. Darmusey, S. Le Guellec, L. Leroy, L. Gaston, N. Desplat, N. Thébault, C. Merle

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): E. Darbo, F. Chibon, G. Pérot, L. Darmusey

Writing, review, and/or revision of the manuscript: E. Darbo, F. Chibon, G. Pérot, L. Darmusey

Administrative, technical or logistic support (i.e., reporting or organizing data, constructing databases): P. Rochaix, T. Valentin, G. Ferron, C. Chevreau, B. Bui, E. Stoeckle, D. Ranchere-Vince, P. Mééus, P. Terrier, S. Piperno-Neumann, F. Colin, G. De Pinieux, F. Duffaud, J.M. Coindre, J.Y. Blay

Study supervision: F. Chibon

Other (data, expertise, and guidance): The French Sarcoma Group and the International Cancer Genome Consortium

Financial support:

The Instituts Thematiques Multiorganismes (ITMO) Cancer and the Claudius Regaud Institut supported this work.

References

1. WHO Classification of Tumours Editorial Board. *Soft Tissue and Bone Tumours* [Internet]. 5th ed. 2020 [cited 2020 Sep 1]. Available from: <https://publications.iarc.fr/Book-And-Report-Series/Who-Classification-Of-Tumours/Soft-Tissue-And-Bone-Tumours-2020>
2. Judson I, Verweij J, Gelderblom H, Hartmann JT, Schöffski P, Blay J-Y, 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;15:415–23.
3. van der Graaf WT, Blay J-Y, Chawla SP, Kim D-W, Bui-Nguyen B, Casali PG, et al. Pazopanib for metastatic soft-tissue sarcoma (PALETTE): a randomised, double-blind, placebo-controlled phase 3 trial. *The Lancet*. 2012;379:1879–86.
4. Ben-Ami E, Barysaukas CM, Solomon S, Tahlil K, Malley R, Hohos M, 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;123:3285–90.
5. Chibon F, Lagarde P, Salas S, Pérot G, Brouste V, Tirode 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;16:781–7.
6. Perou CM, Sørli T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406:747–52.
7. Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med*. 2013;10:e1001453.
8. Nielsen TO, West RB, Linn SC, Alter O, Knowling MA, O'Connell JX, et al. Molecular characterisation of soft tissue tumours: a gene expression study. *The Lancet*. 2002;359:1301–7.
9. Baird K, Davis S, Antonescu CR, Harper UL, Walker RL, Chen Y, et al. Gene Expression Profiling of Human Sarcomas: Insights into Sarcoma Biology. *Cancer Res*. 2005;65:9226–35.
10. Perot G, Derre J, Coindre J-M, Tirode F, Lucchesi C, Mariani O, et al. Strong Smooth Muscle Differentiation Is Dependent on Myocardin Gene Amplification in Most Human Retroperitoneal Leiomyosarcomas. *Cancer Research*. 2009;69:2269–78.
11. Beck AH, Lee C-H, Witten DM, Gleason BC, Edris B, Espinosa I, et al. Discovery of molecular subtypes in leiomyosarcoma through integrative molecular profiling. *Oncogene*. 2010;29:845–54.
12. Guo X, Jo VY, Mills AM, Zhu SX, Lee C-H, Espinosa I, et al. Clinically Relevant Molecular Subtypes in Leiomyosarcoma. *Clinical Cancer Research*. 2015;21:3501–11.
13. Lee Y-F, Roe T, Mangham DC, Fisher C, Grimer RJ, Judson I. Gene expression profiling identifies distinct molecular subgroups of leiomyosarcoma with clinical relevance. *Br J Cancer*. 2016;115:1000–7.
14. Hemming ML, Fan C, Raut CP, Demetri GD, Armstrong SA, Sicinska E, 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, Pérot G, Chibon F, Bonnin S, Lagarde P, Terrier P, 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;223:64–71.
16. Italiano A, Lagarde P, Brulard C, Terrier P, Lae M, Marques B, et al. Genetic Profiling Identifies Two Classes of Soft-Tissue Leiomyosarcomas with Distinct Clinical Characteristics. *Clinical Cancer Research*. 2013;19:1190–6.
17. Watson S, Perrin V, Guillemot D, Reynaud S, Coindre J-M, Karanian M, et al. Transcriptomic definition of molecular subgroups of small round cell sarcomas: Molecular classification of sarcoma subtypes. *J Pathol*. 2018;245:29–40.
18. Amant F, Coosemans A, Debiec-Rychter M, Timmerman D, Vergote I. Clinical management of uterine sarcomas. *The Lancet Oncology*. 2009;10:1188–98.
19. Ayllon-Benitez A, Bourqui R, Thébault P, Mougín F. GSA: an alternative to enrichment analysis for annotating gene sets. *NAR Genomics and Bioinformatics*. 2020;2:lqaa017.
20. The Cancer Genome Atlas Research Network, Abeshouse A, Adebamowo C, Adebamowo SN, Akbani R, Akherdolu T, et al. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell*. 2017;171:950-965.e28.
21. Cordes KR, Sheehy NT, White MP, Berry EC, Morton SU, Muth AN, et al. miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. *Nature*. 2009;460:705–10.
22. Jiang Y, Yin H, Zheng X-L. MicroRNA-1 inhibits myocardin-induced contractility of human vascular smooth muscle cells. *J Cell Physiol*. 2010;225:506–11.
23. Almeida MI, Nicoloso MS, Zeng L, Ivan C, Spizzo R, Gafa R, et al. Strand-Specific miR-28-5p and miR-28-3p Have Distinct Effects in Colorectal Cancer Cells. *Gastroenterology*. 2012;142:886-896.e9.
24. Zhou C, Chen Y, Kang W, Lv H, Fang Z, Yan F, 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, Burkhard FC, Rehrauer H, Aquino Fournier C, Monastyrskaya K. MicroRNA MiR-199a-5p Regulates Smooth Muscle Cell Proliferation and Morphology by Targeting WNT2 Signaling Pathway. *J Biol Chem*. 2015;290:7067–86.
26. Kim J, Kang Y, Kojima Y, Lighthouse JK, Hu X, Aldred MA, et al. An endothelial apelin-FGF link mediated by miR-424 and miR-503 is disrupted in pulmonary arterial hypertension. *Nat Med*. 2013;19:74–82.

27. Horita H, Wysoczynski CL, Walker LA, Moulton KS, Li M, Ostriker A, et al. Nuclear PTEN functions as an essential regulator of SRF-dependent transcription to control smooth muscle differentiation. *Nat Commun.* 2016;7:10830.
28. Evelyn CR, Wade SM, Wang Q, Wu M, Iñiguez-Lluhí JA, Merajver SD, et al. CCG-1423: a small-molecule inhibitor of RhoA transcriptional signaling. *Mol Cancer Ther.* 2007;6:2249–60.
29. Zhou N, Lee J-J, Stoll S, Ma B, Costa KD, Qiu H. 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;44:701.
30. Miettinen M. Smooth muscle tumors of soft tissue and non-uterine viscera: biology and prognosis. *Modern Pathology.* 2014;27:S17–29.
31. Miano JM. Myocardin in biology and disease. *J Biomed Res.* 2015;29:3–19.
32. Miano J. Serum response factor: toggling between disparate programs of gene expression. *Journal of Molecular and Cellular Cardiology.* 2003;35:577–93.
33. Treisman R. Identification of a protein-binding site that mediates transcriptional response of the c-fos gene to serum factors. *Cell.* 1986;46:567–74.
34. Turczyńska KM, Swärd K, Hien TT, Wohlfahrt J, Mattisson IY, Ekman M, et al. Regulation of Smooth Muscle Dystrophin and Synaptopodin 2 Expression by Actin Polymerization and Vascular Injury. *Arterioscler Thromb Vasc Biol.* 2015;35:1489–97.
35. Wang Y, Marino-Enriquez A, Bennett RR, Zhu M, Shen Y, Eilers G, et al. Dystrophin is a tumor suppressor in human cancers with myogenic programs. *Nat Genet.* 2014;46:601–6.
36. Lartigue L, Merle C, Lagarde P, Delespaul L, Lesluyes T, Le Guellec S, et al. Genome remodeling upon mesenchymal tumor cell fusion contributes to tumor progression and metastatic spread. *Oncogene.* 2020;39:4198–211.
37. Yadav RK, Chae S-W, Kim H-R, Chae HJ. Endoplasmic Reticulum Stress and Cancer. *J Cancer Prev.* 2014;19:75–88.
38. Owens GK, Kumar MS, Wamhoff BR. Molecular regulation of vascular smooth muscle cell differentiation in development and disease. *Physiol Rev.* 2004;84:767–801.
39. Vacante F, Denby L, Sluimer JC, Baker AH. The function of miR-143, miR-145 and the MiR-143 host gene in cardiovascular development and disease. *Vascular Pharmacology.* 2019;112:24–30.
40. 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;8:9597–607.
41. Benetatos L, Hatzimichael E, Londin E, Vartholomatos G, Loher P, Rigoutsos I, et al. The microRNAs within the DLK1-DIO3 genomic region: involvement in disease pathogenesis. *Cell Mol Life Sci.* 2013;70:795–814.
42. Alexander MR, Owens GK. Epigenetic Control of Smooth Muscle Cell Differentiation and Phenotypic Switching in Vascular Development and Disease. *Annu Rev Physiol.* 2012;74:13–40.
43. Cuppens T, Annibaldi D, Coosemans A, Trovik J, Haar N ter, Colas E, 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. American Association for Cancer Research;* 2017;23:1274–85.
44. Fourneau B, Chaire V, Lucchesi C, Karanian M, Pineau R, Laroche-Clary A, 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. Impact Journals;* 2016;8:7878–90.
45. Lesluyes T, Blay J-Y, Schoffski P, Italiano A, Le Cesne A, Debiec-Rychter M, et al. Expression and prognostic significance of PDGF ligands (A, B, C, and D) and PDGFR (A, B, and L) in soft-tissue sarcomas and GIST. *JCO.* 2017;35:11067–11067.
46. Lagarde P, Perot G, Kauffmann A, Brulard C, Dapremont V, Hostein I, et al. Mitotic Checkpoints and Chromosome Instability Are Strong Predictors of Clinical Outcome in Gastrointestinal Stromal Tumors. *Clinical Cancer Research.* 2012;18:826–38.
47. Lagarde P, Przybyl J, Brulard C, Pérot G, Pierron G, Delattre O, et al. Chromosome Instability Accounts for Reverse Metastatic Outcomes of Pediatric and Adult Synovial Sarcomas. *JCO.* 2013;31:608–15.
48. Fletcher C, Bridge JA, Hogendoorn P, Mertens F. WHO Classification of Tumours of Soft Tissue and Bone. 4th ed. Lyon: IARC Press; 2013.
49. Goldman M, Craft B, Swatloski T, Cline M, Morozova O, Diekhans M, et al. The UCSC Cancer Genomics Browser: update 2015. *Nucleic Acids Research.* 2015;43:D812–7.
50. Goldman MJ, Craft B, Hastie M, Repčeka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol.* 2020;38:675–8.
51. Bolstad B. preprocessCore: A collection of pre-processing functions [Internet]. 2019. Available from: <https://github.com/bmbolstad/preprocessCore>
52. Csárdi G, Nepusz T. The igraph software package for complex network research. *InterJournal.* 2006;Complex Systems:1695.
53. Scrucca L, Fop M, Murphy T Brendan, Raftery A E. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R Journal.* 2016;8:289.
54. Therneau TM. A Package for Survival Analysis in R [Internet]. 2020. Available from: <https://CRAN.R-project.org/package=survival>
55. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using “ggplot2” [Internet]. 2019. Available from: <https://CRAN.R-project.org/package=survminer>
56. Lê S, Josse J, Husson F. **FactoMineR**: An R Package for Multivariate Analysis. *J Stat Soft [Internet].* 2008 [cited 2020 Jun 29];25. Available from: <http://www.jstatsoft.org/v25/i01/>
57. Kolde R. Pheatmap: pretty heatmaps [Internet].

2019. Available from: <https://CRAN.R-project.org/package=pheatmap>
58. Vu VQ. ggbiplot: A ggplot2 based biplot [Internet]. 2011. Available from: <http://github.com/vqv/ggbiplot>
59. van der Maaten L, Hinton G. Visualizing non-metric similarities in multiple maps. *Mach Learn*. 2012;87:33–55.
60. Vila-Casadesús M, Gironella M, Lozano JJ. MiRComb: An R Package to Analyse miRNA-mRNA Interactions. Examples across Five Digestive Cancers. Campbell M, editor. *PLoS ONE*. 2016;11:e0151127.
61. Xiao F, Zuo Z, Cai G, Kang S, Gao X, Li T. miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Research*. 2009;37:D105–10.
62. Chou C-H, Shrestha S, Yang C-D, Chang N-W, Lin Y-L, Liao K-W, et al. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Research*. 2018;46:D296–302.
63. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology*. 2012;16:284–7.
64. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003;34:267–73.
65. Imrichová H, Hulselmans G, Kalender Atak Z, Potier D, Aerts S. i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic Acids Res*. 2015;43:W57–64.
66. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med*. 2018;10:33.
67. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*. 2019;47:D941–7.
68. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24–6.

Supplementary Material and Methods

Samples collection

Samples used in ICGC cohort were collected prospectively as part of the ICGC program (International Cancer Genome Consortium) with patient consent or came from care, were re-qualified for research and were part of the Biological Resources Center at Bergonie Cancer Institute (CRB-IB).

Other samples used in this study were also 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. These samples were obtained from care and were requalified for research. Patients' written informed consent was approved by the Committee of Protection of Individuals.

DNA extraction

Genomic DNA from frozen samples was isolated using standard phenol-chloroform extraction protocol. DNA was quantified using a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Genomic DNA from blood samples was extracted using customized automated purification of DNA from compromised blood samples on the Autopure LS protocol according to the manufacturer's recommendations (9001340, Qiagen, Hilden, Germany) with increased centrifugation of 10 min for DNA precipitation and DNA wash.

RNA extraction

RNA extraction and quality assessment were performed as previously described (1).

Expression Microarray Data

Microarray data were obtained on two different platforms. 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.

Normalization of Affymetrix and Agilent micro-arrays and gene selection

Selecting consistent genes between Affymetrix and Agilent:

$$cor_{max,i} = \max(cor(g_{Affy,i}, g_{Agi,vj=i}), cor(g_{Affy,i}, g_{Affy,vj=i}), cor(g_{Agi,i}, g_{Agi,vj=i}))$$

$$g_i = \begin{cases} 1 & cor(g_i, g_{j=i}) > 0.8 \vee cor(g_i, g_{j=i}) > cor_{max,i} \\ 0 & otherwise \end{cases}$$

Where $i, j \in \mathcal{A} \{1..n\}$ are the indexes of pairs of

genes to be compared. g_{Affy} and g_{Agi} are genes analyzed on Affymetrix and Agilent micro-arrays and \mathcal{A} the ensemble of all possible indexes between 1 and n the number of genes to be tested. g_i represents the status of gene i that will be kept if 1 or discarded otherwise.

Median harmonization:

$$median_{i,Affy} = median(g_{i,\forall k \in \{Affy\}})$$

$$median_{i,Agi} = median(g_{i,\forall k \in \{Agi\}})$$

$$median_{i,All} = mean(median_{i,Affy}, median_{i,Agi})$$

$$g_{i,AffyHarm} = g_{i,Affy} - median_{i,Affy} + median_{i,All}$$

$$g_{i,AgiHarm} = g_{i,Agi} - median_{i,Agi} + median_{i,All}$$

Where $i \in \{1..n\}$ is the gene index and n the

number of genes to be analyzed, g represents the gene expression either in Affymetrix samples (*Affy*) or Agilent samples (*Agi*).

RNA sequencing

To control the sequencing quality, ERCC RNA Spike-In Mix (4456740, Rockville, MD, USA) were added to each RNA sample as recommended by the manufacturer. Analysis of this quality control was performed as previously described (1).

Libraries from total RNA were prepared using the TruSeq®Stranded Total RNA Sample Preparation kit (RS-122-2301, Illumina Inc., San Diego, CA, USA) according to the manufacturer's protocol. Briefly, 0.5 µg of total RNA was ribo-depleted using the Ribo-Zero Gold Kit. RNA fragmentation resulted in fragments of 80 – 450nt, with a major peak at 160nt. First-strand cDNA synthesis by random hexamers and reverse transcriptase was followed by second-strand cDNA synthesis, performed in presence of dUTP instead of dTTP. Blunt-ended double-stranded cDNA was 3'adenylated and Illumina-indexed adapters were ligated. Resulting libraries were enriched with 15 PCR cycles.

Libraries were sequenced on HiSeq2000 (Illumina Inc., San Diego, CA, USA) in paired-end mode with a read length of 2x75 bp using TruSeq SBS Kit v3-HS (FC-401-3001, Illumina Inc., San Diego, CA, USA). Image analysis, base calling and base quality scoring of the run were processed by integrated primary analysis software – Real Time Analysis (RTA 1.13.48) and followed by generation of FASTQ sequence files with CASAVA (v1.8, Illumina Inc., San Diego, CA, USA). Library construction and RNA sequencing were performed at the Centro Nacional de Análisis Genómico (CNAG, Barcelona, Spain).

Detection of fusion transcripts

RNA bioinformatic analysis (alignment and expression quantification) was performed as previously described (1). Fusion transcripts were detected with Defuse v0.6.1 (2) as previously described (3).

Whole genome sequencing

To construct short-insert paired-end libraries, a no-PCR protocol was used with the TruSeq™DNA Sample Preparation Kit v2 (FC-121-2001/FC-121-2002, Illumina Inc., San Diego, CA, USA) and the KAPA Library Preparation kit (Kapa Biosystems, Basel, Switzerland). Briefly, 2 µg of genomic DNA were sheared on a Covaris™ E220, size-selected and concentrated using AMPure XP beads (A63880, Agencourt, Beckman Coulter, Brea, CA, USA) in order to obtain a fragment size of 220 – 480bp. Fragmented DNA was end-repaired, adenylated and ligated to Illumina-specific indexed paired-end adapters.

DNA sequencing was performed in paired-end mode, 2x100bp or 2x125 bp according to Flowcell version in five or in three sequencing lanes of HiSeq2000 Flowcell v3 or v4 (Illumina Inc., San Diego, CA, USA) to analyze tumor or normal/constitutive samples and to obtain a minimal yield of 145 Gb or 85 Gb, respectively. Two tumor samples were sequenced in 20 lanes to obtain a minimal yield of 560 Gb (LMS2 and LMS5). Library construction and DNA sequencing were performed at the Centro Nacional de Análisis Genómico (CNAG, Barcelona, Spain). Image analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis (RTA 1.13.48) and followed by generation of FASTQ sequence files with CASAVA (Illumina Inc., San Diego, CA, USA).

DNA reads were trimmed of the 5' and 3' low quality bases (phred cut off 20, max trim size 30nt) and sequencing adapters were removed with Sickle2 (<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 (4), 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 (5) and PicardTools v1.118 (<http://broadinstitute.github.io/picard/index.html>), respectively.

Break Point Detection

Structural variants (SV) were detected from paired tumor/normal whole genome high-quality sequencing data. Paired-end reads were aligned using Bowtie v2.2.1.0 (4), 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.

i) Identification: at this step, reads with at least one soft-clipped end were analyzed as singletons. A position was considered as a potential breakpoint if it was covered by at least 4 soft-clipped reads, 5 soft-clipped bases (with at least two occurrences of two different bases), and if they represented more than 5% of the total amount of reads at this position in the tumor sample. We selected potential somatic events by discarding positions covered by at least 1 read and 1 base in a surrounding 5-nucleotide window in the normal sample. We refer to them as the “first side” of the breakpoint.

ii) Characterization: to determine the genomic positions of the soft-clipped sequence from selected reads, we used the UCSC blat server (6). If no match was returned, the reverse complement sequence was pulled to test. If there was still no match, the BAM file was investigated for some soft-clip somatic position around the discordant or oversized-insert read mate (hereafter named abnormal) location from the first side of the breakpoint. Because of the small size of the soft-clipped sequence, multiple matches can be found. We used soft-clipped abnormal read mates to select matches with the most coherent chromosomal locations. We refer to them as the “second side” of the breakpoint.

iii) Selection: Positions detected from both the first and second sides (in a 5-nucleotide window) were defined as the common pool. We considered as artifacts (due to repeat regions for instance) couples of positions covered with reads and associated soft-clipped sequences separated by fewer than 15 nucleotides and discarded them. We classified the breakpoints in three groups: high-confidence breakpoints, breakpoints needing investigation, and unique position breakpoints. If a breakpoint was

covered by reads and associated soft-clipped sequences having both positions belonging to the common pool, it was classified in the first group. If a breakpoint was covered by reads and associated soft-clipped sequences having only one of the positions belonging to the common pool, it was classified in the second group. Then the missing position was searched among the filtered positions. If it was present in the normal sample, the position was discarded and the breakpoint was completed otherwise. Finally, the third group corresponds to breakpoints with both sides outside the common pool and considered as unique: these were discarded. The sides of breakpoints were sorted according to their chromosomal positions to avoid duplicates.

Detection of variants

Variants were detected using *mpileup* from SAMtools v0.1.19 (5), with a minimum of 20 as phred quality score (-Q 20), and *bcftools view -cvg*, SAMtools v0.1.19 (5). Detected variants with fewer than 5 coverage reads were filtered out. Variants detected in tumor RNA, constitutional and tumor DNA 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 with the Annovar v20160314 tool (7). Variants whose alternative allele frequency (AF) in the Caucasian population (CEU) was lower than 0.1% as reported in 1000Genome database (8) were selected.

PCR on genomic DNA

To validate mutations and breakpoints on genomic DNA, PCR primers were designed using the Primer 3 program (<http://frodo.wi.mit.edu/primer3/>) and are presented in Supplementary Table S4D. All PCR were performed as previously described for RT-PCR on 50ng of gDNA using AmpliTaqGold® DNA polymerase (4311820, Applied Biosystems, Foster City, CA, USA) and PCR programs used are indicated in Supplementary Table S4D.

RT-PCR

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 (<http://frodo.wi.mit.edu/primer3/>) (Supplementary Table S4E). All PCR were performed on 50ng of cDNA using AmpliTaqGold® DNA polymerase ((4311820, Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions with the PCR program described in Supplementary

Table S4D.

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).

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 Panoramic 250 Flash II Digital Slide Scanner and analyzed with the Panoramic 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.

Copy number data acquisition and pre-treatment

CGH from Affymetrix cohort (Supplementary Table S5B): Array-CGH for 53 cases was performed using BAC-array as described in Chibon et al., 2010 (9), 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. Genes absent in more than a third of the patients were discarded.

WGS from ICGC cohort: paired tumor/normal whole genome sequencing data were processed using the cn.MOPS R package (10) with default parameters and a 500-nucleotide window. Regions were intersected with TxDb.Hsapiens.UCSC.hg19.knownGene (11) 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.

miRNA sequencing

Libraries were prepared by the Genomics Unit at the Centre de Regulació Genòmica (Barcelona, Spain) using the NEBNext Small RNA Library Prep Set for Illumina (E7330L, New England Biolabs Inc., Ipswich, MA, USA) according to the manufacturer's recommendations. Libraries were sequenced on HiSeq2500 (Illumina Inc., San Diego, CA, USA) in single-read mode with a read length of 50bp (V4 chemistry) to reach a minimal yield of 10M reads.

Reads were then trimmed for adaptors using cutadapt version 1.10 (12) 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 (13) according to the recommendations in (14). We first used the BWA-aln algorithm version 0.7.17 (15) with -n 1 -o 0 -e 0 -k 1 -t 8 parameters and evaluated mapping quality using Qualimap version 2.2.2b. We then used Samtools version 1.9 (5) to discard the reads with a mapping quality under 30 and to count mapped reads with the reference sequences (*samtools idxstats*).

Differential expression analysis

mRNA: Differential expression (DE) analysis was performed using 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 then applied the EdgeR R package (16) 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 (17). 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 R² value between ICGC and TCGA logFC.

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 was diluted in Prep kit 26 (783-2876, Roche, Basel, Switzerland) and incubated for 1h (1:200, Clone 138G6, 9559, Cell Signaling Technology, Danvers, MA, USA). 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 Panoramic 250 Flash II Digital Slide Scanner and analyzed with the Panoramic 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 were NCL-DYS1 (1/25, DYS1CE,

Novacastra, Leica, Wetzlar, Germany) and ab15277 for all the dystrophin isoforms (1/100, Abcam, Cambridge, UK). Secondary antibodies used were Alexa Fluor 488 anti-mouse (1/400, A-11001, Thermo Fisher Scientific, Waltham, MA, USA) and Alexa Fluor 594 anti-rabbit (1/400, A-11072, Invitrogen, Carlsbad, CA, USA).

Cell culture

Human LMS cell lines OC80, OC109, OC88 and OC48 and human UPS cell lines OC98 and OC110 were cultured in RPMI-1640 (524000-025, Life Technologies, Carlsbad, CA, USA) supplied with 10% fetal bovine serum (S1810-500, Dutscher, Brumath, France) at 37°C in a humidified chamber containing 5% CO₂.

Cytotoxicity analysis

LMS and UPS cells were seeded in 96-well microplates at a density of 2.10³ 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, Bertin Bioreagent, Montigny le Bretonneux, France) or CCG-100602 (10787, 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: $\frac{(A_{\text{sample}(570\text{nm})} - A_{\text{sample}(650\text{nm})})}{(A_{\text{control}(570\text{nm})} - A_{\text{control}(650\text{nm})})} * 100$. IC₅₀ was calculated with GraphPad Prism (GraphPad Software, San Diego, CA, USA) using non-linear regression.

Visualization

All generated plots were produced using ggplot2 R package (18) and extension ggpubr (19).

References supplementary:

1. Lesluyes T, Pérot G, Largeau MR, Brulard C, Lagarde P, Dapremont V, et al. RNA sequencing validation of the Complexity INDEX in SARComas prognostic signature. *Eur J Cancer*. 2016;57:104–11.
2. McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MGF, et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput*

Biol. 2011;7:e1001138.

3. Delepaul L, Lesluyes T, Pérot G, Brulard C, Lartigue L, Baud J, et al. Recurrent TRIO Fusion in Nontranslocation-Related Sarcomas. *Clin Cancer Res*. 2017;23:857–67.
4. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
5. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
6. Kent WJ. BLAT---The BLAST-Like Alignment Tool. *Genome Research*. 2002;12:656–64.
7. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.
8. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
9. Chibon F, Lagarde P, Salas S, Pérot G, Brouste V, Tirode 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. *Nature Medicine*. 2010;16:781.
10. Klambauer G, Schwarzbauer K, Mayr A, Clevert D-A, Mitterecker A, Bodenhofer U, 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:e69–e69.
11. Carlson M. TxDb.Hsapiens.UCSC.hg19.knownGene: Annotation package for TxDb object(s). 2015.
12. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j*. 2011;17:10.
13. Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research*. 2011;39:D152–7.
14. Tam S, Tsao M-S, McPherson JD. Optimization of miRNA-seq data preprocessing. *Briefings in Bioinformatics*. 2015;16:950–63.
15. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
16. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–40.
17. Vila-Casadesús M, Gironella M, Lozano JJ. MiRComb: An R Package to Analyse miRNA-mRNA Interactions. Examples across Five Digestive Cancers. Campbell M, editor. *PLoS ONE*. 2016;11:e0151127.
18. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016. Available from: <https://ggplot2.tidyverse.org>
19. Kassambara A. ggpubr: “ggplot2” Based Publication Ready Plots [Internet]. 2020. Available from: <https://CRAN.R-project.org/package=ggpubr>

Supplementary tables are available upon request

Supplementary Tables S1: 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.

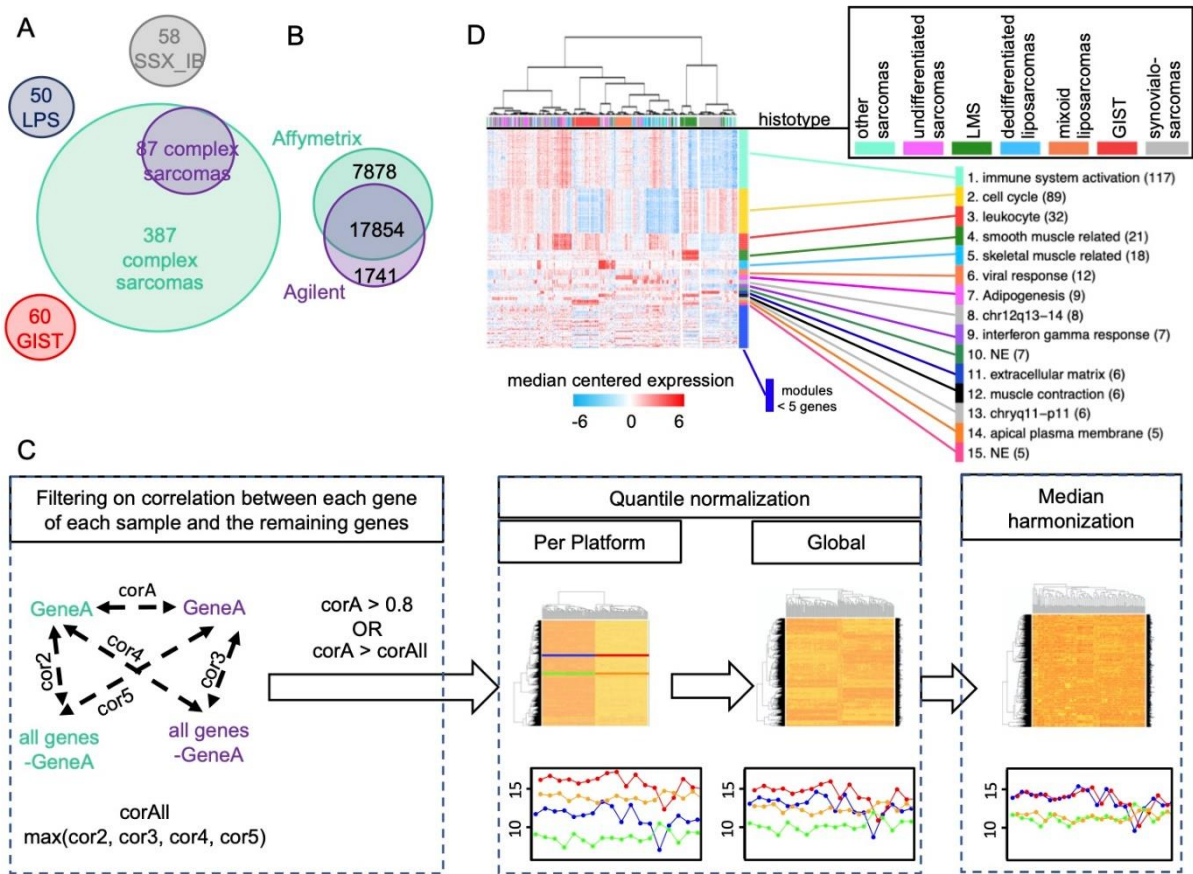
Supplementary Tables S2: Table summarizing differential expression analysis for mRNA in Affymetrix, ICGC and TCGA, for miRNA in ICGC and TCGA and miRComb miRNA-mRNA interaction analysis.

Supplementary Tables S3: Table summarizing copy number alteration analysis by cytogenic band and by genes in merged cohort.

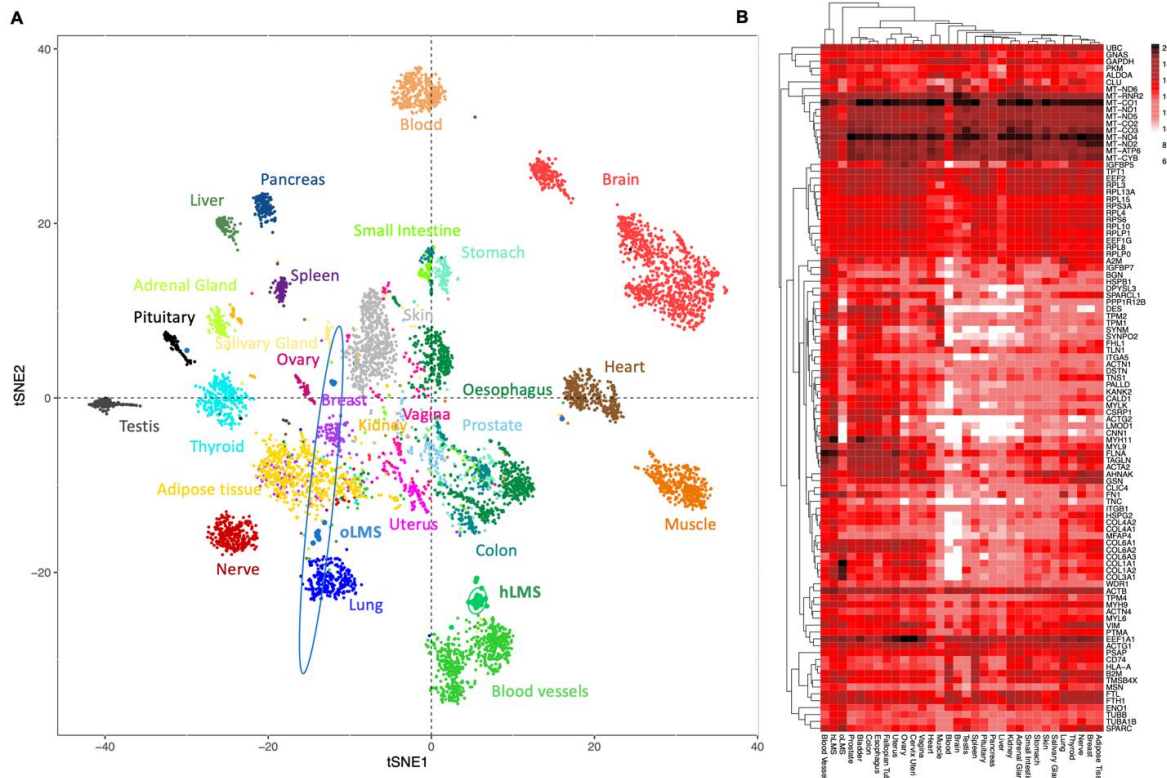
Supplementary Tables S4: A. Table summarizing genetic alterations occurring for TP53, RB1, PTEN, DMD, ATRX and MYOCD in each sample of ICGC cohort as well as expression data of the six genes. **B.** Table presenting mutations detected in TP53, RB1, DMD and ATRX. **C.** Table presenting structural variants 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.

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

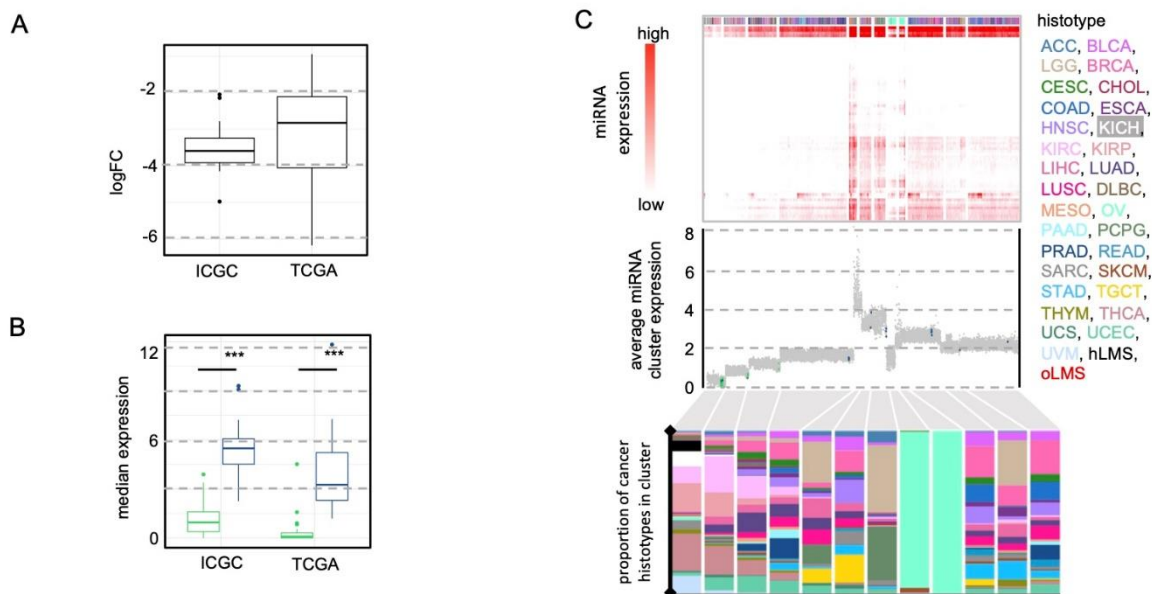
Supplementary Table S6: Clinical annotation summary for 98 Affymetrix, 59 ICGC and 75 TCGA patients.



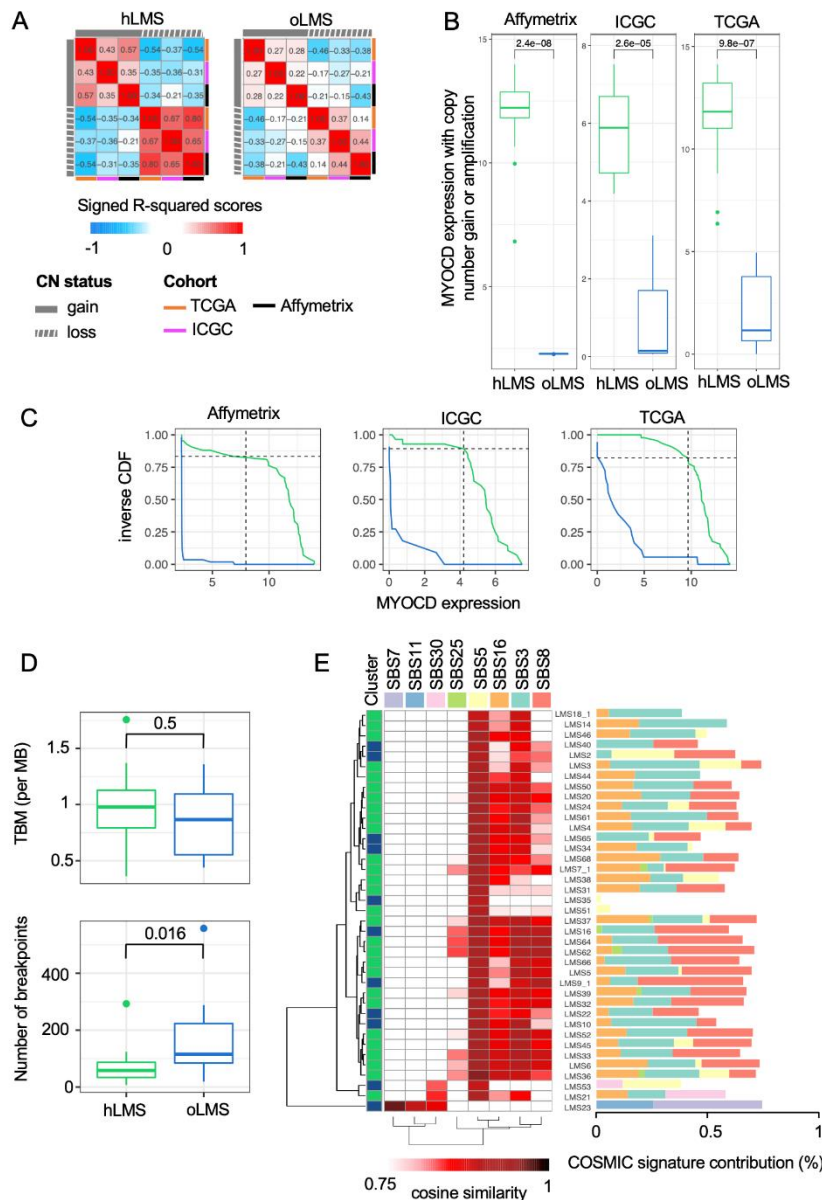
Supplementary Figure S1: Normalization of Affymetrix and Agilent micro-arrays. **A.** Number of patients per cohort. Overlapping circles represent common samples between experiments. **B.** Venn diagram showing intersection of genes between Affymetrix and Agilent micro-arrays. **C.** Normalization pipeline. The 87 samples analyzed on both Agilent and Affymetrix technologies were used to select the 9066 highly correlating genes between the two platforms (Pearson's correlation > 0.8) or correlating better than with any other genes within and between platforms. The micro-arrays were first quantile-normalized per platform and merged to undergo a second round of quantile normalization. Finally, the median across samples for each gene was computed separately per platform and then harmonization was obtained by subtracting the mean of these medians from the gene expression values. Heatmaps (rows: genes, columns: samples) represent unsupervised clustering of samples at each step of normalization process. The two rounds of quantile normalization cluster the samples per platform, while the harmonized data cluster patients with themselves. Plots under heatmaps show expression profiles (x-axis: expression level) of two randomly chosen genes (gene 1: blue in Affymetrix, red in Agilent, gene 2: green in Affymetrix, orange in Agilent) in the 87 patients. **D.** Summary of functional enrichment analysis of co-expressed gene modules. Heatmap showing clustering of 555 sarcoma patients and 467 genes grouped in modules of co-expressed genes for which a representative functional term is associated. Number before term refers to module number in detailed results (Table S1A). Number of genes present in modules is shown in brackets.



Supplementary Figure S2: t-SNE clustering from most expressed hLMS 100 genes in 7414 normal and LMS samples from GTEX dataset. **A.** 2-dimensional t-SNE representation of samples colored by tissue or cancer type. **B.** Heatmap shows median expression level per tissue type of most expressed hLMS 100 genes.



Supplementary Figure S3: Evaluation DIO3-DLK1 miRNA cluster **A.** Boxplot representing distribution of LogFC (y-axis). **B.** Boxplot representing distribution of median expression with comparison between hLMS (green) and oLMS (blue). Stars indicate significant difference between LMS groups (Wilcoxon's p-value < 0.0001). **C.** HCPC clustering on 72 mature miRNAs from DLK1-DIO3 miRNA cluster in the 9564 PANCAN samples. Top: Heatmap showing pan-cancer analysis of 83 mature miRNAs from the DIO3-DLK1 cluster in 9564 HCPC-clustered patients. Column annotation indicates histotypes of samples. Middle: Each dot represents a patient ordered on x-axis as in heatmap above. Values on y-axis represent average expression value of the 83 miRNAs in each patient. Green and blue dots indicate hLMS and oLMS respectively. Bottom: Barplot showing proportion of each histotype in clusters obtained during HCPC analysis.



Supplementary Figure S4: Copy number alterations and mutational burden. **A.** Heatmap showing linear regression r-squared scores representing correlation between gain and loss penetrance profiles of Affymetrix, ICGC and TCGA cohorts. Scores signed according to type of correlation (anti-correlation: negative, positive otherwise). **B.** Boxplot representing *MYOCD* gene expression in patients with *MYOCD* gain or amplification. Numbers represent Wilcoxon's test p-value. **C.** Inverse cumulative distribution function (CDF) of hLMS (green) and oLMS (blue) ranked according to *MYOCD* expression in Affymetrix, ICGC and TCGA cohorts. Vertical dashed lines represent 3rd quartile of all gene expressions present in experiments: Affymetrix: 7.95, ICGC: 4.2, TCGA: 9.66. Horizontal dashed lines indicate proportion of hLMS with *MYOCD* expression above the corresponding threshold and which were considered as high expression: Affymetrix: 83%, ICGC: 89%, TCGA: 82%. **D.** Boxplot distribution of TMB (top: number of somatic mutations per Mb, y-axis) and number of breakpoints (bottom) in hLMS and oLMS. LMS23 was discarded from analysis, TMB=120.35, breakpoints=74. Numbers indicate Wilcoxon's test p-values. **E.** Heatmap showing cosine similarity between LMS samples (rows, hierarchical clustering: Euclidean distance, average agglomeration) and COSMIC Single Base Substitution (SBS) signatures (column, hierarchical clustering: cosine similarity, complete agglomeration). Cosine similarity is set to white if under 0.75, and light to dark red (1) otherwise. Rows in bar chart (right panel) correspond to same sample as in heatmap. Optimal linear contribution (x-axis) of each SBS signature to reconstruct sample mutational profiles. SBS3: defective homologous recombination DNA damage repair, SBS5: clock-like, SBS7: ultraviolet light exposure, SBS8: unknown, SBS11: temozolomide treatment, SBS16: unknown, SBS25: chemotherapy treatment, SBS30: Defective DNA base excision repair due to *NTHL1* mutation.

III. Conclusions

Notre étude diffère de celles jusqu'alors réalisées sur les LMS sur plusieurs aspects : exclusion des LMS utérins, uniquement sur les tumeurs primaires et intégration de données venant de multiples OMICS. La méthode développée dans le papier sépare les LMS en deux groupes, hLMS et oLMS, ayant des origines cellulaires différentes, avec des caractéristiques transcriptomiques et cliniques spécifiques. Au niveau génomique nous montrons quatre altérations récurrentes et communes aux deux groupes : *TP53*, *RB1*, *PTEN* et *ATRX*. Parmi ces quatre gènes, *TP53* (Pérot *et al.*, 2010; Rubio *et al.*, 2010), *RB1* (Calo *et al.*, 2010; Dick and Rubin, 2013) et *PTEN* (Hernando *et al.*, 2007; Gibault *et al.*, 2011; Stefano and Giovanni, 2019) ont des rôles connus dans l'oncogenèse en général et plus spécifiquement dans les LMS. Les altérations de ces trois gènes présentent des spécificités selon le groupe de LMS. *TP53* et *RB1* ont leurs deux allèles altérés mais les mécanismes d'altération diffèrent : dans les oLMS les deux allèles sont perdus et dans les hLMS un allèle est perdu et le second est muté. La protéine codée par *PTEN* est plus souvent totalement absente dans les hLMS que dans les oLMS. Le quatrième gène récurrent est *ATRX* et, contrairement aux trois autres, il présente plus de mutations que d'altérations du nombre de copies ou de structure. Concernant l'effet de ces altérations dans l'oncogenèse des LMS, toutes les études jusqu'alors publiées associent les altérations d'*ATRX* au phénotype ALT (Liau, Lee, *et al.*, 2015; Liau, Tsai, *et al.*, 2015; Chudasama *et al.*, 2018; Ren *et al.*, 2018). Physiologiquement *ATRX* a un rôle dans le maintien de l'hétérochromatine constitutive et spécifique au type cellulaire, son altération pourrait avoir d'autres conséquences dans l'oncogenèse des LMS en dehors du phénotype ATL. C'est ce qui a été étudié dans un projet à part entière, rapporté dans le second article de cette thèse dans la partie « **ATRX : son altération est-elle un driver ?** ».

Dans les hLMS, en plus des spécificités dans les mécanismes d'altérations de *TP53* et *RB1*, nous avons montré une co-expression des gènes et des micro ARN liés aux phénotypes contractile et synthétique. Cette co-expression est associée à deux évènements spécifiques aux hLMS : la surexpression de *MYOCD*, le partenaire de SRF comme cofacteur de transcription de la différenciation musculaire lisse, et la perte de

PTEN, qui, d'après l'hypothèse de Horita *et al.*, permet à SRF de se lier à des cibles dans des promoteurs de gènes de la prolifération (Horita *et al.*, 2016). L'inhibition de l'interaction entre SRF et MYOCD tue les cellules de hLMS plus fortement que celles des oLMS et UPS, validant le rôle essentiel de cet axe dans l'oncogenèse des hLMS. Cependant, plusieurs questions restent en suspens : quelles cibles de SRF et MYOCD ensemble ou séparément sont spécifiques des hLMS et par quels changements l'inhibition de SRF/MYOCD entraîne la mort spécifique des hLMS ? Je me suis efforcée à répondre à ces questions dans la suite de ma thèse développée dans la partie « **Comment la sur-activation de la voie SRF/MYOCD est-elle le *driver* des hLMS ?** ».

B. ATRX : son altération est-elle un *driver* ?

I. Contexte

Notre étude sur les LMS démontre qu'ATRX est le troisième gène le plus muté (12,8%) après *TP53* (48,7%) et *RB1* (17,9%) dans les LMS, confirmant ainsi le travail du TCGA qui l'avait classé dans les trois plus mutés dans l'ensemble des sarcomes à génomique complexe (The Cancer Genome Atlas Research Network *et al.*, 2017). Plus spécifiquement, les études portant sur le séquençage des exomes ou du génome des LMS montrent un taux d'altérations d'ATRX entre 16 et 25% (Watson, Goldberg and Bérubé, 2015; Lee *et al.*, 2017; Chudasama *et al.*, 2018). A ce jour et à ma connaissance, la perte d'ATRX, dans les LMS, n'a été étudiée que dans son association au phénotype ALT (Liau, Tsai, *et al.*, 2015; Chudasama *et al.*, 2018; Ren *et al.*, 2018). Or, ATRX est un remodelleur de la chromatine qui préserve, dans le noyau, les zones d'hétérochromatine constitutive et spécifique au type cellulaire et permet aussi l'expression d'un sous-groupe de facteurs de transcription, les ZNF (Baumann and Fuente, 2009; Valle-García *et al.*, 2016). La perte d'ATRX modifie donc l'expression de nombreux gènes dans des cellules cancéreuses en changeant l'ouverture d'hétérochromatine ou en limitant l'expression de facteurs de transcription. Dans des cellules du cancer du foie, cette perte entraîne de façon très significative une activation de la voie TGF- β et des micro ARN ayant un rôle dans la migration, la survie et l'angiogenèse. Mais aussi une diminution significative de la signalisation des voies des cellules souches et de TNF α (Liang *et al.*, 2019).

Nous nous sommes donc demandé quels sont les impacts fonctionnels des altérations d'ATRX, hormis son rôle dans le phénotype ALT, dans l'oncogenèse des LMS.

II. Article 2 : *ATRX alteration contributes to tumor growth and immune escape in pleomorphic sarcomas*

Lucie Darmusey*, Gaëlle Perot*, Noémie Thébault, Sophie Le Guellec, Nelly Desplat, Laëtitia Gaston, Lucile Delespaul, Tom Lesluyes, Elodie Darbo, Anne Gomez-Brochet, Elodie Richard, Jessica Baud, Laura Leroy, Jean-Michel Coindre, Jean-Yves Blay, Frédéric Chibon

* : co-premier auteurs

Soumis à Cancer Discovery, Octobre 2020

Résumé

L'intégration des mutations et des variants structuraux de la première phase de notre cohorte de LMS (67 cas) donne un taux d'altération d'ATRX de 29,8% dont la majorité amène à la perte de la protéine ATRX au noyau. A noter que 75% des LMS utérins comportent une altération d'ATRX et que celle-ci est ici aussi associée au phénotype ALT. Nous avons par la suite comparé l'expression des gènes entre les LMS avec un ATRX normal et ceux avec un ATRX altéré et mis en évidence la sous-expression des gènes liés à la contraction et au système musculaire. La perte d'ATRX se ferait donc principalement dans des cellules dédifférenciées. Pour comprendre le rôle que peut avoir cette perte dans des tumeurs moins différenciées, nous avons réanalysé une cohorte précédemment publiée composée de 60 sarcomes pléomorphes peu différenciés et caractérisés au niveau de l'ARN. ATRX est altéré dans 16,7% de la cohorte. Le différentiel d'expression montre, dans cette seconde cohorte, une sous-expression des gènes liés à l'immunité lorsqu'ATRX est perdu. Afin de valider les rôles de la perte d'ATRX en dehors de ALT, trois modèles de sous-expression d'ATRX de sarcomes peu différenciés ont été créés. Les expérimentations *in vitro* et *in vivo* ont révélé un phénotype plus agressif des cellules de sarcomes lorsqu'elles sous-expriment ATRX. De plus, une expérimentation *in vivo* dans des hôtes murins immunocompétents démontre une capacité spécifique des cellules sans ATRX à développer des tumeurs. Le séquençage de l'ARN de ces tumeurs et des immunofluorescences sur la tryptase indiquent que cette capacité est probablement liée au non recrutement de mastocytes dans la tumeur.

ATRX alteration contributes to tumor growth and immune escape in pleomorphic sarcomas

Lucie Darmusey^{1,2,3*}, Gaëlle Pérot^{1,4*}, Noémie Thébaud^{1,2}, Sophie Le Guellec^{1,2}, Nelly Desplat⁵, Laëtitia Gaston⁶, Lucile Delespaul¹, Tom Lesluyes¹, Elodie Darbo^{5,7,8}, Anne Gomez-Brouchet^{1,2,4}, Elodie Richard⁵, Jessica Baud⁵, Laura Leroy^{1,2}, Jean-Michel Coindre^{5,9}, Jean-Yves Blay^{10,11}, Frédéric Chibon^{1,2}

ABSTRACT

Whole genome and transcriptome sequencing of a cohort of 67 leiomyosarcomas revealed *ATRX* to be one of the most frequently mutated genes in LMS after *TP53* and *RBI*. While its function is well described in the ALT mechanism, we wondered whether its alteration could have complementary effects on sarcoma oncogenesis. *ATRX* alteration is associated with the down-expression of genes linked to differentiation in LMS, and to immunity in an additional cohort of 60 poorly differentiated sarcomas. *In vitro* and *in vivo* models showed that *ATRX* loss increases tumor growth rate and immune escape by decreasing the immunity load of active mast cells in sarcoma tumors. These data indicate that an alternative to unsuccessful targeting of the adaptive immune system in sarcoma could be to target the innate system. This might lead to a better outcome for sarcoma patients in terms of *ATRX* status.

SIGNIFICANCE: There is still no efficient systemic treatment for pleomorphic sarcomas. Here we show that 1/4 of them have an *ARTX* alteration that diminishes the immune response. This phenotype is related to the inhibition of mast cell recruitment upon *ATRX* alteration, which could be targeted to adapt immunotherapy against pleomorphic sarcomas.

1. OncoSarc, INSERM U1037, Cancer Research Center in Toulouse (CRCT), Toulouse, France. 2. Department of Pathology, Institut Claudius Régaud, IUCT-Oncopole, Toulouse, France. 3. University of Toulouse 3, Paul Sabatier, Toulouse, France. 4. Centre Hospitalier Universitaire (CHU) de Toulouse, IUCT-Oncopole, Toulouse, France. 5. INSERM UMR1218, ACTION, Institut Bergonié, Bordeaux, France. 6. Department of Medical Genetics, CHU de Bordeaux, Bordeaux, France. 7. CNRS UMR5800, LaBRI, Talence, France. 8. Université de Bordeaux, Bordeaux, France. 9. Department of Pathology, Institut Bergonié, Bordeaux, France. 10.

Department of Medical Oncology, Centre Léon Bérard, Lyon, France. 11. Université Claude Bernard Lyon 1, Inserm U1052, CNRS 5286, Cancer Research Center of Lyon, Lyon, France.

Note: * : L. Darmusey and G. Pérot contributed equally to this article

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

Introduction

Pleomorphic sarcomas are a group of rare mesenchymal tumors comprising different histotypes such as undifferentiated pleomorphic sarcoma (UPS), myxofibrosarcoma (MFS), dedifferentiated liposarcoma (DDLPS), osteosarcoma (OS) and leiomyosarcoma (LMS), which is the most frequent subtype (1). LMS has a smooth muscle differentiation and can occur in any anatomical site, although there are three main locations: limbs, trunk and uterus. Currently, the first-line treatment is wide-margin resection for localized tumors and anthracycline-based chemotherapy for advanced tumors (2). However, these treatments are still not effective enough as 48 to 89% of LMS develop metastases depending on the tumor location and the mortality rate is between 50 and 65% with a median survival of around 12 months (3,4). From a genomic standpoint, LMS, like other pleomorphic sarcomas, have a very rearranged and unbalanced karyotype (2).

In a whole genome and whole transcriptome sequencing study conducted by our team, Darbo *et al.* showed that LMS have a low somatic mutation burden but a high level of copy-number alterations (Darbo *et al.*, 2020). But only three genes came out to be recurrently mutated, as also showed by the TCGA study (5), *TP53* (mutated in 48.7%), *RBI* (mutated in 17.9%) and *ATRX* (mutated in 12.8%). *RBI* (6,7) and *TP53* (8,9) are tumor suppressor genes that have long been known to be implicated in the oncogenesis of pleomorphic sarcomas. *ATRX* is a chromatin modifier gene with a Swi/Snf2 domain (10). Its tumor suppressive function has so far been related to its role in the alternative lengthening of the telomere (ALT) mechanism (11), inducing genome instability (12) and

leading to a poor prognosis of *ATRX*-altered tumors(13). Recently, its involvement in senescence (14) and in intrinsic immunity *via* its interactions within promyelocytic leukemia nuclear bodies (PML NBs) (15) was questioned. Here, we investigated whether *ATRX* might have additional impacts in the oncogenesis of pleomorphic sarcomas beyond its role in the ALT mechanism and show how its involvement in oncogenesis is also linked to differentiation, tumor growth and immunity.

Results

Distinct genetic alterations trigger loss of *ATRX* protein

Sixty-seven LMSs (Cohort 1) were sequenced at the whole genome and transcriptome levels (28 LMSs + 39 LMSs from Darbo *et al.*, 2020) and *ATRX* was identified as the third most frequently mutated gene after *TP53* and *RBI*. By integrating mutations and structural variations (SV), it is altered in 20 cases (29.8 %; Figure 1), with 8 point mutations (missense (MS) and non-sense (NS); 40%), 7 frameshifts (FS; 35%) and 5 structural variants (SV; 25%). All mutations and SVs were validated by an independent technique (RNA sequencing and/or Sanger sequencing) (Supplementary Tables 1 and 2; Supplementary Figures 1, 2, and 3). *ATRX* was altered in 75% of uterine LMSs (6/8) and in 23.7% of non-uterine LMSs (14/59), which is significantly lower ($P = 7.002 \times 10^{-3}$; Figure 2A and Supplementary Figure 4). Regarding SVs, 3 out of 5 led to a loss of *ATRX* expression and the other two led to a frameshift (Supplementary Table 1). These alterations were hemizygous in the three males due to the location of *ATRX* on chromosome X (Xq21), and in two females with either a deletion of

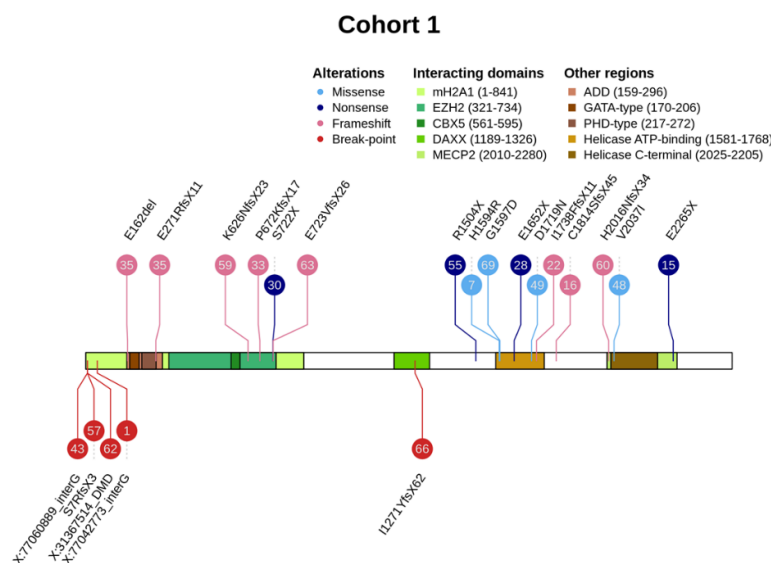


Figure 1: *ATRX* mutations and structural variants. *ATRX* alterations are color-coded according to their type (legend at the top). Numbers in bubbles represent tumor sample. Consequences of all point mutations on *ATRX* protein are annotated above a schematic representation of the protein, or below for two structural variants (LMS57 and LMS66). For the other three structural variants, annotations correspond to the break-point partner in genomic coordinates.

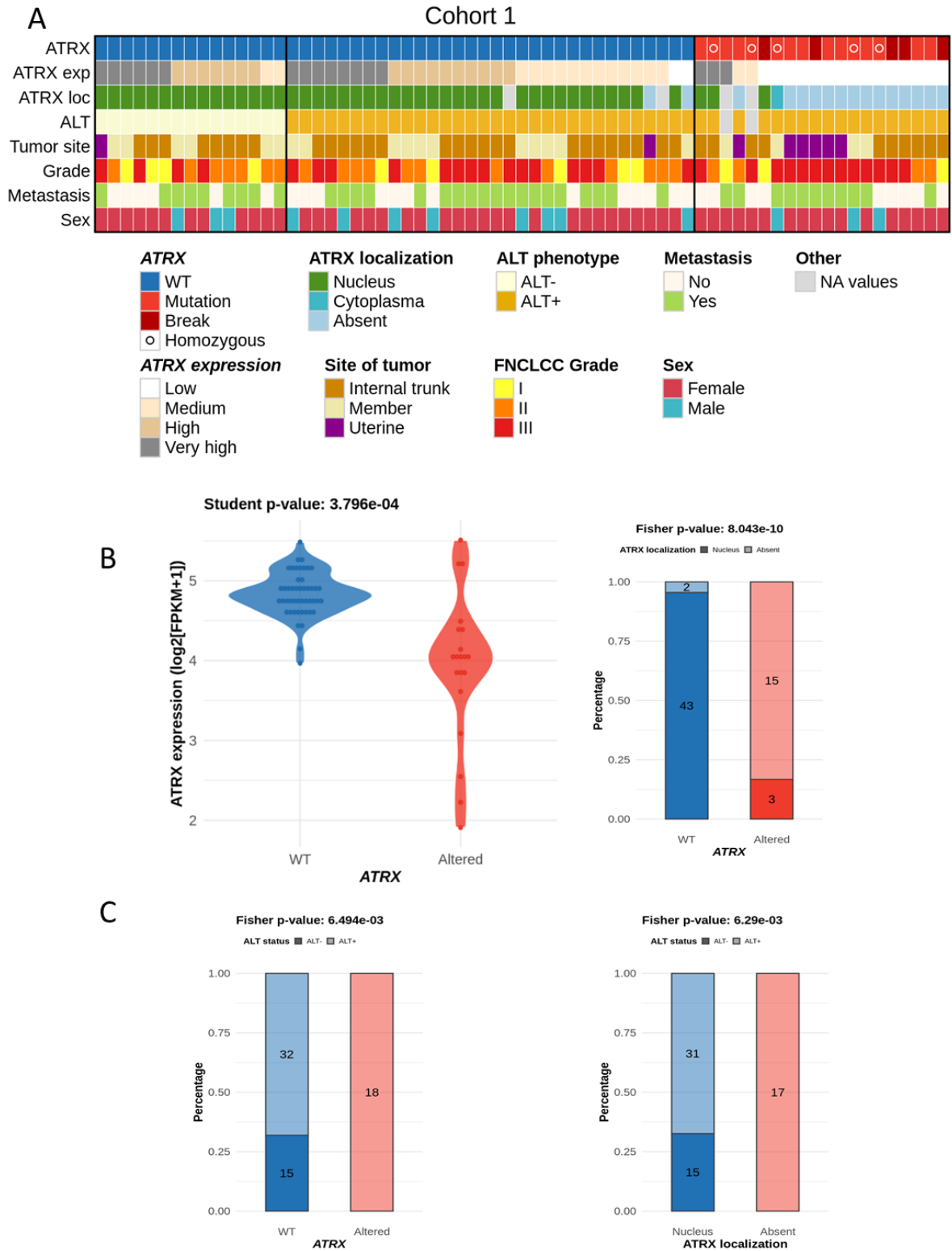


Figure 2: ATRX status and integrated representation. **A** Integrated representation shows ATRX alterations, ATRX mRNA expression (by quartile), ATRX localization, ALT mechanism phenotype, tumor site, FNCLCC grade, presence or not of metastasis and sex of each patient. Tumors are ordered by ATRX status, ALT phenotype, mRNA expression and protein localization. **B** Association between ATRX alteration and its mRNA expression (log₂(FPKM+1)) (left) or its protein localization (right). **C** Relation between ATRX status (left) or its protein localization (right) and ALT mechanism phenotype. For ATRX localization, the “absent” group means “not at the nucleus”, including all cases without expression and the case with a cytoplasmic localization (LMS16). P-values were calculated with Student test for **B** - left and with Fisher test for **B** - right and **C**.

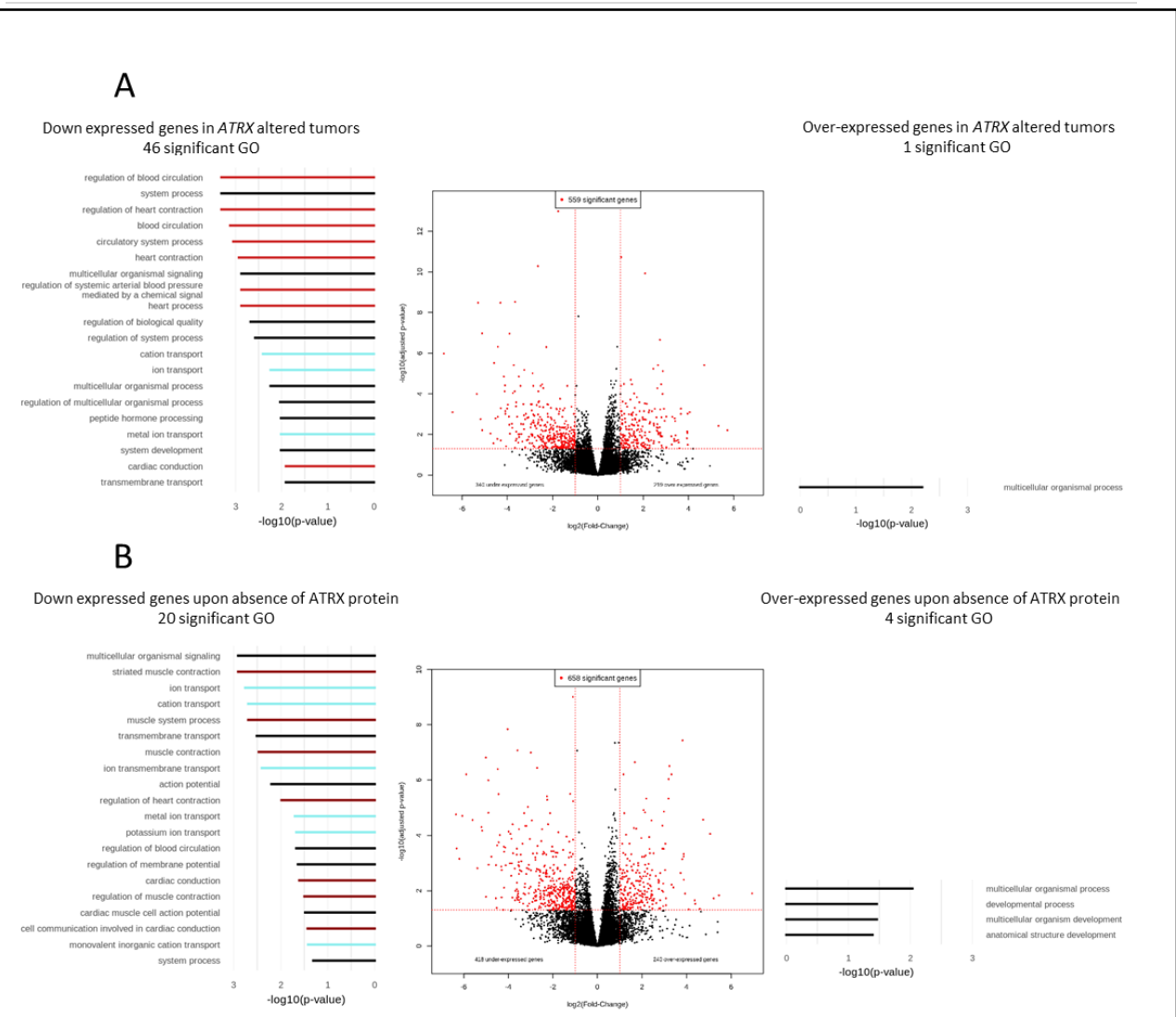


Figure 3: Differential gene expression and Gene Ontology analyses according to *ATRX* alteration in leiomyosarcomas (Cohort 1). Differentially expressed genes according to **A** *ATRX* status (wild-type vs altered) or **B** *ATRX* expression (nucleus vs absent). Red dots indicate significant genes ($P \leq 0.05$ and fold-change ≤ -2 or ≥ 2). Gene Ontology (GO) analyses, represented on the left (under-expressed genes) and the right (over-expressed genes), identified 46 and 1 significant GO terms ($P \leq 0.05$), respectively in **A** and 20 and 4 significant GO terms ($P \leq 0.05$) in **B**. On each side, the 20 most significant GO terms are represented and color-coded by mechanisms; light red, dark red, light blue and black colors indicate “circulatory system process”, “muscle system process”, “ion transport” and general terms, respectively. For *ATRX* localization, the “absent” group means “not at the nucleus”, including all cases without expression and the case with a cytoplasmic localization (LMS16). All p-values adjusted by Benjamini and Hochberg method.

the second allele (LMS69) or an isodisomy (LMS49). In the other 15 females, 93.3% of the alterations (14/15) occurred on the active X, as RNAseq analysis showed the altered transcript expression (Supplementary Table 1). No expression of the mutated allele was detected in LMS48 (Supplementary Table 1 and Figure 2). Alterations were distributed throughout the whole gene but two regions were most frequently affected: one between exons 17 and 21 (40%, 6/15) and the other in exon 9 (33.4%, 5/15). At the mRNA level, mutated cases had a significantly lower *ATRX* expression than wild-type (WT) tumors ($P = 3.79 \times 10^{-4}$; Figure 2B) and at the protein level, alterations led to a loss of nuclear protein ($P = 8.04 \times 10^{-10}$; Figure 2B and Supplementary Figure 5).

Is *ATRX* alteration linked to ALT mechanism?

Since *ATRX* loss is linked to the ALT phenotype (16), the ALT status of tumors was determined. Most LMSs were ALT-positive (ALT+, 76.9%, 50/65) (Figure 2A and Supplementary Figure 6). Both *ATRX* alteration ($P = 6.49 \times 10^{-3}$) and *ATRX* protein loss ($P = 6.29 \times 10^{-3}$) were significantly associated with the ALT mechanism (Figure 2C). However, while all *ATRX*-altered cases were ALT+, most ALT+ cases were *ATRX* WT (64%, 32/50) with 93.3% of cases (28/30) expressing the protein in the nucleus (Figures 2A and 2C).

Is *ATRX* alteration associated with prognosis?

Neither *ATRX* status (altered or WT), mRNA expression (below or above defined cut-off, see material and methods section), protein localization (nuclear or absent) nor ALT phenotype (positive or negative) could split patients into two groups with

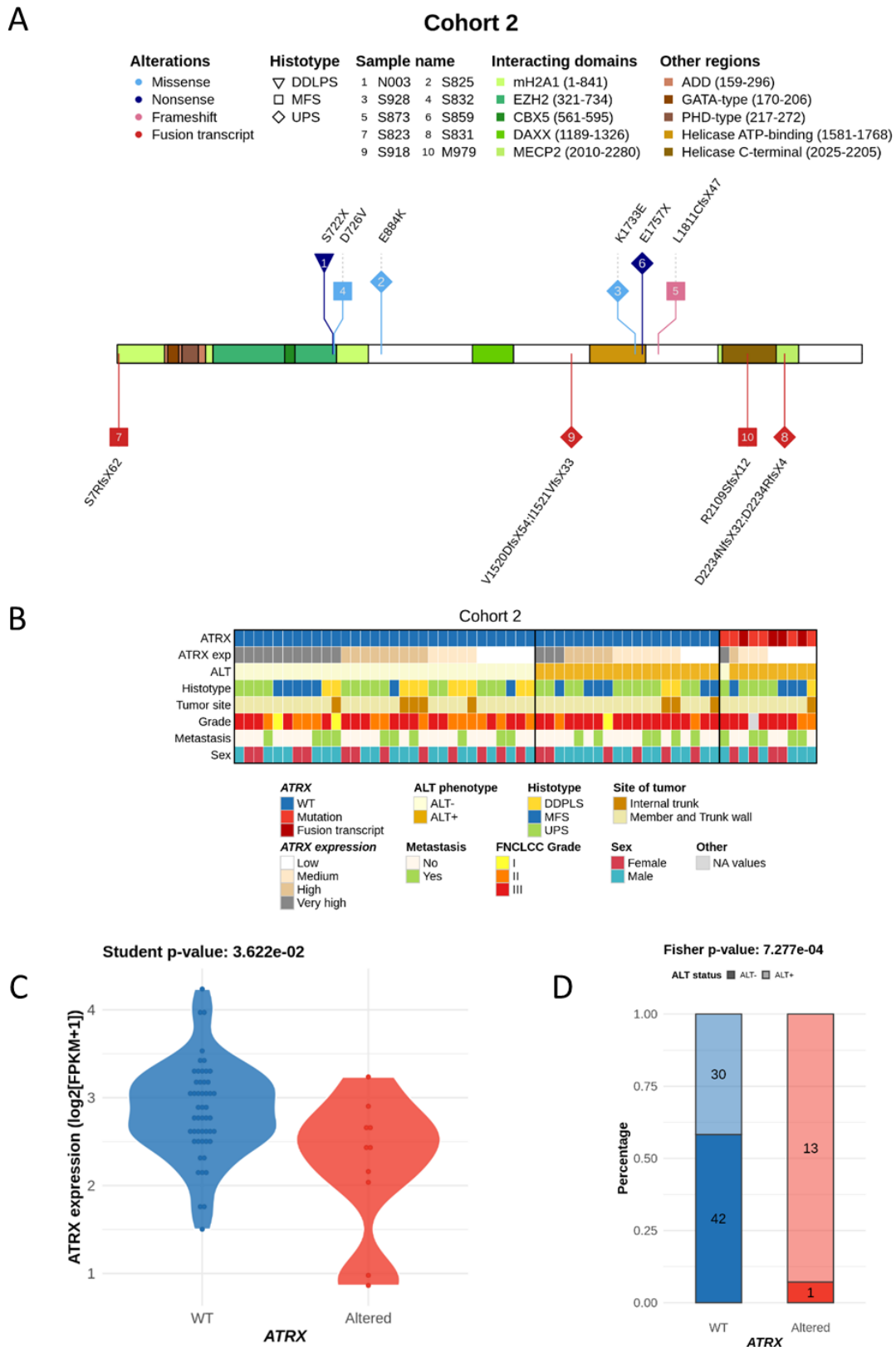


Figure 4: *ATRX* alterations and integrated representation in poorly differentiated pleomorphic sarcomas (cohort 2). **A** *ATRX* alterations are color-coded by their type and shapes represent histotypes. Numbers in bubbles indicate a tumor sample (legend at the top). Translated consequences on *ATRX* protein is annotated above a protein schematic representation for mutations, or below for fusion transcripts. **B** Integrated representation shows *ATRX* alterations, mRNA expression (by quartile), ALT mechanism phenotype, histotypes, tumor site, FNCLCC grade, presence or not of metastasis and sex of each patient. Tumors are ordered by *ATRX* status, ALT phenotype, mRNA expression and histotypes. **C** Association between *ATRX* status and its mRNA expression ($\log_2(\text{FPKM}+1)$). **D** Relation between *ATRX* status and ALT phenotype.

significantly distinct prognoses (Supplementary Figure 7).

Which transcriptomic program is modified upon ATRX alteration?

Searching for the oncogenic impact of these ATRX alterations, we tested whether altered tumors had a distinct transcriptomic program and identified 340 and 219 genes significantly down- and up-expressed in the ATRX-altered group, respectively ($P < 0.05$; Figure 3A). Functional enrichment analysis (Figure 3A) showed that genes down-expressed were significantly involved in blood pressure, heart contraction and striated muscle contraction. These findings were strengthened when patients were grouped according to ATRX protein localization, since genes down-expressed upon protein loss were found to be involved in similar biological mechanisms, *i.e.* muscle system and contraction (Figure 3B).

As expected, clustering based on these 559 differentially expressed genes (Supplementary Figure 8) revealed a group with a high percentage of ATRX-altered patients (85%, 17/20). These patients had tumors that were uterine or of “other” LMS type (Darbo *et al.*, 2020). “Other” LMSs are less differentiated than “homogeneous” LMSs and are thought to derive from fibroblasts rather than smooth muscle cells (SMCs) (Darbo *et al.*, 2020).

Altogether, these data suggested that either ATRX alteration preferentially occurs in partially or undifferentiated cells, or that it may induce dedifferentiation. To explore these hypotheses, we studied the ATRX status in a second cohort comprising

poorly differentiated pleomorphic sarcomas characterized by RNAseq.

What are ATRX alterations in poorly differentiated pleomorphic sarcomas?

RNA sequencing of 60 pleomorphic sarcomas (cohort 2) from a previously published cohort (17) was reanalyzed and 10 ATRX-altered tumors (16.7%) were identified (Supplementary Tables 1 and 3; Supplementary Figures 9 and 10). The types of mutation as well as their functional consequences were similar to those detected in cohort 1 (Figure 4A). Mutated cases were not enriched in any annotation (*i.e.* histotype, tumor site, grade, metastasis or sex) (Figure 4B) but had a significantly lower mRNA expression of ATRX ($P = 3.62 \times 10^{-2}$; Figure 4C) and were significantly associated with ALT ($P = 7.27 \times 10^{-4}$; Figure 4D). ATRX-altered tumors did not have a distinct prognosis in cohort 2 (Supplementary Figure 11A), nor when the two cohorts are merged (Supplementary Figure 11B).

What are the transcriptomic modifications in poorly differentiated sarcomas upon ATRX alteration?

Functional enrichment analysis of differentially expressed genes showed that ATRX alteration induced the overexpression of 76 genes enriched in GO terms related to the metabolic process, and the down-expression of 506 genes enriched in GO related to immunity. The five most significantly enriched GO were (Figure 5) “T cell activation” ($P = 8.41 \times 10^{-19}$), “lymphocyte activation” ($P = 1.13 \times 10^{-17}$), “immune system process” ($P = 4.82 \times 10^{-16}$), “leukocyte

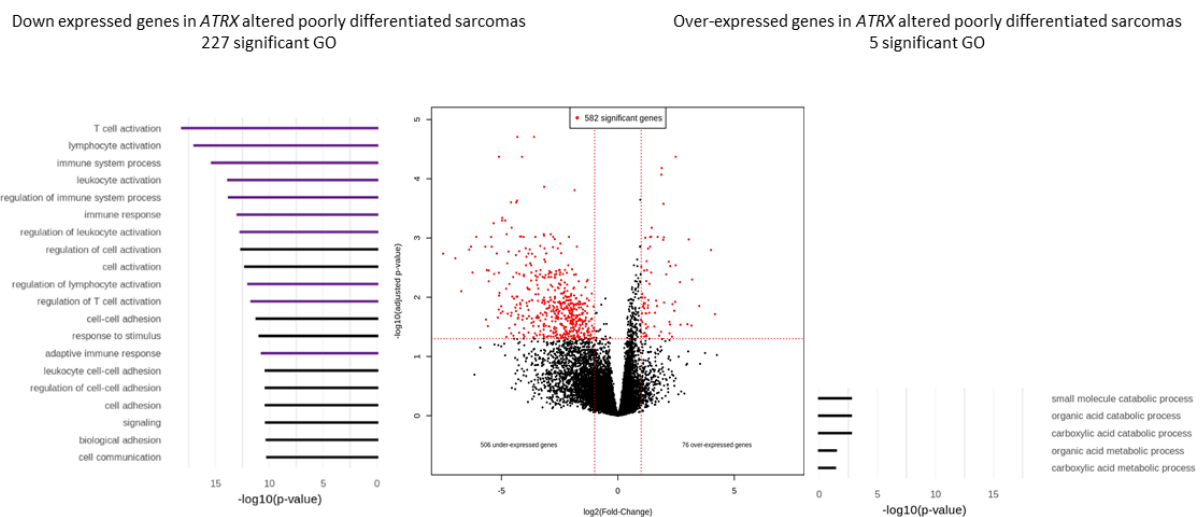


Figure 5: Differential gene expression and Gene Ontology analyses according to ATRX status (wild-type vs altered) in poorly differentiated sarcomas (Cohort 2). Differentially expressed genes in ATRX-altered tumors are represented in red ($P \leq 0.05$ and fold-change ≤ -2 or ≥ 2). Gene Ontology (GO) analyses, represented on the left (under-expressed genes) and the right (over-expressed genes), identified 227 and 5 significant GO terms ($P \leq 0.05$), respectively. On the left, the 20 most significant GO terms are represented and color-coded by mechanism; purple and black groups indicate “immunity system process” and general terms, respectively. All p-values adjusted by Benjamini and Hochberg method.

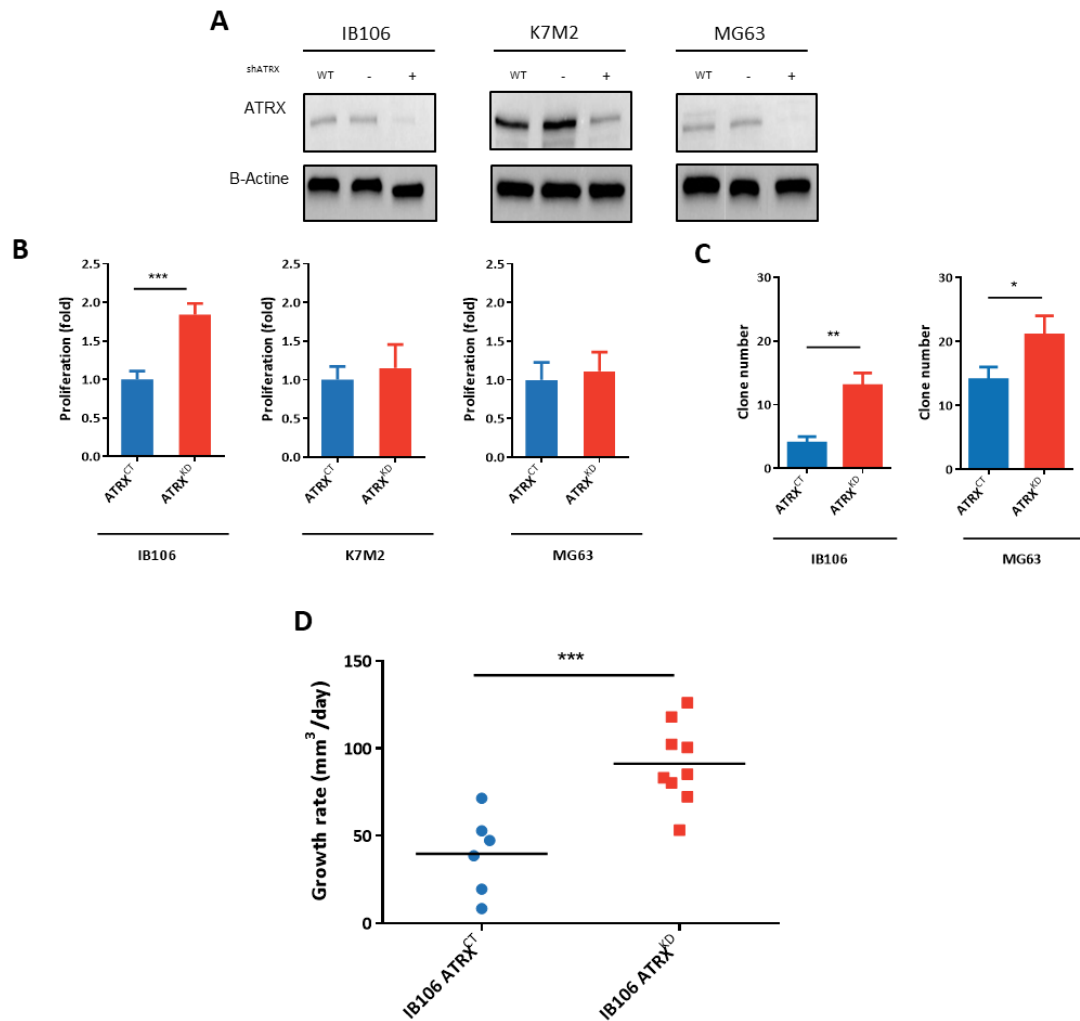


Figure 6: ATRX knock-down increases aggressiveness of sarcoma cells. **A** ATRX knock-down by shRNA validation in western blot in K7M2, MG63 and IB106 cell lines. **B** Proliferation analysis by MTT after 4 days, comparing ATRX^{CT} and ATRX^{KD} cells in K7M2, MG63 and IB106 cell lines (mean \pm s.d.; n = 3 independent experiments). **C** Soft agar assay analysis comparing ATRX^{CT} and ATRX^{KD} cells in K7M2, MG63 and IB106 cell lines (mean \pm s.d.; n = 4 independent experiments). Images were taken after 4 weeks and crystal violet staining. **D** Tumor growth rate analysis of IB106 ATRX^{CT} or IB106 ATRX^{KD} cells sub-cutaneous xenografts on NSG mice (N = 10 in each group). Growth rate is calculated by segmental linear regression with GraphPad. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, p-value was calculated with 2-way ANOVA for **A** and unpaired t-test for **B**, **C** and **D**.

activation” ($P = 1.58 \times 10^{-14}$) and “regulation of immune system process” ($P = 1.82 \times 10^{-14}$).

Results from both cohorts indicated that *ATRX* alteration is associated with differentiation and immunity. Since this is particularly relevant in the current clinical context, we functionally tested the hypothesis that *ATRX* alteration might modify the anti-tumor immune response.

Does *ATRX* knock-down impact oncogenic features?

To functionally test the impact of *ATRX* alterations, three models of *ATRX* knock-down (ATRX^{KD}) were constructed: i) a model to evaluate tumor growth *in vitro* and *in vivo* in a human UPS cell line (IB106), ii) another to study immunity in immunocompetent mice (Balb/c) with a mouse poorly differentiated OS (K7M2; a sarcoma with very close genetics to UPS

and LMS) and iii) a third to compare human and mouse, a human OS cell line (MG63). These cell lines were infected by lentivirus with an *ATRX* shRNA. Western blot evidenced the successful extinction of *ATRX* in each cell line (Figure 6A). ALT analysis showed that *ATRX* shRNA did not change ALT status in any cell line (Supplementary Figure 12).

In vitro, a significant ($P < 0.0001$) increase in proliferation was observed in the UPS cell line IB106 ATRX^{KD} but not in OS cell lines (K7M2 and MG63) (Figure 6B). Colonies formed in soft agar assay revealed that the mouse cell line K7M2 was unable to form any colony with or without *ATRX* expression. In contrast, there was a significant increase in colony number in human cell lines IB106 and MG63 upon *ATRX* down-expression, from a mean of 14 to 21 colonies ($P = 2.6 \times 10^{-3}$) and from 4 to 13 ($P = 1 \times 10^{-3}$), respectively (Figure 6C). Next, IB106 ATRX^{CT}

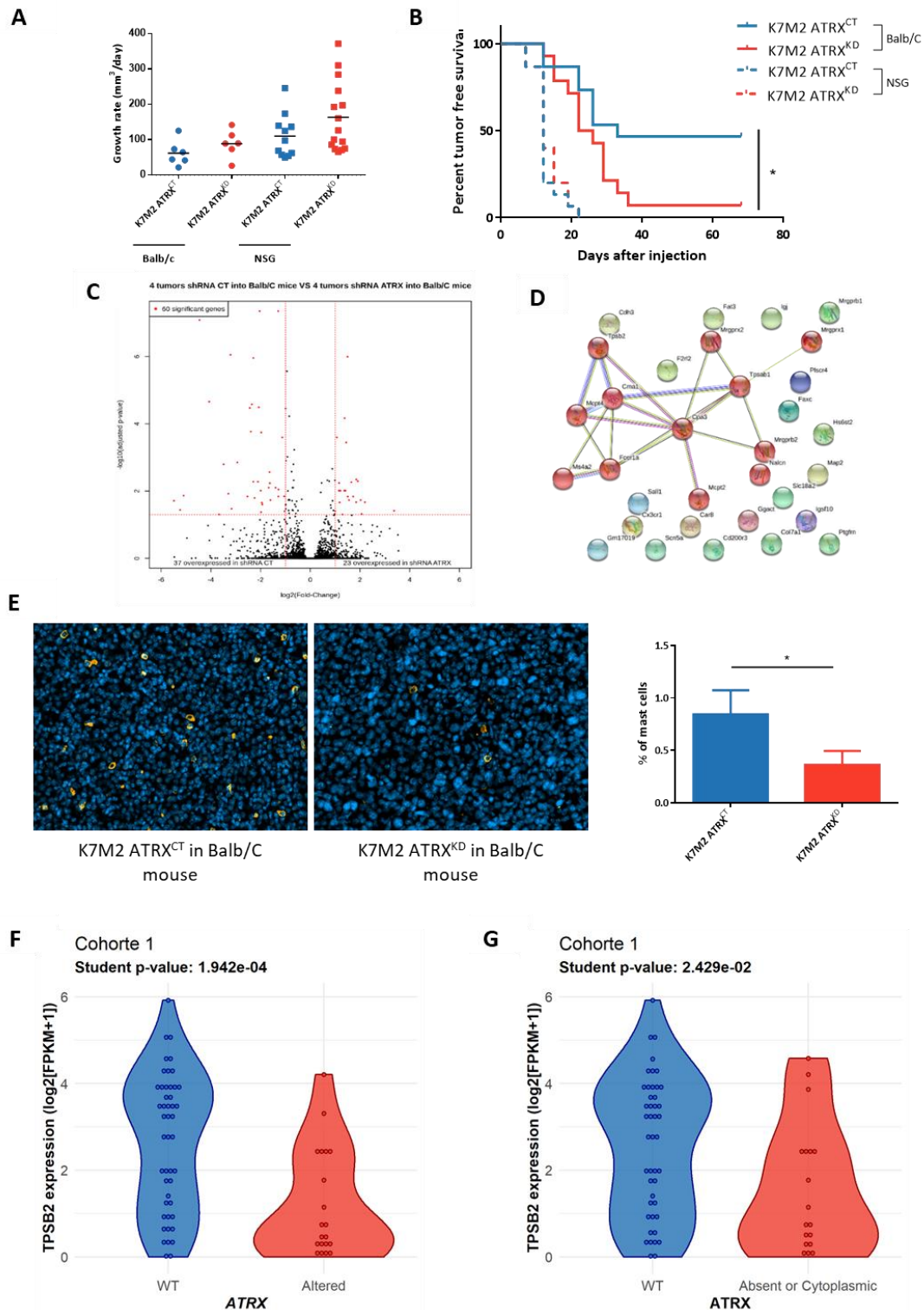


Figure 7: ATRX knock-down allows immune escape of sarcomas *via* non-recruitment of mast cells. **A** Tumor growth rate analysis of K7M2 ATRX^{CT} or K7M2 ATRX^{KD} cells xenografted under the skin of NSG or Balb/c mice (N = 15 in each group) **B** Tumor-free survival curves of K7M2 ATRX^{CT} or ATRX^{KD} tumors in immunodeficient NSG mice and immunocompetent Balb/c mice (N = 15 mice for each condition) using Kaplan-Meier method. **C** Comparison of RNA expression in log₂(FPKM+1) of K7M2 ATRX^{KD} tumors *versus* K7M2 ATRX^{CT} tumors developed in immunocompetent mice (N=4 each) showing 23 and 37 significantly up- and down-expressed genes in K7M2 ATRX^{KD} tumors, respectively. **D** Links between down-expressed genes in K7M2 ATRX^{KD} tumors found by the STRING Database showing one cluster with genes involved in mast cells *via* MCL clustering. **E** Immunostaining of mast cells by targeting tryptase in K7M2 ATRX^{CT} and K7M2 ATRX^{KD} tumor tissues with nucleus marked with DAPI. On the right, percent of mast cells in the two conditions. **F** TPSB2 mRNA expression in log₂(FPKM+1) according to ATRX status in cohort 1. **G** TPSB2 mRNA expression in log₂(FPKM+1) according to ATRX localization in cohort 1. *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, p-value was calculated with Mantel-Cox test for **B** and unpaired t-test for **E**, **F**, **G**.

(control) and ATRX^{KD} cells were subcutaneously grafted in 10 NSG mice each. A tumor grew in 6/9 ATRX^{CT} group and in 9/10 ATRX^{KD} group. Tumor growth rates were three-fold higher in ATRX^{KD} tumors (91.2 ± 7.6 mm³/day) than in ATRX^{CT} tumors (32.9 ± 10.6 mm³/day) ($P = 5 \times 10^{-4}$, Figure 6D).

Does ATRX alteration modify anti-tumor immune response?

The involvement of ATRX alteration in immune escape was tested by grafting K7M2 ATRX^{CT} and ATRX^{KD} cells in immunodeficient NSG mice and in immunocompetent Balb/c mice (N=15 for each group). Growth rate was not significantly increased upon ATRX knock-down in any hosts (Figure 7A). Tumor-free survival in ATRX^{CT} and in ATRX^{KD} models displayed no significant differences in immunodeficient NSG mice, whereas in immunocompetent Balb/c mice there was 53.4% (8/13) of tumor induction with K7M2 ATRX^{CT} versus 92.8% (13/14) with K7M2 ATRX^{KD}. Therefore, tumor-free survival was significantly poorer upon ATRX knock-down ($P = 9.7 \times 10^{-3}$; Figure 7B).

Differential gene expression analysis between ATRX^{KD} vs ATRX^{CT} K7M2 tumors in Balb/c mice revealed that 37 genes were down-regulated and 23 genes were overexpressed in ATRX^{KD} tumors (Figure 7C). The low number of genes precluded any functional enrichment analysis. Consequently, a String Protein Interaction (Szklarczyk *et al.*, 2019) analysis was performed. Whereas no consistent clusters arose with up-regulated genes upon ATRX knock-down (Supplementary Figure 13), one emerged in down-regulated genes, with 12 proteins out of 37 linked to mast cell pathways (including *TPSB2* coding tryptase, a widely used mast cells marker) (Figure 7D).

Immunofluorescence against tryptase on the murine tumors previously processed in RNAseq showed that mast cells expressing tryptase represented a mean of 0.8% of total cells in ATRX^{CT} tumors, whereas they constituted 0.3% of ATRX^{KD} tumors ($P = 0.01$; Figure 7E). This significant difference prompted us to assess whether the proportion of infiltrating mast cells in human sarcomas is also related to ATRX alteration and the absence of ATRX from the nucleus. In the LMS cohort, the only one fully characterized at every level, *TPSB2* was significantly under-expressed in ATRX-altered cases ($P = 1.9 \times 10^{-4}$; Figure 7F) and in tumors with no nuclear ATRX ($P = 0.02$; Figure 7G), which might indicate that there are fewer infiltrating mast cells in these human LMS.

Discussion

This in-depth ATRX genetic analysis revealed that ATRX alteration likely affects a quarter of pleomorphic sarcomas, since it was found in 29.8% of LMS and in 16.7% of undifferentiated sarcomas (US).

Cohort 2 is less deeply characterized (WGseq for cohort 1, RNAseq for cohort 2), so cases might be missed with this RNAseq-based screening (20.7% observed for UPS/MFS/DDLPs in TCGA (5)). The rate of alteration in LMS is consistent with the rate of 24% found by Chudasama and colleagues (19), but it is slightly higher than that generally observed in other LMS cohorts, which is around 16% (5,20,21), probably due to the exhaustiveness of WGseq. ATRX mutations were distributed across the entire gene, as previously observed (5,19,22,23). Three main factors link the two types of sarcomas in the present study: i) in females, all alterations except in LMS48 can be interpreted as occurring on the active X, ii) point mutations are more frequent (75% in LMS and 60% in US) than structural variations (25% in LMS and 40% in US), as previously observed (5,19); and iii) the alterations lead preferentially to a frameshift and thus to a truncated protein in 66.7% of cases (20/30, 65% in LMS and 70% in US), in agreement with previous descriptions in sarcomas (5,19,22,14). Of note, ATRX alterations in the present study were not significantly associated with a poorer prognosis. However, this association depends on the cohorts studied (13,23) and was observed in only one cohort that mainly included missense mutations (23).

ATRX mutated cases were also linked to the location of LMS, i.e. 75% of uterine cases were ATRX-altered (6/8). Loss of ATRX in uterine tumors is a key difference between benign and malignant tumors. In this location, it has been proposed to use ATRX loss as a marker of the highly probable evolution of benign tumors toward malignancy (25). In other LMS locations, ATRX loss is linked to the “other” LMS group. LMSs belonging to this subtype are mainly poorly differentiated and likely originate from fibroblastic cells (Darbo *et al.*, 2020). Furthermore, as ATRX loss in LMS is associated with a lower expression of genes related to smooth muscle activity, we hypothesize that it occurs preferentially in poorly differentiated cells. The degree of cell differentiation may be crucial for the loss of ATRX to confer advantages to the precursor sarcoma cell.

ATRX knock-down modifies tumor cell proliferation and clonogenicity in sarcoma models, as confirmed *in vivo* where ATRX knock-down tumors grew three-fold faster than controls. Interestingly, poorly differentiated sarcomas with ATRX alteration overexpressed genes related to metabolism, whose upregulation is a known hallmark of cancers and supports cell survival and proliferation (26). The hypothesis that ATRX could act through metabolism regulation is a very appealing one that now requires functional validation.

In vivo experimentation revealed a new role of ATRX, as its alteration was associated with a poorer outcome exclusively in an immunocompetent murine host, and

with down-expression of immune-related genes in poorly differentiated human sarcomas. These two findings show that ATRX loss can influence the regulation of immune response in sarcomas, probably by limiting mast cell recruitment, as evidenced by the lower proportion of tumor infiltrating mast cells upon ATRX down-expression. The role of mast cells in tumor control is currently considered as dual and antagonistic, since they can support tumorigenesis or suppress tumor growth. Their role is dependent on the type of tumor (27). To our knowledge, no study has yet investigated the role of mast cells in the oncogenesis of sarcomas. *FceRI* and *Ms4a2* are two down-expressed genes in ATRX^{KD} K7M2 tumors. They are part of the IgE activating mast cell pathway that confers them a protective role in epithelial tumors (28). In addition to their higher proportion, these mast cells present in ATRX^{WT} tumors likely play a suppressor role in which they recruit other immune cells to tumor sites by enhancing vascular permeability and direct chemoattraction (29). In human LMS, the absence of ATRX is linked to the down expression of *TPSB2*, which is a protein produced almost exclusively by mast cells and is widely used to identify them. Furthermore, genes down expressed by ATRX-altered poorly differentiated sarcomas are mostly linked to adaptive immune cell activation, so adaptive immune cells are either less present or less active. This could be achieved by avoiding the release of chemoattractants and hence the recruitment or activation of other immune cells. The precise mechanism involved upon ATRX loss that changes the immune microenvironment of sarcomas needs to be deciphered.

Regarding ATRX expression, 27% of cases (17/63) showed no nuclear ATRX protein, which is consistent with the literature (13,22,23,30). In tumors presenting FS/NS, 87.5% (14/16) exhibited no ATRX protein at all. In these cases, ATRX mRNA level was low, likely meaning that if the truncated protein is expressed (missed by our screening with the C-terminal antibody), it should be very low. Moreover, if truncated proteins are expressed, the lost domains should be the same in all studied sarcoma types, with partial or complete loss of the helicase C-terminal domain in 90% of cases (18/20) and of both helicase domains in 70% of cases (14/20). As the majority of MS mutations occurred in one helicase domain (71.4%, 5/7) and IHC detected a nuclear localization of the protein, a decrease in ATRX enzymatic activity may be hypothesized (31). Collectively, these results suggest that alterations of ATRX preferentially target its enzymatic functions rather than its protein-protein interactions, thus explaining why mutations in ATRX partner genes (*i.e.* *DAXX*, *EZH2*, *SP100*) are not frequent and not an alternative to ATRX alteration in sarcomas. We thus hypothesize that, by modifying its chromatin remodeling action, alterations of ATRX trigger a specific transcriptomic program that

promotes attenuated mast cell recruitment, leading to the observed immune response in models and human tumors.

Our findings show that ATRX alterations are quite frequent in pleomorphic sarcomas (close to 25%) and mostly lead to the loss of ATRX. In addition, they demonstrate that ATRX alterations are not only associated with ALT phenotype but also with differentiation and immune response regulation through non-recruitment of mast cells. Currently, most immunotherapies of sarcomas, which target the adaptive immune system and specifically T cells by helping them to recognize tumors, have a low response rate (32). Indeed, several recent trials have assessed the response to checkpoint inhibitors, which are used to thwart immune system escape by activating CD8+ cytotoxic T cells, with an overall response rate not higher than 25% (33,34). A better response rate with around 66% of response was reached by transferring autologous T cells transduced with a T cell receptor directed against a cancer antigen (CAR-T cell) expressed by the selected tumors, but only 6 patients were involved (35). As targeting the adaptive immune system does not work well in sarcomas, some have tried to target the innate immune system by making therapeutic vaccines which rely on the activation of dendritic cells in the presence of predetermined immunogenic antigen (32). One trial presented 10 out of 23 patients who lived more than 1 year whereas others died after around 7 months (36) and another one showed a 1-year progression-free survival of 70.6% (37). Targeting the innate system might therefore lead to a better outcome for sarcoma patients that could be further improved by assessing ATRX status before testing mast cell-enhancing therapies, as they have been successful in other solid tumors (38). These therapies enhanced local mast cell degranulation by using IgE antibodies, as proposed by Singer and Jensen-Jarolim (39). This strategy could be useful in ATRX WT tumors to enhance the anti-tumoral action of mast cells and, in ATRX-altered sarcomas, to enhance mast cell recruitment and activation (40).

Methods

Collection of samples

Samples used in cohort 1 were collected prospectively by the French Sarcoma Group as part of the ICGC program (International Cancer Genome Consortium). Samples used in cohort 2 were part of the cohort used in Lesluyes *et al.* (2016). ICGC Whole-Genome sequencing and RNA sequencing data for the 67 LMS are available at <https://dcc.icgc.org/projects/LMS-FR>.

Tumor samples and histological classification

One hundred and twenty-seven soft tissue sarcomas were selected according to their histological subtype. Cases were obtained from the collective database of

the French Sarcoma Group (<https://conticabase.sarcomabcb.org>). All cases were systematically reviewed by expert pathologists of the French Sarcoma Group according to the World Health Organization guidelines (41). As such, tumors were classified into four categories: leiomyosarcomas (n=67), undifferentiated pleomorphic sarcomas (n=30), myxofibrosarcomas (n=17) and dedifferentiated liposarcomas (n=13) (Supplementary Table 4). Clinico-pathological data and patient information are summarized in Supplementary Table 4. The 67 leiomyosarcomas are part of the ICGC program (International Cancer Genome Consortium) (cohort 1). Constitutional DNA and tumor DNA and RNA were available for these cases. For the other cases (cohort 2), only tumor DNA and RNA were available.

Validation of *ATRX* alterations

For cohort 1, all FS were verified at both DNA and RNA levels by Sanger sequencing. All SV were verified on DNA by Sanger sequencing and the effect on RNA was detected by RNAseq. Regarding MS and NS mutations, only those not found in both WGS and RNAseq were verified by Sanger sequencing. Complete deletion of *ATRX* due to total loss of chrXq or chrX was seen in 4 females all presenting RNA expression and nuclear protein, implying that the loss occurred on the inactive X. One triploid tumor developed in a male also presented a deletion of the gene but with one copy left. Therefore, they were all considered as WT concerning *ATRX* alteration. For cohort 2, whole *ATRX* cDNA was sequenced by Sanger sequencing for all cases and alterations found at RNA level were verified on DNA.

PCR on genomic DNA

For screening of mutations on genomic DNA, PCR primers were designed using the Primer 3 program (<http://frodo.wi.mit.edu/primer3/>) and are presented in Supplementary Table 5. 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 with the PCR program described in the Supplementary Table 5 Legend.

RT-PCR

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 (<http://frodo.wi.mit.edu/primer3/>). For *ATRX* cDNA screening, primers used are presented in Supplementary Table 6. For fusion transcript detection, control PCR were first performed with different forward and reverse primers for each gene

implicated in the fusion, and then PCR was performed using a forward primer for one gene and reverse primer for the other gene (Supplementary Table 7). All PCR were performed as previously described for PCR on genomic DNA.

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, Rockville, MD, USA). FinchTV software (V1.4.0) was also used (Geospiza, Seattle, WA, USA).

Immunohistochemistry

Sixty-seven tumors in cohort 1 were analyzed on tissue microarrays. Each case was represented by three spots 4-µm-thick and 1mm in diameter. Immunohistochemistry was performed on a BenchMark Ultra instrument (Ventana, Washington D.C, USA). Antigen retrieval was performed using a CCI protocol for 16 min at 98°C (Ventana, Washington D.C, USA), and the anti-*ATRX* antibody (1:1000, Clone BSB-108, Diagomics, Blagnac, France) was diluted in PREPKIT9 for 20 min. Antibody detection was performed using the Optiview detection kit (860-099, Ventana, Washington D.C, USA). Immunohistochemical pictures were taken using a Panoramic 250 Flash II Digital Slide Scanner and analyzed with the Panoramic Viewer (3DHISTECH Ltd., Budapest, Hungary).

Immunolabeling for *ATRX* was considered as positive if tumor cells had nuclear labeling, whatever its intensity (1, 2 or 3), with no evidence of cytoplasmic labeling. Neoplasms were scored as negative for *ATRX* if there was no labeling. One tumor presenting cytoplasmic sequestration with a strong intensity was considered as interpretable. The internal controls (inflammatory and endothelial cells) had to be positive with a nuclear labeling; otherwise the case was considered as not interpretable.

Immunofluorescence

One hundred and twenty-seven tumors from the two cohorts were analyzed on tissue microarrays. Tissues were deparaffinized in xylene and rehydrated in a series of ethanol baths. For antigen retrieval, slides

were incubated in DAKO Target Retrieval Solution, pH 9 (S236784-2, DAKO, Carpinteria, CA, USA), for 20 min in a microwave oven. The primary antibodies and dilutions (dilution in DAKO REAL antibody diluent, S202230-2, DAKO, Carpinteria, CA, USA) used to study ALT were as follows: anti-PML (PG-M3, 1:200, sc-966, Santa-Cruz, Dallas, TX, USA) and anti-TERF2 antibody (1:200, HPA001907, Sigma, St Louis, MO, USA). All primary antibodies were incubated for 1h at room temperature (RT). Secondary antibodies and dilutions used were as follows: anti-Mouse Immunoglobulins/FITC (1:400, F0479, Dako, Carpinteria, CA, USA) and anti-Rabbit IgG (H+L) Alexa Fluor® 594 conjugate (1:500, A-11072, Invitrogen, Carlsbad, CA, USA). Slides were mounted with Vectashield/DAPI medium (H-1200-10, Vector Laboratories, Burlingame, CA, USA) and were then analyzed under a Nikon Eclipse 80i (Nikon, Melville, NY, USA) fluorescent microscope with appropriate filters. Pictures were captured using a Hamamatsu C4742-95 CCD camera (Hamamatsu, Hamamatsu City, Japan).

To study tryptase, tissue sections were blocked with 5% mouse serum PBS1X for 1h30 and incubated with mouse anti-tryptase antibody (1:300, ab2378, ABCam, Cambridge, UK) for 1h at RT. Then Alexa Fluor Plus 594 goat anti-mouse secondary antibody (1:400, A32742, Molecular Probes, Eugene, OR, USA) was incubated for 1h at RT. Slides were mounted using the Vectashield mounting medium plus DAPI (H-1200-10, Vector Laboratories, Burlingame, CA, USA). Images were acquired on a Zeiss Cell Observer microscope (Carl Zeiss, Oberkochen, Germany). Percentage of mast cells was assessed by counting the number of mast cells in 10 same size randomly localized regions of interest (ROI) in each tumor, divided by the total number of cells in these ROI determined by the number of nuclei count with ImageJ 1.51u (NIH, USA).

Statistical analysis

Kaplan-Meier analysis was performed for metastasis-free survival and overall survival. To subdivide *ATRX* expression in two groups, expression was plotted for *ATRX* WT and altered cases, separately. The intersection between these two density curves was 4.45 (log₂ FPKM) and 2.77 for cohort 1 and 2, respectively. Differential gene expression (DGE) analyses were performed by R package DESeq. Gene Ontology (GO) analysis was performed on these differentially expressed genes ($P < 0.05$ and fold-change > 2 or < -2), by R package Goseq. In parallel, significant genes with $P < 0.01$ were used to make a heatmap (R package ComplexHeatMap). Clustering order was used to perform an enrichment of four clinical characteristics: tumor site, sex, FNCLCC grade and LMS status (Darbo *et al*, 2020), with a Fisher test, and p-values were adjusted with the Benjamini and Hochberg method.

Cell culture

Human osteosarcoma cell line MG63 and mouse osteosarcoma cell line K7M2 (42) were cultured in DMEM (31966-021, Life Technologies, Carlsbad, CA, USA), human UPS cell line IB106 (43) was cultured in RPMI-1640 (524000-025, Life Technologies, Carlsbad, CA, USA). Both were supplied with 10% fetal calf serum at 37°C in a humidified chamber containing 5% CO₂.

ShRNA Knockdown of ATRX expression

shRNAs constructs targeting human or mouse *ATRX* were obtained from OriGene (Rockville, MD, USA). The 28 bp human sequence was 5'-CCTTCTAACTACCAGCAGTTGATATGAG -3' and the 29 bp mouse sequence was 5'-CATCAAGTAGATGGTGTTCAGTTTATGTG -3'. A shRNA 29-mer scrambled shRNA was used as a negative control and obtained from OriGene.

Production of lentiviruses

Lentiviruses were produced by co-transfection of pVSVg, psPAX2 and shRNA construct in HEK293T cells. Co-transfection was performed by adding these plasmids, chloroquine at 0.025 mM (C6628, Sigma, St Louis, MO, USA), CaCl₂ at 0.125 M (C5050, Sigma, St Louis, MO, USA) and HeBS 1X (51558, Sigma, St Louis, MO, USA), HEK293T cells were then incubated at 37°C in a humidified chamber containing 5% CO₂. After 6 hours, HEK293T cell medium was changed with RPMI-1640 (524000-025, Life Technologies, Rockville, MD, USA) containing 10% of fetal calf serum.

Lentiviral transduction

HEK293T cell culture medium was filtered with a 0.45 µm PES filter and was mixed at 1:1 ratio with K7M2, MG63 or IB106 culture medium previously seeded. Polybrene (8 µg/ml, H9268, Sigma, St Louis, MO, USA) was also added with the virus. Infected cells were selected with puromycin and cells were sorted by FACS (BDFACSAria, BD Biosciences, San Jose, CA, USA) thanks to their GFP expression when vector with shRNA *ATRX* was integrated.

Western Blot analysis

Protein extracts were separated from each cell line with RIPA protein lysis buffer (R0278, Sigma, St Louis, MO, USA) containing 1X protease cocktail (P8340, Sigma, St Louis, MO, USA). Protein extracts were separated by electrophoresis on acrylamide gel (456-8085, Bio-Rad, Hercules, CA, USA) and transferred onto PVDF membrane. Then they were probed with antibodies against *ATRX* (1:1000, HPA001906, Sigma, St Louis, MO, USA) or actin (1:5000, A5316, Sigma, St Louis, MO, USA). Proteins of interest were detected with HRP-conjugated sheep anti-mouse IgG antibody (1:5000, NAX931V, GE

Healthcare, Piscataway, NJ, USA) or HRP-conjugated donkey anti-rabbit IgG antibody (1:5000, NA934V, GE Healthcare, Piscataway, NJ, USA) and visualized with the ECL prime Western blotting detection reagent (RPN2236, GE Healthcare, Piscataway, NJ, USA), according to the manufacturer's recommendations and using the PXi system (Syngene, Bangalore, India).

Cell proliferation assay

Cells of each cell line with ATRX^{KD} or ATRX^{CT} were seeded onto a 96-well plate (3.10^3 cells/well). After 4 days, cell proliferation was evaluated by adding 20 μ L of MTT (M2128, 5mg/mL, Sigma, St Louis, MO) to cell medium. Two hours later, medium was replaced by 100 μ L of DMSO (5879, Sigma, St Louis, MO, USA) and the optic density (OD) of each well was read with a spectrophotometer (PowerWaveX, Bio-Tel Instrument, Winooski, VT, USA) at 570 and 630 nm. Live cell number was correlated to $\Delta OD = OD_{570nm} - OD_{630nm}$. Experiments were performed independently in triplicate three times.

Soft agar assay

Cells of each cell line with ATRX^{KD} or ATRX^{CT} were seeded (5000 cells/well) in 0.35 % agarose cell medium (16500-500, Invitrogen, Carlsbad, CA, USA) onto a 6-well plate containing a 0.5 % agar base. 0.5 mL of cell culture medium was added and changed every 3-4 days. After incubating for 3 to 4 weeks, colonies were visualized with 0.005 % crystal violet staining (HT90132, Sigma, St Louis, MO, USA). Experiments were performed independently in triplicate four times.

In vivo experimentations

Mice were maintained under specific pathogen-free conditions in the animal facility of University of Bordeaux (France) or at the CREFRE (Centre Régional d'Exploration Fonctionnelle et Ressources Expérimentales, Toulouse, France). Experiments were performed in conformity with the rules of the Institutional Animal Care and Use committee (approval number DAP-APAFiS-2018041617309605) and all efforts were made to minimize animal suffering. For all experiments, 10^6 cells were injected into the dorsal flank of 6-8-week-old NSG (NOD/Shi-scid/IL-2R γ null) mice or Balb/c mice. Tumor sizes were measured twice a week using a caliper and their volume was calculated using the formula: $l^2 \times L/2$. At the end of the experiment, mice were sacrificed by cervical dislocation. Tumors were then weighed and divided in two parts for formalin fixation and nitrogen freezing. Each tumor was stained with HE and analyzed by a pathologist specialized in sarcomas. Growth rate was calculated with the segmental linear regression of GraphPad Prism (GraphPad Software, San Diego, CA, USA) and statistical analyses were done using an unpaired T-

test. Survival curves were analyzed with GraphPad Prism using the Kaplan-Meier method.

Mice tumor RNA sequencing and analysis

Total RNA was extracted, prepared and sequenced as described in supplementary methods to obtain more than 20 million paired-end reads with a length of 75 bp each. Bioinformatic analysis was done as previously described (17).

RNA reads were aligned to the mm10 genome assembly with STAR v2.6.0c (44). Low-quality (score < 20) and duplicated PCR paired-end reads were removed with SAMtools v1.8 (45) and PicardTools v2.18.2

(<http://broadinstitute.github.io/picard/index.html>), respectively. Then, gene expression was quantified with Cufflinks v2.2.1 (46), using RefSeq (47) genes (without miRNA and rRNA) from mm10 UCSC Table Browser (48) fixed on 2019/01.

Differential gene expression was performed by R package DESeq, between ATRX^{KD} and ATRX^{CT} tumors extracted from Balb/c mice. Relationships between proteins overexpressed in ATRX^{KD} and ATRX^{CT} tumors were assessed by the STRING Database (18).

Expression data are available on Gene Expression Omnibus (GEO) under accession GSE157953.

ALT specific c-circle detection

The C-circle assay, which partially detects single-stranded telomeric (CCCTAA)_n DNA circles (C-circles) amplified by the Phi29 polymerase in the absence of dCTP, was performed as previously described (49).

Disclosure of potential conflicts of interest:

The authors declare no potential conflicts of interest.

Authors' contributions:

Conception and design: L. Darmusey, G. Pérot, N. Thébault, F. Chibon

Development of methodology: L. Darmusey, G. Pérot, N. Thébault, F. Chibon

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): L. Darmusey, G. Pérot, N. Thébault, S. Le Guellec, N. Desplat, L. Gaston, L. Delespaul, T. Lesluyes, E. Darbo, A. Gomez-Brouchet, E. Richard, J. Baud, L. Leroy

Analysis and interpretation of data (e.g., statistical analysis, bio-statistics, computational analysis): L. Darmusey, G. Pérot, N. Thébault, F. Chibon

Writing, review, and/or revision of the manuscript: L. Darmusey, G. Pérot, N. Thébault, F. Chibon

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J.M. Coindre, J.Y. Blay

Study supervision: F. Chibon

Other (data, expertise, and guidance): The French Sarcoma Group and the International Cancer Genome Consortium

Acknowledgments:

The authors would like to thank the Centre Nacional d'Anàlisi Genòmica (CNAG, Barcelona, Spain) for WG and RNA sequencing services. We acknowledge the personnel of CREFRE US006 and Animal Facility A2 for their technical assistance. We thank Mickaël Michaud for his technical and theoretical advice and Michel Charbonneau and Nathalie Grandin for their work on ALT detection in cell lines. We acknowledge Françoise Redini for providing us MG63 and K7M2 cell lines.

Financial support:

The Instituts Thematiques Multiorganismes (ITMO) Cancer and the International Cancer Genome Consortium supported this work.

References

- Blay J-Y, Soibinet P, Penel N, Bompas E, Duffaud F, Stoeckle E, et al. Improved survival using specialized multidisciplinary board in sarcoma patients. *Ann Oncol*. 2017;28:2852–9.
- Serrano C, George S. Leiomyosarcoma. *Hematol Oncol Clin North Am*. 2013;27:957–74.
- Nicolas MM, Tamboli P, Gomez JA, Czerniak BA. Pleomorphic and dedifferentiated leiomyosarcoma: clinicopathologic and immunohistochemical study of 41 cases. *Hum Pathol*. 2010;41:663–71.
- Oda Y, Miyajima K, Kawaguchi K, Tamiya S, Oshiro Y, Hachitanda Y, et al. Pleomorphic leiomyosarcoma: clinicopathologic and immunohistochemical study with special emphasis on its distinction from ordinary leiomyosarcoma and malignant fibrous histiocytoma. *Am J Surg Pathol*. 2001;25:1030–8.
- The Cancer Genome Atlas Research Network, Abeshouse A, Adebamowo C, Adebamowo SN, Akbani R, Akredolu T, et al. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell*. 2017;171:950-965.e28.
- Isfort RJ, Cody DB, Lovell GJ, Gioeli D, Weissman BE, Doersen C-J. Analysis of oncogene, tumor suppressor gene, and chromosomal alterations in HeLa × osteosarcoma somatic cell hybrids. *Mol Carcinog*. 1999;25:30–41.
- Calo E, Quintero-Estades JA, Danielian PS, Nedelcu S, Berman SD, Lees JA. Rb regulates fate choice and lineage commitment in vivo. *Nature*. 2010;466:1110–4.
- Pérot G, Chibon F, Montero A, Lagarde P, de Thé H, Terrier P, et al. Constant p53 Pathway Inactivation in a Large Series of Soft Tissue Sarcomas with Complex Genetics. *Am J Pathol*. 2010;177:2080–90.
- Rubio R, García-Castro J, Gutiérrez-Aranda I, Paramio J, Santos M, Catalina P, et al. Deficiency in p53 but not Retinoblastoma Induces the Transformation of Mesenchymal Stem Cells In vitro and Initiates Leiomyosarcoma In vivo. *Cancer Res*. 2010;70:4185–94.
- Gibbons RJ, Picketts DJ, Villard L, Higgs DR. Mutations in a putative global transcriptional regulator cause X-linked mental retardation with α -thalassemia (ATR-X syndrome). *Cell*. 1995;80:837–45.
- Heaphy CM, Wilde RF de, Jiao Y, Klein AP, Edil BH, Shi C, et al. Altered Telomeres in Tumors with ATRX and DAXX Mutations. *Science*. 2011;333:425–425.
- Marzec P, Armenise C, Pérot G, Roumelioti F-M, Basyuk E, Gagos S, et al. Nuclear-Receptor-Mediated Telomere Insertion Leads to Genome Instability in ALT Cancers. *Cell*. 2015;160:913–27.
- Liau J-Y, Tsai J-H, Jeng Y-M, Lee J-C, Hsu H-H, Yang C-Y. Leiomyosarcoma With Alternative Lengthening of Telomeres Is Associated With Aggressive Histologic Features, Loss of ATRX Expression, and Poor Clinical Outcome: *Am J Surg Pathol*. 2015;39:236–44.
- Kovatcheva M, Liu DD, Dickson MA, Klein ME, O'Connor R, Wilder FO, et al. MDM2 turnover and expression of ATRX determine the choice between quiescence and senescence in response to CDK4 inhibition. *Oncotarget*. 2015;6:8226–43.
- Han M, Napier CE, Frölich S, Teber E, Wong T, Noble JR, et al. Synthetic lethality of cytolytic HSV-1 in cancer cells with ATRX and PML deficiency. *J Cell Sci*. 2019;132.
- Clynes D, Jelinska C, Kella B, Ayyub H, Scott C, Mitson M, et al. Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX. *Nat Commun [Internet]*. 2015 [cited 2016 Jul 26];6. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4501375/>
- Lesluyes T, Pérot G, Largeau MR, Brulard C, Lagarde P, Dapremont V, et al. RNA sequencing validation of the Complexity INDEX in SARCOMAS prognostic signature. *Eur J Cancer Oxf Engl* 1990. 2016;57:104–11.
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47:D607–13.
- Chudasama P, Mughal SS, Sanders MA, Hübschmann D, Chung I, Deeg KI, et al. Integrative genomic and transcriptomic analysis of leiomyosarcoma. *Nat Commun [Internet]*. 2018 [cited 2019 Sep 12];9. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5762758/>
- Lee PJ, Yoo NS, Hagemann IS, Pfeifer JD, Cottrell CE, Abel HJ, et al. Spectrum of mutations in leiomyosarcomas identified by clinical targeted next-generation sequencing. *Exp Mol Pathol*. Academic Press; 2017;102:156–61.
- Watson LA, Goldberg H, Bérubé NG. Emerging roles of ATRX in cancer. *Epigenomics*. 2015;7:1365–78.
- Ren X, Tu C, Tang Z, Ma R, Li Z. Alternative lengthening of telomeres phenotype and loss of ATRX expression in sarcomas. *Oncol Lett*. 2018;15:7489–96.
- Yang C-Y, Liau J-Y, Huang W-J, Chang Y-T, Chang M-C, Lee J-C, et al. Targeted next-generation

- sequencing of cancer genes identified frequent TP53 and ATRX mutations in leiomyosarcoma. *Am J Transl Res.* 2015;7:2072–81.
24. Chen X, Bahrami A, Pappo A, Easton J, Dalton J, Hedlund E, et al. Recurrent Somatic Structural Variations Contribute to Tumorigenesis in Pediatric Osteosarcoma. *Cell Rep. Cell Press;* 2014;7:104–12.
 25. Ahvenainen TV, Mäkinen NM, von Nandelstadh P, Vahteristo MEA, Pasanen AM, Bützow RC, et al. Loss of ATRX/DAXX expression and alternative lengthening of telomeres in uterine leiomyomas. *Cancer.* 2018;124:4650–6.
 26. DeBerardinis RJ, Chandel NS. Fundamentals of cancer metabolism. *Sci Adv [Internet]. American Association for the Advancement of Science;* 2016 [cited 2020 May 26];2. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4928883/>
 27. Komi DEA, Redegeld FA. Role of Mast Cells in Shaping the Tumor Microenvironment. *Clin Rev Allergy Immunol [Internet].* 2019 [cited 2019 Sep 18]; Available from: <https://doi.org/10.1007/s12016-019-08753-w>
 28. Crawford G, Hayes MD, Seoane RC, Ward S, Dalessandri T, Lai C, et al. Epithelial damage and tissue $\gamma\delta$ T cells promote a unique tumor-protective IgE response. *Nat Immunol. Europe PMC Funders;* 2018;19:859.
 29. Oldford SA, Haidl ID, Howatt MA, Leiva CA, Johnston B, Marshall JS. A Critical Role for Mast Cells and Mast Cell-Derived IL-6 in TLR2-Mediated Inhibition of Tumor Growth. *J Immunol. American Association of Immunologists;* 2010;185:7067–76.
 30. Koelsche C, Renner M, Johann P, Leiss I, Sahn F, Schimmack S, et al. Differential nuclear ATRX expression in sarcomas. *Histopathology.* 2016;68:738–45.
 31. Mitson M, Kelley LA, Sternberg MJE, Higgs DR, Gibbons RJ. Functional significance of mutations in the Snf2 domain of ATRX. *Hum Mol Genet. Oxford Academic;* 2011;20:2603–10.
 32. Ayodele O, Abdul Razak AR. Immunotherapy in soft-tissue sarcoma. *Curr Oncol [Internet].* 2020 [cited 2020 Jun 18];26. Available from: <https://current-oncology.com/index.php/oncology/article/view/5407>
 33. D'Angelo SP, Mahoney MR, Van Tine BA, Atkins J, Milhem MM, Jahagirdar BN, et al. Nivolumab with or without ipilimumab treatment for metastatic sarcoma (Alliance A091401): two open-label, non-comparative, randomised, phase 2 trials. *Lancet Oncol.* 2018;19:416–26.
 34. Wilky BA, Trucco MM, Subhawong TK, Florou V, Park W, Kwon D, et al. Axitinib plus pembrolizumab in patients with advanced sarcomas including alveolar soft-part sarcoma: a single-centre, single-arm, phase 2 trial. *Lancet Oncol.* 2019;20:837–48.
 35. Robbins PF, Morgan RA, Feldman SA, Yang JC, Sherry RM, Dudley ME, et al. Tumor Regression in Patients With Metastatic Synovial Cell Sarcoma and Melanoma Using Genetically Engineered Lymphocytes Reactive With NY-ESO-1. *J Clin Oncol. American Society of Clinical Oncology;* 2011;29:917–24.
 36. Dillman R, Barth N, Selvan S, Beutel L, de Leon C, DePriest C, et al. Phase I/II Trial of Autologous Tumor Cell Line-Derived Vaccines for Recurrent or Metastatic Sarcomas. *Cancer Biother Radiopharm. Mary Ann Liebert, Inc., publishers;* 2004;19:581–8.
 37. Finkelstein SE, Iclozan C, Bui MM, Cotter MJ, Ramakrishnan R, Ahmed J, et al. Combination of External Beam Radiotherapy (EBRT) With Intratumoral Injection of Dendritic Cells as Neo-Adjuvant Treatment of High-Risk Soft Tissue Sarcoma Patients. *Int J Radiat Oncol.* 2012;82:924–32.
 38. Oldford SA, Marshall JS. Mast cells as targets for immunotherapy of solid tumors. *Mol Immunol.* 2015;63:113–24.
 39. Singer J, Jensen-Jarolim E. IgE-based immunotherapy of cancer: challenges and chances. *Allergy. John Wiley & Sons, Ltd;* 2014;69:137–49.
 40. Teo PZ, Utz PJ, Mollick JA. Using the allergic immune system to target cancer: activity of IgE antibodies specific for human CD20 and MUC1. *Cancer Immunol Immunother.* 2012;61:2295–309.
 41. Fletcher C, Bridge JA, Hogendoorn P, Mertens F. WHO Classification of Tumours of Soft Tissue and Bone. 4th ed. Lyon: IARC Press; 2013.
 42. Khanna C, Prehn J, Yeung C, Caylor J, Tsokos M, Helman L. An orthotopic model of murine osteosarcoma with clonally related variants differing in pulmonary metastatic potential. *Clin Exp Metastasis.* 2000;18:261–71.
 43. Lagarde P, Brulard C, Pérot G, Mauduit O, Delespaul L, Neuville A, 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;2:1014.
 44. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15–21.
 45. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078–9.
 46. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Baren MJ van, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010;28:511–5.
 47. Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* 2012;40:D130–5.
 48. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 2004;32:D493–6.
 49. Fogli A, Demattei M-V, Corset L, Vauris-Barrière C, Chautard E, Biau J, et al. Detection of the alternative lengthening of telomeres pathway in malignant gliomas for improved molecular diagnosis. *J Neurooncol.* 2017;135:381–90.

Supplemental Methods

DNA extraction

For both cohorts, genomic DNA from frozen samples was isolated using standard phenol-chloroform extraction protocol. DNA was quantified using Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Blood material was also available for cohort 1. Genomic DNA from blood samples was extracted using customized automated purification of DNA from compromised blood samples on the Autopure LS device according to the manufacturer's recommendations (9001340, Qiagen, Hilden, Germany) with increased centrifugation of 10 min for DNA precipitation and DNA wash.

Whole genome sequencing and analysis

Whole genome sequencing was performed only on cohort 1. To construct short-insert paired-end libraries, a no-PCR protocol was used with the TruSeq™DNA Sample Preparation Kit v2 (FC-121-2001/FC-121-2002, Illumina Inc., San Diego, CA, USA) and the KAPA Library Preparation Kit (KK8235, Kapa Biosystems, Basel, Switzerland). Briefly, 2 µg of genomic DNA were sheared on a Covaris™ E220, size-selected and concentrated using AMPure XP beads (A63880, Agencourt, Beckman Coulter, Brea, CA, USA) in order to reach a fragment size of 220 – 480 bp. Fragmented DNA was end-repaired, adenylated and ligated to Illumina-specific indexed paired-end adapters.

DNA sequencing was performed in paired-end mode, 2x100 bp or 2x125 bp according to flowcell version, in five or three sequencing lanes of HiSeq2000 flowcell v3 or v4 (Illumina Inc., San Diego, CA, USA) to analyze tumor or normal/constitutive samples and to reach minimal yield of 145 or 85 Gb, respectively. Two tumor samples were sequenced in 20 lanes to reach a minimal yield of 560 Gb. Image analysis, base calling and quality scoring of the run were processed using the manufacturer's Real Time Analysis software (RTA 1.13.48) and followed by generation of FASTQ sequence files by CASAVA (Illumina Inc., San Diego, CA, USA).

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 (<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 (1), 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 (2) and PicardTools v1.118

(<http://broadinstitute.github.io/picard/index.html>), respectively. SNV were detected by SAMtools mpileup v0.1.19 (2), with a minimum of 20 as phred quality score (-Q 20), and by bcftools call -Am (SAMtools v0.2.0, <http://samtools.github.io/bcftools/call-m.pdf>).

RNA extraction, sequencing and analysis from frozen samples

For both cohorts, RNA extraction was performed using standard TRIzol (15596026, Thermo Fisher, Waltham, MA, USA) / chloroform 32211-1L, Fisher Scientific, Hampton, NH, USA) extraction followed by 100% ethanol precipitation and RNA purification using the RNeasy Mini Kit (74104, Qiagen, Hilden, Germany) with DNase treatment (79254, RNase-Free DNase Set, Qiagen, Hilden, Germany). Total RNA was quantified using a Nanodrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA) and qualified with an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) using the Agilent RNA 6000 Nano Kit (5067-1511, Agilent, Santa Clara, CA, USA) according to the manufacturer's instructions.

To control the sequencing quality, ERCC RNA Spike-In Mix (4456740, Life technologies, Carlsbad, CA, USA) were added to each RNA sample as recommended by the manufacturer. Analysis of this quality control was performed as previously described (3).

Regarding cohort 1, libraries from total RNA were prepared using the TruSeq@Stranded Total RNA Gold Library Preparation Kit (RS-122-2301, Illumina Inc., San Diego, CA, USA) according to the manufacturer's protocol. Briefly, 0.5 µg of total RNA was ribo-depleted using the Ribo-Zero Gold Kit. RNA fragmentation resulted in fragments of 80 – 450 nt, with a major peak at 160 nt. First-strand cDNA synthesis by random hexamers and reverse transcriptase was followed by second-strand cDNA synthesis, performed in presence of dUTP instead of dTTP. Blunt-ended double-stranded cDNA was 3'adenylated and Illumina-indexed adapters were ligated. Resulting libraries were enriched with 15 PCR cycles.

Construction of cohort 2 libraries was the same as previously described for frozen samples (3).

Libraries, of these two cohorts were sequenced on HiSeq2000 (Illumina Inc., San Diego, CA, USA) in paired-end mode with a read length of 2x75 bp using TruSeq SBS Kit v3-HS FC-401-3001, Illumina Inc., San Diego, CA, USA). Image analysis, base calling and base quality scoring of the run were processed by integrated primary Real Time Analysis (RTA 1.13.48) software and followed by generation of FASTQ

sequence files by CASAVA (v1.8, Illumina Inc., San Diego, CA, USA). Library construction and RNA sequencing were performed at the Centro Nacional de Análisis Genómico (CNAG, Barcelona, Spain).

RNA bioinformatic analysis (alignment and expression quantification) for these two cohorts was performed as previously described (3). Fusion transcripts were detected with Defuse v0.6.1 (4) as previously described (5).

Regarding cohort 1, SNV (Single Nucleotide Variant) were detected using samtools mpileup (SAMtools v0.1.19 (2)), with a minimum of 20 as phred quality score (-Q 20), and bcftools view -cvg (SAMtools v0.1.19 (6)). Regarding cohort 2, SNV were detected using samtools mpileup (2), with a minimum of 20 as phred quality score (-Q 20), and bcftools view -cgN (6). Detected variants with fewer than 5 coverage reads were filtered out.

Annotation of variants

Regarding cohort 1, variants detected in constitutional, 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.

Since tumor and constitutional DNA were not available for cohort 2, candidate variants could not be filtered based on genotype differences between constitutional DNA and DNA/RNA tumor. Instead, filters applied were the same as for cohort 1 above, but with a minimum coverage of 5 reads due to a lower sequencing depth.

Variants were annotated using the Annovar v20160314 tool (7). Variants were selected whose alternative allele frequency (AF) in the Caucasian population (CEU) is lower than 0.1%, as reported in the 1000Genome database. Finally, variants were kept if they were localized in coding regions and were non-synonymous.

Break Point Detection

For all the 67 LMS (cohort 1), the whole ATRX sequence obtained by WG sequencing was visualized using the Integrative Genomics Viewer (IGV, v2.6.3) (8) and soft-clipped reads were detected for 5 cases.

ATRX fusion partner sequences and breakpoints location could be determined thanks to the blat function on the UCSC website.

Break point detection was performed as described above, and for alignment with Bowtie v2.2.1.0 (1), local and very-sensitive-local options were used. ...

Supplemental References

1. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
2. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
3. Lesluyes T, Pérot G, Largeau MR, Brulard C, Lagarde P, Dapremont V, et al. RNA sequencing validation of the Complexity INDEX in SARComas prognostic signature. *Eur J Cancer Oxf Engl* 1990. 2016;57:104–11.
4. McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MGF, et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput Biol*. 2011;7:e1001138.
5. Delespaul L, Lesluyes T, Pérot G, Brulard C, Lartigue L, Baud J, et al. Recurrent TRIO Fusion in Nontranslocation-Related Sarcomas. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2017;23:857–67.
6. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27:2987–93.
7. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.
8. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol*. Nature Publishing Group; 2011;29:24–6.

Cohort	Sample	Histotype	Location	Sex	Allele 1	Allele 2	DNA (WGSseq)		RNA (RNAseq)		ATRX: log2 (FPKM+1)	Functional ATRX status	gDNA (Ilg19)	cDNA (NM_000489.6)	Protein
							Reads alt/total	Mut. allele freq.	Reads alt/total	Mut. exp. freq.					
1	LMS16T	LMS	Member	M	Frameshift	Y	812	0.67	25.54	0.74	3.61	0	c.5441_5445del	p.(C181456A>A5)	
	LMS21T	LMS	Member	M	Frameshift	Y	1117	0.65	18.36	0.59	4.10	0	c.5212_5232delinsTTTA	p.(E1738F>X1)	
	LMS35T	LMS	Member	F	Frameshift	Y	2159	0.56	1946	0.41	4.06	1/2	c.8070pC>I;A85_487del	p.(E1738F>X1);p.(E162del)	
	LMS33T	LMS	Internal Trunk	F	Frameshift	Y	3261	0.52	1091.31	0.83	4.07	0	c.2014_2055del	p.(E1738F>X1);p.(E162del)	
	LMS53T	LMS	Internal Trunk	F	Frameshift	Y	2852	0.54	39.56	0.70	3.99	0	chrX:g.769399393dupG	p.(E1738F>X1);p.(E162del)	
	LMS59T	LMS	Uterus	F	Frameshift	Y	1252	0.23	34.37	0.92	3.09	0	chrX:g.76938581_76938618dup	p.(E1738F>X1);p.(E162del)	
	LMS50T	LMS	Uterus	F	Frameshift	Y	3054	0.56	4.5	0.80	3.84	0	chrX:g.76849230delG	p.(E1738F>X1);p.(E162del)	
	LMS15T	LMS	Internal Trunk	M	Nonense	Y	2227	0.81	18.25	0.72	3.86	0	chrX:g.76578786C>A	p.(E2265X)	
	LMS38T	LMS	Uterus	F	Nonense	Y	1747	0.56	44.48	0.92	4.42	0	chrX:g.7688956C>A	p.(E1652X)	
	LMS30T	LMS	Uterus	F	Nonense	Y	2044	0.55	34.38	0.89	4.14	0	chrX:g.76938683G>C	p.(E1722X)	
	LMS55T	LMS	Uterus	F	Nonense	Y	1844	0.41	18.18	1.00	4.07	0	chrX:g.76907651G>A	p.(R1504X)	
	LMS48T	LMS	Member	F	Miscense	Y	3267	0.48	0.66	1.00	5.23	1	chrX:g.76849167C>T	p.(R1504X)	
	LMS49T	LMS	Internal Trunk	F	Miscense	Loss	4145	0.91	118.125	0.94	5.51	0	chrX:g.76879980C>T	p.(D1719N)	
	LMS69T	LMS	Internal Trunk	F	Miscense	Loss	2328	0.82	59.61	0.87	4.49	0	chrX:g.76889104C>T	p.(G1597D)	
	LMS71	LMS	Internal Trunk	F	Miscense	Y	1851	0.55	95.111	0.86	5.19	0	chrX:g.76893113T>C	p.(H1594R)	
	LMS57T	LMS	Uterus	F	Del. & FS	Y	NA	NA	39.49	0.89	4.56	0	chrX:g.76882435_76882612del	p.(H1594R)	
	LMS66T	LMS	Internal Trunk	F	Del. & FS	Y	NA	NA	45.48	0.94	3.85	0	chrX:g.76756028_76920646delinsATTCCA	p.(S1761X3)	
	LMS1T	LMS	Internal Trunk	F	Del. Inv. Ins.	Y	NA	NA	NA	NA	2.55	0	c.3810_7479delins3809+1_3809+187	p.(I1271Y>X62)	
	LMS43T	LMS	Internal Trunk	F	Del.	Y	NA	NA	NA	NA	1.91	0	NA	NA	
	LMS62T	LMS	Internal Trunk	F	Trans. Complex	Y	NA	NA	NA	NA	2.22	0	Complex	NA	
2	S873	MFS	Member	M	Frameshift	Y	NA	NA	0.4*	Sanger: B	2.39	0	c.5432_5442del	p.(L1811C>A>A6)	
	N03	DDLPS	Internal Trunk	M	Nonense	Y	NA	NA	0.3*	Sanger: AB	0.86	0	c.2185C>G	p.(S6772X)	
	S859	UPS	Member	M	Nonense	Y	NA	NA	8.8	1.00	2.64	0	c.5269G>T	p.(E1757X)	
	S825	UPS	Trunk wall	M	Miscense	Y	NA	NA	21.21	1.00	3.24	0	c.2650G>A	p.(E884K)	
	S928	UPS	Member	F	Miscense	Y	NA	NA	8.8	1.00	2.67	0	c.5197A>G	p.(K1734E)	
	S832	MFS	Member	F	Miscense	Y	NA	NA	6.6	1.00	2.90	0	c.2177A>T	p.(D1766Z)	
	S823	MFS	Member	M	Fusion transcript	Y	NA	NA	NA	NA	0.98	0	NA	p.(S1706S62)	
	S831	UPS	Trunk wall	M	Fusion transcript	Y	NA	NA	NA	NA	2.47	0	NA	p.(I2234N>S);p.(I2234R>A);p.(S4)	
S916	UPS	Member	F	Fusion transcript	Y	NA	NA	NA	NA	2.16	0	NA	p.(V1520D>S);p.(I321V>X33)		
M979	MFS	Member	F	Fusion transcript	Y	NA	NA	NA	NA	2.04	0	NA	p.(R1209S>X12)		

Supplementary Table 1: ATRX mutations and structural variants in both cohorts.

For each of the 30 validated alterations of ATRX (NM_000489.6), genomic and expression information and annotations are reported in the table. Regarding cohort 2, only gDNA Sanger sequencing data for point mutations are available. *: not detected by RNAseq but detected by Sanger sequencing on cDNA. Mutation expression frequency for these two mutations could not be determined using RNAseq, probably because of tumor heterogeneity, so allele expression in RNA was defined according to Sanger sequencing data. Alt: alternative, Freq.: frequency, Exp.: expression, Mut.: mutation, Loc.: location, M: Male, F: Female, LMS: leiomyosarcoma, DDLPS: dedifferentiated liposarcoma, MFS: myxofibrosarcoma, UPS: undifferentiated pleomorphic sarcoma, Y: chromosome Y, A: reference allele, B: alternative allele, Del: deletion, Ins: insertion, Inv: inversion, FS: Frameshift, SV: structural variant, Trans.: translocation, FPKM: Fragments Per Kilobase Million, NA: not available.

Sample	Histotype	Sex	Gene_chromosome1	Break location	Break position (hg19)	Gene_chromosome2	Gene name 2	Break location 2	Break position 2 (hg19)	Mechanism	Consequence on ATRX gene	Genomic nomenclature (hg19) chrX	cDNA nomenclature (NM_000489.6)	Protein nomenclature	Annotations	
															Fusion partner	Annotations
LMS1T	LMS	F	X	Intron 2	76961352	X	Intergenic	X	77042825	Deletion/inversion-insertion	Loss of 3' part to exon 2	g.76961353_77042808delins:77042774_77042825inv	-	-		
LMS43T	LMS	F	X	Intron 1	77000958	X	Intergenic	X	77060889	Deletion	Loss of 3' part to exon 1	g.77000959_77060888del	-	-		
LMS57T	LMS	F	X	Intron 19	76882434	X	ATRX	Intron 1	76982613	Deletion	Fusion of intron 1 to intron 19	g.76882435_76982612del	c.21_4956del	p.(S786_X3)		
LMS62T	LMS	F	X	Intron 1	77035789	X	DMD	Intron 60	31367514	Complex	Loss of 3' part to exon 1	Complex	-	-		
LMS66T	LMS	F	X	Intron 10	76920646	X	Intergenic	X	76756027	Deletion	Loss of exon 1 to the end	g.76756028_76920645delins:ATTCCA	c.3810_7479delins:3809+_1_3809+187	p.(L1271YfsX62)		

Supplementary Table 2: ATRX breakpoints in cohort 1.

Chromosomal and genomic coordinates of gene breakpoint regarding ATRX and its fusion partners are indicated together with the potential mechanism involved and consequence for the ATRX gene. Nomenclatures of the alteration and its effect on RNA are indicated when an alternative RNA product was detected by RNAseq. Predicted effect on protein is then indicated. M: Male, F: Female, LMS: leiomyosarcoma.

Sample	Histotype	Sex	Gene_chromosome1	Break location	Break position (hg19)	Gene_chromosome2	Gene name 2	Break location 2	Break position 2 (hg19)	Mechanism	ORF conservation	ATRX Protein	Annotations	
													Fusion partner	Annotations
S823	MFS	M	X	<i>ATRX</i> exon 1	77041470	15	<i>DDX11L9</i>	exon 1	102519162	Translocation	No		p.(S7RfsX62)	
S831	UPS	M	X	exon 30	76812922	X	<i>RNU6-974P</i>	downstream	81166026	Eversion	No		p.(D2234NfsX32;D2234RfsX4)	
S918	UPS	F	X	exon 15	76891547	X	<i>RPL279N11.1</i>	upstream	75878708	Deletion	No		p.(V1520DfsX54;I1521VfsX33)	
M979	MFS	F	X	exon 28	76829715	X	<i>FTX</i>	intron	73401759	Inversion	No		p.(R2109SfsX12)	

Supplementary Table 3: ATRX fusion transcripts in cohort 2.

Chromosomal and genomic coordinates of gene breakpoint regarding ATRX and its fusion partners are indicated together with the potential mechanism involved and predicted effect on the protein. MFS: myxofibrosarcoma, UPS: undifferentiated pleomorphic sarcoma, M: Male, F: Female.

	Cohort1 (n=67)	Cohort2 (n=60)
FOLLOW-UP and OUTCOMES		
Follow-up (years)		
Median	5.22	5.94
Range	1.48-19.77	0.01-28.88
PATIENT CHARACTERISTICS		
Age at diagnosis (years)		
Mean	62.94	62.95
Median	64	64.5
Range	22-80	20-87
Gender		
Female	53 (78.46%)	22 (36.67%)
Male	14 (21.54%)	38 (63.33%)
TUMOR CHARACTERISTICS		
Tumor site		
Internal trunk	38	9
Uterine	8	0
Member and Trunk wall	21	51
Tumor depth		
Deep	56	42
Superficial	7	5
Superficial and deep	4	13
Tumor size (cm)		
Median	8	8
Range	1.5-23	1-30
Histotype		
LMS	67	0
UPS	0	30
MFS	0	17
DDLPS	0	13
FNCLCC Grade		
I	12	4
II	23	18
III	32	40
Resection status, margins		
R0	42	24
R1	18	28
R2	1	2
Unknown	6	6

Supplementary Table 4: Clinical characteristics in both cohorts

LMS: leiomyosarcoma, UPS: undifferentiated pleomorphic sarcoma, MFS: myxofibrosarcoma, DDLPS: dedifferentiated liposarcoma.

Primer name	Primer sequence 5'-->3'	Location	Product size (bp)	PCR program	Ref gene	Tumor
ATRXgDNAe87F ATRXgDNAe87R	ACTGTGTGCAATATGCCATTT AGAAAGTCTTCCAAGGGCAGA	Exon 7	366	TD 60°C	NM_000489.6	LMS35T
ATRXgDNAe89F ATRXgDNAe89R	CGGTAATTCTCTGACAGTGCT AGAAGACTCAGACTGGGTTTGT	Exon 9	346	TD 60°C		N003/LMS35T
ATRXgDNAe17F ATRXgDNAe17R	TGCTGTTCTTAGAAGTTTGGTT CATTAGAACCTTCTGCTCAAACA	Exon 17	485	TD 60°C		M987
ATRXgDNAe20F ATRXgDNAe20R	CAACGATGTCATTTATCTCTCG ACCACTGATTATAAAGCATCTCA	Exon 20	452	TD 60°C		S928/S859/LMS22T
ATRXgDNAe21F ATRXgDNAe21R	TGAGCATTTCAATGGGGAAT CCTCAGAAAATATGTTGGATTG	Exon 21	326	TD 60°C		S873/LMS16T
ATRXgDNAe26F ATRXgDNAe26R	CTCCCAAGTCCCAATCAAGTT AGGAAAGGAAAGGAAAGCAACA	Exon 26	377	TD 60°C		LMS48T/LMS60T
ATRXBPLMS1F1 ATRXBPLMS1R1 ATRXBPLMS1F2 ATRXBPLMS1R2	AACCTGGCAATCAAGTCTGTGC TCCA A AATTCATATGGACCAAGC GAGGTCGAGGTTGGAAGGATG CACCTCAGCCTCTCAAGCAG	ATRX intron 2 Upstream ATRX gene	268 153	TD 60°C TD 60°C		LMS1T
ATRXBPLMS43F1 ATRXBPLMS43R1 ATRXBPLMS43F2 ATRXBPLMS43R2	CGTCCTAGCCTCTGTTAAACC TGGAAGTCTATTCAGCCCTT AATGTCCTTCTGTGCTGGC GGCAGAAAGGATCCCTTGAAC	ATRX intron 1 Intergenic chrX	206 182	TD 60°C TD 60°C		LMS43T
ATRXBPLMS57F1 ATRXBPLMS57R1 ATRXBPLMS57F2 ATRXBPLMS57R2	TTCA CCGTGTGAACCAAGAT ACCACTAAATGTGCAATAACA TCCA CACTCTCCAGCATC CCCTATCTGCAATTGGTCA	ATRX intron 19 ATRX intron 1	390 476	TD 60°C TD 60°C		LMS57T
ATRXBPLMS62F1 ATRXBPLMS62R1 ATRXBPLMS62F2 ATRXBPLMS62R2	AGCCTTGCCTGTACTTCTTTG CATGCCTGTAGTCCCAAGCT GGCGACA GGGTGAAGTGC AGCAACA A A A C C T G T A A C C T	DMD intron 60 ATRX intron 1	151 206	TD 60°C TD 60°C		LMS62T
ATRXBPLMS66F1 ATRXBPLMS66R1 ATRXBPLMS66F2 ATRXBPLMS66R2	GGGAATTGAACAATGAGAACACA GTTGCCCACTACTGAC TGCTGTGAATAAACTCTCGT TTTGATGCTGCATAA C C T T C C A	Intergenic chrX ATRX intron 10	118 150	TD 60°C TD 60°C		LMS66T

Supplementary Table 5: Genomic DNA primers used for point mutations and breakpoint validation by Sanger sequencing

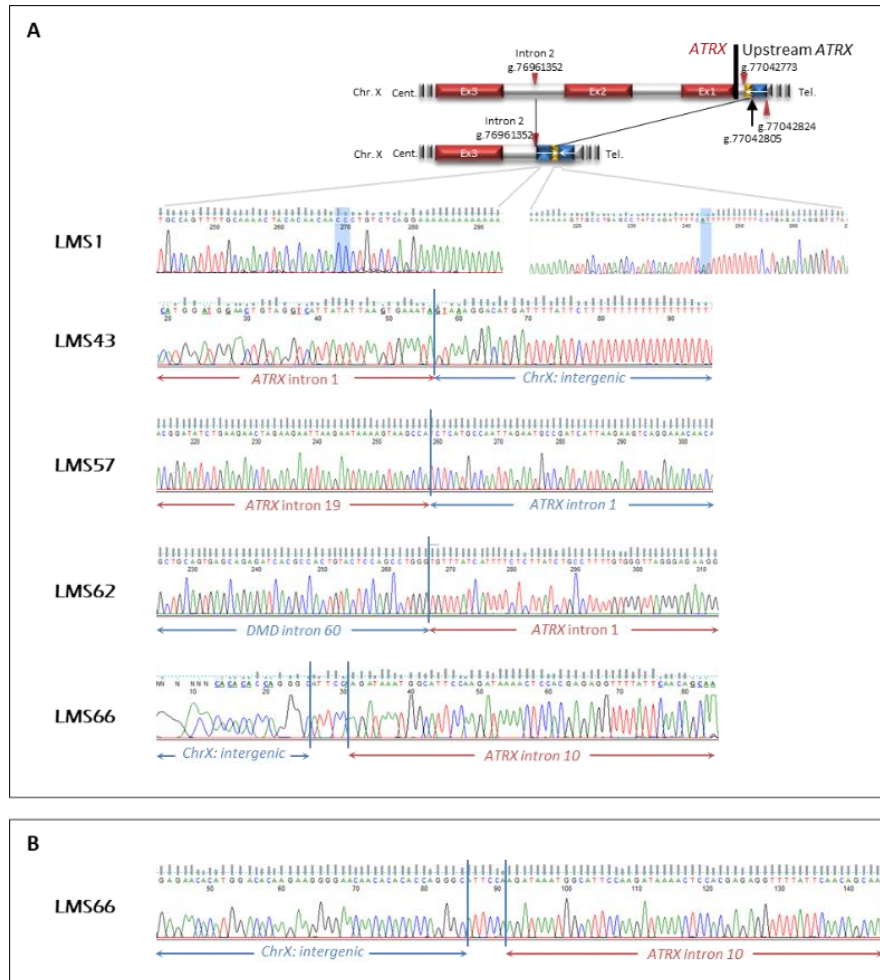
Forward and reverse primers used for mutation and breakpoint (BP) validations on genomic DNA are presented. All mutations were detected by two independent techniques and only those not detected by both Wgseq and RNAseq were validated by Sanger sequencing. RNA primers were used for DNA validation of some ATRX mutations in exon 9, see Supplementary Table 6. For each BP validation in cohort 1, the two primers used to detect the BP are in red: other primers were used to detect a potential reciprocal translocation. PCR was a touch-down 60°C program (TD 60°C) (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). RefSeq annotation used to locate primers for the gene is indicated. Tumors in which each primer couple was used are indicated. Ex: exon, bp: base pair, F: forward, R: reverse.

Primer name	Primer sequence 5'→3'	Product size (bp)	DNA validation	RNA validation
ATRXcDNA1	ACCGCTGACCCATGAGT	143		
ATRXcDNA1	CCACTGATTTTATCTGTGTTTGA			
ATRXcDNA2	TTCCCTGACACTCATCAGAA	399		
ATRXcDNA2	TTTCATTTTACTTCTCTCTAAATTC			
ATRXcDNA3	CAGAGCCAGTCTGAATGAA	627		LMS35T
ATRXcDNA3	CAGAGCCAGAACAGGAATCA			
ATRXcDNA4	CAGTGTGTGACAAAATAAGAA	495		
ATRXcDNA4	TTCCAAAGCACAAAGTTTTC			
ATRXcDNA5	TGGATGCTGTAACAAAGAGAAA	552		
ATRXcDNA5	CAGCACCTTAAATGGGGAAT			
ATRXcDNA6	GGAGGATTAATACTAAACTACAGC	579	LMS33T/LMS59T LMS63T/S832	LMS59T/LMS63T
ATRXcDNA6	CAAAAGTCTTATGGTTGTATGAATTT			
ATRXcDNA7	TGAAATGCTAGCAAATCCTCAA	773	S825	
ATRXcDNA7	CAGTTCCTTTTGTCTCTGC			
ATRXcDNA8	AAGTACAGGATGGCTTATCTGATATT	699		
ATRXcDNA8	GGGAGTTCTCTTTTCTCTCTTG			
ATRXcDNA9	CAAAAGTGGCTCATCATCTG	684		
ATRXcDNA9	TTCACTGCTCACTTCTCTG			
ATRXcDNA10	TGAGTGAAGGAAATCTGGA	696		
ATRXcDNA10	CAGACTCAGCAACAAATCC			
ATRXcDNA11	AAGATGCTTCAACCAAG	802		LMS22T
ATRXcDNA11	TCTGCACACTGCCATTTTGA			
ATRXcDNA12	CATTGATGGTAAATTTATCAAGGAA	761		LMS16T
ATRXcDNA12	TCAGCATCAGCATCTGFAACAA			
ATRXcDNA13	TTTAAACTGGAAAGAAAGTAAAGCTAC	741		LMS48T/LMS60T
ATRXcDNA13	TCCTCTTCTTCTTCTTCTGA			
ATRXcDNA14	TGGAAGCTCTTTTACTAATGAAT	559		
ATRXcDNA14	TCCTGGCTGGCTTGTCTACT			
ATRXcDNA15	ACAGTGTGACAGCATGACGA	545		
ATRXcDNA15	TTGAGTCTGTGTAAGTCACTGATTC			
ATRXcDNALMS33F	GTGAACTGGACAGGAAAACA	690		LMS33T
ATRXcDNALMS33R	TGCTGTGTTCTCATCTCAGAA			
ATRXcDNALMS35F	AACAATGTAGTGGTGTCCG	234		LMS35T
ATRXcDNALMS35R	TCTTATTTTCTGCAACAACCTGT			
ATRXBPcDNALMS66F1	TCATCTCTAGTTTGAAGCAAGG	233		LMS66T
ATRXBPcDNALMS66R2	CCTCAGCAACTAACACAGG			

Supplementary Table 6: Primers used for ATRX cDNA screening and validation of mutations by Sanger sequencing. ATRX forward and reverse primers used in cohort 2 for ATRX cDNA screening in both cohorts for some exon 9 ATRX DNA mutations and in cohort 1 for validation of cDNA mutations are presented. TD60°C PCR program described in Supplementary Table 5 legend was used. Primers used for cDNA validation of fusion transcript in LMS66 are also presented (primers in red). Tumors in which each primer couple was used are indicated. bp: base pair, F: forward, R: reverse.

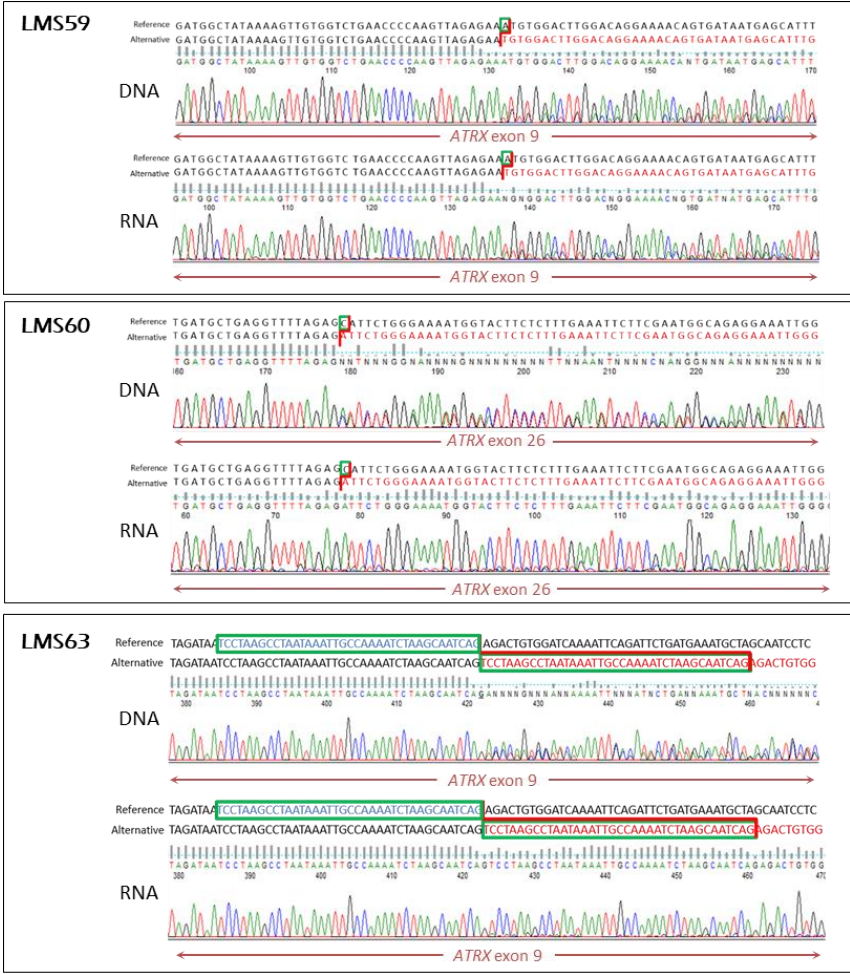
Primer name	Primer sequence 5'→3'	Tumor
ATRXex1F	CAGTGCATTTCTATCGTAAACCG	S823
DDX11L9ex1R	CTCTAAGCATGGCTCCTCTCTC	
ATRXex15F	TGAGAACAGAAACACAAAATGCT	S918
RP1-279N11.1 75878708F	CACA ACTGGCTCAACTGCTT	
ATRXex15F	TGAGAACAGAAACACAAAATGCT	
RP1-279N11.1 75879066F	TCATGTCAGCTTTTGTGACGA	
ATRXex27F	GTCCCTCATATCTCTGGACTTGA	M979
RP1-279N11.1 75913486F	TCTCTGCCAGGCTTCTTGA	
ATRXex27F	GTCCCTCATATCTCTGGACTTGA	S831
FTXintronF	TGGAAGAAATGTAAGGCCAAGG	
ATRXex30F	CAGGTGGAAGCTCATTTTACTA	
RNU6-974P81233619R	CAAGAATGTGGAGAGCCCTTC	S823
ATRXex30F	CAGGTGGAAGCTCATTTTACTA	
RNU6-974P81166026R	GCAGTTGACACTTGCTACAATG	
DDX11L9ex1F	CTCTTAGCCAGACTTCCCG	S918
ATRXex1R	CAGAGTTACTTCCAGAACCACT	
RP1-279N11.1 75913486R	AGTGCTCTCTTGTTTATGGGA	M979
ATRXex27R	CTCGATTAGCAGCTACCAAGAT	
RP1-279N11.1 75878708R	TCAACTTTGCTGAGCCAAACT	
ATRXex15R	TCTTACCAAGGCCATACAGT	
RP1-279N11.1 75879066R	GCCTGGAGCTTACAATCAG	S831
ATRXex15R	TCTTACCAAGGCCATACAGT	
FTXintronR	CAGGGAGCTTTCACATCAACA	S831
ATRXex27R	CTCGATTAGCAGCTACCAAGAT	
RNU6-974P81233619F	TCACTGAGAACATTCACCC	S831
ATRXex30R	GTTGATTACTCATTTGCTGACAGG	
RNU6-974P81166026F	GAGGAAGACAAGTTGGGCA	
ATRXex30R	GTTGATTACTCATTTGCTGACAGG	

Supplementary Table 7: Combination of primers used to detect fusion transcript in cohort 2. First part of table indicates primer combination used to detect each fusion transcript. Second part describes primer combination used to detect potential reciprocal fusion transcript. TD60°C PCR program described in Supplementary Table 5 legend was used. F: forward, R: reverse, Ex: exon.

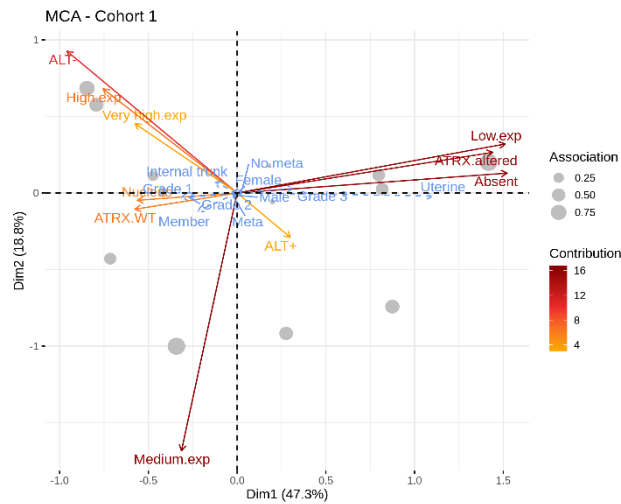


Supplementary Figure 1: Validated ATRX fusions in cohort 1

(A) PCR and sequencing validation on gDNA of ATRX fusions in cohort 1. Regarding LMS 1, the deletion is also shown as a representation of the chromosomal region before and after the deletion of the 5' end of ATRX associated with a duplication-inversion-insertion of an upstream sequence. Sequence direction is represented by an arrow highlighting the complex rearrangement. Duplicated-inverted and inverted sequences are indicated by blue and yellow boxes, respectively. g: genomic coordinate, Cent.: centromere, Tel.: telomere, Ex: exon. Chromatograms are presented with the delimitation of the fusion point. (b) RT-PCR and sequencing validation on cDNA of ATRX fusion transcript in LMS66. Chromatogram is presented with the delimitation of the two fusion points and the insertion between them.

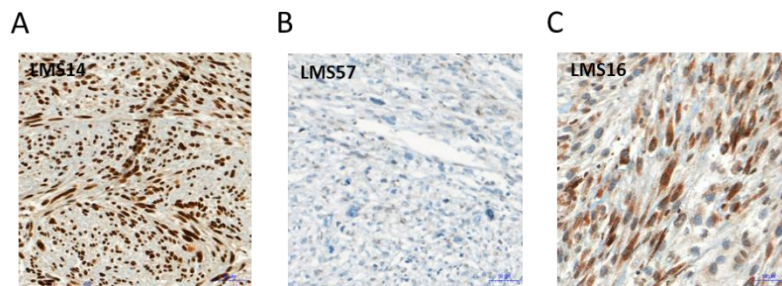


Supplementary Figure 3: Last three validated ATRX frameshifts in cohort 1
 PCR/RT-PCR and Sanger sequencing validation of last three frameshifts detected in cohort 1. Green box surrounds bases deleted in alternative product, and green box borders sequence duplicated in alternative sequence for LMS63. Red lines highlight abnormal junctions between two reference sequences in alternative sequences.



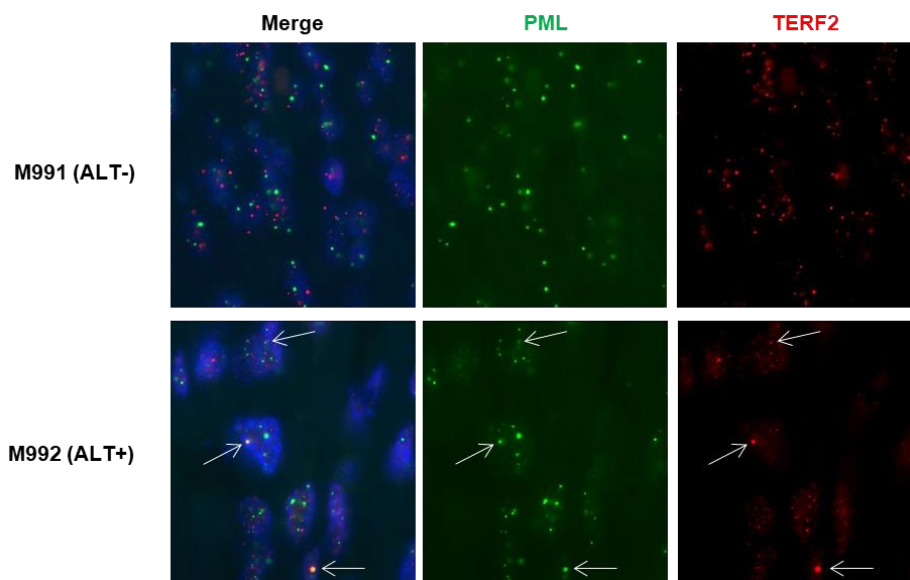
Supplementary Figure 4: Multiple correspondence analysis

Multiple correspondence analysis (MCA) was performed on ATRX status, mRNA expression, ATRX location and ALT phenotype. Additional variables and sample groups are represented in blue arrows and gray points, respectively.



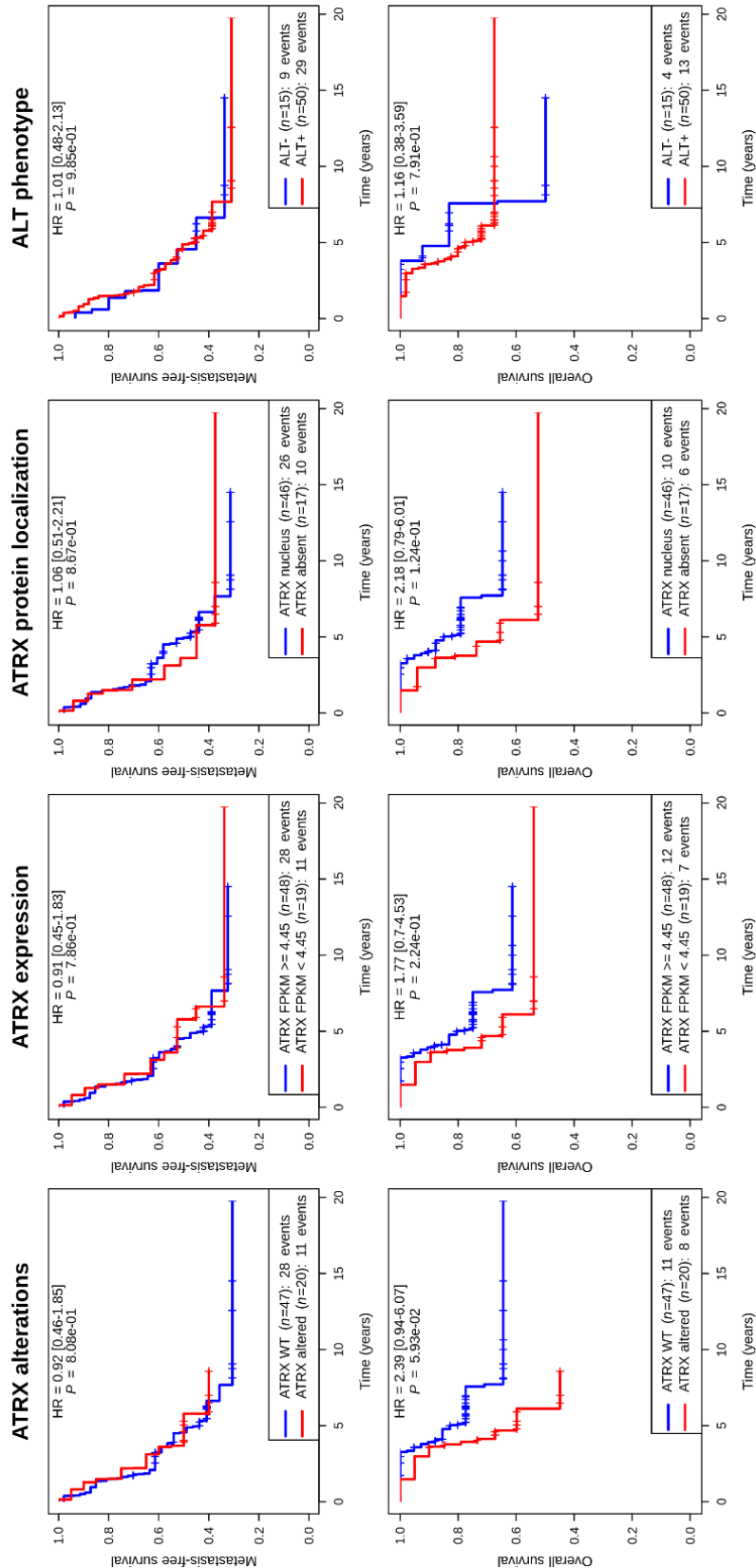
Supplementary Figure 5: Example of ATRX staining by immunohistochemistry

Leiomyosarcomas with (A) strong nuclear staining, (B) no tumorous staining or (C) cytoplasmic staining. In all samples, non-tumorous cells serve as internal control and have nuclear staining.



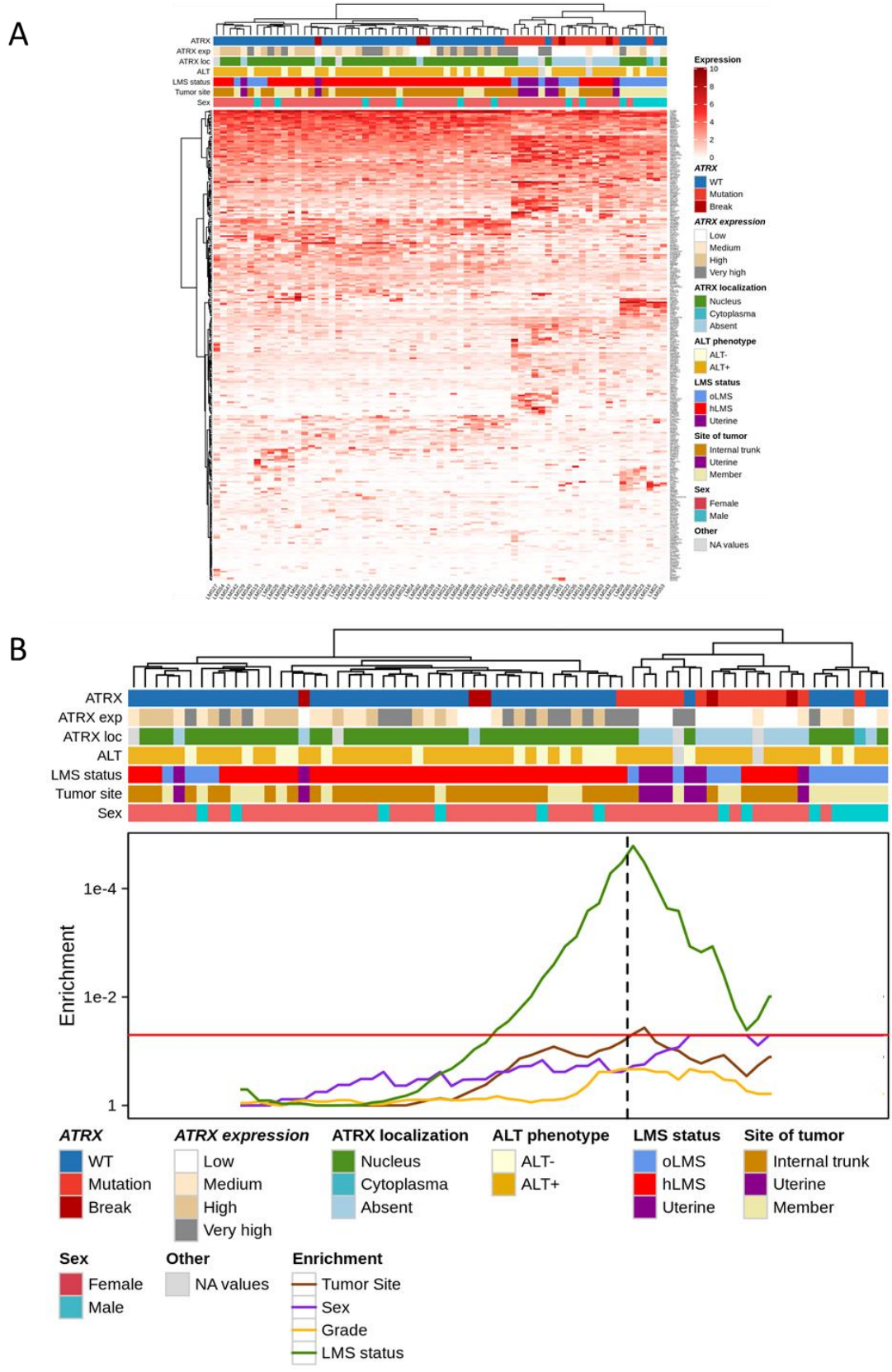
Supplementary Figure 6: ALT status determined by PML/TERF2 immunofluorescence analysis

Immunofluorescence showing absence of co-localization of PML (green) and TERF2 (red) signals in a tumor sample (ALT-) and of PML and TERF2 protein co-localization (ALT+), indicating that the ALT mechanism should be active in this tumor (ALT+). Protein co-localizations are highlighted by arrows. Magnification: X1000.

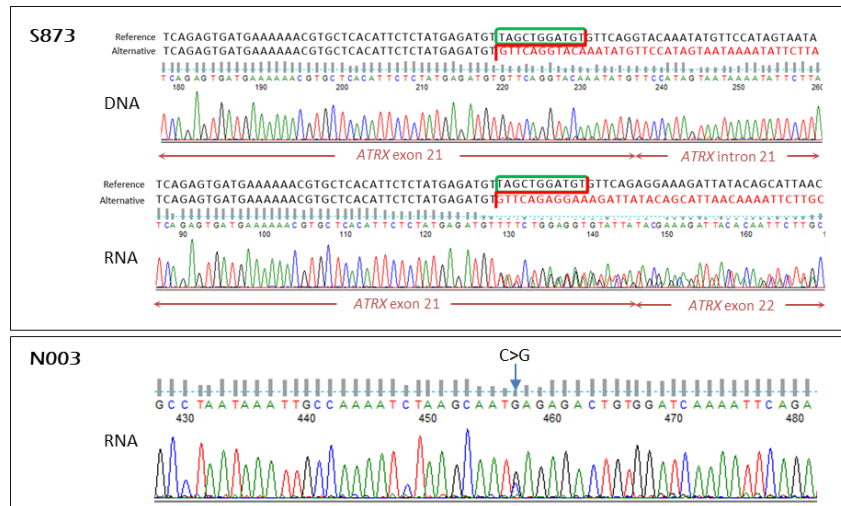


Supplementary Figure 7: Kaplan-Meier analysis in leiomyosarcomas (Cohort 1), according to ATRX status (wild-type vs altered), ATRX expression (high vs low), ATRX localization (nucleus vs absent) and ALT mechanism phenotype (ALT- vs ALT+)

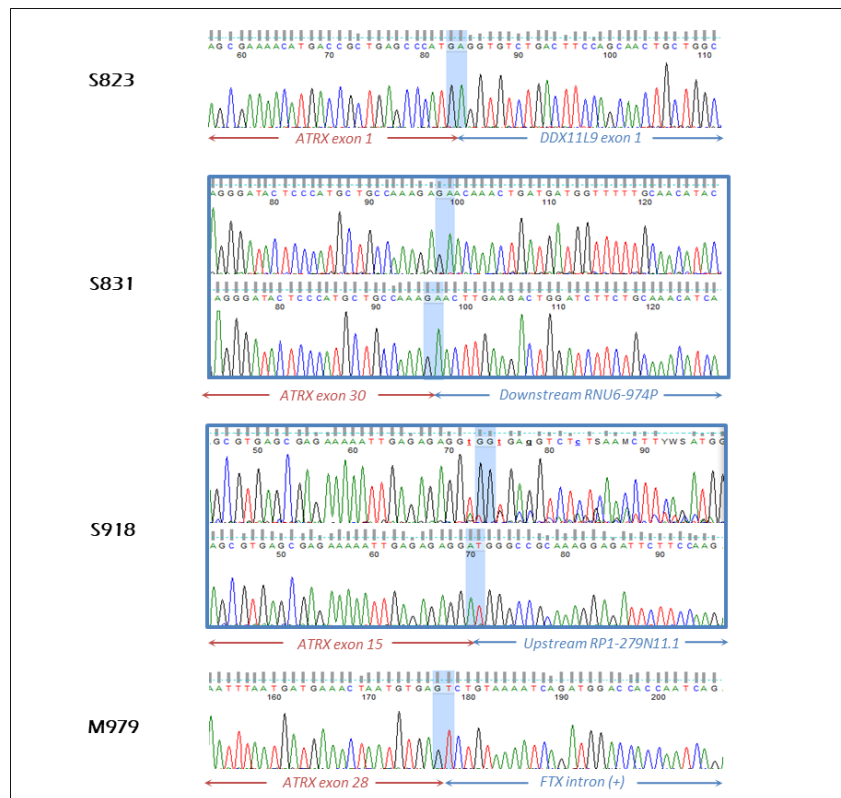
Up: metastasis-free survival. Down: overall survival. To subdivide ATRX expression into two groups, it is plotted for ATRX wild-type and altered cases, separately. Intersection between these two density curves is 4.45 (log₂(FPKM+1)).



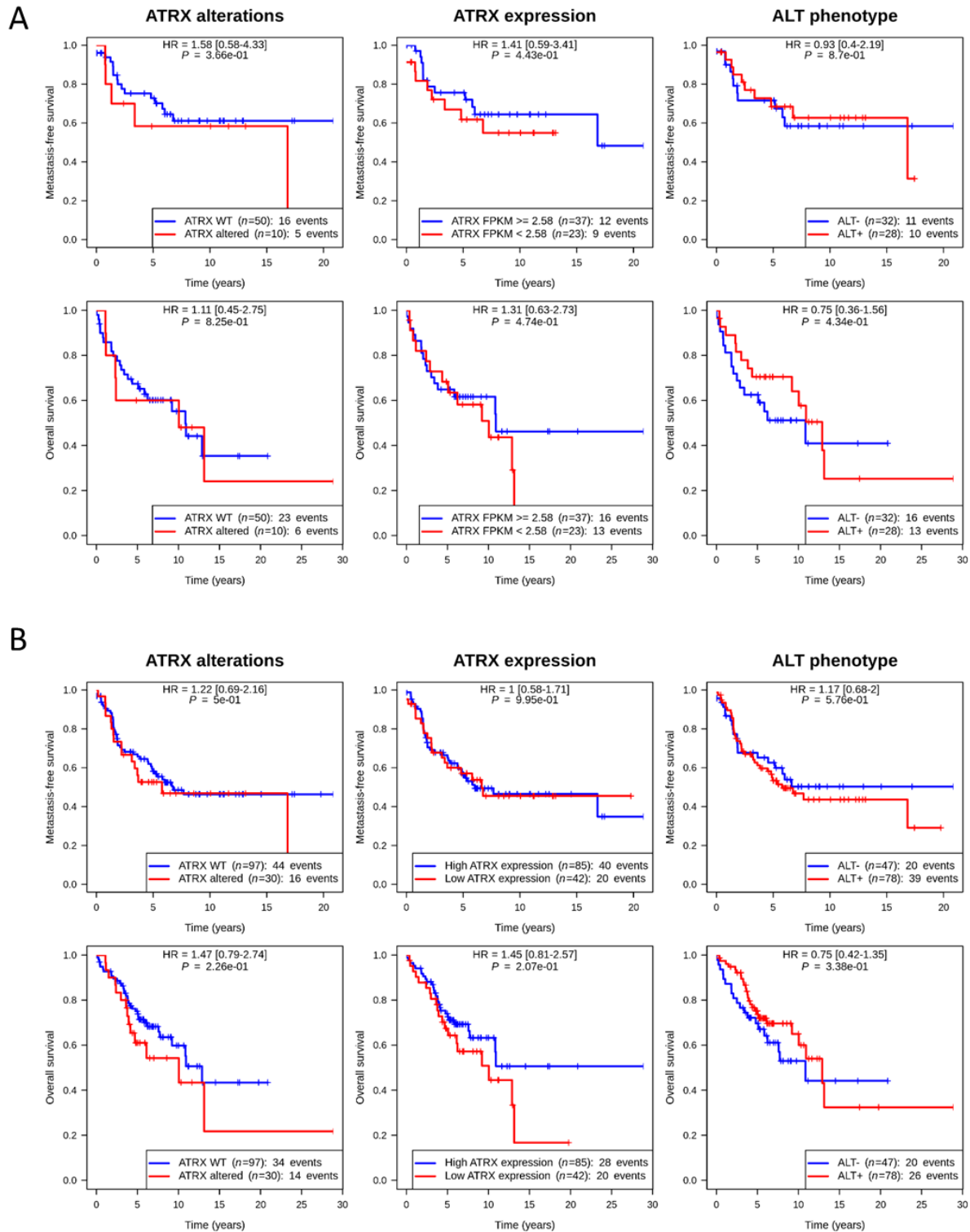
Supplementary Figure 8: Clustering with differentially expressed genes according to ATRX status (A) mRNA expression ($\log_2(\text{FPKM}+1)$) of 279 differentially expressed genes in ATRX altered tumors (see also Figure 3A; $P \leq 0.01$ and fold-change ≤ -2 or ≥ 2). (B) Enrichment analysis with clinical data, according to the cluster above. Color coding of tumor characteristics shown below. P-values obtained by Fisher test were adjusted with Benjamini and Hochberg method. Red line represents threshold ($P = 0.05$) and dotted line represents the two clusters. Concerning LMS status, oLMS and hLMS means “other” and “homogeneous” LMS, respectively (Darbo et al, 2020).



Supplementary Figure 9: Validated ATRX frameshift and non-sense mutation not detected by RNAseq in cohort 2 PCR/RT-PCR and Sanger sequencing. Chromatograms are shown. Green box surrounds bases deleted in alternative product and red lines highlight abnormal junctions between two reference sequences in alternative sequences. Substitution in N003 is highlighted by an arrow.



Supplementary Figure 10: Validated ATRX fusion transcripts in cohort 2 RT-PCR and Sanger sequencing validation of each rearrangement are shown in chromatograms of boundary amplification and Sanger sequencing.

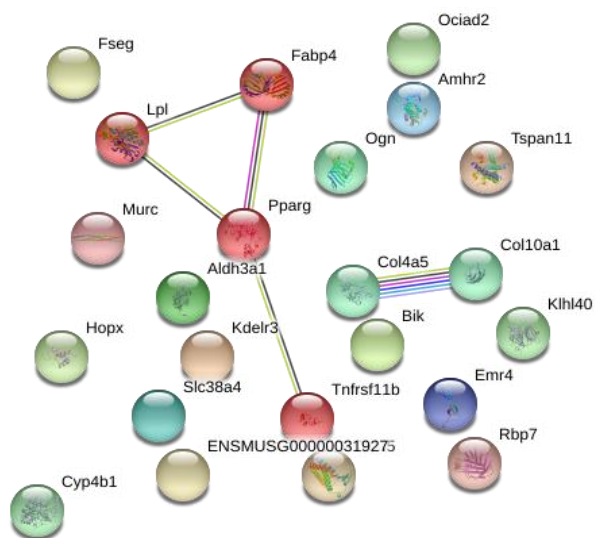


Supplementary Figure 11: Kaplan-Meier analyses

(A) Kaplan-Meier analysis in poorly differentiated sarcomas (Cohort 2) according to ATRX status (wild-type vs altered), ATRX expression (high vs low) and ALT phenotype (ALT- vs ALT+). Up: metastasis-free survival. Down: overall survival. To subdivide ATRX expression into two groups, it is plotted for ATRX wild-type and altered cases separately. Intersection between these two density curves is 2.58 (log₂(FPKM+1)). (B) Kaplan-Meier analysis in leiomyosarcomas and poorly differentiated sarcomas (Cohorts 1 and 2), depending on ATRX status (wild-type vs altered), ATRX expression (high vs low) and ALT mechanism phenotype (ALT- vs ALT+). Up: metastasis-free survival. Down: overall survival. To subdivide ATRX expression into two groups, the two cohorts are taken separately to identify the cut-off (see also Supp. Figures 4A and 6A).



Supplementary Figure 12: C-circle analysis of ALT status upon shRNA against ATRX or in control condition. Positive control is U2OS DNA.



Supplementary Figure 13: Links between proteins encoded by genes overexpressed in K7M2 ATRXKD tumors displayed by STRING Database showing two clusters via MCL cluster.

III. Conclusions

Environ un quart des sarcomes pléomorphes comporteraient donc une altération d'*ATRX*, avec une proportion un peu plus importante dans les LMS que dans les autres sarcomes pléomorphes, menant très majoritairement à la perte de la protéine au noyau. Ces altérations sont, plus souvent, des mutations que des altérations de structure ce qui diffère de la plupart des autres gènes dans les modifications du génome des LMS. Cependant, les altérations d'*ATRX* se trouvent dans des localisations particulières : elles sont retrouvées dans 75% des LMS utérins et dans des oLMS qui sont plutôt dans les membres. Les oLMS sont liés à une différenciation faible et viennent probablement de fibroblastes ou de cellules mésenchymateuses souches. De plus, la perte d'*ATRX* étant liée à la sous-expression de gènes de l'activité musculaire, nous émettons l'hypothèse que la perte d'*ATRX* donne des avantages aux cellules de sarcomes que si celles-ci sont peu différenciées. Cette perte ne serait donc pas un *driver* fort dirigeant toute l'oncogenèse des oLMS, des LMS utérins ou des autres sarcomes peu différenciés mais un *driver* donnant des avantages significatifs aux cellules portant cette altération.

En effet, *in vivo* des UPS sous-exprimant *ATRX* ont une vitesse de croissance plus importante que les UPS exprimant *ATRX* normalement. De plus, dans les modèles murins immunocompétents, la sous-expression d'*ATRX* est associée à plus de risques de développer un sarcome pléomorphe indifférencié et à une sous-représentation des mastocytes dans ces tumeurs. Ces résultats, joints à la sous-expression des gènes liés à l'activation de l'immunité adaptative quand *ATRX* est perdue dans la cohorte de sarcomes peu différenciés, permettent de supposer que la perte d'*ATRX* autorise un échappement au système immunitaire *via* le non-recrutement de mastocytes.

Comme expliqué dans le paragraphe « **L'immunothérapie** », celle-ci n'a pas de résultats très encourageants dans les sarcomes. Or, les thérapies d'augmentation des capacités des mastocytes ont eu de bons résultats dans d'autres tumeurs solides (Oldford and Marshall, 2015) mais n'ont jamais été testées, à ma connaissance, dans les sarcomes. En connaissant l'effet de la perte d'*ATRX*, une stratégie de combinaison de chimiokines, pour recruter les mastocytes, et d'anticorps IgE, pour activer leur rôles

anti-tumoraux par le recrutement d'autres cellules immunitaires (Oldford *et al.*, 2010; Crawford *et al.*, 2018), pourrait être proposé pour le traitement des sarcomes pléomorphes peu différenciés comportant une perte d'ATRX.

C. Comment la sur-activation de la voie SRF/MYOCD est-elle le *driver* des hLMS ?

I. Contexte

Dans l'étude « **Article 1 : SRF-MYOCD axis is the targetable driver of a well differentiated and aggressive subtype of Leiomyosarcomas** », nous avons émis l'hypothèse que la voie centrale de l'oncogenèse des LMS présentant un transcriptome homogène ou hLMS est celle de SRF/MYOCD. Cette voie servirait de *driver* fort car elle semble constamment suractivée de par la surexpression de *MYOCD*, permettant la transcription forte des gènes à l'origine de la différenciation musculaire lisse avec son partenaire SRF. Cette différenciation importante des cellules de hLMS alors que les tumeurs sont plus agressives que celles moins différenciées de oLMS est contradictoire avec ce qui est connu de l'oncogenèse. Dans la grande majorité des tumeurs solides, l'agressivité tumorale est associée à la dédifférenciation des cellules cancéreuses lors de l'oncogenèse (Jögi *et al.*, 2012). Dans un contexte de hLMS déficient en PTEN, il peut être envisagé que les cibles de SRF soient modifiées (Horita *et al.*, 2016) et ceci s'accompagne ainsi de la co-transcription de gènes prolifératifs et de gènes de la différenciation. Le fait que les lignées cellulaires de hLMS soient plus sensibles à l'inhibiteur spécifique de l'interaction de SRF/MYOCD que les lignées de oLMS et d'UPS nous conforte dans l'idée que cette voie est centrale dans l'oncogenèse des hLMS. Il est intéressant d'analyser maintenant les différences dans les cibles de SRF seul, de SRF/MYOCD et de MYOCD seul entre hLMS et oLMS/UPS afin de valider cette hypothèse. Mais aussi de comprendre quels changements s'opèrent dans les hLMS lors du traitement avec l'inhibiteur de SRF/MYOCD qui ne se passe pas dans les oLMS et UPS, permettant ainsi d'aller plus loin dans la validation de cette inhibition comme traitement des hLMS.

II. Résultats

1. Validation des résultats protéiques précédents dans l'ensemble de la cohorte de LMS

Pour valider les caractéristiques biologiques des hLMS (absence de PTEN, surexpression de *MYOCD*, présence de DMD à la membrane) *versus* oLMS dans l'ensemble de la cohorte (120 cas), les analyses d'expression de l'ARN et les immunohistochimies ou immunofluorescences de PTEN, *MYOCD* et DMD ont été réalisées sur 110 LMS présents sur TMA. Parmi lesquels, 14 cas de LMS utérins ont été sortis de l'analyse. Les statuts h ou oLMS ont été déterminés par la méthode des centroïdes comme pour la première phase de la cohorte.

MYOCD est significativement surexprimé dans les hLMS ($P < 1 \times 10^{-4}$), PTEN est significativement perdu dans les hLMS ($P = 1,4 \times 10^{-2}$) et l'ensemble des isoformes de DMD (Dp427 seule ou toutes les isoformes) se trouve préférentiellement à la membrane dans les hLMS ($P = 7 \times 10^{-4}$ et $P = 1 \times 10^{-4}$ respectivement)(**Figure 28 B, C et D**). La distribution des localisations des protéines ou d'expression des ARN entre h et oLMS reste la même en augmentant la taille de la cohorte. Afin d'avoir une connaissance globale de l'expression de l'axe SRF/*MYOCD*, l'immunohistochimie de SRF a aussi été mise au point et effectuée sur les 110 cas des TMA. SRF est exprimé dans tous les LMS mais avec des niveaux d'expression différents. Le niveau d'expression le plus haut est significativement associé aux hLMS ($P = 4,4 \times 10^{-3}$)(**Figure 28 A**). Dans l'optique de déterminer les combinaisons de localisations protéiques ou d'expression ARN les plus probables dans un hLMS, une analyse des correspondances multiples a été faite. Celle-ci montre que dans la plupart des hLMS, il y a une forte expression en ARN de *MYOCD*, une quantité protéique de SRF importante, PTEN absent et DMD à la membrane. De plus, ces caractéristiques sont associées à des LMS bien différenciés et intra-abdominaux validant le statut des tumeurs de hLMS (**Figure 28 E**).

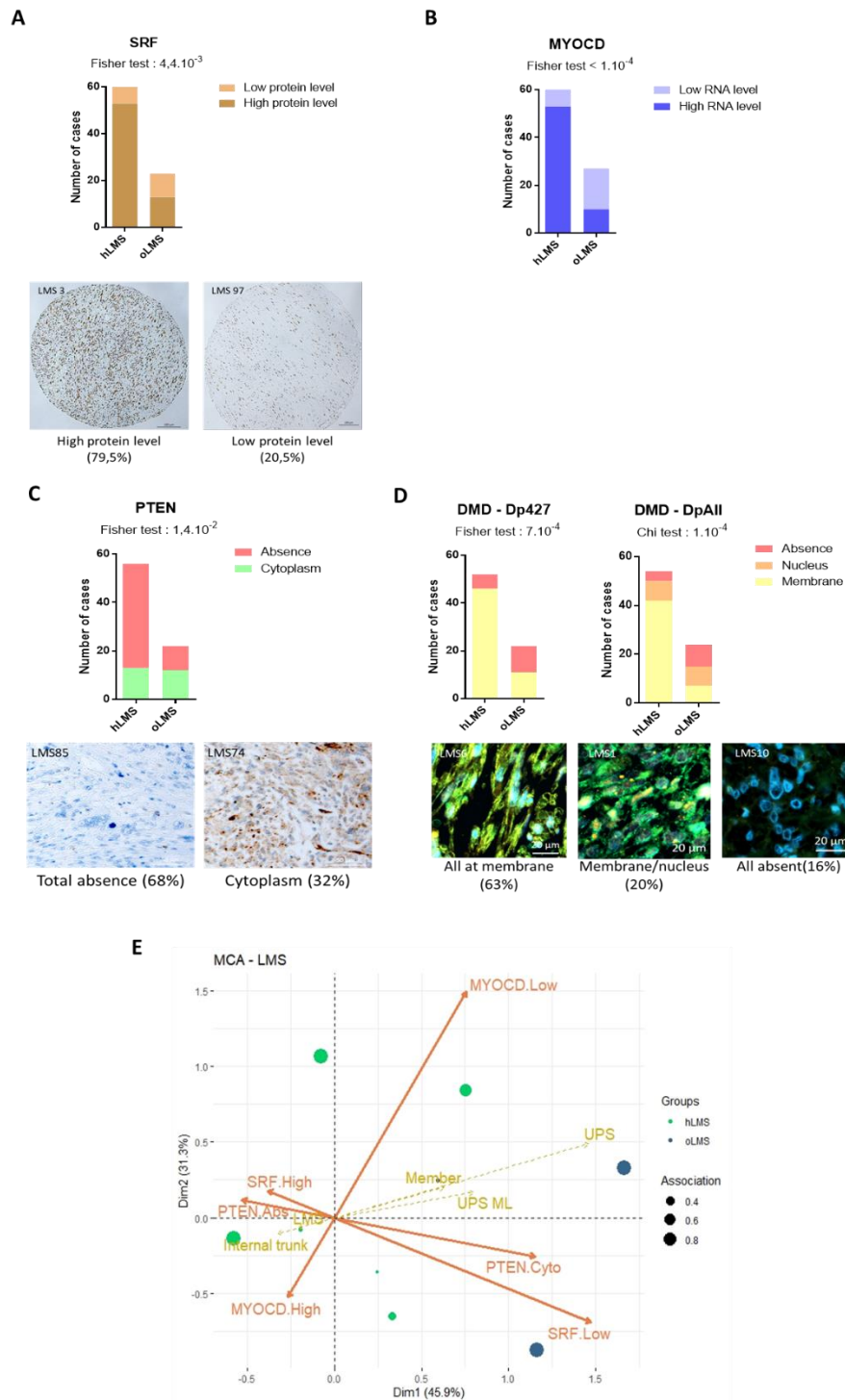


Figure 28 : Caractérisation de la voie SRF/MYOCD dans l'ensemble de la cohorte de LMS

A Immunohistochimie contre SRF et distribution des deux niveaux de protéine parmi les h et oLMS. **B** Distribution des deux niveaux d'expression en ARN, limite entre les groupes à la médiane d'expression en ARN, parmi les h et les oLMS. **C** Immunohistochimie contre PTEN et distribution des localisations de la protéine parmi les h et oLMS. **D** Immunofluorescence contre Dp427 (vert) et l'ensemble des isoformes (orange), noyaux marqués au DAPI (bleu) et distribution des localisations de la protéine parmi les h et oLMS. **E** Analyse des correspondances multiples combinant les résultats des analyses **A**, **B**, **C** et **D**. La significativité des associations est calculée par le test exact de Fisher pour **A**, **B**, **C** et **D** (Dp427) et par le test du Chi² pour **D** (DpAll).

LMS : LMS bien différencié, UPS ML : LMS à différenciation faible, UPS : LMS dédifférencié.

2. Analyse de l'effet de la suractivation de SRF/MYOCD dans les hLMS

a. Stratégie d'analyse

SRF et MYOCD étant des facteurs de transcription travaillant ensemble et séparément, la méthode la plus adaptée pour évaluer les changements de cibles entre SMC, hLMS et oLMS/UPS est la technique de séquençage d'immunoprécipitation de la chromatine (ChIPseq), dont le principe est expliqué dans la **Figure 29**, avec un anticorps dirigé contre SRF et un contre MYOCD. L'analyse des conséquences sur le transcriptome de ces changements peut se faire par la comparaison de chaque ChIPseq avec le séquençage de l'ARN des différentes lignées.

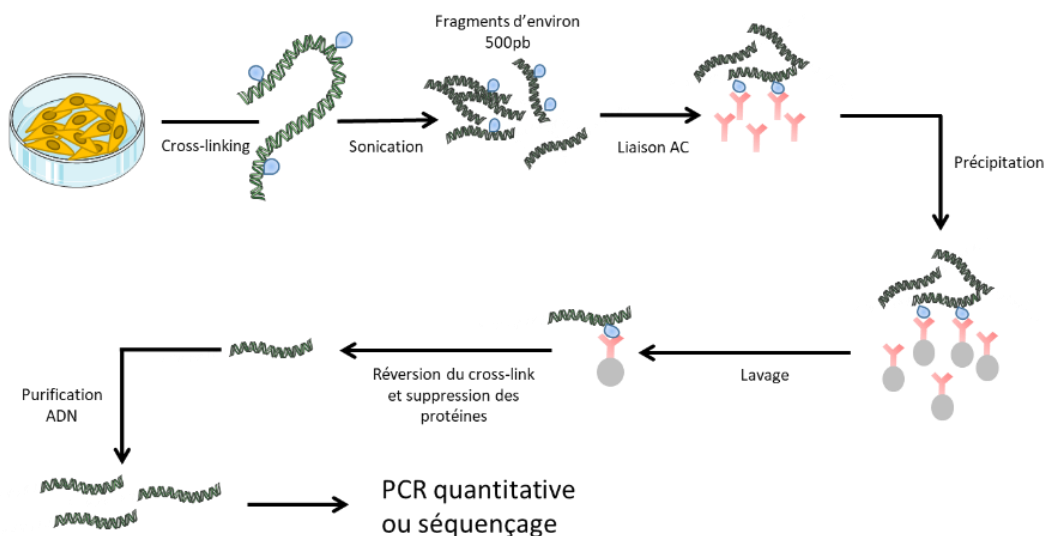


Figure 29 : Principe d'immunoprécipitation de la chromatine (ChIP)

Pour réaliser cette technique, les cinq lignées choisies sont les mêmes que celles utilisées dans l'étude « **Article 1 : SRF-MYOCD axis is the targetable driver of a well differentiated and aggressive subtype of Leiomyosarcomas** » : OC80 un hLMS (avec une amplification de MYOCD), OC88 un oLMS, OC48 un oLMS (avec un gain de MYOCD), OC98 et OC110 deux UPS. Une lignée cellulaire de SMC vasculaire normale a aussi été choisie.

b. Recherche d'anticorps pour MYOCD

Aucune ChIP n'ayant été réalisée avec un anticorps dirigé contre MYOCD, il a d'abord fallu valider la spécificité d'au moins un anticorps du commerce. Pour ce

faire, l'expression de *MYOCD* a, dans un premier temps, été mesurée par qPCR dans trois lignées : OC48 un oLMS qui a un gain de *MYOCD*, OC88 un oLMS et OC98 un UPS (**Figure 30 A**). Puis, quatre anticorps dirigés contre *MYOCD* ont été testés par des western blot (**Figure 30 B**). Dans tous les cas, les anticorps étaient soit non spécifiques avec des bandes à des tailles inattendues, soit avec des résultats aberrants quant à la quantité de *MYOCD* attendue (plus important dans OC88 qui n'exprime pas l'ARN de *MYOCD* que dans OC48).

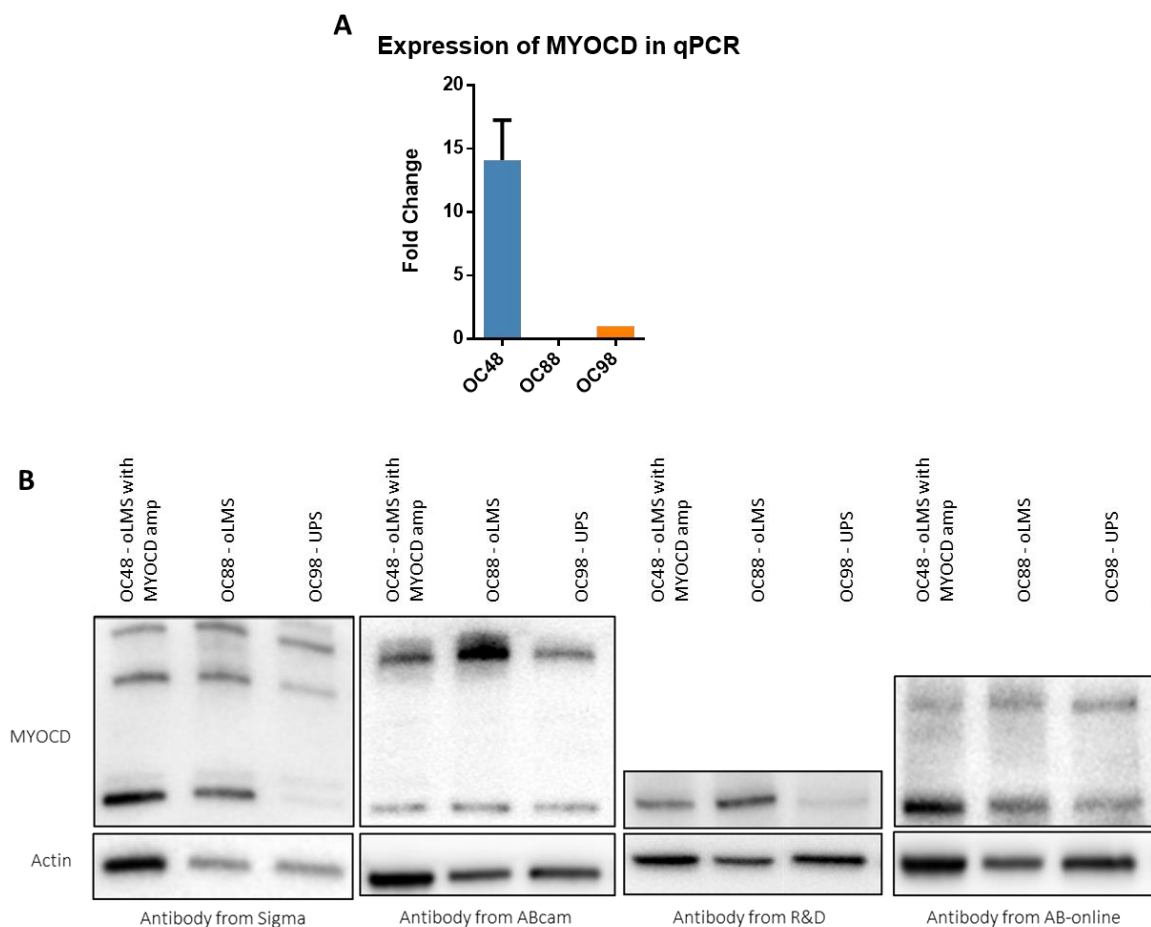


Figure 30 : Recherche d'anticorps spécifiques de MYOCD

A Estimation de l'expression en ARN de OC88 (oLMS avec un gain de *MYOCD*), OC48 (oLMS) et OC98 (UPS). **B** Test de la spécificité contre *MYOCD* de 4 anticorps commerciaux dans les mêmes lignées cellulaires.

En juillet 2018, période durant laquelle j'effectuais ces tests, Lyu et ces collègues ont publié une étude prouvant que *MYOCD* est absent des bandes à 100 kDa, supposées spécifiques, extraites par les anticorps commerciaux. Ils ont donc réalisé un étiquetage, Tag-HA et Tag-FLAG, du gène endogène de *MYOCD* par CRISPR dans des

embryons de souris. Après un développement normal de la souris, ils ont récupéré les SMC et avec des anticorps spécifiques des Tag, récupéré la protéine MYOCD endogène. La bande unique se situait à 120 kDa et la spectrométrie de masse y valide la présence de MYOCD. Les anticorps commerciaux ne sont donc en rien spécifique pour la détection de la protéine MYOCD (Lyu *et al.*, 2018). C'est pourquoi, j'ai arrêté la recherche d'anticorps spécifiques commerciaux pour mettre en place une solution adaptée.

c. Mise au point d'un TAG de MYOCD par CRISPR

Afin de palier à l'absence d'anticorps dirigés contre MYOCD, un étiquetage par un Tag-FLAGx3 de la protéine endogène des lignées cellulaires modèles a été effectué. Le système CRISPR associé à la recombinaison homologue et au système CRE/LOX, afin de supprimer les séquences permettant de sélectionner les cellules recombinées, a été choisi pour réaliser ce marquage (**Figure 31**).

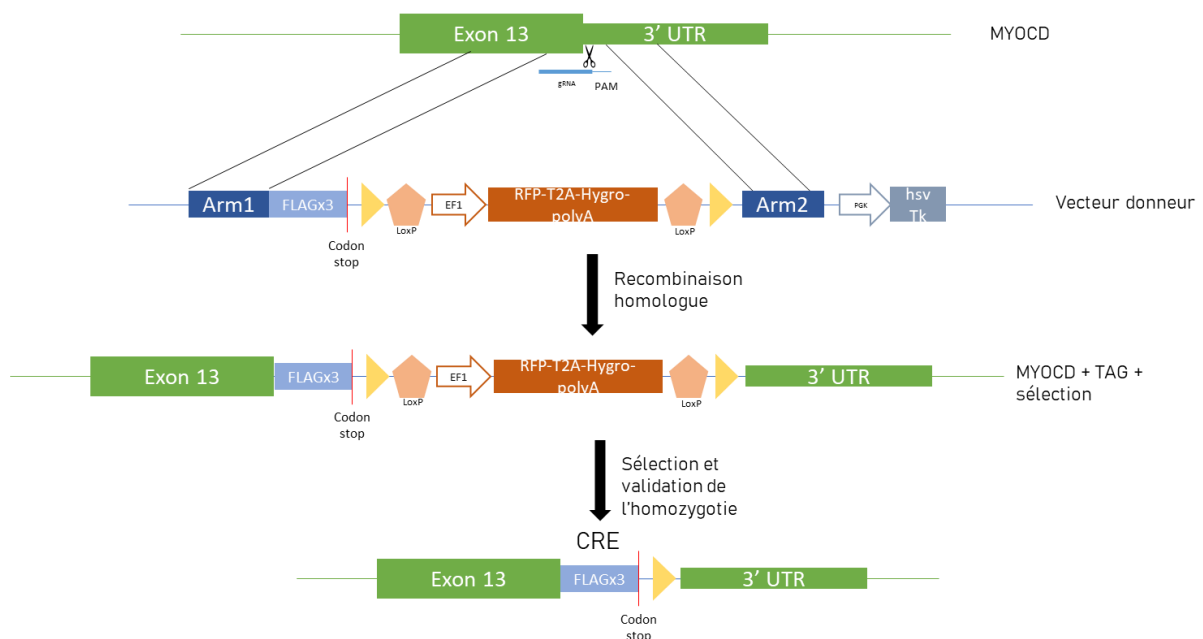


Figure 31 : Représentation schématique de la technique d'étiquetage de MYOCD endogène

Pour limiter au maximum la sélection de cellules ayant intégré, dans une mauvaise cible, le Tag-FLAGx3 et les séquences de sélection par un autre mécanisme que la recombinaison homologue, le vecteur comporte un promoteur et un gène dit de « suicide », c'est-à-dire que si ce gène est exprimé, alors la cellule meurt sous l'effet

d'une drogue, le ganciclovir (**Figure 32**).

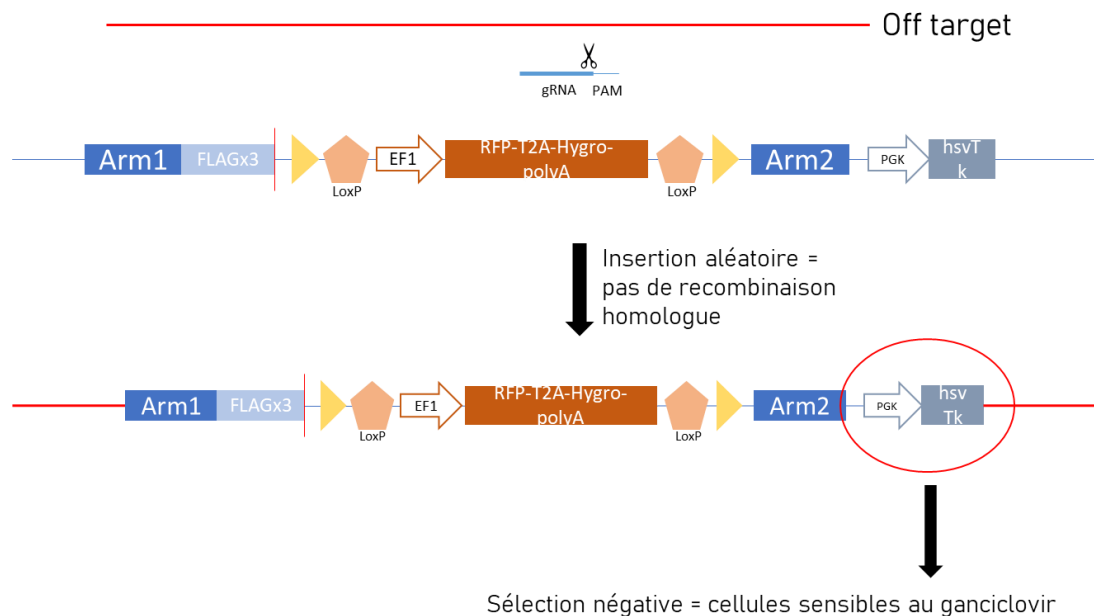


Figure 32 : Représentation schématique d'une insertion dans une mauvaise cible

L'ensemble des séquences et des techniques de préparation des vecteurs, de bactériologie et de transduction des lignées cellulaires utilisées pour créer et sélectionner les cellules portant le Tag-FLAGx3 se trouve dans la « **Partie 4 : Matériels et Méthodes** ».

Après les traitements de sélection et de suicide des cellules, la validation de la bonne intégration de la séquence se fait par deux PCR : une dite out-out et une de jonction (**Figure 33 A**), par un western blot et une immunofluorescence avec un anticorps dirigé contre le Tag-FLAG. Sur le gel d'agarose de la PCR out-out de la première lignée établie, OC48, il est possible de distinguer une bande à environ 800 pb correspondant à l'absence d'insert dans MYOCD et une à 2500 pb correspondant à la taille attendue avec l'insert. Or, la lignée sans CRISPR OC48 présente aussi cette bande, bien que très légère, celle-ci serait donc non spécifique de l'insertion du Tag-FLAG. Cette bande à 2500 pb a aussi été retrouvée sur les gels d'agarose des PCR out-out de deux autres lignées sans CRISPR OC88 et OC98. Bien que dans les lignées après CRISPR la bande à 2500 pb était plus intense, elle n'est pas spécifique. Le résultat de la PCR de jonction était négatif, au vu de l'absence de bande sur le gel d'agarose, pour les trois lignées (**Figure 33 B** ; PCR de jonction non montrée pour les lignées OC88 et

OC98). Le western blot sur la lignée OC48 avec un anti-FLAG révèle une bande non spécifique vers 70 kDa et une bande spécifique de la lignée après CRISPR à 300 kDa. La taille attendue de MYOCD est entre 100 et 120 kDa (taille théorique et taille retrouvée dans la souris après étiquetage)(Figure 33 C). Enfin une immunofluorescence avec l'anticorps anti-FLAG, met à jour la présence spécifique, dans le noyau, du FLAG dans la lignée OC48 après CRISPR et sélection (Figure 33 D). L'ensemble de ces résultats démontre que le Tag-FLAG s'est inséré dans le génome par recombinaison homologue, mais pas dans *MYOCD*. On a émis l'hypothèse qu'une partie suffisante des bras d'homologie soit proche de séquences présentes dans un autre gène. Pour vérifier cette hypothèse, un alignement a été réalisé de chacun des bras d'homologie sur les séquences de référence du génome humain avec le logiciel Blastn (Altschul *et al.*, 1997), permettant d'aligner des séquences en partie similaire. Les résultats montrent un alignement unique du premier bras d'homologie sur *MYOCD*. Le second bras s'aligne sur sa totalité uniquement sur *MYOCD* ainsi que sur 42 nucléotides de *SMARCE1* se situant en 17q21.2 (Figure 33 E). Cette hypothèse semble donc incorrecte. On peut alors proposer deux autres suppositions : la présence de micro-homologies, entre les bras et un autre gène, non détectables par ces algorithmes ou la présence d'un site hors cible non détecté de l'ARN guide dans un gène proche de *MYOCD*.

Par manque de temps, le projet d'étiquetage de *MYOCD* endogène a été arrêté après ces résultats. Par la suite, je me suis concentrée sur la mise au point de la technique de CHIP dans les lignées de sarcome en utilisant uniquement des anticorps dirigés contre SRF.

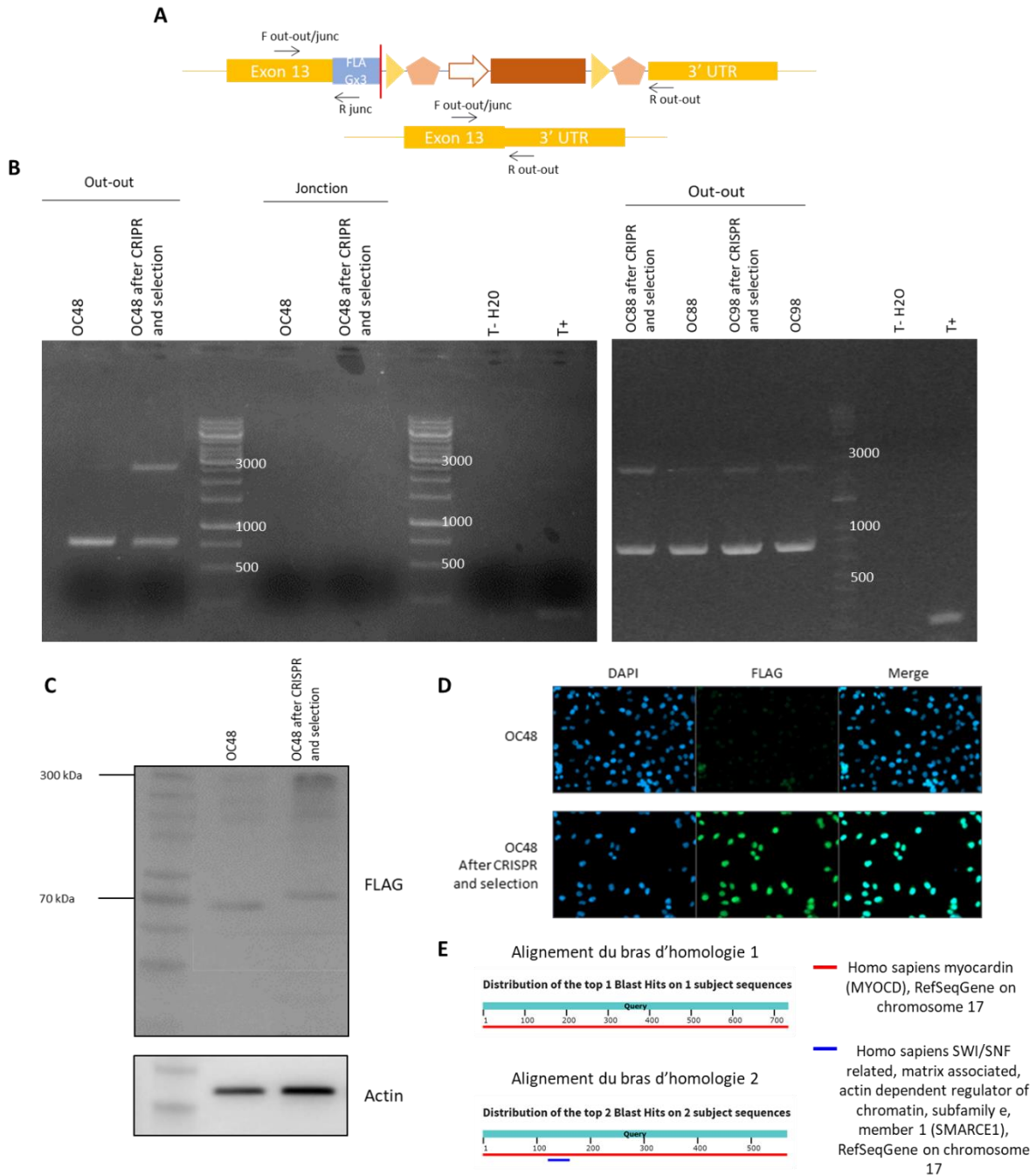


Figure 33 : Validations de l'étiquetage de MYOCD

A Représentation de la localisation des primers utilisés pour tester s'il y a insertion au bon site. **B** Gels d'agarose PCR out-out sur 3 lignées, la taille attendue en présence de l'insert est de 2500 pb, et de la PCR jonction sur une lignée, une absence de bande montre l'absence de la jonction entre l'exon 13 de *MYOCD* et l'insert (résultats identiques pour les 3 lignées – non montré). **C** Western Blot avec les extraits protéiques d'une lignée et un anticorps dirigé contre le FLAG montrant la présence d'un FLAG à environ 300 kDa (taille attendue de *MYOCD* entre 100 et 120 kDa). **D** Immunofluorescence avec un anti-FLAG sur la lignée OC48. **E** Résultat de l'alignement des bras d'homologie sur les séquences de référence du génome humain par le logiciel Blastn.

d. Mise au point de ChIP dans des lignées de sarcomes

La première étape de mise au point de la ChIP, sur des lignées cellulaires sur lesquelles cette technique n'a jamais été réalisée, est d'exécuter une sonication correcte. Pour ce faire, 2 sonicateurs ainsi que plusieurs amplitudes de sondes ont été testées sur une première lignée. Puis, pour chaque lignée, des quantités différentes de cycles ont été effectuées afin de trouver le nombre idéal de cycles pour avoir la majorité de l'ADN cassé entre 200 et 750 pb. Pour l'ensemble des lignées, 20 cycles était le bon nombre (**Figure 34 A** ; uniquement 2 lignées sont représentées ici), sauf pour OC48 pour laquelle il a fallu diluer deux fois plus les culots de cellules et pour qui le nombre idéal de cycles était 25 (**Figure 34 B**).

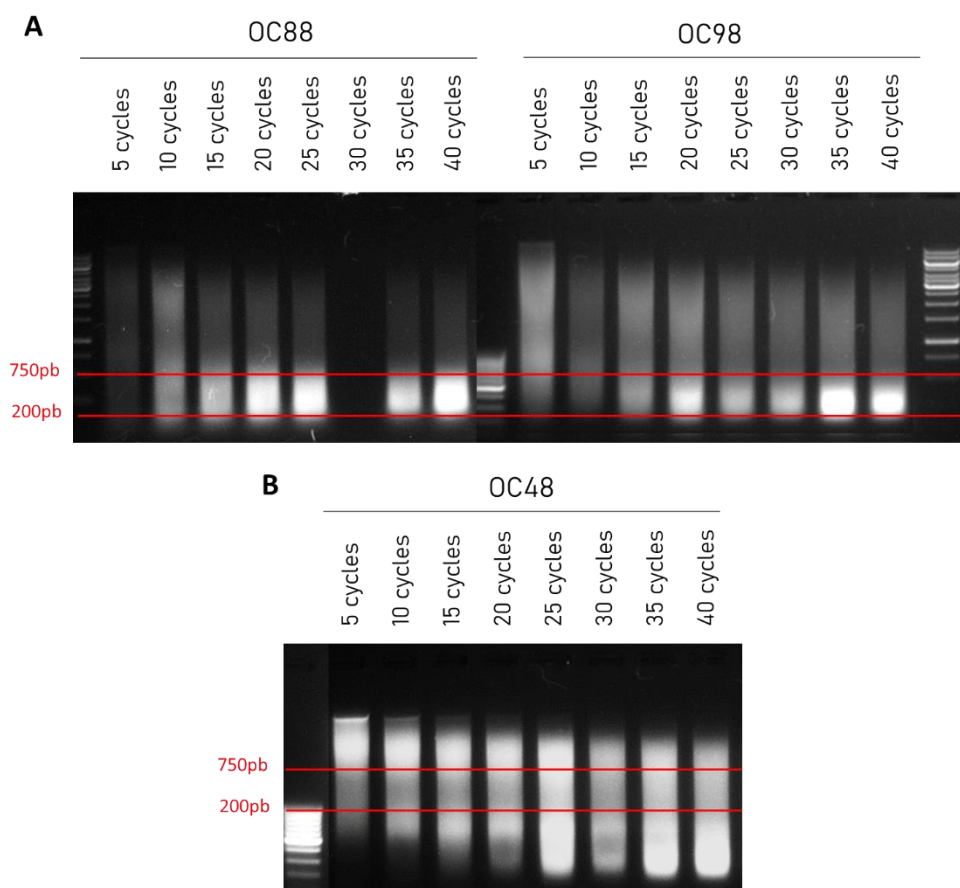


Figure 34 : Mise au point de la sonication pour les ChIP

A Mise au point de la sonication pour les lignées OC88 et OC98 dont le nombre de cycles idéal est 20, résultats similaires aux autres lignées cellulaires modèles sauf OC88. **B** Mise au point de la sonication pour la lignée OC48 avec une dilution 2 fois plus importante des culots cellulaires et pour laquelle le nombre idéal de cycles est 25.

La deuxième étape de la mise au point a été de trouver l'anticorps anti-SRF le plus adapté pour réaliser une ChIP. Deux anti-SRF, utilisés dans le projet ENCODE,

ont été testés. L'efficacité de la ChIP était contrôlée par un anticorps dirigé contre l'ARN polymérase II (PolII), des IgG qui ne doivent reconnaître aucune séquence, et par une immunoprécipitation sans anticorps (Mock). Les primers ayant servi à la qPCR à la fin de la ChIP reconnaissent, pour le contrôle négatif, une séquence intergénique du chromosome 22 (Chr22) et, pour le contrôle positif, la séquence avec un site de liaison à l'ADN de SRF, indépendamment de MYOCD, dans le promoteur de l'actine β (*ACTB*). Les sites de liaison de SRF à l'ADN ont été récupérés *via* le projet ENCODE puis les séquences de référence entourant le site de la boîte CARG ont servi de modèles pour créer les primers. L'anticorps le plus efficace était l'anticorps numéro 2 de chez Bethyl Laboratories (**Figure 35**).

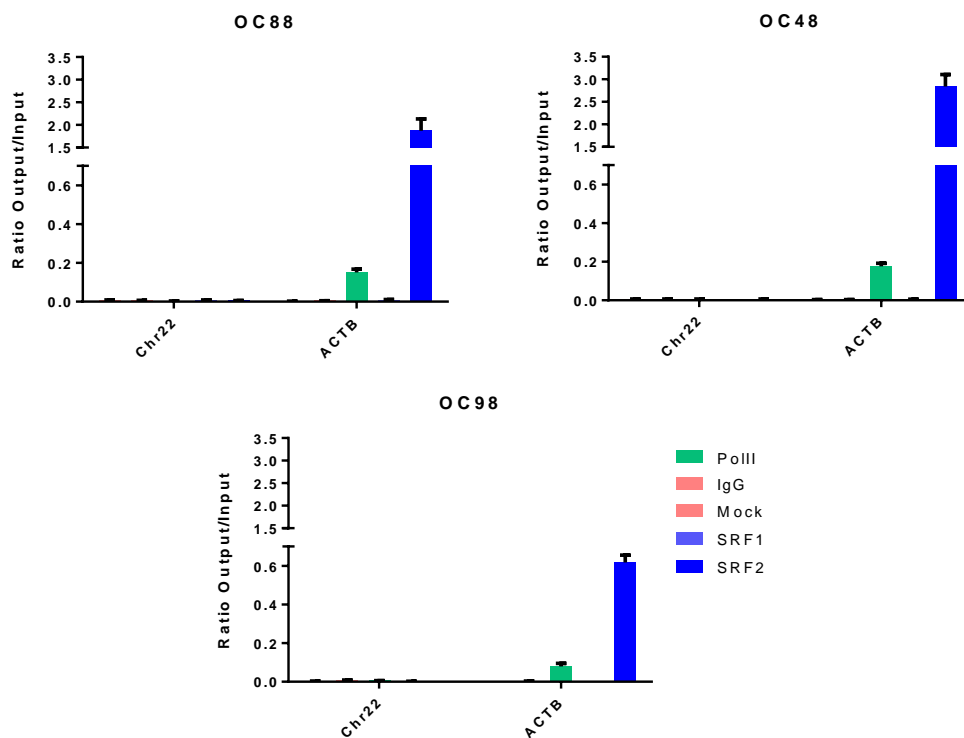


Figure 35 : qPCR de validation des ChIP testant les anticorps dirigés contre SRF

Enfin, pour quantifier la liaison à l'ADN de SRF sur les sites communs avec MYOCD, deux sites de liaison à l'ADN de SRF dans le promoteur de gènes connus pour être spécifiquement surexprimés lors de la surexpression de MYOCD (*TAGLN* et *MYL6*) ont été choisis. De la même manière que pour le site *ACTB* de SRF, les

séquences du site de liaison ont été récupérées par le projet ENCODE, cela a permis de créer des primers de qPCR encadrant les sites afin de voir les différences de liaison de SRF selon la quantité de MYOCD dans les différentes lignées.

e. Comparaison des profils qPCR de ChIP par rapport aux SMC

En comparant les profils qPCR de chaque lignée cellulaire modèle à celui de la lignée de SMC, on peut voir que la lignée de hLMS, OC80, a un profil très similaire à celui des SMC, avec SRF très présent dans le promoteur de l'actine β mais aussi présent dans ceux de *TAGLN* et *MYL6*, confirmant une origine musculaire lisse des hLMS comme évoqué dans « **Article 1 : SRF-MYOCD axis is the targetable driver of a well differentiated and aggressive subtype of Leiomyosarcomas** ». Les oLMS présentent des programmes proches entre eux mais assez différents de celui des SMC. SRF est moins présent dans l'ensemble des sites analysés voir quasiment absent de ceux des promoteurs de *TAGLN* et *MYL6* et cela malgré l'expression de *MYOCD* dans OC48. Ces profils montrent que les oLMS n'ont pas pour origine les SMC. Enfin, les deux UPS ne présentent pas de SRF dans les sites des gènes spécifiques des muscles lisses et une quantité variable de SRF dans le promoteur de l'actine β (**Figure 36**). La ressemblance entre les profils des lignées de oLMS et d'UPS permet de poser l'hypothèse d'origines cellulaires proches.

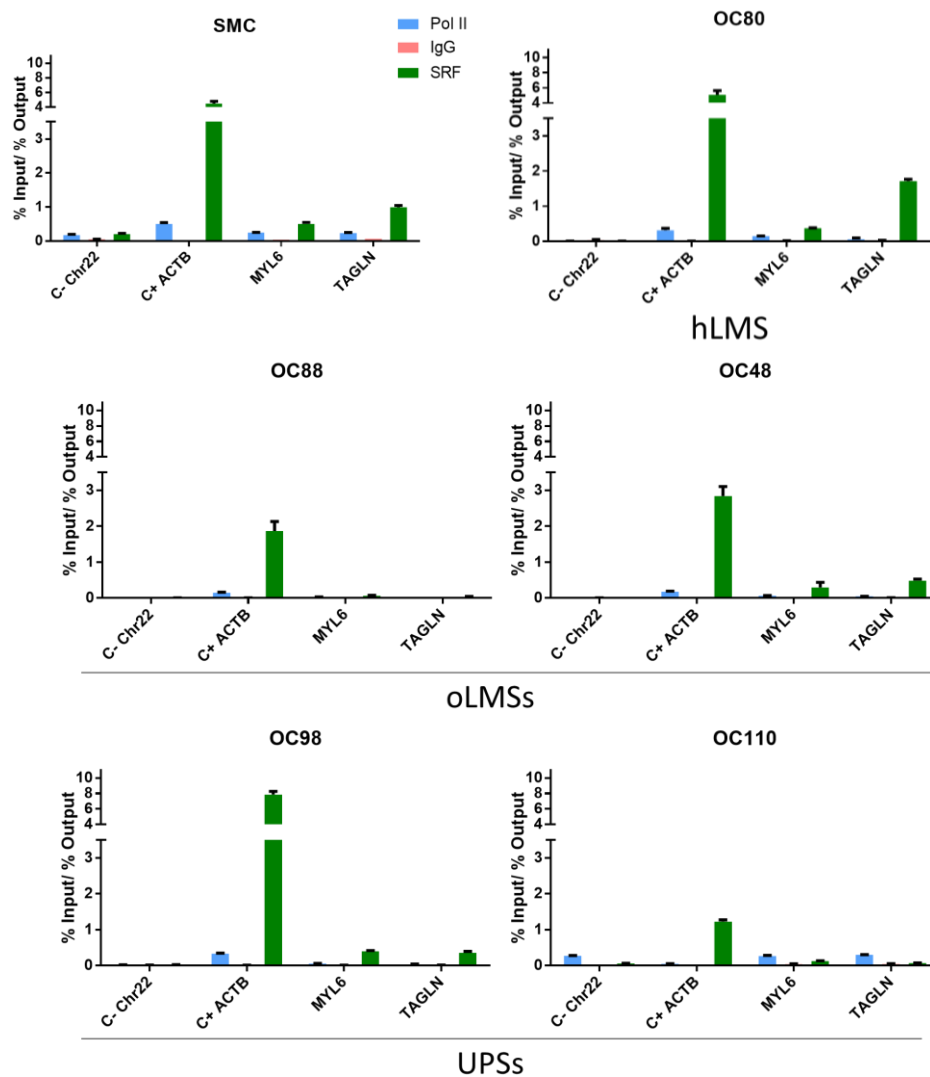


Figure 36 : Profils qPCR des ChIP sur l'ensemble des lignées cellulaires modèles

Le séquençage de ces ChIP et l'analyse comparative avec le séquençage de l'ARN de chacune de ces lignées sont en cours et permettront de connaître quelles voies sont sous contrôle de SRF dans l'oncogenèse des hLMS par rapport aux oLMS et UPS. De plus, cela permettra aussi de voir les différences entre SMC et hLMS et de répondre à la question suivante : est-ce l'augmentation du nombre de cibles différentes de SRF qui permet l'expression simultanément des gènes du phénotype contractile et du phénotype synthétique ?

3. Cibler la voie SRF/MYOCD

Il a été démontré dans l'article présentant les hLMS que ces derniers sont plus

sensibles à un inhibiteur spécifique de l'interaction SRF/MYOCD, le CCG-100602.

Dans un premier temps, des analyses de cytotoxicité ont été réalisées avec une gamme de CCG-100602 sur les cinq lignées cancéreuses. Cette analyse permet de suivre, lors d'une même expérimentation, la prolifération et la mort cellulaire, dans le temps. La prolifération est établie selon un ratio de confluence entre le temps 0 et le temps T. La mort cellulaire par apoptose est déterminée grâce à l'ajout d'un peptide à quatre aminoacides (DEVD) dans le milieu. Ce peptide va dans le noyau et, quand il est coupé par la caspase 3, elle-même activée par les voies apoptotiques, se lie à l'ADN devenant ainsi fluorescent.

Après analyse de la confluence, il est à noter que dans l'ensemble des lignées, le CCG-100602 diminue la prolifération cellulaire par rapport au DMSO de façon significative, après 48h, à partir de 6,25 µg/mL ou de 3,13 µg/mL dans les lignées OC80 et OC88 ou dans les trois autres respectivement (**Figure 37**). Cela signifie que même à faible dose cet inhibiteur agit sur la prolifération de cellules d'origine mésenchymateuse.

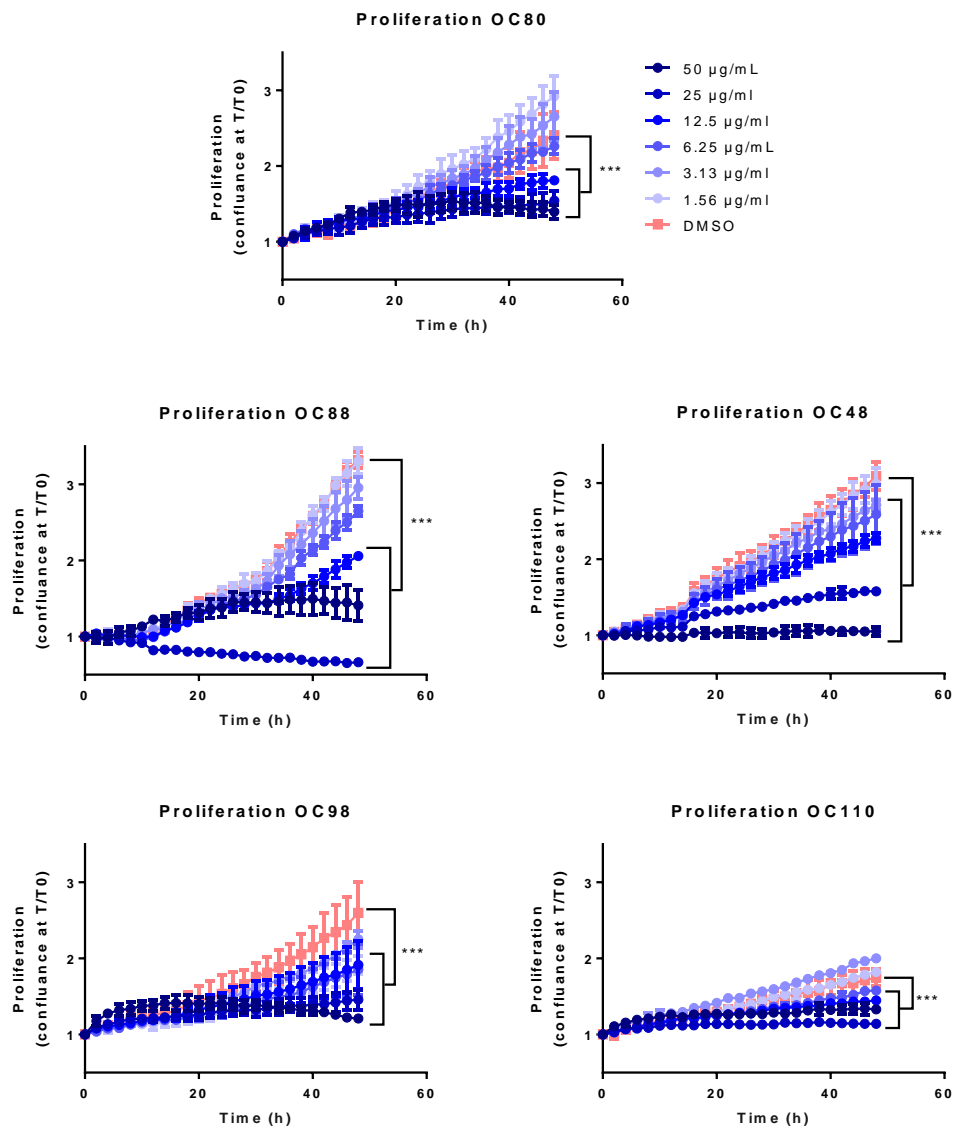


Figure 37 : Prolifération cellulaire selon une gamme de CCG-100602

Confluence au temps T divisé par la confluence durant 48h au temps 0 selon une gamme de CCG-100602 allant de 1,56 à 50 µg/mL.

L'analyse de la mort cellulaire par apoptose montre que la totalité des lignées sont sensibles au CCG-100602 avec une mort significativement augmentée par rapport au DMSO à 50 µg/mL, sauf la lignée OC88 pour qui la significativité commence à 25 µg/mL. Cependant après 48h, le pourcentage de cellules positives à la caspase 3 est significativement plus important dans la lignée de hLMS, OC80, que dans les lignées de oLMS et d'UPS (**Figure 38**). Ces résultats corroborent ceux de l'« **Article 1 : SRF-MYOCD axis is the targetable driver of a well differentiated and aggressive subtype of Leiomyosarcomas** » avec la technique du MTT.

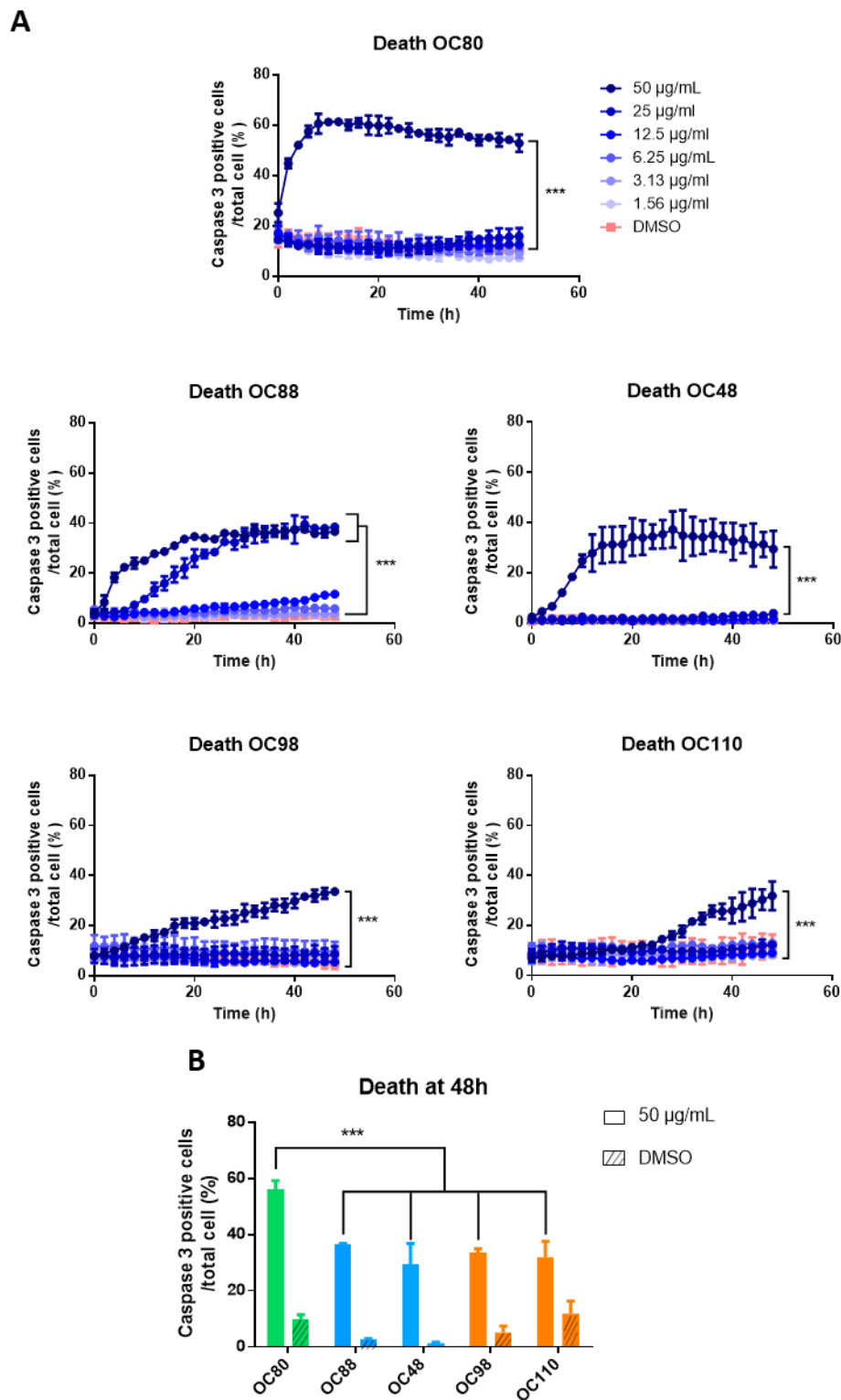


Figure 38 : Pourcentage de cellules caspase 3 positives selon une gamme de CCG-100602

A Pourcentage de cellules caspase 3 positives au cours du temps durant 48h selon une gamme de CCG-100602 allant de 1,56 à 50 µg/mL. **B** Pourcentage de cellules caspase 3 positives après 48h pour la dose 50 µg/mL et le contrôle DMSO.

La question qui en découle est : comment la mort est-elle induite et quels changements induits par l'inhibition de cette interaction induisent la mort cellulaire ?

Pour répondre à cette question, des ChIP ont été réalisées après traitement, avec le CCG-100602, des lignées cellulaires à leur IC50. La diminution de SRF au niveau de ces sites de liaison a été validée par qPCR (**Figure 39**). Puis l'ensemble des ChIP ont été envoyées à séquencer ainsi que l'ARN de chacune des lignées après traitement et dont les analyses sont en cours.

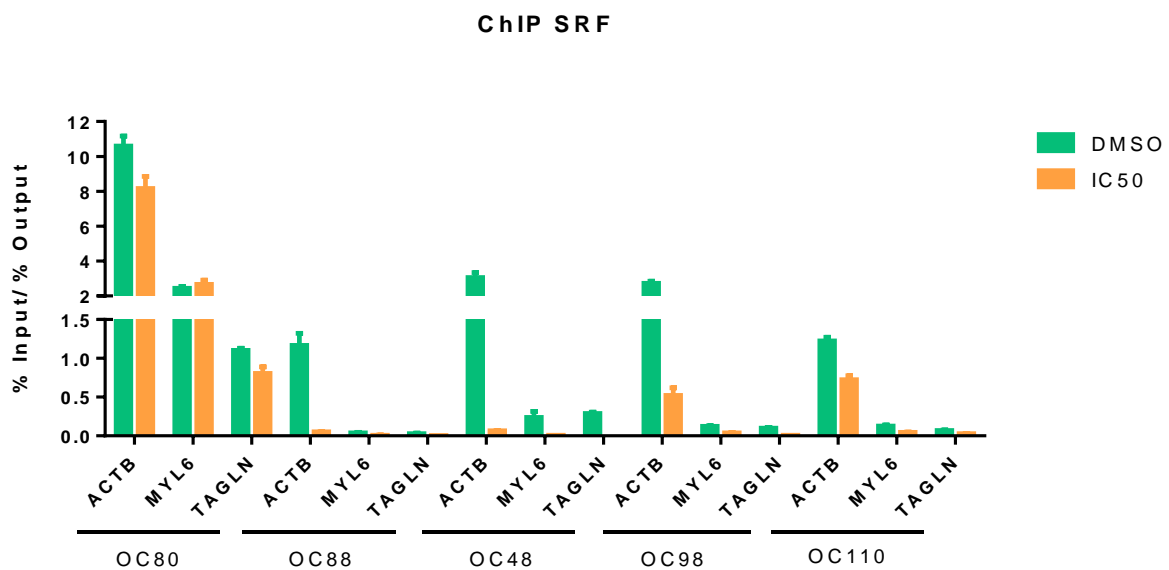


Figure 39 : qPCR de validation des ChIP avec un anti-SRF sur les lignées après traitement à l'IC50 du CCG-100602 ou au DMSO

III. Conclusions

Dans un premier temps, nous avons pu valider, dans la cohorte de 96 cas de LMS, les différences entre hLMS et oLMS quant à la surexpression de *MYOCD*, l'absence de *PTEN* et la présence de *DMD* à la membrane. Ce résultat valide, dans une cohorte plus large, les différences biologiques déterminées dans la cohorte de 37 LMS. En ajoutant l'expression forte de SRF spécifique des hLMS, le rôle central dans l'oncogenèse de l'axe SRF/MYOCD se confirme. Ces résultats associés aux profils transcriptomiques, directement liés aux gènes régulés par SRF avec ou sans MYOCD, renforcent l'intérêt d'étudier plus en détail leurs effets séparément et ensemble.

Malheureusement l'absence d'anticorps spécifiques dirigés contre MYOCD empêche de valider une surexpression protéique de cette dernière dans les tumeurs humaines. Dans les lignées cellulaires modèles, l'étude du rôle de sa surexpression et de ses cibles spécifiques, malgré les tentatives d'étiquetage de MYOCD endogène, reste en suspens et est aujourd'hui dépendante de nouveaux développements techniques.

La mise au point de la technique de CHIP sur cinq lignées cellulaires de sarcomes et sur une lignée cellulaire de SMC vasculaire avec un anticorps dirigé contre SRF permet d'observer des profils de liaison à l'ADN très similaires entre SMC vasculaire et la lignée de hLMS, OC80, donnant un élément supplémentaire à la validation de l'origine musculaire lisse vasculaire supposée des hLMS. Enfin, les profils des oLMS sont beaucoup plus proches de ceux des UPS. Au vu de la faible expression de *MYOCD* dans ces lignées de oLMS comme dans les UPS ainsi que de la faible liaison de SRF sur les cibles associées au complexe SRF/MYOCD, on peut supposer que c'est cette absence de SRF/MYOCD sur les sites des promoteurs des gènes de la différenciation musculaire lisse qui explique la faible différenciation des oLMS. Les séquençages des ARN et de l'ensemble des CHIP permettront de valider ces données préliminaires et de connaître les gènes et voies régulées par la surexpression de SRF et de *MYOCD*, en l'absence de *PTEN*, dans les hLMS.

L'inhibition de l'interaction de SRF/MYOCD par le composé CCG-100602 a

montré une sensibilité plus forte des cellules de hLMS par rapport à celles de oLMS ou d'UPS confirmant le rôle central et spécifique de l'axe SRF/MYOCD dans l'oncogenèse des hLMS. De plus, cet inhibiteur a des effets différents selon la dose utilisée, de faibles doses stoppent la prolifération alors que des doses supérieures entraineront la mort cellulaire par apoptose. Le séquençage des ChIP couplé au séquençage ARN après traitement au CCG-100602 mettra à jour les voies cellulaires modulées qui entraînent cette mort cellulaire. De plus, le CCG-100602 n'étant pas encore développé pour une utilisation chez l'humain, il sera possible de tester, sur les cellules de hLMS, des drogues déjà utilisables chez l'humain et ciblant les voies qui sont modifiées par l'inhibition de SRF/MYOCD.

Partie 3 : Discussion et Perspectives

A. LMS et sarcomes pléomorphes : le remaniement du génome comme *driver* ?

I. Des altérations permissives de modifications génomiques importantes

Parmi les altérations présentes dans notre cohorte, deux sont communes à l'ensemble des tumeurs : *TP53* et *RB1* retrouvées, respectivement, dans 100% et 97,4% des cas (le seul LMS ne présentant pas d'altération dans *RB1* est le seul LMS avec une signature mutationnelle UV). Chudasama et ses collègues ont établi que, dans l'oncogenèse des LMS, l'inactivation de *TP53* et de *RB1* est un événement nécessaire et précoce (Chudasama *et al.*, 2018). Quand ces deux altérations se retrouvent dans une même cellule, cela pourrait être un des événements initiateurs de l'oncogenèse des LMS.

Dans l'ensemble des cellules dites « normales » comme les fibroblastes, chaque nouvelle division cellulaire entraîne l'apparition de 2 à 10 mutations (Martincorena and Campbell, 2015). Il est prédit que chaque cellule somatique accumule entre plusieurs centaines et plusieurs milliers de mutations ponctuelles au cours de sa vie, ce qui est proche du taux de mutation retrouvé dans les cellules cancéreuses (Martincorena and Campbell, 2015). Il existe aussi une sélection positive des cellules somatiques non cancéreuses portant une mutation dans un gène *driver* de l'oncogenèse. Cela mène à une colonisation clonale de ces cellules dans le tissu normal, si bien que vers 52-55 ans entre 5 et 20% des cellules de l'œsophage comportent une mutation de *TP53*. Cette proportion augmente avec l'âge et les facteurs de risque associés au cancer de l'œsophage tels que la consommation de cigarettes et d'alcool (Martincorena, 2019). Pour autant, toutes ces cellules ne sont pas à l'origine de cancer et il existe une grande différence entre les cellules « normales » et les cellules cancéreuses : l'étendue de l'instabilité génomique. En effet, les cellules normales étudiées à ce jour sont très majoritairement diploïdes avec occasionnellement des changements du nombre de copies (Blokzijl *et al.*, 2016; Lee-Six *et al.*, 2018; Martincorena *et al.*, 2018; Moore *et al.*, 2018).

Les LMS surviennent principalement chez des personnes âgées de 70 à 80 ans, le nombre de cellules normales ayant au moins une mutation dans un gène *driver* est important. Il est donc plus probable qu'une cellule accumule plusieurs mutations *drivers* permissives d'une oncogenèse ; dans le contexte des LMS, ces altérations sont inactivatrices de *TP53* et *RB1*. Les protéines codées par ces gènes étant les garantes principales de l'intégrité du génome, leur absence permet la survie des cellules, malgré l'accumulation d'altérations plus importantes comme des pertes ou des gains de bras de chromosomes voire de chromosomes entiers.

Plusieurs mécanismes, non nécessairement exclusifs, peuvent être le point de départ de l'instabilité génomique des LMS comme le système ALT présent dans 76,9% des LMS de notre cohorte. Ce mécanisme conduit à la fois à l'immortalisation de la cellule, en allongeant les télomères, et à un génome instable, en insérant des séquences télomériques intra-chromosomiques, créant ainsi des points sensibles aux cassures et aux translocations (Marzec *et al.*, 2015). Un autre mécanisme à l'origine d'instabilité est le *chromothripsis*, source de réarrangement complexe, qui est retrouvé dans 35% des LMS (Chudasama *et al.*, 2018). Cet événement crée des centaines de réarrangements sur quelques parties de chromosomes et est à l'origine de translocations, pertes, gains et inversions. Enfin, les mécanismes de tétraploïdisation amènent à une cellule à la génomique instable qui, au cours des mitoses, génèrent des cellules aneuploïdes et augmentent l'hétérogénéité génomique intercellulaire (Cowell and Wigley, 1980; Storchová *et al.*, 2006; Wangsa *et al.*, 2018). L'augmentation de la ploïdie est corrélée avec le pronostic, la taille de la tumeur et le grade dans les LMS (Tsushima *et al.*, 1987).

II. Sélection des clones ayant les altérations les plus oncogéniques

Les cellules ayant accumulé des altérations dans des gènes essentiels sont vouées à mourir ou à devenir sénescents (Martincorena *et al.*, 2017). Par contre, les autres seront soumises à de la sélection positive. Cela signifie que les cellules ayant acquis des altérations *drivers* leur permettant de mieux survivre et/ou de mieux proliférer prennent le dessus sur les autres.

1. oLMS, LMS utérins et sarcomes pléomorphes : la perte d'ATRX, un *driver* important

Dans les LMS utérins, les oLMS et les sarcomes pléomorphes peu différenciés l'une de ces altérations *drivers* apportant un ou plusieurs avantages aux cellules cancéreuses est la perte d'ATRX, présente dans environ un quart des tumeurs.

L'effet le plus inattendu de cette perte dans les sarcomes pléomorphes est la diminution de la présence ou de l'activation des cellules du système immunitaire. Nos données sur les tumeurs humaines montrent une non-activation des lymphocytes et des cellules T plus particulièrement. Mes analyses *in vivo* suggèrent que cette non-activation est liée à une limitation du recrutement des mastocytes. Depuis longtemps les sarcomes pléomorphes et en particulier les LMS sont connus pour comporter peu de cellules immunitaires. Les populations de cellules immunitaires les plus présentes dans ces tumeurs sont les lymphocytes T et les macrophages (D'Angelo *et al.*, 2018). La régulation du recrutement des mastocytes, même en nombre restreint, pourrait être à l'origine du non recrutement des autres cellules du système immunitaire. Les mastocytes attirent les autres cellules immunitaires et aident à la maturation des cellules dendritiques en produisant plusieurs chimio-attractants (tels que TNF- α , MMP-9, Cox-2, IL-6, iNOS et VEGF). Les mastocytes inhibent aussi la croissance des cellules tumorales et sont directement cytotoxiques *via* les molécules IL-1, IL-4, IL-8, IL-6, MCP-3, MCP-4, TNF- α , IFN- γ , LTB₄, TGF- β et la chymase (Varricchi *et al.*, 2017; Komi and Redegeld, 2019). Cependant, la présence de mastocytes n'est pas toujours bénéfique pour la tumeur. Ces cellules, lorsqu'elles sont activées, permettent l'angiogenèse grâce à la production de VEGF, IL-8, FGF-2, de l'héparine, de la tryptase et de la chymase, et permettent le remodelage des tissus et la prolifération cellulaire *via* MMP9 et MMP2. Les mastocytes libèrent aussi de l'IL-10, de l'histamine et TNF- α qui participent à l'immunosuppression. Ces différents rôles sont dépendants du type de tumeur, de la micro-localisation des mastocytes dans la tumeur et des molécules activatrices présentes dans le microenvironnement (Varricchi *et al.*, 2017). Aujourd'hui, il n'y a pas, dans la littérature, d'étude portant sur le rôle des mastocytes dans les LMS et plus largement dans les sarcomes, ce qui nous empêche de conclure avec une

certitude absolue quant à leur rôle en absence ou en présence d'ATRX.

Néanmoins, l'augmentation de la prise tumorale quand ATRX est sous-exprimé et l'expression de *FcεRI* et de *Ms4a2* dans les tumeurs exprimant ATRX indiquent que la voie d'activation des mastocytes est celle des IgE, puisque ces deux gènes en font partie. L'activation des mastocytes par cette voie leur confère un rôle protecteur contre les tumeurs épithéliales (Crawford *et al.*, 2018). Cela permet de supposer que les mastocytes, présents dans le microenvironnement des sarcomes pléomorphes lorsqu'ATRX est exprimé, sont anti-tumoraux et capables de recruter d'autres cellules immunitaires. L'absence d'ATRX, de par sa fonction de modification de la chromatine accessible, modifie l'expression de certains gènes et peut entraîner, par exemple, la diminution de l'expression de la voie de signalisation de SCF (Liang *et al.*, 2019). Cette protéine est connue pour recruter et activer les mastocytes anti-tumoraux (Varricchi *et al.*, 2017). Dans la cohorte de LMS, en l'absence d'ATRX, SCF est sous-exprimé, sans atteindre la significativité après correction, cela pourrait donc être à l'origine de la limitation du recrutement des mastocytes qui ne pourront donc plus jouer leur rôle anti-tumoral.

Par ailleurs, notre étude sur l'effet de cette perte, dans certains oLMS et sarcomes pléomorphes, montre aussi que celle-ci permet aux cellules d'être plus agressives en proliférant plus rapidement. Cet effet et ceux sur l'immunité sont indépendants du statut ALT, puisque dans les modèles cellulaires utilisés, la sous-expression d'ATRX n'entraîne pas la mise en place du système ALT. Alors comment la perte d'ATRX peut-elle modifier la biologie de la cellule de sarcome indépendamment du système ALT et des altérations génomiques que cette perte provoque ? La réponse se trouve dans le rôle physiologique primaire d'ATRX : le maintien de l'hétérochromatine constitutive et spécifique ainsi que la régulation de facteurs de transcription ZNF (Baumann and Fuente, 2009; Valle-García *et al.*, 2016; Liang *et al.*, 2019). Dans les gliomes, la perte d'ATRX est liée à des changements profonds dans l'accessibilité de la chromatine modifiant l'expression de sets complets de gènes. Dans ce contexte, ces changements entraînent l'augmentation de la motilité des cellules et diminue la différenciation neuronale (Danussi *et al.*, 2018). Dans le

contexte des LMS, on peut émettre l'hypothèse que la perte d'ATRX joue un rôle indépendant du système ALT en changeant l'accessibilité de la chromatine, augmentant l'agressivité des cellules de sarcome pléomorphe ainsi que leur capacité à échapper au système immunitaire.

La perte d'ATRX se retrouve aussi dans 75% des LMS utérins comme nous avons pu le démontrer. Cette perte est déjà connue comme différenciant les LMS utérins des léiomyomes (Ahvenainen *et al.*, 2018). Ces derniers sont des tumeurs bénignes qui touchent environ 75% des femmes de 50 ans (Bulun, 2013) et bien que les LMS utérins soient rares, ils sont très agressifs. Il est donc nécessaire de pouvoir identifier les lésions classées comme léiomyomes qui pourraient, en réalité, être malignes. Dans l'étude d'Ahvenainen *et al.*, plusieurs des patientes diagnostiquées avec des léiomyomes ont développé une tumeur de type LMS utérin au niveau des poumons quelques temps après l'exérèse de la première lésion. Cette dernière comportait une perte d'ATRX (Ahvenainen *et al.*, 2018), confirmant que les tumeurs utérines avec une perte d'ATRX sont des LMS. L'altération de ce gène dans les tumeurs utérines est probablement un élément important de dédifférenciation des cellules du myomètre semblable à la dédifférenciation neuronale des gliomes perdant ATRX (Danussi *et al.*, 2018). Cette perte est aussi un élément important d'acquisition de l'instabilité génomique. L'absence d'ATRX permet à la fois d'augmenter les cassures chromosomiques au niveau des sites fragiles communs (Pladevall-Morera *et al.*, 2019) et d'aider à la mise en place du système ALT et de son instabilité (Marzec *et al.*, 2015).

2. hLMS : la sur-activation de SRF/MYOCD comme *driver* principal

Dans le cas des hLMS, nous avons suivi l'hypothèse que l'homogénéité des transcriptomes résulte d'un *driver* fort responsable du transcriptome, comme cela a été démontré dans les sarcomes à *driver* fort qui ont un transcriptome homogène (Watson *et al.*, 2018).

Nous avons démontré que ce transcriptome homogène est majoritairement lié à la surexpression des gènes associés au phénotype contractile et au phénotype synthétique des cellules musculaires lisses. L'ensemble de ces gènes sont sous contrôle

du SRF et de ces différents partenaires notamment MYOCD. La surexpression des gènes (p.ex. : *MYH11*, *TAGLN*, *ACTA2*, *CNN1*) et des micro ARN (p.ex. : *mir1*, *mir143*, *mir145*) du phénotype contractile dépendent de la surexpression de MYOCD (Xin *et al.*, 2009; Chen Jie *et al.*, 2011; Pérot *et al.*, 2014). Effectivement, MYOCD amène SRF à se fixer aux cibles spécifiques dans les promoteurs des gènes et des micro ARN liés au phénotype contractile (**Figure 40**). Dans le même temps, la surexpression protéique de SRF et l'absence de PTEN, autorisant une liberté de liaison de SRF à d'autres sites que ceux associés au phénotype contractile, entraînent la surexpression de gènes (p.ex. : *MSN*, *FN1*, *COL1A1/2*, *TNC*) et de micro ARN (p.ex. : *mir28-3p*, *mir455*, *mir199a*) affiliés au phénotype synthétique (**Figure 40**). La co-expression des gènes des deux phénotypes permet la prolifération des cellules de hLMS tout en étant bien différenciées.

Parmi les GO significativement enrichies, associées aux gènes surexprimés dans les hLMS par rapport aux oLMS, on retrouve deux voies métaboliques : la production d'ATP par la mitochondrie et la glycolyse. Dans les SMC vasculaires, le *switch* phénotypique physiologique est associé à un *switch* métabolique. En effet, les SMC vasculaires synthétiques dépendent préférentiellement de la glycolyse pour produire leur ATP, alors que les SMC vasculaires contractiles dépendent plus de la production d'ATP par les mitochondries (Shi *et al.*, 2020). La présence des gènes des deux phénotypes dans les hLMS leur permettrait donc d'augmenter leur production énergétique en multipliant les voies de production d'ATP. La glycolyse facilite aussi la production de nucléotides et d'acides aminés nécessaires à une prolifération importante (Hanahan and Weinberg, 2011).

Dans ce contexte, l'expression des gènes soumis au contrôle de SRF est nécessaire à la survie de la cellule. Ces gènes sont par ailleurs protégés puisqu'ils sont significativement moins cassés dans les LMS que si les cassures arrivaient par hasard (données de l'équipe non montrées). A un autre niveau, l'expression en ARN et protéique de la grande isoforme de DMD n'est pas impactée par des cassures intragéniques de ce gène très long dans les hLMS, alors qu'elle est très majoritairement perdue dans les autres sarcomes myogéniques y compris les oLMS (Wang *et al.*, 2014),

probablement à cause de conflits entre réplication et transcription qui induisent des cassures (Merle *et al.*, 2020). Or, la grande isoforme de DMD est transcrite dans les hLMS donc potentiellement soumise à ces conflits. Par conséquent, la perte d'un ou plusieurs gènes régulés par SRF/MYOCD *via* un mécanisme génomique amènerait probablement un avantage à la cellule. La nécessité de la présence de ces gènes est validée par la mort des cellules de hLMS lors de l'inhibition de SRF/MYOCD. Le rôle exact de la nécessité de la présence de DMD à la membrane ainsi que de l'intégrité du cytosquelette dans l'oncogenèse des hLMS reste à déterminer. Deux hypothèses du rôle de ce cytosquelette peuvent être proposées : maintenir le bon fonctionnement des canaux calciques puisque l'afflux de calcium peut être à l'origine de la limitation de l'apoptose et augmenter la prolifération (Marchi and Pinton, 2016) et/ou maintenir l'intégrité du cytosquelette permettant la motilité cellulaire, au vu des liens décrits entre cytosquelette, motilité et capacité métastatique (Blanchoin *et al.*, 2014) et des liens entre surexpression de MYOCD et migration cellulaire dans les LMS (Pérot *et al.*, 2009).

L'ensemble de ces capacités prolifératives et métaboliques repose sur la surexpression de *MYOCD* et de son partenaire SRF. Il est possible que la perte de PTEN joue un rôle dans la surexpression de *MYOCD*. Sa perte dans des LMS bien différenciés et qui sont des hLMS (la quasi-totalité des LMS de la cohorte de Gibault *et al.* sont ceux utilisés dans la cohorte Affymetrix pour déterminer la signature h/oLMS), est associée à l'activation constante d'AKT (Gibault *et al.*, 2012; The Cancer Genome Atlas Research Network *et al.*, 2017). En effet, PTEN est le répresseur principal d'AKT, son absence entraîne une activation de l'ensemble des voies liées à l'activation d'AKT. L'une des voies inhibées par cette activation est celle des protéines de la famille FoxO. FoxO3a, répresseur de la transcription du gène *MYOCD*, et FoxO4, un répresseur de l'activité de la protéine MYOCD, sont tous les deux inhibés par AKT. Ces inhibitions se font par phosphorylation empêchant l'entrée dans le noyau de FoxO3a ou exportant dans le cytoplasme FoxO4 (Liu *et al.*, 2005; Yang *et al.*, 2013). La perte nucléaire de PTEN permet donc l'expression de gènes pro-prolifératifs et sa perte cytoplasmique permet l'absence d'inhibition de la transcription et de l'activité de MYOCD donc de l'expression de gènes pro-différenciation (**Figure 40**).

B. Comment la différenciation de la cellule d'origine détermine-t-elle l'oncogenèse ?

L'intégration des OMICS avec les différents groupes de LMS permet de mieux comprendre les altérations nécessaires ou aidantes de leur oncogenèse. Ces analyses ont aussi participé à déterminer qu'il existe au moins deux types de cellule d'origine différente au sein des LMS. Les hLMS proviennent de cellules musculaires lisses vasculaires, alors que les oLMS pourraient venir de cellules moins différenciées ou dédifférenciées. Cette différence pourrait être augmentée par la perte d'ATRX ou nécessaire pour que la perte d'ATRX apporte des avantages à la cellule cancéreuse. Dans des cellules progénitrices des hLMS, il est très probable que cette perte protéique d'ATRX arrive au hasard des altérations. Cependant, elle n'apporterait pas les mêmes avantages aux cellules tumorales et ne sera donc pas positivement sélectionnée au cours du développement tumoral.

A l'inverse le taux d'altération d'ATRX de 75% dans les LMS utérins en fait une altération importante dans ce sous-groupe. Ces LMS ont un profil transcriptomique distinct des oLMS et des hLMS. En effet lorsqu'ils sont introduits dans les analyses non supervisées de l'expression de l'ARN, ils forment un groupe spécifique avec une transcriptomique mixte entre les deux autres groupes (données non montrées de l'équipe; Hemming *et al.*, 2020). Au vu de la localisation anatomique de ces LMS, la cellule probablement progénitrice de ce sous-groupe est une cellule du myomètre, le muscle de l'utérus composé de cellules musculaires lisses. Ces cellules sont elles aussi très plastiques et soumises au *switch* phénotypique, mais celui-ci est déclenché par des voies spécifiques principalement liées aux hormones de la grossesse (Shynlova *et al.*, 2009; Tanfin *et al.*, 2011). Dans ce contexte cellulaire, la perte d'ATRX permet aux cellules d'être moins différenciées et de proliférer plus rapidement. Cette même altération d'ATRX dans un autre type de cellule semble avoir d'autres impacts comme dans les autres sarcomes pléomorphes peu différenciés, où la perte d'ATRX a aussi un effet sur l'échappement au système immunitaire. Le contexte cellulaire étant différent les modifications liées à l'absence d'ATRX entre euchromatine et hétérochromatine n'auront pas des conséquences identiques. Une même altération *driver* nécessaire à

l'oncogenèse aura donc des rôles différents selon la cellule d'origine.

Lors de l'étude des différences entre oLMS et hLMS, nous avons montré que les oLMS ont des marques d'histones de fibroblastes et de cellules souches mésenchymateuses mais aussi qu'ils expriment des gènes de la transition épithéliomésenchymateuse. Ces différents résultats ne permettent pas de faire une conclusion sur la cellule d'origine exacte des oLMS mais il est possible d'émettre plusieurs hypothèses. Les oLMS pourraient donc provenir de chacune des cellules suivantes : de cellules transdifférenciées, passant d'une différenciation épithélial à une faible différenciation musculaire lisse *via* la transition épithéliomésenchymateuse, de cellules peu différenciées comme les cellules souches mésenchymateuses, de cellules dédifférenciées comme des fibroblastes qui, au grès des mutations aléatoires et de la sélection positive, exprimeraient certains gènes marqueurs de la différenciation musculaire lisse sans jamais être totalement différenciés.

Les oLMS semblent avoir une oncogenèse basée sur un remaniement important du génome, sans qu'il ne soit possible, à ce jour, de dégager d'autres caractéristiques, sûrement en raison d'une grande hétérogénéité dans ce groupe. Néanmoins, lors d'analyses non-supervisées de l'expression des ARN avec de nombreux sarcomes, les oLMS se regroupent avec d'autres sarcomes présentant une très faible différenciation. Ces sarcomes pléomorphes peu différenciés sont individualisés par les anatomopathologistes par l'expression faible ou sporadique de marqueurs de chaque ligne de différenciation ainsi que par quelques clones de cellules morphologiquement plus différenciées dans les tumeurs (Carvalho *et al.*, 2019). Cette catégorie inclut : les LPS pléomorphes, dont la majorité des cellules sont morphologiquement proches des UPS avec quelques lipoblastes ; les RMS pléomorphes, avec peu de cellules présentant des marqueurs musculaires ; les OS extra-squelettiques, morphologiquement très proches des UPS avec des calcifications ; les myxofibrosarcomes et les UPS. Tous ces sous-types présentent des caryotypes très remaniés mais sans altération génomique spécifique (Hornick, 2018; Carvalho *et al.*, 2019). Ces différents sous-types histologiques sont à l'image des oLMS d'un point de vue transcriptomique, morphologique et génomique. Ils pourraient donc avoir les mêmes cellules d'origine

que les oLMS et le même type d'oncogenèse. Une cellule d'origine commune de ces sarcomes pléomorphes pourrait être les cellules souches mésenchymateuses. Cette hypothèse est soutenue par plusieurs études, dans lesquelles différentes altérations dans les cellules souches mésenchymateuses donnent des types de sarcome pléomorphe différents. Dans ces cellules, l'ajout des oncogènes *hTERT*, *SV40* et *H-Ras* ou l'altération de la voie Wnt donne des UPS alors que les délétions de *TP53* et *RB1* donnent des LMS (Matushansky *et al.*, 2007; Li *et al.*, 2009; Rubio *et al.*, 2010). Ne plus étudier ces tumeurs séparément selon une ligne de différenciation faible dans de petites cohortes mais tous ensemble comme une entité de sarcomes pléomorphes à faible différenciation permettrait de déterminer quelles sont la ou les cellule(s) d'origine et de déterminer les différents *drivers* de leur(s) oncogenèse(s) afin de proposer une thérapie efficace.

D'un autre côté, la cellule d'origine des hLMS est plus clairement identifiée après ce travail puisqu'il est très probable qu'il s'agisse des cellules musculaires lisses vasculaires. L'analyse de la transcriptomique homogène de ces tumeurs permet de caractériser la voie suractivée de SRF/MYOCD comme *driver* fort nécessaire à leur oncogenèse. Cette voie détourne la machinerie biologique du *switch* phénotypique déjà présente dans les cellules musculaires lisses afin de garder une certaine différenciation, qui autorise la cellule à produire beaucoup d'énergie et dans le même temps lui accorde une forte capacité proliférative.

Néanmoins, cette capacité existe chez toutes les cellules musculaires lisses, alors pourquoi les hLMS ne proviendraient que des cellules musculaires lisses vasculaires ? Pourquoi n'avons-nous pas des hLMS dont la transcriptomique est proche de celle des cellules musculaires lisses viscérales alors que ces cellules sont très nombreuses ? Tout d'abord, les LMS du tractus gastro-intestinale existent mais les cas sont extrêmement rares et généralement rapportés sous forme de *case report* (Hashimoto, Tsuneyoshi and Enjoji, 1985; Granero Peiró, 2015; Sahli *et al.*, 2016; Ilias *et al.*, 2017). Il est possible de trouver une littérature conséquente concernant les LMS du tractus gastro-intestinal datant d'avant le début des années 2000. C'est à cette période que la découverte de la mutation de *KIT* dans les tumeurs gastro-intestinales (Hirota *et al.*, 1998) a permis de

différencier précisément les GIST des LMS. Depuis, la grande majorité des tumeurs localisées dans le tractus gastro-intestinal se révèle être des GIST (Hilal *et al.*, 2016). Quelques années plus tard, la découverte, dans le tractus gastro-intestinal humain, des cellules interstitielles de Cajal (Komuro, 2006), cellules mésenchymateuses servant de *pacemaker* aux cellules musculaires lisses, n'exprimant pas les marqueurs des muscles lisses mais exprimant KIT et PDGFR α , a permis de définir que ces dernières étaient à l'origine des GIST (Sanders, Ward and Koh, 2014). La rareté clinique explique une absence des LMS gastro-intestinaux dans les cohortes de recherche. Il serait donc intéressant d'étudier le transcriptome de ces LMS rares afin de voir s'ils se rapprochent de celui des hLMS vasculaires.

Une question subsiste cependant : pourquoi une telle rareté par rapport aux hLMS vasculaires ? Les vaisseaux sanguins sont composés de cellules de tout le spectre de la différenciation. Le *switch* y est très actif et survient à de nombreuses occasions : blessure, grossesse, sport, vaisseau bouché, *etc.*, alors qu'il semble moins actif dans le tractus gastro-intestinal mais cette impression pourrait venir d'un manque d'étude de ce phénomène dans les SMC viscérales (Scirocco *et al.*, 2016). Il est connu que dans les SMC viscérales, le microenvironnement micro-biotique module l'expression des gènes de l'hôte et ces changements peuvent entraîner le *switch* (Hooper *et al.*, 2001). L'activation des deux phénotypes passerait par des voies peut-être plus complexes au sein des cellules musculaires lisses viscérales limitant ainsi le risque oncogénique.

Il est possible de se poser la même question à propos des muscles lisses présents sous la peau, les *arrector pili*. La présentation de LMS sous-cutanés est moins rare que celles des LMS gastro-intestinaux mais plus rare que les vasculaires et survient souvent chez des patients d'âge avancé dans des localisations exposées au soleil (Llombart *et al.*, 2019). Ces tumeurs présentent des signatures mutationnelles liées aux UV comme le LMS du scalp de notre cohorte (LMS23). Ce LMS 23 est classé parmi les oLMS par la méthode des centroïdes et est le seul sans altération de *RB1*, l'oncogenèse des LMS cutanés pourrait en partie dépendre des mutations induites par les UV et être fort probablement différente de celles des autres LMS venant de cellule musculaire lisse.

Dans les LMS comme dans l'ensemble des sarcomes et des cancers, l'oncogenèse

dépend donc de la cellule dans laquelle elle se produit. Par exemple, dans les myoblastes exprimant *MYF5* la fusion de *SS18* et *SSX* donnera un synoviosarcome, alors que dans les myoblastes exprimant *MYF6* cette fusion produira une myopathie et pas de tumeur (Haldar *et al.*, 2007). D'après nos résultats, malgré les altérations communes de *TP53* et *RB1*, la sur-activation de *SRF/MYOCD*, dans une SMC vasculaire, est nécessaire à l'oncogenèse des hLMS (Figure 41). Alors que l'oncogenèse des oLMS passerait par l'acquisition de multiples oncogènes tel que la perte d'*ATRX*, dans des cellules d'origine encore indéterminée (Figure 41).

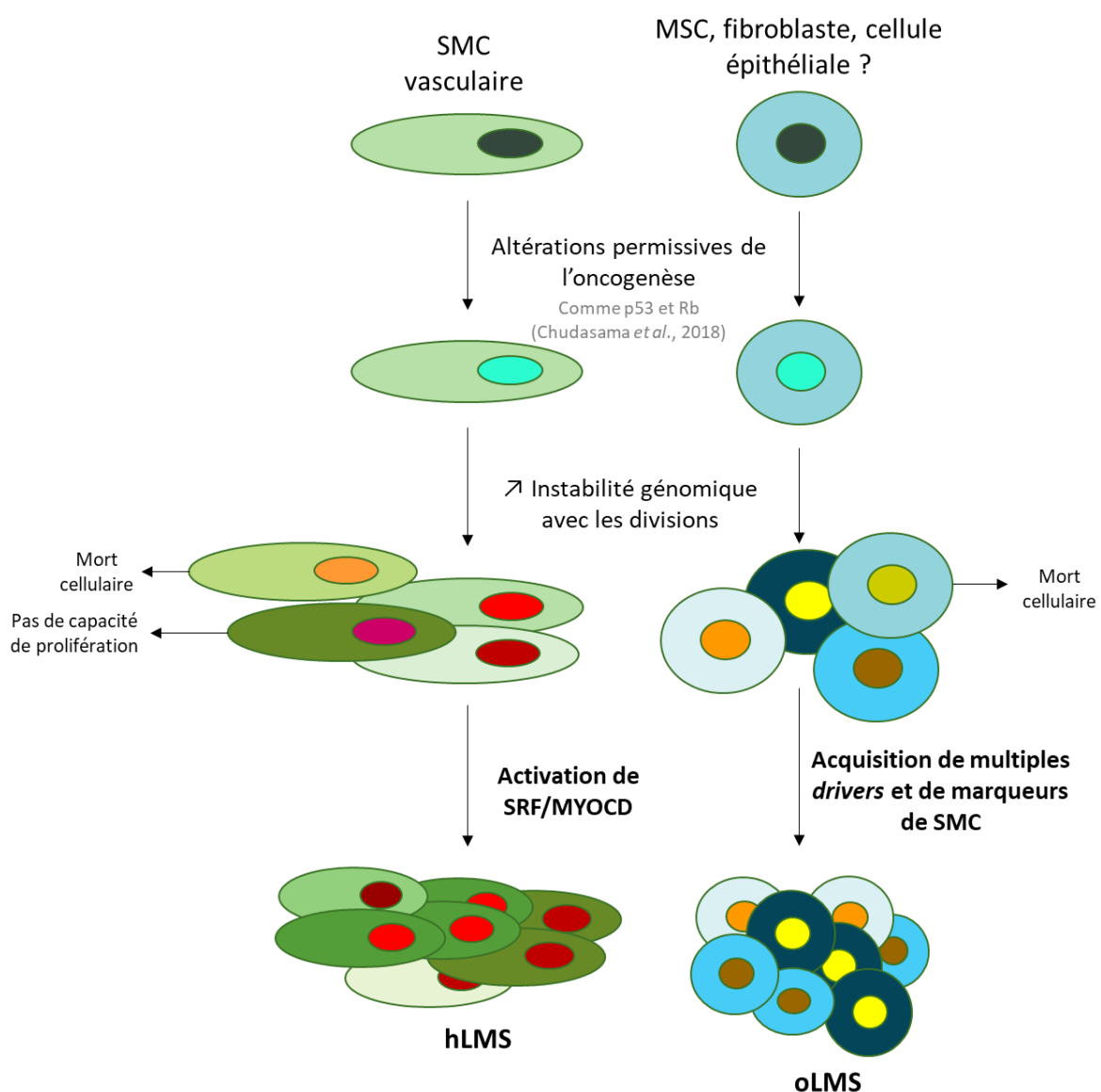


Figure 41 : Modèle d'oncogenèse des hLMS et des oLMS

C. Doit-on proposer des thérapies spécifiques aux altérations des tumeurs ?

Comme expliqué en détail dans le paragraphe « Prises en charge thérapeutiques variées des sarcomes », il existe très peu de traitements adaptés aux sarcomes en général et aux LMS en particulier. L'exérèse chirurgicale avec des marges saines reste le traitement standard (WHO Classification of Tumours Editorial Board, 2020). Mais si la tumeur métastase ou qu'elle n'a été diagnostiquée qu'après avoir déjà métastasé, les probabilités de survie du patient deviennent très faibles (Blay *et al.*, 2017). Le traitement standard actuel d'un LMS métastatique est le même qu'il y a 50 ans : les chimiothérapies à bases d'anthracyclines et notamment la doxorubicine bien que les LMS aient une réponse assez faible (P. G. Casali, Abecassis, *et al.*, 2018b). Depuis quelques années, d'autres chimiothérapies sont venues étoffer la deuxième ligne de traitement comme l'ifosfamide, la gemcitabine et la trabectedine (Serrano and George, 2013). Les connaissances actuelles sur la biologie des LMS ont permis le développement d'inhibiteurs de mTOR, testés suite à la découverte de la perte de PTEN dans de nombreux cas. Néanmoins ces inhibiteurs n'ont pas démontré de preuves cliniques suffisantes pour une utilisation à grande échelle chez les patients (Serrano and George, 2013).

Le travail réalisé dans cette thèse permet de proposer des pistes de traitements adaptées aux spécificités des différentes oncogèneses des LMS.

I. Vers une utilisation clinique des statuts h/oLMS

La signature nous permettant de séparer les deux types de LMS a été une nécessité pour déterminer les voies impliquées dans chaque oncogène et pouvoir proposer des thérapies spécifiques.

La démonstration des différences transcriptomiques des différents sous-types d'un même groupe de tumeur a aussi été le point de départ de grandes avancées thérapeutiques en cancérologie. Dans le cancer du sein, l'identification des différents

sous-types a permis le développement de la thérapie utilisant des anticorps dirigés contre le récepteur codé par *ERBB2* dans le groupe où il est amplifié. Dans le groupe des cancers du sein présentant l'expression des récepteurs à œstrogène une thérapie basée sur les œstrogènes a été développée. La capacité à distinguer les différents sous-groupes est donc à l'origine d'avancées thérapeutiques majeures (Sørli *et al.*, 2001; Waks and Winer, 2019). Dans le cancer du poumon, le traitement dépend des différentes mutations *drivers* retrouvées dans la tumeur. Si une mutation de *EGFR* est présente, le patient sera alors traité avec des inhibiteurs de EGFR, alors que si c'est une fusion impliquant *ALK* qui est retrouvée, il sera utilisé des inhibiteurs des tyrosines kinases (Hirsch *et al.*, 2017). Les différents groupes de cancers du sein et du poumon sont aussi associés à des pronostics plus ou moins favorables, les thérapies ont donc été développées en association avec des chimiothérapies plus ou moins agressives (Hirsch *et al.*, 2017; Waks and Winer, 2019). Déterminer le groupe des LMS dans les prochaines études cliniques améliorerait donc la prise en charge des patients. A la fois au niveau de la spécificité des traitements proposés et au niveau du risque métastatique puisque les hLMS sont de moins bon pronostic. Actuellement la localisation de la tumeur est l'un des seuls facteurs pronostiques utilisés (WHO Classification of Tumours Editorial Board, 2020).

II. Immunothérapie lors de la perte d'ATRAX

La perte d'ATRAX dans les sarcomes pléomorphes peu différenciés, incluant les oLMS, survient dans environ un quart des cas. Nous avons démontré que cette perte est associée à une diminution de la présence ou de l'activation des cellules du système immunitaire notamment des mastocytes, alors que lorsque les mastocytes sont présents, ils semblent avoir un rôle préférentiellement anti-tumoral. Dans les cas présentant la perte d'ATRAX et comme les mastocytes font partie de l'immunité innée, nous proposons de tester une nouvelle forme d'immunothérapie dans les sarcomes qui ne soit pas dirigée vers l'immunité acquise mais pour aider l'immunité innée. En effet, la majorité des immunothérapies jusqu'alors testées dans les sarcomes sont basées sur la modulation du système immunitaire acquis. En théorie, bloquer les

récepteurs tumoraux à l'origine de l'absence de reconnaissance des cellules tumorales par les cellules T permet à ces dernières de reconnaître la tumeur et de devenir cytotoxique (Ayodele and Abdul Razak, 2020). Mais les drogues associées à cette théorie ne donnent pas les effets escomptés sur les sarcomes en clinique. Dans un certain nombre de cancers, les thérapies à base d'immunoglobulines, qui viennent activer les cellules immunitaires dans le microenvironnement tumoral, ont fait leurs preuves et sont de plus en plus utilisées. Plus spécifique du système immunitaire inné, l'utilisation d'IgE a été validée par des modèles *in vitro* et *in vivo* et nécessite maintenant des tests chez l'humain. Cette technique permet l'activation par les récepteurs FcεRI des cellules effectrices des IgE, comprenant les mastocytes, les macrophages et les monocytes (Leoh, Daniels-Wells and Penichet, 2015). Couplée à des chimioattractants, elle permet aussi le recrutement de ces cellules effectrices spécifiques (Jensen-Jarolim *et al.*, 2008).

Le développement d'immunothérapie utilisant des IgE couplée à des chimioattractants dans les sarcomes pléomorphes peu différenciés permettrait, lors de la perte d'ATRAX, de pallier l'absence des mastocytes et de réactiver ainsi le système immunitaire.

III. Une thérapie prometteuse des hLMS : inhibition de la voie SRF/MYOCD

Dans le contexte des hLMS, la connaissance de l'oncogenèse permet de proposer l'inhibition de l'interaction entre SRF et MYOCD comme thérapie ciblée.

Jusqu'alors d'autres thérapies ciblées ont été proposées notamment l'inhibition de la voie Pi3K/AKT/PTEN, soit par une inhibition de mTOR mais qui n'est pas efficace en clinique, soit par une double inhibition de Pi3K et de mTOR. Cette thérapie a eu des résultats précliniques *in vivo* en demi-teinte. En effet, elle a d'abord été testée sur cinq PDX de LMS. Pour 4 des 5 souris, les tumeurs ne progressaient plus et deux ont même régressé. Ces quatre PDX, répondant à la double inhibition, n'exprimaient pas PTEN contrairement à la cinquième PDX (Cuppens *et al.*, 2017), il est donc plus

probable que les PDX répondeuses soient des hLMS. Cependant, dans une seconde étude, cette double inhibition dans des lignées cellulaires de LMS, présentant les caractéristiques des hLMS, a entraîné, malgré une baisse de la prolifération des tumeurs *in vivo*, une résistance par l'activation de la voie KRAS/MEK/ERK, antagoniste de la MYOCD (Fourneaux *et al.*, 2016). Cibler en aval de la voie Pi3K/AKT/PTEN en inhibant directement l'interaction de SRF/MYOCD éviterait la mise en place de cette résistance.

L'inhibiteur de SRF/MYOCD n'ayant pas été développé pour une utilisation chez l'humain, comprendre son fonctionnement dans les hLMS faciliterait son développement clinique. Il pourrait même être possible de cibler des voies elles-mêmes inhibées en l'absence de l'interaction SRF/MYOCD et ayant, possiblement, une drogue dirigée contre elles déjà utilisée en clinique. Par exemple, une des cibles de SRF, PPP2CA, enzyme intervenant dans le métabolisme du glycogène, peut être inhibée par de la Cantharidine. Cette substance est produite par les insectes de la famille des Méloïdés et est utilisée en dermatologie pour traiter le *molluscum contagiosum*. C'est une lésion cutanée due à un virus ayant des effets sur le cycle cellulaire, l'inflammation et la mort cellulaire (Chen, Anstey and Bugert, 2013). Dans les SMC vasculaires, cette drogue a pour effet de diminuer la prolifération et la migration (Qiu *et al.*, 2019). En utilisant ce genre de produit, il serait alors possible d'améliorer le traitement et d'accélérer sa disponibilité pour les patients.

D. Perspectives

Les travaux réalisés au cours de ma thèse ont permis de valider la séparation des LMS en trois maladies distinctes : les LMS utérins, les LMS homogènes provenant des cellules musculaires lisses vasculaires et les autres LMS, peu différenciés et dont l'origine exacte reste à déterminer. Ces trois groupes ont des oncogènes différentes bien qu'ils aient en commun les altérations de *TP53* et de *RB1* qui, d'après une autre étude, arriveraient de façon précoce et en amont des événements de tétraploïdisation (Chudasama *et al.*, 2018). Les LMS utérins, de leur côté, se développent dans la majorité des cas avec l'absence d'ATRAX.

Les oLMS, comme les autres sarcomes pléomorphes peu différenciés, acquièrent de multiples *drivers* avec, dans un quart des cas, la perte d'ATRAX permettant aux cellules de proliférer plus rapidement et d'échapper au système immunitaire. Le non recrutement des mastocytes au sein de la tumeur pourrait être le point de départ de cet échappement. En effet, nous avons montré une baisse significative de la quantité de mastocyte dans les tumeurs en absence d'ATRAX. Pour en comprendre la portée et pouvoir ainsi proposer un traitement adapté, il faut maintenant déchiffrer le rôle exact des mastocytes en présence et en absence d'ATRAX. Dans les mêmes contextes, déterminer comment se répercute l'absence des mastocytes sur le reste du système immunitaire, et spécifiquement les lymphocytes et les macrophages, principales cellules immunitaires présentes dans les sarcomes, devient alors nécessaire. En faisant du séquençage ARN en cellule unique, on pourrait déterminer dans une même tumeur de sarcome pléomorphe avec ou sans altération d'ATRAX, les voies activées par chaque type de cellule immunitaire. Cette technique est en cours de mise au point au sein du laboratoire. Puis il sera nécessaire de mettre en place des modèles cellulaires, à partir de ces données, afin de déterminer comment chaque type de cellule immunitaire interagit entre elles et avec la tumeur dans des contextes tumoraux avec et sans ATRAX. Dans le même temps et pour améliorer l'immunothérapie développée, il faudrait cartographier avec précision les changements, liés à la perte d'ATRAX, dans les voies cellulaires qui limitent l'activation du système immunitaire. Pour cela, il faudrait associer plusieurs éléments : les

modifications de l'accessibilité à la chromatine en présence et absence d'ATRX, grâce à la technique ATAC-seq, les changements d'expression des gènes selon le statut d'ATRX, *via* du séquençage ARN, et la connaissance des points de liaison d'ATRX à l'ADN quand il est présent avec de la ChIPseq. L'ensemble de ces éléments permettront de comprendre les modifications dans le génome associées à la perte d'ATRX et leurs conséquences.

Comme les différents sarcomes pléomorphes à différenciation très faible se regroupent lors des analyses non supervisées de l'expression des gènes, les étudier en détails comme une seule entité serait utile pour déterminer leur oncogénèse. On pourra alors vérifier l'hypothèse que l'expression des différents marqueurs de différenciation est le résultat d'évènements génomiques qui ne donnent pas un avantage fort aux cellules, puisque les clones les ayant ne prennent pas le dessus, et ne les désavantagent pas non plus, puisqu'elles restent en vie. Une analyse, à l'image de celle que nous avons menée sur les LMS, intégrant les différents OMICS aux données cliniques des sarcomes pléomorphes adjoint à des validations fonctionnelles apporterait les réponses nécessaires pour connaître leur(s) cellule(s) d'origine ainsi que leurs *drivers* principaux, en plus des altérations de *TP53* et *RB1*. Ces sarcomes n'ayant, actuellement, aucune option de traitement ciblé, cela ouvrirait la voie au développement d'options de traitements ciblés.

D'un autre côté, les hLMS, par différents mécanismes, sur-activent la voie de SRF/MYOCD permettant l'expression des gènes des deux phénotypes : contractile et prolifératif. Les mécanismes en aval de cette sur-activation restent à déterminer afin de comprendre et de caractériser la totalité de l'oncogénèse des hLMS. La mise au point soit d'un anticorps spécifique de MYOCD, soit d'une technique permettant de marquer MYOCD est une priorité. La technique utilisant le CRISPR-Cas9 et la recombinaison homologue pourrait être efficace en modifiant soit les bras d'homologie, soit l'ARN guide, soit en se plaçant au niveau 3'UTR du gène, tests que je n'ai malheureusement pas eu le temps de réaliser. Ceci permettrait d'aller au bout de l'analyse des changements de cible du complexe SRF/MYOCD et de MYOCD seul entre SMC vasculaire, hLMS et oLMS. Néanmoins, dans les lignées tumorales

présentant un gain ou une amplification de *MYOCD*, l'établissement de tels marquages sur chaque copie sera très probablement impossible. Mais même un marquage partiel de *MYOCD* devrait apporter les réponses attendues. De plus, comme la liaison de facteur de transcription comme SRF/*MYOCD* peut modifier la structure de la chromatine à petite échelle (McDonald *et al.*, 2006), on peut se poser la question suivante : la surexpression de SRF et de *MYOCD* dans les hLMS peut-elle modifier l'organisation de la chromatine et être à l'origine d'altérations spécifiques des hLMS ? Pour y répondre, il pourrait être utilisé, dans des lignées cellulaires de SMC vasculaires et de hLMS, la technique de capture de conformation chromosomique Hi-C dans laquelle les anticorps spécifiques de SRF et de *MYOCD* permettraient de déterminer les régions chromosomiques rapprochées par la surexpression de ces protéines. La comparaison de ces rapprochements aux cassures génomiques retrouvées dans les LMS permettrait de déterminer si SRF/*MYOCD* a aussi un rôle dans les altérations génomiques, en plus de celui dans la modification des profils transcriptomiques vers l'oncogenèse des hLMS.

La confirmation du rôle de *driver* fort de la voie SRF/*MYOCD* par son inhibition permet de proposer un nouveau traitement ciblé et spécifique des hLMS. Ces résultats sont des données *in vitro* qui nécessitent une validation *in vivo*. Cela devra passer par une expérimentation de xénogreffe sous-cutanée de la lignée cellulaire de hLMS. Puis par un test sur des PDX provenant de tumeurs de hLMS afin d'être au plus près de la biologie tumorale humaine. Les validations précliniques de cette inhibition sont une étape indispensable à un développement clinique.

Une fois que le développement clinique de cet inhibiteur commencera, il sera nécessaire de pouvoir proposer une technique en routine qui déterminera le statut des LMS. Une technique objective pour ce faire serait de diminuer la signature aux 50 à 100 gènes les plus discriminants et les utiliser sur une puce à ARN de type NanoString®. Cette technique est déjà utilisée en routine pour la PAM 50 par exemple, mais nécessiterait de la mise au point pour la réduire. Or, on sait que les statuts hLMS et oLMS corrélaient à 87,5% avec le score associé à une combinaison de quatre marquages immunohistochimiques : l' α -actine musculaire lisse, la caldesmone, la

calponine et la transg line (P rot *et al.*, 2014). Ce score est un outil simple   mettre en place et imm diatement utilisable en routine.

Par la suite, le traitement par inhibiteur de SRF/MYOCD pourrait  tre test  sur les autres types de LMS comme les cutan s, gastro-intestinaux ou de l'ut rus. Afin de comprendre comment cette inhibition agit dans des cellules avec un fonctionnement diff rent, il sera int ressant de confronter les profils transcriptomiques de ces tumeurs pour comprendre les points communs et les diff rences avec les hLMS. Dans notre cohorte, l'oncog nese dans les cellules musculaires lisses vasculaires est li e   un d tournement du *switch* ph notypique. On peut donc  mettre l'hypoth se que dans les autres types de cellule musculaire lisse ce *switch* est aussi d tourn  vers une oncog nese. Cependant, on peut voir, dans notre cohorte, que les LMS de l'ut rus sont diff rents des hLMS puisque leurs profils transcriptomiques sont distincts de ceux des hLMS. Chaque type de cellule musculaire lisse pourrait donc avoir une oncog nese d pendante des sp cificit s du *switch* ph notypique de la cellule. Dans les cellules du myom tre, les changements hormonaux y ont un r le pr pond rant et le pic d'incidence des LMS ut rins se situent aux alentours de 50 ans vers la p ri-m nopause des femmes (Miettinen, 2016). Cette p riode de changements hormonaux est associ e   des cellules vieillissantes ayant un taux d'alt ration somatique qui augmente (Martincorena and Campbell, 2015). Cette association pourrait  tre   l'origine d'un *switch* ph notypique qui n'est plus contr l  et ainsi de l'oncog nese des LMS de l'ut rus. Concernant les cellules du tractus gastro-intestinal, les voies de la mise en place du *switch* semblent plus contr l es et moins activ es que dans les cellules musculaires lisses vasculaires (Scirocco *et al.*, 2016), expliquant  ventuellement la raret  de ces cas. Les LMS ont un pic d'incidence entre 70 et 80 ans,   cet  ge, le taux de mutations accumul es dans les cellules non transform es est donc  lev  (Martincorena and Campbell, 2015). La probabilit  qu'une combinaison d'alt ration permissive d'une oncog nese et limitant le contr le du *switch* ph notypique soit pr sente dans une m me cellule musculaire lisse est plus  lev e. Dans ce contexte, un *switch* forc  par un changement physiologique, comme une inflammation aigu e ou chronique, m canique, comme la perte du tonus musculaire, ou encore micro-biotique (Scirocco *et al.*, 2016) pourrait  tre le point de d part de l'oncog nese de ces LMS. Les

LMS sous-cutanés, quant à eux, sont retrouvés dans des localisations exposées aux UV (Llombart *et al.*, 2019). Ce type d'exposition est lié à une augmentation du taux de mutation dans les gènes actifs (Supek and Lehner, 2017). Les gènes responsables du contrôle du *switch* mais aussi de la prolifération et de l'apoptose seraient donc plus altérés et entraîneraient une oncogenèse lors de l'activation du *switch*. Notre compréhension de l'oncogenèse des hLMS, suite à ce travail de thèse, ainsi que ces différentes hypothèses placent le détournement du *switch* phénotypique, caractéristique des cellules musculaires lisses, au centre de leur oncogenèse.

Partie 4 : Matériels et Méthodes

La sur-activation de la voie SRF/MYOCD peut-elle être le driver des hLMS ?

Immunohistochimie

Les IHC ont été réalisées sur TMA (tissu microarray) préalablement préparés sur la cohorte de LMS. La détection de PTEN et de SRF a été réalisée sur BenchMark Ultra (Ventana, Washington D.C, USA). Le démasquage antigénique a été fait en utilisant le programme « CC1 protocol » pendant 4 min à 100°C (Ventana, Washington D.C, USA), l'anticorps dirigé contre PTEN était dilué dans du Prepkits26 et incubé 1h (1:200, Clone 138G6, Cell signaling technology, Danvers, MA, USA), de même pour celui dirigé contre SRF (1:500, A303-172A, Bethyl Laboratories, Montgomery, TX, USA). La détection des anticorps a été faite en utilisant le kit Optiview durant 12 min (Ventana, Washington D.C, USA). Les images d'immunohistochimie ont été prises par le scanner de lames 250 Flash II Digital et analysées avec le logiciel Panoramic Viewer (3DHISTECH Ltd., Budapest, Hungary).

Immunofluorescence sur bloc FFPE

Des sections de 4µm d'épaisseur ont été réalisées sur les blocs de TMA puis déparaffinées et réhydratées dans des bains successifs d'éthanol de 100% à 70%. Après démasquage antigénique par la chaleur pendant 20 min (pH6, S236984-2, DAKO, Agilent Technologies, Santa Clara, CA, USA), les lames ont été incubées 10 min avec le tampon de saturation (5% BSA/PBS 1X). Les lames ont ensuite été incubées avec les anticorps primaires à 4°C, en chambre humide, sur la nuit : anti-Dp427 (1:25, NCL-DYS1, Novacastra, Leica, Wetzlar, Germany) et anti-dystrophine (1:100, ab15277, Abcam, Cambridge, UK). Après lavage avec du PBS 1X, les lames ont été incubées, en chambre humide avec les anticorps secondaires pendant 1h à température ambiante : AlexaFluor 488 anti-*mouse* (1:400, A11001, Thermo Fisher Scientific, Waltham, MA, USA) et AlexaFluor 594 anti-*rabbit* (1:400, A-11072, Thermo Fisher Scientific, Waltham, MA, USA). Pour finir, les lames ont été montées en vectashield/DAPI (H-1200-10, Vector Laboratories, Burlingame, California CA, USA). L'analyse et les images ont été faites sur le vidéomicroscope Cell Observer (Carl Zeiss, Oberkochen, Germany).

Culture cellulaire

Les lignées cellulaires de LMS et d'UPS ont été cultivées dans du RPMI-1640 (524000-025, Life Technologies, Rockville, MD, Carlsbad, CA, USA) contenant 10% de sérum de veau foetal. La lignée cellulaire de SMC vasculaire a été cultivée dans du DMEM (31966021, Life Technologies, Rockville, MD, Carlsbad, CA, USA) contenant 10% de sérum de veau foetal. Toutes les lignées ont été gardées à 37°C dans un incubateur maintenu à 5% de CO₂.

RT-PCR quantitative de MYOCD

Les ARNm ont été extraits par un protocole standard TRIzol (15596026, Thermo Fisher Scientific, Waltham, MA, USA) / chloroforme (32211-1L, Fisher Scientific, Hampton, NH, USA) suivie par une précipitation avec de l'éthanol 100% et une purification avec le kit RNeasy Mini (74104, Qiagen, Hilden, Germany) avec un traitement à la DNase (79254, RNase-Free DNase Set, Qiagen, Hilden, Germany). La rétrotranscription de 1µg d'ARN messagers a été effectuée avec des hexamères aléatoires et le kit *High-Capacity cDNA Reverse Transcription* (4368814, Applied Biosystems, Foster City, CA, USA) en suivant les recommandations du fournisseur. Les *primers* ont été sélectionnés en utilisant le programme Primer 3 (<http://frodo.wi.mit.edu/primer3/>): *forward* : CACCCAACAACCCTCACTTT et *reverse* : TTGGGGATGGAATTGAAAAC. La qPCR a été réalisée sur 1µL dilué au 1/5 d'ADNc en utilisant le *master mix SYBR Green* (4368577, Applied Biosystem, Foster City, CA, USA). Le programme utilisé sur le thermocycleur StepOnePlus (Applied Biosystem) était : 10 minutes 95°C puis 40 cycles : 15 secondes à 95°C et 1 minute à 60°C, une *melt curve* a été réalisée. Les *fold change* entre l'expression de chacun des échantillons par rapport à OC80 (contrôle négatif) ont été calculés par la méthode des $\Delta\Delta C_t$ (Livak and Schmittgen, 2001) avec comme contrôle l'expression du gène de la β -actine.

Western Blotting

Pour l'extraction de protéines, les cellules sont récupérées par grattage dans une solution de RIPA (89901, Thermo Fisher Scientific, Waltham, MA, USA) et d'un cocktail d'inhibiteurs de protéases et de phosphatases (1X, P8340, Sigma, St Louis, MO, USA). Les protéines sont dosées et diluées afin d'obtenir une quantité de 15 ng. Après migration à 100V durant 1h, les protéines ont été transférées sur membrane PVDF par transfert rapide (iBlot2, Thermo Fisher Scientific, Waltham, MA, USA). Après saturation dans du PBS 1X/5% Lait, les membranes ont été incubées à 4°C toute la nuit avec les anticorps primaires suivants : anti-MYOCD (1:1000, SAB4200539, Sigma, St Louis, MO, USA), anti-MYOCD (1:1000, ab107301, Abcam, Cambridge, UK), anti-MYOCD (1:1000, MAB4028, R&D System, Minneapolis, MI, USA), anti-MYOCD (ABIN204415, Antibodies-Online, Atlanta, GA, USA), anti-FLAG (1:500, 14793S, Cell Signaling Technologies, Danvers, MA, USA) et anti- β -actine (1:5000, A200066, Sigma, St Louis, MO, USA). Après lavages au PBS 1X, les membranes ont été incubées avec les anticorps secondaires pendant 1h à température ambiante : anti-*rabbit* et anti-*mouse* HRP (1:5000, 7074S et 7076S, Cell Signaling Technologies, Danvers, MA, USA). Les résultats sont analysés sur le Pxi (SynGene, India).

Préparation des lignées Tag-FLAG de MYOCD

Le vecteur contenant la Cas9, résistant à la néomycine, vient de chez Tebu-Bio (CP-LvC9NU-01, Tebu-Bio, Le Perray-en-Yvelines, France, **Figure 42 A**). Deux ARN guides ont été sélectionnés avec le logiciel CHOPCHOP (<http://chopchop.cbu.uib.no/>), les séquences choisies sont celles contenant le moins de fausses cibles possibles, au plus près de l'exon 13 de MYOCD (**Figure 42 B**). Ces ARN guides ont été synthétisés par Eurofins Genomics (Ebersberg, Germany) et introduit dans le vecteur résistant à la blasticidine EF1 α -Blasticidin-H1-gRNA (CASLV500PA-B, System biosciences, Palo Alto, CA, USA, **Figure 42 C**) selon les recommandations du fournisseur. Les deux vecteurs ont été transformés dans des bactéries thermocompétantes Top10 (C404003, Invitrogen, Carlsbad, CA, USA) et étalées sur LB agar contenant 50 μ g/mL d'ampicilline. Les clones ont ensuite été

cultivés dans 200mL de LB liquide avec 50 µg/mL d'ampicilline et l'extraction d'ADN plasmidique a été réalisée en suivant le kit *EndoFree Plasmid Maxi* (12362, Qiagen, Hilden, Germany). L'insertion des séquences des ARN guides a été validée par séquençage Sanger réalisé par Genoscreen (Lille, France), les *primers* utilisés étaient ceux fournis avec le vecteur (GTCATCAACCCGCTCCAAGG).

Le vecteur donneur, dont l'insert comporte une résistance à l'hygromycine, est le vecteur MCS1-EF1 α -RFP-T2A-Hygro-pA-MCS2-PGK-hsvTK (HR710PA-1, System biosciences, Palo Alto, CA, USA, **Figure 42 D**) auquel 2 bras d'homologies, synthétisés par Eurofins Genomics (Ebersberg, Germany), ont été ajoutés : le premier bras présente une homologie avec la séquence avant le codon stop de l'exon 13 de *MYOCD* suivi par 3 séquences FLAG puis le codon stop et le second bras présente une homologie avec le 3'UTR de *MYOCD* (les séquences sont dans la **Figure 42 E**). Le bras 1 est flanqué par deux séquences reconnus par BsrGI et le bras 2 par celles reconnues par NotI. De plus, le début de la séquence du second bras contient les ARN guides avec une modification de la séquence PAM. Les bras ont été introduits séquentiellement par digestion enzymatique (BsrGI, R0575S et NotI, R3189S, New England Biolabs, Ipswich, MA, USA) et ligation (ligase T4, M0202S, New England Biolabs, Ipswich, MA, USA) en suivant les recommandations du fournisseur des enzymes. Après chaque insertion, les vecteurs ont été transformés, produits et extraits de la même façon que les deux précédents, en ajoutant une étape de validation des colonies ayant bien intégré les bras par PCR (protocole ci-après) en utilisant les *primers* dont les séquences sont dans la **Figure 42 E** avant l'expansion. Pour terminer, les vecteurs ont été validés par séquençage Sanger en utilisant les même *primers* que pour la PCR (Genoscreen, Lille, France).



Figure 42 : Vecteurs et séquences ayant servi à la préparation de l'étiquetage de MYOCD

A Carte du vecteur comportant la Cas9. **B** Capture des écrans de résultats du logiciel CHOPCHOP pour les 2 ARN guides choisis. **C** Carte du vecteur utilisé pour insérer les ARN guides. **D** Carte du vecteur utilisé pour insérer les bras d'homologies et la séquence Tag-FLAG. **E** Séquences des bras d'homologie synthétisés. En vert les parties homologues de la séquence de référence de MYOCD, en rouge les séquences reconnues par une enzyme de restriction (BsrGI pour le bras 1 et NotI pour le 2), en bleu clair la séquence FLAG répétée 3 fois, en bleu foncé l'ARN guide et en orange la modification du PAM (ici pour l'ARN guide 1), les séquences soulignées sont celles des primers utilisés pour la PCR et le séquençage.

Les trois productions de virus, une par vecteur, servant à transduire les lignées cellulaires ont été réalisées par co-transfection de la lignée cellulaire HEK-293T avec un des vecteurs d'intérêt, un vecteur pPax2 et un pVsvg. Chaque lignée cellulaire modèle a ensuite été co-transduite avec les trois types de particules virales. Après 3 jours, les cellules ont été mises en présence de 100 µg/mL d'hygromycine (H0654, Sigma, St Louis, MO, USA). Une semaine après que les cellules contrôles soient mortes, les cellules transduites sont mises en présence de 5 µg/mL de ganciclovir (G2536, Sigma, St Louis, MO, USA), permettant le suicide des cellules ayant intégré l'insert du vecteur donneur par un autre moyen que par recombinaison homologue comme expliqué dans « **Mise au point d'un TAG de MYOCD par CRISPR** ». Comme préconisé par le fournisseur du vecteur, la ganciclovir est retirée après 72h de traitement.

PCR

L'ADN génomique est extrait des cellules par un protocole standard au phénol (GEXPHE03-00U, EuroBio Scientific, Les Ulis, France)/chloroforme (25666, Sigma, St Louis, MO, USA). Les *primers* utilisés sont présentés pour les PCR out-out et de jonctions dans le **Tableau 3** et ont été sélectionnés avec le logiciel Primer 3. Toutes les PCR ont été réalisées sur 50 ng d'ADN en utilisant l'ADN polymérase AmpliTaqGold® (4311820, Applied Biosystems, Foster City, CA, USA) selon les instructions du fournisseur avec le programme décrit dans la légende du **Tableau 3**.

PCR	Forward	Reverse
Out-Out	GGA CCT CAC TCC GCC AAA TTC	GCA CTG GTG CAT TGG GCA TT
Jonction	GGA CCT CAC TCC GCC AAA TTC	TAA TCG TCA TCG TCT TTG TAA TCA ATA T

Tableau 3 : Séquences des primers pour les PCR de validation de l'insertion du Tag-FLAG

Le programme de PCR était un *touch-down* 60°C (2 cycles à une température de 60°C, suivie par 2 cycles à 59°C, 2 cycles à 58°C, 3 cycles à 57°C, 3 cycles à 56°C, 4 cycles à 55°C, 4 cycles à 54°C, 5 cycles à 53°C et 10 cycles à 52°C).

Immunofluorescence sur lignée

Les cellules ont été fixées en PFA 4% pendant 10 min. Après lavages au PBS 1X, les lames ont été saturées avec 5% de BSA/PBS 1X. Le marquage immunofluorescent a été réalisé dans les mêmes conditions que sur les blocs FFPE sauf l'incubation de l'anticorps primaire qui était ici d'1h à température ambiante : anticorps primaire anti-FLAG (1:500, 14793S, Cell Signaling Technologies, Danvers, MA, USA) et anticorps secondaire AlexaFluor 488 anti-rabbit (1:400, A11008, Thermo Fisher Scientific, Waltham, MA, USA). Pour finir, les lames ont été montées en vectashield/DAPI (H-1200-10, Vector Laboratories, Burlingame, CaliforniaCA, USA). L'analyse et les images ont été faites sur le vidéomicroscope Cell Observer (Carl Zeiss, Oberkochen, Germany).

Immunoprécipitation de chromatine

Concernant les ChIP de la partie « Cibler la voie SRF/MYOCD » uniquement, les cellules ont été traitées au CCG-100602 (10787, Bertin Bioreagent, Montigny le Bretonneux, France) ou au DMSO (41639, Sigma, St Louis, MO, USA) à leur IC50 respectifs (OC80: 5 µg/mL, OC48: 15 µg/mL, OC88: 18 µg/mL, OC98 et OC110: 20 µg/mL), 48h avant la fixation. Du formaldéhyde (F8775, Sigma, St Louis, MO, USA) a été ajouté au milieu de culture à une concentration finale de 1% et le *crosslinking* a duré 15min à température ambiante. La réaction a été stoppée par ajout de glycine (50046, Sigma, St Louis, MO, USA) à une concentration finale de 0,125M. Les cellules sont lavées au PBS 1X après 5 min et récupérées en PBS 1X par grattage. Les culots de cellules ont été soumis à une première lyse dans le tampon suivant : Pipes 5mM (80635, Sigma, St Louis, MO, USA), KCl 85mM (P5405, Sigma, St Louis, MO, USA) et NP-40 0.5% (I8896, Sigma, St Louis, MO, USA). La lyse a été homogénéisée avec un *dounce* et les noyaux ont été récupérés par centrifugation. Ils ont ensuite été incubés 10 min sur glace dans un tampon de lyse nucléaire (Tris pH8 50mM (T1503, Sigma, St Louis, MO, USA), EDTA pH8 10mM (15575-020, Invitrogen, Carlsbad, CA, USA), SDS 1% (EU0760-A, Euromedex, Strasbourg, France)) et soniqués sur glace par des cycles de 15 secondes de sonication et 30 secondes de repos (repos de 3min tous les 5 cycles) à 60% de la puissance du sonicateur (Vibra Cell, Sonics & Materials, Newtown, CO, USA)) afin d'obtenir des fragments entre 200 et 750 pb. Les échantillons ont été dilués 10 fois dans un tampon de dilution (SDS 0,01%, Triton X-100 1,1% (93443, Sigma, St Louis, MO, USA), EDTA pH8 1,2 mM, Tris pH8 16,7 mM et NaCl 167 mM (31434, Sigma, St Louis, MO, USA)) et prélavés durant 1h à 4°C avec 100 µL de billes de protéine A (20334, Thermo Fisher Scientific, Waltham, MA, USA) et de protéine G (20399, Thermo Fisher Scientific, Waltham, MA, USA) préalablement bloquées. Le blocage a été réalisé en incubant les billes pendant 4h à 4°C avec 560µg de BSA (B8667, Sigma, St Louis, MO, USA). 200 ng de chromatines prélavées ont été incubées sur la nuit à 4°C avec 5µg d'anti-SRF (A303-172A, Bethyl Laboratories, Montgomery, TX, USA) ou 2 µg d'anti-polIII (A304-405A, Bethyl Laboratories, Montgomery, TX, USA) comme contrôle positif ou 5 µg d'IgG (M0659, Sigma, St Louis, MO, USA) comme contrôle négatif. Les complexes immuns ont été récupérés avec 50 µL de billes bloquées de protéines-A/protéines-G durant 1 h à 4°C sur roue. Les billes ont ensuite été lavées une fois dans

un tampon de dialyse (EDTA pH8 2mM, Tris pH8 50 mM, Sarkosyl 0,2% (L5777, Sigma, St Louis, MO, USA)) et 5 fois avec un tampon de lavage (Tris pH8 100 mM, LiCl 500 mM (65476, Sigma, St Louis, MO, USA), NP-40 1%, NaDoc 1% (D6750, Sigma, St Louis, MO, USA)). Les billes ont été resuspendues dans 150 μ L de tampon TE 1X (Tris pH8 10 mM, EDTA pH8 0,5 mM) et les ARN ont été enlevés du mélange par traitement à la RNase A (12091-021, Invitrogen, Carlsbad, CA, USA) durant 30 min à 37°C. Le *crosslinking* a été inversé en présence de 0,5% de SDS à 70°C sur la nuit. Après 1h30 de traitement à la protéinase K (3115828001, Roche, Basel, Suisse) à 50°C, les ADN immunoprécipités et les ADN d'entrée (*input*) ont été purifiés par un protocole standard phénol (15593031, Invitrogen, Carlsbad, CA, USA) /chloroforme (25666, Sigma, St Louis, MO, USA), précipités à l'éthanol 100% et analysés en triplicat par PCR en temps réel avec le *master mix TBGreen Premix Ex Taq Rox Plus* (RR42LR, Takara Bio,

Gène avec cible SRF dans son promoteur	Forward	Reverse
ACTB	CTC GAG CCA TAA AAG GCA AC	TTC CTC AAT CTC GCT CTC G
TAGLN	GCA GGG TCC TGT CCA TAA AA	AGG CTC CAT ATT TGG GGA AA
MYL6	CAG GTT GGG ATG ACC CTA GT	ACT ACA CTT CTC AGG GCA GGA
C- Chr22	CCC ATC TCA ACC TCC ACA CT	CCC ATC TCA ACC TCC ACA CT

Tableau 4 : Séquences des primers utilisés pour les qPCR de validation des ChIP

Kusatsu, Shiga, Japan) selon les recommandations du fournisseur ; les *primers* utilisés sont dans le **Tableau 4**.

Pour les ChIPseq, les ADN purifiés venant de multiples ChIP ont été rassemblés et précipités à l'éthanol 100% afin d'avoir 5 ng (dosage au qBit avec le kit haute sensibilité de l'ADNdb (Q32851, Invitrogen, Carlsbad, CA, USA)) pour préparer les bibliothèques en suivant le kit *TruSeq ChIP library* (15034288, Illumina, San Diego, CA, USA) selon les recommandations du fournisseur. Le séquençage a été réalisé par le Centro Nacional de Análisis Genómico (CNAG, Barcelona, Spain) sur HiSeq 3000 (Illumina, San Diego, CA, USA).

Analyse de cytotoxicité à l'Incucyte

Les lignées cellulaires de LMS et d'UPS ont chacune étéensemencées dans 21 puits d'une plaque 96 puits à raison de 3000 cellules par puits (sauf pour les oLMS qui ont étéensemencés avec 2500 cellules). Le lendemain, une gamme de l'inhibiteur de SRF/MYOCD, CCG-100602 (10787, Bertin Bioreagent, Montigny le Bretonneux, France), allant de 50 à 1,56 µg/mL, a été déposée en triplicat sur les cellules. Du Molecular Probes™ CellEvent™ Caspase-3/7 Green ReadyProbes™ (R37111, Invitrogen, Carlsbad, CA, USA), réactif exprimant la GFP quand il est coupé par la caspase 3/7, a été ajouté au milieu (2 gouttes par mL de milieu).

Les cellules sont placées dans l'Incucyte (Essen Bioscience, Newark, United Kingdom) durant 48h, et trois images par puits sont prises toutes les 2h. Chaque image est prise en phase et en GFP. Puis les images sont analysées avec le logiciel associé à l'Incucyte. Pour ce faire, pour chaque lignée cellulaire, un masque est élaboré pour reconnaître chaque cellule (analyse *cell-by-cell*), puis une analyse cellule à cellule permet de distinguer les cellules exprimant de la GFP. Cette technique est appliquée à chaque image prise durant les 48h. Le premier masque permet aussi de déterminer la confluence présente dans chaque image.

Bibliographie

Aapola, U. *et al.* (2000) 'Isolation and Initial Characterization of a Novel Zinc Finger Gene, DNMT3L, on 21q22.3, Related to the Cytosine-5- Methyltransferase 3 Gene Family', *Genomics*, 65(3), pp. 293–298. doi: 10.1006/geno.2000.6168.

Abbas, A. R. *et al.* (2005) 'Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data', *Genes & Immunity*, 6(4), pp. 319–331. doi: 10.1038/sj.gene.6364173.

Ahvenainen, T. V. *et al.* (2018) 'Loss of ATRX/DAXX expression and alternative lengthening of telomeres in uterine leiomyomas', *Cancer*, 124(24), pp. 4650–4656. doi: 10.1002/cncr.31754.

Alexandrov, L. B. and Stratton, M. R. (2014) 'Mutational signatures: the patterns of somatic mutations hidden in cancer genomes', *Current Opinion in Genetics & Development*, 24, pp. 52–60. doi: 10.1016/j.gde.2013.11.014.

Altschul, S. F. *et al.* (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Research*, 25(17), pp. 3389–3402. doi: 10.1093/nar/25.17.3389.

Amorim, J. P. *et al.* (2016) 'The Role of ATRX in the Alternative Lengthening of Telomeres (ALT) Phenotype', *Genes*, 7(9), p. 66. doi: 10.3390/genes7090066.

Annala, M. J. *et al.* (2013) 'Fusion genes and their discovery using high throughput sequencing', *Cancer Letters*, 340(2), pp. 192–200. doi: 10.1016/j.canlet.2013.01.011.

Aurias, A. *et al.* (1983) 'Translocation of chromosome 22 in Ewing's sarcoma', *Comptes rendus des seances de l'Academie des sciences. Serie III, Sciences de la vie*, 296(23). Available at: <http://pubmed-ncbi-nlm-nih-gov./6416623/> (Accessed: 11 August 2020).

Ayodele, O. and Abdul Razak, A. R. (2020) 'Immunotherapy in soft-tissue sarcoma', *Current Oncology*, 26. doi: 10.3747/co.26.5407.

Babij, P. (1993) 'Tissue-specific and developmentally regulated alternative splicing of a visceral isoform of smooth muscle myosin heavy chain', *Nucleic Acids Research*, 21(6), pp. 1467–1471. doi: 10.1093/nar/21.6.1467.

Badri, K. R., Zhou, Y. and Schuger, L. (2008) 'Embryological Origin of Airway Smooth Muscle', *Proceedings of the American Thoracic Society*, 5(1), pp. 4–10. doi: 10.1513/pats.200704-049VS.

Baird, K. *et al.* (2005) 'Gene Expression Profiling of Human Sarcomas: Insights into Sarcoma Biology', *Cancer Research*, 65(20), pp. 9226–9235. doi: 10.1158/0008-5472.CAN-05-1699.

- Bakhoun, S. F. and Cantley, L. C. (2018) 'The Multifaceted Role of Chromosomal Instability in Cancer and Its Microenvironment', *Cell*, 174(6), pp. 1347–1360. doi: 10.1016/j.cell.2018.08.027.
- Banito, A. *et al.* (2018) 'The SS18-SSX Oncoprotein Hijacks KDM2B-PRC1.1 to Drive Synovial Sarcoma', *Cancer Cell*, 33(3), pp. 527–541.e8. doi: 10.1016/j.ccell.2018.01.018.
- Barnea, E. *et al.* (1990) 'Specificity of expression of the muscle and brain dystrophin gene promoters in muscle and brain cells', *Neuron*, 5(6), pp. 881–888. doi: 10.1016/0896-6273(90)90348-J.
- Baumann, C. and Fuente, R. D. L. (2009) 'ATRX marks the inactive X chromosome (Xi) in somatic cells and during imprinted X chromosome inactivation in trophoblast stem cells', *Chromosoma*, 118(2), pp. 209–222. doi: 10.1007/s00412-008-0189-x.
- Baylin, S. B. and Jones, P. A. (2016) 'Epigenetic Determinants of Cancer', *Cold Spring Harbor Perspectives in Biology*, 8(9), p. a019505. doi: 10.1101/cshperspect.a019505.
- Bennett, M. R., Sinha, S. and Owens, G. K. (2016) 'Vascular smooth muscle cells in atherosclerosis', *Circulation research*, 118(4), pp. 692–702. doi: 10.1161/CIRCRESAHA.115.306361.
- Benson, E. S. (1961) 'Leukemia and the Philadelphia chromosome', *Postgraduate Medicine*, 30, pp. A22–A28.
- Berman, B. P. *et al.* (2012) 'Exploring the cancer methylome', *BMC Proceedings*, 6(Suppl 6), p. O24. doi: 10.1186/1753-6561-6-S6-O24.
- Beroukhi, R. *et al.* (2010) 'The landscape of somatic copy-number alteration across human cancers', *Nature*, 463(7283), pp. 899–905. doi: 10.1038/nature08822.
- Bhat, H. F. *et al.* (2018) 'ABC of multifaceted dystrophin glycoprotein complex (DGC)', *Journal of Cellular Physiology*, 233(7), pp. 5142–5159. doi: 10.1002/jcp.25982.
- Blake, D. J. *et al.* (2002) 'Function and Genetics of Dystrophin and Dystrophin-Related Proteins in Muscle', *Physiological Reviews*, 82(2), pp. 291–329. doi: 10.1152/physrev.00028.2001.
- Blanchoin, L. *et al.* (2014) 'Actin Dynamics, Architecture, and Mechanics in Cell Motility', *Physiological Reviews*, 94(1), pp. 235–263. doi: 10.1152/physrev.00018.2013.
- Blay, J.-Y. *et al.* (2012) 'Gastrointestinal Stromal Tumors (GIST): a Rare Entity, a Tumor Model for Personalized Therapy, and Yet Ten Different Molecular Subtypes', *Discovery Medicine*, 13(72), pp. 357–367.
- Blay, J.-Y. *et al.* (2017) 'Improved survival using specialized multidisciplinary board in sarcoma patients', *Annals of Oncology*, 28(11), pp. 2852–2859. doi: 10.1093/annonc/mdx484.

Blokzijl, F. *et al.* (2016) 'Tissue-specific mutation accumulation in human adult stem cells during life', *Nature*, 538(7624), pp. 260–264. doi: 10.1038/nature19768.

Bulun, S. E. (2013) 'Uterine Fibroids', *The New England Journal of Medicine*, (369), pp. 1344–1355. doi: 10.1056/NEJMra1209993.

Calin, G. A. *et al.* (2002) 'Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia', *Proceedings of the National Academy of Sciences*, 99(24), pp. 15524–15529. doi: 10.1073/pnas.242606799.

Calo, E. *et al.* (2010) 'Rb regulates fate choice and lineage commitment in vivo', *Nature*, 466(7310), pp. 1110–1114. doi: 10.1038/nature09264.

Campbell, J. and Campbell, G. (1994) 'The role of smooth muscle cells in atherosclerosis', *Current Opinion in Lipidology*, 5(5), pp. 323–30. doi: 10.1097/00041433-199410000-00003.

Campbell Julie H. and Campbell Gordon R. (2012) 'Smooth Muscle Phenotypic Modulation – A Personal Experience', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 32(8), pp. 1784–1789. doi: 10.1161/ATVBAHA.111.243212.

Carvalho, S. D. *et al.* (2019) 'Pleomorphic Sarcomas', *Surgical Pathology Clinics*, 12(1), pp. 63–105. doi: 10.1016/j.path.2018.10.004.

Casali, P. G., Bielack, S., *et al.* (2018) 'Bone sarcomas: ESMO-PaedCan-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up', *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 29(Suppl 4), pp. iv79–iv95. doi: 10.1093/annonc/mdy310.

Casali, P. G., Abecassis, N., *et al.* (2018a) 'Gastrointestinal stromal tumours: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up', *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 29(Suppl 4), pp. iv68–iv78. doi: 10.1093/annonc/mdy095.

Casali, P.G. *et al.* (2018) 'Soft tissue and visceral sarcomas: ESMO–EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up', *Annals of Oncology*, 29, pp. iv51–iv67. doi: 10.1093/annonc/mdy096.

Casali, P. G., Abecassis, N., *et al.* (2018b) 'Soft tissue and visceral sarcomas: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up', *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, 29(Suppl 4), pp. iv51–iv67. doi: 10.1093/annonc/mdy096.

Catalanotto, C., Cogoni, C. and Zardo, G. (2016) 'MicroRNA in Control of Gene Expression: An Overview of Nuclear Functions', *International Journal of Molecular Sciences*, 17(10), p. 1712. doi: 10.3390/ijms17101712.

Chaligné, R. and Heard, E. (2014) 'X-chromosome inactivation in development and cancer', *FEBS Letters*, 588(15), pp. 2514–2522. doi: 10.1016/j.febslet.2014.06.023.

- Chanrion, M. *et al.* (2008) 'A Gene Expression Signature that Can Predict the Recurrence of Tamoxifen-Treated Primary Breast Cancer', *Clinical Cancer Research*, 14(6), pp. 1744–1752. doi: 10.1158/1078-0432.CCR-07-1833.
- Chase, A. and Cross, N. C. P. (2011) 'Aberrations of EZH2 in Cancer', *Clinical Cancer Research*, 17(9), pp. 2613–2618. doi: 10.1158/1078-0432.CCR-10-2156.
- Chelly, J. *et al.* (1990) 'Dystrophin gene transcribed from different promoters in neuronal and glial cells', *Nature*, 344(6261), pp. 64–65. doi: 10.1038/344064a0.
- Chen, F. *et al.* (2011) 'Frequent epigenetic inactivation of Myocardin in human nasopharyngeal carcinoma', *Head & Neck*, 33(1), pp. 54–59. doi: 10.1002/hed.21396.
- Chen Jie *et al.* (2011) 'Induction of MicroRNA-1 by Myocardin in Smooth Muscle Cells Inhibits Cell Proliferation', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 31(2), pp. 368–375. doi: 10.1161/ATVBAHA.110.218149.
- Chen, X. *et al.* (2014) 'Recurrent Somatic Structural Variations Contribute to Tumorigenesis in Pediatric Osteosarcoma', *Cell Reports*, 7(1), pp. 104–112. doi: 10.1016/j.celrep.2014.03.003.
- Chen, X., Anstey, A. V. and Bugert, J. J. (2013) 'Molluscum contagiosum virus infection', *The Lancet Infectious Diseases*, 13(10), pp. 877–888. doi: 10.1016/S1473-3099(13)70109-9.
- Chi, J.-T. *et al.* (2007) 'Gene Expression Programs of Human Smooth Muscle Cells: Tissue-Specific Differentiation and Prognostic Significance in Breast Cancers', *PLoS Genetics*, 3(9). doi: 10.1371/journal.pgen.0030164.
- Chibon, F. *et al.* (2000) 'The RB1 Gene Is the Target of Chromosome 13 Deletions in Malignant Fibrous Histiocytoma', *Cancer Research*, 60(22), pp. 6339–6345.
- Chibon, F. *et al.* (2004) 'ASK1 (MAP3K5) as a potential therapeutic target in malignant fibrous histiocytomas with 12q14–q15 and 6q23 amplifications', *Genes, Chromosomes and Cancer*, 40(1), pp. 32–37. doi: 10.1002/gcc.20012.
- Chibon, F. *et al.* (2010) 'Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity', *Nature Medicine*, 16(7), p. 781. doi: 10.1038/nm.2174.
- Chibon, F. *et al.* (2019) 'CINSARC signature as a prognostic marker for clinical outcome in sarcomas and beyond', *Genes, Chromosomes and Cancer*, 58(2), pp. 124–129. doi: 10.1002/gcc.22703.
- Choucair, K. *et al.* (2019) 'Natural killer cells: a review of biology, therapeutic potential and challenges in treatment of solid tumors', *Future Oncology*, 15(26), pp. 3053–3069. doi: 10.2217/fo-2019-0116.
- Chudasama, P. *et al.* (2018) 'Integrative genomic and transcriptomic analysis of

leiomyosarcoma', *Nature Communications*, 9. doi: 10.1038/s41467-017-02602-0.

Chung, J.-H. *et al.* (2006) 'The ERK1/2 pathway modulates nuclear PTEN-mediated cell cycle arrest by cyclin D1 transcriptional regulation', *Human Molecular Genetics*, 15(17), pp. 2553–2559. doi: 10.1093/hmg/ddl177.

Clark, J. *et al.* (1994) 'Identification of novel genes, SYT and SSX, involved in the t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma', *Nature Genetics*, 7(4), pp. 502–508. doi: 10.1038/ng0894-502.

Clynes, D. *et al.* (2015) 'Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX', *Nature Communications*, 6. doi: 10.1038/ncomms8538.

Coindre, J. M. (1998) 'Recommendations for anatomo-pathologic management of soft tissue sarcomas in the adult. Pathologists of the FNCLCC Sarcoma Group (Fédération Nationale des Centres de Lutte Contre le Cancer)', *Annales De Pathologie*, 18(6), pp. 505–511.

Coindre, J. M. (2010) 'Biologie moléculaire des sarcomes', *Bulletin du Cancer*, 97(11), pp. 1337–1345. doi: 10.1684/bdc.2010.1213.

Cordes, K. R. *et al.* (2009) 'miR-145 and miR-143 regulate smooth muscle cell fate and plasticity', *Nature*, 460(7256), pp. 705–710. doi: 10.1038/nature08195.

Cortés-Ciriano, I. *et al.* (2020) 'Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing', *Nature Genetics*, 52(3), pp. 331–341. doi: 10.1038/s41588-019-0576-7.

Cowell, J. K. and Wigley, C. B. (1980) 'Changes in DNA Content During In Vitro Transformation of Mouse Salivary Gland Epithelium', *JNCI: Journal of the National Cancer Institute*, 64(6), pp. 1443–1449. doi: 10.1093/jnci/64.6.1443.

Crawford, G. *et al.* (2018) 'Epithelial damage and tissue $\gamma\delta$ T cells promote a unique tumor-protective IgE response', *Nature immunology*, 19(8), p. 859. doi: 10.1038/s41590-018-0161-8.

Creemers, E. E. *et al.* (2006) 'Myocardin is a direct transcriptional target of Mef2, Tead and Foxo proteins during cardiovascular development', *Development*, 133(21), pp. 4245–4256. doi: 10.1242/dev.02610.

Cuppens, T. *et al.* (2017) 'Potential Targets' Analysis Reveals Dual PI3K/mTOR Pathway Inhibition as a Promising Therapeutic Strategy for Uterine Leiomyosarcomas – an ENITEC Group Initiative', *Clinical Cancer Research*, 23(5), pp. 1274–1285. doi: 10.1158/1078-0432.CCR-16-2149.

D'Angelo, S. P. *et al.* (2018) 'Nivolumab with or without ipilimumab treatment for metastatic sarcoma (Alliance A091401): two open-label, non-comparative, randomised, phase 2 trials', *The Lancet Oncology*, 19(3), pp. 416–426. doi:

10.1016/S1470-2045(18)30006-8.

Danussi, C. *et al.* (2018) 'Atrx inactivation drives disease-defining phenotypes in glioma cells of origin through global epigenomic remodeling', *Nature Communications*, 9(1), p. 1057. doi: 10.1038/s41467-018-03476-6.

Davoli, T. *et al.* (2013) 'Cumulative Haploinsufficiency and Triplosensitivity Drive Aneuploidy Patterns and Shape the Cancer Genome', *Cell*, 155(4), pp. 948–962. doi: 10.1016/j.cell.2013.10.011.

Delattre, O. *et al.* (1992) 'Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours', *Nature*, 359(6391), pp. 162–165. doi: 10.1038/359162a0.

Delespaul, L. *et al.* (2017) 'Recurrent TRIO Fusion in Nontranslocation-Related Sarcomas', *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, 23(3), pp. 857–867. doi: 10.1158/1078-0432.CCR-16-0290.

Delespaul, L. *et al.* (2019) 'Fusion-mediated chromosomal instability promotes aneuploidy patterns that resemble human tumors', *Oncogene*, pp. 1–12. doi: 10.1038/s41388-019-0859-6.

Demico, E. G. *et al.* (2012) 'New Therapeutic Targets in Soft Tissue Sarcoma', *Advances in Anatomic Pathology*, 19(3), pp. 170–180. doi: 10.1097/PAP.0b013e318253462f.

Derakhshani, A. *et al.* (2019) 'Mast cells: A double-edged sword in cancer', *Immunology Letters*, 209, pp. 28–35. doi: 10.1016/j.imlet.2019.03.011.

Dewhurst, S. M. *et al.* (2014) 'Tolerance of Whole-Genome Doubling Propagates Chromosomal Instability and Accelerates Cancer Genome Evolution', *Cancer Discovery*, 4(2), pp. 175–185. doi: 10.1158/2159-8290.CD-13-0285.

Dick, F. A. and Rubin, S. M. (2013) 'Molecular mechanisms underlying RB protein function', *Nature Reviews Molecular Cell Biology*, 14(5), pp. 297–306. doi: 10.1038/nrm3567.

Diederichs, S. *et al.* (2016) 'The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations', *EMBO Molecular Medicine*, 8(5), pp. 442–457. doi: 10.15252/emmm.201506055.

Ducimetière, F. *et al.* (2011) 'Incidence of Sarcoma Histotypes and Molecular Subtypes in a Prospective Epidemiological Study with Central Pathology Review and Molecular Testing', *PLOS ONE*, 6(8), p. e20294. doi: 10.1371/journal.pone.0020294.

Eichhorn, S. W. *et al.* (2014) 'mRNA Destabilization Is the Dominant Effect of Mammalian MicroRNAs by the Time Substantial Repression Ensues', *Molecular Cell*, 56(1), pp. 104–115. doi: 10.1016/j.molcel.2014.08.028.

Esner, M. *et al.* (2006) 'Smooth muscle of the dorsal aorta shares a common clonal origin with skeletal muscle of the myotome', *Development*, 133(4), pp. 737–749. doi: 10.1242/dev.02226.

Feil Susanne *et al.* (2014) 'Transdifferentiation of Vascular Smooth Muscle Cells to Macrophage-Like Cells During Atherogenesis', *Circulation Research*, 115(7), pp. 662–667. doi: 10.1161/CIRCRESAHA.115.304634.

Feil Susanne, Hofmann Franz and Feil Robert (2004) 'SM22 α Modulates Vascular Smooth Muscle Cell Phenotype During Atherogenesis', *Circulation Research*, 94(7), pp. 863–865. doi: 10.1161/01.RES.0000126417.38728.F6.

Fields, J. P. and Helwig, E. B. (1981) 'Leiomyosarcoma of the skin and subcutaneous tissue', *Cancer*, 47(1), pp. 156–169. doi: 10.1002/1097-0142(19810101)47:1<156::aid-cncr2820470127>3.0.co;2-#.

Fletcher, C. *et al.* (2013) *WHO Classification of Tumours of Soft Tissue and Bone*. 4th edn. Lyon: IARC Press.

Fligman, I. *et al.* (1995) 'Molecular diagnosis of synovial sarcoma and characterization of a variant SYT-SSX2 fusion transcript.', *The American Journal of Pathology*, 147(6), p. 1592.

Fourneaux, B. *et al.* (2016) '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*, 8(5), pp. 7878–7890. doi: 10.18632/oncotarget.13987.

Fridman, W. H. *et al.* (2017) 'The immune contexture in cancer prognosis and treatment', *Nature Reviews Clinical Oncology*, 14(12), pp. 717–734. doi: 10.1038/nrclinonc.2017.101.

Fujiwara, H. *et al.* (2011) 'The Basement Membrane of Hair Follicle Stem Cells Is a Muscle Cell Niche', *Cell*, 144(4), pp. 577–589. doi: 10.1016/j.cell.2011.01.014.

Gagnon, K. T. *et al.* (2014) 'RNAi Factors Are Present and Active in Human Cell Nuclei', *Cell Reports*, 6(1), pp. 211–221. doi: 10.1016/j.celrep.2013.12.013.

Gao, Q. Q. and McNally, E. M. (2015) 'The Dystrophin Complex: Structure, Function, and Implications for Therapy', *Comprehensive Physiology*, pp. 1223–1239. doi: 10.1002/cphy.c140048.

Gibault, L. *et al.* (2011) 'New insights in sarcoma oncogenesis: a comprehensive analysis of a large series of 160 soft tissue sarcomas with complex genomics', *The Journal of Pathology*, 223(1), pp. 64–71. doi: 10.1002/path.2787.

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

doi: 10.1038/modpathol.2011.163.

Gilgenkrantz, H. *et al.* (1992) 'Positive and negative regulatory DNA elements including a CCArGG box are involved in the cell type-specific expression of the human muscle dystrophin gene.', *Journal of Biological Chemistry*, 267(15), pp. 10823–10830.

Ginn-Pease, M. E. and Eng, C. (2003) 'Increased Nuclear Phosphatase and Tensin Homologue Deleted on Chromosome 10 Is Associated with G0-G1 in MCF-7 Cells', *Cancer Research*, 63(2), pp. 282–286.

Goldberg, A. D. *et al.* (2010) 'Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions', *Cell*, 140(5), pp. 678–691. doi: 10.1016/j.cell.2010.01.003.

van der Graaf, W. T. *et al.* (2012) 'Pazopanib for metastatic soft-tissue sarcoma (PALETTE): a randomised, double-blind, placebo-controlled phase 3 trial', *The Lancet*, 379(9829), pp. 1879–1886. doi: 10.1016/S0140-6736(12)60651-5.

Granero Peiró, L. (2015) 'Leiomyosarcoma of the ascending colon: a rare tumor with poor prognosis', *Revista Española de Enfermedades Digestivas*. doi: 10.17235/reed.2015.3606/2014.

Grisaru-Tal, S. *et al.* (2020) 'A new dawn for eosinophils in the tumour microenvironment', *Nature Reviews Cancer*, pp. 1–14. doi: 10.1038/s41568-020-0283-9.

Gualdrini, F. *et al.* (2016) 'SRF Co-factors Control the Balance between Cell Proliferation and Contractility', *Molecular Cell*, 64(6), pp. 1048–1061. doi: 10.1016/j.molcel.2016.10.016.

Guillou, L. and Aurias, A. (2010) 'Soft tissue sarcomas with complex genomic profiles', *Virchows Archiv*, 456(2), pp. 201–217. doi: 10.1007/s00428-009-0853-4.

Guo, H. *et al.* (2010) 'Mammalian microRNAs predominantly act to decrease target mRNA levels', *Nature*, 466(7308), pp. 835–840. doi: 10.1038/nature09267.

Gustafson, P. (1994) 'Soft tissue sarcoma. Epidemiology and prognosis in 508 patients', *Acta Orthopaedica Scandinavica. Supplementum*, 259, pp. 1–31.

Haldar, M. *et al.* (2007) 'A Conditional Mouse Model of Synovial Sarcoma: Insights into a Myogenic Origin', *Cancer Cell*, 11(4), pp. 375–388. doi: 10.1016/j.ccr.2007.01.016.

Hanahan, D. and Weinberg, R. A. (2011) 'Hallmarks of Cancer: The Next Generation', *Cell*, 144(5), pp. 646–674. doi: 10.1016/j.cell.2011.02.013.

Hao, H. *et al.* (2002) 'Heterogeneity of smooth muscle cell populations cultured from pig coronary artery', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 22(7), pp. 1093–1099. doi: 10.1161/01.atv.0000022407.91111.e4.

Hao Hiroyuki, Gabbiani Giulio and Bochaton-Piallat Marie-Luce (2003) 'Arterial Smooth Muscle Cell Heterogeneity', *Arteriosclerosis, Thrombosis, and Vascular Biology*,

23(9), pp. 1510–1520. doi: 10.1161/01.ATV.0000090130.85752.ED.

Hashimoto, H., Tsuneyoshi, M. and Enjoji, M. (1985) 'Malignant smooth muscle tumors of the retroperitoneum and mesentery: A clinicopathologic analysis of 44 cases', *Journal of Surgical Oncology*, 28(3), pp. 177–186. doi: 10.1002/jso.2930280307.

Hata, A. and Lieberman, J. (2015) 'Dysregulation of microRNA biogenesis and gene silencing in cancer', *Science Signaling*, 8(368), pp. re3–re3. doi: 10.1126/scisignal.2005825.

Hautmann, M. B., Madsen, C. S. and Owens, G. K. (1997) 'A transforming growth factor beta (TGFbeta) control element drives TGFbeta-induced stimulation of smooth muscle alpha-actin gene expression in concert with two CArG elements', *The Journal of Biological Chemistry*, 272(16), pp. 10948–10956. doi: 10.1074/jbc.272.16.10948.

Hayashi, K. *et al.* (2004) 'Insulin Receptor Substrate-1/SHP-2 Interaction, a Phenotype-dependent Switching Machinery of Insulin-like Growth Factor-I Signaling in Vascular Smooth Muscle Cells', *Journal of Biological Chemistry*, 279(39), pp. 40807–40818. doi: 10.1074/jbc.M405100200.

He, J. *et al.* (2020) 'Qualitative transcriptional signature for predicting pathological response of colorectal cancer to FOLFOX therapy', *Cancer Science*, 111(1), pp. 253–265. doi: 10.1111/cas.14263.

He, L. *et al.* (2005) 'A microRNA polycistron as a potential human oncogene', *Nature*, 435(7043), pp. 828–833. doi: 10.1038/nature03552.

Heaphy, C. M. *et al.* (2011) 'Altered Telomeres in Tumors with ATRX and DAXX Mutations', *Science*, 333(6041), pp. 425–425. doi: 10.1126/science.1207313.

Heinrich, M. C. *et al.* (2019) 'Genomic aberrations in cell cycle genes predict progression of KIT-mutant gastrointestinal stromal tumors (GISTs)', *Clinical Sarcoma Research*, 9(1), p. 3. doi: 10.1186/s13569-019-0112-7.

Hemming, M. L. *et al.* (2020) 'Oncogenic Gene-Expression Programs in Leiomyosarcoma and Characterization of Conventional, Inflammatory, and Uterogenic Subtypes', *Molecular Cancer Research*. doi: 10.1158/1541-7786.MCR-20-0197.

Henson, J. D. *et al.* (2002) 'Alternative lengthening of telomeres in mammalian cells', *Oncogene*, 21(4), pp. 598–610. doi: 10.1038/sj.onc.1205058.

Hernando, E. *et al.* (2007) 'The AKT-mTOR pathway plays a critical role in the development of leiomyosarcomas', *Nature Medicine*, 13(6), pp. 748–753. doi: 10.1038/nm1560.

Heyden, R. J. *et al.* (2013) '10.8 Smooth Muscle', in *Anatomy and Physiology*. OpenStax. Available at: <https://opentextbc.ca/anatomyandphysiology/chapter/10-8-smooth-muscle/> (Accessed: 24 July 2020).

- Hilal, L. *et al.* (2016) 'Gastrointestinal (GI) leiomyosarcoma (LMS) case series and review on diagnosis, management, and prognosis', *Medical Oncology*, 33(2), p. 20. doi: 10.1007/s12032-016-0730-3.
- Hillmer, A. M. *et al.* (2011) 'Comprehensive long-span paired-end-tag mapping reveals characteristic patterns of structural variations in epithelial cancer genomes', *Genome Research*, 21(5), pp. 665–675. doi: 10.1101/gr.113555.110.
- Hirota, S. *et al.* (1998) 'Gain-of-Function Mutations of c-kit in Human Gastrointestinal Stromal Tumors', *Science*, 279(5350), pp. 577–580. doi: 10.1126/science.279.5350.577.
- Hirsch, F. R. *et al.* (2017) 'Lung cancer: current therapies and new targeted treatments', *The Lancet*, 389(10066), pp. 299–311. doi: 10.1016/S0140-6736(16)30958-8.
- Honoré, C. *et al.* (2015) 'Soft tissue sarcoma in France in 2015: Epidemiology, classification and organization of clinical care', *Journal of Visceral Surgery*, 152(4), pp. 223–230. doi: 10.1016/j.jviscsurg.2015.05.001.
- Hooper, L. V. *et al.* (2001) 'Molecular Analysis of Commensal Host-Microbial Relationships in the Intestine', *Science*, 291(5505), pp. 881–884. doi: 10.1126/science.291.5505.881.
- Horita, H. *et al.* (2016) 'Nuclear PTEN functions as an essential regulator of SRF-dependent transcription to control smooth muscle differentiation', *Nature Communications*, 7, p. 10830. doi: 10.1038/ncomms10830.
- Hornick, J. L. (2018) 'Subclassification of pleomorphic sarcomas: How and why should we care?', *Annals of Diagnostic Pathology*, 37, pp. 118–124. doi: 10.1016/j.anndiagpath.2018.10.006.
- House, S. J. *et al.* (2008) 'The non-excitabile smooth muscle: Calcium signaling and phenotypic switching during vascular disease', *Pflugers Archiv: European journal of physiology*, 456(5), p. 769. doi: 10.1007/s00424-008-0491-8.
- Howard, B. D. and Tessman, I. (1964) 'Identification of the altered bases in mutated single-stranded DNA: II. In vivo mutagenesis by 5-bromodeoxyuridine and 2-aminopurine', *Journal of Molecular Biology*, 9(2), pp. 364–371. doi: 10.1016/S0022-2836(64)80213-8.
- Hugnot, J. P. *et al.* (1992) 'Distal transcript of the dystrophin gene initiated from an alternative first exon and encoding a 75-kDa protein widely distributed in nonmuscle tissues.', *Proceedings of the National Academy of Sciences of the United States of America*, 89(16), pp. 7506–7510.
- Ichii, T. *et al.* (2001) 'Fibrillar collagen specifically regulates human vascular smooth muscle cell genes involved in cellular responses and the pericellular matrix environment', *Circulation Research*, 88(5), pp. 460–467. doi: 10.1161/01.res.88.5.460.
- Ilias, A. S. *et al.* (2017) 'Rare case of primary leiomyosarcoma of sigmoid mesocolon',

Case Reports, 2017, p. bcr. doi: 10.1136/bcr-2017-219826.

Inaki, K. and Liu, E. T. (2012) 'Structural mutations in cancer: mechanistic and functional insights', *Trends in Genetics*, 28(11), pp. 550–559. doi: 10.1016/j.tig.2012.07.002.

International Cancer Genome Consortium (2020) *International Cancer Genome Consortium*. Available at: <https://icgc.org/content/icgc-home-0> (Accessed: 21 July 2020).

Issa, J.-P. J. *et al.* (2004) 'Phase 1 study of low-dose prolonged exposure schedules of the hypomethylating agent 5-aza-2'-deoxycytidine (decitabine) in hematopoietic malignancies', *Blood*, 103(5), pp. 1635–1640. doi: 10.1182/blood-2003-03-0687.

Italiano, A. *et al.* (2013) 'Genetic Profiling Identifies Two Classes of Soft-Tissue Leiomyosarcomas with Distinct Clinical Characteristics', *Clinical Cancer Research*, 19(5), pp. 1190–1196. doi: 10.1158/1078-0432.CCR-12-2970.

Jäger, N. *et al.* (2013) 'Hypermethylation of the Inactive X Chromosome Is a Frequent Event in Cancer', *Cell*, 155(3), pp. 567–581. doi: 10.1016/j.cell.2013.09.042.

Jansen, M. P. H. M. *et al.* (2007) 'HOXB13-to-IL17BR Expression Ratio Is Related With Tumor Aggressiveness and Response to Tamoxifen of Recurrent Breast Cancer: A Retrospective Study', *Journal of Clinical Oncology*, 25(6), pp. 662–668. doi: 10.1200/JCO.2006.07.3676.

Jemaà, M. *et al.* (2016) 'Heterogeneity in sarcoma cell lines reveals enhanced motility of tetraploid versus diploid cells', *Oncotarget*, 8(10), pp. 16669–16689. doi: 10.18632/oncotarget.14291.

Jensen, A. M. (2005) 'Potential roles for BMP and Pax genes in the development of iris smooth muscle', *Developmental Dynamics*, 232(2), pp. 385–392. doi: 10.1002/dvdy.20224.

Jensen-Jarolim, E. *et al.* (2008) 'AllergoOncology: the role of IgE-mediated allergy in cancer', *Allergy*, 63(10), pp. 1255–1266. doi: 10.1111/j.1398-9995.2008.01768.x.

Jögi, A. *et al.* (2012) 'Cancer cell differentiation heterogeneity and aggressive behavior in solid tumors', *Upsala Journal of Medical Sciences*, 117(2), pp. 217–224. doi: 10.3109/03009734.2012.659294.

Jones, P. A., Issa, J.-P. J. and Baylin, S. (2016) 'Targeting the cancer epigenome for therapy', *Nature Reviews Genetics*, 17(10), pp. 630–641. doi: 10.1038/nrg.2016.93.

Kanai Hiroyoshi *et al.* (2001) 'Transforming Growth Factor- β /Smads Signaling Induces Transcription of the Cell Type-Restricted Ankyrin Repeat Protein CARP Gene Through CAGA Motif in Vascular Smooth Muscle Cells', *Circulation Research*, 88(1), pp. 30–36. doi: 10.1161/01.RES.88.1.30.

Karpf, A. R. and Jones, D. A. (2002) 'Reactivating the expression of methylation

silenced genes in human cancer', *Oncogene*, 21(35), pp. 5496–5503. doi: 10.1038/sj.onc.1205602.

Kirsch, D. G. *et al.* (2007) 'A spatially and temporally restricted mouse model of soft tissue sarcoma', *Nature Medicine*, 13(8), pp. 992–997. doi: 10.1038/nm1602.

Komi, D. E. A. and Redegeld, F. A. (2019) 'Role of Mast Cells in Shaping the Tumor Microenvironment', *Clinical Reviews in Allergy & Immunology*. doi: 10.1007/s12016-019-08753-w.

Komuro, T. (2006) 'Structure and organization of interstitial cells of Cajal in the gastrointestinal tract', *The Journal of Physiology*, 576(3), pp. 653–658. doi: 10.1113/jphysiol.2006.116624.

Koulaxouzidis, G. *et al.* (2015) 'Is revisional surgery mandatory when an unexpected sarcoma diagnosis is made following primary surgery?', *World Journal of Surgical Oncology*, 13(1), pp. 1–10. doi: 10.1186/s12957-015-0719-y.

Kulis, M. and Esteller, M. (2010) '2 - DNA Methylation and Cancer', in Herceg, Z. and Ushijima, T. (eds) *Advances in Genetics*. Academic Press (Epigenetics and Cancer, Part A), pp. 27–56. doi: 10.1016/B978-0-12-380866-0.60002-2.

Lander, E. S. *et al.* (2001) 'Initial sequencing and analysis of the human genome', *Nature*, 409(6822), pp. 860–921. doi: 10.1038/35057062.

Lartigue, L. *et al.* (2020) 'Genome remodeling upon mesenchymal tumor cell fusion contributes to tumor progression and metastatic spread', *Oncogene*, 39(21), pp. 4198–4211. doi: 10.1038/s41388-020-1276-6.

Lavin, M. F. and Gueven, N. (2006) 'The complexity of p53 stabilization and activation', *Cell Death & Differentiation*, 13(6), pp. 941–950. doi: 10.1038/sj.cdd.4401925.

Lawlor, R. T. *et al.* (2019) 'Alternative lengthening of telomeres (ALT) influences survival in soft tissue sarcomas: a systematic review with meta-analysis', *BMC Cancer*, 19(1), pp. 1–7. doi: 10.1186/s12885-019-5424-8.

Lee, J.-C. *et al.* (2015) 'Alternative lengthening of telomeres and loss of ATRX are frequent events in pleomorphic and dedifferentiated liposarcomas', *Modern Pathology*, 28(8), pp. 1064–1073. doi: 10.1038/modpathol.2015.67.

Lee, P. J. *et al.* (2017) 'Spectrum of mutations in leiomyosarcomas identified by clinical targeted next-generation sequencing', *Experimental and Molecular Pathology*, 102(1), pp. 156–161. doi: 10.1016/j.yexmp.2017.01.012.

Lee, Y.-F. *et al.* (2016) 'Gene expression profiling identifies distinct molecular subgroups of leiomyosarcoma with clinical relevance', *British Journal of Cancer*, 115(8), pp. 1000–1007. doi: 10.1038/bjc.2016.280.

Lee-Six, H. *et al.* (2018) 'The landscape of somatic mutation in normal colorectal

epithelial cells', *bioRxiv*, p. 416800. doi: 10.1101/416800.

Leoh, L. S., Daniels-Wells, T. R. and Penichet, M. L. (2015) 'IgE Immunotherapy Against Cancer', in Lafaille, J. J. and Curotto de Lafaille, M. A. (eds) *IgE Antibodies: Generation and Function*. Cham: Springer International Publishing (Current Topics in Microbiology and Immunology), pp. 109–149. doi: 10.1007/978-3-319-13725-4_6.

Lesluyes, T. *et al.* (2017) 'The CINSARC signature as a prognostic marker for clinical outcome in multiple neoplasms', *Scientific Reports*, 7(1), pp. 1–8. doi: 10.1038/s41598-017-05726-x.

Lewis, W. H. (1903) 'Wandering pigmented cells arising from the epithelium of the optic cup, with observations on the origin of the M. sphincter pupillae in the chick', *American Journal of Anatomy*, 2(3), pp. 405–416. doi: 10.1002/aja.1000020311.

Li, L. *et al.* (1996) 'Expression of the SM22a Promoter in Transgenic Mice Provides Evidence for Distinct Transcriptional Regulatory Programs in Vascular and Visceral Smooth Muscle Cells', *Journal of Cell Biology*, 132(5), p. 11. doi: 10.1083/jcb.132.5.849.

Li, L. *et al.* (1997) 'Evidence for Serum Response Factor-Mediated Regulatory Networks Governing SM22 α Transcription in Smooth, Skeletal, and Cardiac Muscle Cells', *Developmental Biology*, 187(2), pp. 311–321. doi: 10.1006/dbio.1997.8621.

Li, L. *et al.* (2018) 'miR-145-5p Inhibits Vascular Smooth Muscle Cells (VSMCs) Proliferation and Migration by Dysregulating the Transforming Growth Factor- β Signaling Cascade', *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, 24, pp. 4894–4904. doi: 10.12659/MSM.910986.

Li, N. *et al.* (2009) 'Genetically transforming human mesenchymal stem cells to sarcomas', *Cancer*, 115(20), pp. 4795–4806. doi: 10.1002/cncr.24519.

Li, S. *et al.* (2001) 'Innate diversity of adult human arterial smooth muscle cells: cloning of distinct subtypes from the internal thoracic artery', *Circulation Research*, 89(6), pp. 517–525. doi: 10.1161/hh1801.097165.

Li, X. *et al.* (1997) 'Suppression of smooth-muscle alpha-actin expression by platelet-derived growth factor in vascular smooth-muscle cells involves Ras and cytosolic phospholipase A2', *The Biochemical Journal*, 327 (Pt 3), pp. 709–716. doi: 10.1042/bj3270709.

Li, Y. *et al.* (2012) 'The miR-17-92 cluster expands multipotent hematopoietic progenitors whereas imbalanced expression of its individual oncogenic miRNAs promotes leukemia in mice', *Blood*, 119(19), pp. 4486–4498. doi: 10.1182/blood-2011-09-378687.

Liang, J. *et al.* (2019) 'Global changes in chromatin accessibility and transcription following ATRX inactivation in human cancer cells', *FEBS Letters*, 0(0). doi: 10.1002/1873-3468.13549.

- Liao, X.-H. *et al.* (2014) 'Re-expression and epigenetic modification of maspin induced apoptosis in MCF-7 cells mediated by myocardin', *Cellular Signalling*, 26(6), pp. 1335–1346. doi: 10.1016/j.cellsig.2014.03.001.
- Liau, J.-Y., Lee, J.-C., *et al.* (2015) 'Comprehensive screening of alternative lengthening of telomeres phenotype and loss of ATRX expression in sarcomas', *Modern Pathology*, 28(12), pp. 1545–1554. doi: 10.1038/modpathol.2015.114.
- Liau, J.-Y., Tsai, J.-H., *et al.* (2015) 'Leiomyosarcoma With Alternative Lengthening of Telomeres Is Associated With Aggressive Histologic Features, Loss of ATRX Expression, and Poor Clinical Outcome', *The American Journal of Surgical Pathology*, 39(2), pp. 236–244. doi: 10.1097/PAS.0000000000000324.
- Liu Mingjun and Gomez Delphine (2019) 'Smooth Muscle Cell Phenotypic Diversity', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 39(9), pp. 1715–1723. doi: 10.1161/ATVBAHA.119.312131.
- Liu, Y., Sinha, S. and Owens, G. (2003) 'A Transforming Growth Factor- β Control Element Required for SM α -Actin Expression in Vivo Also Partially Mediates GSKF-dependent Transcriptional Repression', *Journal of Biological Chemistry*, 278(48), pp. 48004–48011. doi: 10.1074/jbc.M301902200.
- Liu, Z.-P. *et al.* (2005) 'Phenotypic Modulation of Smooth Muscle Cells through Interaction of Foxo4 and Myocardin', *Developmental Cell*, 9(2), pp. 261–270. doi: 10.1016/j.devcel.2005.05.017.
- Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method', *Methods*, 25(4), pp. 402–408. doi: 10.1006/meth.2001.1262.
- Llombart, B. *et al.* (2019) 'Leiomyosarcoma and Pleomorphic Dermal Sarcoma: Guidelines for Diagnosis and Treatment', *Actas Dermo-Sifiliográficas (English Edition)*, 110(1), pp. 4–11. doi: 10.1016/j.adengl.2018.11.002.
- Long, J. Z. *et al.* (2014) 'A Smooth Muscle-Like Origin for Beige Adipocytes', *Cell Metabolism*, 19(5), pp. 810–820. doi: 10.1016/j.cmet.2014.03.025.
- Long, X. *et al.* (2008) 'Myocardin is Sufficient for a Smooth Muscle-Like Contractile Phenotype', *Arteriosclerosis, thrombosis, and vascular biology*, 28(8), pp. 1505–1510. doi: 10.1161/ATVBAHA.108.166066.
- Lovejoy, C. A. *et al.* (2012) 'Loss of ATRX, Genome Instability, and an Altered DNA Damage Response Are Hallmarks of the Alternative Lengthening of Telomeres Pathway', *PLoS Genetics*, 8(7). doi: 10.1371/journal.pgen.1002772.
- Lv, J. *et al.* (2018) 'Long noncoding RNA H19-derived miR-675 aggravates restenosis by targeting PTEN', *Biochemical and Biophysical Research Communications*, 497(4), pp. 1154–1161. doi: 10.1016/j.bbrc.2017.01.011.

- Lyu, Q. *et al.* (2018) 'CRISPR-Cas9 Mediated Epitope Tagging Provides Accurate and Versatile Assessment of Myocardin', *Arteriosclerosis, thrombosis, and vascular biology*, p. ATVBAHA-118.
- Ma, X.-J. *et al.* (2004) 'A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen', *Cancer Cell*, 5(6), pp. 607–616. doi: 10.1016/j.ccr.2004.05.015.
- Mack Christopher P. and Owens Gary K. (1999) 'Regulation of Smooth Muscle α -Actin Expression In Vivo Is Dependent on CARG Elements Within the 5' and First Intron Promoter Regions', *Circulation Research*, 84(7), pp. 852–861. doi: 10.1161/01.RES.84.7.852.
- Mahajan, N. *et al.* (2019) 'Maspin binds to cardiolipin in mitochondria and triggers apoptosis', *The FASEB Journal*, 33(5), p. 6354. doi: 10.1096/fj.201802182R.
- Majesky, M. W. *et al.* (2017) 'Differentiated Smooth Muscle Cells Generate a Subpopulation of Resident Vascular Progenitor Cells in the Adventitia Regulated by Klf4', *Circulation Research*, 120(2), pp. 296–311. doi: 10.1161/CIRCRESAHA.116.309322.
- Maki, H. (2002) 'Origins of Spontaneous Mutations: Specificity and Directionality of Base-Substitution, Frameshift, and Sequence-Substitution Mutageneses', *Annual Review of Genetics*, 36(1), pp. 279–303. doi: 10.1146/annurev.genet.36.042602.094806.
- Manabe, I. and Owens, G. K. (2001) 'The Smooth Muscle Myosin Heavy Chain Gene Exhibits Smooth Muscle Subtype-selective Modular Regulation in Vivo', *Journal of Biological Chemistry*, 276(42), pp. 39076–39087. doi: 10.1074/jbc.M105402200.
- Manabe Ichiro and Owens Gary K. (2001) 'Recruitment of Serum Response Factor and Hyperacetylation of Histones at Smooth Muscle-Specific Regulatory Regions During Differentiation of a Novel P19-Derived In Vitro Smooth Muscle Differentiation System', *Circulation Research*, 88(11), pp. 1127–1134. doi: 10.1161/hh1101.091339.
- Mao, X. *et al.* (2012) 'Vascular Smooth Muscle Cell Smad4 Gene is Important for Mouse Vascular Development', *Arteriosclerosis, thrombosis, and vascular biology*, 32(9), pp. 2171–2177. doi: 10.1161/ATVBAHA.112.253872.
- Marais, R., Wynne, J. and Treisman, R. (1993) 'The SRF accessory protein Elk-1 contains a growth factor-regulated transcriptional activation domain', *Cell*, 73(2), pp. 381–393. doi: 10.1016/0092-8674(93)90237-K.
- Marchi, S. and Pinton, P. (2016) 'Alterations of calcium homeostasis in cancer cells', *Current Opinion in Pharmacology*, 29, pp. 1–6. doi: 10.1016/j.coph.2016.03.002.
- Marcozzi, A., Pellestor, F. and Kloosterman, W. P. (2018) 'The Genomic Characteristics and Origin of Chromothripsis', in Pellestor, F. (ed.) *Chromothripsis: Methods and Protocols*. New York, NY: Springer (Methods in Molecular Biology), pp. 3–19. doi: 10.1007/978-1-4939-7780-2_1.

- Mariani, O. *et al.* (2007) 'JUN Oncogene Amplification and Overexpression Block Adipocytic Differentiation in Highly Aggressive Sarcomas', *Cancer Cell*, 11(4), pp. 361–374. doi: 10.1016/j.ccr.2007.02.007.
- Martincorena, I. *et al.* (2017) 'Universal Patterns of Selection in Cancer and Somatic Tissues', *Cell*, 171(5), pp. 1029–1041.e21. doi: 10.1016/j.cell.2017.09.042.
- Martincorena, I. *et al.* (2018) 'Somatic mutant clones colonize the human esophagus with age', *Science*, 362(6417), pp. 911–917. doi: 10.1126/science.aau3879.
- Martincorena, I. (2019) 'Somatic mutation and clonal expansions in human tissues', *Genome Medicine*, 11(1), pp. 1–3. doi: 10.1186/s13073-019-0648-4.
- Martincorena, I. and Campbell, P. J. (2015) 'Somatic mutation in cancer and normal cells', *Science*, 349(6255), pp. 1483–1489. doi: 10.1126/science.aab4082.
- Marzec, P. *et al.* (2015) 'Nuclear-Receptor-Mediated Telomere Insertion Leads to Genome Instability in ALT Cancers', *Cell*, 160(5), pp. 913–927. doi: 10.1016/j.cell.2015.01.044.
- Matheu, A., Maraver, A. and Serrano, M. (2008) 'The Arf/p53 Pathway in Cancer and Aging', *Cancer Research*, 68(15), pp. 6031–6034. doi: 10.1158/0008-5472.CAN-07-6851.
- Matushansky, I. *et al.* (2007) 'Derivation of sarcomas from mesenchymal stem cells via inactivation of the Wnt pathway', *The Journal of Clinical Investigation*, 117(11), pp. 3248–3257. doi: 10.1172/JCI31377.
- Mauduit, O. *et al.* (2019) 'Recurrent DMD Deletions Highlight Specific Role of Dp71 Isoform in Soft-Tissue Sarcomas', *Cancers*, 11(7), p. 922. doi: 10.3390/cancers11070922.
- Maxam, A. M. and Gilbert, W. (1977) 'A new method for sequencing DNA.', *Proceedings of the National Academy of Sciences of the United States of America*, 74(2), pp. 560–564.
- McDonald, O. G. *et al.* (2006) 'Control of SRF binding to CArG box chromatin regulates smooth muscle gene expression in vivo', *The Journal of Clinical Investigation*, 116(1), pp. 36–48. doi: 10.1172/JCI26505.
- McHugh, K. M. (1996) 'Molecular Analysis of Gastrointestinal Smooth Muscle Development', *Journal of Pediatric Gastroenterology and Nutrition*, 23(4), pp. 379–394.
- McLin, V. A., Henning, S. J. and Jamrich, M. (2009) 'The Role of the Visceral Mesoderm in the Development of the Gastrointestinal Tract', *Gastroenterology*, 136(7), pp. 2074–2091. doi: 10.1053/j.gastro.2009.03.001.
- Merle, C. *et al.* (2020) 'Tetraploidization of Immortalized Myoblasts Induced by Cell Fusion Drives Myogenic Sarcoma Development with DMD Deletion', *Cancers*, 12(5), p. 1281. doi: 10.3390/cancers12051281.

Mertens, F. *et al.* (2009) 'Translocation-Related Sarcomas', *Seminars in Oncology*, 36(4), pp. 312–323. doi: 10.1053/j.seminoncol.2009.06.004.

Miano, J. (2003) 'Serum response factor: toggling between disparate programs of gene expression', *Journal of Molecular and Cellular Cardiology*, 35(6), pp. 577–593. doi: 10.1016/S0022-2828(03)00110-X.

Miano J M *et al.* (1994) 'Smooth muscle myosin heavy chain exclusively marks the smooth muscle lineage during mouse embryogenesis.', *Circulation Research*, 75(5), pp. 803–812. doi: 10.1161/01.RES.75.5.803.

Miano, J. M. (2015) 'Myocardin in biology and disease', *Journal of Biomedical Research*, 29(1), pp. 3–19. doi: 10.7555/JBR.29.20140151.

Miettinen, M. (ed.) (2016) *Modern Soft Tissue Pathology: Tumors and Non-Neoplastic Conditions*. 2nd edn. Cambridge: Cambridge University Press. Available at: <https://www.cambridge.org/core/books/modern-soft-tissue-pathology/716C3345EDB53691CEE30DB640C84E7E> (Accessed: 22 July 2020).

Miettinen, M. and Lasota, J. (2006) 'Gastrointestinal stromal tumors: Pathology and prognosis at different sites', *Seminars in Diagnostic Pathology*, 23(2), pp. 70–83. doi: 10.1053/j.semdp.2006.09.001.

Mitelman, F., Johansson, B. and Mertens, F. (2007) 'The impact of translocations and gene fusions on cancer causation', *Nature Reviews Cancer*, 7(4), pp. 233–245. doi: 10.1038/nrc2091.

Moasser, M. M. (2007) 'The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis', *Oncogene*, 26(45), pp. 6469–6487. doi: 10.1038/sj.onc.1210477.

Molchadsky, A. *et al.* (2008) 'p53 Plays a Role in Mesenchymal Differentiation Programs, in a Cell Fate Dependent Manner', *PLOS ONE*, 3(11), p. e3707. doi: 10.1371/journal.pone.0003707.

Momiyama, T. *et al.* (1998) 'Functional Involvement of Serum Response Factor in the Transcriptional Regulation of Caldesmon Gene', *Biochemical and Biophysical Research Communications*, 242(2), pp. 429–435. doi: 10.1006/bbrc.1997.7959.

Mongroo, P. S. and Rustgi, A. K. (2010) 'The role of the miR-200 family in epithelial-mesenchymal transition', *Cancer Biology & Therapy*, 10(3), pp. 219–222. doi: 10.4161/cbt.10.3.12548.

Moore, L. *et al.* (2018) 'The mutational landscape of normal human endometrial epithelium', *bioRxiv*, p. 505685. doi: 10.1101/505685.

Morin, R. D. *et al.* (2010) 'Somatic mutation of EZH2 (Y641) in Follicular and Diffuse Large B-cell Lymphomas of Germinal Center Origin', *Nature genetics*, 42(2), pp. 181–185. doi: 10.1038/ng.518.

- Nakano, K. and Takahashi, S. (2018a) 'Current Molecular Targeted Therapies for Bone and Soft Tissue Sarcomas', *International Journal of Molecular Sciences*, 19(3). doi: 10.3390/ijms19030739.
- Nakano, K. and Takahashi, S. (2018b) 'Translocation-Related Sarcomas', *International Journal of Molecular Sciences*, 19(12). doi: 10.3390/ijms19123784.
- Nemenoff Raphael A. *et al.* (2011) 'SDF-1 α Induction in Mature Smooth Muscle Cells by Inactivation of PTEN Is a Critical Mediator of Exacerbated Injury-Induced Neointima Formation', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 31(6), pp. 1300–1308. doi: 10.1161/ATVBAHA.111.223701.
- Newman, A. M. *et al.* (2015) 'Robust enumeration of cell subsets from tissue expression profiles', *Nature Methods*, 12(5), pp. 453–457. doi: 10.1038/nmeth.3337.
- Ng, C. K. *et al.* (2012) 'The role of tandem duplicator phenotype in tumour evolution in high-grade serous ovarian cancer', *The Journal of Pathology*, 226(5), pp. 703–712. doi: 10.1002/path.3980.
- Nielsen, T. O. *et al.* (2002) 'Molecular characterisation of soft tissue tumours: a gene expression study', *The Lancet*, 359(9314), pp. 1301–1307. doi: 10.1016/S0140-6736(02)08270-3.
- Nikolic, I. *et al.* (2017) 'Discovering cancer vulnerabilities using high-throughput micro-RNA screening', *Nucleic Acids Research*, 45(22), pp. 12657–12670. doi: 10.1093/nar/gkx1072.
- Nittner, D. *et al.* (2012) 'Synthetic lethality between Rb, p53 and Dicer or miR-17–92 in retinal progenitors suppresses retinoblastoma formation', *Nature Cell Biology*, 14(9), pp. 958–965. doi: 10.1038/ncb2556.
- Norman, C. *et al.* (1988) 'Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element', *Cell*, 55(6), pp. 989–1003. doi: 10.1016/0092-8674(88)90244-9.
- O'Donnell, J. S. *et al.* (2018) 'PI3K-AKT-mTOR inhibition in cancer immunotherapy, redux', *Seminars in Cancer Biology*, 48, pp. 91–103. doi: 10.1016/j.semcancer.2017.04.015.
- Ohue, Y. and Nishikawa, H. (2019) 'Regulatory T (Treg) cells in cancer: Can Treg cells be a new therapeutic target?', *Cancer Science*, 110(7), pp. 2080–2089. doi: 10.1111/cas.14069.
- Okano, M. *et al.* (1999) 'DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development', *Cell*, 99(3), pp. 247–257. doi: 10.1016/S0092-8674(00)81656-6.
- Oldford, S. A. *et al.* (2010) 'A Critical Role for Mast Cells and Mast Cell-Derived IL-6 in TLR2-Mediated Inhibition of Tumor Growth', *The Journal of Immunology*, 185(11), pp. 7067–7076. doi: 10.4049/jimmunol.1001137.

- Oldford, S. A. and Marshall, J. S. (2015) 'Mast cells as targets for immunotherapy of solid tumors', *Molecular Immunology*, 63(1), pp. 113–124. doi: 10.1016/j.molimm.2014.02.020.
- Oren, M. and Rotter, V. (2010) 'Mutant p53 Gain-of-Function in Cancer', *Cold Spring Harbor Perspectives in Biology*, 2(2). doi: 10.1101/cshperspect.a001107.
- Ostuni, R. *et al.* (2015) 'Macrophages and cancer: from mechanisms to therapeutic implications', *Trends in Immunology*, 36(4), pp. 229–239. doi: 10.1016/j.it.2015.02.004.
- Ott, V. L. *et al.* (2003) 'Mast cell-dependent migration of effector CD8 + T cells through production of leukotriene B 4', *Nature Immunology*, 4(10), pp. 974–981. doi: 10.1038/ni971.
- Owens, G. K., Kumar, M. S. and Wamhoff, B. R. (2004) 'Molecular Regulation of Vascular Smooth Muscle Cell Differentiation in Development and Disease', *Physiological Reviews*, 84(3), pp. 767–801. doi: 10.1152/physrev.00041.2003.
- Parker, J. S. *et al.* (2009) 'Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes', *Journal of Clinical Oncology*, 27(8), pp. 1160–1167. doi: 10.1200/JCO.2008.18.1370.
- Parmacek, M. S. (2009) 'MicroRNA-modulated targeting of vascular smooth muscle cells', *The Journal of Clinical Investigation*, 119(9), pp. 2526–2528. doi: 10.1172/JCI40503.
- Pedeutour, F. *et al.* (1994) 'Complex composition and co-amplification of SAS and MDM2 in ring and giant rod marker chromosomes in well-differentiated liposarcoma', *Genes, Chromosomes and Cancer*, 10(2), pp. 85–94. doi: 10.1002/gcc.2870100203.
- Pedeutour, F. *et al.* (1996) 'Translocation, t(17;22)(q22;q13), in dermatofibrosarcoma protuberans: a new tumor-associated chromosome rearrangement', *Cytogenetics and cell genetics*, 72(2–3), pp. 171–174. doi: 10.1159/000134178.
- Pelizzola, M. and Ecker, J. R. (2011) 'The DNA methylome', *FEBS letters*, 585(13), pp. 1994–2000. doi: 10.1016/j.febslet.2010.10.061.
- Pérot, G. *et al.* (2009) 'Strong Smooth Muscle Differentiation Is Dependent on Myocardin Gene Amplification in Most Human Retroperitoneal Leiomyosarcomas', *Cancer Research*, 69(6), pp. 2269–2278. doi: 10.1158/0008-5472.CAN-08-1443.
- Pérot, G. *et al.* (2010) 'Constant p53 Pathway Inactivation in a Large Series of Soft Tissue Sarcomas with Complex Genetics', *The American Journal of Pathology*, 177(4), pp. 2080–2090. doi: 10.2353/ajpath.2010.100104.
- Pérot, G. *et al.* (2014) 'Smooth muscle differentiation identifies two classes of poorly differentiated pleomorphic sarcomas with distinct outcome', *Modern Pathology*, 27(6), pp. 840–850. doi: 10.1038/modpathol.2013.205.
- Phillips, D. H. (2018) 'Mutational spectra and mutational signatures: Insights into

cancer aetiology and mechanisms of DNA damage and repair', *DNA Repair*, 71, pp. 6–11. doi: 10.1016/j.dnarep.2018.08.003.

Pladevall-Morera, D. *et al.* (2019) 'Proteomic characterization of chromosomal common fragile site (CFS)-associated proteins uncovers ATRX as a regulator of CFS stability', *Nucleic Acids Research*, 47(15), pp. 8004–8018. doi: 10.1093/nar/gkz510.

Pouget, C., Pottin, K. and Jaffredo, T. (2008) 'Sclerotomal origin of vascular smooth muscle cells and pericytes in the embryo', *Developmental Biology*, 315(2), pp. 437–447. doi: 10.1016/j.ydbio.2007.12.045.

Prins, K. W. *et al.* (2009) 'Dystrophin is a microtubule-associated protein', *Journal of Cell Biology*, 186(3), pp. 363–369. doi: 10.1083/jcb.200905048.

Qin, T. *et al.* (2009) 'Mechanisms of resistance to 5-aza-2'-deoxycytidine in human cancer cell lines', *Blood*, 113(3), pp. 659–667. doi: 10.1182/blood-2008-02-140038.

Qin, T. *et al.* (2011) 'Mechanisms of Resistance to Decitabine in the Myelodysplastic Syndrome', *PLOS ONE*, 6(8), p. e23372. doi: 10.1371/journal.pone.0023372.

Qiu, L. *et al.* (2019) 'Cantharidin Attenuates the Proliferation and Migration of Vascular Smooth Muscle Cells through Suppressing Inflammatory Response', *Biological and Pharmaceutical Bulletin*, 42(1), pp. 34–42. doi: 10.1248/bpb.b18-00462.

Qiu Ping and Li Li (2002) 'Histone Acetylation and Recruitment of Serum Responsive Factor and CREB-Binding Protein Onto SM22 Promoter During SM22 Gene Expression', *Circulation Research*, 90(8), pp. 858–865. doi: 10.1161/01.RES.0000016504.08608.B9.

Rausch, T. *et al.* (2012) 'Genome Sequencing of Pediatric Medulloblastoma Links Catastrophic DNA Rearrangements with TP53 Mutations', *Cell*, 148(1), pp. 59–71. doi: 10.1016/j.cell.2011.12.013.

Regan, C. P. *et al.* (2000) 'Molecular mechanisms of decreased smooth muscle differentiation marker expression after vascular injury', *The Journal of Clinical Investigation*, 106(9), pp. 1139–1147. doi: 10.1172/JCI10522.

Rehman, S. and Ahmed, D. (2020) 'Embryology, Kidney, Bladder, and Ureter', in *StatPearls*. Treasure Island (FL): StatPearls Publishing. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK547747/> (Accessed: 24 July 2020).

Ren, B. *et al.* (2003) 'Gene expression analysis of human soft tissue leiomyosarcomas', *Human Pathology*, 34(6), pp. 549–558. doi: 10.1016/S0046-8177(03)00014-5.

Ren, X. *et al.* (2018) 'Alternative lengthening of telomeres phenotype and loss of ATRX expression in sarcomas', *Oncology Letters*, 15(5), pp. 7489–7496. doi: 10.3892/ol.2018.8318.

Rensen, S. S. M., Doevendans, P. A. F. M. and van Eys, G. J. J. M. (2007) 'Regulation

and characteristics of vascular smooth muscle cell phenotypic diversity', *Netherlands Heart Journal*, 15(3), pp. 100–108.

Rentschler, S. *et al.* (1999) 'The WW Domain of Dystrophin Requires EF-Hands Region to Interact with β -Dystroglycan', *Biological Chemistry*, 380(4), pp. 431–442. doi: 10.1515/BC.1999.057.

Robbins, P. F. *et al.* (2011) 'Tumor Regression in Patients With Metastatic Synovial Cell Sarcoma and Melanoma Using Genetically Engineered Lymphocytes Reactive With NY-ESO-1', *Journal of Clinical Oncology*, 29(7), pp. 917–924. doi: 10.1200/JCO.2010.32.2537.

Robin, Y.-M. *et al.* (2013) 'Transgelin is a novel marker of smooth muscle differentiation that improves diagnostic accuracy of leiomyosarcomas: a comparative immunohistochemical reappraisal of myogenic markers in 900 soft tissue tumors', *Modern Pathology*, 26(4), pp. 502–510. doi: 10.1038/modpathol.2012.192.

Rossi, F. *et al.* (2006) 'Oncogenic Kit signaling and therapeutic intervention in a mouse model of gastrointestinal stromal tumor', *Proceedings of the National Academy of Sciences of the United States of America*, 103(34), p. 12843. doi: 10.1073/pnas.0511076103.

Rubio, R. *et al.* (2010) 'Deficiency in p53 but not Retinoblastoma Induces the Transformation of Mesenchymal Stem Cells In vitro and Initiates Leiomyosarcoma In vivo', *Cancer Research*, 70(10), pp. 4185–4194. doi: 10.1158/0008-5472.CAN-09-4640.

Rupaimoole, R. and Slack, F. J. (2017) 'MicroRNA therapeutics: towards a new era for the management of cancer and other diseases', *Nature Reviews Drug Discovery*, 16(3), pp. 203–222. doi: 10.1038/nrd.2016.246.

Rzucidlo, E. M., Martin, K. A. and Powell, R. J. (2007) 'Regulation of vascular smooth muscle cell differentiation', *Journal of Vascular Surgery*, 45(6), pp. A25–A32. doi: 10.1016/j.jvs.2007.03.001.

Saga, H. *et al.* (1999) 'Phenotype-Dependent Expression of alpha-Smooth Muscle Actin in Visceral Smooth Muscle Cells', *Experimental Cell Research*, 247, pp. 279–292. doi: 10.1006/excr.1998.4339.

Sahli, N. *et al.* (2016) 'Unusual evolution of leiomyosarcoma of the rectum: a case report and review of the literature', *Journal of Medical Case Reports*, 10(1), pp. 1–4. doi: 10.1186/s13256-016-1047-8.

Saiag, P. *et al.* (2015) 'Diagnosis and treatment of dermatofibrosarcoma protuberans. European consensus-based interdisciplinary guideline', *European Journal of Cancer*, 51(17), pp. 2604–2608. doi: 10.1016/j.ejca.2015.06.108.

Sakkal, S., Miller, S. and Nurgali, V. A. and K. (2016) 'Eosinophils in Cancer: Favourable or Unfavourable?', *Current Medicinal Chemistry*, 7(23), pp. 650–666. doi: 10.2174/0929867323666160119094313.

- Salamon, P. *et al.* (2005) 'Human mast cells release Interleukin-8 and induce neutrophil chemotaxis on contact with activated T cells', *Allergy*, 60(10), pp. 1316–1319. doi: 10.1111/j.1398-9995.2005.00886.x.
- Sanders, K. M., Ward, S. M. and Koh, S. D. (2014) 'Interstitial Cells: Regulators of Smooth Muscle Function', *Physiological Reviews*, 94(3), pp. 859–907. doi: 10.1152/physrev.00037.2013.
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) 'DNA sequencing with chain-terminating inhibitors', *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), pp. 5463–5467.
- Sansregret, L., Vanhaesebroeck, B. and Swanton, C. (2018) 'Determinants and clinical implications of chromosomal instability in cancer', *Nature Reviews Clinical Oncology*, 15(3), pp. 139–150. doi: 10.1038/nrclinonc.2017.198.
- Saphirstein, R. J. and Morgan, K. G. (2014) 'The Contribution of Vascular Smooth Muscle to Aortic Stiffness Across Length Scales', *Microcirculation*, 21(3), pp. 201–207. doi: 10.1111/micc.12101.
- Sarbassov, D. D. *et al.* (2006) 'Prolonged Rapamycin Treatment Inhibits mTORC2 Assembly and Akt/PKB', *Molecular Cell*, 22(2), pp. 159–168. doi: 10.1016/j.molcel.2006.03.029.
- Schotta, G. *et al.* (2004) 'A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin', *Genes & Development*, 18(11), pp. 1251–1262. doi: 10.1101/gad.300704.
- Schuster-Böckler, B. and Lehner, B. (2012) 'Chromatin organization is a major influence on regional mutation rates in human cancer cells', *Nature*, 488(7412), pp. 504–507. doi: 10.1038/nature11273.
- Scirocco, A. *et al.* (2016) 'Cellular and Molecular Mechanisms of Phenotypic Switch in Gastrointestinal Smooth Muscle', *Journal of Cellular Physiology*, 231(2), pp. 295–302. doi: 10.1002/jcp.25105.
- Segal, N. H. *et al.* (2003) 'Classification and Subtype Prediction of Adult Soft Tissue Sarcoma by Functional Genomics', *The American Journal of Pathology*, 163(2), pp. 691–700. doi: 10.1016/S0002-9440(10)63696-6.
- Serrano, C. and George, S. (2013) 'Leiomyosarcoma', *Hematology/Oncology Clinics of North America*, 27(5), pp. 957–974. doi: 10.1016/j.hoc.2013.07.002.
- Setlow, R. B. and Carrier, W. L. (1966) 'Pyrimidine dimers in ultraviolet-irradiated DNA's', *Journal of Molecular Biology*, 17(1), pp. 237–254. doi: 10.1016/S0022-2836(66)80105-5.
- Shankman, L. S. *et al.* (2015) 'KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis', *Nature Medicine*,

21(6), pp. 628–637. doi: 10.1038/nm.3866.

Sharma, S., Kelly, T. K. and Jones, P. A. (2010) 'Epigenetics in cancer', *Carcinogenesis*, 31(1), pp. 27–36. doi: 10.1093/carcin/bgp220.

Shaul, M. E. and Fridlender, Z. G. (2019) 'Tumour-associated neutrophils in patients with cancer', *Nature Reviews Clinical Oncology*, 16(10), pp. 601–620. doi: 10.1038/s41571-019-0222-4.

Shen, W. H. *et al.* (2007) 'Essential Role for Nuclear PTEN in Maintaining Chromosomal Integrity', *Cell*, 128(1), pp. 157–170. doi: 10.1016/j.cell.2006.11.042.

Shi, J. *et al.* (2020) 'Metabolism of vascular smooth muscle cells in vascular diseases', *American Journal of Physiology-Heart and Circulatory Physiology*, 319(3), pp. H613–H631. doi: 10.1152/ajpheart.00220.2020.

Shimizu, A. *et al.* (1999) 'The Dermatofibrosarcoma Protuberans-associated Collagen Type I α 1/Platelet-derived Growth Factor (PDGF) B-Chain Fusion Gene Generates a Transforming Protein That Is Processed to Functional PDGF-BB', *Cancer Research*, 59(15), pp. 3719–3723.

Shinohara, Sho *et al.* (2012) 'Regulation of Differentiated Phenotypes of Vascular Smooth Muscle Cells', *Current Basic and Pathological Approaches to the Function of Muscle Cells and Tissues - From Molecules to Humans*. doi: 10.5772/48573.

Shynlova, O. *et al.* (2009) 'Integration of endocrine and mechanical signals in the regulation of myometrial functions during pregnancy and labour', *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 144, pp. S2–S10. doi: 10.1016/j.ejogrb.2009.02.044.

Skytting, B. *et al.* (1999) 'A Novel Fusion Gene, SYT-SSX4, in Synovial Sarcoma', *JNCI: Journal of the National Cancer Institute*, 91(11), pp. 974–975. doi: 10.1093/jnci/91.11.974.

Sørli, T. *et al.* (2001) 'Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications', *Proceedings of the National Academy of Sciences*, 98(19), pp. 10869–10874. doi: 10.1073/pnas.191367098.

Spain, B. H. *et al.* (1999) 'Truncated BRCA2 is cytoplasmic: Implications for cancer-linked mutations', *Proceedings of the National Academy of Sciences of the United States of America*, 96(24), pp. 13920–13925.

Speer Mei Y. *et al.* (2009) 'Smooth Muscle Cells Give Rise to Osteochondrogenic Precursors and Chondrocytes in Calcifying Arteries', *Circulation Research*, 104(6), pp. 733–741. doi: 10.1161/CIRCRESAHA.108.183053.

Stefano, S. and Giovanni, S. (2019) 'The PTEN Tumor Suppressor Gene in Soft Tissue Sarcoma', *Cancers*, 11(8). doi: 10.3390/cancers11081169.

Stephens, P. J. *et al.* (2009) 'Complex landscapes of somatic rearrangement in human

breast cancer genomes', *Nature*, 462(7276), pp. 1005–1010. doi: 10.1038/nature08645.

Stiles, B. *et al.* (2002) 'Essential Role of AKT-1/Protein Kinase B α in PTEN-Controlled Tumorigenesis', *Molecular and Cellular Biology*, 22(11), pp. 3842–3851. doi: 10.1128/MCB.22.11.3842-3851.2002.

Storchová, Z. *et al.* (2006) 'Genome-wide genetic analysis of polyploidy in yeast', *Nature*, 443(7111), pp. 541–547. doi: 10.1038/nature05178.

Straub, V. *et al.* (1999) ' ϵ -Sarcoglycan Replaces α -Sarcoglycan in Smooth Muscle to Form a Unique Dystrophin-Glycoprotein Complex', *Journal of Biological Chemistry*, 274(39), pp. 27989–27996. doi: 10.1074/jbc.274.39.27989.

Sun, S. *et al.* (2011) 'miR-146a and Krüppel-like factor 4 form a feedback loop to participate in vascular smooth muscle cell proliferation', *EMBO reports*, 12(1), pp. 56–62. doi: 10.1038/embor.2010.172.

Supek, F. and Lehner, B. (2017) 'Clustered Mutation Signatures Reveal that Error-Prone DNA Repair Targets Mutations to Active Genes', *Cell*, 170(3), pp. 534–547.e23. doi: 10.1016/j.cell.2017.07.003.

Szklarczyk, D. *et al.* (2019) 'STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets', *Nucleic Acids Research*, 47(D1), pp. D607–D613. doi: 10.1093/nar/gky1131.

Takahashi, R. *et al.* (2013) 'Phase II study of personalized peptide vaccination for refractory bone and soft tissue sarcoma patients', *Cancer Science*, 104(10), pp. 1285–1294. doi: 10.1111/cas.12226.

Tan, Sichuang *et al.* (2016) 'Knocking down Dp71 expression in A549 cells reduces its malignancy in vivo and in vitro', *Cancer Investigation*, 34(1), pp. 16–25. doi: 10.3109/07357907.2015.1084002.

Tanfin, Z. *et al.* (2011) 'Endothelin-1: Physiological and pathological roles in myometrium', *The International Journal of Biochemistry & Cell Biology*, 43(3), pp. 299–302. doi: 10.1016/j.biocel.2010.10.009.

Tang, R. *et al.* (2008) 'Myocardin inhibits cellular proliferation by inhibiting NF- κ B(p65)-dependent cell cycle progression', *Proceedings of the National Academy of Sciences of the United States of America*, 105(9), p. 3362. doi: 10.1073/pnas.0705842105.

Tanić, M. *et al.* (2018) 'Molecular signature of response to preoperative radiotherapy in locally advanced breast cancer', *Radiation Oncology*, 13(1), pp. 1–10. doi: 10.1186/s13014-018-1129-4.

Tap, W. D. *et al.* (2016) 'Olaratumab and doxorubicin versus doxorubicin alone for treatment of soft-tissue sarcoma: an open-label phase 1b and randomised phase 2 trial', *The Lancet*, 388(10043), pp. 488–497. doi: 10.1016/S0140-6736(16)30587-6.

Tawbi, H. A. *et al.* (2017) 'Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial', *The Lancet Oncology*, 18(11), pp. 1493–1501. doi: 10.1016/S1470-2045(17)30624-1.

Taylor, B. S. *et al.* (2011) 'Advances in sarcoma genomics and new therapeutic targets', *Nature Reviews. Cancer*, 11(8), pp. 541–557. doi: 10.1038/nrc3087.

Templeton, A. J. *et al.* (2014) 'Prognostic role of neutrophil-to-lymphocyte ratio in solid tumors: a systematic review and meta-analysis', *Journal of the National Cancer Institute*, 106(6), p. dju124. doi: 10.1093/jnci/dju124.

The Cancer Genome Atlas Program - National Cancer Institute (2018). Available at: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga> (Accessed: 21 July 2020).

The Cancer Genome Atlas Research Network *et al.* (2017) 'Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas', *Cell*, 171(4), pp. 950-965.e28. doi: 10.1016/j.cell.2017.10.014.

The ICGC/TCGA Pan-Cancer (2020) 'Pan-cancer analysis of whole genomes', *Nature*, 578(7793), pp. 82–93. doi: 10.1038/s41586-020-1969-6.

Thorsson, V. *et al.* (2018) 'The Immune Landscape of Cancer', *Immunity*, 48(4), pp. 812–830.e14. doi: 10.1016/j.immuni.2018.03.023.

Tirode, F. *et al.* (2007) 'Mesenchymal Stem Cell Features of Ewing Tumors', *Cancer Cell*, 11(5), pp. 421–429. doi: 10.1016/j.ccr.2007.02.027.

Totoki, Y. *et al.* (2011) 'High-resolution characterization of a hepatocellular carcinoma genome', *Nature Genetics*, 43(5), pp. 464–469. doi: 10.1038/ng.804.

Toulmonde, M. *et al.* (2014) 'Retroperitoneal sarcomas: patterns of care at diagnosis, prognostic factors and focus on main histological subtypes: a multicenter analysis of the French Sarcoma Group', *Annals of Oncology*, 25(3), pp. 735–742. doi: 10.1093/annonc/mdt577.

Townsend Raymond R. *et al.* (2015) 'Recommendations for Improving and Standardizing Vascular Research on Arterial Stiffness', *Hypertension*, 66(3), pp. 698–722. doi: 10.1161/HYP.0000000000000033.

Tsang, D. P. and Cheng, A. S. (2011) 'Epigenetic regulation of signaling pathways in cancer: Role of the histone methyltransferase EZH2', *Journal of Gastroenterology and Hepatology*, 26(1), pp. 19–27. doi: 10.1111/j.1440-1746.2010.06447.x.

Tsumura, H. *et al.* (2006) 'Cooperation of oncogenic K-ras and p53 deficiency in pleomorphic rhabdomyosarcoma development in adult mice', *Oncogene*, 25(59), pp. 7673–7679. doi: 10.1038/sj.onc.1209749.

Tsushima, K. *et al.* (1987) 'Leiomyosarcomas and Benign Smooth Muscle Tumors of the

Stomach: Nuclear DNA Patterns Studied by Flow Cytometry', *Mayo Clinic Proceedings*, 62(4), pp. 275–280. doi: 10.1016/S0025-6196(12)61904-1.

Turczyńska, K. M. *et al.* (2015) 'Regulation of Smooth Muscle Dystrophin and Synaptopodin 2 Expression by Actin Polymerization and Vascular Injury', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 35(6), pp. 1489–1497. doi: 10.1161/ATVBAHA.114.305065.

Turner, E. C. *et al.* (2013) 'Identification of a Klf4-dependent upstream repressor region mediating transcriptional regulation of the myocardin gene in human smooth muscle cells', *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1829(11), pp. 1191–1201. doi: 10.1016/j.bbagr.2013.09.002.

Valle-García, D. *et al.* (2016) 'ATRX binds to atypical chromatin domains at the 3' exons of zinc finger genes to preserve H3K9me3 enrichment', *Epigenetics*, 11(6), pp. 398–414. doi: 10.1080/15592294.2016.1169351.

Varricchi, G. *et al.* (2017) 'Are Mast Cells MASTers in Cancer?', *Frontiers in Immunology*, 8. doi: 10.3389/fimmu.2017.00424.

Venter, J. C. *et al.* (2001) 'The Sequence of the Human Genome', *Science*, 291(5507), pp. 1304–1351. doi: 10.1126/science.1058040.

Villarreal-Silva, M. *et al.* (2011) 'Knockdown of Dystrophin Dp71 Impairs PC12 Cells Cycle: Localization in the Spindle and Cytokinesis Structures Implies a Role for Dp71 in Cell Division', *PLOS ONE*, 6(8), p. e23504. doi: 10.1371/journal.pone.0023504.

Waks, A. G. and Winer, E. P. (2019) 'Breast Cancer Treatment: A Review', *JAMA*, 321(3), p. 288. doi: 10.1001/jama.2018.19323.

Wang, D. and Atanasov, A. G. (2019) 'The microRNAs Regulating Vascular Smooth Muscle Cell Proliferation: A Minireview', *International Journal of Molecular Sciences*, 20(2). doi: 10.3390/ijms20020324.

Wang, G. *et al.* (2015) 'Origin and differentiation of vascular smooth muscle cells', *The Journal of Physiology*, 593(Pt 14), pp. 3013–3030. doi: 10.1113/JP270033.

Wang, H. *et al.* (2015) 'miR-146b-5p promotes VSMC proliferation and migration', *International Journal of Clinical and Experimental Pathology*, 8(10), pp. 12901–12907.

Wang, Y. *et al.* (2014) 'Dystrophin is a tumor suppressor in human cancers with myogenic programs', *Nature Genetics*, 46(6), pp. 601–606. doi: 10.1038/ng.2974.

Wang, Z. *et al.* (2004) 'Myocardin and ternary complex factors compete for SRF to control smooth muscle gene expression', *Nature*, 428(6979), pp. 185–189. doi: 10.1038/nature02382.

Wangsa, D. *et al.* (2018) 'Near-tetraploid cancer cells show chromosome instability triggered by replication stress and exhibit enhanced invasiveness', *The FASEB Journal*,

32(7), pp. 3502–3517. doi: 10.1096/fj.201700247RR.

Watson, L. A., Goldberg, H. and Bérubé, N. G. (2015) 'Emerging roles of ATRX in cancer', *Epigenomics*, 7(8), pp. 1365–1378. doi: 10.2217/epi.15.82.

Watson, S. *et al.* (2018) 'Transcriptomic definition of molecular subgroups of small round cell sarcomas', *The Journal of Pathology*, 245(1), pp. 29–40. doi: 10.1002/path.5053.

Way, M. *et al.* (1992) 'Expression of the N-terminal domain of dystrophin in E. coli and demonstration of binding to F-actin', *FEBS Letters*, 301(3), pp. 243–245. doi: 10.1016/0014-5793(92)80249-G.

Wei, G. *et al.* (2006) 'Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance', *Cancer Cell*, 10(4), pp. 331–342. doi: 10.1016/j.ccr.2006.09.006.

WHO Classification of Tumours Editorial Board (2020) *Soft Tissue and Bone Tumours*. 5th edn. Available at: <https://publications.iarc.fr/Book-And-Report-Series/Who-Classification-Of-Tumours/Soft-Tissue-And-Bone-Tumours-2020> (Accessed: 1 September 2020).

Wouters, M. C. A. and Nelson, B. H. (2018) 'Prognostic Significance of Tumor-Infiltrating B Cells and Plasma Cells in Human Cancer', *Clinical Cancer Research*, 24(24), pp. 6125–6135. doi: 10.1158/1078-0432.CCR-18-1481.

Xie Changqing *et al.* (2011) 'Smooth Muscle Cell Differentiation In Vitro', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 31(7), pp. 1485–1494. doi: 10.1161/ATVBAHA.110.221101.

Xie Wei-Bing *et al.* (2013) 'Smad2 and Myocardin-Related Transcription Factor B Cooperatively Regulate Vascular Smooth Muscle Differentiation From Neural Crest Cells', *Circulation Research*, 113(8), pp. e76–e86. doi: 10.1161/CIRCRESAHA.113.301921.

Xin, M. *et al.* (2009) 'MicroRNAs miR-143 and miR-145 modulate cytoskeletal dynamics and responsiveness of smooth muscle cells to injury', *Genes & Development*, 23(18), pp. 2166–2178. doi: 10.1101/gad.1842409.

Yang, C.-Y. *et al.* (2015) 'Targeted next-generation sequencing of cancer genes identified frequent TP53 and ATRX mutations in leiomyosarcoma', *American Journal of Translational Research*, 7(10), pp. 2072–2081.

Yang, X. *et al.* (2013) 'Spry1 and Spry4 Differentially Regulate Human Aortic Smooth Muscle Cell Phenotype via Akt/FoxO/Myocardin Signaling', *PLOS ONE*, 8(3), p. e58746. doi: 10.1371/journal.pone.0058746.

Yu, F. *et al.* (2007) 'let-7 Regulates Self Renewal and Tumorigenicity of Breast Cancer Cells', *Cell*, 131(6), pp. 1109–1123. doi: 10.1016/j.cell.2007.10.054.

Zhou, N. *et al.* (2017) 'Inhibition of SRF/myocardin reduces aortic stiffness by targeting vascular smooth muscle cell stiffening in hypertension', *Cardiovascular Research*, 113(2), p. 171. doi: 10.1093/cvr/cvw222.

AUTEUR : Lucie Darmusey

TITRE : Caractérisation d'axes *drivers* de l'oncogenèse des léiomyosarcomes : validations fonctionnelles et cibles thérapeutiques

DIRECTEUR DE THESE : Dr. Frédéric Chibon

Les sarcomes à génomique complexe et plus particulièrement les léiomyosarcomes (LMS), qui ont une différenciation musculaire lisse, sont des tumeurs rares et agressives dont l'oncogenèse reste aujourd'hui incomprise. Afin de mieux appréhender cette oncogenèse et de mettre en évidence ces *drivers*, une intégration de plusieurs OMICS et des données cliniques d'une cohorte de 120 LMS a été réalisée. Cela permet de séparer les LMS en deux grands groupes cliniquement différents : hLMS et oLMS. Dans ce dernier groupe ainsi que dans les LMS utérins, un des gènes les plus altérés est *ATRX*. Par l'étude de l'impact de cette altération dans des modèles murins, je montre qu'elle entraîne une croissance plus rapide des tumeurs et un échappement au système immunitaire. A l'opposé des oLMS, les hLMS ont un transcriptome homogène, ils sont aussi plus agressifs, se situent au niveau du tronc et du rétropéritoine, ont des marques d'histones des muscles lisses et des gènes régulés par des facteurs de transcription spécifiques des muscles lisses comme SRF. De plus, les hLMS ont des altérations spécifiques notamment la surexpression de *MYOCD*, le cofacteur de SRF dans les muscles lisses, et la perte de *PTEN*, capable de limiter la fixation SRF à ses gènes cibles de la différenciation des muscles lisses. J'ai alors démontré que l'inhibition de l'interaction entre SRF et *MYOCD* entraîne la mort par apoptose des cellules de hLMS plus fortement que celles de oLMS, démontrant ainsi que l'activation de cet axe est indispensable à la survie des hLMS, cela en fait une cible thérapeutique prometteuse. Pour mieux comprendre comment l'activation pathologique de ce complexe dirige l'oncogenèse et comment ces cellules en deviennent dépendantes, j'ai mis au point une analyse par ChIPseq pour étudier, dans le contexte des LMS, les cibles de la fixation du complexe SRF/*MYOCD* dont l'analyse est en cours. L'ensemble de ces travaux apporte de nouvelles voies thérapeutiques dont l'étude préclinique doit maintenant se poursuivre *in-vivo*.

Mots Clés : Léiomyosarcomes, oncogenèse, génomique, transcriptomique, analyses fonctionnelles

Genomically complex sarcomas and more specifically leiomyosarcomas (LMS), which have a smooth muscle differentiation, are rare and aggressive tumors whose oncogenesis is still misunderstood. In order to comprehend this oncogenesis, an integration of multiple OMICS with clinical data of a cohort of 120 LMS was performed. This allowed to split LMSs into 2 groups clinically distinct: hLMS and oLMS. In this last group as well as in uterine LMS one of the most altered gene is *ATRX*. By studying the impacts of this alteration in mice models, I show that it leads to a higher growth rate and an escape to the immune system. In contrast to oLMSs, hLMSs have homogenous transcriptomes, they are more aggressive, are in the trunk or the retroperitoneum, have histone marks of smooth muscles and express genes regulated by specific transcription factors of smooth muscle cell as SRF. Furthermore, hLMSs have specific alterations such as *MYOCD* overexpression, the co-factor of SRF into smooth muscle, and the loss of *PTEN*, able to restrict SRF fixation to its target genes of smooth muscle differentiation. Then, I could demonstrate that the inhibition of SRF/*MYOCD* interaction kills by apoptosis hLMS cells more than oLMS ones demonstrating that the activation of this axe is necessary to hLMS survival, making it an encouraging therapeutic target. To better understand how the pathological activation of this complex is leading the oncogenesis and how this cells become dependents of it, I developed a ChIPseq analysis to study, in the context of LMS, the targets of SRF/*MYOCD* fixation which analysis is ongoing. Together those works bring new therapeutic paths which preclinical studies must now continue *in vivo*.

Key words: Leiomyosarcomas, oncogenesis, genomic, transcriptomic, functional analysis

Equipe 19 - OncoSarc
Centre de Recherche en Cancérologie de Toulouse
2 Avenue Hubert Curien
31037 Toulouse
