Remodelage génomique des sarcomes pléomorphes : caractérisation transcriptomique et agressivité tumorale

Tom Lesluyes

université de BORDEAUX

THÈSE PRÉSENTÉE POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

SPÉCIALITÉ BIOINFORMATIQUE

PAR TOM LESLUYES

REMODELAGE GÉNOMIQUE DES SARCOMES PLÉOMORPHES : CARACTÉRISATION TRANSCRIPTOMIQUE ET AGRESSIVITÉ TUMORALE

SOUS LA DIRECTION DE : JEAN-MICHEL COINDRE

Soutenue publiquement le 24 Mai 2019 devant le jury composé de :

Pr. Sophie JAVERZAT	Université de Bordeaux	Présidente
Dr. Alain VIARI	INRIA Grenoble-Rhône-Alpes, Grenoble	Rapporteur
Dr. Aurélien DE REYNIÈS	La Ligue contre le cancer, Paris	Rapporteur
Dr. Françoise RÉDINI	INSERM UMR1238, Nantes	Examinatrice
Pr. Florence DUFFAUD	Hôpital de la Timone, Marseille	Examinatrice
Pr. Jean-Michel COINDRE	Université de Bordeaux	Directeur de thèse
Dr. Frédéric CHIBON	INSERM UMR1037, Toulouse	Invité

Résumé / Abstract

Remodelage génomique des sarcomes pléomorphes : caractérisation transcriptomique et agressivité tumorale

Résumé : Les sarcomes pléomorphes sont des tumeurs mésenchymateuses rares caractérisées par de nombreux remaniements chromosomiques. Leur processus d'oncogenèse reste encore mal compris, aucune altération génétique motrice de l'initiation tumorale n'a pu être identifiée de façon récurrente et spécifique à ce jour. Les travaux que j'ai réalisés durant ma thèse avaient pour but de mieux comprendre la biologie de ces tumeurs, notamment les conséquences transcriptomiques de leur remodelage génomique. Nous avons caractérisé les transcrits de fusion exprimés dans ces tumeurs par séquençage haut-débit (RNA-seq). Ceci nous a amené à identifier l'expression de plusieurs transcrits chimériques impliquant le gène *TRIO* (5,1% des tumeurs). De plus, cette analyse ainsi que l'identification de variants exprimés nous ont permis d'identifier de fréquentes mutations de gènes suppresseurs de tumeurs tels qu'*ATRX* (16% des tumeurs) et plus généralement des membres du complexe SWI/SNF (47% des tumeurs). Les altérations de ce complexe majeur de remodelage de la chromatine sont associées à une plus grande instabilité génétique et à un phénotype plus agressif. Dans les sarcomes pléomorphes, l'instabilité génétique est liée à la progression tumorale via l'expression d'une signature transcriptomique pronostique. Cette signature, nommée CINSARC, est impliquée dans le contrôle de la mitose et de la ségrégation chromosomique. Nous avons voulu déterminer l'origine de cette expression via une étude intégrant la génomique et des mécanismes de régulation épigénétique, transcriptionnelle et post-transcriptionnelle. Si ces mécanismes ne semblent pas directement causals de l'expression de CINSARC, d'importantes modifications ont pu être mises en évidences. D'un point de vue clinique, nous avons démontré que l'expression de cette signature est un facteur pronostique universel de l'agressivité tumorale dans de nombreux types de cancers. De plus, CINSARC est un meilleur marqueur pronostique que le grade FNCLCC, actuel standard international d'évaluation du risque métastatique des sarcomes des tissus mous. Nous avons ainsi développé une méthode permettant une application clinique routinière de la signature CINSARC afin d'améliorer la prise en charge thérapeutique de ces tumeurs.

Mots-clés : sarcome, génome, transcriptome, métastase.

Genomic remodeling of pleomorphic sarcomas: transcriptomic characterization and tumor aggressiveness

Abstract: Pleomorphic sarcomas are rare mesenchymal tumors characterized by many chromosomal rearrangements. Their oncogenic process is still poorly understood, no recurrent and specific genetic alteration able to drive the tumor initiation has been identified yet. The work I did during my thesis had the objective to better understand the biology of these tumors, focusing on transcriptomic consequences of their genomic remodeling. We characterized fusion transcripts in these tumors by high-throughput sequencing (RNA-seq). This led us to identify the expression of several chimeric transcripts involving *TRIO* (5.1% of cases). Moreover, this analysis and the identification of expressed variants allowed us to identify frequent mutations of tumor suppressor genes such as *ATRX* (16% of cases) and more generally members of the SWI/SNF complex (47% of cases). Alterations of this major complex of chromatin remodeling are associated with higher genetic instability and more aggressive phenotype. In pleomorphic sarcomas, genetic instability is linked to tumor progression through the expression of a prognostic transcriptomic signature. This signature, termed CINSARC, is involved in mitosis control and chromosome segregation pathways. We wanted to determine the origin of such expression by integrating genomics and epigenetics, transcriptional and post-transcriptional regulation mechanisms. Though these mechanisms do not seem to directly regulate CINSARC expression, important changes have been highlighted. From a clinical standpoint, we demonstrated that the signature expression is a global prognostic factor of tumor aggressiveness in numerous cancer types. In addition, CINSARC is a better prognostic marker than the FNCLCC grading system, the current international standard to evaluate the metastatic risk in soft tissue sarcomas. We consequently developed a method allowing a daily clinical application of the CINSARC signature to improve the therapeutic management of these tumors.

Keywords: sarcoma, genome, transcriptome, metastasis.

Unités de recherche :

2016 - 2017
Génétique et Biologie des Sarcomes
INSERM U1218, Institut Bergonié
229 cours de l'Argonne, 33076 Bordeaux, France

2017 - 2019
Oncogenèse des Sarcomes
INSERM UMR1037, Centre de Recherches en Cancérologie de Toulouse
2 avenue Hubert Curien, 31037 Toulouse, France

« What more powerful form of study of mankind could there be than to read our own instruction book? »

Francis Collins

Remerciements

Je tiens tout d'abord à remercier l'ensemble des membres de mon jury pour avoir pris le temps d'évaluer ces travaux de thèse : Pr. Sophie Javerzat, présidente, Dr. Alain Viari et Dr. Aurélien De Reyniès, rapporteurs ainsi que Dr. Françoise Rédini et Pr. Florence Duffaud, examinatrices.

Je remercie chaleureusement l'école doctorale de Bordeaux : Antoine de Daruvar, son ancien président, pour m'avoir autorisé à m'inscrire en thèse, ainsi que le secrétariat : Catherine Ceyrac et Nathalie Vignais, pour leur précieuse aide administrative lors de mon inscription et du déménagement Bordeaux-Toulouse.

Fred, toute cette aventure n'aurait évidemment pas été possible sans ton aide ni ton soutien. Un immense merci pour m'avoir accordé ta confiance pendant ces six ans (avec sept contrats de travail signés, record à battre). Tu as su me diriger tout en me laissant beaucoup de liberté, ce qui m'a permis de travailler sur beaucoup de projets différents et de rencontrer du (beau) monde. Merci aussi d'avoir joué au plombier avant le déménagement sur Toulouse et au taxi/serveur pour ma soirée de pacs. Je retiendrai bien sûr une phrase fétiche qui a su calmer quelques débats lors des pauses-café : « oui, mais lui il est important ». Merci patron !

Jean-Michel, cette thèse n'aurait pas été possible sans votre aide. Un grand merci pour avoir accepté d'encadrer ces travaux qui auront été riches d'enseignements. Je n'oublie pas la conclusion de votre présentation lors du GSF 2017 chez Mickey : « soyez généreux mais ne lâchez rien ». J'avoue avoir dépensé beaucoup d'énergie sur la seconde partie de ce précepte, je me donnerai à l'avenir l'occasion de l'appliquer dans sa totalité.

Gaëlle, PI, ça a été un grand plaisir de travailler avec toi. La précision et la rapidité légendaire qui te caractérisent ont toujours été une source d'inspiration pour moi. Merci pour tous tes conseils, ta bonne humeur et pour ton dynamisme qui ont été une solide source de motivation pendant toutes ces années. Tu vas beaucoup nous manquer avec Lucile. On te fait de gros bisous ainsi qu'à ta petite famille et nous nous reverrons avec plaisir autour d'une tartiflette !

Sabrina, ça a été un grand plaisir de pouvoir collaborer avec toi. Je garderai bien sûr en mémoire les « *cazzo* » et les « *madonna* » qui ont animé nos réunions. Garde la bonne humeur qui te caractérise et maintiens le cap !

Céline, merci de m'avoir initialement formé et de m'avoir emmené sur des questionnements plus biologiques qu'informatiques. Je conserve aujourd'hui cet état d'esprit dans lequel le code que nous développons doit être une question biologique traduite en langage informatique.

Merci aux drôles de dames : Candice, Lucie et Noémie pour toutes nos discussions très enrichissantes autour de la biologie cellulaire, de la génétique et de la bioinformatique. Ces rencontres ont été très enrichissantes pour moi et je l'espère un peu pour vous aussi. Nos soirées au Ministère et nos *debriefings* de Top Chef me manqueront. Je vous souhaite plein de bonnes choses pour la suite !

Je remercie l'ensemble des membres du laboratoire qui ont contribué à ma stabilité mentale, notamment au travers des *afterworks* : Agnès, Jennifer, Jessica, Laetitia, Laura, Lydia, Vanessa, Benjamin, Olivier, Simon, ainsi que leurs conjoint(e)s rencontré(e)s autour d'un (souvent plusieurs) verre(s). Au plaisir de vous revoir lors d'un congrès, dans un bar (pour changer) ou à d'autres occasions.

Merci à tous les collaborateurs avec qui j'ai eu l'occasion de travailler. Tous ces échanges m'ont permis d'aborder des sujets passionnants et de m'immerger dans le "petit" univers de la recherche sur les sarcomes. Merci aussi aux associations de patients qui mettent en œuvre d'importants moyens afin que nous puissions continuer de travailler avec passion.

Je remercie mes amis, ma belle-famille et ma famille pour leur soutien inconditionnel. Un remerciement tout particulier à mes parents qui se sont questionnés pendant longtemps ce que j'allais faire dans la vie ... Eh ben voilà : c'est ça ! J'espère que vous aurez compris un peu le contenu de ce document. En résumé : tout est compliqué, peu de choses fonctionnent comme on le veut mais on y va toujours à fond !

Lucile, mon amour, toute cette aventure fut riche en rebondissements que nous avons franchi à deux. Merci pour ton soutien inconditionnel depuis toutes ces années, pour ta douceur

et pour toutes nos discussions passionnées entre la bioinformatique, la génétique et au-delà. Merci de m'avoir réconforté dans les moments difficiles, calmé dans mes moments de colère et botté le cul lors de mes (trop nombreux) caprices. Comme dit le proverbe : « seul on va plus vite, ensemble on va plus loin ». J'ai fait beaucoup d'erreurs ces dernières années en me précipitant, il est maintenant temps d'aller plus loin. Je terminerais simplement en reprenant ce que tu as prophétisé : l'avenir nous appartient !

Je n'oublie pas que tout ceci n'est finalement qu'un début et que le chemin sera long, parfois même semé d'embuches. Charge à moi de relever les défis qui se présenteront et de provoquer les bonnes occasions. Y'a plus qu'à !

Sommaire

Liste des figures

Liste des tables

Abréviations

— A —

ADN : acide désoxyribonucléique
ALT : *alternative lengthening of telomeres*
ARN : acide ribonucléique

— C —

cGAS-STING : *cyclic GMP-AMP synthase-stimulator of interferon genes*
CGH-*array* : *comparative genomic hybridization array*
chr : chromosome
CIN : *chromosomal instability*
CINSARC : *Complexity INdex in SARComas*
cm : centimètre

— D —

DDLPS : *dedifferenciated liposarcoma*
DNA : *deoxyribonucleic acid*
DNA-seq : DNA *sequencing*

— E —

ESMO : *European Society for Medical Oncology*

— F —

FFPE : *formalin-fixed, paraffin-embedded*
FISH : *fluorescence in situ hybridization*
FNCLCC : Fédération Française des Centres de Lutte Contre le Cancer

— G —

GEF1 : GTP *exchange factor 1*
GeneSigDB : *Gene Signature DataBase*
GI : *Genomic Index*
GIST : *gastrointestinal stromal tumor*
GSF : Groupe Sarcome Français
GTPase : *guanosine triphosphatase*
Gy : gray

— H —

Hi-C : capture de la conformation des chromosomes à haute résolution
HGP : *the Human Genome Project*

— I —

ICGC : *International Cancer Genome Consortium*
ITK : inhibiteur de tyrosines kinases

— J —

JNK : *c-Jun N-terminal kinase*

— K —

kb : kilobase

— L —

LM-BN : *leiomyoma with bizarre nuclei*
LMS : léiomyosarcome
lncRNA : *long non-coding* RNA
LPS : liposarcome

— M —

MAP-kinases : *mitogen-activated protein kinases*
MFS : myxofibrosarcome
mg : milligramme
miRNA : *micro*RNA
miRNA-seq : *micro*RNA *sequencing*
mm : millimètre
MPNST : *malignant peripheral nerve sheath tumor*
MSigDB : *Molecular Signatures Database*

— N —

NF-κB : *nuclear factor-kappa B*

— O —

OMS : Organisation Mondiale de la Santé

— P —

pb : paire de bases
PCR : *polymerase chain reaction*
PHRCK : programme hospitalier de recherche clinique en cancérologie
PRC2 : *Polycomb Repressive Complex 2*
PRECOG : *PREdiction of Clinical Outcomes from Genomic profiles*

— R —

Rac1 : *ras-related C3 botulinum toxin substrate 1*
RhoG : *ras homolog family member G*
RISC : RNA-*induced silencing complex*
RMS : rhabdomyosarcome
RNA : *ribonucleic acid*
RNA-seq : RNA *sequencing*
RT-qPCR : *quantitative reverse transcription polymerase chain reaction*

— S —

SGC : sarcome à génétique complexe
SGS : sarcome à génétique simple
SRF : *serum response factor*
STM : sarcome des tissus mous
STMA : sarcome des tissus mous de l'adulte
STUMP : *uterine smooth muscle tumor with uncertain malignant potential*
SWI/SNF : *SWItch/sucrose nonfermentable*

— T —

TCGA : *The Cancer Genome Atlas*
TMLU : tumeur musculaire lisse utérine
TSS : *transcription start site*

— U —

UPS : *undifferentiated pleomorphic sarcoma*
UTR : *untranslated transcribed region*
UTROSCT : *uterine tumor resembling ovarian sex cord tumor*

Abréviations des gènes et des protéines

— A —

ABL1 : *ABL proto-oncogene 1, non-receptor tyrosine kinase*
ACTB : *actin beta*
ADAMTS12 : *ADAM metallopeptidase with thrombospondin type 1 motif 12*
ANLN : *anillin actin binding protein*
APC : *APC, WNT signaling pathway regulator*
ARIH2 : *ariadne RBR E3 ubiquitin protein ligase 2*
ASPM : *abnormal spindle microtubule assembly*
ASPSCR1 : *ASPSCR1, UBX domain containing tether for SLC2A4*
ATF2 : *activating transcription factor 2*
ATM : *ATM serine/threonine kinase*
ATR : *ATR serine/threonine kinase*
ATRX : *ATRX, chromatin remodeler*
AURKA : *aurora kinase A*
AURKB : *aurora kinase B*

— B —

B2M : *beta-2-microglobulin*
BAF47 : *SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1*
BCL6 : *BCL6, transcription repressor*
BCOR : *BCL6 corepressor*
BCR : *BCR, RhoGEF and GTPase activating protein*
BIRC5 : *baculoviral IAP repeat containing 5*
BMI1 : *BMI1 proto-oncogene, polycomb ring finger*
BORA : *bora, aurora kinase A activator*
BRCA2 : *BRCA2, DNA repair associated*
BUB1 : *BUB1 mitotic checkpoint serine/threonine kinase*
BUB1B : *BUB1 mitotic checkpoint serine/threonine kinase B*

— C —

c-src : *SRC proto-oncogene, non-receptor tyrosine kinase*
CCNA2 : *cyclin A2*
CCNB1 : *cyclin B1*
CCNB2 : *cyclin B2*
CCNB3 : *cyclin B3*
CCND1 : *cyclin D1*
CCNE1 : *cyclin E1*
CDC6 : *cell division cycle 6*
CDC7 : *cell division cycle 7*
CDC20 : *cell division cycle 20*
CDC42 : *cell division cycle 42*
CDC45 : *cell division cycle 45*
CDCA2 : *cell division cycle associated 2*
CDCA3 : *cell division cycle associated 3*
CDCA8 : *cell division cycle associated 8*
CDH18 : *cadherin 18*
CDK1 : *cyclin dependent kinase 1*
CDK4 : *cyclin dependent kinase 4*
CDK6 : *cyclin dependent kinase 6*
CDKN2A : *cyclin dependent kinase inhibitor 2A*
CDKN2B : *cyclin dependent kinase inhibitor 2B*
CENPA : *centromere protein A*
CENPE : *centromere protein E*
CENPL : *centromere protein L*
CEP55 : *centrosomal protein 55*
CHEK1 : *checkpoint kinase 1*
CITED2 : *Cbp/p300 interacting transactivator with Glu/Asp rich carboxy-terminal domain 2*
CKAP5 : *cytoskeleton associated protein 5*
CKS2 : *CDC28 protein kinase regulatory subunit 2*
CLNK : *cytokine dependent hematopoietic cell linker*
COL1A1 : *collagen type I alpha 1 chain*
COL6A3 : *collagen type VI alpha 3 chain*
CoREST : *REST corepressor 1*
CREB3L2 : *cAMP responsive element binding protein 3 like 2*
CTNNA2 : *catenin alpha 2*
CTNNB1 : *catenin beta 1*
CTSD : *cathepsin D*

— D —

DDIT3 : *DNA damage inducible transcript 3*
DNAH5 : *dynein axonemal heavy chain 5*
DNMT3A : *DNA methyltransferase 3 alpha*
DNMT3B : *DNA methyltransferase 3 beta*

— E —

E2F : *E2F transcription factor*
ECT2 : *epithelial cell transforming 2*

EED : *embryonic ectoderm development*
EGR1 : *early growth response 1*
EMILIN2 : *elastin microfibril interfacer 2*
ERCC2 : *ERCC excision repair 2, TFIIH core complex helicase subunit*
ESPL1 : *extra spindle pole bodies like 1, separase*
Etv1 : *ETS variant 1*
ETV6 : *ETS variant 6*
EWSR1 : *EWS RNA binding protein 1*

— F —

FAM53B : *family with sequence similarity 53 member B*
FANCI : *FA complementation group I*
FBXL7 : *F-box and leucine rich repeat protein 7*
FBXO5 : *F-box protein 5*
FLI1 : *Fli-1 proto-oncogene, ETS transcription factor*
FOXM1 : *forkhead box M1*
FOXO1 : *forkhead box O1*
FUS : *FUS RNA binding protein*

— G —

GAPDH : *glyceraldehyde-3-phosphate dehydrogenase*
GREB1 : *growth regulating estrogen receptor binding 1*

— H —

H2AFX : *H2A histone family member X*
HER2 : *Erb-B2 receptor tyrosine kinase 2*
HOTAIR : *HOX transcript antisense RNA*
HOXB13 : *homeobox B13*
HP1BP3 : *heterochromatin protein 1 binding protein 3*
HPRT1 : *hypoxanthine phosphoribosyltransferase 1*
HRAS : *HRas proto-oncogene, GTPase*

— I —

IFITM10 : *interferon induced transmembrane protein 10*
IFT52 : *intraflagellar transport 52*
IL17RB : *interleukin 17 receptor B*
IQUB : *IQ motif and ubiquitin domain containing*

— J —

JAZF1 : *JAZF zinc finger 1*
JJAZ1 : *joined to JAZF1*
JUN : *Jun proto-oncogene, AP-1 transcription factor subunit*

— K —

KIF2C : *kinesin family member 2C*
KIF4A : *kinesin family member 4A*
KIF11 : *kinesin family member 11*
KIF14 : *kinesin family member 14*
KIF15 : *kinesin family member 15*
KIF18A : *kinesin family member 18A*
KIF20A : *kinesin family member 20A*
KIF23 : *kinesin family member 23*
KIT : *KIT proto-oncogene receptor tyrosine kinase* **Kras** : *KRAS proto-cncogene, GTPase*

— L —

LINC01504 : *long intergenic non-protein coding RNA 1504*
LRFN5 : *leucine rich repeat and fibronectin type III domain containing 5*
LSD1 : *lysine demethylase 1A*

— M —

MAD2L1 : *mitotic arrest deficient 2 like 1*
MAML3 : *mastermind like transcriptional coactivator 3*
MAP3K5 : *mitogen-activated protein kinase kinase kinase 5*
MARCH11 : *membrane associated ring-CH-type finger 11*
MCM2 : *minichromosome maintenance complex component 2*
MCM7 : *minichromosome maintenance complex component 7*
MDM2 : *MDM2 proto-oncogene*
MED12 : *mediator complex subunit 12*
MELK : *maternal embryonic leucine zipper kinase*
MGMT : *O-6-methylguanine-DNA methyltransferase*
MMP2 : *matrix metallopeptidase 2*
MYC : *MYC proto-oncogene, bHLH transcription factor*
Myf5 : *myogenic factor 5*
MYOCD : *myocardin*
MyoD : *myogenic differentiation 1*

— N —

NAB2 : *NGFI-A binding protein 2*
NCAPH : *non-SMC condensin I complex subunit H*
NCOA1 : *nuclear receptor coactivator 1*
NCOA2 : *nuclear receptor coactivator 2*
NDE1 : *nudE neurodevelopment protein 1*
NEK2 : *NIMA related kinase 2*
NF1 : *neurofibromin 1*
NTRK3 : *neurotrophic receptor tyrosine kinase 3*
NUF2 : *NDC80 kinetochore complex component NUF2*

— O —

OIP5 : *Opa interacting protein 5*

— P —

p53 : *tumor protein p53*
PAX3 : *paired box 3*
PAX7 : *paired box 7*
PBK : *PDZ binding kinase*
PD-1 : *programmed cell death 1*
PD-L1 : *programmed cell death 1 ligand 1*
PDGFB : *platelet derived growth factor subunit B*
PDGFD : *platelet derived growth factor D*
PDGFRA : *platelet derived growth factor receptor alpha*
PDGFRα : *platelet derived growth factor receptor alpha*
PDGFRβ : *platelet derived growth factor receptor beta*
PGK1 : *phosphoglycerate kinase 1*
PIK3R1 : *phosphoinositide-3-kinase regulatory subunit 1*
PLK4 : *polo like kinase 4*
PPIA : *peptidylprolyl isomerase A*
PPP2R5E : *protein phosphatase 2 regulatory subunit B'epsilon*
pRb : *RB transcriptional corepressor 1*
PRC1 : *protein regulator of cytokinesis 1*
PRDM10 : *PR/SET domain 10*
PTEN : *phosphatase and tensin homolog*
PTTG1 : *pituitary tumor-transforming 1*

— R —

RAD51AP1 : *RAD51 associated protein 1*
RB1 : *RB transcriptional corepressor 1*
RCC2 : *regulator of chromosome condensation 2*
REST : *RE1 silencing transcription factor*
RICTOR : *RPTOR independent companion of MTOR complex 2*
RNASEH2A : *ribonuclease H2 subunit A*
RPLP0 : *ribosomal protein lateral stalk subunit P0*
RRM2 : *ribonucleotide reductase regulatory subunit M2*

— S —

SDHA : *succinate dehydrogenase complex flavoprotein subunit A*
SGO2 : *shugoshin 2*
SLC9A3 : *solute carrier family 9 member A3*
SMARCA4 : *SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4*
SMARCB1 : *SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1*
SMC2 : *structural maintenance of chromosomes 2*
SNHG18 : *small nucleolar RNA host gene 18*
SPAG5 : *sperm associated antigen 5*
SPC25 : *SPC25, NDC80 kinetochore complex component*
SRC : *SRC proto-oncogene, non-receptor tyrosine kinase*
SS18 : *SS18, nBAF chromatin remodeling complex subunit*
SSX1 : *SSX family member 1*
SSX2 : *SSX family member 2*
SSX4 : *SSX family member 4*
STAG2 : *stromal antigen 2*
STAT3 : *signal transducer and activator of transcription 3*
STAT6 : *signal transducer and activator of transcription 6*
SUZ12 : *SUZ12, polycomb repressive complex 2 subunit*

— T —

TERT : *telomerase reverse transcriptase*
TET2 : *tet methylcytosine dioxygenase 2*
TFE3 : *transcription factor binding to IGHM enhancer 3*
TLE1 : *TLE family member 1, transcriptional corepressor*
TLS : *FUS RNA binding protein*
TOP2A : *DNA topoisomerase II alpha*
TP53 : *tumor protein p53*
TPX2 : *TPX2, microtubule nucleation factor*
TRIO : *trio Rho guanine nucleotide exchange factor*
TRIP13 : *thyroid hormone receptor interactor 13*
TTK : *TTK protein kinase*

— U —

UBE2C : *ubiquitin conjugating enzyme E2 C*
UHRF2 : *ubiquitin like with PHD and ring finger domains 2*

— V —

v-src : *tyrosine-protein kinase transforming protein Src*

— W —

WWTR1 : *WW domain containing transcription regulator 1*

— Y —

YAP1 : *Yes associated protein 1*

— Z —

ZNF131 : *zinc finger protein 131*
ZNF558 : *zinc finger protein 558*
ZWINT : *ZW10 interacting kinetochore protein*

1 | Introduction

1.1 – Descriptif clinique des sarcomes des tissus mous

1.1.1 – Présentation

Les tumeurs des tissus mous sont une famille très hétérogène de cancers, affectant l'ensemble des tissus conjonctifs non osseux de l'organisme. Ces tissus de soutien comprennent entre autres les muscles, les cellules adipeuses, les vaisseaux sanguins, *etc.* Selon le dernier rapport en date de l'Organisation Mondiale de la Santé (OMS), ces tumeurs peuvent être regroupées en 12 grandes lignes de différenciation comprenant un total de 113 sous-types histologiques (**table 1** ; Fletcher *et al.*, 2013).

Tumeurs	**Sous-types**
Adipocytaires	13
Fibroblastiques/myofibroblastiques	32
Fibrohistiocytaires	5
Musculaires lisses	2
Périvasculaires	3
Musculaires striées	5
Vasculaires	13
Chondro-osseuses	2
Stromales gastro-intestinales	1
Des gaines nerveuses	14
À différenciation incertaine	22
Indifférenciées	1

TABLE 1 – Lignes de différenciation des tumeurs mésenchymateuses des parties molles. Table adaptée de Fletcher *et al.*, 2013.

La grande majorité (99%) des cas diagnostiqués sont des tumeurs bénignes pour lesquelles une résection chirurgicale seule conduit presque systématiquement à la rémission com-

plète des patients. Le pourcentage restant sont des tumeurs agressives (localement ou rarement métastatiques) ou hautement malignes (fort potentiel d'invasion locale et/ou à distance), portant l'appellation de sarcomes des tissus mous (STM). Ils représentent près de 80 sous-types histologiques et posent un réel défi clinique, aussi bien par l'établissement d'un diagnostic que par la prise en charge thérapeutique des patients (Casali *et al.*, 2018).

En France, 4000 nouveaux cas de STM de l'adulte (STMA ; 1 à 2 % des cancers totaux) sont diagnostiqués par an, principalement retrouvés chez les personnes de plus de 60 ans (Ducimetière *et al.*, 2011 ; Honoré *et al.*, 2015). Ces tumeurs sont principalement localisées au niveau des membres (60%), puis du tronc (30%) et enfin de la région de la tête et du cou (10% ; Guillou & Aurias, 2010). Elles sont généralement très agressives avec de fréquentes récidives locales et/ou à distance (souvent pulmonaires) et dont 10% des patients présentent des métastases décelables au moment du diagnostic initial.

1.1.2 – Étiologie

La grande majorité des STM se développent *de novo*, sans aucun facteur déclenchant identifié. Néanmoins, plusieurs associations ont pu être mises en évidence.

Agents chimiques : le risque de développement d'un STM a pu être corrélé à une exposition à des dioxines (carcinogènes) contenues dans certains herbicides ou produites par des usines de transformation chimique (Eriksson *et al.*, 1990).

Exposition aux radiations : 1 à 5% des STMA sont diagnostiqués chez des patients traités par radiothérapie pour un précédent cancer (Brady *et al.*, 1992). Ils définissent le groupe de sarcomes développés sur territoires irradiés. Ces tumeurs présentent des caractéristiques cliniques spécifiques et sont plus agressives que les STMA sporadiques (développement spontané). Les études cliniques montrent que ces sarcomes se développent principalement sur des tissus fortement exposés aux radiations (50 Gy), par opposition aux carcinomes qui se développent sur des tissus moins exposés (30 Gy ; Wiklund *et al.*, 1991 ; Lagrange *et al.*, 2000).

Infection virale : associés à une immunodéficience, les virus d'Epstein-Barr et Herpès-virus 8 sont respectivement reliés au développement de tumeurs des muscles lisses (Deyrup *et al.*,

2006) et cutanés (sarcome de Kaposi ; Blauvelt, 1999), dont l'initiation serait permise par l'action des protéines virales oncogènes.

Prédisposition génétique : avec des mutations constitutionnelles de plusieurs gènes (Fletcher *et al.*, 2013 ; Farid & Ngeow, 2016) tels que :

***TP53* :** syndrome de Li-Frauméni, où la moitié des patients atteints de ce syndrome développent des tumeurs avant 30 ans et dont près de 30% sont des STM et des ostéosarcomes (Gonzalez *et al.*, 2009).

***RB1* :** rétinoblastome, associé au développement de STM et d'ostéosarcomes (Kleinerman *et al.*, 2012).

***APC* :** polypose adénomateuse familiale (ou syndrome de Gardner), prédisposant aux tumeurs desmoïdes (Gurbuz *et al.*, 1994).

***NF1* :** neurofibromatose de type 1, associée au développement de tumeurs malignes des gaines nerveuses périphériques (MPNST pour *malignant peripheral nerve sheath tumors* ; Evans *et al.*, 2002).

Autres prédispositions. Une large étude menée sur 848 patients atteints de STMA rapporte une corrélation entre certaines variations génétiques rares et le développement de sarcomes. Ces polymorphismes pathogéniques constitutifs concernent principalement les gènes *BRCA2* (5% des cas), *ATM* (3,1% des cas), *TP53* (1,9% des cas), *ERCC2* (1,8% des cas) et *ATR* (1,7% des cas ; Ballinger *et al.*, 2016).

1.1.3 – Diagnostic

Le diagnostic des STM se réfère à la classification histologique publiée par l'OMS, se basant sur l'observation de la ligne de différenciation tumorale (Fletcher *et al.*, 2013). On distinguera ainsi plusieurs sous-types histologiques correspondant à plusieurs lignes de différenciation possibles (**table 1 page 18**).

Les critères diagnostiques incluent les caractéristiques cliniques (localisation et taille de la tumeur, âge et sexe des patients, *etc.*), l'aspect macroscopique (sur biopsie ou pièce d'exérèse), l'aspect microscopique sur une coupe de tissus fixés au formol et inclus en paraffine (FFPE pour *formalin-fixed, paraffin-embedded* ; **figure 1**), ainsi que la présence de marqueurs

génétiques détectés par différentes techniques de biologie moléculaire comme l'immunohistochimie, le séquençage Sanger, l'hybridation fluorescente *in situ* (FISH pour *fluorescence in situ hybridization*) ou encore l'hybridation génomique comparative (CGH-*array* pour *comparative genomic hybridization array*).

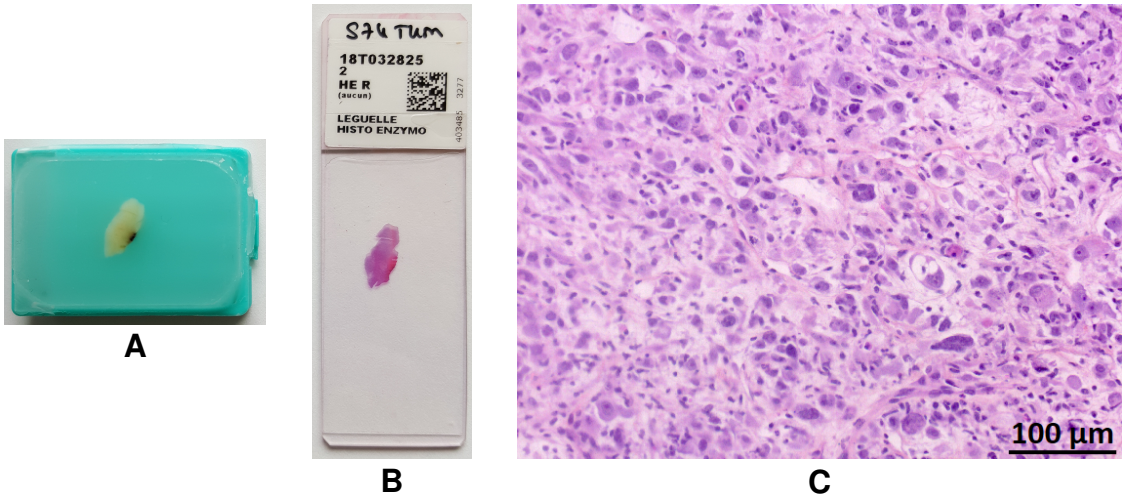


FIGURE 1 – Examen microscopique d'un bloc FFPE. **A** : bloc FFPE. **B** : lame du bloc FFPE avec coloration à l'hématoxyline et à l'éosine. **C** : visualisation au microscope (x200) de la coupe, correspond à un sarcome pléomorphe indifférencié de haut grade. Crédit : Sophie Le Guellec.

En raison de la très grande hétérogénéité de ces tumeurs, une double lecture des cas est systématiquement effectuée dans l'un des centres experts du Groupe Sarcome Français (GSF) au travers du Réseau de Référence en Pathologie des Sarcomes des tissus mous et des viscères (https://rreps.sarcomabcb.org ; Perrier *et al.*, 2018). La prise en charge des cas et leur suivi sont assurés lors des réunions de concertation pluridisciplinaire, faisant intervenir les principaux acteurs médicaux concernés par le traitement des patients (oncologues, chirurgiens, anatomo-pathologistes, *etc.*).

En dehors de ces centres experts, le risque d'erreur diagnostique initiale critique (impliquant une erreur de prise en charge thérapeutique) est estimé entre 10 et 25% et est due à une méconnaissance de la pathologie (Arbiser *et al.*, 2001 ; Ray-Coquard *et al.*, 2012). La relecture d'un cas par un pathologiste expert permet ainsi de reclasser 40% des diagnostics initiaux en reprécisant le sous-type histologique et le pronostic. Les données épidémiologiques démontrent que, hors des centres experts, les dossiers de patients sont moins souvent revus en réunions

de concertation pluridisciplinaire, qu'ils ont moins souvent une imagerie préopératoire de la tumeur et une moindre probabilité d'avoir un diagnostic établi avant la résection chirurgicale (Mathoulin-Pélissier *et al.*, 2014).

1.1.4 – Pronostic

Ces cancers sont généralement de mauvais pronostic, avec des taux de rechute locale de 20 à 30%, des taux d'événement métastatique de 30 à 50% (souvent pulmonaires) et des taux de survie globale d'environ 50% à 5 ans (variable selon le sous-type histologique ; Fletcher *et al.*, 2013). Les trois principaux facteurs cliniques, en plus du sous-type histologique en lui-même, intervenant dans l'agressivité de ces tumeurs sont la qualité de la résection chirurgicale, la localisation de la tumeur et sa taille. Classiquement, les tumeurs du rétropéritoine sont volumineuses, rendant l'acte chirurgical plus compliqué. À l'inverse les tumeurs des membres sont généralement de meilleur pronostic.

Le facteur pronostique de référence des STM est celui défini par la Fédération Française des Centres de Lutte Contre le Cancer (FNCLCC ; Trojani *et al.*, 1984 ; Coindre, 2006). Il se base sur trois critères qui sont : la différenciation tumorale, l'index mitotique et la nécrose (**table 2**). L'association de ces trois critères, sous forme d'un score, stratifie les tumeurs en 3 groupes représentant des niveaux d'agressivité significativement différents (**figure 2**). Ainsi, les tumeurs de grade 1 sont de bon pronostic, celles de grade 3 sont de mauvais pronostic et le grade 2 correspond à une agressivité intermédiaire.

Cependant, l'application clinique du grade FNCLCC est limitée par plusieurs facteurs : il est difficilement applicable sur des micro-biopsies et/ou sur des tissus traités en situation néoadjuvante, non applicable sur tous les sous-types histologiques, peu informatif pour les tumeurs de grade 2 (40% des cas) et sa reproductibilité peut légèrement varier entre pathologistes (Coindre *et al.*, 1986, 2001 ; Coindre, 2006).

De plus, si l'apport du grade FNCLCC a été majeur pour la prise en charge thérapeutique des patients, il repose intégralement sur une observation phénotypique des cellules tumorales. Il devient alors nécessaire de déterminer les causes de ces caractéristiques phénotypiques, per-

Différenciation tumorale :	
Score 1	Forte ressemblance histologique avec du tissu mésenchymateux normal
Score 2	Ressemblance histologique incertaine
Score 3	Sarcome indifférencié ou certains types histologiques spécifiques
Index mitotique [a] :	
Score 1	0 à 9 mitoses pour 10 champs
Score 2	10 à 19 mitoses pour 10 champs
Score 3	$\geq$ 20 mitoses pour 10 champs
Nécrose tumorale [b] :	
Score 0	Absence
Score 1	$<$ 50%
Score 2	$\geq$ 50%
Grade histologique :	
Grade 1	Total : 2 ou 3
Grade 2	Total : 4 ou 5
Grade 3	Total : 6, 7 ou 8

TABLE 2 – Critères et seuils d'évaluation du grade FNCLCC. [a] : un champ représente 0.1734mm^2. [b] : évaluée au niveau histologique et macroscopique. Table adaptée de Trojani *et al.*, 1984 et de Coindre, 2006.

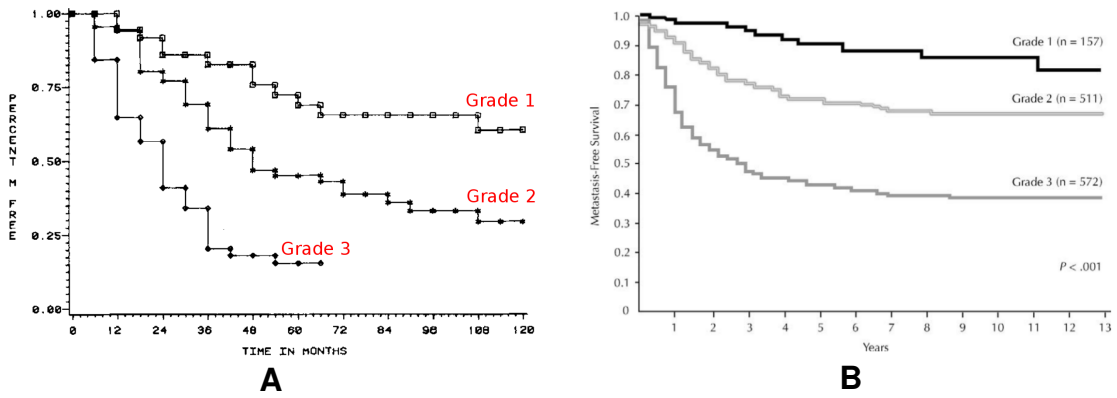


FIGURE 2 – Valeur pronostique du grade FNCLCC. **A** : première évaluation pronostique sur une cohorte de 155 patients. **B** : seconde évaluation pronostique sur une cohorte de 1240 patients. Ces analyses montrent des survies sans métastase (ordonnée) au cours du temps (abscisse) significativement différentes selon le grade FNCLCC. Images adaptées de Trojani *et al.*, 1984 et Coindre, 2006

mettant ainsi d'affiner le diagnostic mais aussi le pronostic par une meilleure compréhension de la nature de ces tumeurs. Ceci permettrait par exemple de comprendre pourquoi le grade FNCLCC est pronostique sur une grande variété de sous-types histologiques mais pas sur les rhabdomyosarcomes (RMS) et les MPNST.

Il est à noter que le pronostic des tumeurs stromales gastro-intestinales (GIST pour *gastrointestinal stromal tumors*) n'est pas évalué par le grade FNCLCC mais par le système d'évaluation défini par l'*Armed Forces Institute of Pathology* (Miettinen *et al.*, 2006). Ce système

est basé sur trois critères : la taille de la tumeur, le nombre de mitoses observées et le site anatomique tumoral. Il permet ainsi de classer les GIST selon quatre groupes associés à des agressivités tumorales distinctes : très faible, faible, intermédiaire et élevée.

1.1.5 – Traitements

La prise en charge thérapeutique des patients atteints de sarcomes est discutée au cas par cas lors des réunions de concertation pluridisciplinaire mais les principales recommandations sont données par les rapports réguliers de l'*European Society for Medical Oncology* (ESMO ; Casali *et al.*, 2018). Les sarcomes étant un groupe de tumeurs aux caractéristiques cliniques très hétérogènes, de nombreux traitements peuvent être envisagés. Néanmoins, cette pathologie a relativement peu bénéficié d'opportunités thérapeutiques majeures depuis plusieurs décennies (Demicco *et al.*, 2012).

Quatre principales situations cliniques sont définies dans les recommandations thérapeutiques des STM et dépendent de deux facteurs : si la tumeur est localisée ou métastatique et si elle est résécable ou pas (Casali *et al.*, 2018). Chacune de ces situations cliniques fait l'objet d'un protocole thérapeutique distinct et qui est ajusté au cas par cas, principalement en fonction du sous-type histologique. La valeur pronostique du grade FNCLCC est utilisée dans le cadre des tumeurs non métastatiques et résécables, où elle constitue la première branche décisionnelle associée au traitement (**figure 3**).

1.1.5.1 – La chirurgie

La résection, si possible avec marges chirurgicales saines à l'examen microscopique (dite R0), reste le traitement de référence et ayant le meilleur impact pronostique pour ces tumeurs lorsqu'elles restent localisées, ce d'autant si le chirurgien s'est spécialisé dans la chirurgie des sarcomes (Toulmonde *et al.*, 2014). Ainsi, une première chirurgie inadaptée n'est pas rattrapée par un quelconque traitement ultérieur (Honoré *et al.*, 2015). Selon la localisation et/ou le volume de la tumeur et/ou la qualité de la chirurgie, une résection incomplète peut être planifiée. Ces résections incomplètes sont de mauvais pronostic et une seconde chirurgie peut être envisa-

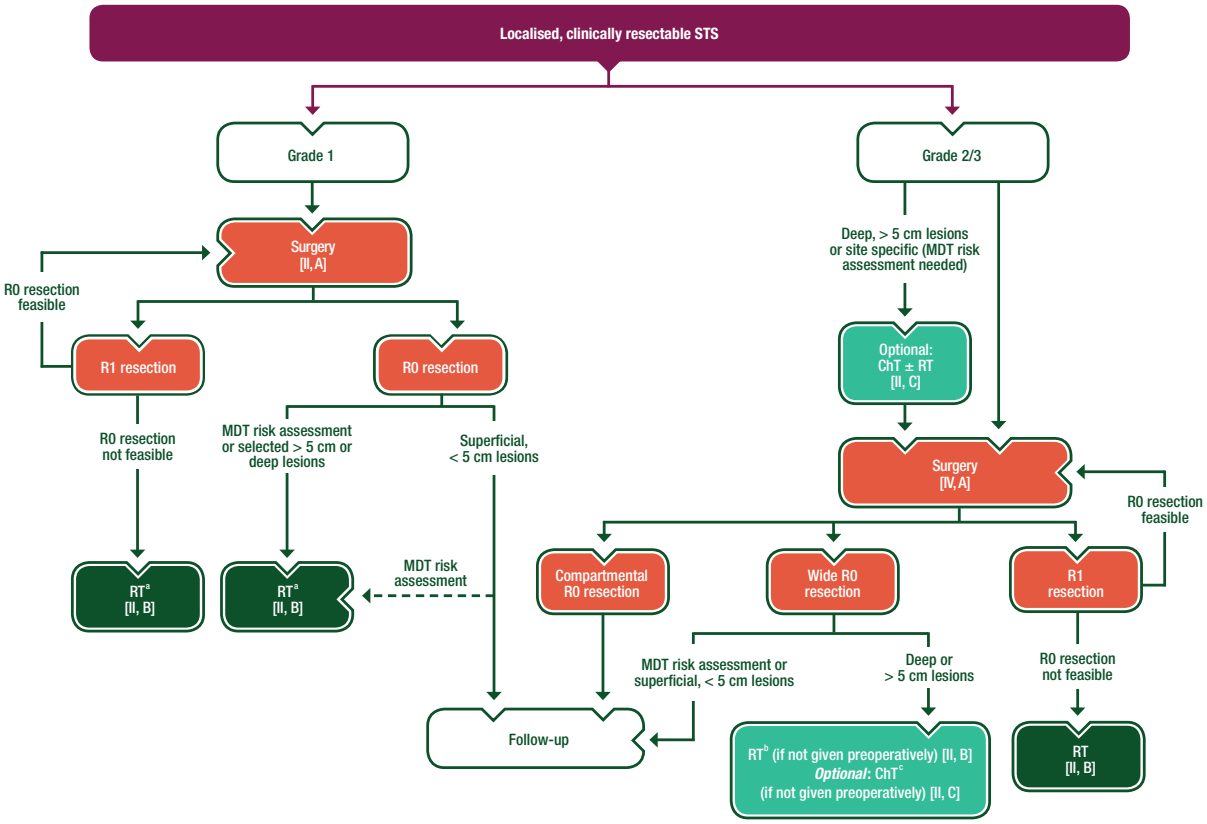


FIGURE 3 – Recommandations thérapeutiques des STM localisés et résécables. La première branche décisionnelle correspond au grade FNCLCC, actuel facteur pronostique de référence de la pathologie. MDT : équipe pluridisciplinaire, RT : radiothérapie, ChT : chimiothérapie. Image reprise de Casali *et al.*, 2018.

gée par la suite. Une partie de l'agressivité des STM du rétropéritoine s'explique par la taille et la localisation de ces tumeurs, qui ne permettent pas toujours une résection chirurgicale idéale (R0).

Une erreur de diagnostic initial, notamment la confusion entre tumeur bénigne et maligne, entraîne une mauvaise prise en charge chirurgicale dans les centres non spécialisés. En amont, elle implique une absence d'imagerie pré-opératoire, sans biopsie et/ou sans prévision de marges chirurgicales adéquates. Lors de l'opération, cela se traduit par des chirurgies dites "oups" (*whoops procedures*) avec effraction tumorale dont les conséquences cliniques sont désastreuses pour les patients qui sont souvent opérés une seconde fois dans les centres experts (Koulaxouzidis *et al.*, 2015).

1.1.5.2 – La radiothérapie

Elle est fréquemment proposée pour les sarcomes de haut grade en situation adjuvante afin d'améliorer le contrôle local de la maladie (Toulmonde *et al.*, 2014). Fractionnée, une irradiation totale de 50 à 66 Gy est administrée en fonction de la résection. Elle est recommandée en situation néoadjuvante dans deux cas de figure : si les complications peuvent être trop importantes pour une administration post-opératoire (elle est alors accompagnée de chimiothérapie), ou dans le cas d'une résection incomplète (R1 ou R2) qui ne sera pas réopérée.

Certains sous-types histologiques sont bien connus comme étant de bons ou mauvais répondeurs vis à vis de la radiothérapie. Par exemple, les synovialosarcomes, les tumeurs fibreuses solitaires et les liposarcomes (LPS) sont généralement radio-sensibles. Par opposition, la majorité des sarcomes pléomorphes répondent mal à la radiothérapie (Rhomberg, 2006 ; Casali *et al.*, 2018).

1.1.5.3 – La chimiothérapie

Elle est fortement recommandée pour les tumeurs métastatiques ou à risque élevé. Elle est administrée en situation néoadjuvante pour réduire le volume tumoral et ainsi faciliter l'exérèse, et en situation adjuvante pour tuer les cellules tumorales circulantes. En première ligne, une chimiothérapie à base d'anthracyclines est recommandée et peut être additionnée d'ifosmamide. Le cas échéant, une chimiothérapie spécifique au sous-type histologique de sarcome peut être envisagée (comme la combinaison doxorubicine et dacarbazine pour les léiomyosarcomes ; LMS).

Néanmoins, une étude clinique de phase 3 sur des sarcomes de haut risque, comprenant notamment des LMS et des sarcomes pléomorphes indifférenciés (UPS pour *undifferentiated pleomorphic sarcomas*), démontre que ce type de chimiothérapie spécifique, en situation néoadjuvante, n'apporte pas de bénéfice comparativement à une chimiothérapie standardisée (Gronchi *et al.*, 2017). Ces résultats poussent à redéfinir ce qui est alors considéré comme les traitements de référence (Baldini *et al.*, 2018).

1.1.5.4 – La thérapie ciblée

Hormis les inhibiteurs de tyrosines kinases (ITK) approuvés dans les GIST (imatinib, regorafenib et sunitinib), les sarcomes bénéficient de peu de thérapies ciblées (Nakano & Takahashi, 2018a). Ceci inclut certains sarcomes qui possèdent pourtant des altérations génétiques motrices et spécifiques (qui seront précisées dans la **partie 1.2.2 page 35**) mais qui restent actuellement sans traitement spécifique. Quelques molécules sont en cours d'essai, telles que des inhibiteurs d'histone désacétylases qui restreignent la modification de la structure de la chromatine permise par le complexe protéique SS18-SSX/TLE1/ATF2 dans les synovialosarcomes (Laporte *et al.*, 2017), ou des anti-CDK4/6 dans les LPS bien différenciés et LPS dédifférenciés (DDLPS pour *dedifferenciated liposarcomas* ; Dickson *et al.*, 2016).

Elle n'est pour l'instant pas envisageable pour les sarcomes pléomorphes puisqu'aucun levier thérapeutique récurrent n'a pu être identifié dans ces tumeurs. Des premières études de médecine personnalisée rapportent que 20 à 40% des sarcomes seraient éligibles pour une thérapie spécifique, ciblant une altération qui peut être identifiée par des analyses de biologie moléculaire (Gounder *et al.*, 2017 ; Ben Ami *et al.*, 2018). Toutefois, de grosses limitations thérapeutiques subsistent : une absence de molécule à tester, une inéligibilité des patients aux essais cliniques et un résultat trop tardif au regard d'un traitement déjà débuté ou du statut vital du patient. Cette approche reste néanmoins possible si une association pathologie-altération-traitement est déjà connue (Tlemsani *et al.*, 2018a,b).

1.1.5.5 – L'immunothérapie

Elle n'est pour l'instant pas un traitement de référence dans les STMA, excepté l'olaratumab pour les GIST (ciblant PDGFRα ; Tap *et al.*, 2016). De nombreuses études cliniques, surtout en phases 1 et 2 sont en cours pour évaluer la pertinence de l'immunothérapie dans le traitement des STMA (NCT03463408, NCT03138161, NCT03116529, *etc.*).

Ces dernières années, de nombreuses molécules ont été développées afin de bloquer la liaison entre PD-L1 (exprimée à la surface des cellules cancéreuses) et PD-1 (récepteur présent à la surface des lymphocytes T). Cette absence de liaison conduit à l'activation des lympho-

cytes et de nombreux essais cliniques y ont démontré un intérêt thérapeutique (cancers du poumon, gastriques, colorectaux, urothéliaux, lymphomes, *etc.* ; Gong *et al.*, 2018). Des essais sont en cours pour évaluer l'impact de ces molécules dans les sarcomes mais les premières études montrent un faible taux de réponse objective (Mir *et al.*, 2017).

1.2 – Descriptif génétique des sarcomes des tissus mous

En plus de leur classification histologique (**partie 1.1.3 page 20**), les STM peuvent être catégorisés sur la base de leurs altérations génétiques (Guillou & Aurias, 2010 ; Fletcher *et al.*, 2013). La première catégorie regroupe différents sous-types histologiques mais dont les points communs sont une anomalie génétique motrice de l'oncogenèse à la fois récurrente et spécifique, ainsi qu'un remaniement génomique globalement mineur (voire intermédiaire). Ce groupe est ainsi nommé "sarcomes à génétique simple" (SGS). La seconde catégorie regroupe différents sous-types histologiques définis par une très grande instabilité génétique, comprenant de nombreux remaniements génomiques mais sans altération spécifique, récurrente et motrice de l'oncogenèse. Ce groupe est ainsi nommé "sarcomes à génétique complexe" (SGC) et regroupe les sarcomes pléomorphes.

Avant de définir ces deux groupes, il est intéressant de noter que les premières grandes découvertes concernant la génétique tumorale synthétisent les principales altérations retrouvées dans les sarcomes (par ordre chronologique) : l'aneuploïdie, les translocations et les mutations ponctuelles. Nous verrons dans un premier temps dans quels contextes ces altérations ont été identifiées puis comment ces altérations participent à définir la biologie des sarcomes.

1.2.1 – Génétique et cancer : des premières découvertes au *hallmark* du cancer

Dès le début du XX^{ème} siècle, Theodor Boveri est l'un des premiers à faire l'hypothèse que l'origine des cancers pourrait être génétique (Boveri, 1914). Ce zoologiste allemand remarque en effet que la fécondation d'un œuf d'oursin par deux spermatozoïdes aboutit à une ségrégation chromosomique anormale dans les cellules filles (Boveri, 1904). Si la plupart de ces cellules filles ne survivent pas à cette anomalie, celles qui y parviennent présentent des développements anormaux. Ces résultats font écho aux observations de David von Hansemann qui, quelques années auparavant, rapporte que la plupart des divisions cellulaires de carcinomes

n'aboutissent pas à une répartition équitable des chromosomes dans les cellules filles (von Hansemann, 1890). Ces résultats conduisent ainsi Boveri à faire le lien entre altérations génétiques et oncogenèse. Il propose alors que des régulateurs chromosomiques existent (aujourd'hui nommés gènes suppresseurs de tumeurs), hérités des cellules parentales, permettant un contrôle de la prolifération cellulaire en condition normale (Boveri, 1914). L'inactivation de ces régulateurs lors de mitoses anormales permettrait aux cellules de proliférer sans cesse. Cette théorie fondatrice posera les bases de la contribution chromosome-cancer et du concours potentiel de l'aneuploïdie (nombre anormal de chromosomes dans une cellule) sur l'oncogenèse (**figure 4**; Bignold *et al.*, 2006).

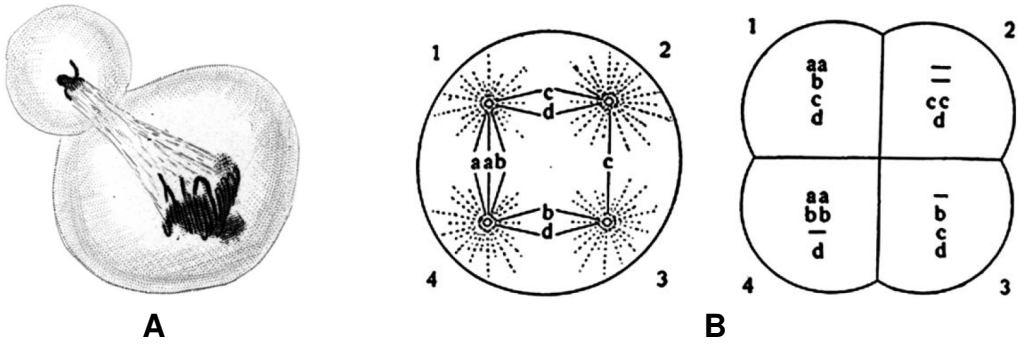


FIGURE 4 – Premières observations de l'aneuploïdie. **A** : illustration d'une mitose anormale observée dans un carcinome (von Hansemann, 1890). **B** : représentation de la répartition asymétrique des chromosomes (représentés par les lettres a, b, c et d) lors d'une mitose tétra-polaire (pôles 1 à 4; Boveri, 1914). Images reprises de Bignold *et al.*, 2006.

Les travaux de Boveri furent initialement accueillis avec scepticisme et il fallut attendre plusieurs décennies avant que les anomalies génétiques ne deviennent un support d'étude majeur afin de comprendre comment les cellules peuvent accéder à l'oncogenèse (**figure 5**; Balmain, 2001). La découverte de la structure d'ADN, par les efforts conjoints de James Watson, Francis Crick, Maurice Wilkins et Rosalind Franklin (Watson & Crick, 1953 ; Wilkins *et al.*, 1953 ; Franklin & Gosling, 1953) puis l'amélioration des techniques de cytogénétique ont permis d'initier des études ayant réhabilité les observations et théories de Boveri, constituant ainsi les fondations de la génétique tumorale.

La première translocation identifiée dans les tumeurs humaines fut l'échange spécifique entre les chromosomes 9 et 22, nommé chromosome de Philadelphie, dans les leucémies myé-

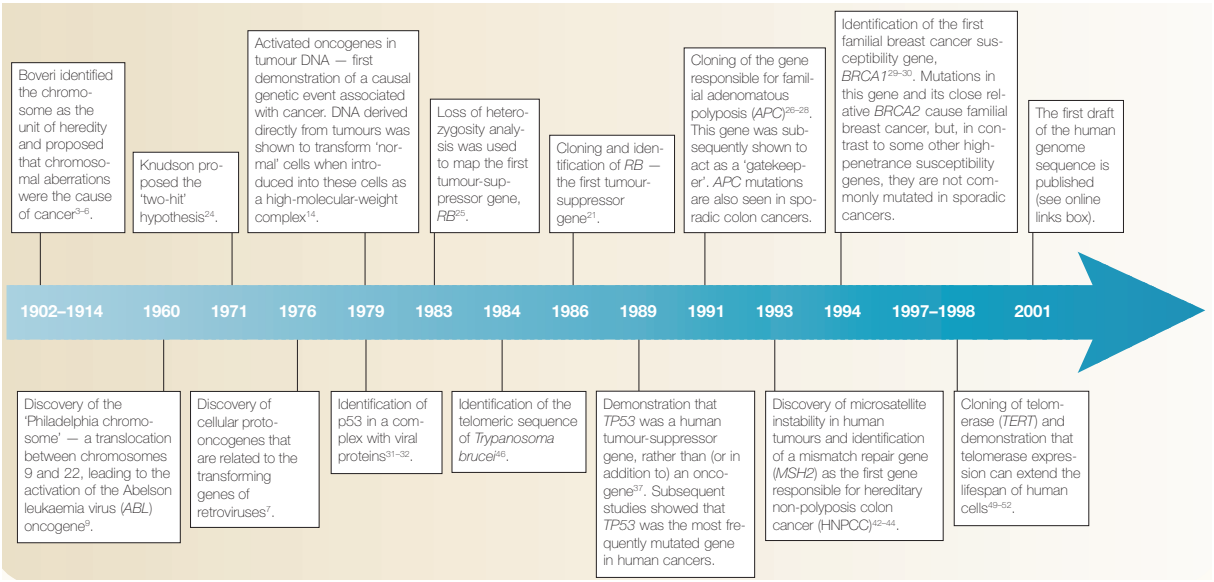


FIGURE 5 – Histoire de la génétique tumorale, de Boveri au séquençage du génome humain. Image reprise de Balmain, 2001.

loïdes chroniques (**figure 6A** ; Nowell & Hungerford, 1960 ; Rowley, 1973). Dans les tumeurs solides, la première translocation identifiée fut celle entre les chromosomes 11 et 22 dans les sarcomes d'Ewing (**figure 6B** ; Turc-Carel *et al.*, 1983 ; Aurias *et al.*, 1984). Il faudra encore attendre plusieurs années, grâce à une meilleure compréhension de la génétique tumorale et à l'établissement de techniques plus résolutives, avant de comprendre quelles sont les conséquences génétiques de ces deux translocations, avec respectivement la fusion des gènes *BCR* avec *ABL1* (Groffen *et al.*, 1984) et *EWSR1* avec *FLI1* (Delattre *et al.*, 1992).

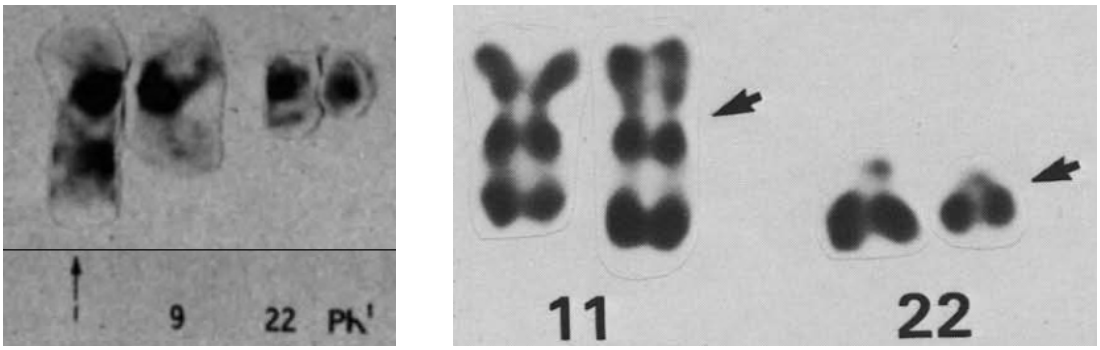


FIGURE 6 – Premières observations des translocations spécifiques. **A** : identification d'une translocation t(9;22) dans une cellule de leucémie myéloïde chronique. Image adaptée de Rowley, 1973. **B** : identification d'une translocation t(11;22) dans une cellule de sarcome d'Ewing. Image reprise de Aurias *et al.*, 1984).

Fin des années 70, des travaux menés sur le virus du sarcome de Rous (Rous, 1910, 1911) permettent d'identifier la première protéine virale oncogène (v-src) puis l'analogue humain de cette protéine (c-src ; Stehelin *et al.*, 1976 ; Oppermann *et al.*, 1979 ; Czernilofsky *et al.*, 1980 ; Smart *et al.*, 1981). Ces travaux précurseurs ont ainsi permis d'identifier la première oncoprotéine, donnant alors la définition formelle du premier oncogène (*SRC*) : une gène dont l'expression contribue à l'initiation tumorale. L'oncogène *HRAS* sera identifié quelques années après et notamment sa mutation p.G12V qui participe à l'oncogenèse des tumeurs de la vessie (Reddy *et al.*, 1982 ; Tabin *et al.*, 1982 ; Taparowsky *et al.*, 1982). D'autres découvertes viendront par la suite étoffer les connaissances accumulées sur la génétique tumorale. Un exemple notable est l'identification des délétions génomiques d'un locus (13q14), contenant notamment *RB1* (gène suppresseur de tumeur), altération récurrente retrouvée dans les rétinoblastomes (Friend *et al.*, 1986). Ces études princeps ont collectivement montré que l'initiation tumorale était associée à des altérations génétiques spécifiques.

En 1986, Renato Dulbecco déclare : « Nous avons deux options : soit essayer de découvrir les gènes importants dans la malignité par une approche au coup par coup, soit séquencer le génome entier d'une espèce animale sélectionnée » (Dulbecco, 1986). Face aux enjeux médicaux d'une telle importance, le gouvernement américain démarre le projet de séquençage du génome humain (HGP pour *the Human Genome Project*) en 1990, alors principalement financé par son département de la santé et de la recherche médicale (le *National Institute of Health*) qui planifie sa durée à 15 ans. La réalisation de ce projet devient rapidement un objectif international et ce ne sont pas moins de 20 laboratoires répartis dans six pays (Allemagne, Royaume-Uni, Chine, France, Japon et les États-Unis d'Amérique) qui contribueront à sa réalisation. La complétion de ce projet en 2003 permit de pratiquement caractériser la totalité (99%) de notre génome, apportant de nouvelles informations jusqu'alors inconnues telles que des événements de duplications génomiques, des millions de polymorphismes, l'identification de nouveaux gènes, la diversité des transcrits, *etc.* (International Human Genome Sequencing Consortium, 2001, 2004).

L'obtention de ces informations a été capitale pour de très nombreuses avancées dans la génétique tumorale puisqu'elles ont permis de définir la première référence du génome humain sur laquelle se baser (Bell, 2010 ; Wheeler & Wang, 2013). Au-delà des enjeux oncologiques, tous les domaines reliés de près ou de loin à la génétique ont profité de ces nouvelles connais-

sances (génétique des populations, maladies héréditaires, immunologie, métabolisme, criminologie, *etc.*). Ce projet a également été une preuve de concept de l'utilisation de technologies de séquençage haut débit (la première génération) pour une meilleure compréhension de la génétique, à la base nucléique près.

Bien que très résolutif, le séquençage haut débit était initialement limité par un facteur majeur : son coût. En effet, le HGP a coûté près de $3.000.000.000 (soit $1 le nucléotide) mais des améliorations successives de la technique ont permis de nettes diminutions budgétaires. Le génome de Craig Venter, fondateur de Celera Genomics et concurrent du projet HGP dans la course au séquençage du génome humain (Pennisi, 1999 ; Venter *et al.*, 2001), fut obtenu pour $100.000.000 (Levy *et al.*, 2007). La seconde génération de séquençage haut débit fut amorcée par la mise au point du séquençage 454 (Roche ; Margulies *et al.*, 2005) et permit de séquencer le génome de James Watson en deux mois pour $1.000.000 (Wheeler *et al.*, 2008). Ceci conduit à la théorisation du séquençage du génome complet à $1.000 (Mardis, 2006). Cet objectif est aujourd'hui proche d'être atteint (**figure 7**), le coût de l'analyse de ces données serait alors supérieur à celui du séquençage brut (Mardis, 2010).

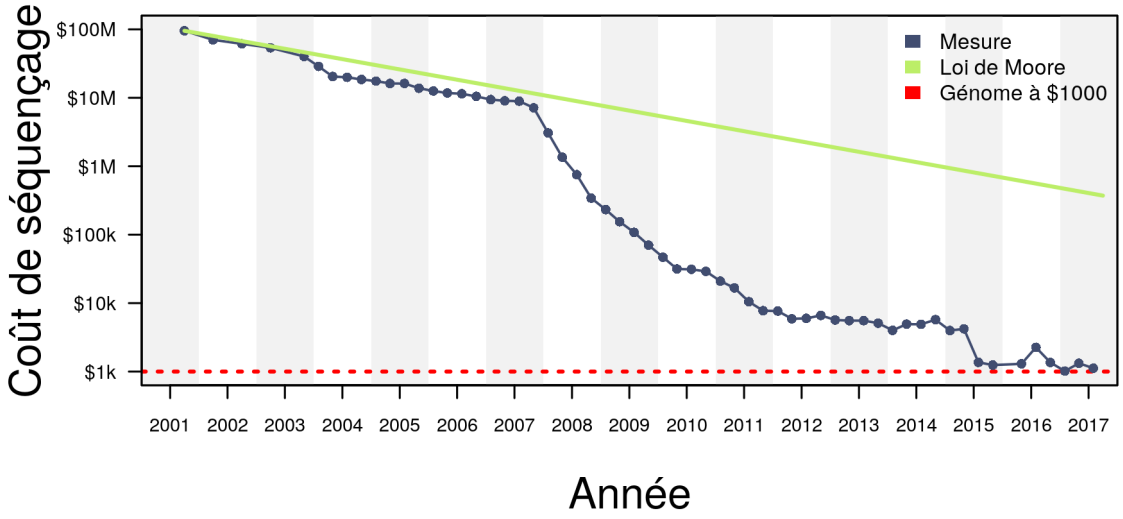


FIGURE 7 – Évolution du coût du séquençage du génome humain. Ce graphique retrace l'évolution du coût depuis 2001, comparativement à l'adaptation de la loi de Moore (1975) selon laquelle la puissance de calcul des ordinateurs double tous les deux ans. Ici, elle représente une diminution du coût de séquençage d'un facteur deux en deux ans à partir du premier point. Cette loi s'applique dans de nombreux autres domaines et suivait effectivement la tendance du coût de séquençage jusqu'à fin 2007. Le début de l'année 2008 marque la mise en service de la seconde génération de séquenceurs haut débit. Données obtenues sur https://www.genome.gov/sequencingcostsdata.

À partir du principe du séquençage ADN (DNA-seq pour DNA *sequencing*), de nombreuses variantes ont par la suite été développées afin d'étudier de façon toujours plus exhaustive l'ensemble des facettes de la génétique tumorale (**figure 8** ; Reuter *et al.*, 2015). On peut par exemple citer le séquençage ARN (RNA-seq pour RNA *sequencing*) afin de quantifier l'expression génique (Bainbridge *et al.*, 2006 ; Mortazavi *et al.*, 2008) ou la capture de la conformation des chromosomes à haute résolution (Hi-C) afin de modéliser l'architecture spatiale en trois dimensions des chromosomes au sein des noyaux (Lieberman-Aiden *et al.*, 2009). L'accessibilité de ces techniques, ainsi que d'autres méthodes permettant par exemple d'étudier les variations du nombre de copies ou de la méthylation de l'ADN par le design de puces spécifiques, ont contribué à l'émergence d'imposants projets, notamment au travers de deux groupes centraux que sont l'*International Cancer Genome Consortium* (ICGC ; https://icgc.org) et *The Cancer Genome Atlas* (TCGA ; https://cancergenome.nih.gov).

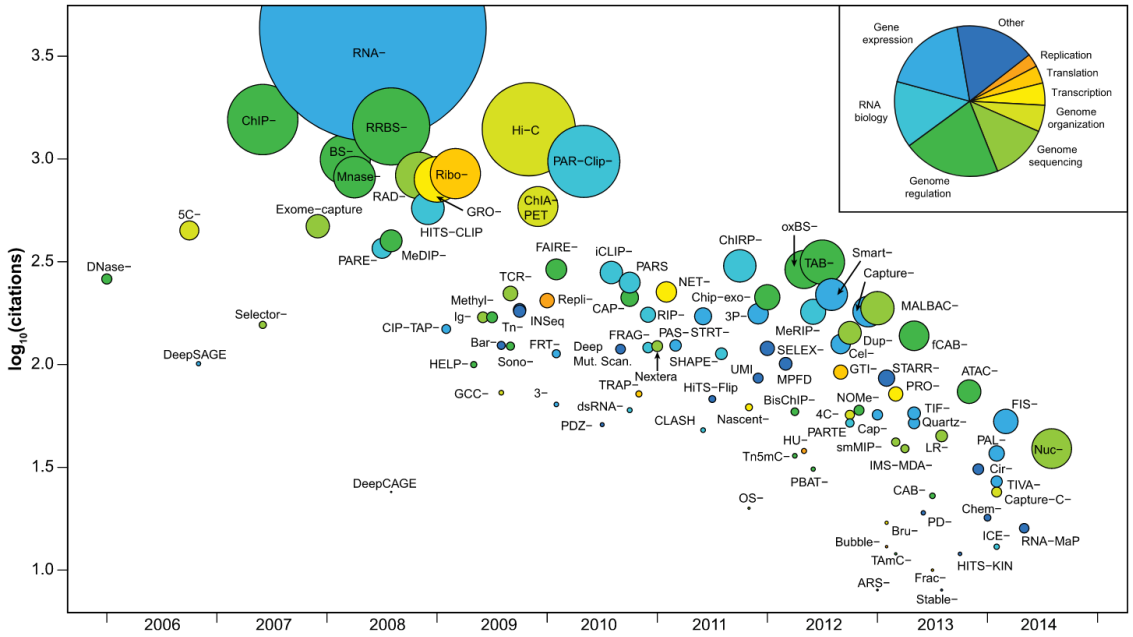


FIGURE 8 – Champs d'étude couverts par le séquençage haut débit. Les points représentent chacun une technologie de séquençage (RNA-seq, Hi-C, *etc.*) identifiée par son année de mise au point de 2006 à 2014 (abscisse), son nombre de citations scientifiques (ordonnée), ainsi que son taux de citation par mois (taille des points). La coloration des points indique le champ d'étude correspondant parmi : réplication, traduction, transcription, conformation chromosomique, séquençage ADN, régulation épigénétique, biologie des ARN, expression génique et autres (respectivement de couleurs orange à bleu foncé ; Reuter *et al.*, 2015).

L'implication des altérations génétiques sur le développement tumoral n'est aujourd'hui plus à démontrer et elles constituent une caractéristique universelle de la cellule cancéreuse

(Hanahan & Weinberg, 2000, 2011). Ces altérations sont ainsi capables de conférer des caractéristiques majeures pour la survie cellulaire, la prolifération, la dissémination et pour la résistance aux traitements des cellules tumorales (Vogelstein *et al.*, 2013). Il est donc d'un intérêt majeur d'identifier quelles sont les altérations génétiques présentes dans les tumeurs, y compris dans les sarcomes, afin de mieux comprendre la biologie qui les animent.

1.2.2 – Principales altérations génétiques des sarcomes

Comme précédemment évoqué, les STMA peuvent être classés en deux catégories en fonction de leurs caractéristiques génétiques. Nous allons dans cette partie d'abord voir les sarcomes dits "à génétique simple", puis ceux dits "à génétique complexe", pour enfin s'intéresser à une caractéristique commune à ces deux groupes : l'instabilité génétique et sa répercussion sur la progression tumorale.

Les SGS (50% des STMA) sont caractérisés par une altération oncogénique récurrente et spécifique à chaque sous-type histologique (**table 3** ; Taylor *et al.*, 2011 ; Fletcher *et al.*, 2013). Ces altérations sont de deux principaux types : les ponctuelles et les translocations. On parle alors de sarcomes dit "à mutation" ou "à translocation".

1.2.2.1 – Les sarcomes "à mutation"

Les sarcomes dits "à mutation" (20% des STM) sont caractérisés par des altérations ponctuelles très spécifiques sur quelques gènes cibles en permettant leur activation ou leur inactivation. Les altérations ponctuelles sont des modifications localisées de la séquence nucléique et sont réparties en deux catégories : les substitutions, qui consistent à modifier un nucléotide par un autre, et aux insertions/délétions de moins de 100pb, qui consistent à ajouter ou retirer des nucléotides.

Le principal représentant de cette catégorie de tumeurs est le GIST par mutations activatrices d'un récepteur à tyrosine kinase, activant les voies MAP-kinases et Akt. Ces altérations concernent soit *KIT* (85% ; Hirota *et al.*, 1998), soit *PDGFRA* (10% ; Heinrich *et al.*, 2003). La plupart des mutations de *KIT* se produisent au niveau de ses exons 9 (15% des cas ; domaine ex-

	Type histologique	Gène	Fréquence	Référence
Alt. ponctuelles	GIST	*KIT*†	85%	Hirota *et al.*, 1998
		PDGFRA†	10%	Heinrich *et al.*, 2003
	Tumeur desmoïde	*CTNNB1*†	85%	Shitoh *et al.*, 1999
		APC‡	15%	Miyaki *et al.*, 1993
	Tumeur rhabdoïde	*SMARCB1*‡	30%	Versteege *et al.*, 1998
Translocations	Synovialosarcome	*SS18-SSX1*◇	60%	Clark *et al.*, 1994
		SS18-SSX2◇	35%	Fligman *et al.*, 1995
	Fibrosarcome C	*ETV6-NTRK3*◇	>80%	Knezevich *et al.*, 1998
	RMS alvéolaire	*PAX3-FOXO1*◇	65%	Barr *et al.*, 1993
		PAX7-FOXO1◇	20%	Davis *et al.*, 1994
	DFS	*COL1A1-PDGFB*△	>90%	Pedeutour *et al.*, 1996
	Liposarcome myxoïde	*FUS-DDIT3*◇	>90%	Rabbitts *et al.*, 1993
	Sarcome APM	*ASPSCR1-TFE3*◇	>90%	Ladanyi *et al.*, 2001
	Sarcome FBG	*FUS-CREB3L2*◇	>90%	Storlazzi *et al.*, 2003

TABLE 3 – Exemples d'altérations génétiques motrices retrouvées dans les SGS. Alt : altérations, C : congénital, DFS : dermatofibrosarcome de Darier-Ferrand, APM : alvéolaire des parties molles, FBG : fibromyxoïde de bas grade. † : mutations activatrices. ‡ : mutations inactivatrices. ◇ : protéine chimérique avec nouvelle fonction. △ : oncogène sous contrôle d'un promoteur fort.

tracellulaire) et 11 (70% des cas ; domaine juxta-membranaire), quelques rares mutations étant également rapportées au niveau des exons 13 et 17 (Heinrich *et al.*, 2003 ; Debiec-Rychter *et al.*, 2006 ; Heinrich *et al.*, 2008a). Il a pu être démontré que les mutations de l'exon 9 sont de plus mauvais pronostic que celles de l'exon 11 (Heinrich *et al.*, 2003 ; Debiec-Rychter *et al.*, 2006 ; Heinrich *et al.*, 2008b) et qu'au sein même de l'exon 11, les délétions (même en phase *in-frame*) sont associées à des comportements cliniques plus agressifs que les substitutions (Miettinen *et al.*, 2006 ; Lasota & Miettinen, 2008). La réponse au traitement est également différente selon la nature de l'altération motrice. De hautes doses d'imatinib, un ITK, ont montré des effets bénéfiques sur les GIST mutés dans l'exon 9 de *KIT* par rapport à de plus faibles doses (800mg et 400mg par jour ; Debiec-Rychter *et al.*, 2006), alors qu'aucun bénéfice n'a pu être mesuré pour les mutations de l'exon 11. Aussi, les patients développant des tumeurs mutées dans l'exon 9 répondent mieux au sunitinib (un autre ITK) que celles mutées dans l'exon 11 (Heinrich *et al.*, 2008a).

D'autres tumeurs mésenchymateuses, plus rares, présentent aussi des altérations ponctuelles. Les tumeurs demoïdes présentent des mutations qui vont soit inactiver *APC*, soit activer *CTNNB1* (Miyaki *et al.*, 1993 ; Shitoh *et al.*, 1999). Ces deux types d'altérations vont conduire à l'accumulation de la β-caténine dans le noyau (codée par *CTNNB1* et dégradée par *APC*), ce qui active la voie Wnt/β-caténine impliquée dans la prolifération et la différenciation. Les tumeurs rhabdoïdes (Versteege *et al.*, 1998) et certaines tumeurs thoraciques indifférenciées (Le Loarer *et al.*, 2015) présentent respectivement des mutations inactivatrices de *SMARCB1* et de *SMARCA4*, deux membres du complexe *SWItch/sucrose nonfermentable* (SWI/SNF) du remodelage de la chromatine. Cette modification permet alors la reprogrammation de l'expression génique.

1.2.2.2 – Les sarcomes "à translocation"

Les translocations sont des échanges inter-chromomiques jouant un rôle majeur dans le processus d'oncogenèse car on estime que près de 20% des cancers seraient dus à des translocations (Mitelman *et al.*, 2007 ; Mertens *et al.*, 2015). La *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer* (http://cgap.nci.nih.gov.gate2.inist.fr/Chromosomes/Mitelman) recense près de 22000 gènes de fusion identifiés, tous cancers confondus.

Les translocations oncogéniques sont relativement rares dans les carcinomes mais sont en revanche communes dans les tumeurs hématopoïétique (>60% ; Mitelman *et al.*, 2007) et les sarcomes, où elles caractérisent près de 20% des STMA. Ces sarcomes sont dits "à translocation" et correspondent à plus d'une dizaine de sous-types histologiques (Mitelman *et al.*, 2007 ; Taylor *et al.*, 2011 ; Fletcher *et al.*, 2013 ; Mertens *et al.*, 2016). Près de 300 gènes de fusion (issus de tous types d'altérations structurales) sont rapportés dans les sarcomes "à translocation" (Nakano & Takahashi, 2018a). Si la majorité ne sont pas ou peu ($\leq$1% des cas ; Mertens *et al.*, 2015) récurrents, certains sont très spécifiques (**figure 9**) et leur validation constitue une indication diagnostique.

Les translocations permettent l'expression d'un transcrit chimérique par la formation d'un gène de fusion. Ces derniers sont de deux types : ils permettent soit de placer l'expression d'un

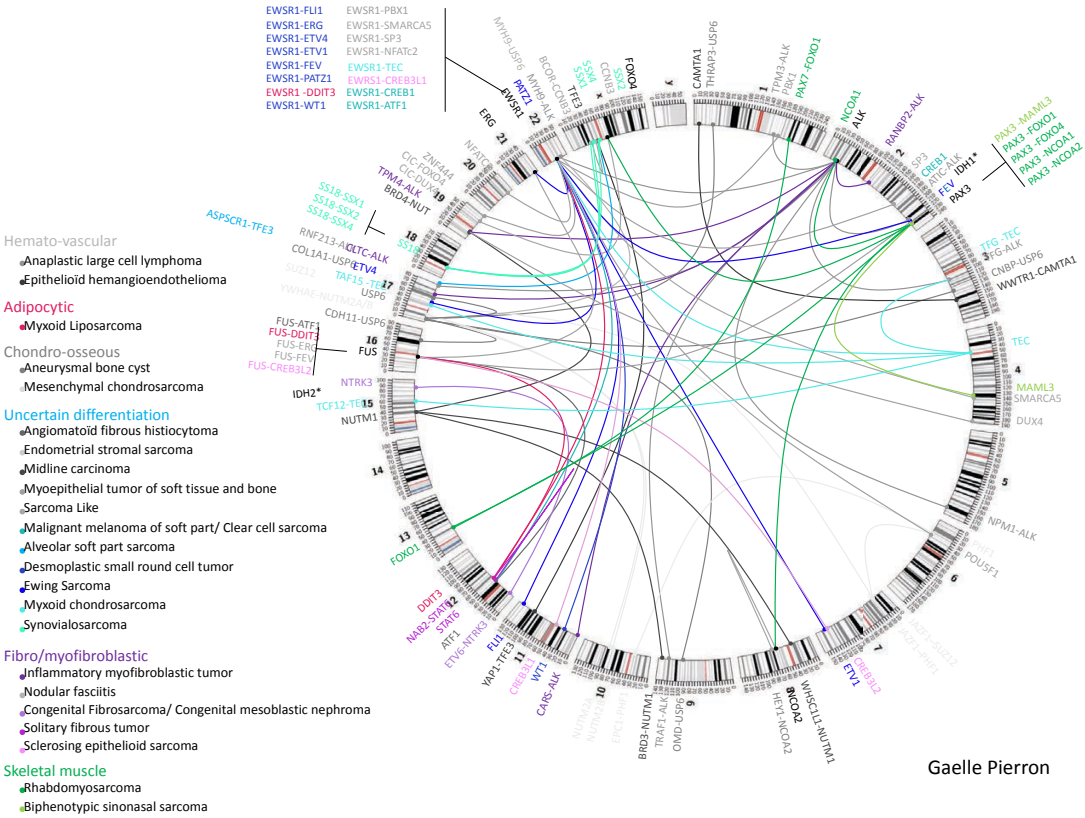


FIGURE 9 – Gènes de fusion récurrents retrouvés dans les sarcomes. Ce schéma inclut les gènes de fusion non issus de translocations (comme *NAB2-STAT6* ou *BCOR-CCNB3*) et retrouvés dans les sarcomes des tissus mous et tumeurs osseuses. Crédits : Olivier Delattre et Gaëlle Pierron.

oncogène sous le contrôle d'un promoteur fort, soit de coder pour une protéine chimérique en combinant des domaines d'intérêt des deux gènes (Mertens *et al.*, 2015).

Les dermatofibrosarcomes de Darier-Ferrand (5% des STM) sont des tumeurs superficielles qui sont localement invasives mais rarement métastatiques (Saiag *et al.*, 2015). Ces tumeurs présentent une translocation récurrente t(17;22)(q22;q13) qui permet l'expression d'un gène de fusion spécifique : *COL1A1-PDGFB* (Pedeutour *et al.*, 1996). La sur-expression de *PDGFB* permise par le promoteur fort de *COL1A1* stimule de manière autocrine PDGFRβ, un récepteur à tyrosine kinase et puissant mitogène (Shimizu *et al.*, 1999).

Parmi les sarcomes "à translocation", le synovialosarcome fait partie des entités les plus communément rencontrées (5 à 10% des STM). Il est principalement retrouvé chez les enfants ainsi que les jeunes adultes et son agressivité semble corrélée à l'âge au diagnostic car la survie globale des enfants atteints de synovialosarcomes est estimée à 90% à 5 ans alors qu'elle est estimée à 60% à 5 ans chez les adultes (Ferrari *et al.*, 2004, 2015). Ces tumeurs sont caractérisées

par la présence (exclusive) de trois translocations t(X;18)(p11;q11) aboutissant à l'expression de *SS18-SSX1* (65% des cas ; Clark *et al.*, 1994), *SS18-SSX2* (35% des cas ; Fligman *et al.*, 1995) ou *SS18-SSX4* (rares ; Skytting *et al.*, 1999). Le message oncogénique délivré par la protéine chimérique n'est pas encore totalement compris, SS18-SSX ne possède pas de domaine de liaison à l'ADN mais permet la reprogrammation cellulaire en modifiant les propriétés de la chromatine via le complexe SWI/SNF de remodelage de la chromatine. En effet, SS18 est un membre de ce complexe et la protéine chimérique SS18-SSX qui s'y intègre conduit à l'exclusion de BAF47, protéine codée par *SMARCB1* (gène suppresseur de tumeurs), modifiant ainsi l'action du complexe (différence d'accessibilité à la chromatine, de méthylation des histones et de l'ADN, *etc.* ; Banito *et al.*, 2018 ; McBride *et al.*, 2018).

On attribue souvent la génération de gènes de fusion aux translocations car elles ont été les premières altérations identifiées (Nowell & Hungerford, 1960 ; Aurias *et al.*, 1984). Toutefois, d'autres altérations structurales peuvent aussi conduire à l'expression d'un gène de fusion telles que les délétions, les duplications et les inversions (**figure 10** ; Annala *et al.*, 2013 ; Mertens *et al.*, 2016). Dans les sarcomes, plusieurs inversions ont pu être reliées à l'expression de gènes de fusion : inv(X)(p11.4;p11.22) dans les sarcomes indifférenciés à cellules rondes (*BCOR-CCNB3* ; Pierron *et al.*, 2012) ou inv(12)(q13.3) dans les tumeurs fibreuses solitaires (*NAB2-STAT6* ; Robinson *et al.*, 2013).

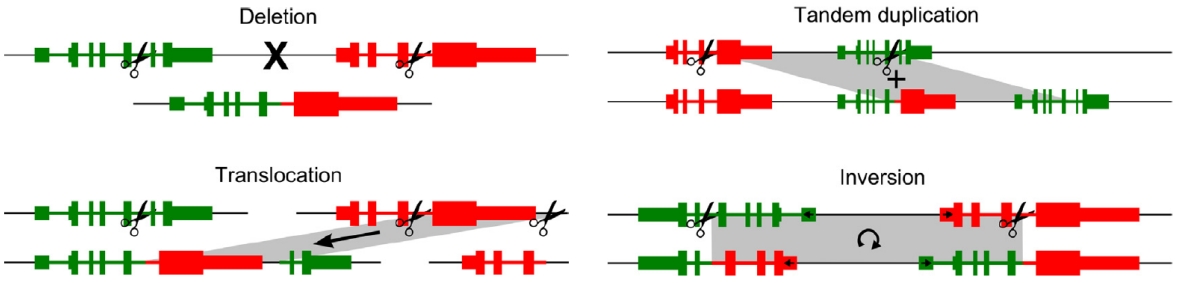


FIGURE 10 – Schéma des altérations structurales conduisant à l'expression de gènes de fusion. De gauche à droite et de haut en bas : délétion, duplication, translocation et inversion. Les deux gènes partenaires sont colorés en vert et en rouge. Image adaptée de Annala *et al.*, 2013.

Récemment, l'implication du mécanisme de *chromoplexy* a pu être identifiée par séquençage haut débit (combinaison de RNA-seq, DNA-seq et exome-seq) dans plusieurs tumeurs mésenchymateuses : un synovialosarcome, un fibrome chondromyxoïde, une tumeur mésen-

chymateuse phosphaturique et ce phénomène semble également fréquent dans les sarcomes d'Ewing (Anderson *et al.*, 2018). La *chromoplexy* est un réarrangement chromosomique complexe, initialement retrouvé dans le cancer de la prostate, dans lequel des translocations en chaîne s'opèrent sur plusieurs chromosomes (Baca *et al.*, 2013). Cette altération a permis de générer plusieurs translocations dans un synovialosarcome, dont une t(X;18)(p11;q11) permettant l'expression de *SS18-SSX1* (**figure 11**).

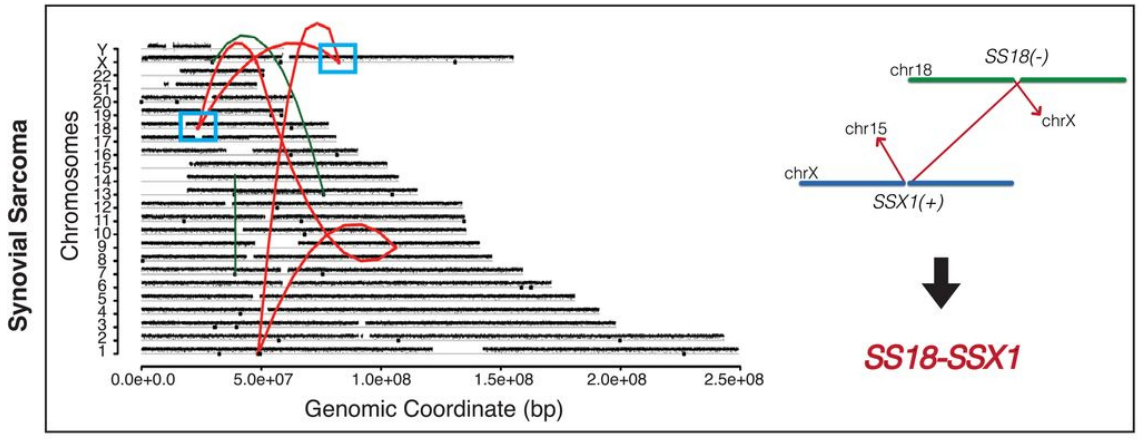


FIGURE 11 – *Chromoplexy* et formation de gènes de fusion. Dans ce synovialosarcome, de multiples translocations en chaîne se produisent (en rouge) et aboutissent à la formation du gène de fusion *SS18-SSX1*. Image adaptée de Anderson *et al.*, 2018.

Ces dernières années, l'accessibilité des techniques de séquençage haut débit et notamment le RNA-seq, a permis d'identifier non seulement de nouveaux gènes de fusion mais également de nouveaux types d'altérations génétiques comme la *chromoplexy* décrite ci-dessus. *BCOR-CCNB3* a ainsi été mis en évidence dans des sarcomes d'Ewing non *EWSR1-FLI1* et permet aujourd'hui de considérer ces tumeurs comme une entité propre parmi les sarcomes indifférenciés à cellules rondes (Ewing-*like* ; Pierron *et al.*, 2012). Le message oncogénique de *BCOR-CCNB3* n'est pas encore entièrement déchiffré mais *BCOR* est un corépresseur de *BCL6* (régulateur négatif de la transcription), *CCNB3* est un régulateur positif des kinases cyclines-dépendantes (cycle cellulaire) et le profil transcriptomique associé à l'expression de *BCOR-CCNB3* montre une activation des voies Wnt et Hedgehog (prolifération et différenciation). On peut aussi citer *NAB2-STAT6* qui est spécifique aux tumeurs fibreuses solitaires, conduisant à une augmentation de la prolifération cellulaire et l'expression des gènes cibles d'EGR1 via le domaine de liaison entre NAB2 et EGR1 (Robinson *et al.*, 2013).

1.2.2.3 – Les sarcomes à génétique complexe

Présentation des différents sous-types

Les SGC regroupent de nombreux sous-types histologiques ayant plusieurs lignes de différenciation (Ducimetière *et al.*, 2011 ; Fletcher *et al.*, 2013). Les principales tumeurs diagnostiquées sont :

Les LMS (20% des SGC) présentent une différenciation musculaire lisse mais cette différenciation est variable d'une tumeur à une autre. Ils sont communément retrouvés au niveau des membres, de l'utérus et du rétropéritoine. Leur agressivité dépend principalement de leur localisation (site anatomique), de leur profondeur (superficiel ou profond) et de leur différenciation (Pérot *et al.*, 2014).

Les UPS (10% des SGC) ne présentent pas de ligne de différenciation et constituent un diagnostic d'exclusion. Ils sont généralement de haut grade et agressifs.

Les myxofibrosarcomes (MFS ; 10% des SGC) ne montrent pas de différenciation particulière mais présentent des lésions fibroblastiques et un stroma myxoïde. Cette absence de différenciation peut les rendre difficilement identifiables parmi les UPS. Ils sont généralement superficiels et localisés au niveau des extrémités.

Les DDLPS (10% des SGC) présentent une différenciation adipocytaire plus ou moins marquée selon leur stade de dédifférenciation. Ce sont des sarcomes qui se développent principalement au niveau du rétropéritoine et qui présentent plus de rechutes locales (50%) que de métastases (15%). Il est estimé que 10% de ces tumeurs sont une évolution d'un LPS bien différencié (Fletcher *et al.*, 2013).

Les LPS pléomorphes (<5% des SGC) présentent aussi une différenciation adipocytaire. Ils se développent principalement au niveau des membres et du tronc et sont plus agressifs que les DDLPS avec 40% de rechutes locales et 40% d'événements métastatiques.

Les RMS pléomorphes (<5% des SGC) présentent une différenciation musculaire striée. Ce sont des tumeurs de haut grade à prédominance masculine et le plus souvent retrouvées au niveau des extrémités inférieures.

Leur génome : aneuploïde et remanié

Là où les SGS sont caractérisés par des altérations ponctuelles ou structurales très spécifiques, les SGC (50% des STM) sont définis par une importante aneuploïdie (**figure 12**). Ces sarcomes ne sont caractérisés ni par une altération structurale, ni par une altération ponctuelle (The Cancer Genome Atlas Research Network, 2017) mais ils sont caractérisés par une très haute complexité génomique et notamment par un nombre anormal de chromosomes. Si l'aneuploïdie est essentiellement délétère pour une cellule normale (Torres *et al.*, 2008), elle est toutefois l'une des propriétés de la cellule cancéreuse car on retrouve de l'aneuploïdie dans plus de 80% des cancers (Rajagopalan & Lengauer, 2004 ; Weaver & Cleveland, 2006). Elle participe à l'oncogenèse en faisant quantitativement varier le génome. Ainsi, l'augmentation du nombre de copies de gènes dits triplo-sensibles augmente leur expression, tandis que la diminution du nombre de copies de gènes dit haplo-insuffisants diminue leur expression (Davoli *et al.*, 2013).

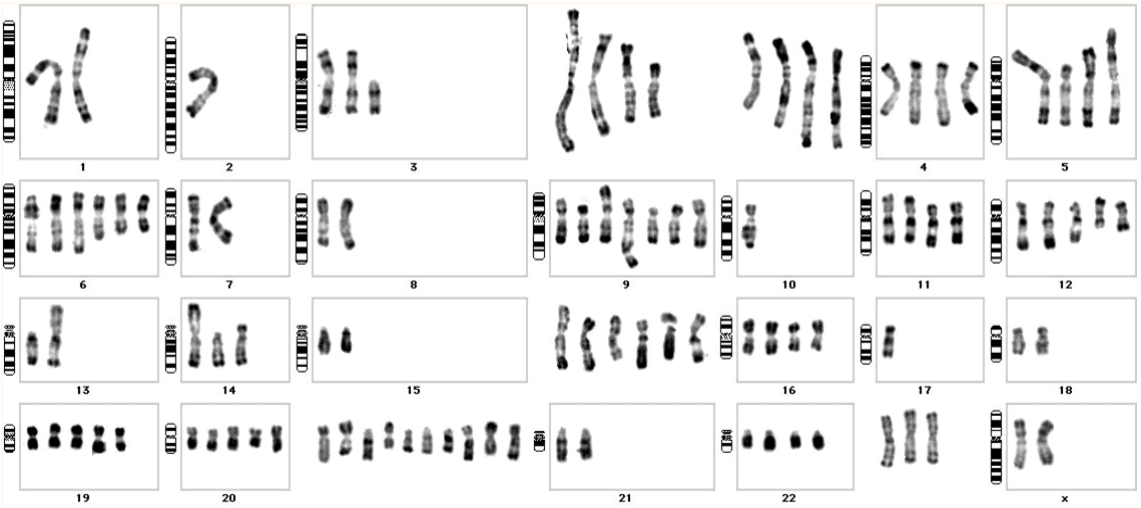


FIGURE 12 – Caryotype caractéristique d'une cellule de sarcome pléomorphe indifférencié. De nombreux chromosomes ne sont pas identifiables dans ce caryotype. Image adaptée de Mairal *et al.* (2000).

L'aneuploïdie a pu être mesurée dans les SGC, notamment dans les LMS où 55% des cas sont retrouvés sous forme tétraploïde (Chudasama *et al.*, 2018) ainsi que dans les UPS où 89% des cas montrent des caractéristiques d'au moins un doublement génomique (Steele *et al.*, 2019). Ceci est particulièrement intéressant car le doublement du génome peut permettre aux cellules de supporter l'aneuploïdie. Il a en effet pu être démontré qu'une augmentation de la

ploïdie permettait aux cellules de mieux résister à des conditions de stress (Rutledge *et al.*, 2016 ; Dewhurst *et al.*, 2014). De plus, la tétraploïdie peut être retrouvée à des stades très précoces du développement tumoral, en étant initiatrice de l'aneuploïdie selon le modèle "diploïde-tétraploïde-aneuploïde" (**figure 13** ; Carter *et al.*, 2012 ; Dewhurst *et al.*, 2014 ; Gerlinger *et al.*, 2014), et est également retrouvée dans l'ensemble des cancers (40% des tumeurs mais le taux est variable entre types de cancers ; Fujiwara *et al.*, 2005 ; Zack *et al.*, 2013). L'ensemble de ces données soulignerait la contribution du doublement du génome dans l'oncogenèse d'une partie des sarcomes pléomorphes.

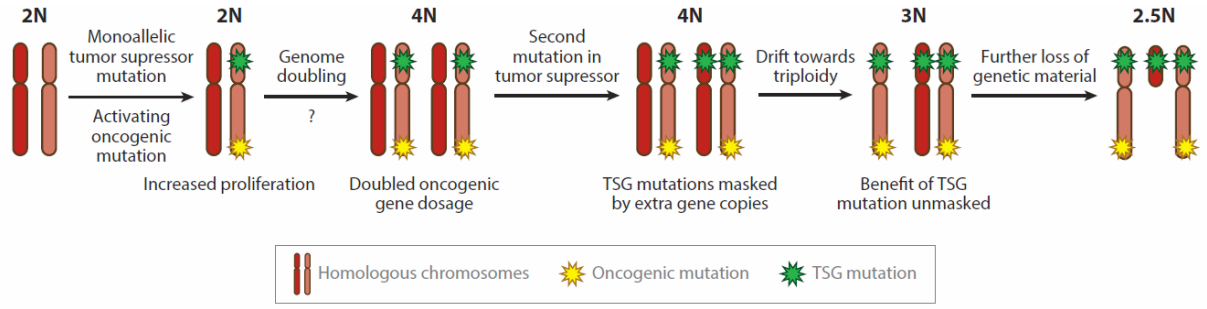


FIGURE 13 – Représentation du modèle "diploïde-tétraploïde-aneuploïde". Suite à l'acquisition d'altérations majeures dans une cellule diploïde (2N), le doublement du génome (4N) conduit à dupliquer l'ensemble du génome (altérations comprises). Le retour à l'état pseudo-diploïde (3N puis 2.5N) s'effectue par perte de matériel génétique ne conférant pas d'avantage sélectif (régions non altérées). Image reprise de Gerlinger *et al.*, 2014.

Le *chromothripsis* est également un acteur majeur du remaniement génomique de ces tumeurs, qui est généralement peu observé dans les cancers (<5% des cas) mais il est fréquent dans les tumeurs osseuses (20%) et les LMS (35% ; Chudasama *et al.*, 2018). Il est un événement catastrophique au cours duquel une portion d'un chromosome, un chromosome entier ou même plusieurs chromosomes (jusqu'à cinq) sont fragmentés en centaines de morceaux. Certains de ces morceaux vont être pris en charge par les mécanismes de réparation de l'ADN qui tenteront de les rassembler de manière anarchique (**figure 14** ; Stephens *et al.*, 2011). Il se caractérise ainsi par de nombreux réarrangements génétiques localisés et par une perte de matériel chromosomique due aux fragments qui sont perdus.

Du fait des nombreuses cassures génomiques induites par le *chromothripsis*, ce mécanisme pourrait être un événement à l'origine d'autres altérations comme la formation de gènes de fusion, la sur-expression d'oncogènes par gain/amplification ou encore l'inactivation de

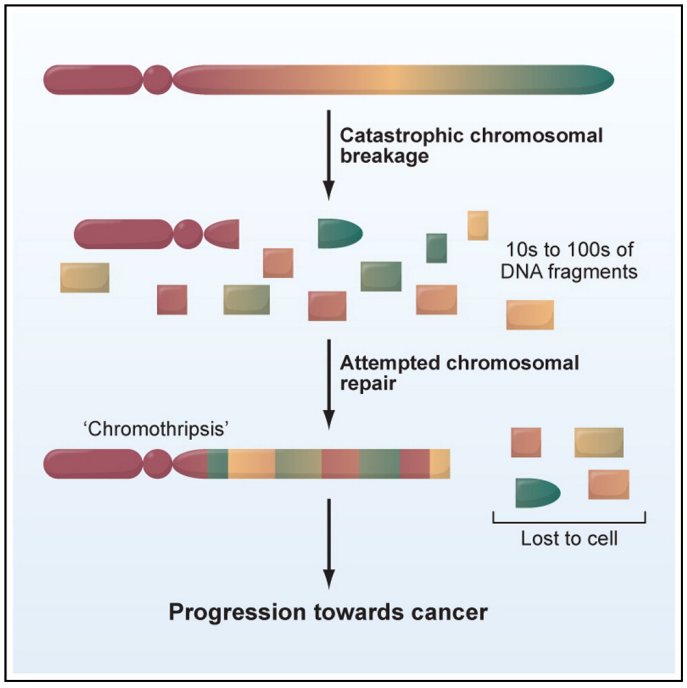


FIGURE 14 – Représentation du mécanisme de *chromothripsis*. Suite à un événement catastrophique conduisant à la pulvérisation d'un chromosome en dizaines voire en centaines de fragments, le réassemblage des fragments aboutit à la restructuration du chromosome. Image reprise de Stephens *et al.*, 2011.

gènes suppresseurs de tumeurs par perte de matériel génomique ou altération de leur séquence codante. (Rode *et al.*, 2016 ; Ly & Cleveland, 2017).

En plus de l'aneuploïdie et des réarrangements génomiques permis par un événement de *chromothripsis*, les sarcomes pléomorphes sont aussi caractérisés par de nombreuses translocations (**figure 15**). À l'inverses des SGS pour lesquels des gènes de fusion spécifiques ont pu être associés à la présence de translocations, les SGC n'ont jusqu'ici pas été reliés à l'expression d'un gène de fusion qui puisse être moteur de l'oncogenèse.

Quelques altérations récurrentes identifiées

Parmi ces tumeurs hautement aneuploïdes, quelques altérations récurrentes ont été identifiées : un gain du chromosome 1 et des pertes des chromosomes 10, 13 et 16 (Mairal *et al.*, 1999 ; Chibon *et al.*, 2003 ; Guillou & Aurias, 2010 ; Gibault *et al.*, 2011 ; Italiano *et al.*, 2013). Les SGC possèdent également deux caractéristiques génétiques très fréquentes : les inactivations systématiques des voies p53 et pRb, deux voies clés de la répression tumorale (Hanahan & Weinberg, 2000, 2011). Ces deux voies sont toujours altérées via un petit nombre de gènes (Chibon *et al.*, 2000 ; Pérot *et al.*, 2010) : *RB1* (80%, par délétion), *TP53* (50%, par mutation

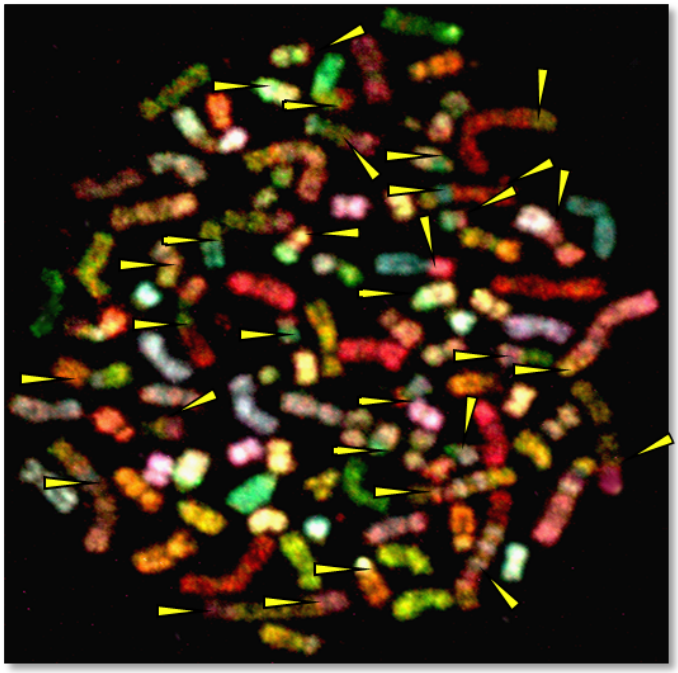


FIGURE 15 – FISH 24 couleurs d'une cellule de sarcome pléomorphe indifférencié. Ceci met en lumière de nombreux réarrangements inter-chromosomiques à grande échelle, indiqués par des flèches jaunes. Image adaptée de Mairal *et al.* (2000).

et/ou délétion) et *CDKN2A*/*CDKN2B* (25%, par délétion). Les altérations de *RB1* et de *TP53* semblent essentielles pour l'initiation tumorale, comme le suggère un modèle murin dans lequel ces deux gènes sont inactivés et qui permet le développement d'ostéosarcomes et de sarcomes indifférenciés (Lin *et al.*, 2009). Le développement de sarcomes indifférenciés, reproduisant les mêmes caractéristiques cliniques que les tumeurs humaines, est également permis par d'autres combinaisons d'altérations (Tp53 et Pten ou Tp53 et Kras ; Dodd *et al.*, 2010). Toutefois, le génome de ces tumeurs n'est pas remanié, suggérant que si ces altérations sont essentielles au développement tumoral, elles ne peuvent pas reproduire seules la génétique caractéristique des SGC.

Hormis les altérations récurrentes mais non spécifiques de *RB1* et *TP53*, les SGC ne présentent à ce jour aucune altération motrice à la fois récurrente et spécifique identifiée (The Cancer Genome Atlas Research Network, 2017). D'autres altérations récurrentes, plus rares et non restreintes à un sous-type histologique, ont pu être identifiées telles que les amplifications des loci 5p13-p15, associées à la sur-expression de *TRIO* (Adamowicz *et al.*, 2006) et du 17p12, associées à la sur-expression de *MYOCD* (assez spécifique des LMS du rétropéritoine ; Pérot *et al.*, 2009). Des pertes du 10q (*PTEN* ; Gibault *et al.*, 2011) et gains du 5p (*RICTOR* ;

Gibault *et al.*, 2012) ont également été observés dans ces tumeurs.

Un des mécanismes pouvant conduire à la dérégulation de ces gènes pourrait être le *chromothripsis*. Il est d'ailleurs rapporté que le chromosome le plus souvent affecté par ce type d'altération est le 17 (contenant notamment *TP53* et *MYOCD* ; Chudasama *et al.*, 2018). Cette même étude montre aussi une forte association entre la présence de *chromothripsis* et l'inactivation de *TP53* (92% des cas). Ceci concorde avec la génétique des SGC dans lesquels *TP53* est fréquemment inactivé (Pérot *et al.*, 2010).

Une première étude portant sur la caractérisation de l'expression de gènes de fusion a été menée sur 84 STM par séquençage haut débit (Hofvander *et al.*, 2015). Elle permit l'identification, dans trois UPS, des transcrits chimériques *MED12-PRDM10* (2 cas) et *CITED2-PRMD10* (1 cas). Il faut toutefois noter que ces trois tumeurs sont de bas grade, non métastatiques et les patients affectés étaient en rémission complète après 3 ans minimum, ce qui en font des cas très atypiques. L'évaluation de la spécificité histologique des réarrangements de *PRMD10* est délicate car, dans cette étude, les autres sous-types ne comportaient que peu de cas (22 MFS et moins d'une dizaine de cas pour les autres sous-types histologiques).

Le seul sous-type histologique de SGC pour lequel une partie significative de l'oncogenèse a pu être identifiée est le DDLPS. Ces sarcomes présentent une amplification de la région 12q14-q15 (Chibon *et al.*, 2002). Cette amplification conduit à la sur-expression de *MDM2* et *CDK4*, gènes qui vont (entre autres) cibler et inactiver p53 et pRb. De plus, deux régions mutuellement exclusives se retrouvent co-amplifiées avec la région 12q14-q15 : le locus 1p32.1 (Mariani *et al.*, 2007) ou 6q23.3 (Chibon *et al.*, 2004). Ces amplifications permettent respectivement la sur-expression de *JUN* et de *MAP3K5* et aboutissent à l'activation de la voie JNK, conférant ainsi une capacité tumorigène et supposée responsable de la dédifférenciation des DDLPS par blocage de la différenciation adipocytaire. Toutefois, si le rôle oncogène de l'activation de la voie JNK a pu être démontré, celle-ci ne semble pas être suffisante pour conduire seule à la dédifférenciation de ces tumeurs puisque l'on retrouve également cette voie activée dans le compartiment bien différencié de ces tumeurs (Snyder *et al.*, 2009).

L'instabilité génétique

Les SGC sont caractérisés par une importante instabilité génétique. Ce processus est défini comme l'accumulation de nouvelles altérations génétiques au cours du temps (Yates & Campbell, 2012). Elle est universellement rencontrée dans les cancers sous deux formes : l'instabilité nucléotidique et l'instabilité chromosomique (CIN pour *chromosomal instability*). L'instabilité nucléotidique se traduit par un nombre important d'altérations ponctuelles tandis que la CIN induit de nombreuses altérations structurales et/ou une aneuploïdie. Ces deux types sont communément retrouvés sous forme d'un équilibre : les familles de cancers présentent soit de faibles niveaux des deux, soit fortement l'une et faiblement l'autre (mais pas fortement des deux ; Ciriello *et al.*, 2013).

Si les sarcomes ne présentent pas une forte instabilité nucléotidique (comme les mélanomes et les cancers du poumon), la CIN est bien leur signature génomique (**figure 16** ; Taylor *et al.*, 2011 ; Fletcher *et al.*, 2013 ; The Cancer Genome Atlas Research Network, 2017). Cela est d'autant plus vrai pour les SGC qui présentent de très nombreuses altérations structurales (**figure 12 page 42** ; **figure 15 page 45**).

L'une des possibles origines de cette instabilité dans les sarcomes pourrait passer par une crise télomérique. L'immortalisation cellulaire peut en effet être assurée par deux mécanismes exclusifs. Elle peut s'effectuer soit par la réactivation de la télomérase (85% des cancers), soit par l'activation du mécanisme alternatif d'élongation des chromosomes (ALT pour *alternative lengthening of telomeres*), qui est le mécanisme actif dans 60% des SGC (Henson *et al.*, 2005 ; Chudasama *et al.*, 2018). Dans ces tumeurs, l'activation du mécanisme ALT est associée à une défaillance du complexe SWI/SNF, passant principalement par *ATRX* (inactivé dans 20 à 30% des cas ; Liau *et al.*, 2015 ; The Cancer Genome Atlas Research Network, 2017). Dans les SGC, le mécanisme ALT est associé à l'agressivité tumorale (grade FNCLCC) et pourrait contribuer au statut instable de ces tumeurs par insertion de séquences télomériques (TTAGGG et/ou TCAGGG) de manière anarchique dans le génome (Marzec *et al.*, 2015).

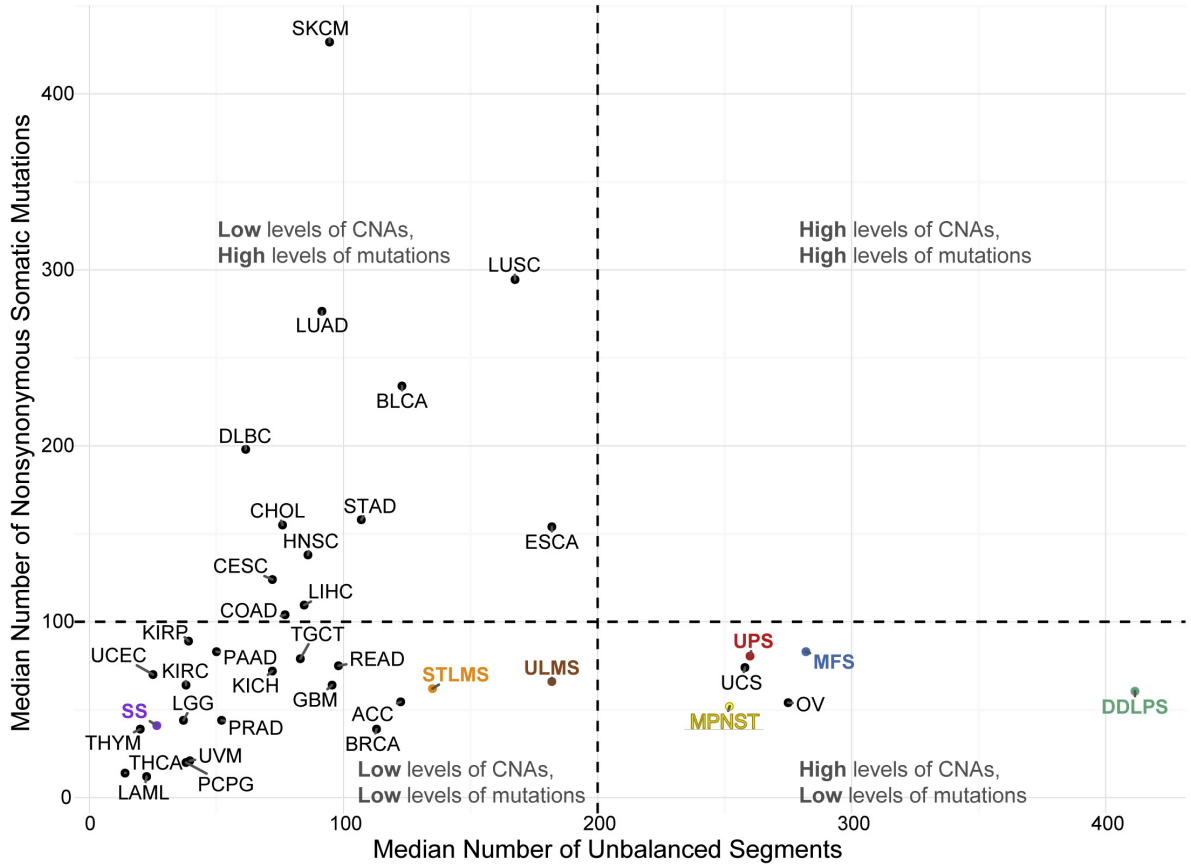


FIGURE 16 – Analyse pan-cancer de la charge mutationnelle et variation du nombre de copies. Le graphique représente, pour chaque type de cancer, la charge mutationnelle (ordonnée) et la variation du nombre de copies (abscisse). Les sous-types histologiques de sarcomes sont indiqués en couleur. SS : synovialosarcome, STLMS : LMS des tissus mous, ULMS : LMS utérins. Image reprise de The Cancer Genome Atlas Research Network, 2017.

Un modèle d'oncogenèse pour les UPS

L'ensemble des données accumulées sur la génétique des UPS, combinant la mesure de la variation du nombre de copies, la fréquence allélique et la détection d'altérations structurales, a récemment conduit à l'élaboration d'un modèle d'oncogenèse pour ces tumeurs (**figure 17** ; Steele *et al.*, 2019). Dans ce modèle, les altérations de *TP53* et de *RB1* seraient initiatrices et le développement tumoral serait assuré par quatre grandes voies d'oncogenèse qui sont, pour la plupart, associées à un doublement du génome (ces voies sont présentées ci-dessous). Il est à noter que trois mécanismes cellulaires peuvent être à l'origine d'un doublement du génome mais dont un seul est explicitement nommé dans ce modèle : l'endoréduplication. Les deux autres mécanismes sont la cytokinèse abortive et la fusion cellulaire.

L'haploïdisation. Cette voie serait permise par une erreur majeure de ségrégation chromosomique lors de la mitose. Cette ségrégation anormale pourrait conduire à la formation

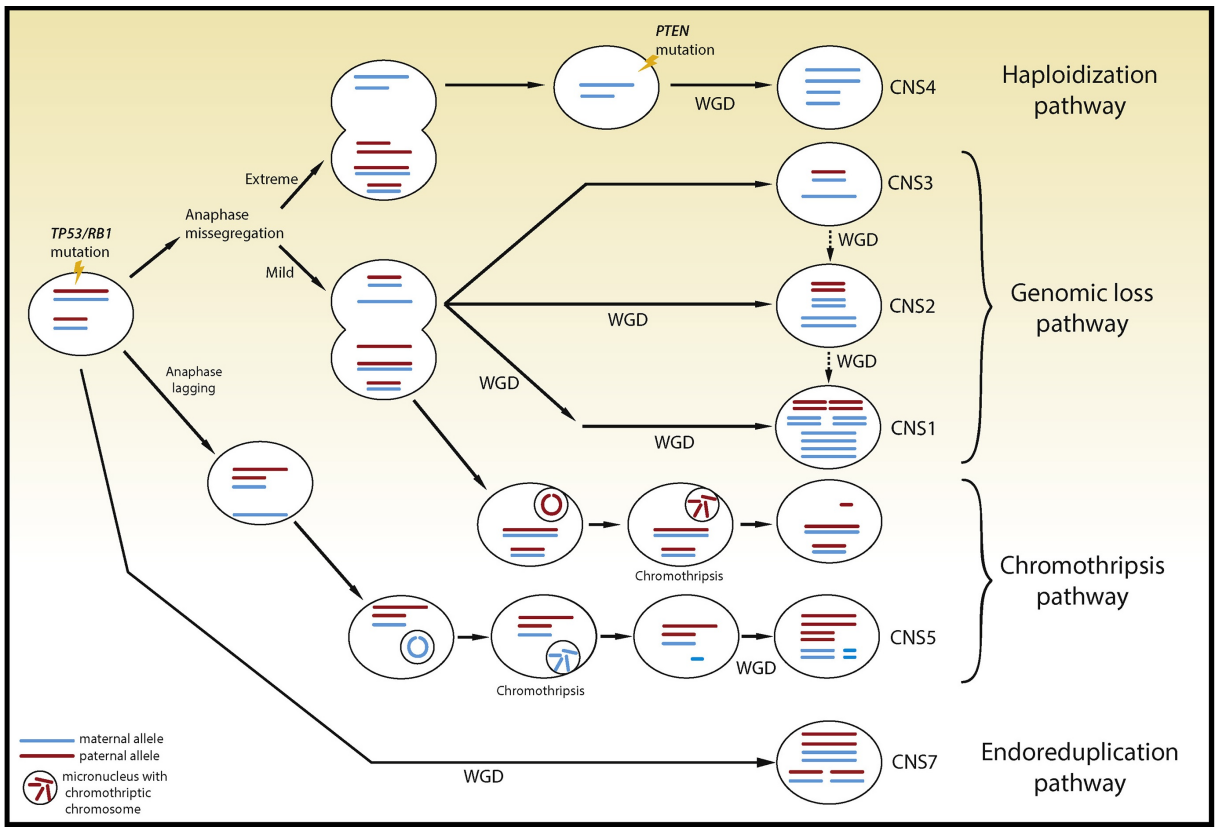


FIGURE 17 – Modèle de l'initiation tumorale des UPS. L'analyse de données génomiques de 52 UPS a permis de mettre en lumière sept signatures de variation du nombre de copies (CNS pour *copy-number signatures*). CNS1 est marquée par une importante perte d'hétérozygotie et une nette augmentation du nombre de copies, suggérant plusieurs doublements du génome. CNS2 est sensiblement similaire à CNS1 mais présente une plus faible augmentation du nombre de copies, suggérant un seul doublement du génome. CNS3 est sensiblement similaire à CNS2 mais sans augmentation du nombre de copies (pseudo-diploïde). CNS4 est caractérisée par une perte d'hétérozygotie en absence de variation du nombre de copies. CNS5 présente de grandes régions gagnées/amplifiées avec conservation de l'hétérozygotie et de grandes régions homozygotes sans variation du nombre de copies. CNS6 est une signature combinant de nombreuses variations, aussi bien du nombre de copies que de statut d'hétérozygotie (CNS6 n'est pas représentée sur le schéma). CNS7 est associée à des gains de grandes régions génomiques. WGD : doublement du génome. Image reprise de Steele *et al.*, 2019.

d'une cellule fille quasi-haploïde, associée à une perte d'hétérozygotie de la quasi-totalité du génome et dont la survie pourrait être permise grâce à un doublement du génome.

La perte de matériel génétique. Cette voie serait assurée par une erreur de ségrégation chromosomique lors de la mitose où une cellule fille se retrouverait pseudo-diploïde et présenterait donc d'importantes pertes d'hétérozygotie. Des événements de polyploïdisation pourraient ensuite générer des cellules pseudo-tétraploïdes voire pseudo-octoploïdes.

Le *chromothripsis*. Cette voie serait basée sur un événement catastrophique conduisant à la pulvérisation d'un ou plusieurs chromosomes, comme précédemment mentionné.

L'endoréduplication. Cette voie correspondrait au mécanisme d'endoréduplication, où deux phases réplicatives successives ne sont pas séparées par une phase mitotique. Ceci per-

mettrait également à une cellule de doubler son génome.

1.2.2.4 – Instabilité génétique et progression tumorale

La contribution de l'instabilité génétique dans l'initiation et la progression tumorale est bien documentée. Elle participe activement à l'acquisition de nouvelles altérations génétiques, conférant des avantages prolifératifs aux cellules qui seront positivement sélectionnées par expansion clonale (McGranahan & Swanton, 2017). Il existe ainsi un lien étroit entre instabilité génétique, hétérogénéité intra-tumorale, résistance aux traitements et capacité métastatique (**figure 18** ; Campbell *et al.*, 2010 ; Yates & Campbell, 2012).

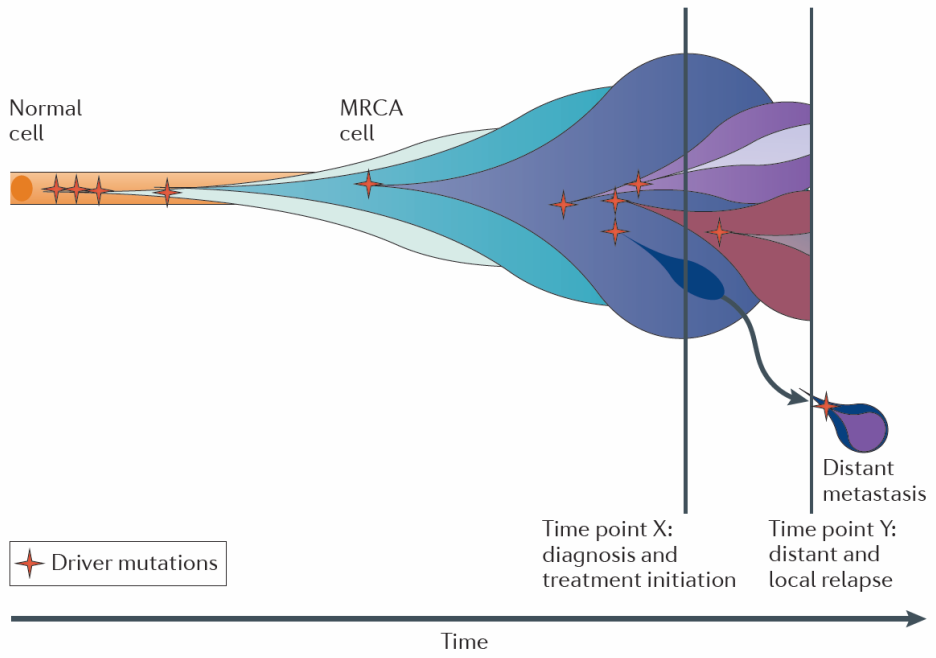


FIGURE 18 – Instabilité génétique, hétérogénéité, résistance aux traitements et capacité métastatique. Ce schéma représente l'acquisition successive d'altérations génétiques qui guide la dynamique de l'hétérogénéité intra-tumorale et l'émergence de sous-clones capables de résister aux traitements ainsi que de se disséminer. Ce schéma représente un modèle d'évolution branchée mais d'autres modèles d'évolution non exclusifs (neutre, linéaire, parallèle, *etc.*) ont été formulés (Venkatesan & Swanton, 2016). MRCA : ancêtre commun le plus récent. Image reprise de Yates & Campbell, 2012.

De précédents travaux de l'équipe ont permis de corréler le niveau d'instabilité génétique des GIST et des synovialosarcomes à leur agressivité tumorale (Lagarde *et al.*, 2012, 2013). Cette instabilité est évaluée par l'index génomique (GI pour *Genomic Index*), correspondant au nombre d'altérations observées (en CGH-*array*) porté au carré sur le nombre de chromosomes

affectés (autosomes uniquement). Cet indicateur génomique est significativement associé au développement métastatique en identifiant un groupe à haut risque (GI≥10 pour les GIST et GI≥1 pour les synovialosarcomes) et un groupe à faible risque (GI<10 pour les GIST et GI<1 pour les synovialosarcomes).

L'agressivité tumorale a également pu être associée au niveau d'instabilité génétique d'autres sous-types histologiques. En effet, les LPS bien différenciés sont des tumeurs peu agressives mais 10% d'entre elles présentent une dédifférenciation, conduisant à la formation d'un DDLPS, tumeur agressive et dont le génome est très remanié (**figure 16 page 48** ; Fletcher *et al.*, 2013). De plus, une étude menée dans les MFS montre une augmentation du nombre d'altérations génomiques et du grade FNCLCC dans les rechutes locales comparativement aux tumeurs primaires (Willems *et al.*, 2006). Il est également à noter que 17% des sarcomes d'Ewing présentent des mutations inactivatrices de *STAG2*, codant pour une sous-unité du complexe de la cohésine (protéine régulant la séparation des chromatides sœurs lors de la ségrégation chromosomique ; Solomon *et al.*, 2011), et sont associées à une augmentation du nombre d'altérations structurales ainsi qu'à la progression tumorale (évaluée en survie globale ; Tirode *et al.*, 2014).

Dans les SGC, le niveau d'instabilité génétique a aussi pu être relié à l'expression d'une signature transcriptomique particulière (Chibon *et al.*, 2010). Avant de la décrire plus en détails (**partie 1.2.3.4 page 57**), nous allons voir par quels moyens le profil transcriptomique peut être obtenu puis comment les altérations génétiques impactent le profil transcriptomique des tumeurs et en quoi l'étude du transcriptome est importante pour mieux comprendre la biologie des sarcomes, notamment au travers des signatures d'expression qui ont pu être établies.

1.2.3 – Le transcriptome : miroir des altérations génétiques

1.2.3.1 – La mesure du transcriptome

Le transcriptome représente l'ensemble des ARN exprimés, issus du processus de transcription. La modification de l'expression génique traduit des changements qui peuvent être reliés à tous les processus biologiques physiologiques tels que l'inflammation, la vascularisation,

l'apoptose, la prolifération, la différenciation, *etc.* (Rosenwald *et al.*, 2003 ; Chen *et al.*, 2012 ; Carter *et al.*, 2012). Puisque le transcriptome est le lien entre le génotype et le phénotype, les altérations génétiques ont un impact mesurable sur le transcriptome.

La mesure des profils transcriptomiques s'est démocratisée avec la technologie de puces d'expression (Schena *et al.*, 1995 ; Golub *et al.*, 1999), mais elle nécessitait de connaitre la séquence des ARN afin de générer les sondes complémentaires. Cette technologie s'est vue être progressivement remplacée par des méthodes de séquençage haut débit et en l'occurrence le RNA-seq (Wang *et al.*, 2009), sans *a priori* sur le transcriptome exprimé. Après les premières études restreintes à l'analyse quantitative de l'expression génique (Bainbridge *et al.*, 2006 ; Mortazavi *et al.*, 2008), l'analyse des séquences des ARN obtenues par des approches haut débit a permis d'identifier des gènes chimériques induits par la transcription (nommés *read-throughs* ; Akiva *et al.*, 2006), de nouveaux gènes et sites d'épissage (Sultan *et al.*, 2008), des altérations ponctuelles (Chepelev *et al.*, 2009), structurales (sous forme de gènes de fusion ; Maher *et al.*, 2009) et enfin de mieux appréhender l'univers des ARN non codants (ENCODE Project Consortium *et al.*, 2007 ; Hafner *et al.*, 2008).

De nombreux protocoles ont pu être mis au point pour le séquençage et l'analyse de données de RNA-seq, que ce soit pour l'analyse globale du transcriptome (déplétion des ARN ribosomiques), du transcriptome codant (sélection par la queue poly(A) des ARN messagers) ou encore l'analyse d'ARN qui peuvent être dégradés (comme ceux extraits de blocs FFPE ; Casamassimi *et al.*, 2017). Les dernières avancées ont permis de quantifier le transcriptome de cellules uniques : l'analyse *single-cell* (Tang *et al.*, 2009), dont la résolution offre l'opportunité de mieux comprendre les effets du micro-environnement, de l'hétérogénéité et de l'évolution tumorale (Cieślik & Chinnaiyan, 2018). Le RNA-seq est ainsi la technique actuelle de choix pour analyser l'ensemble du transcriptome tumoral (Costa *et al.*, 2013).

1.2.3.2 – La spécificité transcriptomique de l'oncogenèse

De nombreuses études transcriptomiques dans les sarcomes (des tissus mous et osseux) ont mis en évidence l'étroite relation entre une altération génétique motrice de l'oncogenèse et le profil transcriptomique qui en résulte. Ainsi, des différences majeures sont observées entre dif-

férents sous-types histologiques, de telle sorte que le transcriptome puisse être envisagé comme le miroir des altérations génétiques qui gouvernent le phénotype tumoral (**figure 19** ; Nielsen *et al.*, 2002 ; Segal *et al.*, 2003 ; Baird *et al.*, 2005 ; Davicioni *et al.*, 2006 ; Francis *et al.*, 2007 ; Nakayama *et al.*, 2007 ; Davicioni *et al.*, 2009 ; Pierron *et al.*, 2012 ; Watson *et al.*, 2018).

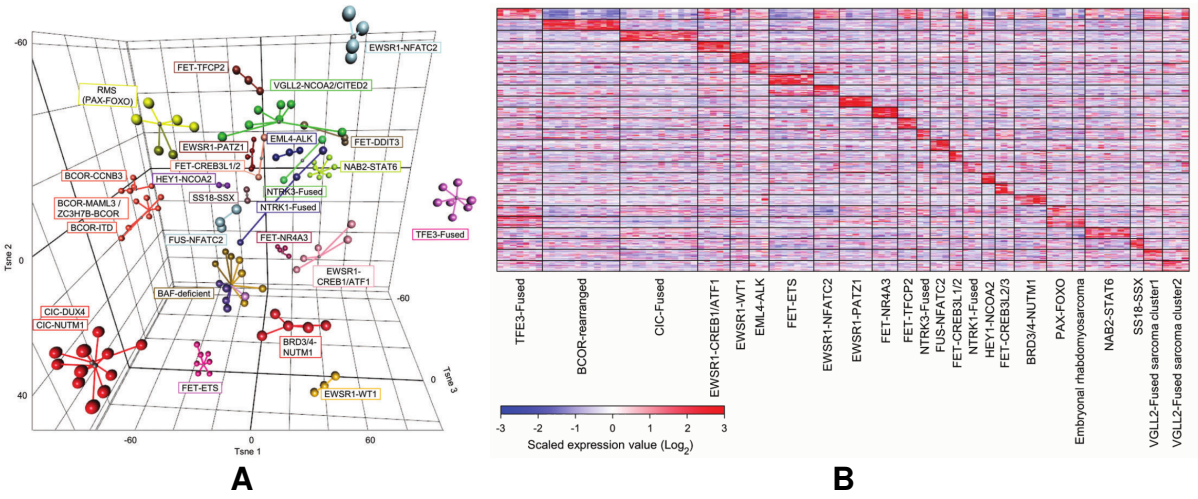


FIGURE 19 – Profils transcriptomiques spécifiques des SGS. **A** : analyse t-*distributed stochastic neighbor embedding* des profils d'expression par déconvolution des données d'expression sur 3 axes. Les tumeurs (points) se regroupent bien selon la présence d'une altération génétique dictant le profil transcriptomique. **B** : top 100 des gènes différentiellement exprimés au travers des 24 groupes de tumeurs. Images reprises de Watson *et al.*, 2018.

Cette spécificité est due à deux facteurs non exclusifs. Le premier facteur est la nature du message oncogénique fort qui est porté par l'altération motrice. Ce message est spécifique et permet d'activer d'importantes voies de signalisation pour l'initiation tumorale, comme les tumeurs qui expriment *BCOR-CCNB3* ou *NAB2-STAT6* et qui sont associées à des oncogenèses distinctes (précisées **partie 1.2.2.2 page 37**). Ceci est principalement dû à la nature des oncogènes forts des SGS qui sont souvent des facteurs de transcription ou des protéines kinases, ainsi qu'à l'activation de la cascade de signalisation qui découle de leurs altérations (Mertens *et al.*, 2015). Le second facteur est le contexte cellulaire qui doit présenter certaines spécificités afin d'initier le développement tumoral. Par exemple, les mutations oncogènes de Kit dans un modèle murin ne permettent le développement de GIST à partir de cellules interstitielles de Cajal que lorsque celles-ci expriment fortement Etv1 (Kwon *et al.*, 2009 ; Chi *et al.*, 2010). Sans ce prérequis, l'initiation tumorale semble exclue. Dans un autre modèle murin, le développement de synovialosarcomes ne se produit qu'à partir de l'expression de *SS18-SSX1* dans

des myoblastes exprimant Myf5 et non dans des cellules musculaires plus avancées dans la différenciation (Haldar *et al.*, 2007).

Par opposition aux SGS, pour lesquels des profils transcriptomiques distincts sont observés, les différents sous-types de SGC ne possèdent que peu de spécificités transcriptomiques (**figure 20** ; Segal *et al.*, 2003 ; Baird *et al.*, 2005), à l'exception des LMS amplifiés pour le gène *MYOCD* (Pérot *et al.*, 2009) qui promeut un message transcriptomique fort et associé à une différenciation musculaire lisse très marquée (Beck *et al.*, 2010 ; The Cancer Genome Atlas Research Network, 2017).

Ainsi, les différents sous-types histologiques de SGC ne sont caractérisés ni par une altération oncogénique spécifique, ni par un profil transcriptomique global distinct. Afin de mieux comprendre la biologie de ces tumeurs mais également de préciser celle des SGS, plusieurs signatures transcriptomiques ont alors été établies.

1.2.3.3 – Les signatures transcriptomiques

Les signatures transcriptomiques sont des regroupements de gènes d'intérêt, pouvant avoir de nombreuses significations biologiques comme une appartenance à une voie de signalisation, un locus génomique, un comportement métabolique, un réseau de co-expression, *etc.* Près de 18000 signatures sont recensées dans la *Molecular Signatures Database* (MSigDB v6.2 ; Subramanian *et al.*, 2005 ; Liberzon *et al.*, 2011, 2015) et près de 3000 dans la *Gene Signature DataBase* (GeneSigDB v4.0 chez l'humain ; cette base n'est plus maintenue depuis 2011 ; Culhane *et al.*, 2010).

Dès la fin des années 90, l'analyse du transcriptome s'est distinguée par sa potentielle application clinique au travers de trois intérêts majeurs : le diagnostic, le pronostic et la prédiction de la réponse au traitement (Roychowdhury & Chinnaiyan, 2016 ; Cieślik & Chinnaiyan, 2018). Dans ces trois catégories, on peut par exemple citer les études pionnières qui ont permis d'établir une signature de 50 gènes capable de distinguer les leucémies myéloïdes aigües et les leucémies lymphoblastiques aigües (Golub *et al.*, 1999), une signature de 70 gènes prédictive du risque métastatique dans les cancers du sein (van de Vijver *et al.*, 2002) et un ratio d'expression entre *HOXB13* et *IL17RB* permettant d'évaluer la réponse au tamoxifène chez des patientes

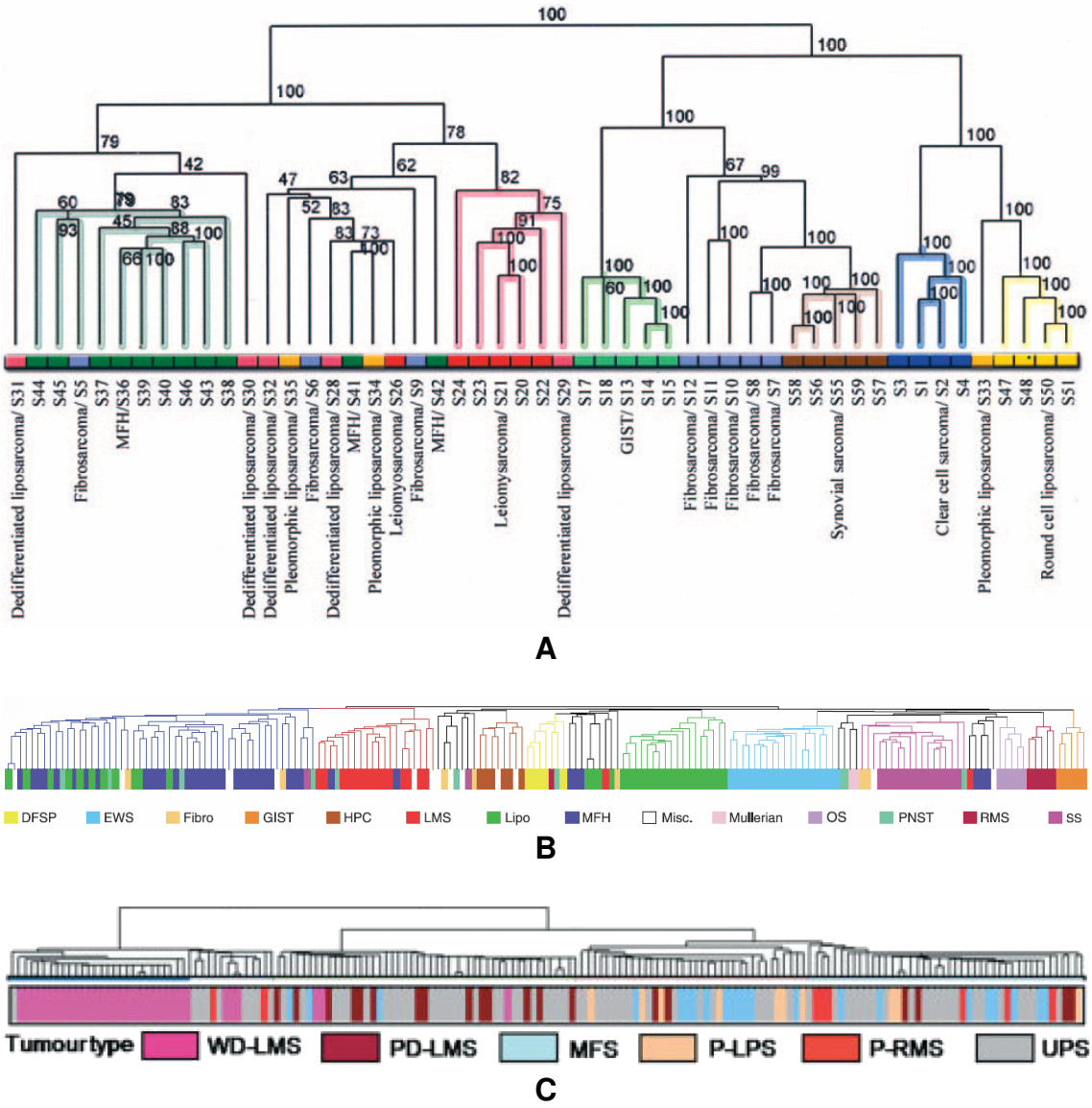


FIGURE 20 – Profils transcriptomiques des sarcomes. **A** : analyse transcriptomique non supervisée d'une série de 51 sarcomes. Image reprise de Segal *et al.*, 2003. **B** : analyse transcriptomique non supervisée d'une série de 181 sarcomes. Image reprise de Baird *et al.*, 2005. Ces deux analyses montrent des similarités transcriptomiques très fortes pour les SGS et des classifications moins distinctes pour les SGC. **C** : analyse transcriptomique non supervisée d'une série de 160 SGC. Image adaptée de Gibault *et al.*, 2011.

atteintes d'un cancer du sein (Ma *et al.*, 2004).

Afin d'aider à la classification des STM, plusieurs signatures ont été établies pour permettre l'identification de différents sous-types histologiques, notamment ceux à génétique simple car ils présentent des profils transcriptomiques bien spécifiques (Nielsen *et al.*, 2002 ; Segal *et al.*, 2003 ; Davicioni *et al.*, 2006 ; Francis *et al.*, 2007 ; Nakayama *et al.*, 2007 ; Davicioni *et al.*, 2009). Toutefois, ces analyses n'ont pas permis de mettre en évidence des marqueurs différentiels entre sous-types histologiques de SGC, excepté pour les LMS présentant

une amplification du gène *MYOCD*.

Malgré l'importance clinique de la prédiction du risque métastatique des SGC (50% d'événements métastatiques à 5 ans), peu d'études se sont attachées à établir des signatures pronostiques dans ces tumeurs. On peut par exemple citer une signature établie à partir de 27 LMS (20 cas non métastatiques et 7 cas métastatiques) et composée de 335 gènes (dont 228 sont sur-exprimés dans les tumeurs métastatiques et 107 sont sur-exprimés dans les tumeurs non métastatiques), mêlant plusieurs voies de signalisation (cycle cellulaire, apoptose, transduction du signal, *etc.* ; Lee *et al.*, 2004). Un autre exemple est une signature établie sur 89 sarcomes pléomorphes (50 cas non métastatiques et 39 cas métastatiques) et composée de 244 gènes, mêlant aussi plusieurs voies de signalisation (hypoxie, métabolisme, adhésion, *etc.* ; Francis *et al.*, 2007). Ces études sont toutefois limitées par plusieurs facteurs majeurs : elles se basent sur des cohortes de petites tailles (<100 cas), définissent de larges signatures (>100 gènes, ce qui est source de sur-interprétation des résultats ; Venet *et al.*, 2011), sans validation sur une cohorte indépendante et ne comparent pas la valeur pronostique obtenue par rapport au grade en analyse multivariée. Il est également à noter qu'aucun gène n'est commun entre ces deux signatures.

Plutôt que de se baser sur les caractéristiques cliniques des tumeurs pour prédire leur agressivité, une étude s'est appuyée sur une caractéristique génétique d'intérêt : l'instabilité chromosomique (Carter *et al.*, 2006). L'analyse menée sur 18 jeux de données d'expression (provenant de différents types de cancers), conduit à définir une signature de 70 gènes ayant d'importantes fonctions de régulation de la réplication et de la ségrégation des chromosomes. La sur-expression de ces gènes est associée à une agressivité tumorale accrue pour les adénocarcinomes du poumon, les mésothéliomes, les gliomes, les médulloblastomes et les cancers du sein (en survie globale et/ou survie sans métastase). D'intérêt clinique prometteur et en adéquation avec la génétique des SGC, cette signature n'était toutefois pas associée à l'agressivité des sarcomes pléomorphes (Chibon *et al.*, 2010). Des travaux de l'équipe ont alors établis un lien entre instabilité génétique et agressivité tumorale dans ces tumeurs.

1.2.3.4 – CINSARC : instabilité génétique et agressivité

La définition d'une signature d'instabilité propre aux SGC s'est articulée autour de trois caractéristiques principales (**figure 21**). La première est le niveau d'instabilité génétique (évalué par CGH-*array*), en distinguant les tumeurs peu réarrangées (<20 altérations) et réarrangées (>35 altérations). La seconde est le grade FNCLCC, valeur pronostique de référence, en distinguant les tumeurs de bas grades (1 et 2) et de haut grade (3). La troisième est le groupe pronostique donné par la signature d'instabilité décrite ci-dessus (Carter *et al.*, 2006) en distinguant les tumeurs qui expriment faiblement la signature et celles qui l'expriment fortement. Le regroupement de tous les gènes différentiellement exprimés entre ces conditions a permis d'établir une signature de 67 gènes nommée *Complexity INdex in SARComas* (CINSARC ; **table 4** ; Chibon *et al.*, 2010).

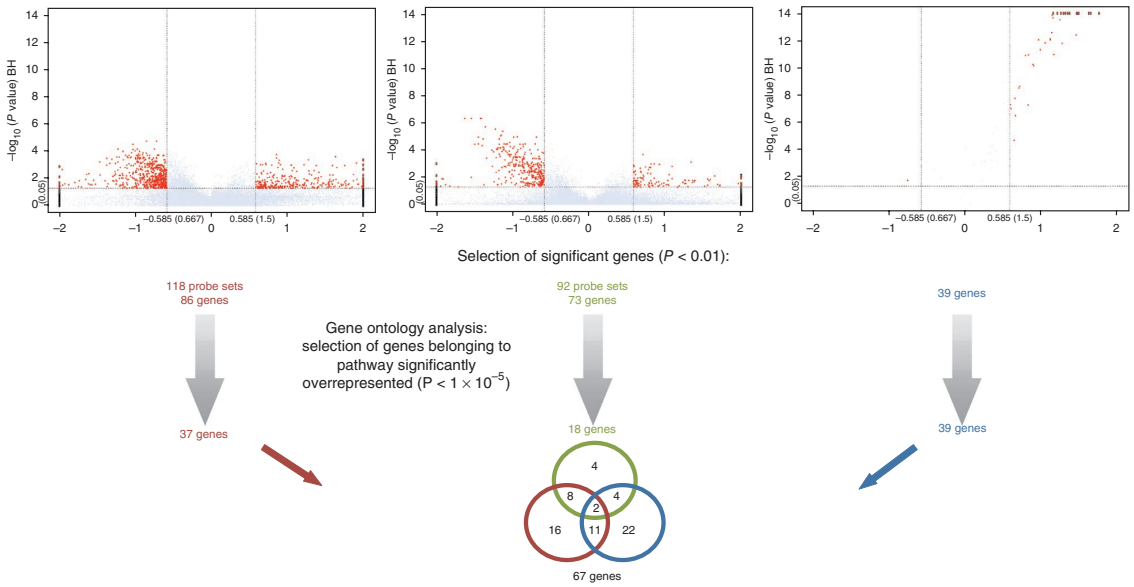


FIGURE 21 – Définition de la signature CINSARC. L'établissement de la signature a été effectuée en se basant sur l'union de trois analyses d'expression différentielle en fonction du nombre d'altérations génomiques (à gauche ; <20 contre >35), du grade FNCLCC (au milieu ; grades 1 et 2 contre 3) et d'une signature d'instabilité (à droite ; faible contre forte expression ; Carter *et al.*, 2006). Les gènes obtenus par les deux premières analyses ont été filtrés pour ne conserver que ceux qui étaient les plus associés à une voie biologique (sans *a priori* sur ces voies). Image reprise de Chibon *et al.*, 2010.

Dans deux cohortes indépendantes de SGC (183 et 127 cas), CINSARC était significativement associée au devenir métastatique des patients en stratifiant ceux à faible risque (groupe C1) et ceux à haut risque métastatique (groupe C2 ; **figure 22A**). De plus, sa valeur pronostique était significativement supérieure à celle du grade FNCLCC (**figure 22B**), démontrant un po-

ANLN	*CDC7*	*CKAP5*	*KIF14*	*NUF2*	*SPC25*
ASPM	*CDC20*	*CKS2*	*KIF15*	*OIP5*	*TOP2A*
AURKA	*CDC45*	*ECT2*	*KIF18A*	*PBK*	*TPX2*
AURKB	*CDCA2*	*ESPL1*	*KIF20A*	*PLK4*	*TRIP13*
BIRC5	*CDCA3*	*FANCI*	*KIF23*	*PRC1*	*TTK*
BORA	*CDCA8*	*FBXO5*	*MAD2L1*	*PTTG1*	*UBE2C*
BUB1	*CDK1*	*FOXM1*	*MCM2*	*RAD51AP1*	*ZWINT*
BUB1B	*CENPA*	*H2AFX*	*MCM7*	*RNASEH2A*	
CCNA2	*CENPE*	*HP1BP3*	*MELK*	*RRM2*	
CCNB1	*CENPL*	*KIF2C*	*NCAPH*	*SGO2*	
CCNB2	*CEP55*	*KIF4A*	*NDE1*	*SMC2*	
CDC6	*CHEK1*	*KIF11*	*NEK2*	*SPAG5*	

TABLE 4 – Les 67 gènes de la signature CINSARC. Table adaptée de Chibon *et al.*, 2010.

tentiel intérêt clinique. Elle fut aussi testée et validée dans d'autres pathologies : des cancers du sein, des lymphomes diffus à grandes cellules B (Chibon *et al.*, 2010) et par la suite dans deux études sur des sous-types de SGS : des GIST (Lagarde *et al.*, 2012) et des synovialosarcomes (Lagarde *et al.*, 2013).

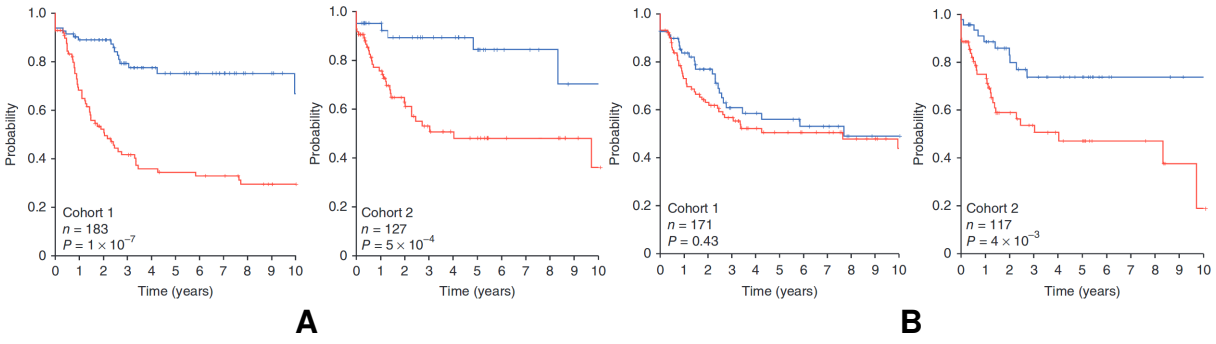


FIGURE 22 – Valeur pronostique de CINSARC comparativement au grade FNCLCC. **A** : valeur pronostique de CINSARC dans deux cohortes indépendantes. **B** : valeur pronostique du grade FNCLCC sur les cas évaluables des mêmes cohortes. L'analyse multivariée démontre aussi une meilleure valeur pronostique avec la signature CINSARC qu'avec le grade FNCLCC. Images reprises de Chibon *et al.*, 2010.

Ces gènes sont impliqués dans le contrôle de la mitose et de la ségrégation des chromosomes, ce qui concorde avec le niveau d'instabilité où plus les tumeurs expriment cette signature et plus leur génome est remanié (Chibon *et al.*, 2010). En outre, la sur-expression individuelle de plusieurs de ses gènes tels que *MAD2L1*, *BUB1*, *CCNB1*, *CCNB2* ou *ESPL1* provoque des

défauts de ségrégation chromosomique, conduisant à des altérations génétiques et à des tumeurs chez la souris (Sotillo *et al.*, 2007 ; Ricke *et al.*, 2011 ; Nam & van Deursen, 2014 ; Mukherjee *et al.*, 2014).

À l'ère du développement de très nombreuses signatures pronostiques, une étude émit des doutes sur la pertinence biologique et clinique de la plupart d'entre elles (Venet *et al.*, 2011). Ainsi, sur 47 signatures pronostiques définies dans le cancer du sein, 28 (60%) ne prédisaient pas plus le devenir clinique que des signatures générées aléatoirement (de taille identique aux marqueurs testés). De plus, 11 étaient moins bonnes prédictrices que la médiane de ces signatures aléatoires. Finalement, plus de 90% des signatures aléatoires de plus de 100 gènes étaient significativement associées au devenir clinique. En conséquence, la robustesse de la signature CINSARC fut testée dans quatre jeux de données indépendants et aucune signature générée aléatoirement n'obtint une valeur pronostique supérieure dans l'ensemble de ces cohortes (Brulard & Chibon, 2013).

Puisque CINSARC est une signature robuste et un meilleur marqueur de l'agressivité que le grade FNCLCC, elle pourrait tout à fait avoir sa place dans la prise en charge thérapeutique des patients en évaluant le risque métastatique des STMA (Neuville *et al.*, 2014). Toutefois, la mise en application clinique d'une signature transcriptomique est soumise à d'importantes validations, notamment au travers d'essais cliniques. Ainsi, parmi les très nombreuses signatures pronostiques établies dans l'ensemble des cancers, peu sont parvenues jusqu'à une application clinique effective (Parkinson *et al.*, 2014 ; Schneider *et al.*, 2015 ; Michiels *et al.*, 2016).

1.2.3.5 – Les signatures pronostiques utilisées dans un contexte clinique

À ce jour, il n'existe aucune signature pronostique présentant un avantage clinique démontré au sein d'un essai clinique dans les STM. Néanmoins, l'analyse du transcriptome d'autres tumeurs a pu mettre en évidence l'existence de profils dont le niveau d'expression était corrélé à la progression tumorale. Il existe ainsi un petit nombre de signatures (moins d'une vingtaine) qui sont entrées dans les habitudes cliniques afin d'apporter des précisions sur le risque d'agressivité tumorale (**table 5**).

Type de cancer	Signature	Gènes	Prédiction	Référence
Sein	MammaPrint	70	Métastase	van de Vijver *et al.*, 2002
	Oncotype DX	21[a]	Métastase	Paik *et al.*, 2004
	Prosigna (PAM50)	50	Métastase	Parker *et al.*, 2009
Poumon	GeneFX *Lung*	15	Rechute	Der *et al.*, 2015
Protate	Prolaris	46[b]	Survie globale	Cuzick *et al.*, 2011
	Decipher	22	Métastase	Erho *et al.*, 2013
Côlon	ColoPrint	18	Métastase	Salazar *et al.*, 2011
	GeneFx *Colon*	482	Rechute	Kennedy *et al.*, 2011

TABLE 5 – Exemples de signatures pronostiques utilisées en clinique. [a] : 16 gènes de test et cinq de référence. [b] : 31 gènes de test et 15 de référence.

Il est important de préciser que chacune de ces signatures doit être replacée dans son contexte clinique. La MammaPrint (van de Vijver *et al.*, 2002) est, par exemple, indiquée pour les patientes atteintes d'un cancer du sein de stades I/II, de moins de 5cm et avec au plus trois ganglions lymphatiques atteints. L'Oncotype DX (Paik *et al.*, 2004) est, quant à elle, indiquée pour les patientes atteintes d'un cancer du sein de stades I/II, positif pour les récepteurs aux œstrogènes, statut HER2 négatif et avec au plus trois ganglions lymphatiques atteints. Ces signatures aident ainsi à évaluer l'agressivité des tumeurs et orienter le traitement qui sera proposé aux patientes.

Au-delà de modifier les pratiques cliniques en identifiant les patientes à risque de façon plus précise, ces signatures permettent aussi de modifier les pratiques thérapeutiques. Par exemple, les signatures MammaPrint et Oncotype DX ont été récemment utilisées dans des essais cliniques de phase 3 (avec respectivement près de 6700 et 9700 patientes ; Cardoso *et al.*, 2016 ; Sparano *et al.*, 2018) dans le but de tester l'apport d'une chimiothérapie adjuvante (en plus de l'hormonothérapie) dans la prise en charge des patientes. À titre informatif, la MammaPrint et Prosigna disposent d'une autorisation de mise sur le marché américain depuis respectivement 2007 et 2013.

Savoir quelle est la meilleure signature reste une question en suspens. De nombreuses études rétrospectives ont été menées en ce sens sur le cancer du sein et ont collectivement démontré qu'aucune ne surpassait les autres (Fan *et al.*, 2006 ; Weigelt *et al.*, 2010 ; Dowsett *et al.*, 2013 ; Sestak *et al.*, 2018). De plus, puisque ces signatures sont capables de déduire

plusieurs sous-types tumoraux (luminal A, luminal B, *basal-like*, statut HER2 positif, *etc.*), les similitudes de classifications inter-signatures ont été mesurées, relevant de 20 à 40% de taux de discordance (Kelly *et al.*, 2012 ; Alvarado *et al.*, 2015 ; Bartlett *et al.*, 2016).

Il est également à noter que ces signatures utilisent différentes technologies de quantification d'expression : RT-qPCR pour OncotypeDX, système nCounter (NanoString) pour Prosigna et puces d'expression (Agendia) pour MammaPrint. Enfin, les différentes méthodes d'application de ces signatures prennent en charge l'analyse d'ARN extraits de blocs FFPE, afin de s'adapter aux pratiques des laboratoires (Chang *et al.*, 2017).

Au-delà de démontrer des bénéfices médicaux avérés lors d'essais cliniques, un marqueur moléculaire doit ainsi pouvoir être compatible avec les pratiques courantes d'analyses. Dans les sarcomes, l'utilisation de la signature pronostique CINSARC serait donc soumise à cette même contrainte si elle est amenée à être quotidiennement appliquée (Neuville *et al.*, 2014).

1.2.4 – La régulation transcriptomique en absence d'altération génomique

Les messages transcriptomiques exprimés dans les cellules tumorales ne sont pas simplement le reflet des altérations génétiques qui y sont présentes. Le transcriptome est hautement dynamique et de nombreux facteurs interviennent entre une altération de l'ADN et le phénotype observé. La biologie des ARN est ainsi soumise à de nombreuses régulations qui s'opèrent à plusieurs niveaux : épigénétique, transcriptionnel et post-transcriptionnel (**figure 23** ; Blair & Yan, 2012). Nous allons ici nous intéresser aux régulations épigénétiques et post-transcriptionnelles les mieux décrites, qui modulent l'expression génique et qui ont un impact démontré dans la biologie des sarcomes.

1.2.4.1 – L'épigénétique

L'épigénétique est définie par des modifications stables et héritables de l'expression génique qui ne sont pas dues à une altération génomique (Berger *et al.*, 2009). Les acteurs les plus décrits dans cette catégorie sont la méthylation de l'ADN et les modifications d'histones

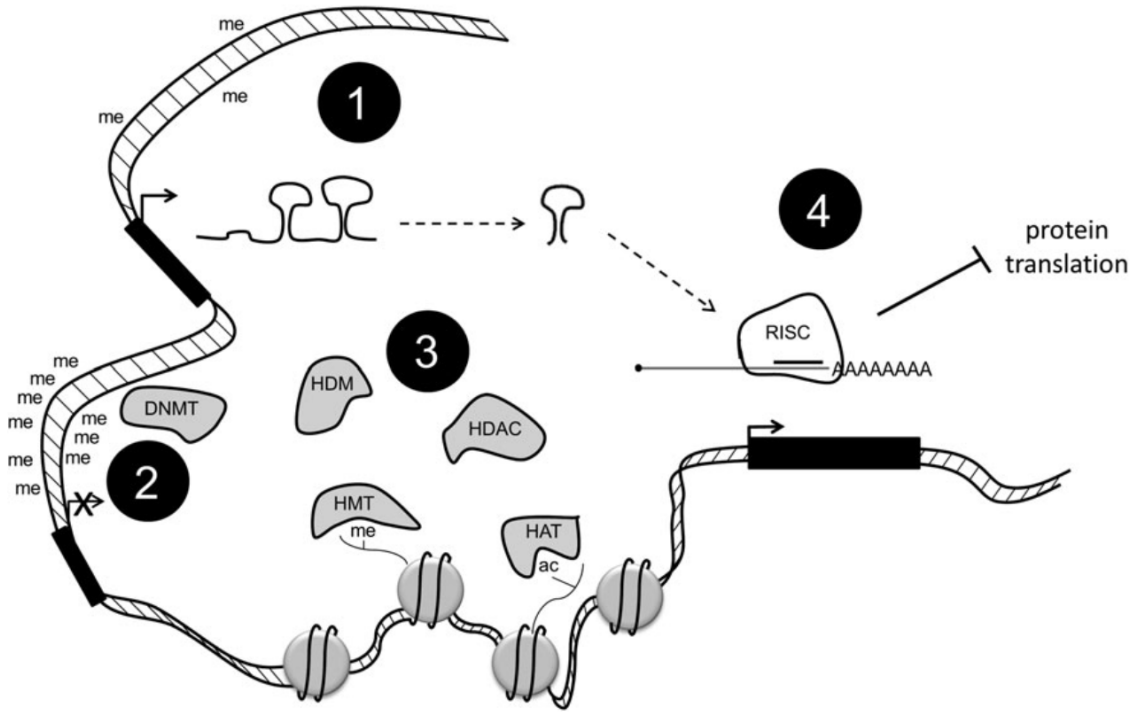


FIGURE 23 – Schéma des principaux acteurs de la régulation de la transcription. **1** : l'hypométhylation des régions promotrices des gènes est associée à leur expression. **2** : l'hyperméthylation des promoteurs est par opposition associée à la répression d'expression. Cette méthylation est rendue possible par les méthyltransférases d'ADN (DNMT). **3** : la modification d'histones modifie la condensation de la chromatine et module de ce fait l'expression génique. Ces modifications sont principalement la méthylation (par ajout via les méthyltransférases ou retrait par les déméthylases, respectivement HMT et HDM) et l'acétylation (par ajout via les acétyltransférases ou retrait par les désacétylases, respectivement HAT et HDAC). **4** : modulation de l'expression génique au niveau post-transcriptionnel par un ARN interférant. Dans cet exemple, l'expression d'un micro-ARN conduit à la dégradation de sa cible, un ARN messager, via le complexe RISC clivant cet ARN et empêchant donc sa traduction en une protéine fonctionnelle. me : méthylation, ac : acétylation. Image reprise de Blair & Yan, 2012.

(Blair & Yan, 2012 ; Allis & Jenuwein, 2016). Ces deux processus vont conduire à une variation d'expression par modification des nucléotides de l'ADN sans en modifier la séquence ou par modification de son état de condensation. L'épigénétique jouerait un rôle crucial dans les premières étapes de l'initiation tumorale en facilitant l'oncogenèse (inactivation de gènes suppresseurs de tumeurs, pression d'acquisition d'altérations spécifiques, *etc.* ; Baylin & Ohm, 2006 ; Feinberg *et al.*, 2006).

La méthylation de l'ADN

La méthylation de l'ADN consiste en l'ajout d'un groupement méthyle sur une cytosine en 5' d'une guanine, formant le dinucléotide CpG. Ce dinucléotide est globalement sous-représenté dans le génome humain (1%) mais certaines de ses séquences, de 500pb à 4kb, sont enrichies en CpG (>50%) et sont ainsi nommées îlots CpG (Bird, 2002 ; Takai & Jones, 2002). Près de 70%

des gènes ont des îlots CpG localisés à proximité (quelques kb) de leurs régions promotrices et qui participent à leur régulation (Saxonov *et al.*, 2006 ; Deaton & Bird, 2011). La méthylation de l'ADN agit en effet en tant que répresseur d'expression par deux mécanismes : les méthylations peuvent directement perturber la fixation de la machinerie de transcription ou bien recruter des protéines possédant une forte affinité pour les 5-méthylcytosines (*methyl-CpG-binding domain proteins* ; Moore *et al.*, 2013). Il est également décrit que la déméthylation globale du génome tumoral est une caractéristique très fréquente et supposée très précoce dans l'initiation tumorale (Feinberg & Vogelstein, 1983 ; Eden *et al.*, 2003 ; Kanai, 2010).

La modification de méthylation de l'ADN la plus décrite dans les sarcomes est celle mesurée dans les synovialosarcomes. L'expression de la protéine chimérique SS18-SSX modifie en effet la conformation de la chromatine via son intégration au sein du complexe SWI/SNF, ce qui conduit à une modification de l'accessibilité de la chromatine, aboutissant à un profil de méthylation d'ADN spécifique dans ces tumeurs (The Cancer Genome Atlas Research Network, 2017 ; Wu *et al.*, 2017). Une analyse intégrative menée sur les MFS révèle d'importantes variations de méthylation au sein de ce sous-type histologique, avec d'un côté un groupe globalement hyperméthylé, de bon pronostic, associé à des amplifications de *MDM2* ainsi qu'à des mutations non-sens de *TET2*, et de l'autre côté un groupe globalement hypométhylé, de mauvais pronostic, associé à des délétions de *CDKN2A*, de *RB1* ainsi qu'à des altérations de *TP53* (ponctuelles et structurales ; Ogura *et al.*, 2018).

D'importantes différences de méthylation entre sous-types histologiques sont également relevées mais dont leur origine n'est pas encore totalement compris, pouvant tout aussi bien être associée à des altérations génétiques et/ou des contextes cellulaires différents). Par exemple, les LMS utérins (tumeurs hautement malignes) ont un niveau de méthylation de l'ADN inférieur à ceux des léiomyomes utérins (tumeurs bénignes) et des LMS des tissus mous (tumeurs malignes ; Miyata *et al.*, 2015 ; The Cancer Genome Atlas Research Network, 2017). Dans les LMS, des inactivations de *CDKN2A* et de *PTEN* ont également été rapportées via l'hyperméthylation de leurs régions promotrices, ce qui pourrait se substituer à leurs délétions génomiques (Kawaguchi *et al.*, 2003, 2005).

En plus de son implication dans la régulation de la transcription, la méthylation de l'ADN

peut avoir des conséquences sur l'acquisition de nouvelles altérations, participant ainsi à entretenir un état d'instabilité génétique. La désamination spontanée des 5-méthylcytosines aboutit à la formation d'une thymine (Pfeifer, 2006 ; Alexandrov *et al.*, 2013). La protéine codée par *MGMT* convertie les 6-O-méthylguanine (qui s'apparient avec une thymine et non pas à une cytosine) en guanine. La répression épigénétique de *MGMT* dans 15% à 30% des STM (Kawaguchi *et al.*, 2006 ; Kim *et al.*, 2009) limite cette reconversion, augmentant ainsi le taux de substitution G>A. Les méthylations aberrantes ont aussi été mises en cause dans la formation de cassures génomiques. Les G-quadruplex sont des structures secondaires stables que peuvent adopter des segments d'ADN et d'ARN riches en guanine. L'hypométhylation rapportée dans ces régions pourrait sensibiliser l'ADN à des cassures génomiques dans les cancers colorectaux, du sein et les ostéosarcomes (De & Michor, 2011).

La modification des histones

Les histones sont les protéines autour desquelles l'ADN s'enroule et qui interviennent dans la formation des nucléosomes (octamères d'histones), permettant la condensation et la décondensation de l'ADN (Strahl & Allis, 2000). Les histones possèdent des extrémités N-terminales qui peuvent être modifiées afin de moduler le niveau de condensation de l'ADN et participer à l'expression des gènes. Les modifications d'histones consistent en l'ajout de divers groupements tels que les acétyles, les méthyles (mono, di ou triméthyles), les phosphates, les ubiquitines, *etc.* (Khorasanizadeh, 2004). La permissivité de la transcription ou sa répression dépend de la nature de la modification, de l'acide aminé impacté et de l'histone affectée (Füllgrabe *et al.*, 2011). Par exemple, la triméthylation de la lysine 4 de l'histone H3 (H3K4me3) permet la transcription alors que la même modification sur la lysine 9 (H3K9me3) a un effet répresseur.

Des modifications d'histones particulières sont observées dans les synovialosarcomes, dans lesquels le complexe SWI/SNF intégrant SS18-SSX permet la perte de la tri-méthylation de H3K27m3 (associée à la répression de la transcription ; McBride *et al.*, 2018). Ceci est expliqué par la capacité du complexe SWI/SNF à évincer le complexe *Polycomb Repressive Complex* 2 (PRC2 ; fonction de méthylation d'histones) de sa localisation génomique, sachant que le remplacement de SS18 par SS18-SSX dans le complexe SWI/SNF augmente cette capacité

d'éviction (Kadoch *et al.*, 2017). Des mutations de type perte de fonction des gènes *EED* et *SUZ12*, impliqués dans le complexe PRC2, sont rapportées dans les MPNST, ce qui conduit aussi à une diminution de la tri-méthylation de H3K27m3 (Pekmezci *et al.*, 2017). À l'inverse, les autres sous-types histologiques (LMS, RMS, DDLPS, GIST, *etc.*) ne présentent pas cette diminution de méthylation d'histone particulière (Mito *et al.*, 2017) et la sur-expression de *PRC2* dans ces sous-types est associée à un mauvais pronostic (Cho *et al.*, 2018).

1.2.4.2 – La régulation post-transcriptionnelle

La régulation post-transcriptionnelle s'effectue lors de nombreuses étapes dans la vie des ARN : leur maturation, leur export du noyau vers le cytoplasme, leur édition, leur dégradation, *etc.* (Corbett, 2018). Ces dernières années et notamment grâce aux techniques de séquençage haut débit, d'importantes régulations géniques par les ARN non codants ont été observées et particulièrement pour deux classes d'entre eux : les micro-ARN (miRNA pour *micro*RNA) et les longs ARN non codants (lncRNA pour *long non-coding* RNA ; Suzuki *et al.*, 2013 ; Fang & Fullwood, 2016). L'expression de ces ARN a ainsi pu être corrélée à des phénotypes tumoraux et représente un moyen de mieux comprendre les étapes menant à l'oncogenèse en l'absence d'altération génétique.

Les micro-ARN

Les miRNA sont de courts (20 à 25pb) ARN non codants qui participent activement à la régulation négative de l'expression génique par hybridation avec l'ARN du gène cible (en 3' ou en 5' UTR ; Lee *et al.*, 2009 ; Bartel, 2009). Cette hybridation empêche alors la machinerie de traduction de synthétiser la protéine. Alternativement, l'hybridation entre le miRNA et l'ARN cible peut s'effectuer au sein du complexe RNA-*induced silencing complex* (RISC), dont la fonction est de cliver les ARN cibles. D'autres types d'ARN non codants peuvent également utiliser ce complexe (*small interfering* RNA, *Piwi-interacting* RNA, *etc.* ; Pratt & MacRae, 2009).

On dénombre plus de 1900 miRNA dans le génome humain et leurs séquences sont très bien conservées entre les espèces (Kozomara *et al.*, 2019). Ils permettent la régulation d'une très grande variété de processus biologiques (prolifération, différenciation, apoptose, *etc.*) en

ciblant des gènes bien spécifiques et leur expression est modifiée dans les cellules cancéreuses (Suzuki *et al.*, 2013). De nombreuses associations entre l'expression de certains "oncomirs" et l'activation/inactivation de voies biologiques (apoptose, différenciation, prolifération, *etc.*) ont été relevées (Esquela-Kerscher & Slack, 2006).

Dans les sarcomes, de nombreuses associations entre miRNA et pronostics ont été établies (Smolle *et al.*, 2017). On retrouve par exemple 12 miRNA (mir-181b, mir-532, let-7d, mir-98, mir-660, mir-106b, mir-185, mir-93, mir-425, mir-501, mir-224 et mir-15b) dont la sur-expression est significativement associée à un devenir clinique péjoratif dans les LMS (The Cancer Genome Atlas Research Network, 2017). Des spécificités d'expression de miRNA sont également observées en fonction de la ligne de différenciation tumorale. Par exemple, les LMS sur-expriment mir-143 et mir-145 (même locus génomique), dont l'expression est sous le contrôle de la myocardine et du SRF, facteurs de transcription permettant la différenciation musculaire lisse (Cordes *et al.*, 2009 ; The Cancer Genome Atlas Research Network, 2017). Les RMS quant à eux, sur-expriment mir-206 et mir-133-1 car ils sont sous le contrôle de MyoD et de la myogénine, facteurs de transcription contrôlant la différenciation musculaire striée (Rota *et al.*, 2011).

Les longs ARN non codants

Les lncRNA sont des ARN de plus de 200pb qui ne codent pas pour une protéine (Zampetaki *et al.*, 2018). Il existe plus de 16000 lncRNA chez l'Homme (Frankish *et al.*, 2019) dont seulement 11% seraient exprimés, mais leur profil d'expression est plus spécifique de l'origine tissulaire que les gènes codants (Derrien *et al.*, 2012).

Les fonctions des lncRNA peuvent être définies selon cinq grandes catégories (Ørom *et al.*, 2010 ; Wang & Chang, 2011). Les lncRNA de type *signal* régulent une activité transcriptionnelle ou de signalisation moléculaire, servant ainsi d'intermédiaire entre différentes voies de signalisation. Les lncRNA de type *decoy* ont une fonction de leurre pour la régulation de l'expression génique et peuvent par exemple servir d'appât pour des miRNA. Les lncRNA de type *guide* permettent de guider des protéines vers une localisation cellulaire spécifique. Ceux de type *scaffold* organisent la structuration de complexes protéiques ou induisent de simples rapprochements spatiaux. Enfin, les lncRNA de type *enhancer* contrôlent la structure de la

chromatine, permettant ainsi la transcription de la même manière que les *enhancers*.

Les lncRNA peuvent être impliqués dans des processus d'oncogenèse et le rôle de quelques uns a pu être décrit dans les sarcomes. Par exemple, le lncRNA *HOTAIR* va permettre de relier deux complexes : le complexe PRC2 et le complexe LSD1/CoREST/REST. Ces deux complexes interviennent dans la modification d'histones, où PRC2 est impliqué dans la formation de H3K27me3 alors que LSD1/CoREST/REST va permettre la déméthylation de H3K4me2, également répresseur de la transcription. Le complexe *HOTAIR*/PRC2/LSD1 ainsi formé (type *scaffold*) est alors un puissant régulateur négatif de la transcription via ces modifications épigénétiques (Tsai *et al.*, 2010). La sur-expression de *HOTAIR* est observée dans plusieurs cancers (tumeurs gastriques, du poumon, du côlon, *etc.* ; Hajjari & Salavaty, 2015) dont les sarcomes (Milhem *et al.*, 2011) et est associée à une acquisition de capacités métastatiques. Un autre exemple est celui de *TUG1* (type *decoy*) dans les ostéosarcomes dont l'expression permet la régulation d'expression de *POU2F1*. Alors que *POU2F1* est régulé négativement par mir-9, *TUG1* va pouvoir se lier à ce miRNA, favorisant ainsi l'expression de *POU2F1* ce qui participe ainsi à la progression tumorale de ces tumeurs (Xie *et al.*, 2016).

1.3 – En résumé

Les SGC forment un groupe très hétérogène de tumeurs dont l'oncogenèse reste à ce jour inexpliquée car elles ne présentent pas d'altération génétique récurrente et motrice qui puisse expliquer leur développement tumoral. Alors que les analyses transcriptomiques menées sur l'ensemble des STM ont permis d'identifier des profils spécifiquement exprimés dans les SGS, en relation avec le message oncogène délivré par leurs altérations motrices, le message oncogène des SGC reste encore incompris. Notamment, les analyses transcriptomiques menées dans ces tumeurs, principalement par puces d'expression, n'ont pas permis de mettre en évidence une quelconque spécificité oncogénique parmi l'ensemble des sous-types histologiques des SGC.

Une des récurrences notables du génome de ces tumeurs est leur important niveau de remaniement. L'impact de ce remaniement génomique global sur le devenir métastatique des SGC a d'ailleurs pu être établi. Ainsi, l'une des avancées majeures de ces dernières années fut l'établissement d'une signature transcriptomique, associée à l'instabilité génétique des SGC et permettant de prédire le devenir clinique des patients en estimant le risque d'un événement métastatique. La signature CINSARC est donc à la fois d'intérêt génétique, afin de mieux comprendre l'impact de ces remaniements génomiques, et clinique, afin de mieux estimer le risque métastatique des sarcomes. Toutefois, le rôle biologique de l'expression de cette signature ainsi que son applicabilité clinique restent encore à être définis.

2 | Objectifs

Mise en application clinique de la signature CINSARC

De précédents travaux de l'équipe ont établi un lien entre l'importante instabilité génétique observée dans les SGC et leur agressivité tumorale. La signature transcriptomique CINSARC est en effet le reflet transcriptomique de ces altérations génomiques et permet de stratifier les tumeurs en fonction de leur risque métastatique. La valeur pronostique de cette signature étant supérieure à celle du grade FNCLCC, actuelle référence internationale, il devenait alors nécessaire de pouvoir appliquer CINSARC en routine clinique. Le premier objectif de cette thèse vise à rendre possible une application clinique de la signature CINSARC, de façon à ce qu'elle puisse être compatible avec les pratiques courantes des centres de soins. Ceci implique de travailler avec du matériel FFPE, standard de conservation des tissus tumoraux dans un contexte diagnostique mais dont le processus de fixation dégrade les acides nucléiques. De plus, la technologie visant à déterminer l'expression de la signature doit pouvoir être accessible en routine clinique. Répondre à ces deux contraintes a ainsi été à la base de mes travaux de thèse.

Identification de gènes de fusion dans les SGC

Les connaissances acquises sur la génétique des sarcomes pléomorphes ne permettent pas d'expliquer par quel(s) moyen(s) ces tumeurs sont initiées. Les analyses de cytogénétique employées jusqu'alors ont bien mis en évidence des remaniements génomiques, notamment des translocations, mais sans pouvoir précisément déterminer s'ils étaient reliés à une altération spécifique. Dans le but d'identifier des altérations spécifiques et récurrentes dans les SGC, nous nous sommes intéressés aux gènes de fusion. En effet, l'expression de gènes de fusion oncogènes, principalement permise par translocation, est une voie d'oncogenèse observée dans 20% des sarcomes. Le second objectif de cette thèse est ainsi de déterminer si les réarrangements génomiques des SGC peuvent permettre l'expression d'un gène de fusion oncogène.

3 | Résultats

3.1 – Instabilité génétique, CINSARC et progression tumorale

3.1.1 – Contexte

L'analyse pan-cancer des altérations ponctuelles et structurales a mis en évidence trois catégories de tumeurs (**figure 16 page 48** ; The Cancer Genome Atlas Research Network, 2017). Une première catégorie, représentée par une majorité de cancers, présente à la fois une faible charge mutationnelle et une faible variation du nombre de copies. Une seconde catégorie, principalement représentée par les mélanomes et cancers du poumon, présente une très importante charge mutationnelle et peu de remaniements génomiques. À l'opposé, la dernière catégorie, principalement représentée par les sarcomes, présente peu d'altérations ponctuelles mais de très nombreuses altérations structurales.

À l'interface entre le niveau d'instabilité génétique des SGC et le devenir clinique des patients, de précédents travaux de l'équipe ont permis de mettre en évidence l'expression d'une signature transcriptomique spécifique (**partie 1.2.3.4 page 57** ; Chibon *et al.*, 2010). Cette signature, nommée CINSARC, permet de stratifier ces tumeurs en deux groupes ayant des pronostics significativement différents (groupes C1 et C2 respectivement associés à un bon et à un mauvais pronostic). Puisque CINSARC est un meilleur marqueur pronostique que le grade FNCLCC, nous nous sommes ainsi intéressés au développement de cette signature afin de la rendre accessible en routine clinique. Dans un second temps, nous nous sommes penchés sur les causes de l'expression de cette signature pour tenter de déterminer quels facteurs peuvent conditionner son expression.

3.1.2 – Vers une application clinique de la signature

Jusqu'à présent, la prise en charge thérapeutique des patients atteints de sarcomes localisés et résécables est guidée par le grade FNCLCC (**figure 3 page 25** ; Casali *et al.*, 2018) dont la valeur pronostique reste la référence internationale (**partie 1.1.4 page 22**). Néanmoins, ce système d'évaluation comporte plusieurs limitations : il n'est pas applicable à tous les sous-types histologiques de STMA, sa reproductibilité peut varier entre pathologistes, il est peu informatif pour les tumeurs de grade 2 (40% des cas) et il est difficile à évaluer sur micro-biopsies et/ou les pièces d'exérèses traitées en situation néoadjuvante (Coindre *et al.*, 1986, 2001 ; Coindre, 2006).

L'utilisation de la signature CINSARC, en plus d'être un meilleur marqueur que le grade FNCLCC (Chibon *et al.*, 2010 ; Lagarde *et al.*, 2013), permettrait aussi d'outrepasser les limites actuelles du grade. Toutefois, avant de pouvoir l'utiliser de façon quotidienne, un travail d'adaptation de cette signature devait être réalisé car son utilisation était alors limitée aux tissus congelés et restreinte aux puces d'expression. Ces pratiques ne sont pas en adéquation avec un contexte d'application clinique pour deux raisons (**partie 1.2.3.5 page 59**). La première est que le support matériel de référence est le bloc FFPE mais dont le processus de fixation des tissus et de conservation conduit à la dégradation des acides nucléiques, ce qui est particulièrement délétère pour l'intégrité des ARN. La seconde est que, bien que des protocoles aient été mis au point pour réaliser des puces d'expression sur ces blocs (Fedorowicz *et al.*, 2009 ; Wimmer *et al.*, 2018), cette technologie est peu adaptée à une activité clinique quotidienne.

3.1.2.1 – Article 1 : *RNA sequencing validation of the Complexity INdex in SARComas prognostic signature*

Lesluyes T, Pérot G, Largeau MR, Brulard C, Lagarde P, Dapremont V, Lucchesi C, Neuville A, Terrier P, Vince-Ranchère D, Mendez-Lago M, Gut M, Gut I, Coindre JM & Chibon F. Article publié dans *European Journal of Cancer* (2016).

Contexte

L'évolution des méthodes de séquençage haut débit a permis de progressivement intégrer ces technologies dans un contexte clinique. Le RNA-seq en particulier (Wang *et al.*, 2009), constitue une précieuse aide aussi bien diagnostique (par l'identification d'une altération spécifique) que pronostique (par l'utilisation d'un marqueur d'agressivité ou de réponse au traitement) ou encore thérapeutique (par l'identification d'une cible génétique ; Byron *et al.*, 2016 ; Serratì *et al.*, 2016). Nous nous sommes alors demandé si cette technologie pourrait correctement analyser les profils transcriptomiques à partir d'ARN extraits de blocs FFPE et donc potentiellement très dégradés. Ceci pourrait ainsi définir la technologie de référence pour la future application clinique routinière de la signature CINSARC.

Résumé de l'article

En vue d'une future application clinique de la signature CINSARC, nous devions trouver une technologie capable de satisfaire les contraintes techniques et matérielles imposées par le contexte clinique de la pathologie. Le matériel de référence étant le bloc FFPE, nous avions pour objectif de trouver une technologie permettant de reproduire le profil transcriptomique de CINSARC à partir d'ARN dégradés extraits de blocs FFPE. Nous avons ainsi challengé le RNA-seq sur ce terrain.

Afin de porter cette signature en situation clinique, nous avons défini deux niveaux de transfert. Le premier consiste à évaluer la valeur pronostique de CINSARC sur deux technologies différentes (les puces d'expression comme référence actuelle et le RNA-seq) à partir d'ARN intacts pour ne mesurer que la variation technologique. Le second niveau de transfert consiste à comparer, sur la même technologie (le RNA-seq en l'occurrence), les deux supports matériel que sont les ARN extraits de tissus congelés et de blocs FFPE pour ne mesurer que la variation due à la dégradation des ARN.

Pour la première partie, une cohorte d'apprentissage (217 cas en puce d'expression) nous a permis d'obtenir les pronostics d'une cohorte de validation (95 cas) sur les deux technologies (puces d'expression et RNA-seq). Ces dernières montrent des stratifications significatives en matière de survie sans métastase (P=2,77e-2 pour les puces d'expression et P=1,34e-2 pour

le RNA-seq) et un taux de valeur pronostique identique de 77% (73 cas sur 95). La validité du RNA-seq a également été confirmée sur une cohorte indépendante : celle du projet TCGA sur les sarcomes (246 cas ; *P*=4,14e-3). Afin d'optimiser le processus d'analyse, nous avons également démontré que nous pouvons réduire la profondeur de séquençage de 40 à 10 millions de *reads* appariés tout en conservant une valeur pronostique significative (*P*=3,2e-2).

Pour la seconde partie, nous avons séquencé 41 cas extraits de blocs FFPE dont nous avions l'équivalent en tissus congelés (dans la cohorte des 95 cas). Les profils de dégradation des ARN (déterminés au Bioanalyzer) corrèlent avec la qualité du séquençage, nous permettant de définir un seuil minimal à partir duquel nous obtenions un profil interprétable. Les 24 cas (59%) en dessous de ce seuil n'ont ainsi pas été interprétables à cause d'un séquençage massif d'ARN synthétiques correspondant au contrôle qualité ajouté dans la librairie (*External* RNA *Controls Consortium* Spike-In ; Jiang *et al.*, 2011). Parmi les 17 cas (41%) au-dessus de ce seuil, 15 cas (88%) partagent des pronostics similaires entre ARN extraits de tissus congelés et de blocs FFPE.

Ces résultats montrent que l'expression de la signature CINSARC peut être correctement évaluée à partir de séquençage haut débit et que le pronostic déduit est conforme à celui obtenu sur puce d'expression. Les différents profils de dégradation des ARN nous ont permis de définir un seuil afin d'anticiper une trop importante dégradation d'ARN provenant de blocs FFPE. Les cas au-dessus de ce seuil ont des pronostics similaires entre tissus congelés et blocs FFPE. Néanmoins, 59% des cas testés ont dû être exclus car présentant une dégradation d'ARN trop importante.

L'ensemble des résultats de cette étude sont inclus ci-après (Lesluyes *et al.*, 2016).

European Journal of Cancer 57 (2016) 104–111

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

RNA sequencing validation of the Complexity INdex in SARComas prognostic signature

Tom Lesluyes [a,b,c], Gaëlle Pérot [a,d], Marine Roxane Largeau [a,b], Céline Brulard [a], Pauline Lagarde [a], Valérie Dapremont [d], Carlo Lucchesi [a,c], Agnès Neuville [a,d], Philippe Terrier [e], Dominique Vince-Ranchère [f], Maria Mendez-Lago [g], Marta Gut [g], Ivo Gut [g], Jean-Michel Coindre [a,d,h], Frédéric Chibon [a,d,*]

[a] *Inserm U916, Institut Bergonié, Bordeaux, France*
[b] *Aquitaine Science Transfert, Centre Condorcet, Pessac, France*
[c] *Site de Recherche Intégrée sur le Cancer, Bordeaux Recherche Intégrée en Oncologie, Bordeaux, France*
[d] *Département de pathologie, Institut Bergonié, Bordeaux, France*
[e] *Département de pathologie, Institut Gustave Roussy, Villejuif, France*
[f] *Département de pathologie, Centre Léon Bérard, Lyon, France*
[g] *Centro Nacional de Análisis Genómico (CNAG), Barcelona, Spain*
[h] *Univ. Bordeaux, F-33000 Bordeaux, France*

Received 26 October 2015; received in revised form 29 December 2015; accepted 31 December 2015
Available online 23 February 2016

KEYWORDS
Sarcoma;
Transcriptome;
Cancer;
High-throughput RNA sequencing;
Metastasis

Abstract ***Background:*** Prognosis of metastatic outcome in soft tissue sarcomas is an important clinical challenge since these tumours can be very aggressive (up to 50% of recurring events). A gene expression signature, Complexity INdex in SARComas (CINSARC), has been identified as a better prognostic factor compared to the current international grading system defined by the *Fédération Nationale des Centres de Lutte Contre le Cancer*. Since CINSARC has been established on frozen tumours analysed by microarrays, we were interested in evaluating its prognostic capacity using next generation sequencing (NGS) on formalin-fixed, paraffin-embedded (FFPE) blocks to better fit laboratory practices.
Methods: Metastatic-free survivals (training/validation approach with independent datasets) and agreement values in classification groups were evaluated. Also, RNA degradation threshold has been established for FFPE blocks and differences in gene expression due to RNA degradation were measured.
Results: CINSARC remains a strong prognostic factor for metastatic outcome in both microarray and RNA-seq technologies ($P < 0.05$), with similar risk-group classifications (77%). We

* *Corresponding author:* Inserm U916, Institut Bergonié, 229 cours de l'Argonne, 33076 Bordeaux, France. Tel.: +33 556 330 443.
E-mail address: f.chibon@bordeaux.unicancer.fr (F. Chibon).

http://dx.doi.org/10.1016/j.ejca.2015.12.027

defined quality threshold to process degraded RNA extracted from FFPE blocks and measured similar classifications with frozen tumours (88%).
Conclusion: These results demonstrate that CINSARC is a platform and material independent prognostic signature for metastatic outcome in various sarcomas. This result opens access to metastatic prognostication in sarcomas through NGS analysis on both frozen and FFPE tumours via the CINSARC signature.

1. Introduction

Adult soft tissue sarcomas are rare (<1%) and aggressive tumours presenting a wide spectrum of clinical behaviour [1]. Since the 1980s, therapeutic management of patients has been widely determined by histological type and a grading system defined by the *Fédération Nationale des Centres de Lutte Contre le Cancer* (FNCLCC) [2]. However, the latter suffers from major limitations: 40% of cases are classified as intermediate grade, it is difficult to apply to microbiopsies, and its reproducibility may vary among pathologists [3].

To overcome these problems, our group identified a set of 67 genes involved in mitosis control and chromosome integrity pathways [4,5]. This transcriptomic signature, named Complexity INdex in SARComas (CINSARC), has been proven to be a valuable prognostic factor in sarcomas with complex genetics, gastrointestinal stromal tumours (GIST), synovial sarcomas and other tumours (breast carcinomas and lymphomas) [4–9]. In prospective, CINSARC could be routinely used by clinical centres to determine the risk of developing metastasis and administrate (neo)-adjuvant therapy to prevent such an event or reduce its growth rate [9].

In recent years, next generation sequencing (NGS) has been increasingly set up in laboratory benches and is often viewed as a competitor to microarray. Even if RNA-seq is still more expensive than chips, it allows getting both genes and transcripts expression, mutations and fusions for expressed genes [10]. The aim of the present study was to assess the prognostic value of CINSARC using RNA-seq in order to adapt new laboratory practices.

To compare the performance of RNA-seq and microarrays, we used expressions (from both technologies) of various sarcomas and compared predictive values based on CINSARC risk groups and clinical annotations from the Conticabase (European sarcoma database). We also performed RNA-seq on formalin-fixed, paraffin-embedded (FFPE) blocks and matching frozen samples to compare the level of agreement between these two material supports.

2. Materials and methods

2.1. Cohorts presentation

To transfer CINSARC from microarrays from frozen tissues to RNA-seq from FFPE blocks, we used three independent cohorts (Table 1). The first one (#1) is a training set of 217 sarcomas with expression computed by microarrays on frozen tissues (current CINSARC conditions). The second one (#2) is a validation set of 95 sarcomas analysed by both microarray and RNA-seq technologies on frozen tissues. These two cohorts are part of the CINSARC technological transfer. The last one (#3) is a set of 41 sarcomas with RNA-seq expression on both FFPE and frozen tissues. This cohort is used to evaluate if CINSARC remains prognostic using different material supports. Additional information about sample selection, RNA extraction, sequencing, analysis, and quality controls can be found in the Supplementary material and methods section.

2.2. Access to data

Microarray and RNA-seq expressions are available on Gene Expression Omnibus under accession GSE71121. RNA-seq raw files (FastQ) will be available on May 15, 2016 on sequence read archive under accessions: SRP057793 and SRP059588 (cohorts #2 and #3, respectively).

3. Results

3.1. From microarray to RNA-seq

We first aimed to test the prognostic value of CINSARC expression as evaluated by RNA sequencing extracted from frozen samples (Table 2; Supplementary Fig. 1).

We counted the number of expressed genes based on different thresholds to make sure that libraries reached the depth to cover almost the whole transcriptome (RNA-seq technology). Among 19,179 annotated coding genes [11], 12,671 (±362) genes were expressed at >1 FPKM (expected Fragments Per Kilobase of transcript per Million fragments sequenced) level [12], 15,097

Table 1
Characteristics of processed cohorts.

Cohort	#1 (n = 217)	#2 (n = 95)	#3 (n = 41)
Expression quantification	Microarray	Microarray and RNA-seq	RNA-seq
Material	Frozen tissue	Frozen tissue	FFPE and frozen tissue
Median follow-up (months) [95% CI]	52.44 [40.15–66.69]	50.92 [29.96–66.60]	19.12 [8.74–26.25]
Median age at diagnosis (years) [95% CI]	62 [60–64]	65 [63–67]	63 [58–66]
Histotypes (%)			
Undifferentiated sarcomas	56 (25.81)	33 (34.74)	9 (21.95)
Leiomyosarcomas	66 (30.41)	23 (24.21)	13 (31.71)
GIST	0	0	10 (24.39)
Dedifferentiated liposarcomas	31 (14.29)	13 (13.68)	5 (12.2)
Myxofibrosarcomas	31 (14.29)	12 (12.63)	3 (7.32)
Others	33 (15.21)	14 (14.74)	1 (2.44)
Genders (%)			
Males	109 (50.23)	54 (56.84)	15 (36.59)
Females	108 (49.77)	41 (43.16)	26 (63.41)
Metastasis (%)	80 (36.87)	44 (46.32)	6 (14.63)
Surgical margins (%)			
R0	100 (46.08%)	30 (31.58%)	25 (60.98%)
R1	43 (19.82%)	33 (34.74%)	13 (31.71%)
R2	10 (4.61%)	4 (4.21%)	0
Unknown	64 (29.49%)	28 (29.47%)	3 (7.32%)
Local recurrences (%)	49 (22.58)	32 (33.68)	8 (19.51)
FNCLCC grade (%)			
I and II	96 (44.24)	27 (28.42)	6 (14.63)
III	110 (50.69)	64 (67.37)	23 (56.1)
Unknown	11 (5.07)	4 (4.21)	2 (4.88)
NA (GIST)	0	0	10 (24.39)

FFPE, formalin-fixed, paraffin-embedded; CI, confidence interval; GIST, gastrointestinal stromal tumours; FNCLCC, *Fédération Nationale des Centres de Lutte Contre le Cancer*; NA, not applicable.
Others histotype for cohort #1 represents four adult fibrosarcomas, one osteosarcoma, nine pleomorphic liposarcomas, ten well differentiated liposarcomas, one low grade myofibroblastic sarcoma, one malignant mesenchymoma, two mesenchymal chondrosarcomas and five pleomorphic rhabdomyosarcomas. *Others* histotype for cohort #2 represents five pleomorphic liposarcomas, one well differentiated liposarcoma, one low grade fibromyxoid sarcoma, five pleomorphic rhabdomyosarcomas and two synovial sarcomas. *Others* histotype for cohort #3 represents one pleomorphic liposarcoma.

(±374) were expressed at >0.1 FPKM level and 16,480 (±398) were expressed at >0.01 FPKM level.

Considering both lowly-expressed genes and those above (>0.001 FPKM for RNA-seq and >3 log2 intensity for microarray), we observed moderate expression correlations between these two technologies (mean Pearson's r is 0.6232; Supplementary Fig. 2) but good correlations when only considering CINSARC genes (mean Pearson's r is 0.8402; Fig. 1). Since the CINSARC approach does not require value but profile similarities, we expected highly- and lowly-expressed CINSARC genes using microarray technology to be highly- and lowly-expressed, respectively using RNA-seq.

Table 2
Cohort metrics in terms of sequenced reads, read loss at each analysis step and Spike-in quality control.

RNA-seq series	Validation cohort #2 (n = 95)	Frozen cohort #3 (n = 41)	FFPE cohort #3 quality 1 and 2 (n = 24)	FFPE cohort #3 quality 3 and 4 (n = 17)
Sequenced reads (million)	79.14 (17.06)	75 (17.06)	73.11 (40.82)	78.69 (16.82)
% loss at pre-alignment step	4.19 (0.82)	4.69 (0.71)	7.92 (2.89)	7.65 (6.53)
% loss at alignment step	23.93 (4.16)	27.51 (4.76)	20.74 (9.43)	17.5 (7.82)
% non-standard alignments	9.52 (0.9)	9.16 (0.79)	12.72 (1.71)	12.98 (4.06)
% PCR duplicates	11.19 (2.77)	9.21 (2.07)	52.09 (10.53)	38.59 (14.69)
Total loss (%)	48.83 (2.57)	50.57 (3.26)	93.46 (6.37)	76.72 (17.25)
Spike-in sequenced in library (%)	0.44 (0.18)	0.46 (0.18)	85.73 (21.41)	38.14 (29.7)
Correlation expression/quantity	0.9601 (0.0066)	0.958 (0.0082)	0.9831 (0.0023)	0.9834 (0.0039)
Detected Spike-ins per sample	69.03 (4.29)	67.71 (4.35)	89.04 (3.2)	85.53 (10.86)

FFPE, formalin-fixed, paraffin-embedded; PCR, polymerase chain reaction.
Numbers in brackets are standard deviations. Pre-alignment step discards short reads (<30 bp) after low-quality trim (Phred score <20) and overlap correction. Alignment step discards reads that do not map to reference Hg19 transcriptome. Non-standard alignments include multiple location mapping, read mate not mapped and discordant strand orientation. Spike-ins results for frozen cohort #3 are based on the 21 samples that include Spike-ins.

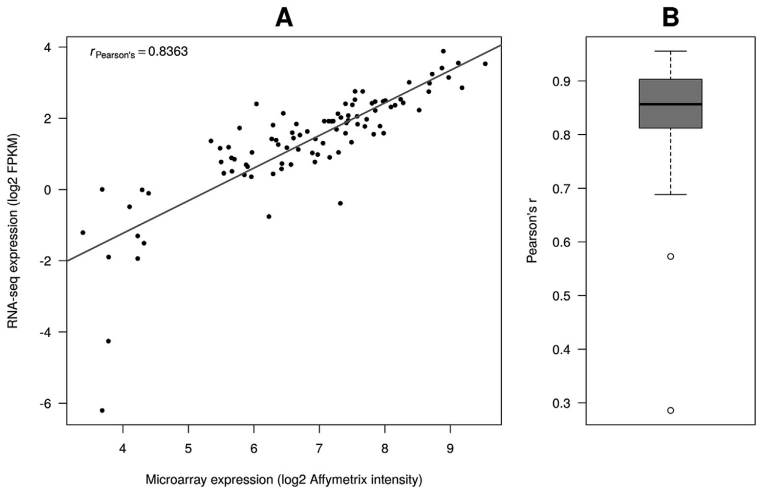


Fig. 1. Raw comparison of CINSARC gene expression between microarray and RNA-seq technologies (performed on validation cohort #2). (A) Example of a typical CINSARC gene analysed by both technologies (points are samples). Mean Pearson's r for all samples is 0.8402 ± 0.0998. (B) Boxplot of correlations obtained in the whole validation cohort (#2). Results indicate good agreements when comparing CINSARC expressions across these two technologies. CINSARC, Complexity INdex in SARComas.

Thus, we performed metastasis-free survival analysis using centroids computed in microarray training cohort #1 and applied on validation cohort #2 with both microarray (Fig. 2A; Table 3) and RNA-seq expressions (Fig. 2B; Table 3) and observed a significant prognostic value with both approaches ($P = 2.8 \times 10^{-2}$ and $P = 1.3 \times 10^{-2}$, respectively). Additionally, we performed the reverse analysis: we computed RNA-seq centroids on validation cohort #2 and applied them to the microarray expression of training cohort #1 (Fig. 2C; Table 3). Again, despite technological differences, the prognostic value was significant ($P = 9.7 \times 10^{-5}$). Agreement test between prognostic groups by both RNA-seq and microarray resulted in a similar CINSARC classification for 73 out of 95 transcriptomic profiles (77% agreement).

To confirm our experiments, we decided to test our prognostic signature on another independent sarcoma dataset from The Cancer Genome Atlas (TCGA) and kept 246 samples with both clinical annotations and RNA-seq expression. We applied RNA-seq centroids on validation cohort #2 to the available dataset of 246 sarcomas and measured significant prognostic value ($P = 4.1 \times 10^{-3}$; Fig. 2D; Table 3).

Also, since RNA sequencing technology gives the opportunity to estimate transcript abundance, we measured CINSARC prognostic at the transcript level. For each CINSARC gene, we selected the transcript with the better measurement of metastatic outcome and computed transcripts centroids in validation cohort (#2). We applied these centroids on TCGA expression data at the transcript level and observed a significant prognostic value ($P = 7.9 \times 10^{-4}$), but a higher number of tumours classified in the high-risk group (79%; Supplementary Fig. 3).

Finally, we were interested in validating the CINSARC signature in dependence of sequencing depth. We reduced, *in silico*, read numbers (20, 10 and 2 million reads representing 10, 5 and 1 million paired-ends, respectively) for RNA-seq cohort #2, processed the same way than normal sequencing depth and compared CINSARC classifications for the 95 tumours. Correlations between original sequencing depth and reduced ones tend to slightly decrease as read number falls (mean Pearson's r are 0.9892 ± 0.0014, 0.9817 ± 0.0016 and 0.9492 ± 0.0036 for 20, 10 and 2 million reads, respectively). Additionally, non-expressed (below 0.001 FPKM) genes in reduced depth (non-expressed genes below this threshold in both reduced and normal depth are not considered) tend to increase as read number falls (1015 ± 207, 1580 ± 238 and 3083 ± 286 for 20, 10 and 2 million, respectively). CINSARC classification remains identical for 90 (95%), 89 (94%) and 76 (80%) tumours for 20, 10 and 2 million reads respectively with significant prognostic values for 20 and 10 million reads ($P = 1.6 \times 10^{-2}$ and $P = 3.2 \times 10^{-2}$, respectively) but not significant for 2 million reads ($P = 5.7 \times 10^{-2}$).

3.2. From frozen tissues to FFPE blocks

We selected a set of 41 FFPE blocks aged up to 4 years (cohort #3) to estimate agreement prediction between

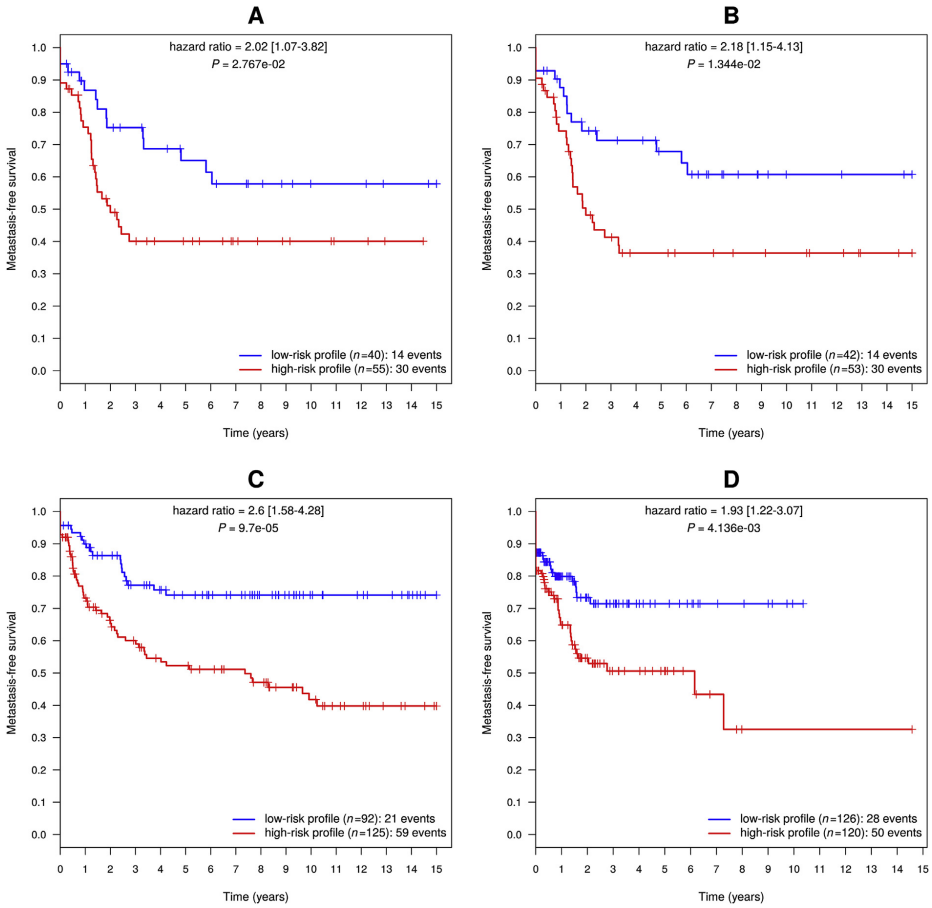


Fig. 2. Kaplan–Meier survival analysis according to CINSARC expression. (A) Prognostic value using training cohort (#1) centroids on validation cohort (#2) with microarray expression. (B) Prognostic value using training cohort (#1) centroids on validation cohort (#2) with RNA-seq expression. (C) Prognostic value using validation cohort (#2) RNA-seq centroids on training cohort (#1). (D) Prognostic value using validation cohort (#2) RNA-seq centroids on TCGA cohort with RNA-seq expression. Each combination is significant so prognostic value of CINSARC remains good with a technological transfer. CINSARC, Complexity INdex in SARComas.

frozen tissues (current CINSARC reference) and FFPE blocks (challenger). Using cut-offs, we assigned a quality index ranging from 1 (worst quality) to 4 (best quality) to each of the 41 RNA extractions (Supplementary material; Supplementary Fig. 4). Twelve samples were assigned to each quality 1, 2 and 3, and five samples were assigned to quality 4.

Though sequencing depth is similar between cohorts, we measured a massive Spike-in sequencing [13] and a high percentage of total read loss in FFPE tumours compared to frozen tumours (Table 2; Supplementary Figs. 1and 5), consistent with RNA degradation. We also measured expressions of two housekeeping genes, *ACTB* and *RPLP0*, and observed a significant expression loss in FFPE tumours compared to frozen tumours, result also consistent with degraded RNA (Supplementary Fig. 6).

We computed correlations between expression from frozen tissues and from FFPE blocks. Interestingly, 22 samples showed very low to moderate correlations (Pearson's r from 0.2 to 0.6) and 19 samples showed moderate to strong correlations (Pearson's r from 0.6 to 0.9) with a significant correlation difference if RNA quality is considered (Wilcoxon's $P = 4.06 \times 10^{-4}$, higher correlations in groups 3 and 4). No dependency was found between quality and block age, extraction

Table 3
Survival analysis metrics corresponding to Fig. 2 Kaplan–Meier curves.

Condition		(A) microarray centroids, microarray expression						(B) microarray centroids, RNA-seq expression					
Time (years)		0	2	4	6	8	10	0	2	4	6	8	10
C1	N risk	40	26	21	17	12	8	42	27	23	18	10	5
	Events	2	9	11	13	14	14	3	10	11	13	14	14
	MFS	0.95	0.75	0.69	0.61	0.58	0.58	0.93	0.74	0.71	0.64	0.61	0.61
C2	N risk	55	23	15	12	7	5	53	22	13	11	9	8
	Events	6	26	30	30	30	30	5	25	30	30	30	30
	MFS	0.89	0.49	0.40	0.40	0.40	0.40	0.91	0.48	0.36	0.36	0.36	0.36
Condition		(C) RNA-seq centroids, microarray expression						(D) RNA-seq centroids, RNA-seq expression					
Time (years)		0	2	4	6	8	10	0	2	4	6	8	10
C1	N risk	92	68	50	40	28	16	126	40	20	12	7	1
	Events	4	12	20	21	21	21	15	27	28	28	28	28
	MFS	0.96	0.86	0.76	0.74	0.74	0.74	0.88	0.73	0.71	0.71	0.71	0.71
C2	N risk	125	64	48	42	34	22	120	32	17	7	1	1
	Events	9	38	49	52	55	58	22	46	48	48	50	50
	MFS	0.93	0.66	0.55	0.51	0.47	0.42	0.82	0.55	0.51	0.51	0.33	0.33

C1 (Classification 1) is the low-risk profile group, *C2* (Classification 2) is the high-risk profile group, *N risk* is the number of patients in groups, *Events* is the cumulative number of metastasis in groups and *MFS* refers to metastasis-free survival.

date or any known clinical parameter (e.g. histotype, size of tumour, metastatic event, local recurrence event, FNCLCC grade, necrosis, surgical margin, site of tumour, radiotherapy and chemotherapy) suggesting that RNA degradation in FFPE block is not associated to any of the tested parameters and that 4-year blocks are viable for CINSARC processing.

Since results we obtained with low-quality RNA from FFPE blocks (groups 1 and 2) demonstrated a high total read loss, a massive Spike-in sequencing, poor correlations with RNA from frozen tumours and low housekeeping gene expression, we consequently pursued experiments on the 17 better quality RNA (groups 3 and 4).

Given that FFPE blocks were not older than 4 years and that the FFPE good quality RNA cohort #3 comprised 17 cases, it was difficult to process survival analysis. Instead, we computed CINSARC groups using RNA-seq centroids on validation cohort #2 applied to the 17 high-quality samples from cohort #3 (Fig. 3). We measured a similar prediction between the two material supports (88% agreement, 90% sensibility and 86% sensitivity).

4. Discussion

The development of a reliable and reproducible, investigator-independent method to predict sarcoma aggressiveness may be essential to optimise the clinical management of patients, especially in the context of sarcomas with complex genomics for which surgery is currently the most effective treatment. We report here the possibility of using a transposable gene expression signature, CINSARC, from microarrays to NGS and from frozen tumours to FFPE blocks, with a proven usefulness on metastatic prognosis in several cancer types and particularly in sarcomas. RNA sequencing, a reproducible technique across laboratories [14], can be used as a new tool by clinical centres to prospectively test this signature and identify other alterations (i.e. other candidate mutant/translocation/signature) to manage patients with a personalised medicine approach.

In the first part of the CINSARC transfer, we measured significant prognosis values with two expression quantification technologies ($P = 2.7 \times 10^{-2}$, $P = 1.3 \times 10^{-2}$ and $P = 9.7 \times 10^{-5}$ for microarray centroids/microarray expression, microarray centroids/RNA-seq expression and RNA-seq centroids/microarray expression, respectively) and similar risk-group classifications (77%), consistent with observed

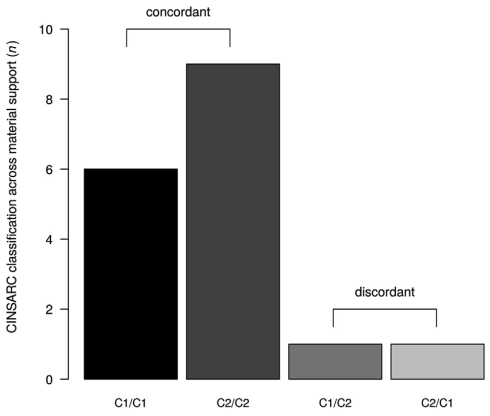


Fig. 3. Material support classification based on CINSARC profiles. Supports are frozen tissue and FFPE blocks before and after slash character, respectively. This shows 88% agreement, 90% sensibility and 86% specificity. CINSARC, Complexity INdex in SARComas; FFPE, formalin-fixed, paraffin-embedded.

moderate to high correlations between these two methods [15–19]

From a raw comparison, CINSARC expression correlated well using both quantification technologies (mean Pearson's r is 0.8402; Fig. 1). However, correlation was poor for two genes (*HP1BP3* and *SGOL2*, Pearson's r are 0.2859 and 0.5729, respectively). Additional investigations demonstrated that microarray probe selection (maximum interquartile range for each CINSARC gene) did not always reflect RNA-seq expression: switching probes for these two genes significantly improved correlation (0.7142 and 0.8181 for *HP1BP3* and *SGOL2*, respectively). Thus, RNA sequencing allows overcoming probe selection which may not reflect global gene expression as probes could be transcript specific.

TCGA enabled us to validate our expression signature on a large independent dataset of annotated sarcomas, with a significant measurement of CINSARC prognostic at gene ($P = 4.1 \times 10^{-3}$) and transcript ($P = 7.9 \times 10^{-4}$) levels. We hypothesise that we would have an even better prediction of metastatic outcome with similar bioinformatics pipeline to produce results that could be perfectly comparable, without technical artifacts due to differences in expression computation (i.e. reference transcriptome, alignment and quantification software, mapping stringency, etc.).

The transcript selection we used (one per gene) may not be the best possible combination for metastatic prognosis. However, considering that the number of combinations is extremely high (8.77×10^{17} possible transcript permutations), we did not have enough statistical power to evaluate all of them. In terms of risk classification, tumours were equally balanced between low- and high-risk groups at the gene level (51% and 49%, respectively) but not at the transcript level (21% and 79%, respectively), suggesting that the transcript selection either does not reflect CINSARC expression (although significant) or is more sensible but less specific compared to gene expression. Also, since new splices are commonly discovered over time, transcripts signature should be computed repeatedly.

We demonstrated that it is possible to obtain similar CINSARC classification when reducing, *in silico*, sequenced reads (from 79 to 10 million) with high correlations (mean Pearson's r is 0.9817), low number of non-expressed (below 0.001 FPKM) genes (1580) and significant prognostic values ($P = 3.2 \times 10^{-2}$). However, a too much read reduction (from 79 to 2 million) produces a high number of non-expressed (below 0.001 FPKM) genes (3083) and non-significant prognostic value. Though a necessary sequencing depth of 10 million reads (5 million paired-ends) is thus mandatory for a correct CINSARC prognostication, targeted sequencing focussing on the 67 CINSARC genes (and some additional control genes) would require a lower coverage. However, several million reads produced by sequencing centres is enough to have a reliable gene/transcript expression for CINSARC prognostication (typically >20 million reads are sequenced) [14].

In the second part of CINSARC transfer, we wanted to test whether the technological transfer could be applied on FFPE samples. Since we had too few and too recent cases in cohort #3 to process survival analysis, we could only use the CINSARC classification. As we previously demonstrated that CINSARC is a prognostic factor in GIST [6], we could include ten of them in cohort #3. For this purpose, we identified RNA quality thresholds based on Bioanalyzer profiles that define whether RNA, extracted from FFPE blocks, is viable for transcriptomic analysis or not. In accordance to these profiles, we showed that RNA sequencing on quality groups 1 and 2 produced massive duplicated reads due to RNA degradation responsible for the polymerase chain reaction duplication of a small RNA fraction less degraded. Using a minimum quality threshold of 3, RNA sequencing produced a reliable CINSARC expression with similar risk group classifications across material support on 17 tumours (88% agreement). Since we observed a very important read loss (>75%; Table 2) and massive Spike-in sequencing (>30%) in RNA-seq from FFPE tumours, library size should not be reduced under 50 million reads (25 million paired-ends).

Tumour heterogeneity can play a role in a different CINSARC classification even if we realised internal controls to choose similar tumour areas (>60% of tumour cells, no or low necrosis) between frozen/FFPE tumours and microarray/RNA-seq technologies (Supplementary material and methods). However, since we observed similar risk-group classifications (77% and 88% for technological and material support transfers, respectively) this would suggest that similar tumour area produces similar CINSARC profile and/or that the CINSARC signature is expressed by the tumour bulk despite intra-tumoural heterogeneity, making CINSARC expression an early event in tumour development.

It has been described that storage conditions (e.g. effects of sunlight, temperature, fixation time, etc.) play a key role in RNA degradation in FFPE blocks [20]. We did not have measurements of these parameters to evaluate what specific condition(s) are responsible for this phenomenon. A challenge, to routinely use CINSARC, will be to identify these factors and produce/store FFPE blocks that least degrades RNA.

To our knowledge, this is the first report on a simultaneous transfer of a gene expression signature in terms of expression quantification technology and material support. To go further into the clinical applicability of the CINSARC signature, we foresee three major steps: to improve the technical aspect of FFPE sequencing by testing new RNA extraction kits adapted to FFPE samples, to set up a CINSARC targeted sequencing to manage multiple patients per sequencing

run, and to prospectively validate CINSARC in the clinical context (two Europe-wide clinical trials are currently ongoing).

Grant support

This work was supported by Aquitaine Science Transfert, euroSARC (FP7-HEALTH-2011), SIRIC BRIO and Inserm.

Conflict of interest statement

None declared.

Acknowledgements

Part of the Bioinformatics computation was performed at the Bordeaux Bioinformatics Center (CBiB), Université de Bordeaux, and computing facilities MCIA (Mésocentre de Calcul Intensif Aquitain), Université de Bordeaux and of the Université de Pau et des Pays de l'Adour. The authors are grateful to the people involved in the French Sarcoma Group (GSF-GETO) for clinical annotations and access to their samples: L. Guillou, F. Collin, A. Leroux, Y-M. Robin, M.C. Chateau, I. Peyrottes, F. Mishellany and J-B. Courrège. The authors thank J. Iriondo of Institut Bergonié for providing medical editorial assistance and English correction.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ejca.2015.12.027.

References

[1] Fletcher CDM, Unni KK, Mertens F. Pathology and genetics of tumours of soft tissue and bone. Lyon: IARC Press; 2002.
[2] Trojani M, Contesso G, Coindre JM, Rouesse J, Bui NB, de Mascarel A, et al. Soft-tissue sarcomas of adults; study of pathological prognostic variables and definition of a histopathological grading system. Int J Cancer 1984;33:37–42.
[3] Coindre JM, Terrier P, Guillou L, Le Doussal V, Collin F, Ranchère D, et al. Predictive value of grade for metastasis development in the main histologic types of adult soft tissue sarcomas: a study of 1240 patients from the French Federation of Cancer Centers Sarcoma Group. Cancer 2001;91:1914–26.
[4] 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. http://dx.doi.org/10.1038/nm.2174.
[5] Chibon F, Coindre JM, Aurias A. Prognostic molecular signature of sarcomas, and uses thereof. International patent CA 2759417, 2010.
[6] Lagarde P, Pérot 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. Clin Cancer Res 2012;18:826–38. http://dx.doi.org/10.1158/1078-0432.CCR-11-1610.
[7] 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. J Clin Oncol 2013;31:608–15. http://dx.doi.org/10.1200/JCO.2012.46.0147.
[8] Chakiba C, Lagarde P, Pissaloux D, Neuville A, Brulard C, Pérot G, et al. Response to chemotherapy is not related to chromosome instability in synovial sarcoma. Ann Oncol 2014;25:2267–71. http://dx.doi.org/10.1093/annonc/mdu362.
[9] Lartigue L, Neuville A, Lagarde P, Brulard C, Rutkowski P, Dei Tos P, et al. Genomic index predicts clinical outcome of intermediate-risk gastrointestinal stromal tumours, providing a new inclusion criterion for imatinib adjuvant therapy. Eur J Cancer 2015;51:75–83. http://dx.doi.org/10.1016/j.ejca.2014.10.014.
[10] Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57–63. http://dx.doi.org/10.1038/nrg2484.
[11] 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. http://dx.doi.org/10.1093/nar/gkr1079.
[12] Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 2010;28:511–5. http://dx.doi.org/10.1038/nbt.1621.
[13] Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, et al. Synthetic spike-in standards for RNA-seq experiments. Genome Res 2011;21:1543–51. http://dx.doi.org/10.1101/gr.121095.111.
[14] 't Hoen PAC, Friedländer MR, Almlöf J, Sammeth M, Pulyakhina I, Anvar SY, et al. Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. Nat Biotechnol 2013;31:1015–22. http://dx.doi.org/10.1038/nbt.2702.
[15] Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 2008;18:1509–17. http://dx.doi.org/10.1101/gr.079558.108.
[16] Leimena MM, Wels M, Bongers RS, Smid EJ, Zoetendal EG, Kleerebezem M. Comparative analysis of *Lactobacillus plantarum* WCFS1 transcriptomes by using DNA microarray and next-generation sequencing technologies. Appl Environ Microbiol 2012;78:4141–8. http://dx.doi.org/10.1128/AEM.00470-12.
[17] Kogenaru S, Qing Y, Guo Y, Wang N. RNA-seq and microarray complement each other in transcriptome profiling. BMC Genomics 2012;13:629. http://dx.doi.org/10.1186/1471-2164-13-629.
[18] Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One 2014;9:e78644. http://dx.doi.org/10.1371/journal.pone.0078644.
[19] Chavan SS, Bauer MA, Peterson EA, Heuck CJ, Johann DJ. Towards the integration, annotation and association of historical microarray experiments with RNA-seq. BMC Bioinformatics 2013;14(Suppl. 14):S4. http://dx.doi.org/10.1186/1471-2105-14-S14-S4.
[20] von Ahlfen S, Missel A, Bendrat K, Schlumpberger M. Determinants of RNA quality from FFPE samples. PLoS One 2007;2:e1261. http://dx.doi.org/10.1371/journal.pone.0001261.

RNA sequencing validation of the CINSARC prognostic signature

Tom Lesluyes, Gaëlle Pérot, Marine Roxane Largeau, Céline Brulard, Pauline Lagarde, Valérie Dapremont, Carlo Lucchesi, Agnès Neuville, Philippe Terrier, Dominique Vince-Ranchère, Maria Mendez-Lago, Marta Gut, Ivo Gut, Jean-Michel Coindre & Frédéric Chibon.

Supplementary material and methods

1. *Ethics statement*

The samples used in this study are part of the Biological Resources Center of Bergonié Cancer Institute (CRB-IB). In accordance with the French Public Health Code (articles L. 1243-4 and R. 1243-61), the CRB-IB has received the agreement from the French authorities to delivered samples for scientific research (number AC-2008-812, on February 2011). These samples are from care and re-qualified for research. The project was approved by the Bergonié ethic committee (scientific advisory board).

2. *Tumor samples*

Tumors were operated in expert centers of the French Sarcoma Group and come from surgical fragments. Every case was histologically reviewed by the pathologist subgroup from the French Sarcoma Group and classified according to the 2013 World Health Organization classification by histology, immunohistochemistry, molecular genetics and cytogenetics when needed.

Before RNA extraction, tumor cellularity and lack of necrosis areas of each frozen sample were verified by a pathologist (J-M. C.) using a mirror block or a frozen section stained with hematoxylin, eosin and safran (HES). Only frozen samples with at least 60% of tumor cells and no or few necrosis (maximum of 5%) were used. In the same manner: FFPE blocks were first verified by a pathologist (J-M. C.) on the corresponding HES slide and only areas with a minimum of 60% of tumor cells and no or few necrosis (maximum of 5% in two cases) were selected to perform RNA extraction on FFPE samples.

3. *RNA extraction*

RNA extraction from frozen samples (around 50 mg of tumor) was performed using a standard TRIzol (Life Technologies) / chloroform extraction followed by 70% ethanol precipitation and RNA purification using the RNeasy Mini Kit (Qiagen) with a DNase treatment (RNase-Free DNase Set, Qiagen).

RNA extraction from FFPE tissue (six to ten carrots obtained using a punch of 1.5 mm diameter, Laboderm) was performed using a Deparaffinization Solution (Qiagen) as recommended by the manufacturer followed by a DNase treatment (RNase-Free DNase Set, Qiagen) and RNA purification using the RNeasy FFPE kit (Qiagen) according to the manufacturer's recommendations.

RNA were quantified using a Nanodrop 1000 spectrophotometer (Thermo Scientific) and were qualified with the Agilent 2100 Bioanalyzer (Agilent) using the Agilent RNA 6000 Nano Kit (Agilent) according to the manufacturer's instructions.

An ERCC RNA Spike-In Mix (Life technologies) was added to each RNA as recommended by the manufacturer [13].

RNA extractions used in RNA-seq and microarrays experiments come from different extractions of the same tumor fragments.

4. *Sample preparation and RNA sequencing*

The RNA-seq libraries from FFPE tissue samples total RNA were prepared using a modified TruSeq RNA Sample Prep Kit v2 protocol. Briefly, rRNA was depleted from 0.5 µg of total RNA using the RiboZero Magnetic Gold Kit (Human/Mouse/Rat, Epicentre). Ribosomal RNA-depleted RNA samples were purified using Agencourt RNA Clean XP beads (Beckman Coulter Genomics) and RNA was eluted with the Elute, Prime, Fragment Mix from the TruSeq RNA Sample Prep Kit v2 (Illumina Inc.). The RNA fragmentation time (2.5 to 5 minutes) varied depending on the quality of the initial total RNA (assessed by Eukaryote Total RNA Nano Bioanalyzer assay, Agilent). Following the fragmentation, first and second strand synthesis, the Illumina barcoded adapters were ligated at 1/10 dilution of the recommended concentration. Libraries were enriched with 15 cycles of PCR. The size and quality of the libraries were assessed in a High Sensitivity DNA Bioanalyzer assay (Agilent).

Number of PCR cycles is a parameter that cannot compensate for low input material (less than the required quantity of total RNA for standard sequencing). If material quantity is not sufficient for standard RNA-seq protocol, then another specific RNA sequencing protocol must be employed.

Starting input material for the libraries construction was DNA free total RNA from frozen tissue using the TruSeq™ RNA Sample Prep Kit v2 (Illumina Inc.,) according to manufacturer's protocol. In short, 0.5 µg of total RNA was used for poly-A based mRNA enrichment selection followed by fragmentation by divalent cations resulting into fragments of 80-250nt, with the major peak at 130nt. Double stranded cDNA was end repaired, 3´adenylated and the 3´-"T" nucleotide of barcoded Illumina adapters was used for the adapter ligation. The ligation product was amplified with 12 cycles of PCR and quality controlled in Agilent DNA 7500 Bioanalyzer assay (Agilent).

Each library was sequenced using TruSeq SBS Kit v3-HS, in paired end mode with the read length 2x76bp. We generated minimally 23 million paired end reads passing filter for each FFPE RNA-seq library or at least 35 million paired end reads passing filter for each frozen tissue RNA-seq library in a fraction of a sequencing lane on HiSeq2000 (Illumina, Inc) following the manufacturer's protocol. 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 by CASAVA (v1.8).

5. *Bioinformatics analysis pipeline for RNA-seq*

From a bioinformatics standpoint, we treated RNA from frozen tissues and FFPE blocks the same way. We estimated general read quality using FastQC (v0.10.1), trimmed low-quality 5' and 3' ends (Phred score <20) using Sickle (v1.2) and used a home-made routine to treat overlapping reads because sometimes, insert size of paired-end reads is too small so reads may overlap each other. In such cases, mutations and gene fusions located in these overlap may be counted twice, generating a bias in detection. To overcome this problem, we merged overlapping reads (at least 10 bp) using SeqPrep (v1.1) and split merged reads in half with a Awk routine. Finally, paired-ends with one read shorter than 30 bp were discarded.

Transcriptomic (coding RefSeq [11] genes from Hg19 UCSC Table Browser [21] fixed on 2014/01) alignments were performed by TopHat2 (v2.0.1) [22]. Once aligned, we removed paired-end reads that mapped at multiple locations on transcriptome, low-quality (score <20) and discordant paired-end reads using SAMtools (v0.1.19) [23]. MarkDuplicates (PicardTools v1.99) removed PCR duplicates and Cufflinks (v2.1.1) [12] estimated gene and transcript abundances. HTSeq (v0.5.4) [24] numbered reads mapped to Spike-in and reference chromosomes. Gene fusions were reported by deFuse (v0.6.1) [25]. Miscellaneous filters, statistics and plots were performed by R (v.3.1.1).

6. *Microarray expression*

We performed microarray expression using Affymetrix Human Genome U133 Plus 2.0 on 312 samples. We normalized chips with GCRMA algorithm from *justGCRMA* routine implemented in gcrma (v2.34) R package. To select probes that better describe CINSARC expression in our cohorts, we computed interquartile range (IQR) of each probe associated to CINSARC genes (release 33 of HG-U133 Plus 2.0 annotations) and kept, for each gene, the probe that maximizes IQR value.

7. *The Cancer Genome Atlas data analysis*

We obtained clinical annotations and gene expressions from The Cancer Genome Atlas (TCGA at http://cancergenome.nih.gov/) data portal for the sarcoma project. Based on metastatic and follow-up information, we built annotation file to process metastasis-free survival analysis on 246 samples. We defined values (metastatic events and dates) as followed. If new tumor event type was set to distant metastasis in follow-up, metastatic event was set to positive and event date was set to the earliest metastatic event. If metastatic disease annotation was confirmed without follow-up information, metastatic event was also set to positive and event date was set to 0 (*e.g.* corresponds to tumor diagnosis date). If metastatic disease was negative, metastatic event was set to negative and event date was set to the last contact day. If metastatic disease annotation was not available and there was no follow-up information, metastatic event was set to negative and event date was set to the last contact day.

8. FFPE blocks quality analysis

Based on Agilent 2100 Bioanalyzer profiles, samples were categorized according to the following criteria: 1) length of the longest RNA fragments, 2) amount of sample (fluorescence intensity) and 3) slope/profile of samples (broad/narrow peak). This lead to the establishment of four quality groups ranging from 1 (worst quality) to 4 (best quality). Low quality RNA refers to groups 1 and 2, high quality RNA refers to groups 3 and 4.

9. ERCC Spike-in quality control

To assess whether RNA-seq expression is consistent with RNA quantity, sequenced libraries included the External RNA Controls Consortium (ERCC) Spike-in [13]. According to manufacturer recommendations and observations of sequenced Spike-ins in different studies [26], three quality criteria have to be fulfilled: 1) the correlation between Spike-in expression and concentration has to be high (Pearson's r >0.95); 2) at least 60 Spike-ins have to be expressed (the most concentrated ones); and 3) Spike-in reads should represent about 1% of total reads.

RNA-seq validation cohort #2 (95 samples) satisfied these requirements with expression correlated to RNA quantity (mean Pearson's r is 0.9601), a good sequencing depth (number of detected Spike-in mean is 69.03) and a good Spike-in proportion (mean is 0.44% of total reads; **Table 2**; **Supplementary Figure 1**).

RNA-seq on frozen cohort #3 (41 samples) also satisfied these requirements with expression correlated to RNA quantity (mean Pearson's r is 0.958), a good sequencing depth (number of detected Spike-in mean is 67.71) and a good Spike-in proportion (mean is 0.46% of total reads).

RNA-seq on FFPE cohort #3 (21 samples among 41 included Spike-ins) showed good correlations between expression and RNA quantity (means Pearson's r are 0.9831 and 0.9834 for quality 1 or 2 and 3 or 4 respectively), high number of detected Spike-ins (means are 89.04 and 85.53 for quality 1/2 and 3/4, respectively) and high percentage of

Spike-ins reads in libraries (means are 85.73% and 38.14% for quality 1/2 and 3/4, respectively). These results are consistent with RNA degradation measured in FFPE blocks: Spike-in RNA, not degraded since added in library preparation, are more likely to be sequenced compared to tumor RNA.

References

[21] 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. doi:10.1093/nar/gkh103.

[22] Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013;**14**:R36. doi:10.1186/gb-2013-14-4-r36.

[23] 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. doi:10.1093/bioinformatics/btp352.

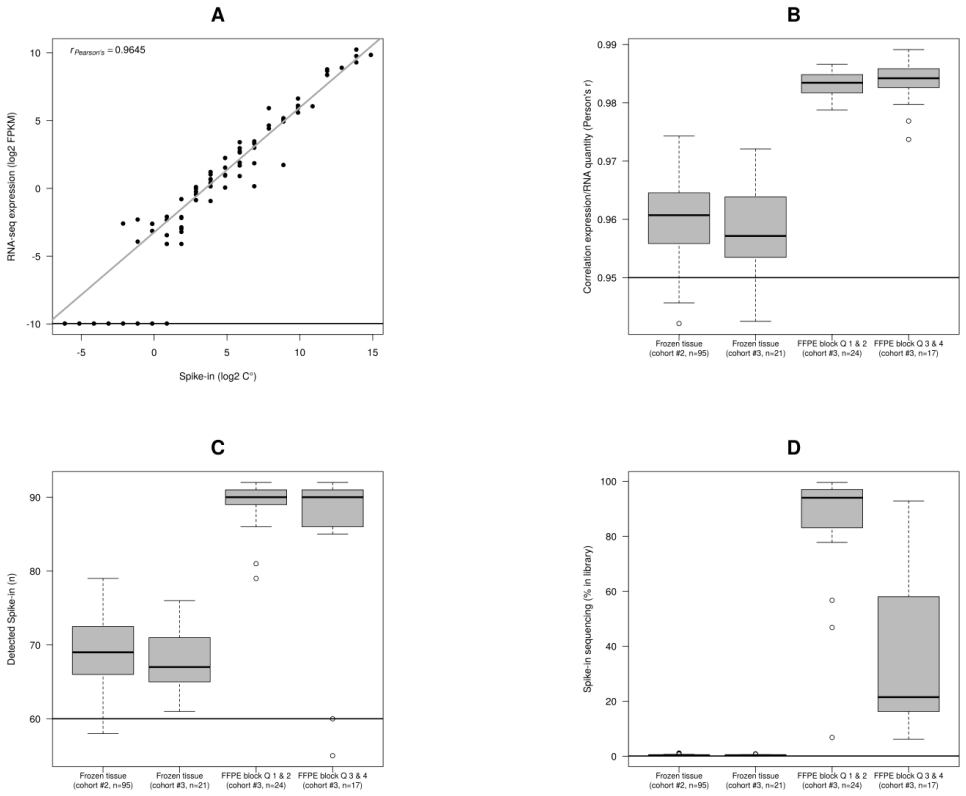
[24] Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;**31**:166-9. doi:10.1093/bioinformatics/btu638.

[25] 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. doi:10.1371/journal.pcbi.1001138.

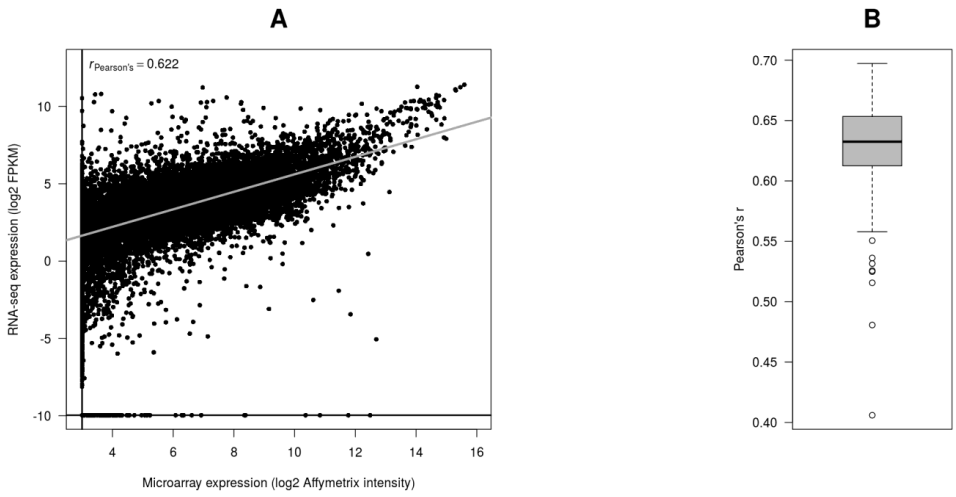
[26] Qing T, Yu Y, Du T, Shi L. mRNA enrichment protocols determine the quantification characteristics of external RNA spike-in controls in RNA-Seq studies. *Sci China Life Sci* 2013;**56**:134-42. doi:10.1007/s11427-013-4437-9.

RNA sequencing validation of the CINSARC prognostic signature

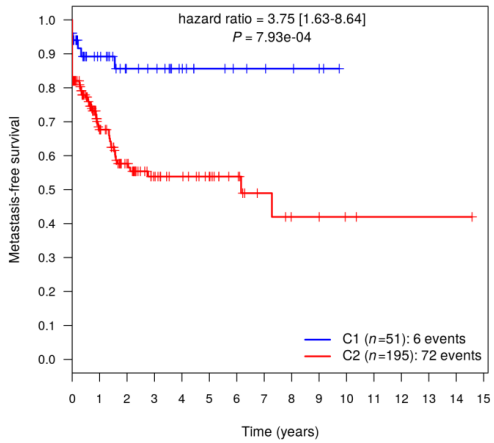
Tom Lesluyes, Gaëlle Pérot, Marine Roxane Largeau, Céline Brulard, Pauline Lagarde, Valérie Dapremont, Carlo Lucchesi, Agnès Neuville, Philippe Terrier, Dominique Vince-Ranchère, Maria Mendez-Lago, Marta Gut, Ivo Gut, Jean-Michel Coindre & Frédéric Chibon.



Supplementary Fig. 1 ERCC Spike-in quality control evaluation. **(A)** Example of correlation between RNA-seq expression and RNA quantity for a given sample. Mean Pearson's *r* for all samples is 0.9601 ± 0.0066. **(B)** Correlations obtained in the different cohorts. **(C)** Numbers of detected Spike-ins in the different cohorts. **(D)** Sequencing in library in the different cohorts. Horizontal black line in **A** represents our expression threshold. Horizontal black lines in **B**, **C** and **D** represent Spike-in metric thresholds recommended by manufacturer. Legend: *FPKM* is gene expression unit as: expected Fragments Per Kilobase of transcript per Million fragments sequenced, *C°* is RNA quantity in Mix given by manufacturer and *Q* refers to RNA quality.

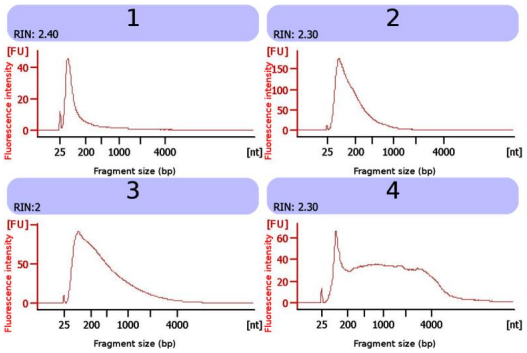


Supplementary Fig. 2 Raw comparison of gene expression between microarray and RNA-seq technologies (performed on validation cohort #2). **(A)** Example of a typical sample where only expressed genes in both technologies are considered (FPKM >0.001 and log2 intensity >3, thresholds represented as black lines, points are genes). Mean Pearson's *r* for all samples is 0.6232 ± 0.0463. **(B)** Boxplot of correlations obtained in the whole validation cohort (#2). Results indicate moderate agreements when comparing gene expressions across these two technologies.

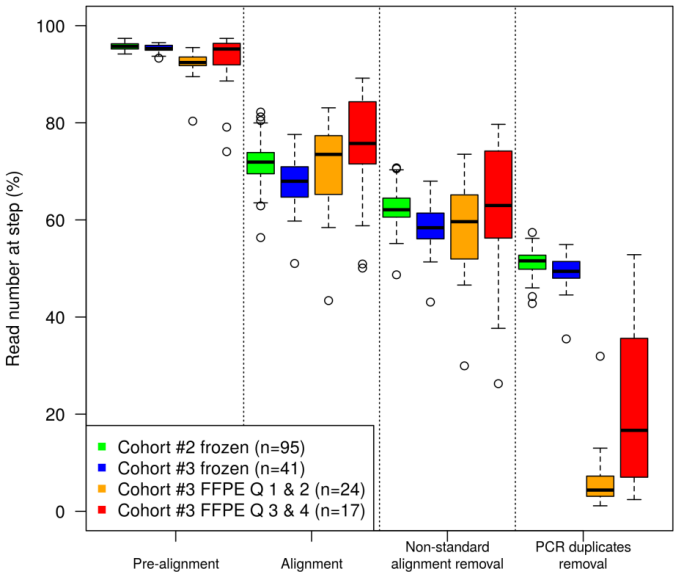


	Time (years)	0	2	4	6	8	10
C1	N risk	51	20	12	6	4	0
	Events	2	6	6	6	6	6
	MFS	0.96	0.86	0.86	0.86	0.86	0.86
C2	N risk	195	52	25	13	4	2
	Events	35	67	70	70	72	72
	MFS	0.82	0.58	0.54	0.54	0.42	0.42

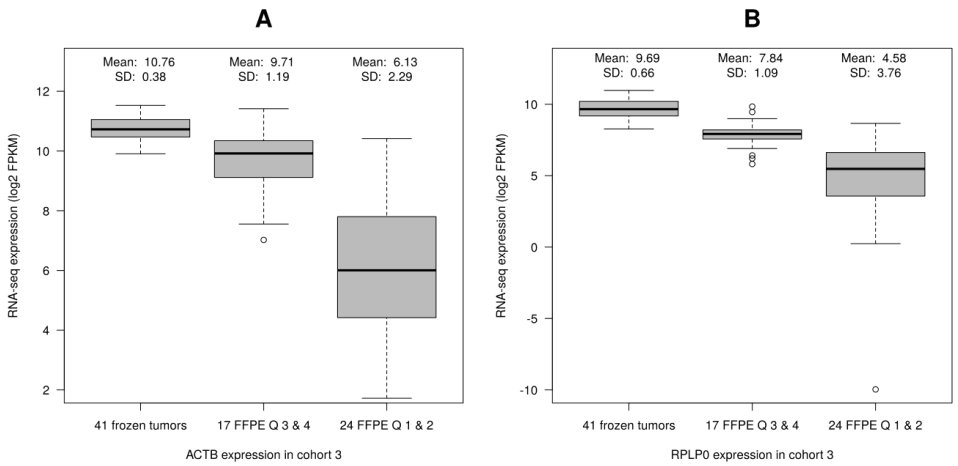
Supplementary Fig. 3 Kaplan-Meyer survival analysis according to CINSARC expression at transcript level on TCGA dataset with validation cohort (#2) centroids. This shows a significant metastatic prognosis using transcript expression. Legend: *C1* (Classification 1) is the low-risk profile group, *C2* (Classification 2) is the high-risk profile group, *N risk* is the number of patients in groups, *Events* is the cumulative number of metastasis in groups and *MFS* refers to metastasis-free survival.



Supplementary Fig. 4 Typical Agilent 2100 Bioanalyzer (Agilent Technologies) profiles of RNA extracted from FFPE blocks for each assigned quality from lowest to highest (1 to 4, left to right and top to bottom, respectively). Legend: *FU* is fluorescence unit, *nt* is nucleotides and *RIN* is RNA integrity number.



Supplementary Fig. 5 Read loss at each pipeline step. This shows read evolution through pipeline analysis at each main step. Comparing both material supports: there are significant read losses at stages: pre-alignment/PCR duplicates (higher losses in FFPE groups, Wilcoxon's $P<10^{-10}$) and alignment step (higher loss in frozen groups, Wilcoxon's $P=0.012$). No group difference at non-standard alignment removal step (Wilcoxon's $P=0.91$). Legend: *Q* refers to RNA quality.



Supplementary Fig. 6 Housekeeping gene expressions in cohort #3. (**A**) *ACTB* and (**B**) *RPLP0* expressions depending on material support and quality group (FFPE only). This shows a significant expression loss in FFPE quality 3 & 4 compared to frozen tumors (Wilcoxon's *P* values are 7.2×10^{-5} and 6.7×10^{-9} for *ACTB* and *RPLP0*, respectively) and an even more significant expression loss in FFPE quality 1 & 2 compared to frozen tumors (Wilcoxon's *P* values are $<10^{-10}$ for *ACTB* and *RPLP0*, respectively). Legend: *Q* refers to RNA quality.

Apport de l'étude

Ces résultats relatent le premier transfert de la signature CINSARC vers une plateforme technologique compatible avec les pratiques cliniques courantes : le RNA-seq. Nos résultats montrent qu'il est un procédé parfaitement viable pour l'obtention de la valeur pronostique de CINSARC sur tissus congelés, où les pronostics obtenus sont compatibles à ceux provenant de puces d'expression.

L'analyse d'ARN extraits de blocs FFPE est plus problématique. Si les cas évalués (au-dessus du seuil de dégradation) ont des profils transcriptomiques viables et des pronostics similaires à ceux obtenus à partir de tissus congelés, 59% des cas sont exclus car trop dégradés pour être analysés. Ceci constitue une limitation critique dans le contexte d'une application clinique. Il deviendrait nécessaire d'optimiser le processus de fixation des tissus ainsi que son stockage pour déterminer si des facteurs (âge du bloc, exposition au soleil, température de stockage, temps de fixation, *etc.*) peuvent être corrélés à la dégradation mesurée (von Ahlfen *et al.*, 2007). Le contrôle de ces paramètres permettrait une meilleure conservation de l'intégrité des acides nucléiques et confèrerait alors une meilleure applicabilité au RNA-seq sur blocs FFPE.

D'autres facteurs inhérents au RNA-seq peuvent limiter son application clinique : le prix du séquençage, le délai d'obtention du résultat (comprenant le séquençage et les analyses), la complexité de sa mise en place (moyens bioinformatiques/biostatistiques, stockage des données) ou encore sa reproductibilité (pipeline d'analyse utilisée). Il faut toutefois noter que ces contraintes se sont drastiquement atténuées depuis une décennie avec l'établissement de protocoles et d'outils de référence, rendant le RNA-seq beaucoup plus abordable (en matière de rapidité de mise en place, de simplicité d'analyse et de coût) que par le passé (Casamassimi *et al.*, 2017 ; Cieślik & Chinnaiyan, 2018).

Cette technologie n'étant, en l'état, pas adaptée à une activité clinique standard, nous avons alors testé une autre technologie qui permettrait d'outrepasser les limitations techniques dues à la dégradation des ARN.

3.1.2.2 – Article 2 : *Validation of the Complexity INdex in SARComas prognostic signature on formalin-fixed, paraffin-embedded, soft-tissue sarcomas*

Le Guellec S, Lesluyes T, Sarot E, Valle C, Filleron T, Rochaix P, Valentin T, Pérot G, Coindre JM & Chibon F. Article publié dans *Annals of Oncology* (2018).

Contexte

La dégradation des ARN extraits de blocs FFPE étant un facteur limitant pour le RNA-seq, nous nous sommes tournés vers le système nCounter (NanoString) qui propose un protocole adapté à de tels échantillons. En effet, son processus de quantification d'expression ne passe ni par une étape de rétro-transcription des ARN, ni par une étape d'amplification de séquences. Il se base sur un système d'hybridation entre les ARN et des sondes spécifiques contenant un code-barres coloré (enchaînement de couleurs spécifiques), dont le produit d'hybridation va être fixé sur un support solide. Chaque molécule d'ARN capturée peut alors être détectée par une caméra à fluorescence en peu d'étapes, réduisant ainsi les possibles sources de biais. Ceci permet une quantification d'expression très automatisée, sans réaction enzymatique ni de préparation particulière de librairie (Geiss *et al.*, 2008).

Ce système a d'ores et déjà rendu certains *CodeSets* d'intérêt (sélection de sondes s'hybridant aux gènes spécifiés) accessibles en routine clinique par l'analyse de blocs FFPE. On peut par exemple citer Prosigna (PAM50), signature transcriptomique de 50 gènes, prédictive du développement métastatique des cancers du sein à dix ans (**partie 1.2.3.5 page 59** ; Parker *et al.*, 2009 ; Nielsen *et al.*, 2014). Il existe aussi plusieurs *CodeSets* qui proposent la détection simultanée de dizaines de transcrits de fusion classiquement retrouvés dans les différents sous-types de sarcomes "à translocation" (Chang *et al.*, 2018 ; Sheth *et al.*, 2018). La nature de cette technologie ne permettant pas de mesurer l'intégralité du transcriptome, le design de *CodeSets* est fait sur demande jusqu'à un maximum de 800 cibles.

Résumé de l'article

Nous avons généré un *CodeSet*, nommé NanoCind, comprenant les 67 gènes de CINSARC et huit gènes de ménage (*ACTB*, *B2M*, *GAPDH*, *HPRT1*, *PGK1*, *PPIA*, *RPLP0* et *SDHA*)

utilisés pour la normalisation des échantillons et comme contrôle qualité interne.

Afin de tester l'applicabilité clinique de NanoCind avec cette technologie, nous avons défini deux cohortes. La première cohorte est constituée de 124 cas passés à la fois en RNA-seq et en puces NanoString sur tissus congelés afin de comparer la valeur pronostique de chacune de ces méthodes. La seconde cohorte est constituée de 67 cas extraits de blocs FFPE (dont 20 micro-biopsies) pour challenger les effets de la dégradation des ARN sur l'obtention des pronostics.

Sur la cohorte de 124 cas, CINSARC évalué avec la puce NanoCind a une valeur pronostique significative (*P*=1,01e-2) alors que celle obtenue avec le RNA-seq est non significative (*P*=9,68e-2). Les 67 cas FFPE, incluant les 20 micro-biopsies, ont tous pu être analysés par NanoCind avec une évaluation correcte du risque métastatique (*P*=1,13e-2).

Parmi les 116 cas, passés à la fois en RNA-seq et en puces NanoString, pour lesquels l'information du grade FNCLCC était disponible, seule la technologie NanoString donne une stratification significativement pronostique (*P*=9,39e-3 contre *P*=1,16e-1 et *P*=6,58e-1 pour le RNA-seq et le grade FNCLCC). De plus, NanoCind retrouve une stratification significative parmi les tumeurs de grade 2 par son évaluation du risque métastatique (43 cas ; *P*=5,16e-4).

Le système nCounter développé par NanoString nous a ainsi permis de créer une puce de référence pour l'évaluation pronostique de CINSARC : NanoCind. Cette solution technologique propose alors l'obtention du pronostic de la signature sur ARN provenant de tissus congelés et également de blocs FFPE où la dégradation n'est pas un facteur limitant.

L'ensemble des résultats de cette étude sont inclus ci-après (Le Guellec *et al.*, 2018).

ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

Annals of Oncology 29: 1828–1835, 2018
doi:10.1093/annonc/mdy194
Published online 31 May 2018

ORIGINAL ARTICLE

Validation of the Complexity INdex in SARComas prognostic signature on formalin-fixed, paraffin-embedded, soft-tissue sarcomas

S. Le Guellec[1,2*], T. Lesluyes[2,3,4,5], E. Sarot[5,6], C. Valle[6], T. Filleron[7], P. Rochaix[1,2], T. Valentin[2,8], G. Pérot[3,9], J.-M. Coindre[4,9] & F. Chibon[1,2]

[1]Department of Pathology, Institut Claudius Regaud, IUCT-Oncopole, Toulouse; [2]INSERM U1037, Cancer Research Center of Toulouse (CRCT), Toulouse; [3]INSERM U1218, Institut Bergonié, Bordeaux; [4]University of Bordeaux, Bordeaux; [5]Institut Claudius Regaud, IUCT-Oncopole, Toulouse; [6]Plateau Génomique et Transcriptomique, INSERM U1037, Cancer Research Center of Toulouse (CRCT), Pôle technologique, Toulouse; Departments of [7]Biostatistics; [8]Oncology, Institut Claudius Regaud, IUCT-Oncopole, Toulouse; [9]Department of Biopathology, Institut Bergonié, Bordeaux, France

Correspondence to: Dr Sophie Le Guellec, Department of Pathology, Institut Claudius Regaud, Institut Universitaire du Cancer Toulouse-Oncopole, 1 avenue Irène Joliot-Curie, 31059 Toulouse Cedex 9, France. Tel: +33-531-156-594; E-mail: LeGuellec.sophie@iuct-oncopole.fr

Background: Prediction of metastatic outcome in sarcomas is challenging for clinical management since they are aggressive and carry a high metastatic risk. A 67-gene expression signature, the Complexity INdex in SARComas (CINSARC), has been identified as a better prognostic factor than the reference pathological grade. Since it cannot be applied easily in standard laboratory practice, we assessed its prognostic value using nanoString on formalin-fixed, paraffin-embedded (FFPE) blocks to evaluate its potential in clinical routine practice and guided therapeutic management.

Methods: A code set consisting of 67 probes derived from the 67 genes of the CINSARC signature was built and named NanoCind®. To compare the performance of RNA-seq and nanoString (NanoCind®), we used expressions of various sarcomas ($n = 124$, frozen samples) using both techniques and compared predictive values based on CINSARC risk groups and clinical annotations. We also used nanoString on FFPE blocks ($n = 67$) and matching frozen and FFPE samples ($n = 45$) to compare their level of agreement. Metastasis-free survival and agreement values in classification groups were evaluated.

Results: CINSARC strongly predicted metastatic outcome using nanoString on frozen samples (HR = 2.9, 95% CI: 1.23–6.82) with similar risk-group classifications (86%). While more than 50% of FFPE blocks were not analyzable by RNA-seq owing to poor RNA quality, all samples were analyzable with nanoString. When similar (risk-group) classifications were measured with frozen tumors (RNA-seq) compared with FFPE blocks (84% agreement), the CINSARC signature was still a predictive factor of metastatic outcome with nanoString on FFPE samples (HR = 4.43, 95% CI: 1.25–15.72).

Conclusion: CINSARC is a material-independent prognostic signature for metastatic outcome in sarcomas and outperforms histological grade. Unlike RNA-seq, nanoString is not influenced by the poor quality of RNA extracted from FFPE blocks. The CINSARC signature can potentially be used in combination with nanoString (NanoCind®) in routine clinical practice on FFPE blocks to predict metastatic outcome.

Key words: sarcoma, FFPE, prognosis, cancer, nanoString, NanoCind®

Introduction

Adult soft-tissue sarcomas (STSs) are rare tumors (<1%) that form a heterogeneous group with more than 100 different pathological subtypes and an aggressive clinical course (40/50% metastases within 5 years of diagnosis) [1]. Clinical management consists in surgical resection with optional adjuvant therapies that depend on clinical characteristics, tumor histological type and histological grade [2]. The histological grading system of the FNCLCC (*Fédération Nationale des Centres de Lutte Contre le Cancer*) comprises three grades: low, intermediate and high risk of metastasis. It is at present the

best predictor of metastatic relapse in early-stage sarcoma and the most influential for deciding on adjuvant chemotherapy [3]. However, it has several limitations. It is not applicable in all pathological subtypes, its reproducibility may vary between pathologists, it is poorly informative grade 2, which represents ~40% of cases, and it is difficult to apply on tumor microbiopsies and on surgical specimens post neo-adjuvant therapy [4]. In 2010, our group identified and validated a 67-gene expression signature named CINSARC (Complexity Index in SARComas) that is a valuable prognostic factor in several sarcoma histotypes and outperforms histological grade [5–7]. CINSARC, in which genes are involved in mitotic control and chromosomal integrity, stratifies tumor prognosis into two groups ("low-risk, C1" and "high-risk, C2"), instead of three with the FNCLCC system, thus facilitating clinical management. It was first used on frozen tumors analyzed by microarrays and was subsequently applied on formalin-fixed, paraffin-embedded (FFPE) RNA-sequencing [8]. However, the RNA is degraded in more than half of all tested FFPE tumors so its routine use for clinical and therapeutic purposes has been limited until now. Indeed, the availability of fresh-frozen tumor material is very low in daily routine practice. Archived FFPE tumor-tissue samples are available in anatomical pathology laboratories and tissue banks. Unfortunately, formalin fixation induces chemical modifications such as crosslinking between nucleic acids and proteins and RNA degradation, so the subsequent use of frozen tissue for genomic and transcriptomic applications such as microarrays and RNA sequencing is limited [9]. NanoString, a recently developed probe-based technique using direct digital measurement of transcript abundance with multiplexed color-coded probe pairs, has been shown to quantify mRNA expression accurately in FFPE and fragmented material [10, 11]. This feature of nanoString offers advantages compared with PCR-based methods, including the absence of amplification bias, which may be higher when using fragmented RNA isolated from FFPE blocks. Several publications have demonstrated the accuracy and precision of nanoString with FFPE material [10, 12, 13]. A code set comprising 67 probes derived from the 67 genes of the CINSARC signature was built and named NanoCind®. Thus, the aim of the present study was to evaluate the prognostic value of the CINSARC signature established by nanoString (NanoCind®) on FFPE tumor blocks in order to use it in routine clinical practice to determine the risk of developing metastases and to guide personalized therapeutic management.

To compare the performance of RNA-seq and nanoString in establishing CINSARC-based risk, we applied both techniques on sarcoma cohorts with clinical annotations from the Conticabase (European sarcoma database). We tested a large cohort of 124 frozen samples and a subgroup of 45 samples together with matching FFPE blocks plus 20 additional FFPE core needle biopsies to test the performance of NanoCind® on degraded material typical of that used in daily practice.

Materials and methods

More details are available in the Supplementary Methods file, available at *Annals of Oncology* online.

Cohorts and design

To evaluate the potential transfer of the CINSARC signature from RNA-seq on fresh-frozen samples (FFSs) to nanoString on FFPE blocks, we used two independent cohorts. The first one which comprised 95 sarcomas was used in 2016 by Chibon's team to validate the transfer of CINSARC from microarrays to RNA-seq on FFS [8]. Of the 95 sarcomas described in that publication, FFS or RNA were still available for 77 cases. A second set of 47 sarcomas was added since RNA-seq expression on FFS ($n=47$) and FFPE ($n=45$) had already been tested by our team. Therefore, for the transfer from RNA-seq to nanoString, we analyzed a set of 124 (77 and 47) sarcomas (named cohort #1) by both RNA-seq and nanoString (NanoCind®) on FFS. To evaluate whether CINSARC had a prognostic value on FFPE tissue, we analyzed the 47 sarcomas plus 20 microbiopsies (cohort #2) with NanoCind® on FFPE blocks.

Tumor samples

Tumors were operated in expert centers of the French Sarcoma Group and came from surgical resections or microbiopsies of untreated primary tumors. RNA was extracted from all FFPE blocks ($n=67$) in cohort #2 using the High Pure FFPET RNA Isolation Kit (Roche). To optimize the extraction protocol, we tested a second extraction kit (Roche FFPET RNA Isolation Kit, used in Prosigna Breast Cancer Prognostic Gene Signature). Finally, Lesluyes et al. used the RNeasy FFPE kit (Qiagen) in 2016 for RNA extraction in 45 cases in cohort #2, but only 17 were analyzable by RNA-seq, owing to poor RNA quality as a result of formalin fixation.

NanoString CodeSet design and expression quantification

Our nCounter code set (NanoCind®) consisted of a panel of 75 probes, including 67 distinct test probes derived from 67 distinct genes (Supplementary Table S1, available at *Annals of Oncology* online) and 8 housekeeping genes for biological normalization purposes.

Results

Cohorts

Tumor characteristics and follow-up of patients in both cohorts are presented in Table 1.

NanoString nCounter gene expression system performance

First the reproducibility of the nCounter system in measuring the NanoCind® panel was examined. The raw counts for all 75 genes on the same sample from two independent hybridizations of RNA from FFS are shown in Supplementary Figure S1A, available at *Annals of Oncology* online (between two independent cartridges) and S1B (in the same cartridge). A linear fit to the data resulted in a median Pearson's correlation coefficient of 0.9849 (inter- and intra-cartridges). Then, we examined the linearity and dynamic range of two FFS. Supplementary Figure S1, available at *Annals of Oncology* online shows the results for all 75 genes from each hybridization reaction with inputs of 100 ng (reference input), 300, 600 and 1000 ng of RNA from the same FFS. Raw points demonstrated a linear correlation between different inputs of the same sample (Supplementary Figure S1C, available at *Annals of Oncology* online).

Table 1. Characteristics of processed cohorts

Cohort	#1 ($n = 124$)	#2 ($n = 67$)
Expression quantification	RNA-seq and NanoString	RNA-seq and NanoString
Material	Fresh-frozen tissue	FFPE blocs
Median age at diagnosis (years) [95% CI]	63 [60–65]	64 [62–68]
Histotype (%)		
Dedifferentiated liposarcoma	15 (12.10)	17 (25.37)
Gastrointestinal stromal tumor	12 (9.68)	11 (16.42)
Leiomyosarcoma	32 (25.81)	17 (25.37)
Myxofibrosarcoma	11 (8.87)	4 (5.97)
Undifferentiated pleomorphic sarcoma	39 (31.45)	17 (25.37)
Others	15 (12.10)	1 (1.49)
Gender (%)		
Female	61 (49.19)	37 (55.22)
Male	63 (50.81)	30 (44.78)
FNCLCC grade (%)		
1	5 (4.03)	1 (1.49)
2	27 (21.77)	16 (23.88)
3	74 (59.68)	37 (55.22)
Unknown	6 (4.84)	2 (2.99)
NA (GIST)	12 (9.68)	11 (16.42)
Surgical margin (%)		
R0	53 (42.74)	37 (55.22)
R1	42 (33.87)	21 (31.34)
R2	3 (2.42)	0
Unknown or NA	26 (20.97)	9 (13.43)
Median follow-up (month) [95% CI]	57.31 [45.27–66.60]	44.35 [33.18–52.49]
Local recurrence (%)	39 (31.45)	13 (19.4%)
Metastasis (%)	48 (38.71)	15 (22.39)

Among the 67 cases in cohort #2: all were processed with nanoString and 17 were processed with RNA-seq. The intersection between nanoString on frozen tissues and nanoString on FFPE blocs is 45 cases. This cohort includes 47 regular FFPE blocs and 20 FFPE core needle biopsies. See Supplementary Figure S4, available at *Annals of Oncology* online for full details. Others cohort #1: two low-grade fibromyxoid sarcomas, two malignant solitary fibrous tumors, six pleomorphic liposarcomas, four pleomorphic rhabdomyosarcomas and one synovial sarcoma. Others cohort #2: one pleomorphic liposarcoma.
FFT, fresh-frozen tissue; FFPE, formalin-fixed, paraffin-embedded; R, RNA sequencing; N, NanoString.

Technological transfer: from RNA-seq to nanoString (NanoCind®)

We carried out comparisons of the same FFS (cohort #1, $n = 124$) between both techniques. Pearson's correlation of 67 CINSARC gene expression values for the same high quality RNA samples between RNA-seq and nanoString was 0.797 ± 0.130 (Supplementary Figure 2B, available at *Annals of Oncology* online). Thus, we aimed to test the prognostic value of CINSARC. An agreement test between the prognostic groups (CINSARC C1, low-risk and CINSARC C2, high-risk) by nanoString on FFS resulted in a similar CINSARC classification for 107 of 124 transcriptomic profiles (86% agreement; Cohen's kappa = 0.687; $P < 10^{-10}$) compared with RNA-seq classification on FFS. By combining associated clinical data outcome, we carried out metastasis-free survival analysis. NanoString had a significant prognostic value ($P = 1.01 \times 10^{-2}$, HR = 2.9, 95% CI: 1.23–6.82) (Figure 1A) while RNA-seq did not ($P = 9.68 \times 10^{-2}$, HR = 1.69, 95% CI: 0.89–3.2) (Figure 1B, Table 2, Supplementary Figure S4, available at *Annals of Oncology* online).

Material transfer: from fresh-frozen tissue to FFPE blocks

To optimize the quantity and quality of RNA extraction, we carried out different extractions on the same FFPE blocks with different commercial kits. Mean Pearson's correlation of the raw counts with the NanoCind® panel for the High Pure FFPET RNA Isolation Kit, Roche (reference kit) compared with the Roche FFPET RNA Isolation Kit (used in Prosigna) ($n = 2$) was 0.993 (Supplementary Figure S3A and B, available at *Annals of Oncology* online). Mean Pearson's correlation of the raw counts with the NanoCind® panel for the High Pure FFPET RNA Isolation Kit, Roche (reference kit) compared with the RNeasy FFPE kit, Qiagen Kit ($n = 27$) was 0.786 (range, 0.104–0.984) (Supplementary Figure S3C, available at *Annals of Oncology* online). Moreover, we assessed the stability of RNA from unstained slides following extraction at different times after cutting with a view to using this technique in routine practice. On the same FFPE block, we carried out RNA extractions on unstained slides the same day of the cut (J0, reference protocol) and 21 days room

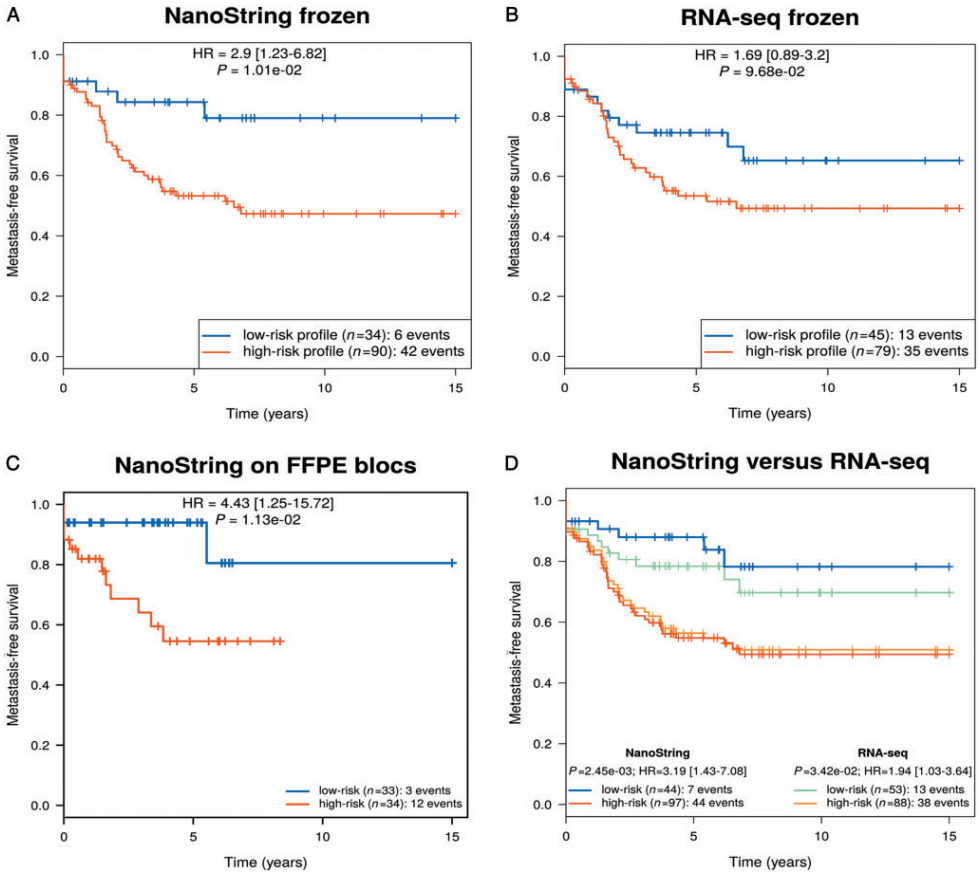


Figure 1. Metastasis-free survival analysis according to CINSARC signature. (A) Prognostic value of nanoString expression data (cohort #1, fresh-frozen samples). (B) Prognostic value of RNA-seq expression data (cohort #1, fresh-frozen samples). (C) Prognostic value of nanoString expression data (NanoCind®) (cohort #2, FFPE samples). (D) Comparison of prognostic values of the two techniques. Fresh-frozen and FFPE samples used with each technique were pooled.

temperature storage after the day of cutting (J21). Pearson's correlation for interval extraction replicate was 0.990 (Supplementary Figure S3D, available at *Annals of Oncology* online).

Thus, we carried out comparisons of paired FFS and FFPE samples ($n = 45$) with nanoString. NanoString provided a good correlation between all 45 pairs of fresh-frozen and FFPE RNA (mean Pearson's $r = 0.651 \pm 0.118$; Supplementary Figure S2A, available at *Annals of Oncology* online). Moreover, Pearson's correlation of 67 CINSARC gene expression values for the same samples ($n = 45$) between RNA-seq on FFS (reference material and technique) and nanoString on FFPE blocks was 0.687 ± 0.104 (Supplementary Figure S2C, available at *Annals of Oncology* online). Then, we tested the prognostic value of the CINSARC signature evaluated by nanoString on FFPE samples. The agreement test between prognostic groups (CINSARC C1, low-risk and CINSARC C2, high-risk) by nanoString on FFPE samples resulted in a similar CINSARC classification for 38 of 45 transcriptomic profiles (84% agreement; Cohen's kappa = 0.688; $P = 3.79 \times 10^{-6}$) compared with RNA-seq classification on FFS (reference technique and materials). Moreover, nanoString had a significant prognostic value ($P = 1.13 \times 10^{-2}$, HR = 4.43, 95% CI: 1.125–15.72) (Figure 1C, Table 2, Supplementary Figure S4, available at *Annals of Oncology* online) on FFPE samples (surgical resections ($n = 47$) and microbiopsies ($n = 20$).

Finally, we carried out metastasis-free survival analysis of the pooled FFS and FFPE samples ($n = 141$) with both RNA-seq and nanoString analyses (Figure 1D). Both techniques provided a significant prognostic value but the CINSARC signature with nanoString (NanoCind®)($P = 2.45 \times 10^{-3}$, HR = 3.19, 95% CI: 1.43–7.08) was

Table 2. Prognostic analyses for metastasis-free survival according to CINSARC signature and to FNCLCC grading system according to cohorts, materials and technologies

Characteristics	*Total, n*	Prognosis, *n*	*P*-value	HR [95% CI]
Cohort #1				
CINSARC, RNA-seq, fresh-frozen tissue	124		0.0968	1.69 [0.89–3.2]
C1 (*n* metastasis)		45 (13)		
C2 (*n* metastasis)		79 (35)		
CINSARC, nanoString (NanoCind®), fresh-frozen tissue	124		**0.0101**	2.9 [1.23–6.82]
C1 (*n* metastasis)		34 (6)		
C2 (*n* metastasis)		90 (42)		
Cohort #2				
CINSARC, RNA-seq, FFPE	17		0.134	5.12 [1.61–5.49]
C1 (*n* metastasis)		8 (0)		
C2 (*n* metastasis)		9 (3)		
CINSARC, nanoString (NanoCind®), FFPE	67		**0.0113**	4.43 [1.25–15.72]
C1 (*n* metastasis)		33 (3)		
C2 (*n* metastasis)		34 (12)		
Pooled cohort #1 and #2				
FNCLCC grade	116		0.658	1.16 [0.6–2.26]
G1/G2 (*n* metastasis)		35 (12)		
G3 (*n* metastasis)		81 (32)		
CINSARC, RNA-seq	116		0.116	1.74 [0.88–3.53]
C1 (*n* metastasis)		36 (10)		
C2 (*n* metastasis)		80 (34)		
CINSARC, nanoString (NanoCind®)	116		**0.00939**	3.58 [1.28–10]
C1 (*n* metastasis)		27 (4)		
C2 (*n* metastasis)		89 (40)		

Significant *P*-values <0.05 are in bold.
FNCLCC, Fédération Nationale des Centres de Lutte Contre le Cancer; G1, grade 1 FNCLCC; G2, grade 2 FNCLCC; G3, grade 3 FNCLCC; CINSARC, Complexity INdex in SARComas; C1, low-risk CINSARC; C2, high-risk CINSARC; HR, hazard ratio; CI, confidence interval; FFPE, formalin-fixed, paraffin-embedded.

more discriminating than that obtained by RNA sequencing ($P = 3.42 \times 10^{-2}$, HR = 1.94, 95% CI: 1.03–3.64). For discordant samples, we found a lower strength of association with the nearest centroid using RNA-seq versus nanoString ($P = 9.49 \times 10^{-3}$; Supplementary Figure S5A, available at *Annals of Oncology* online) compared with the same experiment with concordant samples ($P = 4.36 \times 10^{-2}$; Supplementary Figure S5B, available at *Annals of Oncology* online). Among the pooled FFS and FFPE samples ($n = 141$) with data obtained by RNA-seq and nanoString, the FNCLCC histological grade was available for 116 cases. The CINSARC signature with nanoString (NanoCind®) discriminated two groups of low- and high-risk subjects with clearly significant and different outcomes ($P = 9.39 \times 10^{-3}$, HR = 3.58, 95% CI: 1.28–10, Figure 2A), whereas its prognostic value with RNA-seq was not significant ($P = 1.16 \times 10^{-1}$, HR = 1.74, 95% CI: 0.86–3.53, Figure 2B) and histological grade (G1/G2 versus G3) did not predict outcome ($P = 6.58 \times 10^{-1}$, HR = 1.16, 95% CI: 0.6–2.26, Figure 2C). Consequently, we also investigated the performance of CINSARC in subjects of the same histological grade and in grade 2 tumors, the CINSARC signature (NanoCind®) identified two groups of different outcomes with a significant prognostic value ($P = 5.16 \times 10^{-4}$, HR = 15.73, 95% CI: 2.01–123.37, Figure 2D).

Discussion

STSs are a group of rare, heterogeneous, aggressive tumors with high metastatic risk. The CINSARC signature outperforms histological grade for metastatic prognosis in several cancer types and especially in STS [5–7]. Nevertheless, it is not applicable in routine clinical practice because it was developed and validated in microarrays and RNA-sequencing, which are techniques requiring large amounts of high-quality RNA, especially from fresh-frozen tumor tissue. In the current study, the CINSARC signature established by nanoString (NanoCind®) on FFPE blocks was a strong predictor for metastatic events (which outperformed histological grade) and is therefore usable in routine clinical settings.

In the first part of the CINSARC transfer, CINSARC gene expression values correlated well using these two techniques (mean Pearson's *r* is 0.797 between RNA-seq and nanoString), which is consistent with already published methodological correlations [10–13]. Furthermore, similar risk-group classifications (86%) were noted. However, while its prognostic value with RNA-seq was nonsignificant (reference technique, $P = 9.68 \times 10^{-2}$), it was significant with nanoString (e$P = 1.01 \times 10^{-2}$).

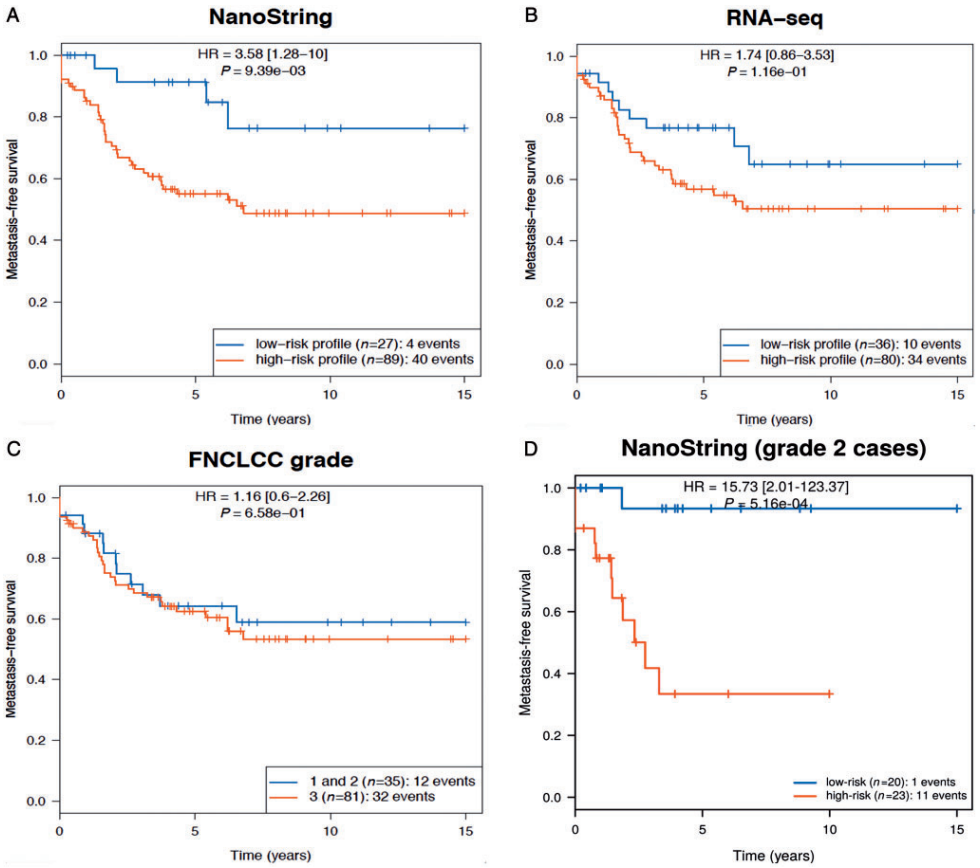


Figure 2. Metastasis-free survival analysis (fresh-frozen samples, cohort #1 and FFPE samples, cohort #2) according to CINSARC signature using nanoString expression data (NanoCind®) (A), RNA-seq expression data (B) and FNCLCC grade (C). (D) Metastasis-free survival analysis according to CINSARC signature using nanoString expression data (NanoCind®) on grade 2 FNCLCC cases.

In recent decades, high-throughput molecular analyses, especially gene expression profiling, have led to the identification of specific gene expression signatures as a prognostic factor in several malignancies such as breast cancer and lymphoma [14, 15]. For example, several multigene signatures such as Oncotype™ or Prosigna™ are marketed for early-stage breast cancer and are used in routine practice for therapeutic management [16]. No similar signature is currently available in routine practice for sarcoma patients. Indeed, although Chibon et al. [5] identified and validated a prognostic molecular signature in sarcomas (CINSARC) in 2010, they worked on frozen tumor material that was evaluated by microarrays, a technique that is not widely available. Moreover, several studies have shown the validity of the CINSARC signature [5–7, 17, 18], in accordance with the REMARK criteria for replicability of biomarker signatures [19]. In 2016, Lesluyes et *al.* successfully transferred the signature using RNA-seq with frozen tissues. However, their attempt to adapt it on FFPE blocks was disappointing with more than 50% of FFPE cases that could not be analyzed with RNA-seq owing to degradation of RNA induced by formalin [8]. In our study, all ($n = 67$) FFPE blocks (the same cases as those tested by Lesluyes et *al.*, except microbiopsies) were analyzable with nanoString. Moreover, in terms of prognosis, the CINSARC signature obtained with nanoString (NanoCind®) from RNA extracted from FFPE blocks had a significant prognostic value ($P = 1.13 \times 10^{-2}$, HR = 4.43, 95% CI: 1.125–15.72), allowing its use in routine practice. The low quality of RNA extracted from FFPE blocks did not influence the efficiency of the results when nanoString was used. These findings are in accordance with previous studies reporting the robust performance of nanoString on RNA extracted from FFPE material with results comparable to those obtained with RNA from fresh-frozen tissue by microarrays [13]. Indeed, nanoString

does not require reverse transcription of mRNA and subsequent cDNA amplification. This feature offers advantages over NGS and PCR-based methods, including the absence of amplification bias, which may be higher when using fragmented RNA isolated from FFPE specimens. This could explain the few discordant classifications we observed and particularly why poor prognosis patients were classified as "good" prognosis by the RNA-seq approach. The value and the superiority of nanoString (NanoCind®) were confirmed by three of our results: first, it gave a significant prognostic value which was more discriminating than that established by RNA-seq ($P = 2.45 \times 10^{-3}$ versus $P = 3.42 \times 10^{-2}$); second, it outperformed the current histological grade (FNCLCC) ($P = 9.39 \times 10^{-3}$ versus $P = 6.58 \times 10^{-1}$) and third, in grade 2 (FNCLCC) tumors, it clearly and significantly differentiated two groups of different outcomes ($P = 5.16 \times 10^{-4}$). Its advantage over histologic grading is that it stratifies patients into two groups instead of three, which is essential for facilitating clinical management. A major point that we have started to test is its applicability to microbiopsies, which is at present the gold standard in terms of STS management [1]. All 20 cases (FFPE, microbiopsies) could be analyzed using nanoString, despite somewhat degraded samples with limited amounts of RNA. Moreover, among the 20 FFPE microbiopsies, three patients have metastases and these three tumors were classified as "poor outcome" by nanoString.

To be able to use this signature in routine clinical practice, we tested it in several conditions. It showed excellent reproducibility (replicates averaging R^2 of 0.983) and robustness, consistent with previous reports [10]. Moreover, we tested three different commercial nucleic acid extraction kits for FFPE material. The two we tested demonstrated good results with high (Pearson's $r = 0.786$ for RNeasy FFPE kit, Qiagen) to excellent (Pearson's $r = 0.993$ for Roche FFPET RNA Isolation Kit, used in Prosigna) correlation. Finally, the test turnaround time with nanoString is shorter than for NGS approaches; library preparation and nucleic acid amplification are not required and the complexity of data analysis is negligible compared with NGS, which requires a specialized bioinformatics pipeline and sequence analysis strategies.

Conclusion

In conclusion, the CINSARC signature, which can now be determined using FFPE samples with nanoString (NanoCind®), has reached another step towards its use in routine practice. Following on from recent data demonstrating that it is an excellent prognostic marker of clinical outcome in many other human malignancies [17], it can now be tested and validated on other independent cancer-specific cohorts with FFPE blocks. In patients with STS, its prognostic value has been widely confirmed and it seems to be better than FNCLCC grade, as suggested by the present study and previously observed [5, 7]. Above all, the CINSARC signature could overcome one of the major limitations of FNCLCC grading by splitting patients with FNCLCC grade 2 STS into two separate prognostic groups that could then benefit from different therapeutic strategies. Furthermore, a French Sarcoma Group study aiming to tailor treatment based on the CINSARC signature will soon begin. Finally, as expected since several publications have demonstrated the accuracy and precision of nanoString in small samples with limited amounts of RNA such as tissue biopsies [10, 12, 13], our preliminary data on 20 cases suggest that Nanocind® is feasible in FFPE microbiopsies, i.e. the standard approach in STS diagnosis [1]. Further NanoCind® evaluation in a larger cohort of FFPE microbiopsies is currently ongoing and will be the final step to consider full CINSARC routine evaluation and global use.

Acknowledgements

The authors thank Mme L. M'Hamdi and Mme L. Puydenus from *SBEC-BEC* (Support Biopathologique aux Etudes Cliniques - Bureau des Etudes Cliniques) from Institut Claudius Regaud-IUCT-Oncopole for their technical support. The data used in this publication were provided by the Conticanet (Connective Tissue Cancer Network) database (https://conticabase.sarcomabcb.org) and the French sarcoma network RRePS (https://rreps.sarcomabcb.org). These databases are financially supported by Conticanet and INCa (Institut National du Cancer). The following centers participated in the study: Paul Papin Center, Angers; Bergonié Institute, Bordeaux; Jean Perrin Center, Clermont-Ferrand; Georges-Francois Leclerc Center, Dijon; Oscar Lambret Center, Lille; Léon-Bérard Center, Lyon; Hôtel-Dieu CHU, Nantes; Antoine Lacassagne Center, Nice; CHU La Timone, Marseille, CHU-Guy de Chauliac, Montpellier; Institut Gustave Roussy, Paris; Paul Strauss Center, Strasbourg.

Funding

«La Ligue Nationale Contre le Cancer-Comité de la Haute Garonne» and "EUROSARC" ("FP7-HEALTH-2011").

Disclosure

CINSARC is filed under the number "WO 2010/122243 A1". NanoCind® is filed under the number «EP18305190.3».

References

1. Fletcher CDM, Bridge JA, Hogendoorn PCW et al. World Health Organization Classification of Tumours: Pathology and Genetics of Tumours of Soft Tissue and Bone. Lyon, France: IARC Press 2012.
2. Trojani M, Contesso G, Coindre JM et al. Soft-tissue sarcomas of adults; study of pathological prognostic variables and definition of a histopathological grading system. Int J Cancer 1984; 33(1): 37–42.
3. Italiano A, Delva F, Mathoulin-Pelissier S et al. Effect of adjuvant chemotherapy on survival in FNCLCC grade 3 soft tissue sarcomas: a multivariate analysis of the French Sarcoma Group Database. Ann Oncol 2010; 21(12): 2436–2441.
4. Coindre JM, Terrier P, Guillou L et al. Predictive value of grade for metastasis development in the main histologic types of adult soft tissue sarcomas: a study of 1240 patients from the French Federation of Cancer Centers Sarcoma Group. Cancer 2001; 91(10): 1914–1926.
5. Chibon F, Lagarde P, Salas S et al. Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. Nat Med 2010; 16(7): 781–787.
6. Lagarde P, Pérot G, Kauffmann A et al. Mitotic checkpoints and chromosome instability are strong predictors of clinical outcome in gastrointestinal stromal tumors. Clin Cancer Res 2012; 18(3): 826–838.

7. Lagarde P, Przybyl J, Brulard C et al. Chromosome instability accounts for reverse metastatic outcomes of pediatric and adult synovial sarcomas. J Clin Oncol 2013; 31(5): 608–615.
8. Lesluyes T, Pérot G, Largeau MR et al. RNA sequencing validation of the Complexity INdex in SARComas prognostic signature. Eur J Cancer 2016; 57: 104–111.
9. Von Ahlfen S, Missel A, Bendrat K, Schlumpberger M. Determinants of RNA quality from FFPE samples. PLoS One 2007; 2(12): e1261.
10. Geiss GK, Bumgarner RE, Birditt B et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol 2008; 26(3): 317–325.
11. Malkov VA, Serikawa KA, Balantac N et al. Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter Assay System. BMC Res Notes 2009; 2(1): 80.
12. Northcott PA, Shih DJ, Remke M et al. Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. Acta Neuropathol 2012; 123(4): 615–626.
13. Reis PP, Waldron L, Goswami RS et al. mRNA transcript quantification in archival samples using multiplexed, color-coded probes. BMC Biotechnol 2011; 11(1): 46.
14. Perou CM, Sørlie T, Eisen MB et al. Molecular portraits of human breast tumours. Nature 2000; 406(6797): 747–752.
15. Lossos IS, Czerwinski DK, Alizadeh AA et al. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. N Engl J Med 2004; 350(18): 1828–1837.
16. Gnant M, Sestak I, Filipits M et al. Identifying clinically relevant prognostic subgroups of postmenopausal women with node-positive hormone receptor-positive early-stage breast cancer treated with endocrine therapy: a combined analysis of ABCSG-8 and ATAC using the PAM50 risk of recurrence score and intrinsic subtype. Ann Oncol 2015; 26(8): 1685–1691.
17. Lesluyes T, Delespaul L, Coindre JM, Chibon F. The CINSARC signature as a prognostic marker for clinical outcome in multiple neoplasms. Sci Rep 2017; 7(1): 5480. 7: 5480.
18. Brulard C, Chibon F. Robust gene expression signature is not merely a significant P value. Eur J Cancer 2013; 49(12): 2771–2773.
19. McShane LM, Altman DG, Sauerbrei W et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). Eur J Cancer 2005; 41(12): 1690–1696.

Validation of the Complexity INdex in SARComas prognostic signature on formalin-fixed, paraffin-embedded, soft tissue sarcomas

S. Le Guellec [1,2], T. Lesluyes [2,4,5,6], E. Sarot [3,6], C. Valle [3], T. Filleron[7], P. Rochaix [1,2], T. Valentin [2,8], G. Pérot [4,9], J-M.Coindre [5,9] and F. Chibon [1,2]

Supplementary methods

Tumor samples

Tumors were operated in expert centers of the French Sarcoma Group (FSG) and came from surgical resections of untreated primary tumors. Every case was histologically reviewed by the pathologist subgroup from the FSG and classified according to the 2013 World Health Organization classification by histology, immunohistochemistry and molecular genetics when needed. FFPE blocks were first verified by a pathologist (SLG or JMC) on the corresponding HES slide and only areas with a minimum of 60% of tumor cells were selected for RNA extraction. RNA was no longer available in 45 cases in cohort #1, so a new extraction was performed from fresh-frozen samples using the same protocols as those described by Lesluyes *et al.* [1]. RNA was extracted from all FFPE blocks (n=47) in cohort #2 (10-micron-thick sections) using a deparaffinization solution and RNA purification using the High Pure FFPET RNA Isolation Kit (Roche), according to the manufacturer's recommendations. One to three (for surgical resections) and three to eight (for core needle biopsies) unstained slides were obtained depending on the tumor

cellularity and tumor surface area. To optimize the extraction protocol, we tested a second extraction kit (Roche FFPET RNA Isolation Kit, used in Prosigna Breast Cancer Prognostic Gene Signature). Finally, Lesluyes and al. used the RNeasy FFPE kit (Qiagen) in 2016 for RNA extraction in 41 cases in cohort #2, but only 17 were analyzable by RNA-seq. These 17 tumors could be processed by RNA sequencing using RNAs from FFPE blocs. As previously described, we performed unstranded and rRNA-depleted (to overcome degraded polyA tails) sequencing with at least 20 million paired ends [1]. RNA-seq expressions are available on Gene Expression Omnibus under accession number GSE71121. RNA-seq raw files (FastQ) are available on Sequence Read Archive under accession numbers SRP057793 and SRP059588 (frozen tissues and FFPE blocs, respectively). RNA extracted from both fresh-frozen and FFPE tissues was assessed for concentration using Nanodrop 1000 (Thermo Scientific). To assess RNA quality, all samples were run on an Agilent Bioanalyzer 2100 (Agilent Technologies, Canada). The degree of RNA integrity was assessed using the smear analysis function in the Agilent 2100 Expert Software to measure the percentage of RNA molecules between 50bp and 300bp (DV_{50-300} metric). Data regarding patients' characteristics and tumor description were obtained by a retrospective review of the medical records. These records and histologic data were entered into a centralized database (https://conticabase.sarcomabcb.org). Ethical approval was obtained from the appropriate committees.

Tissue samples, RNA sample yield and quality

Bioanalyzer analysis of fresh-frozen tissue samples (n=124) showed a good RNA quality with a median RIN (RNA Integrity Number) of 8.18 (range, 3.3-9.9). Concerning FFPE blocks (n=47), data on RNA quality are presented in Supplementary Table 2. FFPE samples (surgical resections) were degraded with a median RIN of 2.24 (range, 1-2.6) and a median DV_{50-300} of 37.8% (range, 21-62%). FFPE samples (core needle biopsies) were

degraded with a median DV_{50-300} of 53% (range, 30-68%). This result was expected since FFPE samples were archival tissues. The FFPE samples used in this study were archived between 2009 and 2013.

NanoString CodeSet design and expression quantification

This technique has already been described by Geiss et al. [2]. Briefly, unique multiplexed probes were made with two sequence-specific probes complementary to a 100-base target region per target mRNA. The capture probe comprised a target-specific oligonucleotide coupled with a short sequence linked to biotin. The reporter probe consisted of a second 50mer target-specific oligonucleotide linked to a unique chain of dye-labelled segments for detection by the system. Our nCounter code set, named NanoCind®, consisted of a panel of 75 probes, including 67 distinct test probes derived from 67 distinct genes (Supplementary Table 1) and eight housekeeping genes for biological normalization purposes. Probe sets were designed and synthesized by nanoString. Additional information concerning probes can be found in the patent, which is filed under the number EP18305190.3. We used 100 ng of total RNA isolated from fresh-frozen tissues, as suggested by the manufacturer. FFPE tissues required a higher amount of total RNA, corresponding to "100 ng adjusted" (adjusted input) according to the DV_{50-300} value, i.e. the percentage of RNA fragments between 50-300 nucleotides. All FFPE cases were included by modifying the adjusted input. Cases were assessed in cartridges of 12 cases per run. The detailed protocol for mRNA transcript quantification, including sample preparation, hybridization, detection and scanning, followed the manufacturer's recommendations, using the nCounter PrepStation and nCounter Digital Analyzer. Data were extracted and validated using the nSolver v3.0 software.

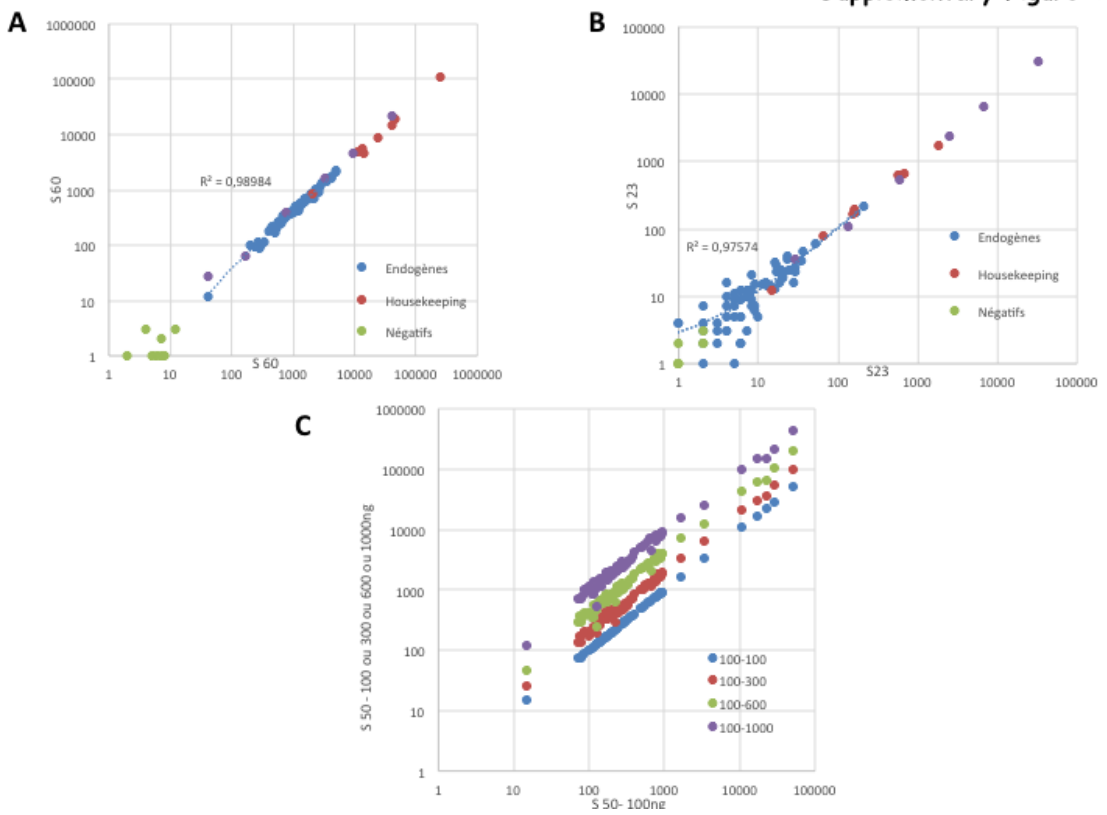
Statistical analysis

NanoString data were normalized using the manufacturer's software (nSolver v3.0) and recommendations. Background subtraction was performed on negative control probes and normalization with positive control probes and housekeeping genes was performed with the geometric mean method. Both RNA-seq and nanoString expression values were log2-transformed. As previously described, we used the nearest centroid method to determine CINSARC risk groups. To apply the signature independently on each sample, we performed leave-one-out cross-validation by computing distances from a sample to centroids (good and poor prognoses) of all other samples using the same technique [3]. Metastasis-free survival was analyzed by the Kaplan-Meier estimator from the original diagnosis to diagnosis of the metastatic event or the last follow-up. Significance was expressed by the Cox proportional hazards regression model implemented in the survival package (v2.41-3) available in R (v3.4.2).

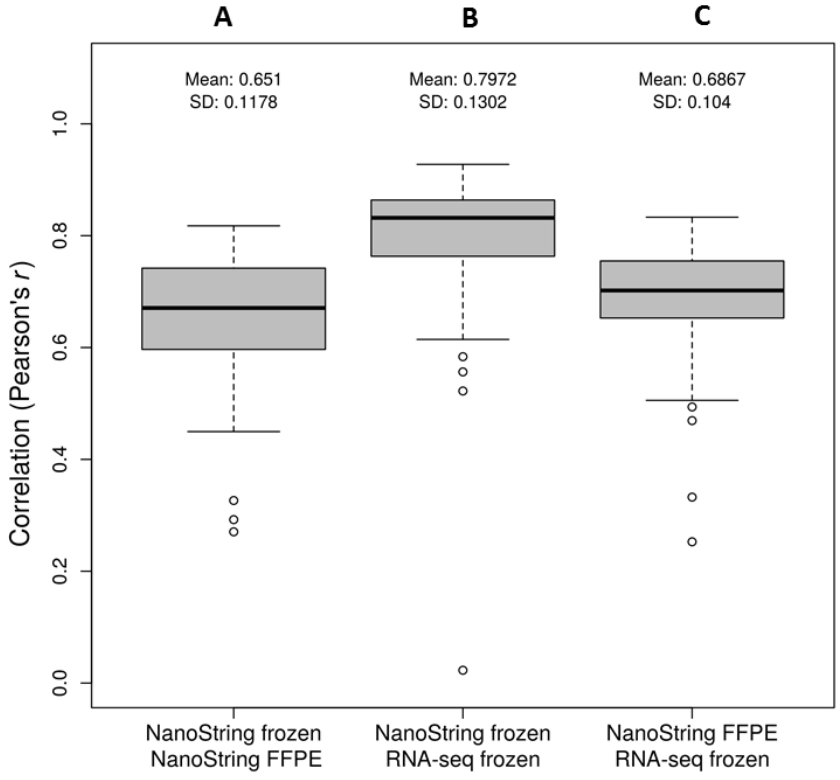
REFERENCES

1- Lesluyes T, Pérot G, Largeau MR, et al. RNA sequencing validation of the Complexity INdex in SARComas prognostic signature. Eur J Cancer 2016; 22:104-111.
2- Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, et al. (2008) Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol 26(3): 317–325.
3- Chibon F, Lagarde P, Salas S 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.

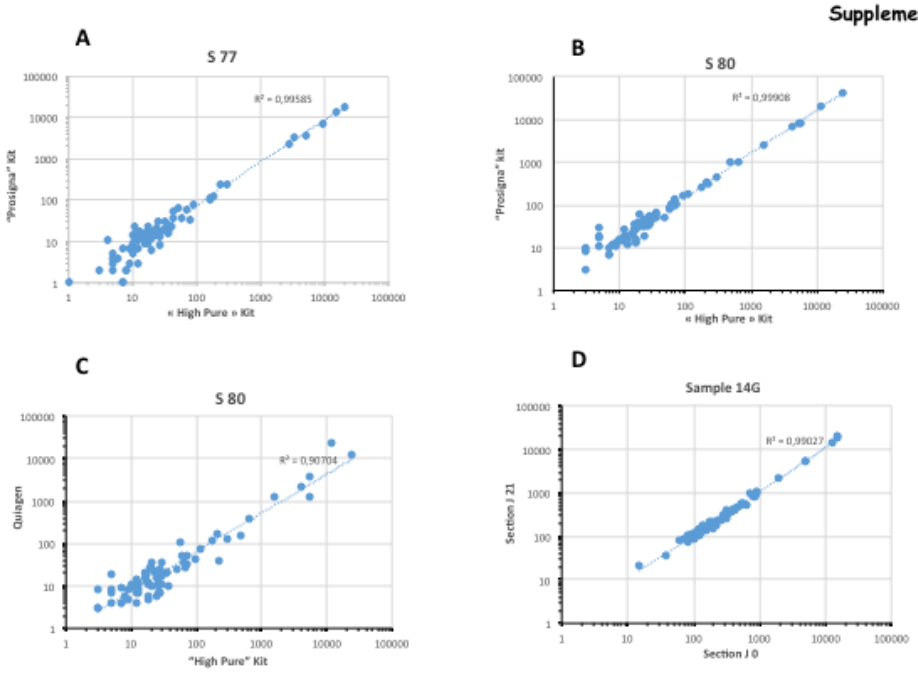
Supplementary Figure 1



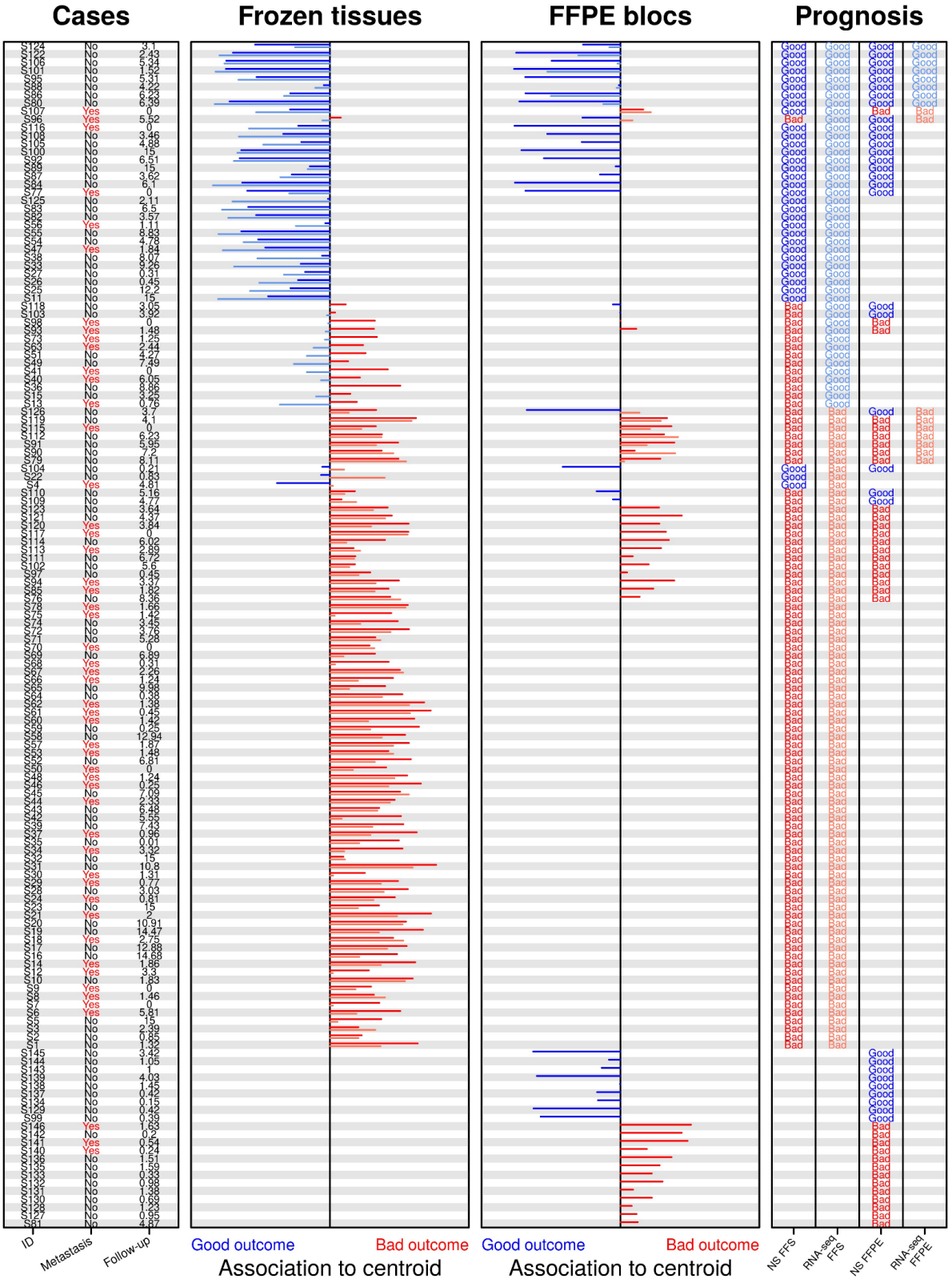
Supplementary Figure 2



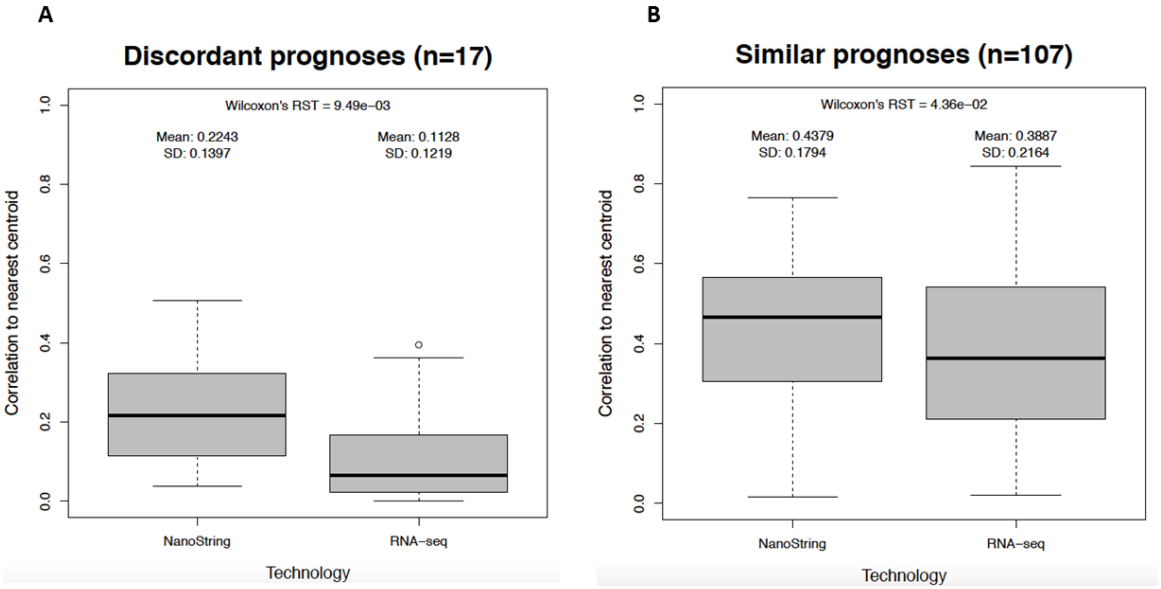
Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure 5



Supplementary Table 1. Location and cytoband of 67 genes of the CINSARC signature

Genes	Chromosome	Cytoband
ANLN	7	p14.2
ASPM	1	q31.3
AURKA	20	q13.2
AURKB	17	p13.1
BIRC5	17	q25.3
BORA	13	q22.1
BUB1	2	q13
BUB1B	15	q15.1
CCNA2	4	q27
CCNB1	5	q13.2
CCNB2	15	q22.2
CDC20	1	p34.2
CDC45	22	q11.21
CDC6	17	q21.2
CDC7	1	p22.2
CDCA2	8	p21.2
CDCA3	12	p13.31
CDCA8	1	p34.3
CDK1	10	q21.2
CENPA	2	p23.3
CENPE	4	q24
CENPL	1	q25.1
CEP55	10	q23.33
CHEK1	11	q24.2
CKAP5	11	p11.2
CKS2	9	q22.2
ECT2	3	q26.31
ESPL1	12	q13.13
FANCI	15	q26.1
FBXO5	6	q25.2
FOXM1	12	p13.33
H2AFX	11	q23.3
HP1BP3	1	p36.12
KIF11	10	q23.33
KIF14	1	q32.1
KIF15	3	p21.31
KIF18A	11	p14.1
KIF20A	5	q31.2
KIF23	15	q23
KIF2C	1	p34.1
KIF4A	X	q13.1
MAD2L1	4	q27
MCM2	3	q21.3
MCM7	7	q22.1
MELK	9	p13.2
NCAPH	2	q11.2
NDE1	16	p13.11
NEK2	1	q32.3
NUF2	1	q23.3
OIP5	15	q15.1
PBK	8	p21.1
PLK4	4	q28.2
PRC1	15	q26.1
PTTG1	5	q33.3
RAD51AP1	12	p13.32
RNASEH2A	19	p13.2
RRM2	2	p25.1
SGO2	2	q33.1
SMC2	9	q31.1
SPAG5	17	q11.2
SPC25	2	q31.1
TOP2A	17	q21.2
TPX2	20	q11.21
TRIP13	5	p15.33
TTK	6	q14.1
UBE2C	20	q13.12
ZWINT	10	q21.1

Supplementary Table 2. RIN, DO260/280, DO260/230 and DV_{50-300} of FFPE tumor samples

Identification number	Material (S/MB)	RIN	DO 260/280	DO 260/230	DV_{50-300}(%)
S120	FFPE (S)	2.3	1.85	1.7	29
S121	FFPE (S)	2.5	1.94	1.88	45
S122	FFPE (S)	2.3	1.88	1.7	35
S124	FFPE (S)	2.4	1.91	1.8	32
S126	FFPE (S)	2.3	1.92	1.83	40
S111	FFPE (S)	2.4	1.9	1.7	23
S76	FFPE (S)	2.3	1.88	2.01	41
S77	FFPE (S)	2.5	1.84	1.4	22
S79	FFPE (S)	2.4	1.95	1.97	46
S80	FFPE (S)	2.3	1.84	1.7	33
S81	FFPE (S)	2.5	1.93	2.04	33
S123	FFPE (S)	2.3	1.92	1.7	44
S84	FFPE (S)	1	1.75	1.06	54
S85	FFPE (S)	2.3	1.91	1.93	32
S86	FFPE (S)	1.7	1.7	0.9	41
S87	FFPE (S)	2.3	1.9	1.7	37
S88	FFPE (S)	1.3	1.9	1.5	48
S89	FFPE (S)	2.3	1.87	1.9	41
S90	FFPE (S)	2	1.73	1.1	49
S91	FFPE (S)	2.3	1.9	1.7	34
S92	FFPE (S)	1.8	1.9	1.5	22
S93	FFPE (S)	2.2	1.9	2.1	27
S94	FFPE (S)	2.1	1.9	2	40
S95	FFPE (S)	2.3	1.9	1.7	41
S96	FFPE (S)	2.1	1.8	1.3	39
S97	FFPE (S)	2.5	2	2	29
S98	FFPE (S)	2.3	1.9	1.7	33
S99	FFPE (S)	2.4	1.9	1.6	41
S100	FFPE (S)	2.2	1.8	1.2	40
S101	FFPE (S)	2.2	1.8	1.3	41
S102	FFPE (S)	2.4	2	2.1	40
S103	FFPE (S)	2.2	1.9	1.3	31
S104	FFPE (S)	2.3	2	2.1	54
S105	FFPE (S)	2.2	1.9	1.5	44
S106	FFPE (S)	2.6	1.8	1.3	42
S107	FFPE (S)	2.3	1.91	1.7	62
S108	FFPE (S)	2.4	1.6	0.9	48
S109	FFPE (S)	2.6	1.93	2.1	21

S110	FFPE (S)	2.4	1.92	2.02	21
S112	FFPE (S)	2.3	1.9	1.8	44
S113	FFPE (S)	2.1	1.8	1.5	23
S114	FFPE (S)	2.2	1.9	1.6	60
S115	FFPE (S)	2.2	1.91	2.07	31
S116	FFPE (S)	2.5	1.93	2.01	25
S117	FFPE (S)	2.3	1.91	1.84	51
S118	FFPE (S)	2.3	1.95	2.04	43
S119	FFPE (S)	2.4	1.29	1.92	24
S127	FFPE (MB)	1.1	1.48	0.7	61
S128	FFPE (MB)	2.1	1.9	1.2	68
S129	FFPE (MB)	NA	1.8	1.6	53
S130	FFPE (MB)	1	2	1.5	44
S131	FFPE (MB)	2.7	1.9	1.5	66
S132	FFPE (MB)	1.7	1.8	1.3	55
S133	FFPE (MB)	1.3	1.9	1.4	48
S139	FFPE (MB)	2.1	1.9	1.4	53
S134	FFPE (MB)	NA	1.8	0.6	30
S135	FFPE (MB)	1	1.9	1.5	41
S136	FFPE (MB)	1.4	1.9	1.7	54
S137	FFPE (MB)	1.3	1.8	1.3	44
S138	FFPE (MB)	NA	2	1.8	62
S140	FFPE (MB)	NA	1.9	1.4	44
S141	FFPE (MB)	NA	1.8	1.2	62
S142	FFPE (MB)	NA	1.8	1.3	57
S143	FFPE (MB)	NA	1.7	1.9	62
S144	FFPE (MB)	NA	1.6	0.8	48
S145	FFPE (MB)	NA	1.8	1	65
S146	FFPE (MB)	NA	1.9	1.4	43

Apport de l'étude

Ces résultats nous confortent dans la future application clinique routinière de CINSARC. La technologie proposée par NanoString, nCounter, nous a permis de définir un *CodeSet* représentatif de l'expression de cette signature : NanoCind. Cette technique procède ainsi, de façon conjointe, les ARN extraits de tissus congelés et de blocs FFPE. La qualité des ARN n'est alors plus un facteur limitant l'obtention de l'expression de la signature.

Ces résultats ont également démontré que CINSARC peut prédire le développement métastatique parmi les sarcomes de grade 2, des résultats similaires ayant déjà été rapportés (Chibon *et al.*, 2010). Ce résultat est intéressant d'un point de vue clinique car les tumeurs de grade 2 ont une agressivité intermédiaire comparativement aux tumeurs de bas grade (bon pronostic) et de haut grade (mauvais pronostic). Ces tumeurs sont nombreuses et leur prise en charge thérapeutique n'est pas standardisée, même si les recommandations de l'ESMO tendent à grouper les tumeurs de grade 2 avec celles de grade 3, ce qui réunirait plus de 80% des STMA dans une catégorie à haut potentiel métastatique (Casali *et al.*, 2018). Le reclassement des STMA de grade 2 selon leur potentielle agressivité constituerait une précieuse aide décisionnelle pour le traitement des patients (Charville & Lazar, 2018).

3.1.2.3 – Article 3 : *Chemotherapy in localized soft tissue sarcoma: will we soon have to treat grade 1 tumors? Update on CINSARC performances*

Valentin T, Lesluyes T, Le Guellec S & Chibon F. Lettre à l'éditeur publiée dans *Annals of Oncology* (2019).

Contexte

Une chimiothérapie néoadjuvante n'est recommandée que pour les tumeurs de hauts grades (2 et 3), profondes et de plus de 5cm. Par opposition, une chirurgie seule est recommandée pour les tumeurs de grade 1 qui bénéficient d'une résection R0 (marges chirurgicales saines à l'examen microscopique). La première branche décisionnelle thérapeutique recommandée pour les STMA non métastatiques et résécables est donc définie par le grade FNCLCC (**partie 1.1.5 page 24** ; Casali *et al.*, 2018). Le second critère majeur est la qualité de la chirurgie,

où les traitements seront plus importants en cas d'effraction tumorale (planifiée ou non).

Nous nous sommes alors demandé quel pourrait être l'apport pronostique de la signature CINSARC si, pour convenir aux pratiques actuelles et avant de devenir le premier critère d'évaluation de l'agressivité des STMA, elle s'insérait entre l'évaluation du grade et la qualité de la chirurgie. Nous avons donc réalisé une méta-analyse des cohortes pour lesquelles nous avions déterminé le statut CINSARC et pour lesquelles le grade FNCLCC était disponible.

Résumé de l'article

Nous avons démontré dans plusieurs études que CINSARC est un meilleur marqueur de l'agressivité des STMA que le grade et que la signature est capable de re-stratifier les tumeurs de grade intermédiaire en fonction du risque métastatique. Afin de statuer sur l'apport de la signature par rapport au standard actuel, en accord avec les pratiques cliniques, une étude exhaustive sur chacun des différents grades est alors nécessaire.

Nous avons constitué une cohorte de 605 tumeurs en compilant les données de plusieurs études précédemment réalisées. Les cas passés sur plusieurs technologies ont été filtrés de telle sorte que seule la méthodologie la plus récente soit conservée, à savoir : le système nCounter (priorité pour les cas FFPE puis aux tissus congelés ; Le Guellec *et al.*, 2018), puis le RNA-seq (tissus congelés uniquement ; Lesluyes *et al.*, 2016, incluant aussi la cohorte du TCGA sarcomes) et finalement les puces d'expression (tissus congelés uniquement ; Chibon *et al.*, 2010 ; Lagarde *et al.*, 2013). Ceci inclut par conséquent 128 cas passés sur puces NanoString (54 FFPE et 74 tissus congelés), 214 RNA-seq (19 de notre cohorte et 195 du TCGA) et 263 puces d'expression (220 SGC et 43 synovialosarcomes). Les valeurs pronostiques de CINSARC utilisées sont celles données par leurs études respectives.

Sur l'ensemble de la cohorte, le grade FNCLCC et CINSARC sont tous les deux des facteurs prédictifs significatifs du développement métastatique. Toutefois, une meilleure stratification est obtenue par CINSARC (P=5,8e-11) comparée à celle donnée par le grade FNCLCC (P=3,81e-4). Parmi l'ensemble des trois grades distincts, CINSARC permet d'identifier les tumeurs en fonction de leur potentiel métastatique (P=7,9e-4 pour le grade 1 ; P=1,78e-6 pour le grade 2 ; P=3,35e-3 pour le grade 3).

Il y a donc bien un intérêt majeur à utiliser la signature CINSARC, que ce soit en complément du grade FNCLCC, pour lequel chaque groupe est bien re-stratifié en fonction du risque métastatique, ou en remplacement de celui-ci.

Ces résultats sont inclus ci-après (Valentin *et al.*, 2019).

Letter to the editor

Chemotherapy in localized soft tissue sarcoma: will we soon have to treat grade 1 tumors? Update on CINSARC performances

The debate concerning perioperative chemotherapy use in localized soft tissue sarcoma (STS) is mired in a dead-end, following the results of a dozen of randomized studies completed by two meta-analyses. In the current guidelines [1], chemotherapy is not yet considered as a standard of care and may be proposed as an option in high-risk patients, currently imperfectly identified with classical factors as tumor size, depth and (Fédération Française des Centres de Lutte Contre le Cancer) FNCLCC grade [2]. This uncertainty may reflect the limits of these factors (in particular FNCLCC grade) for the precise identification of the high-risk patients' population.

When developed in 2010, CINSARC transcriptomic signature clearly appeared to be a promising way to overcome these limitations, as a tool able to identify high-risk STS patients independently of classical prognostic factors, in particular FNCLCC grade [3]. Initially incompatible with routine use (performed on frozen tumors analyzed by microarrays), CINSARC has been optimized [4, 5] is now fully evaluable with FFPE tumors using NanoString

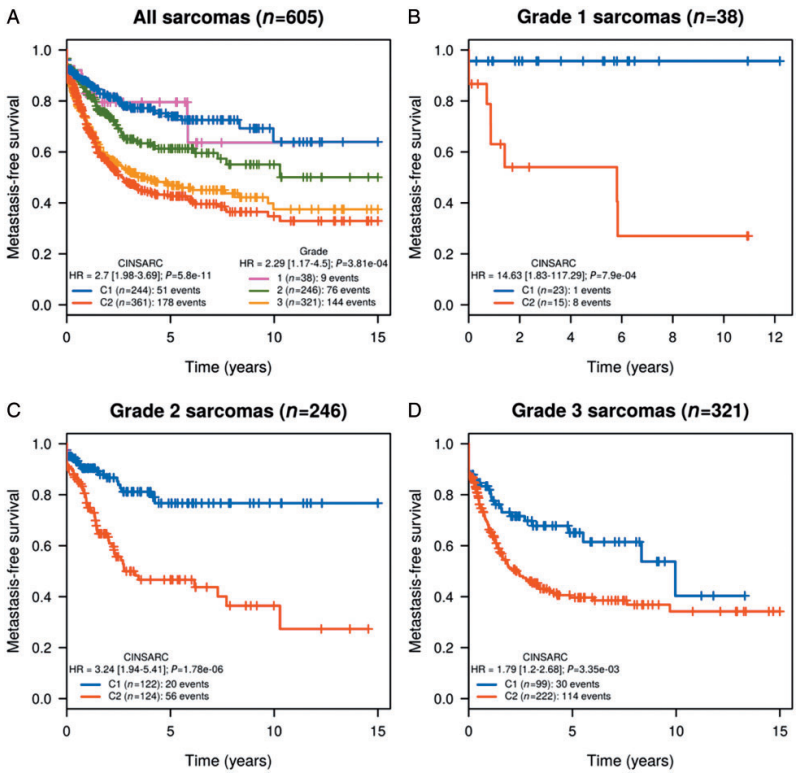


Figure 1. Metastasis Free Survival of patients with STS according to FNCLCC grade and CINSARC signature (A) Metastasis Free Survival of patients with FNCLCC grade 1 (B) grade 2 (C), or grade 3 (D) STS, according to CINSARC signature.

Annals of Oncology

technology (NanoCind®, patent number ≪EP18305190.3≫) [5]. This reproducible, low-cost array only needs small inputs of RNA, making perfectly feasible in core needle biopsies samples, gold standard for STS diagnosis.

Following its technical optimization, CINSARC performances showed a remarkable constancy over successive works. Since its first development, CINSARC signature has indeed been applied to more than 600 STS samples, including 367 samples from the three pivotal studies, 195 samples from TCGA program, and a series of 44 synovial sarcoma samples. The pooled results of these analyses confirm that CINSARC perfectly splits STS patients into two separate prognostic groups (in term of Metastasis Free Survival) more accurately than FNCLCC grade [HR 2.7 (1.98; 3.69) versus 2.29 (1.17; 5.5)] (Figure 1). More interesting, this meta-analysis allows to demonstrate that CINSARC identifies, in each FNCLCC grade, a subgroup of high-risk patients. This feature has already been proposed as a way to overcome the classical limitation of FNCLCC grade in grade 2 patients. However, this meta-analysis reveals that CINSARC also identifies in patients with grade 1 STS, typically considered as low-risk patients, high-risk patients carrying poor prognosis.

As many expert teams consider that high-risk STS patients are likely those supposed to take benefit from perioperative chemotherapy, we foresee CINSARC, able to identify them more accurately than the current histological evaluation, to be the key of this 'never-ending story'. Now applicable in routine settings, the next step is to evaluate the efficacy of perioperative chemotherapy in CINSARC high-risk STS patients. Several clinical trials based on CINSARC are about to begin with the French Sarcoma Group.

Currently, chemotherapy is only offered as an option to patient with large, grade 3 STS! In the near-future, will chemotherapy be given as a standard treatment to patient with grade 1, CINSARC high-risk STS? Unthinkable nowadays, isn't it?

T. Valentin[1,2*], T. Lesluyes[2,3,4], S. Le Guellec[2,5] & F. Chibon[2,5]
[1]Department of Medical Oncology, Institut Claudius Regaud, Toulouse; [2]INSERM U1037, Cancer Research Center of Toulouse (CRCT), Toulouse; [3]INSERM U1218, Institut Bergonié, Bordeaux; [4]University of Bordeaux, Bordeaux; [5]Department of Pathology, Institut Claudius Regaud, Toulouse, France
(*E-mail: valentin.thibaud@iuct-oncopole.fr)

Funding

None declared.

Disclosure

The authors have declared no conflicts of interest.

References

1. Casali PG, Abecassis N, Bauer S et al. Soft tissue and visceral sarcomas: ESMO–EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2018; 29(Suppl 4): iv51–iv67.
2. Coindre JM, Terrier P, Bui NB. Prognostic factors in adult patients with locally controlled soft tissue sarcoma. A study of 546 patients from the French Federation of Cancer Centers Sarcoma Group. J Clin Oncol 1996; 14(3): 869–877.
3. Chibon F, Lagarde P, Salas S et al. Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. Nat Med 2010; 16(7): 781–787.
4. Lesluyes T, Pérot G, Largeau MR et al. RNA sequencing validation of the Complexity INdex in SARComas prognostic signature. Eur J Cancer 2016; 57: 104–111.
5. Le Guellec S, Lesluyes, T, Sarot E et al. Validation of the Complexity INdex in SARComas prognostic signature on formalin-fixed, paraffin-embedded, soft tissue sarcomas. Ann Oncol 2018; 29(8):1828–1835.

doi:10.1093/annonc/mdy465
Published online 18 October 2018

3.1.2.4 – Conclusion

Historiquement limitée à l'évaluation de son expression sur puces d'expression avec du matériel congelé, nos récentes études ont étendu l'applicabilité de la signature CINSARC à de plus récentes technologies : le RNA-seq (Lesluyes *et al.*, 2016) puis le système nCounter (NanoString ; Le Guellec *et al.*, 2018). Cette dernière méthode est actuellement la plus à même d'analyser les extraits tumoraux issus de blocs FFPE afin de s'adapter aux pratiques courantes des laboratoires.

Ces résultats nous confortent dans l'applicabilité clinique de la signature, où elle pourrait, en première intention, seconder le grade FNCLCC pour orienter la décision thérapeutique des sarcomes localisés et résécables (Valentin *et al.*, 2019). La prochaine étape sera alors de tester la valeur de la signature, non plus dans un but rétrospectif mais en condition prospective au cours d'essais cliniques. À l'issue de ces essais et selon ses performances, CINSARC pourrait alors devenir le principal critère d'évaluation du risque métastatique dans les sarcomes.

3.1.3 – Une signature pronostique universelle ?

CINSARC a initialement été validée sur deux cohortes indépendantes de 183 et 127 SGC (Chibon *et al.*, 2010). Cette première étude a également démontré sa valeur pronostique dans une cohorte de 32 GIST (en survie sans métastase), deux cohortes de 78 et 295 cancers du sein (en survie sans métastase) et deux cohortes de 136 et 278 lymphomes diffus à larges cellules B (en survie globale). Deux études supplémentaires, sur 60 GIST (Lagarde *et al.*, 2012) et 58 synovialosarcomes (Lagarde *et al.*, 2013), sont venues par la suite démontrer son intérêt clinique dans les SGS (en survie sans métastase).

CINSARC est donc un marqueur de l'agressivité tumorale (en survie globale et/ou survie sans métastase) dans plusieurs types de cancers. Cette signature est composée de gènes impliqués dans la ségrégation des chromosomes et son expression est corrélée à un phénotype d'instabilité chromosomique. Celle-ci étant universellement rencontrée dans les cancers (même si son degré est variable selon les familles de cancers ; Carter *et al.*, 2012 ; Sansregret *et al.*, 2018), il est donc d'intérêt de regarder si CINSARC peut être un marqueur universel de l'agressivité tumorale.

3.1.3.1 – Article 4 : *The CINSARC signature as a prognostic marker for clinical outcome in multiple neoplasms*

Lesluyes T, Delespaul L, Coindre JM & Chibon F. Article publié dans *Scientific Reports* (2017).

Contexte

L'émergence et le développement des approches pan-cancers ont conduit à l'établissement d'outils et de bases de données regroupant plusieurs familles de cancers. Notamment, la base *PREdiction of Clinical Outcomes from Genomic profiles* (PRECOG) qui compile des données cliniques (en survie globale) et transcriptomiques de près de 18000 tumeurs représentant 39 types tumoraux (Gentles *et al.*, 2015). En interrogeant les données assemblées dans PRECOG, nous avons entrepris une large étude pan-cancer de la valeur pronostique de la signature

CINSARC pour déterminer, de façon exhaustive, si son expression est associée à un phénotype d'agressivité tumorale transversal à l'ensemble des cancers.

Résumé de l'article

Nous avons exploré la ressource PRECOG dans le but de mesurer la valeur pronostique de la signature CINSARC dans la majorité des familles de cancers (Gentles *et al.*, 2015). PRECOG compile les valeurs pronostiques de 23287 gènes dans 166 études, regroupées en 39 types de cancers. Ces valeurs pronostiques sont fournies brutes, pondérées par type de cancer et pondérées en analyse pan-cancer (comprenant alors tous les types tumoraux).

Nous avons d'abord analysé les données pondérées en analyse pan-cancer où nous avons mesuré l'enrichissement de CINSARC en gènes individuellement pronostiques, comparativement à 15499 autres signatures définies dans la GeneSigDB et la MSigDB ($\geq$25 gènes ; Culhane *et al.*, 2010 ; Liberzon *et al.*, 2015). Ces mesures montrent que CINSARC est dans le top 5 des signatures les plus enrichies en gènes individuellement pronostiques. L'analyse dans les différents types de cancers montre que CINSARC est significativement enrichie en gènes les mieux corrélés au devenir clinique dans 21 types différents et que toutes les autres signatures ne couvrent pas autant de familles de cancers.

Nous avons finalement étudié la valeur de CINSARC en analyse de survie dans 83 études ($\geq$ 50 patients, $\geq$ 25 gènes de CINSARC et entre 10 et 90% de taux de survie). Nous avons démontré que CINSARC était bien associée au devenir clinique des patients dans 33 études (40%), couvrant 17 types tumoraux. Plus les études comportent de patients et plus CINSARC avait une valeur pronostique significative, suggérant que de plus importantes tailles de cohortes aboutiraient à une évaluation pronostique plus robuste de la signature.

Ces résultats montrent que l'expression de CINSARC, comparativement à plusieurs milliers d'autres signatures transcriptomiques, est universellement associée à un devenir clinique péjoratif en étant pronostique dans de nombreux types de cancers.

L'ensemble des résultats de cette étude sont inclus ci-après (Lesluyes *et al.*, 2017). Les figures additionnelles 3 et 4 ainsi que la table additionnelle 1, trop volumineuses pour être incluses, sont accessibles en ligne (https://doi.org/10.1038/s41598-017-05726-x).

www.nature.com/scientificreports

OPEN

The CINSARC signature as a prognostic marker for clinical outcome in multiple neoplasms

Received: 20 March 2017
Accepted: 1 June 2017
Published online: 14 July 2017

Tom Lesluyes[1,2], Lucile Delespaul[1,2], Jean-Michel Coindre[2,3] & Frédéric Chibon[1,3]

We previously reported the CINSARC signature as a prognostic marker for metastatic events in soft tissue sarcomas, breast carcinomas and lymphomas through genomic instability, acting as a major factor for tumor aggressiveness. In this study, we used a published resource to investigate CINSARC enrichment in poor outcome-associated genes at pan-cancer level and in 39 cancer types. CINSARC outperformed more than 15,000 defined signatures (including cancer-related), being enriched in top-ranked poor outcome-associated genes of 21 cancer types, widest coverage reached among all tested signatures. Independently, this signature demonstrated significant survival differences between risk-groups in 33 published studies, representing 17 tumor types. As a consequence, we propose the CINSARC prognostication as a general marker for tumor aggressiveness to optimize the clinical managements of patients.

From the first report of gene expression quantification method by Schena *et al.* in 1995[1], to RNA sequencing (RNA-seq), extensively used nowadays by international consortia to decipher transcriptomic abnormalities[2,3], gene expression has become an essential tool in cancer research. For two decades, microarrays provided much information, both at gene and transcript levels, on various oncogenic factors[4–7]. Following on, the next-generation sequencing (NGS) permitted sequencing of RNA fragments, to a single base-pair resolution, where RNA abundance is directly related to the proportion of sequenced reads mapped to a given gene[8]. Moreover, dedicated RNA-seq algorithms allow obtaining genomic information such as point mutations, insertions/deletions, translocations and genomic integrations from foreign organisms: well-known oncogenic and tumor progression mechanisms[9–11]. However, standard measurements are yet to be defined since numerous software exist for RNA-seq processing, including many different gene expression normalization methods[12].

Such transcriptomic investigations generated a large amount of data. Consequently, databases were set up to standardize all information associated with expression matrices, notably: organism, platform identifier, normalization, unit measurement and study-specific information (*i.e.* cell types, treatments, time series, sampling and culture conditions, etc.). Among the many available gene expression databases, the two most used are Gene Expression Omnibus from the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/geo)[13] and ArrayExpress from the European Bioinformatics Institute (http://www.ebi.ac.uk/arrayexpress)[14]. These resources, by gathering results from microarrays and RNA-seq experiments, provide an easy access to millions of cancer-related transcriptomic profiles (cell lines, primary tumors and metastases/relapses).

Since Golub *et al.* identified a specific gene set capable of distinguishing acute myeloid leukemia from acute lymphoblastic leukemia in the late 90s[15], establishment of gene expression signature remains a key part of cancer research. This first study demonstrated the possibility to use gene expression as a new *in silico* classifier, whereas previous options were limited to clinical observations and immunohistochemistry experiments. Subsequently, multiple gene sets were defined, not only to differentiate entities, but also to try predicting disease evolution. Two publications in early 2000s demonstrated the usefulness of transcriptomic profile as a survival indicator in breast cancer by focusing on specific genes[16,17]. Few years later, a two-gene expression ratio was found to be a good predictor of tamoxifen response for breast cancer[18]. Then, inferring chromosomal instability from gene expression has become a promising predictor of clinical outcome in various cancers[19]. To date, the prognostic

[1]INSERM U1218, Institut Bergonié, F-33000, Bordeaux, France. [2]University of Bordeaux, F-33000, Bordeaux, France. [3]Department of Pathology, Institut Bergonié, F-33000, Bordeaux, France. Correspondence and requests for materials should be addressed to T.L. (email: t.lesluyes@bordeaux.unicancer.fr) or F.C. (email: f.chibon@bordeaux.unicancer.fr)

values of specific gene expression signatures have been demonstrated for breast cancer, lymphoma, leukemia, hepatocellular carcinoma, sarcoma, etc.[20–24].

In 2010, we defined a set of 67 genes as a predictor for metastatic events in sarcomas with complex genetics[24], with a better prognosis compared to the standard FNCLCC (*Fédération Nationale des Centres de Lutte Contre le Cancer*) grading system based on tumor differentiation, mitotic index and necrosis[25]. These genes were identified based on differential expression analyses with three different classifiers: FNCLCC grade, genomic alteration number and chromosomal instability signature[19]. Gene ontologies associated with these genes are mitosis and chromosome integrity pathways. This signature, called CINSARC (Complexity INdex in SARComas), was also informative in diffuse large B-cell lymphomas and breast carcinomas. We extended its application scope to sarcomas with simple genetics (harboring recurrent and specific genomic alteration): gastrointestinal stromal tumors and synovial sarcomas[26,27]. Venet *et al*. demonstrated that randomly generated gene sets could be better predictors compared to published signatures in breast cancers[28]. Based on this observation, we compared the prognostic value of CINSARC to 1000 equal-sized randomly generated gene sets on four independent sarcoma datasets. We reported that no randomly generated signature was a better predictor than CINSARC[29].

Recently, Gentles *et al*. reported an extensive data-mining analysis and compiled their results into a new resource named PRECOG (PREdiction of Clinical Outcomes from Genomic profiles; https://precog.stanford.edu)[30]. They evaluated the prognostic value of 23,287 genes across 39 cancer types from 166 published expression datasets, approximately representing 18,000 tumors. Experiments were performed on genes associated with good outcome, particularly genes involved in the immune system with a particular interest for the CIBERSORT (Cell type Identification By Estimating Relative Subsets Of known RNA Transcripts) approach[31]. Gentles *et al*. also observed that prognostic genes are significantly more shared by different tumor types than expected by chance, either associated with good or poor outcome. This highlights the possibility of using a non-cancer-specific signature as an outcome predictor, consisting of genes sharing same expression patterns across several tumor types.

Since we previously demonstrated that CINSARC is a significant prognostic factor in multiple malignancies, we analyzed the PRECOG resource and evaluated its prognostic ability in the available 39 cancer types. Furthermore, this signature has been compared to other gene sets defined by various methods (transcription factor regulated genes, cancer expression patterns, chromosomal positions, co-expression networks, specific immune processes, etc.). We address the following questions: is CINSARC a signature enriched in genes associated with poor outcome in a pan-cancer overview? Could CINSARC be applied to multiple cancer-specific studies such that, with additional validations, this signature could be tested in prospective studies to optimize the clinical management of patients?

Results

Signature enrichments at pan-cancer level. To obtain an exhaustive set of gene expression signatures, we used the MSigDB (Molecular Signatures DataBase; v5.2), including 18,026 gene sets, and the GeneSigDB (Gene Signature DataBase; v4.0), including 2,951 human gene sets[32,33]. CINSARC has already been included in GeneSigDB (named *SARCOMA CHIBON10 67GENES GENOMECOMPLEXITYPREDICTOR*), but lacks several genes (62 instead of 67). Consequently, we included the full signature named *SARCOMA CHIBON10 67GENES GENOMECOMPLEXITYPREDICTOR CURATED*. We performed the popular Gene Set Enrichment Analysis (GSEA) method to determine signature enrichments in genes significantly associated to poor outcome (see methods)[34]. As recommended by GSEA documentation, we removed signatures with less than 25 known genes in PRECOG database to avoid inflated scorings for very small gene sets. This filtered out 4,394 (24%) and 1,085 (37%) signatures from MSigDB and GeneSigDB, respectively. Of note, *BORA* (formerly known as *C13orf34*) was not evaluated in PRECOG and is therefore missing in both full (curated) and reduced (62 genes) CINSARC.

In the database of 15,499 gene sets, 1,273 (8%) were significantly enriched in genes associated with poor prognosis (family-wise error rate $P < 0.05$; Supplementary Table 1: sheet 1). Sorted by decreasing normalized enrichment scores (NES; primary statistic for examining gene set enrichment results; see methods), reduced and full CINSARC ranked 9 and 17, respectively (Figs 1 and 2). In addition to a high NES, leading-edges metrics (extension of the NES results; see methods) showed high sensitivity (tags are 97% and 95%, ranked 4 and 11 for reduced and full CINSARC, respectively), low false-negative rate (lists are 2% and 3%, ranked 5 and 25 for reduced and full CINSARC, respectively) and high enrichment signal strength (signals are 99% and 98%, ranked 11 and 15 for reduced and full CINSARC, respectively). All these four metrics combined (NES and the three leading-edge metrics), other signatures in the top 20 performed worst compared to CINSARC with a mean sensitivity of 76% (range: 65–97%), a mean false-negative rate of 7% (range: 2–13%) and a mean enrichment signal strength of 81% (range: 69–99%). Scored by the mean ranks of the four metrics, CINSARC obtained scores of 7.25 and 17, corresponding to ranks 1 and 4 for reduced and full CINSARC, respectively (Fig. 2). Of note, the top 10 consisted of another breast cancer signature defined by Reyal *et al.* (ranked 3 with a score of 15.5)[35]. Other signatures in the top 10 were co-expression networks of the following genes: *CCNA2*, *HMMR*, *CCNB2*, *CDC20*, *CDC2*, *CENPF* and *RRM2*. Taken together, the top 10 enriched signatures are composed of a limited number of 92 genes strongly involved in mitosis and chromosome segregation pathways ($P < 10^{-30}$; see methods). This result demonstrates that CINSARC is strongly enriched in genes associated with poor prognosis at pan-cancer level, with a high sensitivity and low false-negative rate compared to other signatures.

Signature enrichments at cancer type level. In PRECOG, many cancer types have few outcome-associated genes (*e.g.* at $Q < 0.05$). Fifteen cancer types have no such genes and seven cancer types have from 1 to 9 such genes, for a total of 22 cancer types with less than 10 outcome-associated genes. We wondered how signature enrichments perform in cancer types with low versus high outcome-associated gene contents. We split the 39 cancer types into two subgroups using a threshold of 10 outcome-associated genes: 17 types have at least 10 such genes and 22 have less.

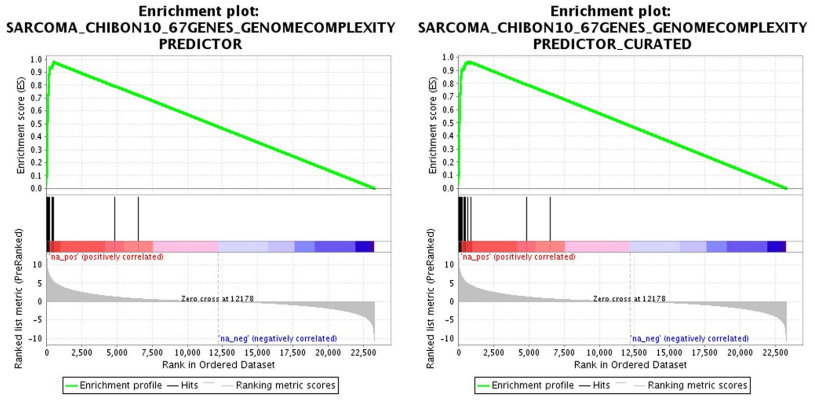


Figure 1. Enrichment plots of the reduced (left) and full (right) CINSARC signatures. Genes are ranked according to their individual prognosis given in PRECOG (poor and good prognoses in the left and right sides, respectively), represented by red and blue segments. Across these segments, CINSARC genes are marked by vertical black lines. The enrichment score (green plot) corresponds to a running-sum statistics: for each gene, if part of the signature, a positive value is added and a negative one otherwise.

In 10 out of the latter 22 types (45%), CINSARC (both full and reduced) was significantly enriched in top-ranked genes (though these genes are not significantly associated with poor prognosis), the widest type coverage reached by all tested signatures (Supplementary Table 1: sheets 2 to 23). Two other cancer signatures performed similar to CINSARC: *BREAST REYAL08 72GENES* and *ROSTY CERVICAL CANCER PROLIFERATION CLUSTER*. Furthermore, eight co-expression networks in MSigDB performed similar to CINSARC: *CCNA2*, *CCNB2*, *CDC20*, *CENPF*, *MCM4*, *PCNA*, *RRM1* and *RRM2*. These 12 signatures are predominantly enriched in the same tumor types: adrenocortical cancer, medulloblastoma, gastric cancer, Burkitt's lymphoma, liver cancer, primary liver cancer, melanoma, mesothelioma, pancreatic cancer and Ewing's sarcoma. The lack of prognosis-associated genes in these 22 types is also measured by different indicators. They present less enriched signatures, less gene ontology enrichments and less protein-protein interactions (Wilcoxon rank sum tests are 1.71×10^{-4}, 1.03×10^{-2} and 7.28×10^{-5}, respectively; Supplementary Figure 1; see methods) compared to the 17 others (those with at least 10 outcome-associated genes).

In these 17 cancer types (with at least 10 outcome-associated genes), CINSARC (both full and reduced) was significantly enriched in top-ranked genes (significantly associated with poor prognosis) of 11 types (65%), the widest type coverage reached by all tested signatures (Supplementary Table 1: sheets 24 to 40). Several other signatures performed equally: BREAST THORNER09 217GENES, *BREAST TROESTER06 134GENES UP DOXTREATED-HME-CC*, *BREAST TROESTER06 81GENES UP SHAM HME-CC*, *CAIRO HEPATOBLASTOMA CLASSES UP*, *GSE18893 TCONV VS TREG 24H TNF STIM UP*, *NAKAYAMA SOFT TISSUE TUMORS PCA2 UP*, *RODRIGUES THYROID CARCINOMA POORLY DIFFERENTIATED UP*, STEMCELL SHATS10 100GENES CONSENSUS STEMNESS RANKING, *WHITEFORD PEDIATRIC CANCER MARKERS* and *ZHOU CELL CYCLE GENES IN IR RESPONSE 24HR*. These 12 signatures are predominantly enriched in the same tumor types: bladder cancer, astrocytoma, glioma, neuroblastoma, breast cancer, germ cell tumor, Mantle cell lymphoma, multiple myeloma, lung adenocarcinoma, ovarian cancer and prostate cancer.

Considering the full set of 39 tumor types, CINSARC covers the largest possible spectrum of tested cancer types, being enriched in top-ranked genes of 21 types: adrenocortical cancer, medulloblastoma, gastric cancer, Burkitt's lymphoma, liver cancer, primary liver cancer, melanoma, mesothelioma, pancreatic cancer, Ewing's sarcoma, bladder cancer, astrocytoma, glioma, neuroblastoma, breast cancer, germ cell tumor, Mantle cell lymphoma, multiple myeloma, lung adenocarcinoma, ovarian cancer and prostate cancer. Aggregated into groups (brain, blood, head and neck and solid), no significant association was observed between CINSARC enrichments and cancer families. Other 18 types, without CINSARC enrichments, are: glioblastoma, colon cancer, acute myeloid leukemia, chronic lymphoid leukemia, diffuse large B-cell lymphoma, melanoma metastasis, meningioma, head and neck cancer, hypopharyngeal cancer, oesophageal cancer, oral squamous cell carcinoma, B-cell acute lymphoblastic leukemia, follicular lymphoma, kidney cancer, large cell carcinoma of the lung, squamous cell carcinoma of the lung, small cell lung cancer and osteosarcoma. As a consequence of these results, we pursued experiments by focusing on the CINSARC signature.

CINSARC prognostic value with survival analysis. Being enriched in genes associated with poor outcome does not essentially make a signature a prognostic factor. This can be explained by the fact that collective information of each individual gene prognostic value may be different from the prognostic value given by a whole set of genes. Accordingly, we investigated whether CINSARC was associated with survival differences in

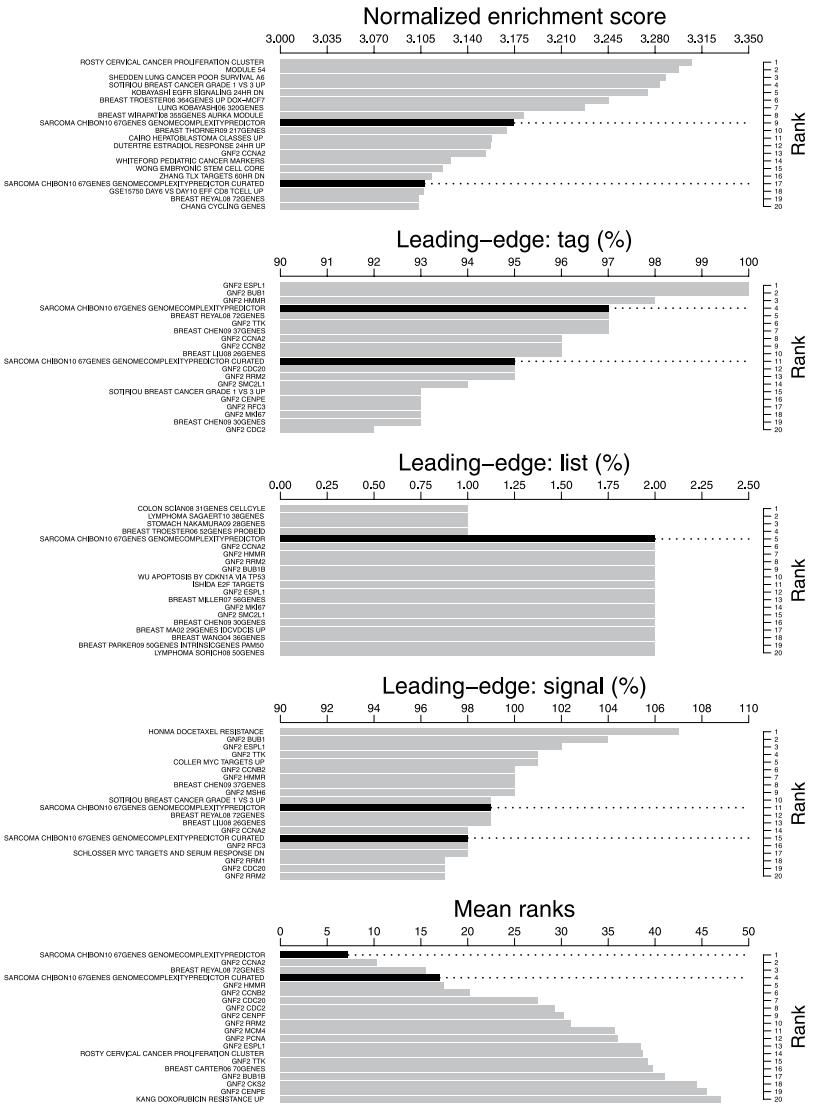


Figure 2. Pan-cancer overview of the top 20 signatures by several metrics (from top to bottom): normalized enrichment score, tag (sensitivity), list (false-negative rate), signal (enrichment strength) and the mean ranks of the previous metrics. Histograms are sorted to display best to worst ranks from top to bottom, respectively. CINSARC signatures are highlighted in black.

studies used to build the PRECOG resource. To obtain a robust evaluation, we filtered out datasets with less than 50 annotated cases, less than 25 CINSARC genes and <10% or >90% death rates. Accordingly, 83 datasets (out of 166; 50%) were investigated using the Kaplan-Meier estimator, covering 27 different cancer types. Since the investigated datasets did not permit generation of training-validation cohorts (median number of samples was 111), we performed leave-one-out cross-validation as the classifier method.

A total of 33 datasets (40%) produced significant survival differences according to the CINSARC classification (Supplementary Table 2, Supplementary Figures 2 and 3). The 17 cancer types associated with these 33 datasets

are: neuroblastoma, breast cancer, Mantle cell lymphoma, diffuse large B-cell lymphoma, ovarian cancer, prostate cancer, acute myeloid leukemia, lung adenocarcinoma, bladder cancer, liver cancer, multiple myeloma, follicular lymphoma, colon cancer, squamous cell carcinoma of the lung, oral squamous cell carcinoma, chronic lymphoid leukemia and B-cell acute lymphoblastic leukemia. Importantly, cancer types where CINSARC was significantly enriched in top-ranked genes demonstrated higher proportion of significant survival differences (22 among 42 datasets; 52%) compared to the others (11 among 41; 27%). Also, cancer types with at least 10 genes associated with poor prognosis demonstrated higher proportion of significant survival differences (25 among 56 datasets; 45%) compared to the others (8 among 27; 30%).

Though no correlation has been measured between survival difference and microarray platform or the number of evaluated CINSARC genes, clear evidence of cohort size effect has been observed: the larger cohort, the higher survival difference (Wilcoxon rank sum test $P = 1.7e\text{-}5$). Among the 83 tested datasets, 46 contain at least 100 tumors and CINSARC classification is significantly associated with the outcome in 26 of them (57%). There are 30 datasets with at least 150 tumors and CINSARC classification is significantly associated with the outcome in 19 of them (63%). At 200 tumors threshold, CINSARC classification is significantly associated with the outcome in 11 out of 16 (69%) datasets.

In these 11 datasets, we wondered how would perform CINSARC if less tumors were taken into account. In fact, we previously reported CINSARC robustness as it was a clinical marker in independent sarcoma datasets[29] but we did not evaluate its robustness in term of cohort subsampling[36]. To firmly evaluate CINSARC robustness regarding cohort size, we investigated the 11 datasets with a least 200 tumors where CINSARC is a prognostic factor. For each dataset, we randomly generated 10,000 sub-cohorts at 75% and 50% of the total cohort size with similar death rates (±10%; Supplementary Figure 4). All of the 11 datasets demonstrate that better prognoses are obtained with 75% compared to 50% subsampling. In eight datasets, CINSARC still classify tumors according to their aggressiveness in most of random trials (>50% at $P < 0.05$). However, due to the decrease in statistical power, CINSARC mostly produces non-significant prognoses in three datasets (PMID: 17410195, 21720365 and 17023574, cancer types are: B-cell acute lymphoblastic leukemia, multiple myeloma and ovarian tumors, respectively). These results confirm that cohort sizes affect prognosis values and that, nonetheless, CINSARC is a robust signature as it remains a prognostic marker for tumor aggressiveness in subsampling analyses.

Discussion

Cancer gene expression signatures have been widely used as tools to classify tumors into specific subtypes, as a prognostic factor for clinical outcomes or as a predictive factor for treatment responses[37]. Using the GSEA algorithm, we demonstrated that CINSARC, among 15,499 signatures established by various methods, is strongly enriched in prognostic genes at pan-cancer level, with high sensitivity (>95%) and low false-negative rate (<3%). At tumor-specific level, CINSARC covers the largest possible spectrum of tumor types among all tested cancer gene expression signatures (enriched in 21 cancers types among 39 tested). As it was not evaluated in PRECOG, we hypothesize that *BORA* gene would increase CINSARC prognostic value since it has been described as an important Aurora kinase A (*AURKA*) activator, required for centrosome maturation, spindle assembly and asymmetric protein localization during mitosis[38]. Moreover, *AURKA* alone has been identified as a prognostic marker in gastrointestinal stromal tumors[26]. Originally, CINSARC is a metastatic marker to predict which sarcomas with complex genetics have high relapse risk (on average 50% at 5-year)[24]. The results we present demonstrate its usefulness as a generic tumor aggressiveness predictor since overall/disease-specific survivals were considered in PRECOG.

Importantly, several PRECOG factors could restrict signature prognostic values, CINSARC included, in the different cancer types. First: it is questionable that a single tumor dataset represents the classical cancer-specific transcriptomic profile and therefore another dataset could produce different results. Second: some studies comprise very high or very low (typically > 90% or < 10%, respectively) death rates which could be a limiting factor in PRECOG survival evaluation since cohorts are split by median gene expression. Third: several cancer types, due to microarray models used in the different studies, may lack important genes so prognostic values were limited to a restricted number of genes (*i.e.* GPL80: Affymetrix Hu6800, GPL91: Affymetrix HG-U95A, non-commercial: GPL257 and GPL3906, etc.). Fourth: it has been described that cohort size is a parameter affecting survival estimation[39]. To bypass these limitations and confirm GSEA results, we performed survival analyses using CINSARC classification. Most datasets (>57%) with at least 100 samples presented a significantly survival difference whereas smaller datasets performed worst. This can be explained as statistical power depends on sample size so the larger dataset, the more statistical power. Moreover, the classification method (leave-one-out cross-validation) computes more robust gene expression patterns with large datasets. Finally, we assessed CINSARC robustness in terms of cohort subsampling and, previously reported, dataset independence[29].

Of note, 483 signatures (out of 15,499; 3%) were investigated at pan-cancer level with unexpected small intersections (<90%) with known genes in PRECOG, corresponding to 10,004 unique genes not reported in PRECOG. A large proportion is punctual: 92% of them only appear in maximum two signatures. As for the recurring genes (806), majority are uncharacterized and/or non-coding genes: 414 are genes predicted by open reading frame analyses, 92 are genes with unknown functions and unidentified orthologs, 60 are L and R ribosomal protein pseudogenes and 28 are long intergenic non-protein coding RNA. Considering the recurrent and curated genes, only 212 of them were not evaluated in PRECOG resource, suggesting high coverage of annotated genes.

We thus demonstrated that CINSARC is enriched in genes having a significant impact on tumor aggressiveness and that this signature can be used as a prognostic marker in multiple malignancies. This highlights the point that mitosis and chromosome segregations are two key pathways leading tumor aggressiveness through genomic instability, a well-known hallmark mechanism[40]. Part of the CINSARC signature, it has been observed that overexpression of several individual genes disrupts chromosomal segregation, generating genomic alterations and tumors in mice (*MAD2L2*, *BUB1*, *CCNB1*, *CCNB2* and *ESPL1*)[41–44]. A recent study highlighted an

aberrant persistence of several CINSARC proteins beyond mitosis in tetraploid versus diploid cells from sarcoma cell lines[45]. This was correlated to high motility (migration and invasion) capabilities measured in tetraploid versus diploid cells, whereas proliferation levels were similar. Though clear evidence of anaphase-promoting complex (APC) deficiency has to be established, a question arises about the impact of this protein persistence and the clinical prognostic interest of this signature.

Recently, we granted CINSARC a better clinical applicability with two major improvements[46]. First: the ability to perform RNA-seq rather than microarray technique, overcoming probe selection as probes can be transcript-specific and could not reflect full gene expression. Second: the ability to analyze FFPE (Formalin-Fixed, Paraffin-Embedded) tumor, daily material used by pathologists as opposed to frozen tissues previously required. As a consequence, two ongoing French and Europe-wide clinical trials will prospectively test the CINSARC signature in various sarcomas. Our results increase its application scope, demonstrating that this signature can be considered as a poor outcome predictor in multiple human malignancies. Finally, though it remains a very informative resource, we highlighted some limitations of PRECOG. Consequently, this study incites additional validations on other independent cancer-specific datasets.

Methods

Gene Set Enrichment Analysis (GSEA) usage. To evaluate the prognostic value of different gene sets as the collective information of individual genes, we used the *GSEAPreranked* tool from GSEA (Gene Set Enrichment Analysis; v2.2.3) algorithm[34]. The algorithm ranks genes based on the available PRECOG *z*-scores, which represents the significance of the association of a given gene to the prognosis of a given cancer type. All combined cancer types were summarized into meta-*z*-scores also available in PRECOG. Based on gene ranks, the algorithm computes a running-sum statistics to determine the enrichment score of each gene set. The enrichment scores are then normalized to take into account different gene set sizes, giving the normalized enrichment score (NES), and significance levels were determined with the family-wise error rate, a more conservative method than the false-discovery rate. Signatures with $P < 0.05$ (family-wise error rate) were considered significantly enriched.

Three leading-edges statistics are also computed. First: tag corresponds to the percentage of genes in a given signature hit before the maximum enrichment score is attained. By analogy, it represents the sensitivity since a high tag value indicates a higher proportion of the signature participating in the core enrichment. Second: list corresponds to the percentage of total genes before the maximum enrichment score. By analogy, it represents the false-negative rate (*e.g.* 1-specificity) since a low list value indicates a high gene set enrichment purity. Third: signal corresponds to the enrichment signal strength combining the two previous metrics. A high signal value indicates a top-ranked narrow enrichment.

Other statistical methods. Biological pathways were determined using GOseq (v1.26.0) package from Bioconductor on the top 200 genes ranked by decreasing meta-*z*-scores or *z*-scores[47]. *P* are given by the Wallenius noncentral hypergeometric distribution method and adjusted with Benjamini-Hochberg correction for multiple comparisons.

Protein-protein interactions were determined using expected/existing interaction ratios from STRINGdb (1.14.0) package from Bioconductor with database version 10 and score threshold set to 400 (removes low-confidence interactions) on the top 200 genes ranked by decreasing meta-*z*-scores or *z*-scores. The lower expected/existing interaction ratios, the higher are the number of protein-protein interactions[48].

The leave-one-out cross-validation classification was performed as follows: each sample was categorized using the nearest centroid method according to outcomes and transcriptomic profiles of all other samples.

Survival analyses were measured using survival (v2.40-1) package with the Kaplan-Meier estimator, where significance levels are given by the log-rank test (*P*) and survival differences given by hazard ratios (HR).

Miscellaneous computations, filters, statistics and plots were performed with R (3.3.2).

References

1. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
2. Zhang, J. *et al.* International Cancer Genome Consortium Data Portal–a one-stop shop for cancer genomics data. *Database* (Oxford) 2011, bar026 (2011).
3. Cancer Genome Atlas Research Network. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
4. DeRisi, J. *et al.* Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* **14**, 457–460 (1996).
5. Schuldiner, O. & Benvenisty, N. A. DNA microarray screen for genes involved in c-MYC and N-MYC oncogenesis in human tumors. *Oncogene* **20**, 4984–4994 (2001).
6. Khan, J. *et al.* cDNA microarrays detect activation of a myogenic transcription program by the PAX3-FKHR fusion oncogene. *Proc. Natl. Acad. Sci. USA* **96**, 13264–13269 (1999).
7. Moch, H. *et al.* High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal cell carcinoma. *Am. J. Pathol.* **154**, 981–986 (1999).
8. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
9. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
10. Maher, C. A. *et al.* Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc. Natl. Acad. Sci. USA* **106**, 12353–12358 (2009).
11. Sung, W.-K. *et al.* Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat. Genet.* **44**, 765–769 (2012).
12. SEQC/MAQC-III Consortium. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* **32**, 903–914 (2014).
13. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
14. Brazma, A. *et al.* ArrayExpress–a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* **31**, 68–71 (2003).

15. Golub, T. R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537 (1999).
16. van't Veer, L. J. *et al.* Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536 (2002).
17. van de Vijver, M. J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
18. Ma, X.-J. *et al.* A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* **5**, 607–616 (2004).
19. Carter, S. L., Eklund, A. C., Kohane, I. S., Harris, L. N. & Szallasi, Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat. Genet.* **38**, 1043–1048 (2006).
20. Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
21. Lossos, I. S. *et al.* Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N. Engl. J. Med.* **350**, 1828–1837 (2004).
22. Valk, P. J. M. *et al.* Prognostically useful gene-expression profiles in acute myeloid leukemia. *N. Engl. J. Med.* **350**, 1617–1628 (2004).
23. Hoshida, Y. *et al.* Gene expression in fixed tissues and outcome in hepatocellular carcinoma. *N. Engl. J. Med.* **359**, 1995–2004 (2008).
24. Chibon, F. *et al.* Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. *Nat. Med.* **16**, 781–787 (2010).
25. Trojani, M. *et al.* Soft-tissue sarcomas of adults; study of pathological prognostic variables and definition of a histopathological grading system. *Int. J. Cancer* **33**, 37–42 (1984).
26. Lagarde, P. *et al.* Mitotic checkpoints and chromosome instability are strong predictors of clinical outcome in gastrointestinal stromal tumors. *Clin. Cancer Res.* **18**, 826–838 (2012).
27. Lagarde, P. *et al.* Chromosome instability accounts for reverse metastatic outcomes of pediatric and adult synovial sarcomas. *J. Clin. Oncol.* **31**, 608–615 (2013).
28. Venet, D., Dumont, J. E. & Detours, V. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput. Biol.* **7**, e1002240 (2011).
29. Brulard, C. & Chibon, F. Robust gene expression signature is not merely a significant P value. *Eur. J. Cancer* **49**, 2771–2773 (2013).
30. Gentles, A. J. *et al.* The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat. Med.* **21**, 938–945 (2015).
31. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
32. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2015).
33. Culhane, A. C. *et al.* GeneSigDB–a curated database of gene expression signatures. *Nucleic Acids Res.* **38**, D716–725 (2010).
34. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
35. Reyal, F. *et al.* A comprehensive analysis of prognostic signatures reveals the high predictive capacity of the proliferation, immune response and RNA splicing modules in breast cancer. *Breast Cancer Res.* **10**, R93 (2008).
36. Li, J. *et al.* Identification of high-quality cancer prognostic markers and metastasis network modules. *Nat Commun* **1**, 34 (2010).
37. Chibon, F. Cancer gene expression signatures - the rise and fall? *Eur. J. Cancer* **49**, 2000–2009 (2013).
38. Hutterer, A. *et al.* Mitotic activation of the kinase Aurora-A requires its binding partner Bora. *Dev. Cell* **11**, 147–157 (2006).
39. Machin, D., Campbell, M. J., Tan, S. B. & Tan, S. H. Sample Size Tables for Clinical Studies. (Wiley-Blackwell, 2008).
40. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
41. Sotillo, R. *et al.* Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. *Cancer Cell* **11**, 9–23 (2007).
42. Ricke, R. M., Jeganathan, K. B. & van Deursen, J. M. Bub1 overexpression induces aneuploidy and tumor formation through Aurora B kinase hyperactivation. *J. Cell Biol.* **193**, 1049–1064 (2011).
43. Nam, H.-J. & van Deursen, J. M. Cyclin B2 and p53 control proper timing of centrosome separation. *Nat. Cell Biol.* **16**, 538–549 (2014).
44. Mukherjee, M. *et al.* MMTV-Espl1 transgenic mice develop aneuploid, estrogen receptor alpha (ERα)-positive mammary adenocarcinomas. *Oncogene* **33**, 5511–5522 (2014).
45. Jemaà, M. *et al.* Heterogeneity in sarcoma cell lines reveals enhanced motility of tetraploid versus diploid cells. *Oncotarget* doi:10.18632/oncotarget.14291 (2016).
46. Lesluyes, T. *et al.* RNA sequencing validation of the Complexity INdex in SARComas prognostic signature. *Eur. J. Cancer* **57**, 104–111 (2016).
47. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
48. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* doi:10.1093/nar/gkw937 (2016).

Acknowledgements

Computer time for this study was provided by the computing facilities MCIA (Mésocentre de Calcul Intensif Aquitain) of the Université de Bordeaux and of the Université de Pau et des Pays de l'Adour. The authors thank Dr. Ravi Nookala for the medical writing service. The authors thank Candice Merle for critically reading the manuscript. This work was supported by *Fondation ARC* and euroSARC (FP7-HEALTH-2011).

Author Contributions

F.C. supervised the study. T.L. and F.C. designed experiments. T.L. acquired the data and performed experiments. T.L. and F.C. wrote the manuscript. L.D. and J.M.C. reviewed experiments and the manuscript. All authors read and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-05726-x

Competing Interests: The authors declare that they have no competing interests.

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

Supplementary legends

Supplementary table 1: Significantly enriched signatures (family-wise error rate $P<0.05$) at pan-cancer level (sheet 1) and for each tumor type (sheets 2 to 40), sorted by mean ranks order. NES refers to Normalized Enrichment Score.

Supplementary table 2: Survival analyses on 83 published studies, ranked by significance levels. PMID refers to PubMed Identifier. *P* refers to nominal log-rank test significance. HR refers to Hazard Ratio. HR CI refers to Hazard Ratio Confidence Interval.

Supplementary figure 1: Differences between low and high (dark grey and light grey, respectively) outcome-associated gene contents in the different cancer types. Tested metrics are (from left to right): the number of significantly enriched signatures, the number of enriched gene ontologies and the number of protein-protein interactions (measured by expected/existing ratios; the lower ratio, the higher interactions).

Supplementary figure 2: Overview of survival information in the 33 published studies with significant survival differences according to the CINSARC signature. Low and high risk-groups are represented in blue and red colors, respectively. Complete details of these survival analyses can be found in supplementary figure 3. *P* refers to nominal log-rank test significance. HR refers to Hazard Ratio.

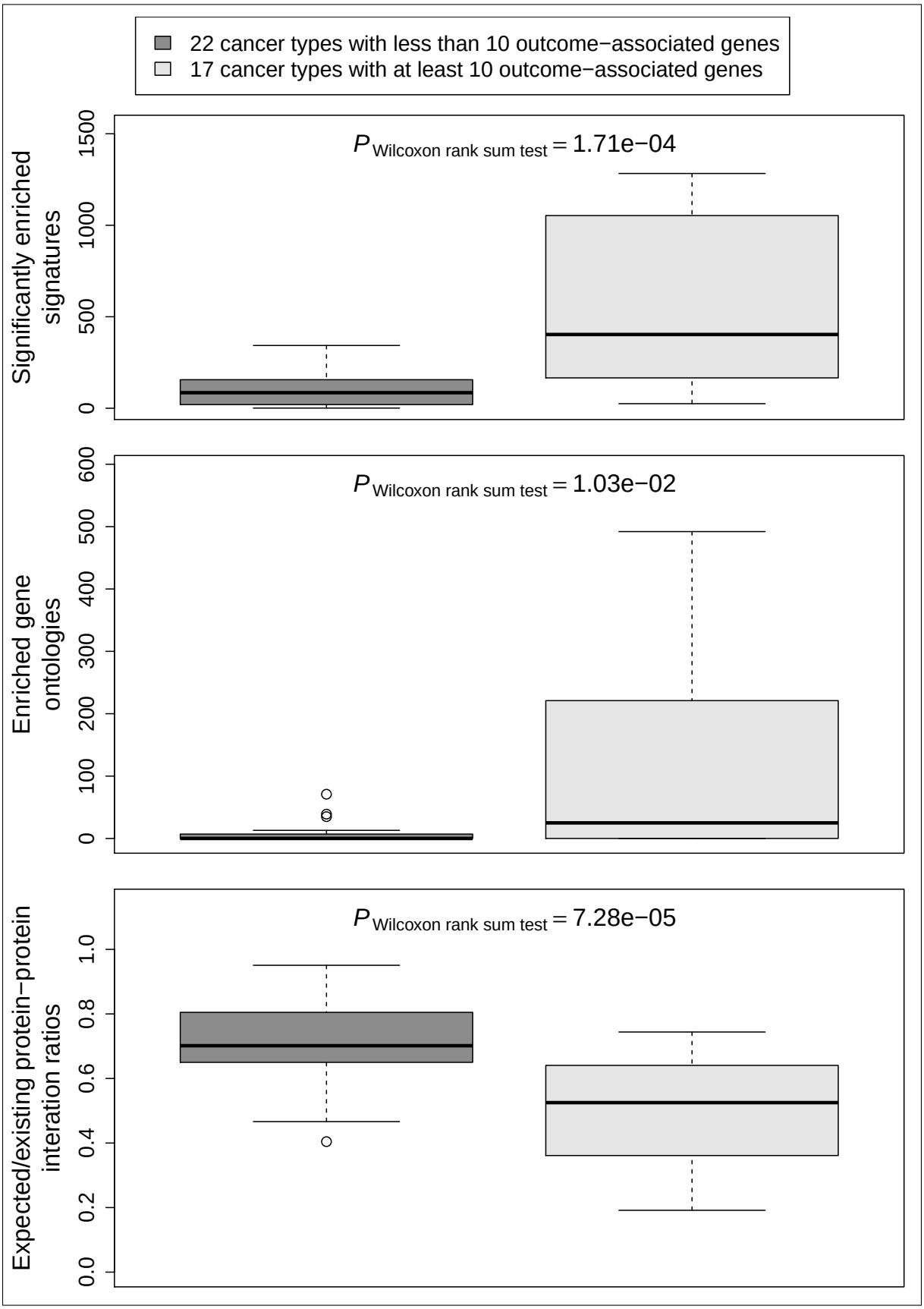
Supplementary figure 3: Complete survival information in the 33 published studies with significant survival differences according to the CINSARC signature. Survival table, below every Kaplan-Meier, describes survival information at several times: from 0 to 5 years. *P* refers to nominal log-rank test significance. HR refers to Hazard Ratio. C1 and C2 refer to the low and high-risk CINSARC classifications, respectively. N risk refers to the remaining population size.

Supplementary figure 4: Evaluation of cohort size effect on prognosis value. Compared to the significance level of CINSARC in the full cohort (P_{Ref}) and to the standard significance threshold (P=0.05; vertical dashed lines), significance levels for 75% and 50% subsampling are given by the red and blue densities, respectively. Distributions of prognoses compared to 0.05 and P_{Ref} are indicated in the table below each density plot. $P>0.05$ represent non-significant subsampling, $P_{Ref}<P<0.05$ represent significant subsampling with lower prognosis than the complete cohort and $P<P_{Ref}$ represent better prognosis than the complete cohort. The sum of the two latter categories represents significant subsampling.

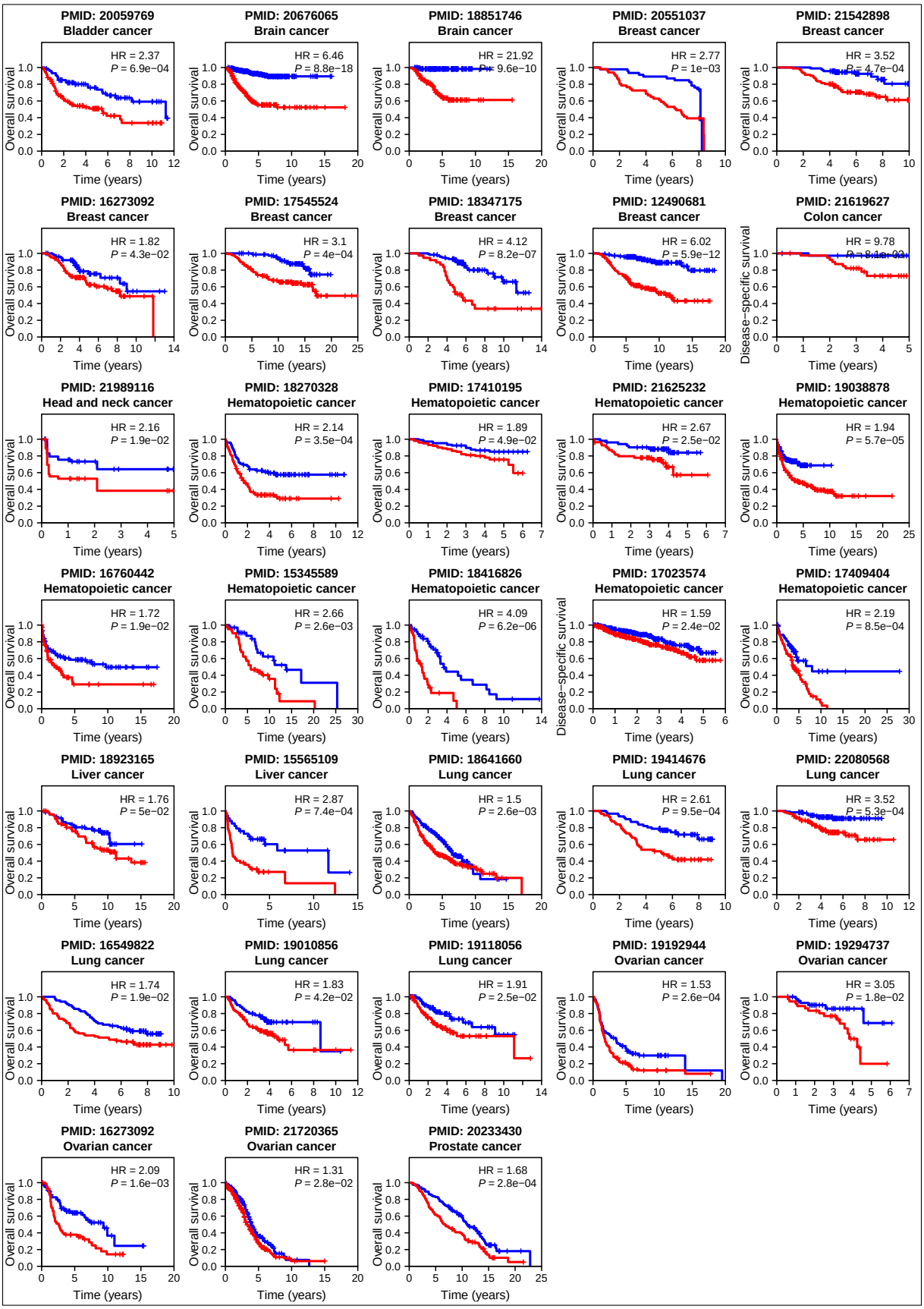
Supplementary table 2

PRECOG ID	PMID	Accession	Platform	Cancer type	Cancer subtype	Outcome	P	HR	HR CI	CINSARC genes	Cohort size
30	20551037	GSE20486	GPL6947	Breast cancer		OS	1.04e−03	2.77	1.47−5.24	67	97
17	12670911	Nutt Glioma	HGU95Av2 EntrezCDF	Brain cancer	Glioma	OS	1.12e−01	1.82	0.86−3.82	46	50
111	11707567	ca00191	HGU95Av2 EntrezCDF	Lung cancer	ADENO	OS	1.18e−01	1.46	0.91−2.34	46	125
27	19435916	GSE10885	GPL887	Breast cancer		OS	1.23e−01	2.22	0.8−6.2	62	52
14	18565887	GSE7696	HGU133Plus2 EntrezCDF	Brain cancer	Glioblastoma	OS	1.29e−01	0.68	0.42−1.12	67	80
166	21372215	GSE21257	GPL10295	Sarcoma	Osteosarcoma	OS	1.48e−01	1.88	0.79−4.46	67	53
65	23634996	TCGA AML	HGU133Plus2 EntrezCDF	Hematopoietic cancer	AML	OS	1.50e−01	1.33	0.9−1.96	67	182
153	16273092	GSE3149	HGU133A EntrezCDF	Ovarian cancer		OS	1.58e−03	2.09	1.31−3.34	62	133
143	18505921	GSE8401	HGU133A EntrezCDF	Melanoma	Metastasis	OS	1.62e−01	0.64	0.33−1.21	62	67
125	20421987	GSE19188	HGU133Plus2 EntrezCDF	Lung cancer	SCC	OS	1.63e−01	1.5	0.85−2.66	67	82
28	19435916	GSE10885	GPL1390	Breast cancer		OS	1.75e−01	1.94	0.73−5.14	50	93
150	19294737	GSE14764	HGU133A EntrezCDF	Ovarian cancer		OS	1.79e−02	0.33	0.12−0.86	62	80
129	17082175	GSE5123	GPL3877	Lung cancer	SCC	OS	1.80e−01	1.68	0.78−3.59	56	51
128	16549822	GSE11969	GPL7015	Lung cancer	SCC	OS	1.89e−02	1.74	1.09−2.77	54	149
75	16760442	GSE4475	HGU133A EntrezCDF	Hematopoietic cancer	DLBCL	OS	1.90e−02	1.72	1.09−2.72	62	159
58	21989116	GSE31056	HGU133Plus2 EntrezCDF	Head and neck cancer	Oral SCC	OS	1.94e−02	2.16	1.12−4.17	67	96
136	16885343	GSE4573	GPL96	Lung cancer	SCC	OS	1.97e−01	0.73	0.45−1.18	61	130
159	20644708	GSE21501	GPL4133	Pancreatic cancer		OS	2.16e−01	1.36	0.83−2.21	66	102
126	21742808	GSE29013	HGU133Plus2 EntrezCDF	Lung cancer	SCC	OS	2.20e−01	1.84	0.69−4.9	67	55
32	21364938	GSE19783	GPL6480	Breast cancer		OS	2.28e−01	1.41	0.81−2.45	67	112
109	12118244	ca00153	Hu6800 EntrezCDF	Lung cancer	ADENO	OS	2.32e−01	1.65	0.72−3.8	29	86
21	20015288	GSE16581	HGU133Plus2 EntrezCDF	Brain cancer	Meningioma	OS	2.34e−01	0.4	0.08−1.92	67	67
87	17023574	GSE24080	HGU133Plus2 EntrezCDF	Hematopoietic cancer	Multiple myeloma	DSS	2.42e−02	1.59	1.06−2.38	67	553
72	21625232	GSE22762	HGU133Plus2 EntrezCDF	Hematopoietic cancer	CLL	OS	2.52e−02	2.67	1.09−6.52	67	107
135	19118056	Roepman LungCancer	Roepman	Lung cancer	SCC	OS	2.54e−02	1.91	1.07−3.41	55	171
114	18641660	ca00182	HGU133A EntrezCDF	Lung cancer	ADENO	OS	2.58e−03	1.5	1.15−1.95	62	442
157	19192944	GSE13876	GPL7759	Ovarian cancer		OS	2.64e−04	1.53	1.21−1.93	66	415
81	15345589	Glas FL	Glas FL	Hematopoietic cancer	FL	OS	2.65e−03	2.66	1.38−5.12	38	75
158	21720365	TCGA Ovarian	HTHGU133A EntrezCDF	Ovarian cancer		OS	2.79e−02	1.31	1.03−1.67	62	503
161	20233430	GSE16560	GPL5474	Prostate cancer		OS	2.82e−04	1.68	1.27−2.23	41	281
132	20643781	GSE17710	GPL9053	Lung cancer	SCC	OS	2.86e−01	1.5	0.71−3.17	66	56
66	19171880	GSE14468	HGU133Plus2 EntrezCDF	Hematopoietic cancer	AML	OS	2.98e−01	1.17	0.87−1.59	67	262
31	20098429	GSE16446	HGU133Plus2 EntrezCDF	Breast cancer		OS	3.19e−01	0.59	0.21−1.67	67	107
64	18270328	GSE10358	HGU133Plus2 EntrezCDF	Hematopoietic cancer	AML	OS	3.46e−04	2.14	1.4−3.29	67	178
155	22241791	GSE32062	GPL6480	Ovarian cancer		OS	3.59e−01	1.19	0.82−1.71	67	260
74	18615101	E−TABM−346	HGU133A EntrezCDF	Hematopoietic cancer	DLBCL	OS	3.62e−01	1.39	0.68−2.85	62	53
76	16760443	GSE4732	GPL3706	Hematopoietic cancer	DLBCL	OS	3.62e−01	1.16	0.84−1.61	27	272
29	21586611	GSE29174	GPL3676	Breast cancer		OS	3.76e−01	0.72	0.34−1.5	56	96
123	18486272	GSE10245	HGU133Plus2 EntrezCDF	Lung cancer	SCC	OS	3.81e−01	0.69	0.3−1.6	67	58
36	17545524	GSE7390	HGU133A EntrezCDF	Breast cancer		OS	4.04e−04	3.1	1.6−6	62	198
134	19010856	GSE8894	GPL570	Lung cancer	SCC	OS	4.17e−02	1.83	1.01−3.31	66	138
37	16273092	GSE3143	HGU95Av2 EntrezCDF	Breast cancer		OS	4.31e−02	1.82	1.01−3.28	46	158
148	19962670	GSE18521	HGU133Plus2 EntrezCDF	Ovarian cancer		OS	4.33e−01	0.77	0.41−1.44	67	53
39	21542898	GSE24450	GPL6947	Breast cancer		OS	4.66e−04	3.52	1.66−7.45	67	183
133	17601969	GSE5828	GPL3877	Lung cancer	SCC	OS	4.72e−01	0.77	0.37−1.58	54	59
46	21619627	GSE24549	GPL5175	Colon cancer		DSS	4.83e−01	0.8	0.43−1.48	67	83
68	17410195	GSE11877	HGU133Plus2 EntrezCDF	Hematopoietic cancer	B ALL	OS	4.92e−02	1.89	0.99−3.61	67	207
93	18923165	GSE10143	GPL5474	Liver cancer		OS	4.96e−02	1.76	0.99−3.13	41	144
59	18716133	GSE12417	HGU133Plus2 EntrezCDF	Hematopoietic cancer	AML	OS	5.19e−01	0.83	0.46−1.47	67	79
149	16204010	GSE31245	HGU95Av2 EntrezCDF	Ovarian cancer		OS	5.19e−01	0.77	0.34−1.73	46	57
113	22080568	GSE31210	HGU133Plus2 EntrezCDF	Lung cancer	ADENO	OS	5.32e−04	3.52	1.65−7.52	67	226
130	20823422	GSE14814	HGU133A EntrezCDF	Lung cancer	SCC	OS	5.49e−01	0.82	0.44−1.56	62	90
154	18593951	GSE26712	HGU133A EntrezCDF	Ovarian cancer		DSS	5.67e−01	0.9	0.63−1.28	62	185
80	19038878	GSE10846	HGU133Plus2 EntrezCDF	Hematopoietic cancer	DLBCL	OS	5.73e−05	1.94	1.4−2.69	67	414
67	17312329	GSE5314	GPL3999	Hematopoietic cancer	B ALL	OS	5.85e−01	1.19	0.64−2.22	54	53
41	12490681	Vijver BreastCancer	Vijver Agilent	Breast cancer		OS	5.88e−12	6.02	3.37−10.75	57	295
142	20460471	GSE22155	GPL6102	Melanoma		OS	6.08e−02	1.75	0.97−3.18	65	54
83	18416826	GSE10793	GPL3278	Hematopoietic cancer	Mantle cell lymphoma	OS	6.21e−06	4.09	2.13−7.83	25	71
139	15016488	GSE1037	GPL962	Lung cancer	SCLC	OS	6.33e−02	2.41	0.93−6.3	55	61
20	11807556	Pomeroy CNS	GPL80	Brain cancer	Medulloblastoma	OS	6.37e−01	0.8	0.32−1.99	28	60
156	18698038	GSE9899	HGU133Plus2 EntrezCDF	Ovarian cancer		OS	6.46e−01	1.09	0.75−1.59	67	278
44	19399471	GSE12945	HGU133A EntrezCDF	Colon cancer		OS	6.72e−01	0.78	0.25−2.46	62	62
6	20059769	GSE13507	GPL6102	Bladder cancer		OS	6.86e−04	2.37	1.42−3.95	65	165
131	16273092	GSE3141	HGU133Plus2 EntrezCDF	Lung cancer	SCC	OS	7.17e−01	0.91	0.54−1.53	67	111
152	20300634	GSE17260	GPL6848	Ovarian cancer		OS	7.37e−01	1.1	0.62−1.98	66	110
92	15565109	GSE1898	GPL1528	Liver cancer		OS	7.38e−04	2.87	1.52−5.44	66	76
47	19914252	GSE17536	HGU133Plus2 EntrezCDF	Colon cancer		OS	7.40e−01	1.08	0.68−1.73	67	177
34	19204204	GSE10886	GPL1390	Breast cancer		OS	7.77e−01	0.91	0.47−1.75	60	178
45	21619627	GSE24550	GPL5175	Colon cancer		DSS	8.15e−03	0.1	0.01−0.81	67	77
137	22960745	TCGA LUSC	HTHGU133A EntrezCDF	Lung cancer	SCC	OS	8.18e−01	1.07	0.61−1.88	62	132
35	18347175	GSE9893	GPL5049	Breast cancer		OS	8.20e−07	4.12	2.24−7.58	32	155
54	20458058	Thurlow HNSCC	HGU133Plus2 EntrezCDF	Head and neck cancer		DSS	8.36e−01	1.08	0.52−2.24	67	74
85	17409404	GSE6477	HGU133A EntrezCDF	Hematopoietic cancer	Multiple myeloma	OS	8.47e−04	2.19	1.37−3.52	62	145
43	19914252	GSE17537	HGU133Plus2 EntrezCDF	Colon cancer		OS	8.73e−01	0.93	0.39−2.25	67	55
24	20676065	E−MTAB−179	NBCustomAgilent4x44K	Brain cancer	Neuroblastoma	OS	8.84e−18	6.46	3.96−10.54	67	478
16	18772890	TCGA Glioblastoma	HTHGU133A EntrezCDF	Brain cancer	Glioblastoma	OS	9.01e−01	1.01	0.8−1.29	62	348
151	19047114	GSE8842	GPL5689	Ovarian cancer		OS	9.12e−01	0.94	0.34−2.65	42	83
63	16597596	ca00119	HGU95Av2 EntrezCDF	Hematopoietic cancer	AML	OS	9.14e−02	1.33	0.95−1.85	46	169
77	11786909	Shipp DLBCL	Hu6800 EntrezCDF	Hematopoietic cancer	DLBCL	OS	9.34e−01	1.03	0.47−2.26	29	58
112	19414676	GSE13213	GPL6480	Lung cancer	ADENO	OS	9.46e−04	2.61	1.45−4.7	67	117
23	18851746	E−TABM−38	A−MEXP−255	Brain cancer	Neuroblastoma	OS	9.60e−10	21.92	5.27−91.23	55	251
138	22960745	TCGA LUSC	AgilentG4502A Gene	Lung cancer	SCC	OS	9.67e−01	0.99	0.58−1.68	57	151
18	19244127	ca00037	HGU133Plus2 EntrezCDF	Brain cancer	Glioma	OS	9.87e−01	1.01	0.52−1.94	67	147

Supplementary figure 1



Supplementary figure 2



3.1.3.2 – Conclusion

Les données contenues dans PRECOG (Gentles *et al.*, 2015) nous ont permis de démontrer que CINSARC est significativement enrichie en gènes individuellement pronostiques en analyse pan-cancer et que, parmi 15499 signatures, elle est la plus associée au devenir clinique des patients. En ré-analysant les données en survie globale, nous avons pu confirmer qu'elle est bien un facteur pronostique dans 33 études, couvrant 17 types de cancers. Il semble donc y avoir un potentiel intérêt à utiliser cette signature comme un reflet de l'agressivité tumorale au-delà des sarcomes.

Ce travail constitue une première évaluation exhaustive hors-sarcomes de CINSARC. Bien que très informative, plusieurs limitations de PRECOG peuvent artificiellement limiter la valeur pronostique de la signature. L'un de ces facteurs limitants, et le plus mesurable, est la taille des cohortes (<100 patients) pour de nombreuses études. Ainsi, des analyses plus fines (contrôle des tailles des cohortes, des sous-types de tumeurs, des taux de survie, *etc.*) sur chaque type de cancer permettront de statuer sur le bénéfice du pronostic de CINSARC.

Pour chaque type de cancer, un système de référence a justement été établi pour permettre une évaluation correcte de l'agressivité tumorale (sous forme de survie sans rechute, sans métastase et/ou survie globale). Des analyses similaires seront nécessaires afin de tester la valeur pronostique de CINSARC comparativement au système de référence de chaque type de cancer. Dans le cas où, comme dans les sarcomes (Valentin *et al.*, 2019), CINSARC serait un meilleur marqueur que le système de référence d'une pathologie, elle aura alors un intérêt clinique majeur.

Par ailleurs, en collaboration avec Sabrina Croce (pathologiste à l'Institut Bergonié, Bordeaux), des analyses sont en cours pour déterminer l'apport de CINSARC dans les tumeurs musculaires lisses utérines (TMLU). De récents travaux nous ont en effet conduits à nous intéresser à cette famille de cancers gynécologiques dont l'instabilité génétique a une valeur à la fois diagnostique et également pronostique (Croce *et al.*, 2018 en **annexe I page 228**). Les critères morphologiques ne permettent pas toujours d'identifier un sous-type histologique précis parmi les TMLU, conduisant au diagnostic d'exclusion de tumeur musculaire lisse de malignité

incertaine (STUMP pour *uterine smooth muscle tumor with uncertain malignant potential*). L'analyse génomique que nous avons menée sur 77 TMLU nous a permis de déterminer que le nombre d'altérations génomiques (évalué par le GI) était relié à la malignité des tumeurs. Les remaniements mesurés permettent ainsi de re-stratifier les STUMP sur leur caractère bénin ou malin, ayant alors un comportement clinique similaire aux léiomyomes (tumeurs bénignes ; GI<10) ou aux LMS utérins (tumeurs hautement malignes ; GI$\geq$10). De plus, le profil génomique, au sein des tumeurs malignes, est pronostique du statut vital des patientes (*P*=3,5e-2 ; seuil de GI=35). Cet apport pronostique est d'autant plus important qu'il re-stratifie aussi de façon significative les tumeurs de stade I (*P*=1,93e-2), évaluées selon les critères définis par la Fédération Internationale des Gynécologues-Obstétriciens. En plus des TMLU, CINSARC sera aussi prochainement testée dans d'autres pathologies : les tumeurs de l'endomètre (en collaboration avec Éliane Mery-Lamarche, pathologiste à l'Institut Universitaire du Cancer de Toulouse) et les tumeurs de l'ovaire (en collaboration avec Laurence Gladieff et Gwenael Ferron, respectivement clinicienne et chirurgien à l'Institut Universitaire du Cancer de Toulouse).

3.1.4 – Causes possibles de l'expression de CINSARC

Si l'association entre l'expression de la signature et l'agressivité tumorale est maintenant bien établie, l'origine de cette expression reste inexpliquée. La composition de la signature CINSARC permet néanmoins de poser deux constats. Premièrement, ces gènes font partie d'un réseau transcriptomique extrêmement bien co-régulé (Su *et al.*, 2004 ; Subramanian *et al.*, 2005 ; Szklarczyk *et al.*, 2015). Deuxièmement, ces gènes ne sont que très rarement retrouvés génétiquement altérés dans les tumeurs (Gordon *et al.*, 2012 ; Potapova *et al.*, 2013 ; Sansregret & Swanton, 2017 ; The Cancer Genome Atlas Research Network, 2017). Puisque la régulation de l'expression de cette signature ne semble pas être associée aux altérations de ses propres gènes, nous avons cherché à déterminer quels facteurs permettraient son expression conditionnelle afin de mieux comprendre quels mécanismes régissent l'agressivité des SGC.

La complétion du projet TCGA sur les sarcomes a justement rendu publique un important volume de données dont certaines que nous pourrions utiliser afin de mieux comprendre les possibles origines de la signature CINSARC. Ces données consistent, entre autres, au niveau de méthylation du génome et à l'expression de miRNA. Ces deux régulateurs majeurs de l'expression génique sont de plus en plus impliqués dans les processus d'initiation et de progression tumorale (**partie 1.2.4 page 61**) et seraient de bons candidats pour expliquer l'expression différentielle de CINSARC.

3.1.4.1 – Présentation des données et sélection des cas du TCGA

Le projet TCGA sarcomes correspond à une analyse intégrative de 206 tumeurs comprenant six sous-types histologiques (LMS, DDLPS, UPS, MFS, MPNST et synovialosarcomes ; The Cancer Genome Atlas Research Network, 2017). Des données non publiées de cas additionnels sont également rendues publiques pour un total de 261 cas (https://portal.gdc.cancer.gov/projects/TCGA-SARC). Pour chacun de ces cas, différents types de données sont accessibles : données cliniques (caractéristiques des patients et des tumeurs aux niveaux macroscopique et microscopique), profils transcriptomiques (RNA-seq), altérations somatiques ponctuelles (exome-seq et quelques DNA-seq), méthylome (puces Infinium

450K, Illumina), expression de miRNA (miRNA-seq pour *micro*RNA *sequencing*) et variations du nombre de copies (puces SNP6, Affymetrix).

Avant d'explorer les données de méthylome et de miRNA, le premier objectif est de déterminer si nous retrouvons les associations entre CINSARC, agressivité tumorale et remaniements génomiques dans la cohorte du TCGA. La validation de cette étape permettrait d'établir une solide base de travail pour pouvoir effectuer une analyse exploratoire exhaustive de l'origine de l'expression de cette signature. Nous avons déjà montré que CINSARC est un marqueur pronostique dans une grande partie de cette cohorte (P=4,1e-3 sur 246 cas ; Lesluyes *et al.*, 2016). Toutefois, ce groupe est très hétérogène et le temps de suivi de nombreux patients est insuffisant (quelques mois) pour précisément définir le statut métastatique. La sélection d'un sous-groupe aux caractéristiques contrôlées devenait alors nécessaire.

Sur l'ensemble des 261 cas, nous avons retenu les principaux sous-types de SGC (UPS, MFS et LMS) dans le but de constituer un groupe homogène et représentatif de la pathologie. De plus, les cas non métastatiques devaient avoir un suivi de plus d'un an et les données de RNA-seq devaient être accessibles pour l'ensemble de ces cas. Ceci nous a permis de constituer une cohorte initiale de 130 cas. À partir des données cliniques et l'expression de la signature CINSARC, nous avons déterminé le pronostic de chaque tumeur par une méthode analogue au *leave-one-out cross-validation*, où la valeur pronostique de chaque tumeur est obtenue en utilisant la méthode du plus proche centroïde sur l'ensemble des autres tumeurs comme référence (en deux groupes de bon et de mauvais pronostic selon le statut métastatique ; Lesluyes *et al.*, 2017). Sur ces 130 tumeurs, CINSARC est un facteur pronostique significativement associé au développement métastatique (test du *log-rank* P=1,89e-2 ; HR=1,82 [1,1-3,03]). Pour les besoins des futures analyses et afin de maximiser les différences entre groupes, nous n'avons conservé que les cas les plus fortement associés à leurs plus proches centroïdes (ρ>0,2).

La cohorte finale est composée de 84 tumeurs, dont 36 sont de bon pronostic (CINSARC C1) et 48 sont de mauvais pronostic (CINSARC C2 ; **figure 24A**). Ces deux groupes partagent des caractéristiques cliniques similaires en matière de genre, sous-type histologique et résection chirurgicale (enrichissements de Fisher P>0,05). Le groupe C1 est significativement enrichi en sarcomes localisés dans le tronc par rapport au groupe C2 dans lequel les tumeurs sont principa-

lement localisées aux niveaux des extrémités (enrichissements de Fisher P<0,05). Les différents taux d'événements métastatiques dans les groupes se traduisent par des survies sans métastase significativement différentes (P=2,89e-03 ; **figure 24B**). Ces taux de survie sans métastase à cinq ans pour les deux groupes sont de 66% [52-84] pour le groupe C1 et 24% [13-43] pour le groupe C2. Les profils transcriptomiques de cette cohorte permettent ainsi, via la signature CINSARC, de bien stratifier les tumeurs en fonction de leur agressivité en deux groupes de pronostics différents. Nous avons ensuite analysé ces profils transcriptomiques afin de déterminer si les gènes de CINSARC sont bien significativement plus exprimés dans le groupe de mauvais pronostic.

Sur l'ensemble des 55594 gènes annotés (codants et non codants), l'analyse d'expression différentielle (réalisée avec DESeq2 ; Love *et al.*, 2014) montre que les 67 gènes de la signature sont bien sur-exprimés dans le groupe C2 comparativement au groupe C1 (P<0,05 pour les 67 gènes ; P<0,05 et *log2 fold-change*>1 pour 61 gènes ; **figure 24C** où les gènes de CINSARC sont colorés en orange). Cette analyse permet aussi d'identifier 136 gènes qui suivent la tendance différentielle des gènes de CINSARC (**figure 24C** : trainée de points noirs se confondant avec les points oranges). Ces gènes font partie intégrante du réseau d'expression de la signature CINSARC (Su *et al.*, 2004 ; Subramanian *et al.*, 2005 ; Szklarczyk *et al.*, 2015). Dans sa globalité, l'analyse d'expression différentielle identifie 2037 gènes sur-exprimés dans le groupe CINSARC C2 et 2028 gènes sur-exprimés dans le groupe CINSARC C1. Les principales voies biologiques associées au groupe C2 sont la ségrégation chromosomique et la régulation de la division cellulaire (P<1e-15 ; analyse réalisée avec GOseq ; Young *et al.*, 2010), ce qui concorde avec une sur-expression globale du réseau dans les tumeurs de mauvais pronostic. Les principales voies biologiques associées au groupe C1 sont la contraction musculaire (P=1,35e-10) et l'homéostasie calcique (P=7,57e-9). Ce dernier résultat pourrait être associé à une différenciation musculaire lisse des LMS du groupe C1 qui sont plus nombreux (et peut être aussi plus différenciés) que dans le groupe C2. Les gènes de la signature sont ainsi bien sur-exprimés dans le groupe de mauvais pronostic.

En se basant sur les données de variation du nombre de copies, nous avons regardé si l'expression de CINSARC est associée aux remaniements génomiques comme précédemment décrit (Chibon *et al.*, 2010 ; Lagarde *et al.*, 2012, 2013). Les tumeurs de mauvais pronostic

A

	CINSARC C1 (*n*=36)	CINSARC C2 (*n*=48)
Median age at diagnosis [CI]	58.5 [56-66]	62.5 [58-67]
Median follow-up in years [CI]	3.02 [1.75-4.43]	2.2 [1.51-3.28]
Gender:		
Female	18 (50%)	32 (67%)
Male	18 (50%)	16 (33%)
Histological subtypes:		
LMS	26 (72%)	28 (58%)
UPS	4 (11%)	14 (29%)
MFS	6 (17%)	6 (12%)
Tumor site:		
Chest wall	25 (69%)	8 (17%)
Gynecological	5 (14%)	12 (25%)
Lower extremity	5 (14%)	25 (52%)
Upper extremity	1 (3%)	2 (4%)
Head and neck	0 (0%)	1 (2%)
Residual tumor:		
R0	24 (67%)	29 (60%)
R1	9 (25%)	13 (27%)
R2	0 (0%)	0 (0%)
RX	3 (8%)	6 (12%)
Metastasis	13 (36%)	34 (71%)
Death	13 (36%)	17 (35%)

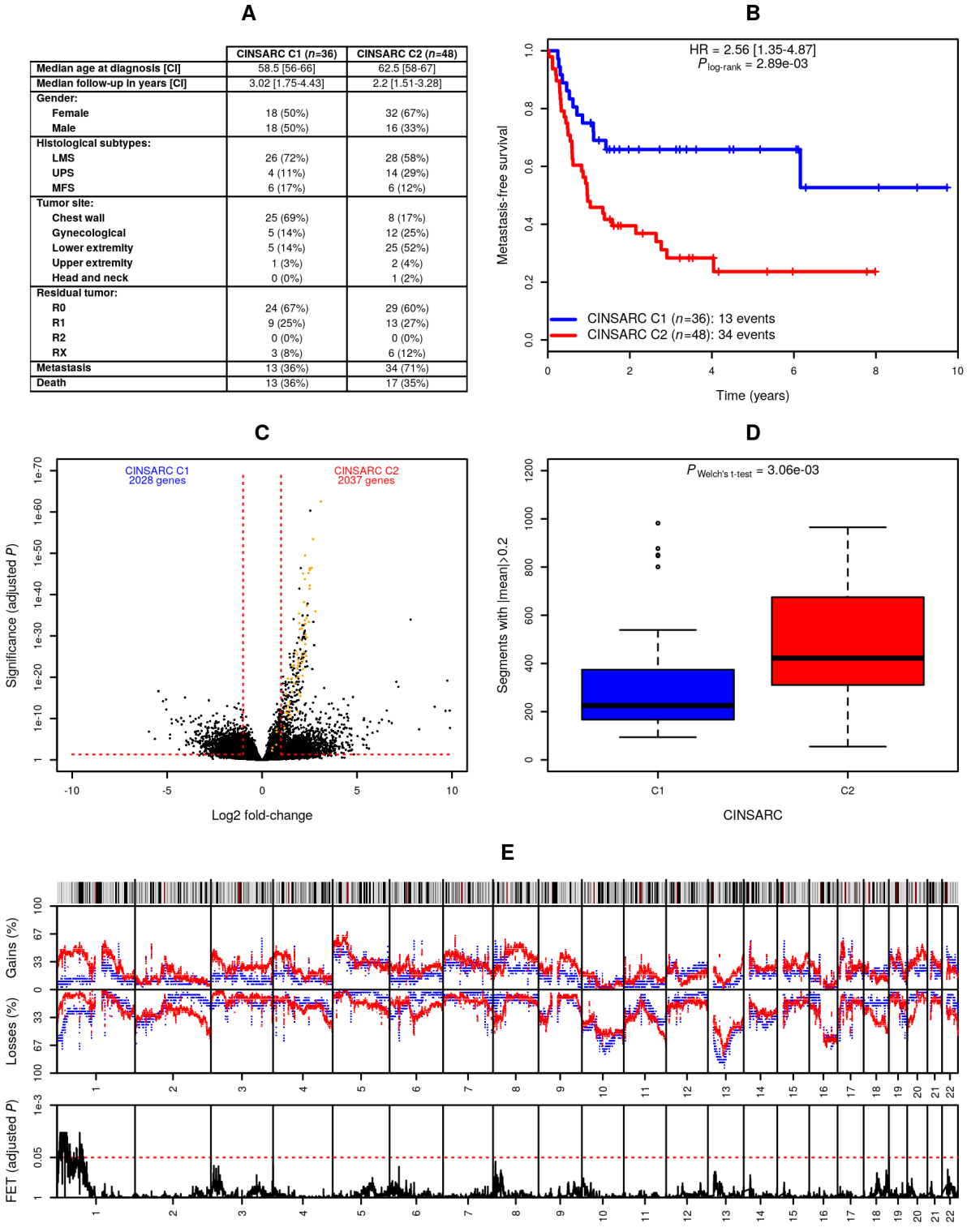


FIGURE 24 – Sélection des cas du TCGA et analyses génomiques. **A** : présentation des caractéristiques cliniques des cas retenus. **B** : courbes de survie sans métastase des groupes de bon et de mauvais pronostic de la cohorte. **C** : analyse d'expression différentielle entre CINSARC C1 et C2, les 67 gènes de la signature sont colorés en orange. **D** : variation différentielle du nombre de copies détectées dans les deux groupes. **E** : variations récurrentes du nombre de copies dans les deux groupes de différents pronostics. Le niveau de variation différentielle est mesuré en utilisant des tests d'enrichissement de Fisher ajustés par la méthode de Benjamini-Hochberg.

présentent significativement plus de variations quantitatives que les tumeurs de bon pronostic (*P*=3,06e-3 ; **figure 24D**). Quant à la localisation génomique, la seule région différentiellement altérée est le 1p, région plus fréquemment gagnée dans le groupe de mauvais pronostic que dans celui de bon pronostic (**figure 24E**). Il est à noter que cinq gènes de la signature se situent sur la région 1p : *CDC20* (1p34.2), *CDC7* (1p22.2), *CDCA8* (1p34.3), *HP1BP3* (1p36.12) et *KIF2C* (1p34.1). De précédentes études ont permis de rapporter des variations récurrentes dans les SGC : gain du chromosome 1 et perte des chromosomes 10, 13 et 16 (Mairal *et al.*, 1999 ; Chibon *et al.*, 2003 ; Guillou & Aurias, 2010 ; Gibault *et al.*, 2011 ; Italiano *et al.*, 2013). Nous retrouvons ici ces altérations comme les plus récurrentes, mais les pertes des chromosomes 10, 13 et 16 ne sont pas significativement associées à un groupe CINSARC. Le gain du 1p est la seule région qui soit reliée à un phénotype plus agressif. À titre comparatif : neuf gènes de CINSARC sont sur le 1q et les chromosomes 10, 13 et 16 contiennent respectivement six, deux et deux gènes de la signature.

Le doublement du génome est considéré comme l'étape conduisant à l'aneuploïdie et à l'instabilité génétique, condition nécessaire à l'acquisition de nouvelles mutations et à l'augmentation de l'hétérogénéité intra-tumorale (Dewhurst *et al.*, 2014 ; Gerlinger *et al.*, 2014 ; Venkatesan & Swanton, 2016). Tous ces paramètres ont été mesurés dans le cadre du projet TCGA sarcomes. La charge mutationnelle a été déterminée avec l'exome-seq/DNA-seq entre la tumeur et le tissu sain du patient pour ne considérer que les altérations ponctuelles somatiques. Le doublement du génome, la ploïdie et la sous-clonalité ont été déterminés d'après les données de puces génomiques (algorithme ABSOLUTE ; Carter *et al.*, 2012). Afin de mieux définir les caractéristiques génétiques des tumeurs CINSARC C1 et C2, nous avons alors comparé les mesures de ces paramètres dans nos deux groupes de pronostics opposés.

Sur les 84 tumeurs de la cohorte étudiée, ces informations sont disponibles pour 61 cas (73%). Les analyses statistiques montrent que le groupe CINSARC C2 est significativement enrichi en tumeurs ayant doublé leur génome comparativement au groupe CINSARC C1 (*P*=3,98e-2 ; **figure 25A**). Un résultat similaire est obtenu par l'analyse de la ploïdie, où les tumeurs de bon pronostic sont majoritairement diploïdes alors que les tumeurs de mauvais pronostic ont des ploïdies plus élevées (*P*=1,87e-2 ; **figure 25B**). L'analyse de la charge mutationnelle indique que les tumeurs CINSARC C2 présentent significativement plus de mutations

somatiques que les tumeurs CINSARC C1 (P=1,67e-2 ; **figure 25C**). Finalement, l'estimation de la fraction de génomes sous-clonaux indique une plus grande hétérogénéité intra-tumorale pour le groupe CINSARC C2 que le groupe CINSARC C1 (P=5,91e-5 ; **figure 25D**).

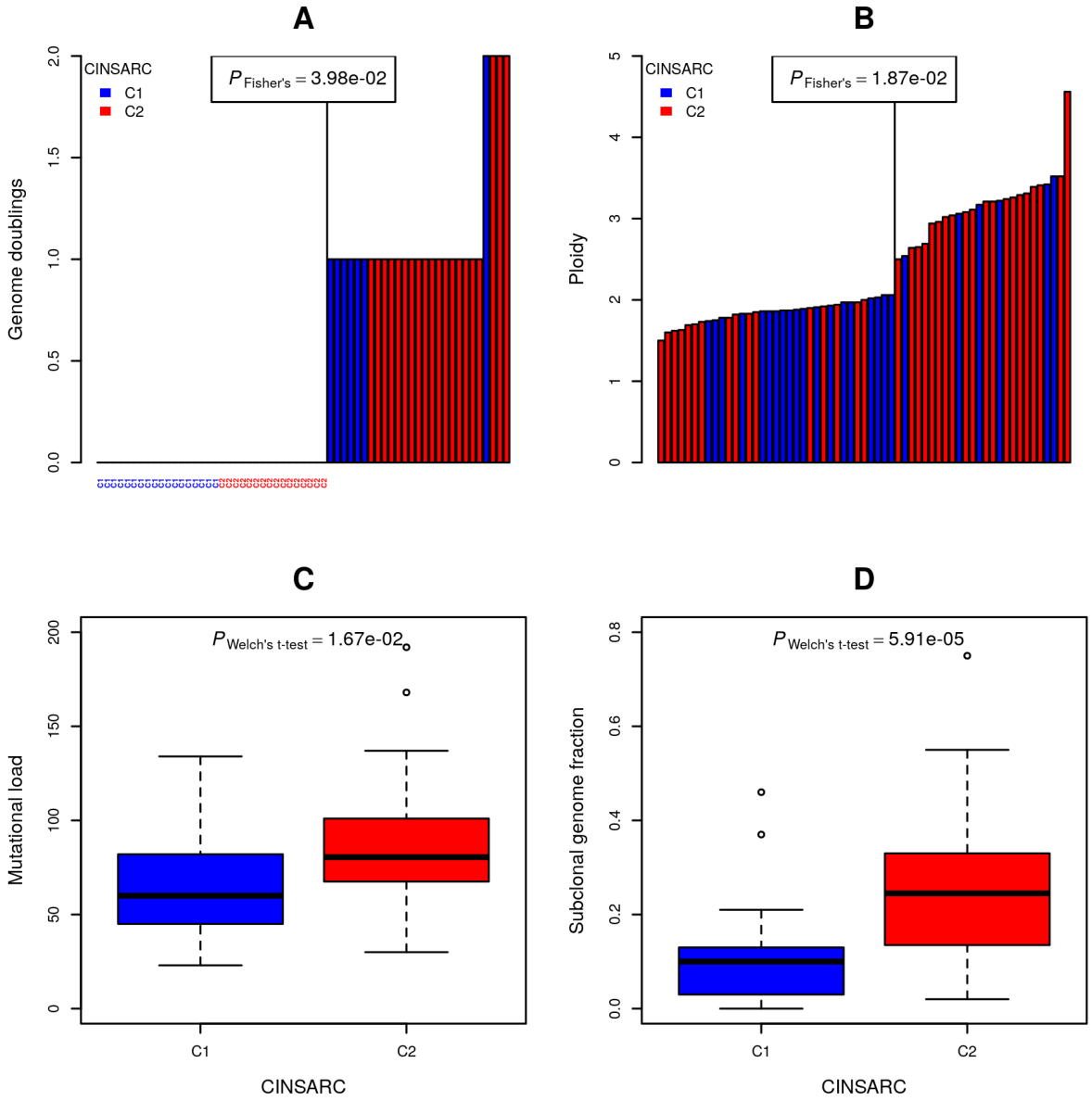


FIGURE 25 – Relation entre CINSARC, ploïdie, charge mutationnelle et hétérogénéité intra-tumorale. **A** : relation entre le doublement du génome et le statut CINSARC. Le seuil de définition des groupes a été fixé à 1 (0 doublement contre au moins 1). **B** : relation entre la ploïdie et le statut CINSARC. Le seuil de définition des groupes a été fixé à 2,28 (correspond au gap observé entre tumeurs pseudo-diploïdes et pseudo-triploïdes). **C** : relation entre la charge mutationnelle et le statut CINSARC. **D** : relation entre l'hétérogénéité intra-tumorale et le statut CINSARC.

L'ensemble des résultats acquis sur cette cohorte confirme les précédentes observations : l'expression de la signature CINSARC est reliée au pronostic et à la complexité génomique de ces tumeurs. Cette nouvelle cohorte a également apporté de nouveaux résultats en précisant que

les tumeurs de mauvais pronostic ont une ploïdie (associée à un doublement du génome), une charge mutationnelle ainsi qu'une hétérogénéité plus importantes que les tumeurs de bon pronostic. Ces données sont en accord avec la littérature, de nombreuses études ayant démontré que ces facteurs sont associés à une plus grande agressivité tumorale (**partie 1.2.2.4 page 50**).

Grâce aux données disponibles dans le projet TCGA sarcomes, nous pouvons donc rechercher si l'expression de CINSARC peut être directement régulée via le méthylome (régulation épigénétique) et/ou l'expression de miRNA (régulation post-transcriptionnelle), pistes jusqu'alors inexplorées et pouvant expliquer l'origine de l'expression différentielle de cette signature.

3.1.4.2 – Exploration des données de méthylation

La méthylation d'îlots CpG dans les régions promotrices permet de réguler négativement l'expression génique et la déméthylation globale du génome est fréquemment observée dans les cancers (**partie 1.2.4.1 page 61** ; Feinberg & Vogelstein, 1983 ; Podlaha *et al.*, 2012). Sur la base de ce constat, nous nous sommes demandé si la régulation de CINSARC pourrait avoir le méthylome comme origine.

Le matériel et méthodes associés à cette partie sont disponibles en **annexe II page 241**.

En analyse non supervisée, lorsque l'on classe les tumeurs par niveau de méthylation croissante (moyenne des β-*values* de chaque tumeur), nous observons un enrichissement significatif des tumeurs de mauvais pronostic dans le groupe hypométhylé (P=1,24e-4 ; **figure 26A**). Un tel enrichissement est aussi observé en matière d'événement métastatique mais cette association, bien que significative, est moindre (P=4,66e-2 ; **figure 26B**).

En analyse supervisée (test t de Welch et ajustement par la méthode de Benjamini-Hochberg ; Du *et al.*, 2010), nous avons identifié 3369 sites (0,94%) hyperméthylés dans le groupe de mauvais pronostic et 41827 sites (11,64%) hyperméthylés dans le groupe de bon pronostic (**figure 26C**). Par opposition, aucune sonde n'est différentiellement méthylée selon le statut métastatique des tumeurs (**figure 26D**).

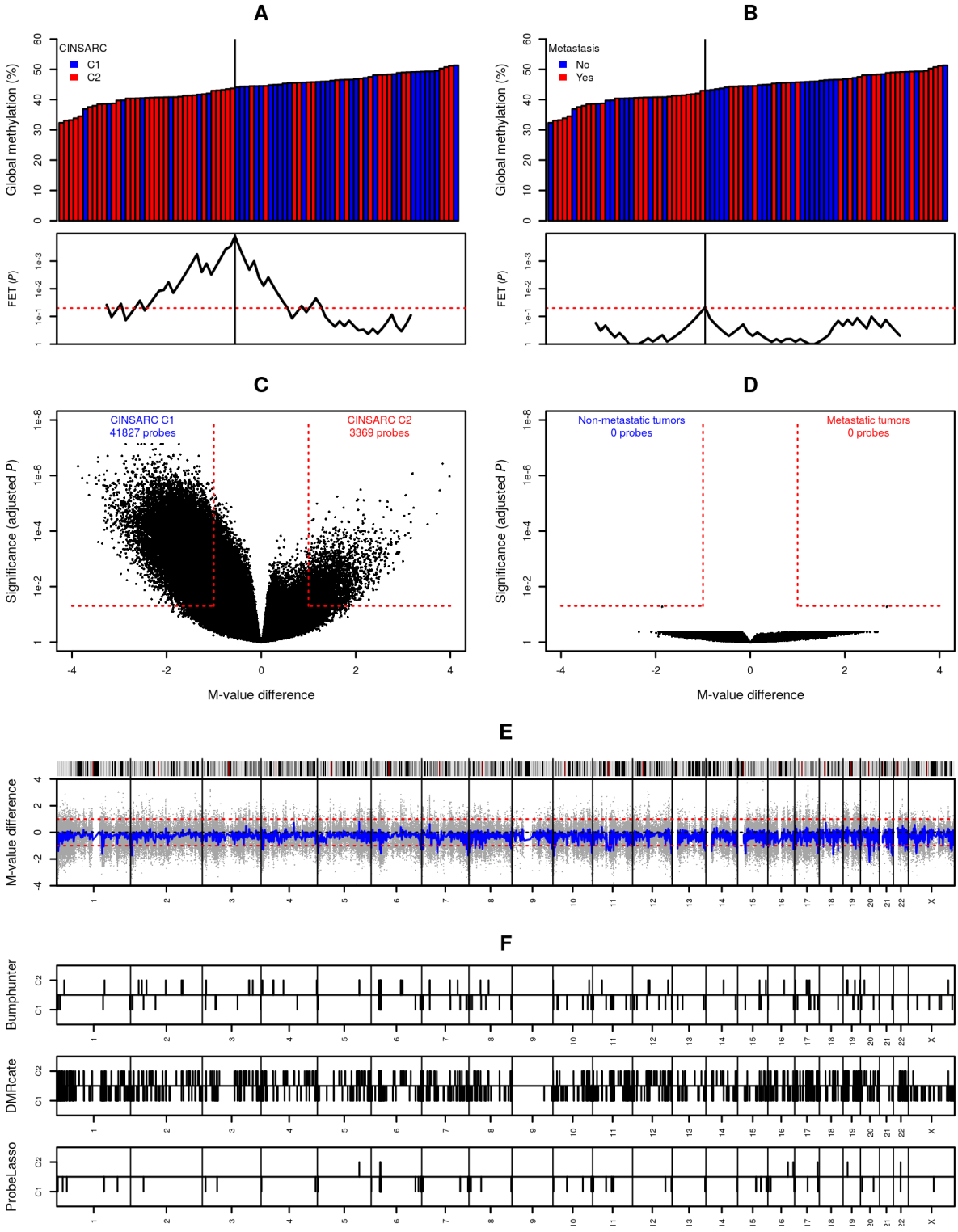


FIGURE 26 – Expression de la signature CINSARC et niveau de méthylation. **A** et **B** : mesure du niveau de méthylation global en fonction du pronostic de la signature CINSARC (A) et du statut métastatique (B). Les graphiques du bas montrent les enrichissements de Fisher sur des fenêtres glissantes. **C** et **D** : sondes différentiellement méthylées en fonction du pronostic de la signature CINSARC (C) et du statut métastatique (D). **E** : visualisation génomique des sondes différentiellement méthylées en fonction du pronostic de la signature CINSARC (C1 et C2 sont respectivement représentés vers le bas et le haut). **F** : visualisation génomique de régions identifiées comme différentiellement méthylées en fonction du pronostic de la signature CINSARC (C1 et C2 sont respectivement représentés vers le bas et le haut) par trois algorithmes implémentés dans ChAMP (Morris *et al.*, 2014 ; Tian *et al.*, 2017).

Puisque des sondes sont différentiellement méthylées selon le statut CINSARC, nous avons cherché à identifier quelles régions génomiques (regroupant un ensemble de sondes) sont préférentiellement impactées. Ces sondes sont réparties sur la totalité du génome humain mais certaines régions semblent sélectivement méthylées selon le statut CINSARC (comme le 1p ou les régions 6p22.2-p21.33 ; **figure 26E**). Afin de déterminer si ces régions sont bien sélectivement méthylées, nous avons utilisé plusieurs algorithmes dédiés à la détermination de telles régions (paquet ChAMP qui implémente trois algorithmes différents : BumpHunter, DMRcate, et ProbeLasso ; Jaffe *et al.*, 2012 ; Morris *et al.*, 2014 ; Peters *et al.*, 2015 ; Butcher & Beck, 2015 ; Tian *et al.*, 2017).

Ces algorithmes ont permis l'identification de nombreuses régions génomiques différentiellement méthylées (**figure 26F**). Toutefois, aucune de ces régions n'impacte l'un des gènes de la signature CINSARC. De plus, l'analyse des gènes localisés dans ces régions ne permet pas de détecter un enrichissement d'une voie biologique particulière (régions C1 et C2 analysées séparément). En l'absence de résultat significatif au niveau de régions génomiques, potentiellement trop courtes pour être détectées par les algorithmes, nous sommes retournés sur les données de sondes différentielles.

Les puces contiennent 881 sondes localisées à l'intérieur, ou proches (<4kb), des îlots CpG reliés aux gènes de la signature CINSARC (pouvant être proches ou lointaines des régions promotrices). Parmi ces 881 sondes, 12 (1,36%) sont significativement différentiellement méthylées (P<0,05 et différence absolue de M-*value*>1). Ces 12 sondes correspondent à six gènes et sont toutes hyperméthylées pour les tumeurs CINSARC C1 (**table 6**). En analyse non supervisée, ces 12 sondes permettent de retrouver le statut CINSARC en fonction de leur niveau de méthylation (enrichissement de Fisher P=1,77e-8 ; **figure 27**).

Si le niveau de méthylation de ces sondes est bien associé au statut CINSARC, les distances entre ces sondes et les sites d'initiation de la transcription (TSS pour *transcription start site*) des gènes ne sont pas en faveur d'une potentielle régulation. En effet, le niveau de méthylation des régions promotrices, proches des TSS, permet de moduler l'expression génique. Ces sondes ne sont pas situées sur ces régions promotrices mais à plus d'1kb en aval, dans les régions codantes ou introniques des gènes. Parmi les 881 sondes associées à CINSARC, 91

Sonde	Localisation	Dist. îlot	Localisation de l'îlot	Gène	Dist. TSS
cg06373377	chr1:43360074	+686	chr1:43358463-43359388	*CDC20*	+1089
cg05525368	chr1:43360233	+845			+1248
cg00277165	chr5:912268	-111	chr5:912379-912746	*TRIP13*	+19624
cg01577755	chr5:912318	-61			+19674
cg14499678	chr5:912484	Dedans			+19840
cg11421768	chr5:912597	Dedans			+19953
cg17510385	chr5:912706	Dedans			+20062
cg12705693	chr5:912745	Dedans			+20101
cg17367077	chr5:160423146	+1040	chr5:160421772-160422106	*PTTG1*	+1217
cg05005217	chr8:25460122	+610	chr8:25457770-25459512	*CDCA2*	+989
cg03489186	chr15:89247613	+3153	chr15:89243767-89244460	*FANCI*	+3299
cg14398957	chr15:90992950	-1056	chr15:90994006-90994988	*PRC1*	+1668

Table 6 – Sondes de CINSARC différentiellement méthylées selon le pronostic de la signature. Les coordonnées génomiques sont indiquées selon l'annotation du génome hg38. Dist îlot : distance entre la sonde et l'îlot CpG. Dist TSS : distance moyenne entre la sonde et les sites d'initiation de la transcription (*transcription start site* ; TSS) des différents transcrits du gène.

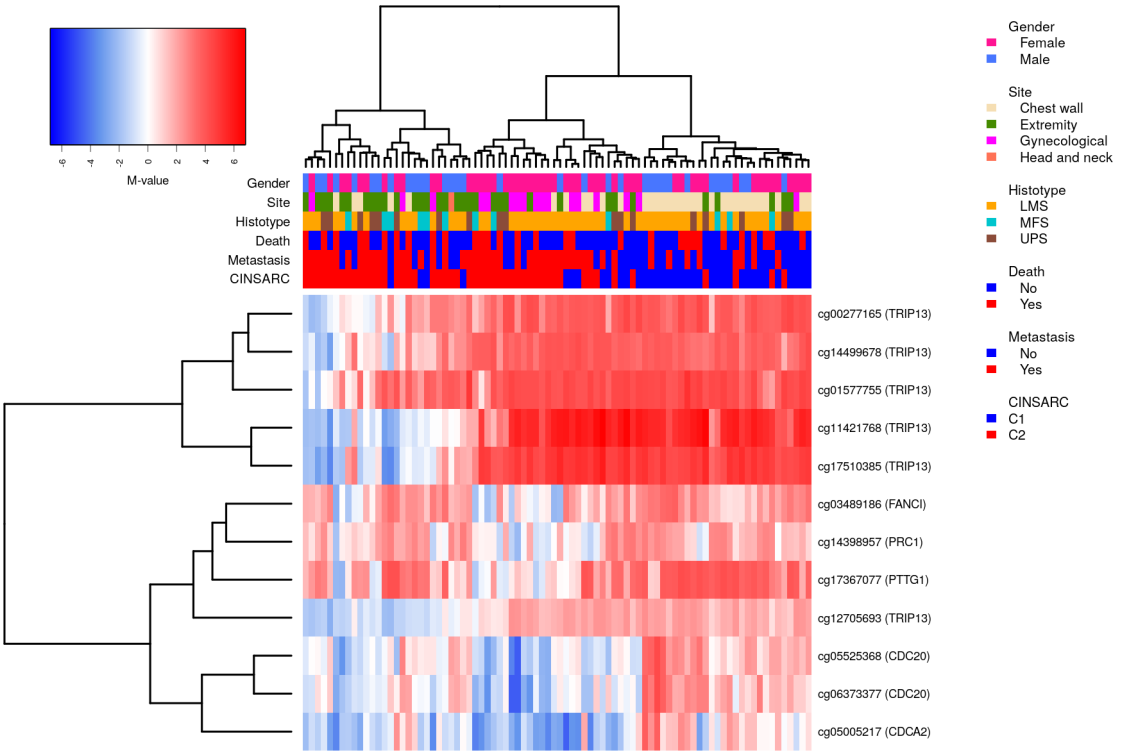


FIGURE 27 – Sondes de CINSARC différentiellement méthylées selon le pronostic de la signature. La *P* indiquée dans le texte correspond à l'enrichissement différentiel du statut CINSARC entre la branche de 28 tumeurs la plus à droite (sondes hyperméthylées) comparativement aux deux branches de gauche regroupant 56 tumeurs (sondes hypométhylées).

(9,68%) correspondent à ces six gènes et sont majoritairement non significatives (79 sur 91 ; 87%), incluant des sondes plus proches des TSS que celles significatives. Le nombre de sondes non significatives pour chacun des six gènes, informant aussi de leur distance par rapport aux TSS, est donné à la suite. *CDC20* : huit (dont sept à moins de 500pb des TSS), *TRIP13* : 24

(dont 20 à moins de 1500pb des TSS), *PTTG1* : neuf (dont huit à moins de 1000pb des TSS), *CDCA2* : 15 (dont dix à moins de 1000pb des TSS), *FANCI* : 11 (toutes à moins de 1000pb des TSS) et *PRC1* : 12 (dont dix à moins de 1000pb des TSS).

Le candidat le plus statistiquement associé au statut CINSARC est *TRIP13* dont l'îlot CpG chr5:912379-912746 est hyperméthylé en condition C1, information supportée par six sondes sur les neuf reliées à cet îlot. Cette méthylation différentielle affecterait l'intron 10 de *TRIP13* (sur 12 ; NM_004237), à une distance de 20kb par rapport au TSS de ce transcrit, ce qui serait une régulation atypique. Les différences de méthylation à l'échelle du génome sont majeures entre les statuts CINSARC C1 et C2. Néanmoins, les résultats obtenus ne sont pas en faveur d'une régulation directe de l'expression de la signature CINSARC par un niveau de méthylation différentielle. Le méthylome ne semble alors pas directement à l'origine de la sur-expression de la signature.

3.1.4.3 – Exploration des données de miRNA-seq

Puisque les miRNA sont des régulateurs négatifs de l'expression génique et dont certains ont été mis en cause dans l'initiation et la progression tumorale (**partie 1.2.4.2 page 65**), nous nous sommes intéressés à leur expression en tant que cause possible à la régulation d'expression de la signature CINSARC. Les données de miRNA-seq compilent des données d'expression de 1881 miRNA pour 82 tumeurs sur les 84 de la cohorte (les deux cas ainsi retirés sont un LMS utérin et un UPS de la jambe chez deux femmes, deux CINSARC C2). Le différentiel d'expression en fonction du statut CINSARC permet d'identifier 73 miRNA sur-exprimés dans le groupe C2 et trois miRNA sur-exprimés dans le groupe C1 (**figure 28A**). Par opposition, aucun miRNA n'est différentiellement exprimé selon le statut métastatique des tumeurs (**figure 28B**).

Afin de déterminer si ces miRNA (significativement conditionnellement exprimés) sont reconnus impliqués dans un processus oncogène, nous avons analysé les données bibliographiques qui les concernent, compilées avec miRCancer (Xie *et al.*, 2013). Les trois miRNA sur-exprimés en condition C1 (et donc sous-exprimés en condition C2 ; mir-143, mir-29c et mir-204) sont tous décrits comme ayant des propriétés anti-tumorales (diminution de la migration et/ou de la prolifération ; **figure 28C**). Par opposition, sur les 25 miRNA sur-exprimés en

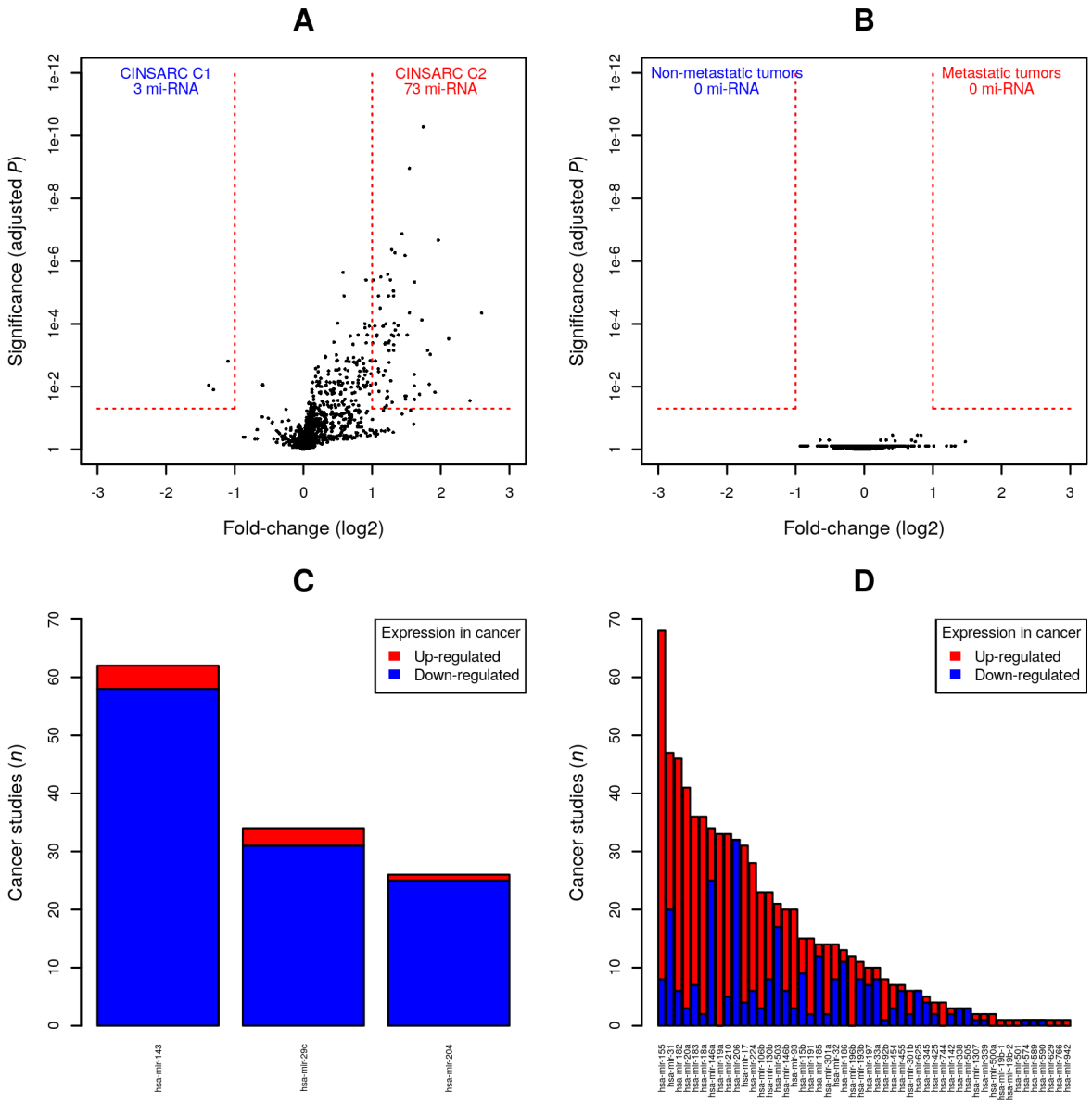


FIGURE 28 – Phénotype CINSARC et expression de miRNA. **A** et **B** : différentiel d'expression sur les 1881 miRNA annotés dans le TCGA en fonction du pronostic de la signature CINSARC (A) et du statut métastatique (B). **C** et **D** : données bibliographiques, obtenues avec miRCancer (Xie *et al.*, 2013), sur les miRNA sur-exprimés dans le groupe CINSARC C1 et CINSARC C2.

condition C2 et à plus de dix études répertoriées dans miRCancer, 17 (68%) sont majoritairement (>50% des études) connus pour leurs rôles pro-tumoraux (**figure 28D**).

Puisque la fonction des miRNA est une régulation négative d'expression post-transcriptionnelle par hybridation avec leurs ARN cibles, nous avons émis l'hypothèse que les trois miRNA sur-exprimés en condition C1 peuvent être une source de régulation négative de l'expression de CINSARC. Les validations expérimentales des interactions miRNA-ARN, compilées dans miRTarBase (Chou *et al.*, 2018), montrent que ces miRNA ont peu de cibles

connues envers la signature CINSARC. Mir-143 ne cible aucun gène de la signature. Mir-204 cible *CDCA3* (Whisnant *et al.*, 2013 ; Memczak *et al.*, 2013) et *FOXM1* (Sun *et al.*, 2015) mais l'expression de ce miRNA n'est pas significativement corrélée à l'expression de ces deux gènes (r=-0,25 et P=0,16 pour *CDCA3* ; r=-0,30 et P=0,09 pour *FOXM1*). Mir-29c cible *CCNA2* (Hafner *et al.*, 2010 ; Kishore *et al.*, 2011 ; Whisnant *et al.*, 2013 ; Memczak *et al.*, 2013) et l'expression de ce miRNA est significativement anti-corrélée à l'expression de ce gène (r=-0,38 et P=6,78e-3).

Nous avons alors exploré les interactions miRNA-ARN prédites, au cas où les données expérimentales n'aient pas été exhaustives pour nos gènes d'intérêt. Plusieurs outils dédiés à la prédiction d'interactions miRNA-ARN existent et se basent principalement sur les similarités d'appariements entre un miRNA et une cible potentielle. Chaque algorithme affine sa prédiction en intégrant des variables plus complexes comme la conservation du miRNA entre espèces ou encore la stabilité prédite du possible appariement. Nous avons donc interrogé la base miRGate qui intègre cinq de ces outils (RNAhybrid, MicroTar, Pita, TargetScan et miRanda ; Krüger & Rehmsmeier, 2006 ; Thadani & Tammi, 2006 ; Kertesz *et al.*, 2007 ; Friedman *et al.*, 2009 ; Betel *et al.*, 2010 ; Andrés-León *et al.*, 2015).

Selon ces prédictions (d'au moins un outil), mir-143, mir-204 et mir-29c ont respectivement 15 cibles possibles (*BIRC5*, *CDC6*, *CDK1*, *CENPL*, *FANCI*, *KIF11*, *KIF14*, *KIF4A*, *MCM2*, *MCM7*, *PBK*, *PRC1*, *SMC2*, *TOP2A* et *TPX2*), 14 (*AURKB*, *BUB1B*, *CCNA2*, *CDC6*, *CDC7*, *CDCA3*, *CKAP5*, *HP1BP3*, *KIF20A*, *KIF4A*, *PBK*, *RRM2*, *SMC2* et *TRIP13*) et 8 (*CCNA2*, *CDC7*, *ECT2*, *H2AFX*, *KIF4A*, *NDE1*, *TPX2* et *TRIP13*) parmi la signature CINSARC (les gènes soulignés seraient régulés par plusieurs miRNA). L'union de ces trois miRNA pourrait négativement réguler l'expression d'un collectif de 28 gènes de la signature (42%).

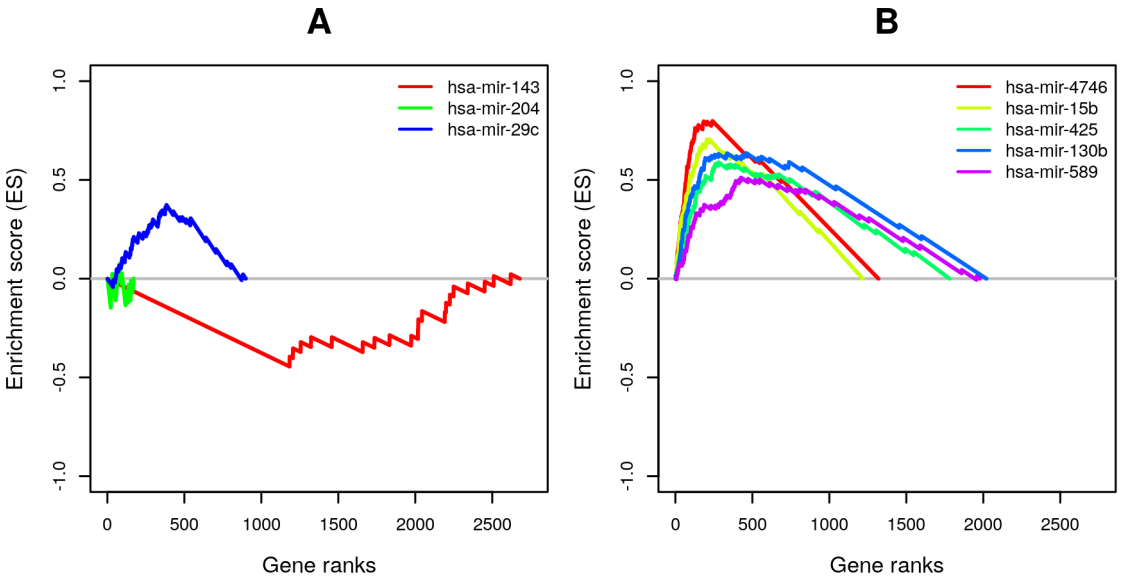
Nous avons alors cherché à établir un lien entre l'expression de ces trois miRNA et l'expression de gènes cibles par régression linéaire afin d'obtenir le niveau de corrélation (r de Pearson) mais également la valeur statistique de cette association (P ajustées par la méthode de Benjamini-Hochberg). Pour réaliser cette analyse, nous avons retenu 11626 gènes codants et exprimés dans la série des 82 tumeurs (expression moyenne >1 en *fragments per kilobase of transcript, per million mapped reads*) parmi l'ensemble des 55594 gènes du transcriptome.

Cette analyse permet d'identifier 2679 (dont 20 CINSARC), 169 (dont 19 CINSARC) et 899 (dont 54 CINSARC) gènes significativement anti-corrélés avec l'expression de mir-143, mir-204 et mir-29c.

Nous obtenons des relations d'anti-corrélations significatives entre expression de miRNA et expression d'ARN, dont une partie de ces gènes correspondent à la signature CINSARC. Nous devons donc déterminer si les gènes de CINSARC sont les mieux anti-corrélés parmi tous ces gènes, ou si aucune force d'anti-corrélation n'existe dans la série. Pour ce faire, nous avons utilisé une méthode d'enrichissement par rangs, analogue à l'algorithme implémenté dans *Gene Set Enrichment Analysis* (**figure 29A** ; Subramanian *et al.*, 2005). Pour mir-204, les expressions des 19 gènes de CINSARC ne font pas partie des gènes les mieux anti-corrélés parmi les 169 (P=2,55e-1). Pour mir-29c, nous avons observé un enrichissement significatif des 54 gènes de CINSARC, démontrant que l'expression d'une importante partie de la signature (81%) est significativement anti-corrélée à l'expression de ce miRNA (P=3,58e-6). Pour mir-143, le constat est inverse : les expressions des 20 gènes de CINSARC sont celles les moins anti-corrélées (mais toujours significatives) à celle du miRNA (P=3,44e-4).

Nous avons également mesuré des corrélations significatives entre l'expression de miRNA sur-exprimés en condition C2 et l'expression de nombreux gènes (**figure 29B**). Ainsi, sur le top 5 de ces miRNA (P ajustées décroissantes en différentiel d'expression), 1319 (dont les 67 CINSARC), 1217 (dont 66 CINSARC), 1781 (dont 62 CINSARC), 2020 (dont 64 CINSARC) et 1978 (dont 59 CINSARC) gènes sont significativement corrélés à l'expression de ces miRNA (respectivement : mir-4746, mir-15b, mir-425, mir-130b et mir-589). Les mêmes tests d'enrichissement ont été effectués pour déterminer si, parmi tous ces gènes, ceux qui composent CINSARC sont les plus corrélés à l'expression des miRNA. Les résultats montrent que ceci est bien le cas pour tous ces miRNA (P<1e-10).

Ces résultats indiquent que les tumeurs CINSARC C2 expriment une plus grande diversité de miRNA (n=73) que les tumeurs CINSARC C1 (n=3). La majorité de ces miRNA ont des rôles connus en faveur (groupe C2) ou défaveur (groupe C1) d'un message oncogène. Peu de données expérimentales obtenues sont en faveur d'une régulation négative d'expression de CINSARC par la sur-expression des miRNA que nous avons identifiés dans le groupe C1. Néanmoins, un



miRNA	# CINSARC	Max ES	Rank at max ES	*P*
hsa-mir-143	20 (29.85%)	-0.44	1183 (44.16%)	3.44e-04
hsa-mir-204	19 (28.36%)	-0.15	22 (13.02%)	2.55e-01
hsa-mir-29c	54 (80.6%)	0.37	384 (42.71%)	3.58e-06

miRNA	# CINSARC	Max ES	Rank at max ES	*P*
hsa-mir-4746	67 (100%)	0.80	241 (18.27%)	4.61e-35
hsa-mir-15b	66 (98.51%)	0.71	221 (18.16%)	3.67e-28
hsa-mir-425	62 (92.54%)	0.59	280 (15.72%)	1.12e-20
hsa-mir-130b	64 (95.52%)	0.64	461 (22.82%)	1.32e-26
hsa-mir-589	59 (88.06%)	0.51	426 (21.54%)	3.55e-15

FIGURE 29 – Association entre expression de miRNA et d'ARN. **A** : enrichissements de la signature CINSARC parmi les gènes significativement anti-corrélés à l'expression des trois miRNA sur-exprimés dans les tumeurs C1. **B** : enrichissements de la signature CINSARC parmi les gènes significativement corrélés à l'expression du top 5 (*P* ajustées décroissantes en différentiel d'expression) des miRNA sur-exprimés dans les tumeurs CINSARC C2. Les *P* indiquées sont calculées par la méthode du *mean-rank gene set enrichment test* (Michaud *et al.*, 2008). ES : score d'enrichissement.

possible régulateur pourrait être mir-29c dont l'expression est significativement anti-corrélée à celle de 54 gènes de CINSARC (81% de la signature) et, parmi les autres gènes anti-corrélés, ces 54 sont les plus associés à l'expression de ce miRNA.

3.1.4.4 – Conclusion

Nous avons en premier lieu confirmé les précédentes observations autour de la signature CINSARC sur la cohorte du TCGA : l'expression de la signature est un facteur pronostique (*P*=2,89e-3 ; **figure 24B page 140**) et est associée à une augmentation des altérations quantitatives (*P*=3,06e-3 ; **figure 24D page 140**). Les données disponibles ont apporté un nouveau regard en précisant d'autres facteurs associés à l'instabilité génétique. Les tumeurs CINSARC C2 ont plus souvent doublé leur génome (*P*=3,98e-2 ; **figure 25A page 142**) conduisant à une aug-

mentation générale de la ploïdie (P=1,87e-2 ; **figure 25B page 142**). Les altérations ponctuelles somatiques augmentent également dans le groupe de tumeurs CINSARC C2 (P=1,67e-2 ; **figure 25C page 142**). Toutes ces altérations participeraient à l'augmentation de l'hétérogénéité intra-tumorale, plus importante dans les tumeurs les plus agressives (P=5,91e-5 ; **figure 25D page 142**).

Ces associations statistiques entre doublement du génome, hétérogénéité et agressivité font écho à des résultats que nous avons obtenus sur deux lignées de sarcomes CINSARC C2 (provenant de deux UPS distincts ; Jemaà *et al.*, 2017 en **annexe III page 242**). Ces deux lignées sont principalement composées de cellules diploïdes mais également de sous-populations aneuploïdes/tétraploïdes. Les analyses expérimentales menées sur des clones diploïdes et tétraploïdes ont montré que la différence de ploïdie ne conférait pas d'avantage prolifératif mais était associée à une mobilité accrue chez les cellules tétraploïdes. De plus, alors que l'expression de la signature CINSARC était similaire entre clones diploïdes et tétraploïdes, l'expression protéique de la signature était nettement plus importante dans les cellules tétraploïdes. Ces données suggèreraient que, parmi les sous-clones présents dans la tumeur, les cellules tétraploïdes auraient un potentiel métastatique supérieur aux cellules diploïdes.

Nous cherchions à déterminer si le méthylome et l'expression de miRNA pouvaient être à l'origine de l'expression de la signature CINSARC. Si d'importantes différences à l'échelle du génome entier sont observées entre statuts C1 et C2 (P=1,24e-4 ; **figure 26A page 144**), le méthylome ne semble pas être directement responsable de l'expression de la signature car aucune région promotrice de l'un des gènes n'est conditionnellement méthylée. Concernant les miRNA, mir-29c pourrait être un bon candidat à la régulation négative d'expression de CINSARC, dont l'expression de 54 gènes est significativement anti-corrélée avec celle de ce miRNA.

L'ensemble de ces résultats nous a alors permis d'établir les cartes d'identité des tumeurs CINSARC C1 et C2 (**figure 30**). Ces données serviront ainsi de base pour de futurs travaux permettant de mieux comprendre les tenants et les aboutissants du remodelage génomique des sarcomes pléomorphes, notamment autour de l'expression de la signature CINSARC.

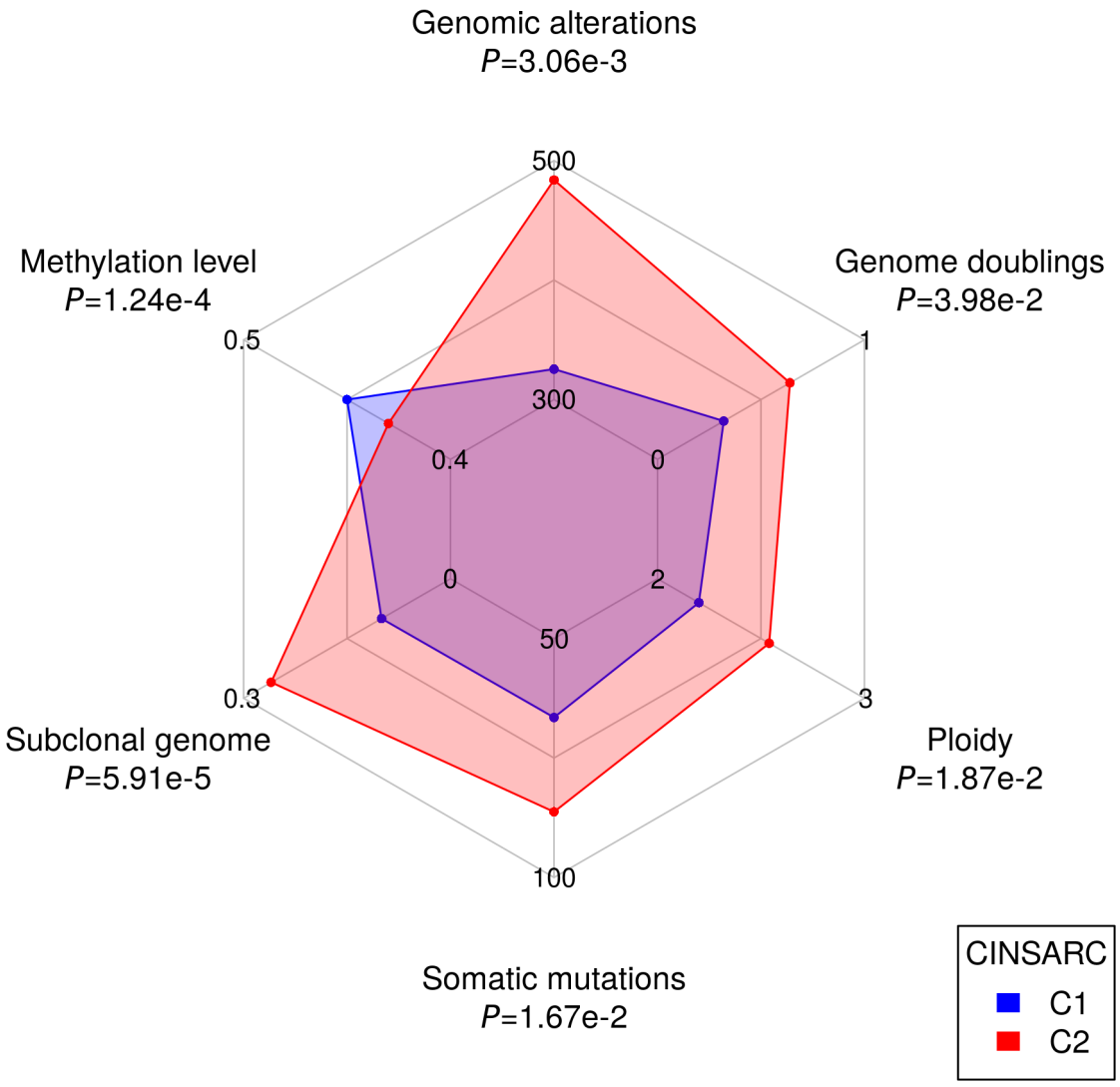


FIGURE 30 – Associations génomiques et pronostics de la signature CINSARC. Ce graphique représente les associations différentielles identifiées entre les tumeurs CINSARC C1 et C2. Les tumeurs exprimant fortement la signature ont significativement plus d'altérations quantitatives, de doublements génomiques, de copies de chromosomes, d'altérations ponctuelles et d'hétérogénéité intra-tumorale que les tumeurs C1. Les tumeurs CINSARC C1 ont des niveaux de méthylation augmentés comparativement aux tumeurs CINSARC C2.

3.2 – Expression de gènes de fusion dans les sarcomes pléomorphes

3.2.1 – Contexte

Au-delà du travail qui a été réalisé autour de la signature CINSARC, le RNA-seq constituait également une formidable opportunité de mieux définir les sarcomes pléomorphes car les altérations génétiques menant leur oncogenèse sont encore mal comprises. Jusqu'ici, les techniques de cytogénétique (notamment la FISH 24 couleurs et la CGH-*array*) ont pu mettre en évidence certains réarrangements structuraux dans ces tumeurs (**partie 1.2.2.3 page 41**). Ceci inclut les inactivations de *RB1* par délétion (dans 80% des cas ; Chibon *et al.*, 2000), de *TP53* par délétion ou mutation ponctuelle (dans 50% des cas ; Pérot *et al.*, 2010), de *PTEN* par délétion (50% des cas ; Gibault *et al.*, 2012), de *CDKN2A* par délétion (dans 30% des cas ; Pérot *et al.*, 2010) ou encore de *NF1* par délétion ou mutation ponctuelle (dans 10% des cas ; Barretina *et al.*, 2010). Les inactivations récurrentes de ces gènes suppresseurs de tumeurs sont bien décrites dans la littérature : elles ne sont pas spécifiques, sont partagées par d'autres types de cancers et notamment dans plusieurs sous-types histologiques de SGC (The Cancer Genome Atlas Research Network, 2017 ; Bailey *et al.*, 2018). Ces altérations ne peuvent donc pas, à elles seules, expliquer la diversité histologique de ces tumeurs. Ainsi, aucune altération récurrente et spécifique d'un sous-type histologique de SGC n'a pu être identifiée.

Par opposition, les SGS expriment un oncogène fort et spécifique, participant activement à leur transformation. Parmi les altérations identifiées dans ces tumeurs, la formation de gènes de fusion, principalement par translocation, est une mécanique chromosomique fréquemment rencontrée (20% des STM, couvrant plus d'une dizaine de sous-types histologiques ; **partie 1.2.2.2 page 37** ; Taylor *et al.*, 2011 ; Fletcher *et al.*, 2013 ; Mertens *et al.*, 2016). Elles permettent l'expression d'un transcrit chimérique, codant pour une protéine chimérique oncogène.

Dans les SGC, aucune étude n'a permis d'identifier la présence de tels gènes de fusion. Si de nombreuses translocations s'opèrent dans ces tumeurs (Mairal *et al.*, 2000), l'expression de gènes de fusion qui pourrait y être associée reste ainsi à déterminer. En effet, l'identification

de telles altérations est rendue difficile avec les techniques de cytogénétique précédemment employées. La FISH 24 couleurs permet d'identifier des translocations et des altérations quantitatives de grandes tailles (plusieurs dizaines de kilobases). La CGH-*array* renseigne, quant à elle, sur les altérations quantitatives (de quelques kilobases) mais n'informe pas sur la réorganisation génomique observée au sein de ces tumeurs. La résolution de ces deux méthodes ne permet donc pas d'identifier avec précision la présence d'un gène de fusion dans un génome aussi remanié que celui des sarcomes pléomorphes. La nature de la dérégulation transcriptomique qui s'opère suite à l'expression d'un gène de fusion n'est pas non plus évaluable par ces techniques.

La précédente constitution d'une cohorte de SGC passée en RNA-seq (Lesluyes *et al.*, 2016) était ainsi une formidable opportunité de pouvoir conduire cette étude. Les avancées technologiques successives en matière de séquençage haut débit ont en effet rendu possible une telle résolution d'analyse (Maher *et al.*, 2009), le RNA-seq étant à ce jour la meilleure technologie permettant d'identifier de nouveaux transcrits chimériques (avec plusieurs dizaines d'algorithmes dédiés ; https://omictools.com/gene-fusion-detection-category). Cette méthode a déjà permis de mieux comprendre la biologie de certains types histologiques de sarcomes en identifiant l'expression de nouveaux gènes de fusion et en les reliant à des dérégulations transcriptomiques spécifiques. On peut par exemple citer l'identification du gène de fusion *BCOR-CCNB3* dans des sarcomes d'Ewing non *EWSR1-FLI1* ou l'identification du gène de fusion *NAB2-STAT6* dans les tumeurs fibreuses solitaires (**partie 1.2.2.2 page 37** ; Pierron *et al.*, 2012 ; Robinson *et al.*, 2013).

Ces données n'ayant jamais été explorées à grande échelle dans les SGC, nous avons entrepris une caractérisation des gènes de fusion exprimés dans ces tumeurs. L'objectif de l'étude était alors de déterminer si un transcrit chimérique, codant pour une protéine chimérique oncogène, pouvait expliquer une partie de l'oncogenèse des SGC.

3.2.2 – Article 5 : *Recurrent TRIO fusion in nontranslocation–related sarcomas*

Delespaul L, Lesluyes T, Pérot G, Brulard C, Lartigue L, Baud J, Lagarde P, Le Guellec S, Neuville A, Terrier P, Vince-Ranchère D, Schmidt S, Debant A, Coindre JM & Chibon F. Article publié dans *Clinical Cancer Research* (2017).

Résumé de l'article

Afin de caractériser les transcrits chimériques exprimés dans les SGC, nous avons réalisé le séquençage ARN d'une cohorte de 112 tumeurs primaires (cas avec de l'ARN en réserve pour aller à la validation expérimentale parmi la cohorte de Lesluyes *et al.*, 2016) et de cinq lignées. Sur ces données de séquençage, l'algorithme deFuse (McPherson *et al.*, 2011) nous a permis de détecter l'ensemble des transcrits chimériques exprimés dans ces échantillons. Les différentes étapes de filtrage (élimination des *read-throughs*, conservation du cadre de lecture et validation en séquençage Sanger) ont alors menées à l'identification d'un seul gène de fusion récurrent exprimé dans la cohorte : *TRIO-TERT*. Le même résultat fut obtenu avec deux autres outils : TopHat-Fusion (Kim & Salzberg, 2011) et ChimeraScan (Iyer *et al.*, 2011).

Par cette analyse, trois cas ont été identifiés présentant l'expression de gènes de fusion impliquant la fusion de *TRIO* (exons 33 ou 34) avec le gène *TERT*. Ces cas expriment chacun deux transcrits chimériques par épissage alternatif de *TERT* (au niveau de l'exon 2), l'un conservant le cadre de lecture et l'autre ne le conservant pas. Un cas supplémentaire fut identifié présentant la fusion entre les gènes *TRIO*, *CDH18* et *TERT*, également avec perte du cadre de lecture. Nous avons alors recherché, puisque les domaines de *TERT* ne sont pas conservés dans la plupart des transcrits chimériques, si *TRIO* pouvait être fusionné avec d'autres partenaires. Ceci nous a amenés à identifier deux autres cas présentant des réarrangements de *TRIO*, impliquant toujours les exons 33 et 34, avec respectivement les gènes *LINC01504* et *ZNF558*.

L'expression de ces transcrits est spécifique aux SGC car ils n'ont pas été détectés dans des SGS (testé expérimentalement dans des GIST et des synovialosarcomes) ni dans d'autres types de tumeurs (données bibliographiques). L'analyse du transcriptome des tumeurs qui ex-

priment un *TRIO* tronqué, comparativement aux autres, a montré une sur-expression de gènes impliqués dans l'inflammation, la prolifération et le cycle cellulaire. Nous avons alors expérimentalement démontré que l'inhibition de l'expression du transcrit chimérique *TRIO-TERT*, dans une lignée d'UPS, diminue la prolifération de façon significative.

L'ensemble des résultats de cette étude sont inclus ci-après (Delespaul *et al.*, 2017). Les tables additionnelles 5 et 6, trop volumineuses pour être incluses, sont accessibles en ligne (https://doi.org/10.1158/1078-0432.CCR-16-0290).

Biology of Human Tumors

Clinical Cancer Research

Recurrent *TRIO* Fusion in Nontranslocation–Related Sarcomas

Lucile Delespaul[1,2], Tom Lesluyes[1,2], Gaëlle Pérot[1,3], Céline Brulard[1], Lydia Lartigue[1,2], Jessica Baud[1,2], Pauline Lagarde[1], Sophie Le Guellec[4], Agnès Neuville[1,3], Philippe Terrier[5], Dominique Vince-Ranchère[6], Susanne Schmidt[7], Anne Debant[7], Jean-Michel Coindre[1,2,3], and Frédéric Chibon[1,3]

Abstract

Purpose: Despite various differences, nontranslocation-related sarcomas (e.g., comprising undifferentiated pleomorphic sarcoma, leiomyosarcoma, myxofibrosarcoma) are unified by their complex genetics. Extensive analysis of the tumor genome using molecular cytogenetic approaches showed many chromosomal gains, losses, and translocations per cell. Genomic quantitative alterations and expression variations have been extensively studied by adapted high-throughput approaches, yet translocations still remained unscreened. We therefore analyzed 117 nontranslocation-related sarcomas by RNA sequencing to identify fusion genes.

Experimental design: We performed RNA sequencing and applied a bioinformatics pipeline dedicated to the detection of fusion transcripts. RT-PCR and Sanger sequencing were then applied to validate predictions and to search for recurrence and specificity.

Results: Among the 6,772 predicted fusion genes, 420 were in-frame. One recurrent rearrangement, consistently involving *TRIO* with various partners, was identified in 5.1% of cases. *TRIO* translocations are either intrachromosomal with *TERT* or interchromosomal with *LINC01504* or *ZNF558*. Our results suggest that all translocations led to a truncated TRIO protein either directly or indirectly by alternative splicing. *TRIO* rearrangement is associated with a modified transcriptomic program to immunity/inflammation, proliferation and migration, and an increase in proliferation.

Conclusions: *TRIO* fusions have been identified in four different sarcoma histotypes, likely meaning that they are not related to a primary oncogenic event but rather to a secondary one implicated in tumor progression. Moreover, they appear to be specific to nontranslocation-related sarcomas, as no such rearrangement was identified in sarcomas with simple genetics. More cases could lead to a significant association of these fusions to a specific clinical behavior. *Clin Cancer Res; 23(3); 857–67. ©2016 AACR.*

Introduction

Sarcomas represent a heterogeneous group of rare tumors accounting for approximately 1% of adult cancers with more than 50 histologic subtypes. These tumors, derived from mesenchymal tissues (i.e., soft tissue, bone, or muscle), are classified on the basis of their line of differentiation, following the World Health Organization (WHO) recommendations (1). Along with this histologic heterogeneity, sarcoma genetics is also highly heterogeneous in terms of oncogene-specific driver alterations and whole-genome instability. Sarcomas can be classified into two groups depending on genome stability. One group consists of sarcomas with a relatively stable genome with specific oncogenic alterations (rare gains and losses involving mainly full chromosomes). Among these, gastrointestinal stromal tumors (GIST; *KIT* or *PDGFRA* mutations; ref. 2) and sarcomas with a specific translocation (e.g., Ewing sarcoma, synovial sarcoma, and alveolar rhabdomyosarcoma; refs. 3–5) represent 20% and 30% of sarcomas, respectively. The second group consists of nontranslocation-related sarcomas with higher chromosomal instabilities and pleomorphic histologic patterns. There is no specific genetic alteration described so far in those sarcomas. Nevertheless, we can distinguish well-differentiated and dedifferentiated liposarcoma (20% of sarcomas) with recurrent *MDM2* and *CDK4* amplification within a moderately rearranged profile (6, 7). The remaining 30% mainly comprise leiomyosarcoma, undifferentiated pleomorphic sarcoma, myxofibrosarcoma, pleomorphic liposarcoma, and pleomorphic rhabdomyosarcoma. Although their genomic instability has not been deciphered yet, some recurring genetic alterations have been identified in these tumors: the 13q14-21 region is frequently lost in leiomyosarcomas and undifferentiated pleomorphic sarcoma targeting *RB1* deletion/inactivation (8); the TP53 pathway is consistently inactivated, mainly by *TP53* deletions/mutations (9); and a gain of 5p region is also frequent and triggers *TRIO* amplification (10, 11). Extensive

[1]Inserm U916, Bergonié Cancer Institute, Bordeaux, France. [2]University of Bordeaux, Bordeaux, France. [3]Department of Pathology, Bergonié Cancer Institute, Bordeaux, France. [4]Department of Pathology, University Cancer institute Toulouse-Oncopole, Toulouse, France. [5]Department of Pathology, Gustave Roussy Institute, Villejuif, France. [6]Department of Pathology, Léon Bérard Center, Lyon, France. [7]Centre de Recherche en Biologie Cellulaire de Montpellier, CRBM, CNRS-UMR 5237, University of Montpellier, Montpellier, France.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Corresponding Author: Frédéric Chibon, Institut Bergonié, 229 Cours de l'Argonne, Bordeaux 33000, France. Phone: 055-633-4443; Fax: 055-633-4438; E-mail: F.chibon@bordeaux.unicancer.fr

doi: 10.1158/1078-0432.CCR-16-0290

Translational Relevance

Nontranslocation-related sarcoma is one of the most common and aggressive subtype of soft tissue sarcomas. Despite extensive quantitative and structural characterization of their genome and transcriptome, it remains the largest sarcoma group without any specific genetic alteration that can be targeted for therapy. Therefore, treatment is still based on surgery with wide margins. Previous high-throughput studies based on quantitative analysis failed to identify such specific alteration. We thus applied RNA sequencing on a large cohort of nontranslocation-related sarcomas from the French Sarcoma Group and identified recurrent, and so far specific, *TRIO* fusions with various partners. Fusion-positive tumors belong to four different histotypes, likely meaning that fusions are not related to a specific primary oncogenic event but rather to a secondary one implicated in tumor progression. Our data suggest that fusion triggers a modified transcriptomic and phenotypic program that could lead to an increase in cell proliferation.

analysis of the tumor genome by molecular cytogenetic approaches showed many breakpoints [array comparative genomic hybridization (aCGH)] and translocations per cell (chromosome painting; refs. 12, 13). Breakpoints located in coding regions are supposed to produce chimeric transcripts with potent oncogenic effects. It has been demonstrated that RNA sequencing (RNA-seq) is an efficient approach to identify fusion transcripts in tumors, particularly in sarcomas (14–16). Hence, we subjected 117 nontranslocation-related sarcomas to RNA-seq to identify fusion transcripts.

Materials and Methods

Ethics statement

The samples used in this study are part of the Biological Resources Center of Institut Bergonié (CRB-IB; Bordeaux, France). In accordance with the French Public Health Code (articles L. 1243-4 and R. 1243-61), the CRB-IB has received the agreement from the French authorities to deliver samples for scientific research (number AC-2008-812). The samples come from care and are requalified for research. The project was approved by the Institut Bergonié ethics committee (scientific advisory board). Every case was histologically reviewed by the pathologist subgroup from the French Sarcoma Group and classified according to the 2013 WHO classification by histology, IHC, and molecular genetics and cytogenetics when needed.

Sample description

The screened series was composed of 112 sarcomas with complex genomics and five cell lines (CL; CL1 derived from an undifferentiated pleomorphic sarcoma T1; CL2 derived from a dedifferentiated liposarcoma; CL3 derived from a myxofibrosarcoma; and CL4 and CL5 derived from two leiomyosarcomas) for the first cohort.

The second and the third cohorts are composed of 27 synovial sarcomas and 74 GIST, respectively. All tumors have been histologically reviewed by a pathologist panel from the French Sarcoma Group.

Cell line establishment and culture

Cell line establishment was performed as described previously (17). Authentication of cell line is analyzed by aCGH by comparing the cell line with the corresponding tumor origin.

The culture medium was composed of RPMI1640 with GlutaMAX (Gibco BRL, Life Technologies) supplemented with 10% FCS and 1% antibiotics (penicillin/streptomycin, Gibco BRL, Life Technologies). All cell lines were tested for mycoplasma by PCR (Sigma; LookOut Mycoplasma PCR Detection Kit) according to the manufacturer's recommendations.

RNA extraction, sample preparation, and RNA-seq

RNA extraction from frozen and formalin-fixed paraffin-embedded (FFPE) samples, sample preparation, and RNA-seq were performed as described previously (18).

Bioinformatics pipeline for fusion detection and gene expression

Bioinformatics analysis was performed as described previously (18).

Briefly, deFuse (v0.6.1; ref. 19) was used on curated FastQ files with ENSEMBL GRCh37.74 annotations. We kept candidate fusions with the following criteria: probability >50%, non-read-through, non-read-through like (same chromosome, indicated as deletion, coding/coding gene region, and genomic distance <20,000) and breakpoint homology <30.

TopHat-Fusion (v2.1.0; ref. 20) was used on curated FastQ files with the same annotations as described by Lesluyes et al. in 2016 (18). Fusion minimum distance was set to 10,000, and fusion anchor length was set to 13. Other parameters were similar to the expression processing.

ChimeraScan (v0.4.5; ref. 21) algorithm was used on raw FastQ files (equal-length reads restriction) with UCSC (fixed on 2016/02/09) annotations. Multiple hits parameter was set to 50. We kept fusions with the following criteria: non-read-through, score >5, spanning reads >2, and non-HLA/HLA fusions (due to sequence homology).

For the three detection algorithms, duplicated entries (same genes) and non-ORF conservation fusions were discarded.

Unsupervised clustering method was performed with Wald algorithm and Euclidean distance available in the *cluster* R package (v2.0.3). Differential gene expression analysis was performed with *DESeq2* R package (1.8.1; ref. 22). Gene interaction was performed using BioGRID (23). Gene Set Enrichment Analysis (GSEA; v2.2.0; refs. 24, 25) was used with Hallmarks gene sets from MSigDB (v5.0), gene set permutation type, RefSeq chip platform without gene symbol collapse, and a minimum set size of 10. Differentially exonic expression was performed by DEXSeq (6 vs. 40; ref. 26).

RT-PCR and Sanger sequencing

Total RNA was reverse transcribed using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's instructions.

For PCR, primers were designed using Primer3 program (http://frodo.wi.mit.edu/primer3/) and are presented in Supplementary Table S1. Touchdown 60°C program has been used (TD 60°C; two cycles at 60°C, followed by two cycles at 59°C, two cycles at 58°C, three cycles at 57°C, three cycles at 56°C, four cycles at 55°C, four cycles at 54°C, five cycles at 53°C, and finally 10 cycles at 52°C). PCR was performed on 50 ng of cDNA using AmpliTaqGold DNA

Polymerase (Applied Biosystems). PCR products were then purified using ExoSAP-IT PCR Purification Kit (GE Healthcare), and sequencing reactions were performed with the Big Dye Terminator V1.1 Kit (Applied Biosystems) according to the manufacturer's recommendations. Samples were then purified using the Big Dye XTerminator Purification Kit (Applied Biosystems) according to the manufacturer's instructions, and sequencing was performed on a 3130xl Genetic Analyzer (Applied Biosystems). Sequence analysis was performed with SeqScape software v2.5 (Applied Biosystems).

Quantitative RT-PCR

Expression of different *TRIO-TERT* and *TRIO* transcripts was studied with probes and primers listed in Supplementary Table S2. Normalization of expression was carried out with two reference genes: *ACTB* and *RPLP0*. Probes and primers (TaqMan Gene Expression Assay, Life Technologies) of these genes are Hs99999903_m1 and Hs99999902_m1, respectively. Reaction was performed with TaqMan Universal PCR Master Mix (Life Technologies) using StepOnePlus (Life Technologies) according to the manufacturer's recommendations. Expression was calculated with the formula: $\Delta C_t = C_{t_{TRIO-TERT}} - C_{t_{\text{reference gene}}}$ (27). Subsequently, difference of expression between both transcripts was calculated with the formula $2^{\Delta C_{t_{\text{transcript 1}-\text{transcript 2}}}}$.

FISH

FISH assay on FFPE cases T1, S867, and S909 was performed using the Histology FISH Accessory Kit (Dako) as described previously (28) on a 4-μm paraffin-embedded section. Regarding metaphase chromosome spreading on CL1, cells were incubated with colchicine overnight and resuspended in isotonic KCl buffer. Cells were finally fixed in standard 3:1 methanol:acetic acid final fixative. After several washes, cells were spread on a glass slide. Denaturation (5 minutes at 82°C for paraffin-embedded tissues and 1 minute at 73°C for chromosome spreading) and hybridization (overnight at 37°C) were achieved by placing the slides into a hybridizer (Dako). For the identification of the *TRIO* fusion, one BAC clone covering 5′ part of *TRIO* (RP11-1134M22; red signal) and two BAC clones covering 3′ part of *TRIO* (RP11-81P9, RP11-1079G4; green signals) were used; green and red fluorescent signals were analyzed in tumors using a Nikon Eclipse 80i fluorescence microscope with appropriate filters. Pictures were captured using a Hamamatsu C4742-95 CCD camera and analyzed with the Genikon software. A *TRIO* rearrangement was detected when red and green signals were separated in the nucleus. A *TRIO* rearrangement was considered as present if at least 10% of tumor cells showed a rearrangement pattern. Unbalanced rearrangements (fusion signal associated with at least one supernumerary red signal) are considered as positive case when chimeric transcript has been validated by RT-PCR and Sanger sequencing and with at least 10% rearranged signals.

Affymetrix CytoScan HD Array

Genomic DNA was isolated using a standard phenol–chloroform extraction protocol. Affymetrix CytoScan HD Array (Affymetrix) was used according to the manufacturer's instructions. The test was conducted on DNA samples from *TRIO* translocated cases. CEL files obtained by scanning the CytoScan arrays were analyzed using the Chromosome Analysis Suite software (Affymetrix) and the annotations of the genome version GRCH37 (hg19).

Immunofluorescence

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 (DAKO), for 20 minutes in a microwave oven. The primary antibodies and dilutions (dilution in DAKO REAL antibody diluent, DAKO) used in this study are as follows: anti-PML (PG-M3, 1:200, sc-966, Santa Cruz Biotechnology) and anti-TERF2 antibody (1:200, HPA001907, Sigma). All primary antibodies were incubated for 1 hour at room temperature. Secondary antibodies and dilutions used are as follows: anti-mouse immunoglobulins/FITC (1/400, Dako) and anti-rabbit IgG (H+L) Alexa Fluor 594 conjugate (1/500, Invitrogen). Slides were mounted with VECTASHIELD/DAPI medium (Vector Laboratories) and were then analyzed using a Nikon Eclipse 80i fluorescent microscope with appropriate filters. Pictures were captured using a Hamamatsu C4742-95 CCD camera.

Design and infection of *TRIO-TERT* shRNA

Short hairpin RNA (shRNA) sequences targeting $TRIO_{33}$-$TERT_2$ and a nontargeting control sequence (Supplementary Table S3) were cloned into the doxycycline-inducible pLKO.1-neo lentiviral shRNA vector (Addgene). VSV-G–pseudotyped lentiviral particles were produced by cotransfection of 293T cells with pLKO.1 constructs and the compatible packaging plasmids psPAX2 and pVSVg. Cell lines were incubated overnight with lentiviral supernatants in the presence of 8 μg/mL polybrene (Sigma H9268), and stably transduced cells were selected with neomycin (0.1–1 mg/mL, Sigma 4727894001) for 10 days. Downregulation of $TRIO_{33}$-$TERT_{2/3}$ mRNA expression was controlled by qRT-PCR.

Proliferation assay

Proliferation measurement was performed using the IncuCyte Live Cell Imaging System (Essen BioScience). CL1s were seeded at 3,000 cells per well in a 96-well plate with eight replicates for each conditions. Cells were pretreated with or without doxycycline (1 μg/mL, Sigma D3072) for 48 hours before proliferation analysis. Treatment was maintained during 5 days of experiment. Proliferation was monitored according to the manufacturer's recommendations (two pictures every 2 hours for each well).

Data access

RNA-seq raw files (FastQ) are available on Sequence Read Archive under the accession numbers SRP057793 (contains 112 tumors) and SRP059536 (contains the five sarcoma cell lines: CL1 is M965, CL2 is M969, CL3 is M961, CL4 is M963, and CL5 is M964).

Expression data for the 117 samples are available on Gene Expression Omnibus under accession number GSE75885.

Results

RNA-seq analysis and fusion prediction

To identify fusion transcripts, we performed RNA-seq on 112 sarcomas with complex genetics (Table 1) and 5 sarcomas cell lines. Means of 78.34 (± 18.71) and 74.97 (± 17.83) million reads per sample were obtained after library sequencing following low quality 5′ and 3′ trims and overlap correction, respectively. DeFuse (19) analysis predicted a total of 10,382 fusion transcripts for all samples with a fusion mean of 88.74 (± 52.7) per sample

Table 1. Nontranslocation-related sarcomas cohort description ($N = 112$)

Characteristics	Cohort ($N = 112$)
Median follow-up (years)	2.94 (95% CI, 2.11–3.77)
Median age (years)	64.5 (95% CI, 62–67)
Sex	
Female	56 (50%)
Male	56 (50%)
FNCLCC grade (%)	
1	3 (2.68%)
2	29 (25.89%)
3	76 (67.86%)
Unknown	4 (3.57%)
Histologic type (%)	
UPS	36 (32.14%)
LMS	33 (29.46%)
DDLPS	18 (16.07%)
MFS	15 (13.39%)
PLPS	4 (3.57%)
PRMS	4 (3.57%)
Other	2 (1.79%)
Location	
Lower limb	45 (40.18%)
Internal trunk	25 (22.32%)
Trunk wall	18 (16.07%)
Upper limb	16 (14.29%)
GI tract	3 (2.68%)
Gynecologic area	3 (2.68%)
Head and neck	2 (1.79%)
Relapse events (%)	
Metastasis	38 (33.93%)
Local recurrences	34 (30.36%)
Median size of tumor (mm; $n = 109$)	100 (95% CI, 80–100)

Abbreviations: CI, confidence interval; DDLPS, dedifferentiated liposarcoma; FNCLCC, *Fédération Française des Centres de Lutte Contre le Cancer*; GI, gastrointestinal; LMS, leiomyosarcoma; MFS, myxofibrosarcoma; PLPS, pleomorphic liposarcoma; PRMS, pleomorphic rhabdomyosarcoma; UPS, undifferentiated pleomorphic sarcoma.

(Fig. 1). Among them, 3,610 detected chimeric transcripts were actually read-throughs (chimeric transcripts between two adjacent genes and induced by transcription; ref. 29) according to deFuse and were consequently discarded from further processing. Following this selection, 6,772 potential fusion transcripts were kept. To identify in-frame fusion transcripts, we filtered this selection and focused on the 420 in-frame predicted fusions with conserved open reading frame (ORF) for each partner. Among these transcripts, eight presented a potential recurrence: *CTSC-RAB38* (13 samples), *GSE1-RP11-680G10.1* (12 samples), *NSUN4-FAAH* (8 samples), *CTBS-GNG5P2* (8 samples), *TRIO-TERT* (4 samples), *FARSA-SYCE2* (2 samples), *RPL11-TCEB3* (2 samples), and *TFG-GPR128* (2 samples). Among these eight transcripts, five are known to be read-throughs but evaded the deFuse filter: *CTSC-RAB38* and *TFG-GPR128* have been reported as read-throughs (30, 31), whereas *CTBS-GNG5P2*, *RPL11-TCEB3*, and *NSUN4-FAAH* are referenced in the ConjoinG database (32), which contains detailed information about identified read-throughs in the human transcriptome. *FARSA-SYCE2* and *GSE1-RP11-680G10.1* fusions have been identified in normal human tissues (RT-PCR and Sanger sequencing; Supplementary Fig. S1) and the two genes implicated in the fusion are on the same chromosome and transcribed on the same genomic strand. As they share similar characteristics as read-throughs, we discarded them from further analysis. The last potential in-frame recurrent chimeric transcript predicted by deFuse, $TRIO_{33/34}$-$TERT_{2/3}$ (fusion between *TRIO* either exon 33 or 34 and either *TERT* exon 2 or 3), occurred in four samples (CL1, S867, S829, and S822) and has not been detected in normal tissues (Fig. 2A and B). In addition, *TRIO-TERT* gene fusion has been validated by RT-PCR and Sanger sequencing in T1, the primary tumor of CL1 (Fig. 2A). No reciprocal transcript has been detected for these fusions (data not shown). Same

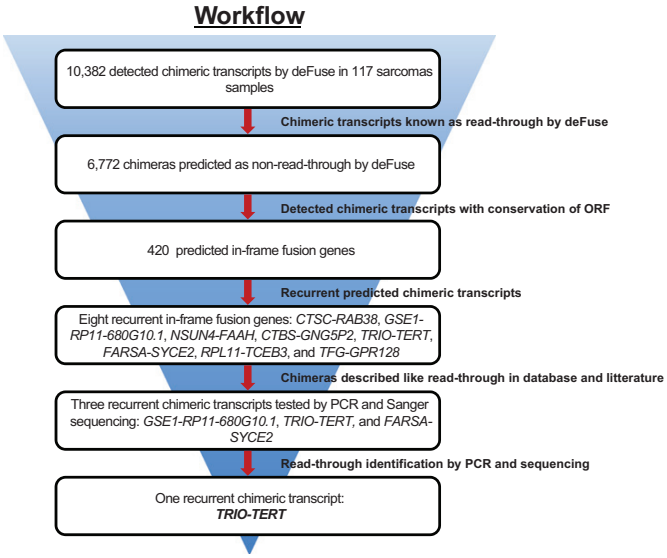


Figure 1.
Candidate fusion gene selection.

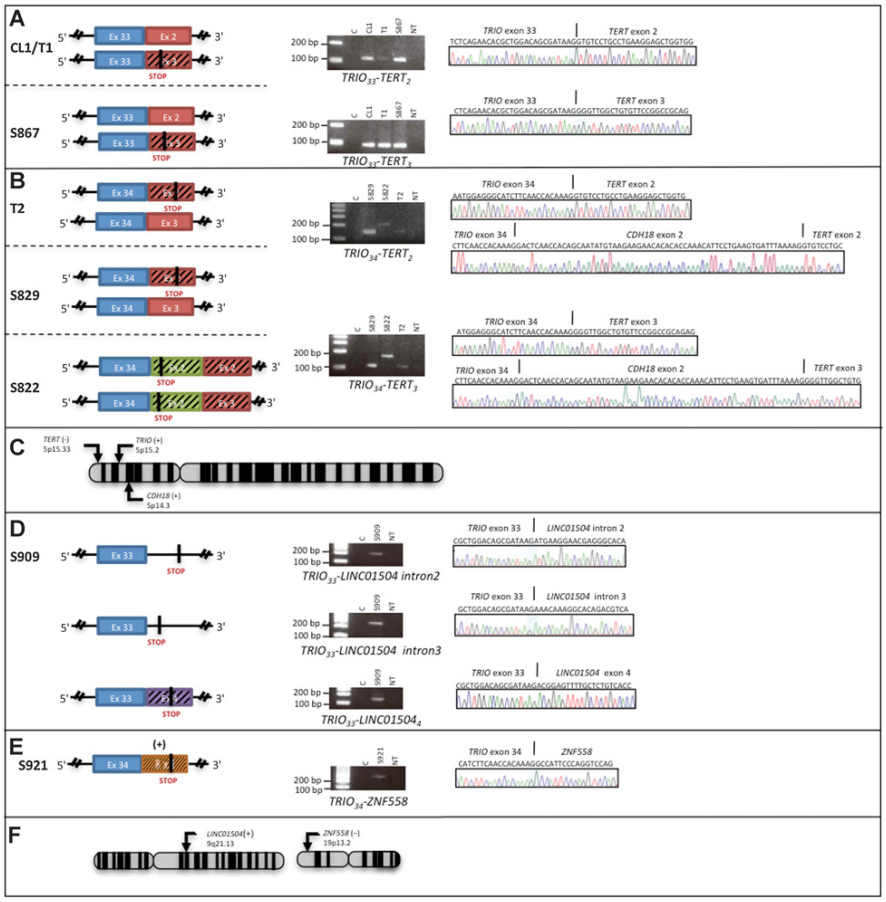


Figure 2.
Identification of seven fused *TRIO* cases. RT-PCR and chromatograms obtained with the four exon combinations involving *TRIO, TERT, LINC01504*, and *ZNF558* genes. **A** and **B,** Cases with $TRIO_{33}$-$TERT_{2/3}$ (**A**), $TRIO_{34}$-$TERT_{2/3}$, $TRIO_{34}$-*CDH18*-$TERT_{2/3}$ (**B**) expressions. **C,** Genomic location of *TRIO, TERT*, and *CDH18*. **D** and **E,** $TRIO_{33}$-*LINC01504* (**D**) and $TRIO_{34}$-*ZNF558* (**E**) expressions. **F,** Genomic location of *LINC01504* and *ZNF558*. **C,** PCR control; NT, normal tissue; *bp*, base pairs. Vertical black lines above chromatograms indicate sequence fusion between the two genes. All chimeric transcripts are represented by scheme for each case. Out-of-frame exons are indicated by black hatches. New STOP codons are indicated by a vertical black line on the scheme. *TRIO, TERT, CDH18, ZNF558*, and *LINC01504* exons are indicated in blue, red, green, orange, and purple respectively. + and − symbols (brackets), genomic strands.

analyses have been performed with two supplementary algorithms, TopHat-Fusion (20) and ChimerScan (21). *TRIO-TERT* is also the only recurrent in-frame chimeric transcript detected with these algorithms.

Consequently, *TRIO-TERT* fusion is the only recurring in-frame fusion identified in this cohort.

What is the genomic alteration that results in the *TRIO-TERT* fusion transcript?

aCGH analysis of three translocated cases (T1, S829, and S822) with high quality DNA available demonstrated that the *TRIO* (5p15.2) and *TERT* (5p15.33) loci (Fig. 2C) were rearranged (gain and/or amplification) in two of three cases (e.g., presented in

Supplementary Fig. S2A and S2B). Interestingly, in the *TRIO* gene, the copy number variation break is located in the region of fusion. We therefore hypothesized that the frequently observed chromosome 5p/*TRIO* amplification in sarcomas (11) could be associated, at least in some cases, to *TRIO* rearrangements. We thus reexamined the genomic profile of the 112 sarcomas tumors of RNA-seq cohort and 24 supplementary cases of nontranslocation-related sarcomas (Supplementary Table S4). Three cases presented a comparable *TRIO* amplification profile (T2, S909, and S921). Specifically, in T2, $TRIO_{34}$-$TERT_3$ fusion was detected by RT-PCR and Sanger sequencing (Fig. 2B). The two new remaining cases with *TRIO* amplification did not express *TRIO-TERT* gene fusion. We performed RNA-seq on these two additional samples, and deFuse analysis indicated that they both expressed a fusion gene involving *TRIO*, but not *TERT*. In sample S909, $TRIO_{33}$ is fused to *LINC01504*, and in sample S921, $TRIO_{34}$ is fused to *ZNF558*; both were subsequently validated by RT-PCR and Sanger sequencing (Fig. 2D–F).

To validate the genomic origin of the fusion, we performed FISH analysis in four cases with available and good enough quality FFPE blocks. A break-apart approach targeting *TRIO* validated rearrangement in the three interpretable cases (CL1/T1, S867, and S909) with 84%, 36%, and 58% of positive cells, respectively (Fig. 3).

Analysis of *TRIO* fusion isoforms

In all cases, regardless of the partner, either *TRIO* exon 33 or exon 34 is fused. In five cases, fusion joined either *TRIO* exon 33 ($TRIO_{33}$-*TERT*: T1 and S867) or exon 34 ($TRIO_{34}$-*TERT*: cases S829, S822, and T2) to *TERT*, an exception being case S822, in which *CDH18* exon 2 is inserted in between *TRIO* exon 34 and *TERT*. Whatever the *TRIO* (or *CDH18* for case S822) exon, fusions occurs either with *TERT* exon 2 or exon 3 (*TRIO*-$TERT_2$ or *TRIO*-$TERT_3$).

As it has been reported that *TERT* exon 2 is subjected to alternative splicing (33), the expression of both isoforms should be due to alternative splicing instead of two translocation events (a much rarer genomic event).

Among these six distinct chimeric transcripts of the *TRIO-TERT* fusion, four ($TRIO_{33}$-$TERT_3$, $TRIO_{34}$-$TERT_2$, $TRIO_{34}$-$CDH18_2$-$TERT_2$, and $TRIO_{34}$-$CDH18_2$-$TERT_3$) did not conserve the *TERT* ORF with a stop codon after the last fused *TRIO* exon at 77, 353, 23, and 23 bp, respectively (Fig. 2A and B).

Given that all cases expressed both isoforms ($TRIO_{33/34}$-$TERT_2$ and $TRIO_{33/34}$-$TERT_3$), we sought to determine which isoform is preferentially expressed. Isoform-specific quantitative RT-PCR analysis showed that chimeric transcripts with *TERT* exon 3 were at least 15 times more expressed regardless the *TERT* ORF conservation (Supplementary Fig. S3).

In case S909, *TRIO* exon 33 is fused to *LINC01504*, leading to the expression of three isoforms: $TRIO_{33}$-$LINC01504_{intron2}$, $TRIO_{33}$-$LINC01504_{intron3}$, or to $TRIO_{33}$-$LINC01504_4$, resulting in the loss of the ORF after the breakpoint with a stop codon after the last *TRIO* exon at 146, 84, and 113 bp, respectively (Fig. 2D). Different transcripts observed in S909 could be explained by alternative splicing mechanisms as sequences near the fusion break-points in *LINC01504* introns 2 and 3 are predicted to be splicing sequences by Human Splicing Finder (34) with 66.9% and 74% of consensus value, respectively (Supplementary Fig. S4).

In case S921, *TRIO* exon 34 is fused to *ZNF558* exon 8 ($TRIO_{34}$-*ZNF558*) but on the *ZNF558* opposite strand resulting in the nonconservation of ORF of *ZNF558* with a stop codon after the breakpoint at 287 bp (Fig. 2E).

Seven cases have been identified: five expressing two isoforms of *TRIO-TERT* and two others cases expressing *TRIO* gene fusion with out-of frame interchromosomal gene partner. For *TRIO-TERT* cases, the preferential expressed isoform is $TRIO_{33/34}$-$TERT_3$ and not the isoform with *TERT* ORF conservation. All these results indicated that *TRIO* gene fusion lead to the formation of truncated TRIO protein.

Unfortunately, we could not detect these fusion proteins by Western blot analysis of sarcoma cell line lysates, due to the lack of sensitivity/specificity of the TRIO antibody available that recognize these fusion proteins (data not shown).

Are *TRIO* fusions specific to sarcomas with complex genetics?

The seven identified translocated cases belong to different histotypes: three undifferentiated pleomorphic sarcomas, two dedifferentiated liposarcomas, one pleomorphic rhabdomyosarcoma, and one myxofibrosarcoma. According to the *Fédération Française des Centres de Lutte Contre le Cancer* grading system, five of them were classified as grade 3 (T1, S867, S829, S822, and S909) and two as grade 2 tumors (T2 and S921). Three tumors presented a metastatic relapse and two presented a local recurrence (Table 2). These results indicate a trend toward tumor aggressiveness without reaching significance considering the number of cases.

Specificity of all *TRIO* chimeric transcripts has been tested on two independent cohorts of 74 GISTs and 24 synovial sarcomas by RT-PCR. None have been identified in these nonpleomorphic

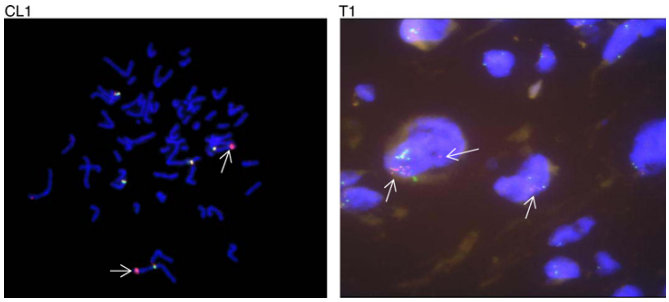


Figure 3.
FISH on one *TRIO* fusion gene case. On the representative picture of FISH, we could observe unbalanced rearrangement of *TRIO* highlighted by arrows in metaphase cell of CL1 and interphase cell of T1. BAC clone covering 5′ part of *TRIO* (red signal) and two BAC clones covering 3′ part of *TRIO* (green signals) were used.

Table 2. Clinical and pathologic data for individuals harboring *TRIO* gene fusion

Case	Histotype	Grade	Location	Metastasis	Local recurrence	Genre	Chimeric transcripts	Chromosome gene 1	Chromosome gene 2	Break position gene 1	Break position gene 2	ORF conservation	New aa before stop codon
T1	UPS	3	Trunk wall	Y	N	F	*TRIO*(ex33)-*TERT*(ex2)	5	5	14406781	1294781	Y	—
							TRIO(ex33)-*TERT*(ex3)	5	5	14406781	1282739	N	24
S867	DD LPS	3	Lower limb	N	N	M	*TRIO*(ex33)-*TERT*(ex2)	5	5	14406781	1294781	Y	—
							TRIO(ex33)-*TERT*(ex3)	5	5	14406781	1282739	N	24
T2	DD LPS	2	Extremities	N	N	M	*TRIO*(ex34)-*TERT*(ex2)	5	5	14420130	1294781	N	117
							TRIO(ex34)-*TERT*(ex3)	5	5	14420130	1282739	Y	—
S829	PRMS	3	Lower limb	Y	Y	M	*TRIO*(ex34)-*TERT*(ex2)	5	5	14420130	1294781	N	117
							TRIO(ex34)-*TERT*(ex3)	5	5	14420130	1282739	Y	—
S822	UPS	3	Upper limb	N	Y	M	*TRIO*(ex34)-*CDH18*(ex2)-*TERT* (ex2)	5	5	14420130	1294781	N	7
							TRIO(ex34)-*CDH18*(ex2)-*TERT* (ex3)	5	5	14420130	1282739	N	7
S909	UPS	3	Internal trunk	Y	N	F	*TRIO*(ex33)-*LINC01504*(intron2)	5	9	14406781	74939905	N	37
							TRIO(ex33)-*LINC01504*(intron3)	5	9	14406781	74948537	N	28
							TRIO(ex33)-*LINC01504*(exon4)	5	9	14406781	74950306	N	98
S921	MFS	2	Lower limb	N	N	F	*TRIO*(ex34)-*ZNF558*	5	19	14420130	8932224	N	196

Abbreviations: aa: amino acid; DDLPS, dedifferentiated liposarcoma; F, female; M, male; MFS, myxofibrosarcoma; N, no; PRMS, pleomorphic rhabdomyosarcoma; UPS, undifferentiated pleomorphic sarcoma; Y, yes.

sarcoma cohorts associated to a specific oncogenic event (*KIT*/*PDGFRA* mutations and t(X;18) translocation, respectively; refs. 2–5).

TRIO gene fusion appears, thus, specific so far to nontranslocation-related sarcomas.

Do *TRIO* fusions induce a specific transcriptomic program?

To test whether the presence of chimeric *TRIO* transcripts could be associated with a peculiar transcriptional profile of the fused *TRIO* samples (FTS), we applied unsupervised clustering method on 117 sarcomas with RNA-seq expression, including the six fused *TRIO* cases and the 111 nonfused *TRIO* samples (non-FTS; Supplementary Fig. S5). Our results indicate that FTSs do not have similar transcriptomic profiles as they did not cluster together. FTSs are localized within a subgroup of nonmuscular sarcomas, located in the extremities (adjusted Fisher $P = 0.014$ for both tumor differentiation and location).

We then performed a supervised differential gene expression analysis comparing fused and nonfused cases and identified 423 significant genes (Supplementary Fig. S6), out of which 69 were overexpressed in the FTS (Supplementary Table S5) and 354 overexpressed in non-FTS (Supplementary Table S6). We observed that *TRIO* and *TERT* were significantly more expressed (4.88 and 16.20 times, respectively) in FTS (adjusted Wald's $P = 4 \times 10^{-7}$ and 4.66×10^{-3}, respectively). To establish whether *TRIO* overexpression is due to wild-type or fused *TRIO*, we measured *TRIO* differential exonic expression of FTS versus non-FTS using DEXSeq (Supplementary Fig. S7; ref. 26). As the analysis could not be done on the whole non-FTS cohort ($n = 111$), we selected the 40 (maximum analysis capacity) samples that highly expressed *TRIO*. This analysis did not measure any differentially expressed exon in FTS versus non-FTS. However, we can see a slight exonic usage switch in layer "exon usage" starting at exon 34. Exons before this point are a bit more expressed in FTS, where *TRIO* expression is due to wild-type and fused *TRIO*. The trend is inversed after the breakpoint; exons 34 at 58 are a bit more expressed in non-FTS. In addition, this result shows that as no reciprocal fusions were observed in FTS and exons beyond 34 are expressed, wild-type *TRIO* is expressed in FTS. Genes interacting with *TRIO* and *TERT* were not identified as differentially expressed according to BioGRID (23).

GSEA (24, 25) was then applied on the Hallmarks new collection included in the MSidDB v5.0 database (containing 50 referenced gene sets). In fused *TRIO* samples, 16 significantly enriched gene sets were identified, mainly implicated in immunity/inflammation (6/16 gene sets), regulation of cell cycle (3/16), and cell proliferation and migration (3/16; Supplementary Fig. S8). Immunity gene set enrichment due to immune cell infiltration has been ruled out following a histologic review indicating that there was no significant difference between translocated and nontranslocated cases in terms of immune cell infiltration. In non-FTS, overexpressed gene sets are implicated, notably in myogenesis and adipogenesis (Supplementary Fig. S9).

These data suggest that *TRIO* fusion events are associated with overexpression of genes involved in immunity and inflammation, regulation of cell cycle, cell proliferation, and migration.

Do fusions trigger TERT reactivation?

In a tumoral context, telomere length is maintained by two exclusive ways, either by telomerase reactivation or by the alternative lengthening of telomeres mechanism (ALT mechanism),

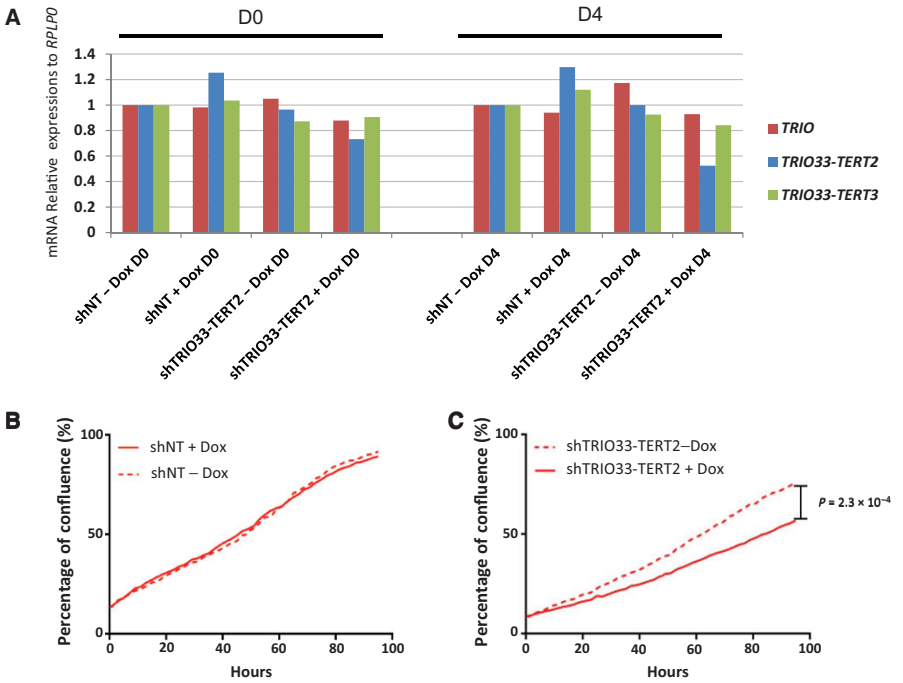


Figure 4.
Effect of $TRIO_{33}$-$TERT_2$ knockdown on cell proliferation. **A,** $TRIO_{33}$-$TERT_2$, $TRIO_{33}$-$TERT_3$, and *TRIO* mRNAs relative expression levels have been determined by qRT-PCR at day 0 (D0) and day 4 (D4) in the cell line expressing doxycycline-inducible shRNA-targeting $TRIO_{33}$-$TERT_2$ isoform (*shTRIO33-TERT2*) and the cell line expressing control shRNA (shNT, nontargeting shRNA). These cell lines have been treated (+) or not (−) with doxycycline (Dox) for 2 days before D0. ShNT–Dox is used as control. **B** and **C,** Proliferation assay has been performed during 4 days on CL1 with shRNA control (**B**) or sh*TRIO33-TERT2* (**C**) treated or not with doxycycline (representative of one experiment; $n = 2$, 8 replicates per condition). *P* values between sh*TRIO33-TERT2* – and + doxycycline conditions are significantly different (Wilcoxon's *P* with Benjamini–Hochberg adjustment < 0.05) from 27 hours of experiment. *P* value at 96 hours is indicated by the vertical line in graph.

and both mechanisms appear to be mutually exclusive (35, 36). In pleomorphic sarcomas, the ALT mechanism is very frequent (60% of cases; ref. 37). To test whether *TRIO-TERT* fusion could be associated with TERT reactivation, we evaluated the colocalization of PML and TERF2 as a hallmark of the ALT mechanisms, which, if negative, means that it is the telomerase reactivation pathway (38). The four *TRIO-TERT* interpretable cases were negative for PML/TERF2 colocalization and, consequently, negative for ALT mechanism. Moreover, S921 ($TRIO_{34}$-*ZNF558*) presented the same results and was also declared ALT negative (Supplementary Fig. S10).

Do *TRIO* fusions trigger phenotype modifications?

To uncover the role of *TRIO* fusions, we performed $TRIO_{33}$-$TERT_{2/3}$ knockdowns in CL1 using lentivirally delivered shRNA contructs. Two shRNAs for each isoform were tested at day 0 and day 4 and only one allowed the specific decrease of $TRIO_{33}$-$TERT_2$ (50%) at 4 days without decreasing *TRIO* and $TRIO_{33}$-$TERT_3$ expressions (Fig. 4A). With respect to transcriptomic analysis, we thus used it to run proliferation assays in CL1. As shown in Fig. 4B and C, contrary to the nontargeting shRNA, knocking down $TRIO_{33}$-$TERT_2$ significantly reduces cell proliferation.

Discussion

Sarcomas with complex genetics are tumors that harbor a high number of chromosomal rearrangements. To better understand the biology of these tumors, we searched for recurrent chimeric transcripts.

Chimeric transcripts in nontranslocation-related sarcomas

In the current study, 10,382 potential chimeric transcripts have been predicted, and 3,604 of them were considered read-throughs by the detection algorithm deFuse. Although read-throughs were

not considered in this study, it has been reported that tumor-specific read-through expression could be involved in oncogenesis (39).

More than 80 filtered fusion transcripts were predicted per case in nontranslocation-related sarcomas, and most of the chimeric transcripts detected arose from intrachromosomal rearrangements and were not in-frame (96%). This is in agreement with data reported by Yoshihara and colleagues, regarding fusion transcript prediction in a large series of 4,366 cancers (various types excluding sarcomas; ref. 40). They showed that the high complexity of tumor genome is associated with a higher number of predicted fusion transcripts. Our results are consistent with the high genomic complexity levels of nontranslocation-related sarcomas. On the contrary, according to the data presented by Yoshihara and colleagues, approximately 36% of detected fusion transcripts were predicted to be in-frame in their cohort compared with 4.13% in our study. This suggests that chimeric transcripts could have other role(s) in nontranslocation-related sarcomas as they would mainly produce a truncated protein or no protein product at all.

TRIO gene fusions lead to truncated TRIO protein expression

Here, we report recurrent *TRIO* fusions identified by RNA-seq and validated in 6 poorly differentiated pleomorphic sarcomas out of 117 (5.1%) and one dedifferentiated liposarcoma (T2) identified in aCGH analysis out of 26 nontranslocation-related sarcomas (3.8%). Three different *TRIO* fusion partners have been identified in this study, *TERT*, *LINC01504*, and *ZNF558*. Whatever the fusion partner, breakpoints in *TRIO* are consistently either after exons 33 or 34. This likely suggests that either these conserved domains, in the fused TRIO proteins, are essential or that genomic location of the breakpoint is specifically targeted by an unknown mechanism. Furthermore, two of the partners (*LINC01504* and *ZNF558*) are fused out-of-frame, suggesting that translocation leads to a truncated TRIO protein and that this truncated form is of a selective advantage for tumor cell. This hypothesis is strengthened by the isoform-specific quantitative analysis, showing that chimeric transcripts with an out-of-frame *TERT* fusion are the most expressed. It is even more strengthened by the differential expression analysis. For statistical consideration, GSEA (50 cancer Hallmarks) was applied instead of Gene Ontology (41) analysis, and we evidenced that *TRIO*-translocated cases significantly overexpressed genes involved in immunity/inflammation, cell proliferation, and migration.

TRIO protein has two GEF domains (42). While GEF1 mediates GDP to GTP exchange, leading to Rac1 and RhoG activation, GEF2 domain triggers RhoA activation. In addition to these domains, TRIO has a SEC14 domain, several spectrin repeats, two SH3 domains, an Ig-like domain, and a serine kinase domain. Unlike the GEF domains, the roles of these latter domains are still unclear. Rac1 and RhoG are members of Rho proteins controlling several pathways leading to cytoskeletal rearrangements, kinase activation, and gene transcription. They are implicated in cell proliferation and motility (43).

All *TRIO* fusions encode for TRIO N-terminal domains Sec14, spectrin, and GEF1 domain (Supplementary Fig. S11), which triggers Rac1 activation. Rac1 has been demonstrated to be directly involved in transformation and tumor progression (44). Rac1 acts by promoting anchorage-independent growth (45), proinflammatory pathway (46), proliferation, cell spreading, and migration by regulating lamellipodia formation (47, 48). Interestingly, *TRIO* gene fusions lead to the loss of the C-terminal GEF2, suggesting that the GEF2 domain is not associated to the process of tumor progression in sarcoma. This is consistent with the observation that most processes induced by TRIO depend on the activation of Rac1 by GEF1 and not on the activation of RhoA by GEF2 (49). Moreover, the depletion of $TRIO_{33}$-$TERT_2$ isoform demonstrates that *TRIO* fusion has an impact on cell proliferation. As this isoform is the least expressed in all cases, we hypothesize that the effect would be stronger if the $TRIO_{33}$-$TERT_3$ isoform was downregulated concomitantly (no shRNA available).

TRIO encodes at least seven different isoforms (49) and two of them (B and C), containing only the Sec14, spectrin, GEF1, and SH3 domains, are exclusively expressed in nervous system (50). Interestingly, the *TRIO* exons involved in the fusions (exons 1 to 33-34) are exactly the same as the ones observed in the two neuronal isoforms B and C, and as a consequence, they share the same protein domains. The role(s) of these two isoforms is unknown. Salhia and colleagues (51) have demonstrated that the transcriptional activation of the nervous system–specific *TRIO* isoform, through transfer in a different cellular context, could be an efficient way to hijack and remodel existing metabolic pathways for the benefit of tumor cells as demonstrated in glioblastoma (51).

Telomerase reactivation in _TRIO_-fused cases

TRIO gene has been identified in fusion with three partners: *TERT*, *LINC01504*, and *ZNF558*. Although *TERT*'s role on elongation of telomeres is well known, the role of *LINC01504* and *ZNF558* is unknown. Stransky and colleagues (52) identified $TRIO_{33}$-$TERT_2$ fusion in two dedifferentiated liposarcomas. The authors suggest that this gene fusion could permit the expression of *TERT* in these tumors. In our study, the length of telomeres is maintained by telomerase reactivation and not by ALT mechanism in the *TRIO* gene fusions. *TERT* is overexpressed in these samples regardless of the *TRIO* fusion partner (i.e., *TERT*, *LINC01504*, and *ZNF558*), suggesting that, in these cases at least, telomerase activation is not associated with *TRIO* rearrangements. Moreover, in-frame *TERT* chimeric transcript loses the TEN domain of TERT permitting the fixation of *TR* RNA, the matrix for telomeres elongation (32). This prompts us to conclude that *TRIO-TERT* fusions do not trigger TERT reactivation in these tumors.

TRIO gene fusion is specific to nontranslocation-related sarcomas

TRIO fusion transcripts have been detected exclusively in a cohort of sarcomas with complex genetics. We did not detect any *TRIO* chimeric transcript in large cohorts of GISTs and synovial sarcomas, which are both sarcoma subtypes associated with a specific genetic alteration and a very low genomic complexity. Moreover, in Yoshihara and colleagues' study (40), chimeric RNA-seq data of 4,366 tumor samples have been submitted to fusion detection, and *TRIO* fusion transcripts have not been detected in the whole cohort of various tumors (excluding sarcomas). *TRIO* fusions could therefore be quite specific to nontranslocation-related sarcomas, with an incidence of around 5%. It is the unique recurrent gene fusion so far identified in these aggressive nontranslocation-related sarcomas. Within this large sarcoma category, *TRIO* fusions are not limited to a specific histologic

subtype. This suggests that, unlike sarcomas with simple genetics and a recurrent specific translocation (e.g., Ewing sarcoma, synovial sarcoma, and alveolar rhabdomyosarcomas), which are associated to a specific transcriptomic profile (14), this chromosomal event is certainly not an initiating one but a secondary one caused by genetic instability of nontranslocation-related sarcomas and involved in tumor progression via cell proliferation enhancement. However, we did not have sufficient statistical evidence with seven cases to highlight this point, and we cannot exclude the possibility that more cases would permit to refine the biological and/or clinical context associated with these *TRIO* rearrangements.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: L. Lartigue, J.-M. Coindre, F. Chibon
Development of methodology: S. Schmidt, F. Chibon
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): L. Delespaul, T. Lesluyes, G. Pérot, J. Baud, P. Lagarde, S. Le Guellec, A. Neuville, P. Terrier, A. Debant, J.-M. Coindre, F. Chibon, D. Vince-Ranchère
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): L. Delespaul, T. Lesluyes, G. Pérot, C. Brulard, L. Lartigue, J. Baud, J.-M. Coindre, F. Chibon
Writing, review, and/or revision of the manuscript: L. Delespaul, T. Lesluyes, L. Lartigue, S. Le Guellec, S. Schmidt, A. Debant, F. Chibon
Study supervision: F. Chibon

Acknowledgments

Computer time for this study was provided by the computing facilities MCIA (Mésocentre de Calcul Intensif Aquitain) of the Université de Bordeaux and of the Université de Pau et des Pays de l'Adour. The authors thank Andreas Bikfalvi laboratory for granting access to IncuCyte. We are grateful to the people involved in the French Sarcoma Group (GSF) for clinical annotations and access to their samples: L. Guillou, F. Collin, A. Leroux, Y-M. Robin, M. C. Chateau, I. Peyrottes, and F. Mishellany. The authors would also like to thank Dr. Ravi Nookala of Institut Bergonié for the medical writing services.

Grant Support

This work was supported by euroSARC (FP7-HEALTH-2011), Fondation ARC pour la recherche Contre le Cancer (to A. Debant), Ligue Contre le Cancer (Comité régional Gironde to F. Chibon and Comité régional Languedoc Roussillon to A. Debant) Inserm, and Integrated Cancer Research Site (SIRIC) of Bordeaux (BRIO-Bordeaux Oncology research).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

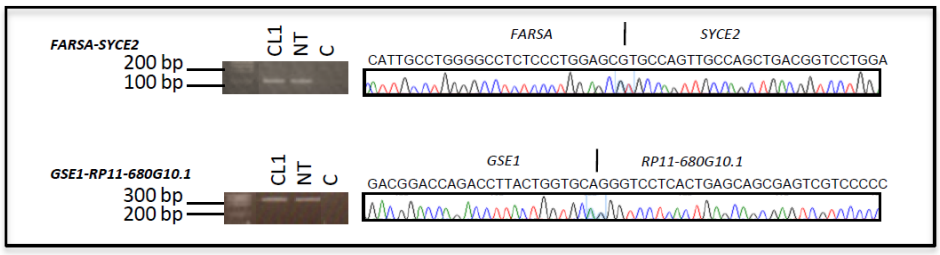
Received February 2, 2016; revised June 27, 2016; accepted July 27, 2016; published OnlineFirst August 15, 2016.

References

1. Fletcher CDM, Unni KK, Mertens F, editors. WHO Classification of Tumours of Soft Tissue and Bone. 4th ed. Lyon, France: IARC Press; 2013.
2. Nakahara M, Isozaki K, Hirota S, Miyagawa J, Hase-Sawada N, Taniguchi M, et al. A novel gain-of-function mutation of c-kit gene in gastrointestinal stromal tumors. Gastroenterology 1998;115:1090–5.
3. Delattre O, Zucman J, Melot T, Garau XS, Zucker JM, Lenoir GM, et al. The Ewing family of tumors–a subgroup of small-round-cell tumors defined by specific chimeric transcripts. N Engl J Med 1994;331:294–9.
4. Guillou L, Coindre J, Gallagher G, Terrier P, Gebhard S, de Saint Aubain Somerhausen N, et al. Detection of the synovial sarcoma translocation t (X;18) (SYT;SSX) in paraffin-embedded tissues using reverse transcriptase-polymerase chain reaction: a reliable and powerful diagnostic tool for pathologists. A molecular analysis of 221 mesenchymal tumors fixed in different fixatives. Hum Pathol 2001;32:105–12.
5. Sorensen PHB, Lynch JC, Qualman SJ, Tirabosco R, Lim JF, Maurer HM, et al. PAX3-FKHR and PAX7-FKHR gene fusions are prognostic indicators in alveolar rhabdomyosarcoma: a report from the children's oncology group. J Clin Oncol 2002;20:2672–9.
6. Chibon F, Mariani O, Derré J, Malinge S, Coindre JM, Guillou L, et al. A subgroup of malignant fibrous histiocytomas is associated with genetic changes similar to those of well-differentiated liposarcomas. Cancer Genet Cytogenet 2002;139:24–9.
7. Coindre J-M, Mariani O, Chibon F, Mairal A, De Saint Aubain Somerhausen N, Favre-Guillevin E, et al. Most malignant fibrous histiocytomas developed in the retroperitoneum are dedifferentiated liposarcomas: a review of 25 cases initially diagnosed as malignant fibrous histiocytoma. Mod Pathol 2003;16:256–62.
8. Chibon F, Mairal A, Fréneaux P, Terrier P, Coindre JM, Sastre X, et al. The RB1 gene is the target of chromosome 13 deletions in malignant fibrous histiocytoma. Cancer Res 2000;60:6339–45.
9. 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.
10. Adamowicz M, Radlwimmer B, Rieker RJ, Mertens D, Schwarzbach M, Schraml P, et al. Frequent amplifications and abundant expression of TRIO, NKD2, and IRX2 in soft tissue sarcomas. Genes Chromosomes Cancer 2006;45:829–38.
11. 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.
12. Chibon F, Mariani O, Mairal A, Derré J, Coindre J-M, Terrier P, et al. The use of clustering software for the classification of comparative genomic hybridization data. an analysis of 109 malignant fibrous histiocytomas. Cancer Genet Cytogenet 2003;141:75–8.
13. Mairal A, Chibon F, Rousselet A, Couturier J, Terrier P, Aurias A. Establishment of a human malignant fibrous histiocytoma cell line, COMA. Characterization by conventional cytogenetics, comparative genomic hybridization, and multiplex fluorescence *in situ* hybridization. Cancer Genet Cytogenet 2000;121:117–23.
14. Pierron G, Tirode F, Lucchesi C, Reynaud S, Ballet S, Cohen-Gogo S, et al. A new subtype of bone sarcoma defined by BCOR-CCNB3 gene fusion. Nat Genet 2012;44:461–6.
15. Chmielecki J, Crago AM, Rosenberg M, O'Connor R, Walker SR, Ambrogio L, et al. Whole-exome sequencing identifies a recurrent NAB2-STAT6 fusion in solitary fibrous tumors. Nat Genet 2013;45:131–2.
16. Robinson DR, Wu Y-M, Kalyana-Sundaram S, Cao X, Lonigro RJ, Sung Y-S, et al. Identification of recurrent NAB2-STAT6 gene fusions in solitary fibrous tumor by integrative sequencing. Nat Genet 2013;45: 180–5.
17. 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 2015;2:1014.
18. 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.
19. 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.
20. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol 2011;12:R72.
21. Iyer MK, Chinnaiyan AM, Maher CA. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. Bioinformatics 2011; 27:2903–4.
22. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014; 15:550.

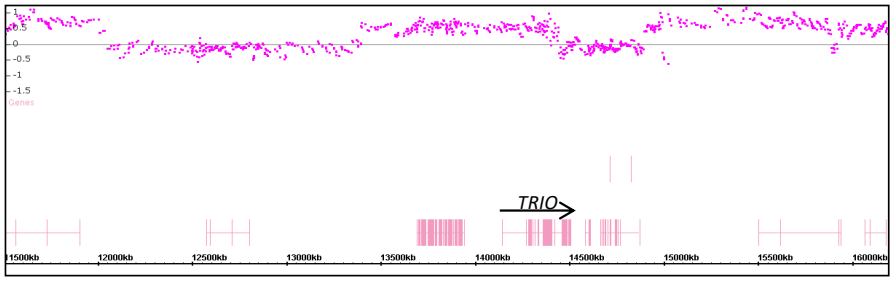
23. Chatr-Aryamontri A, Breitkreutz B-J, Oughtred R, Boucher L, Heinicke S, Chen D, et al. The BioGRID interaction database: 2015 update. Nucleic Acids Res 2015;43:D470–8.
24. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet 2003; 34:267–73.
25. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005;102:15545–50.
26. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. Genome Res 2012;22:2008–17.
27. De Preter K, Speleman F, Combaret V, Lunec J, Laureys G, Eussen BHJ, et al. Quantification of MYCN, DDX1, and NAG gene copy number in neuroblastoma using a real-time quantitative PCR assay. Mod Pathol 2002; 15:159–66.
28. Terrier-Lacombe MJ, Guillou L, Chibon F, Gallagher G, Benhattar J, Terrier P, et al. Superficial primitive Ewing's sarcoma: a clinicopathologic and molecular cytogenetic analysis of 14 cases. Mod Pathol 2009;22:87–94.
29. Akiva P, Toporik A, Edelheit S, Peretz Y, Diber A, Shemesh R, et al. Transcription-mediated gene fusion in the human genome. Genome Res 2006;16:30–6.
30. Grosso AR, Leite AP, Carvalho S, Matos MR, Martins FB, Vítor AC, et al. Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma. Elife 2015;4:pii: e09214.
31. Nacu S, Yuan W, Kan Z, Bhatt D, Rivers CS, Stinson J, et al. Deep RNA sequencing analysis of readthrough gene fusions in human prostate adenocarcinoma and reference samples. BMC Med Genomics 2011;4:11.
32. Prakash T, Sharma VK, Adati N, Ozawa R, Kumar N, Nishida Y, et al. Expression of conjoined genes: another mechanism for gene regulation in eukaryotes. PLoS One 2010;5:e13284.
33. Withers JB, Ashvetiya T, Beemon KL. Exclusion of exon 2 is a common mRNA splice variant of primate telomerase reverse transcriptases. PLoS One 2012;7:e48016.
34. Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 2009;37:e67.
35. Autexier C, Lue NF. The structure and function of telomerase reverse transcriptase. Ann Rev Biochem 2006;75:493–517.
36. Cesare AJ, Reddel RR. Alternative lengthening of telomeres: models, mechanisms and implications. Nat Rev Genet 2010;11:319–30.
37. 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.
38. Hu J, Hwang S-S, Liesa M, Gan B, Sahin E, Jaskelioff M, et al. Antitelomerase therapy provokes ALT and mitochondrial adaptive mechanisms in cancer. Cell 2012; 148:651–63.
39. Varley KE, Gertz J, Roberts BS, Davis NS, Bowling KM, Kirby MK, et al. Recurrent read-through fusion transcripts in breast cancer. Breast Cancer Res Treat 2014;146:287–97.
40. Yoshihara K, Wang Q, Torres-Garcia W, Zheng S, Vegesna R, Kim H, et al. The landscape and therapeutic relevance of cancer-associated transcript fusions. Oncogene 2015;34:4845–54.
41. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 2009;10:48.
42. Debant A, Serra-Pagès C, Seipel K, O'Brien S, Tang M, Park SH, et al. The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase, contains a protein kinase domain, and has separate rac-specific and rho-specific guanine nucleotide exchange factor domains. Proc Natl Acad Sci USA 1996;93:5466–71.
43. Hall A. Rho GTPases and the actin cytoskeleton. Science 1998;279:509–14.
44. Jaffe AB, Hall A. Rho GTPases in transformation and metastasis. Adv Cancer Res 2002;84:57–80.
45. Seipel K, Medley QG, Kedersha NL, Zhang XA, O'Brien SP, Serra-Pages C, et al. Trio amino-terminal guanine nucleotide exchange factor domain expression promotes actin cytoskeleton reorganization, cell migration and anchorage-independent cell growth. J Cell Sci 1999;112:1825–34.
46. Van Rijssel J, Timmerman I, Van Alphen FPJ, Hoogenboezem M, Korchynskyi O, Geerts D, et al. The Rho-GEF Trio regulates a novel pro-inflammatory pathway through the transcription factor Ets2. Biol Open 2013; 2:569–79.
47. Vaqué JP, Dorsam RT, Feng X, Iglesias-Bartolome R, Forsthoefel DJ, Chen Q, et al. A genome-wide RNAi screen reveals a Trio-regulated Rho GTPase circuitry transducing mitogenic signals initiated by G protein-coupled receptors. Mol Cell 2013;49:94–108.
48. van Rijssel J, Hoogenboezem M, Wester L, Hordijk PL, Van Buul JD. The N-terminal DH-PH domain of Trio induces cell spreading and migration by regulating lamellipodia dynamics in a Rac1-dependent fashion. PLoS One 2012;7:e29912.
49. Schmidt S, Debant A. Function and regulation of the Rho guanine nucleotide exchange factor Trio. Small GTPases 2014;5:e29769.
50. Portales-Casamar E, Briançon-Marjollet A, Fromont S, Triboulet R, Debant A. Identification of novel neuronal isoforms of the Rho-GEF Trio. Biol Cell 2006;98:183–93.
51. Salhia B, Tran NL, Chan A, Wolf A, Nakada M, Rutka F, et al. The guanine nucleotide exchange factors trio, Ect2, and Vav3 mediate the invasive behavior of glioblastoma. Am J Pathol 2008;173:1828–38.
52. Stransky N, Cerami E, Schalm S, Kim JL, Lengauer C. The landscape of kinase fusions in cancer. Nat Commun 2014;5:4846.

Supplementary Figure 1: Identification of *FARSA-SYCE2* and *GSE1-RP11-680G10.1* read-throughs in sarcoma CL1 and normal tissue. RT-PCR and chromatograms obtained with *FARSA-SYCE2* or *GSE1-RP11-680G10.1* primers. Vertical black lines indicate sequence merge between the two genes. Legend: *C* refers to PCR control, *NT* is normal tissue and *bp* is base pairs.

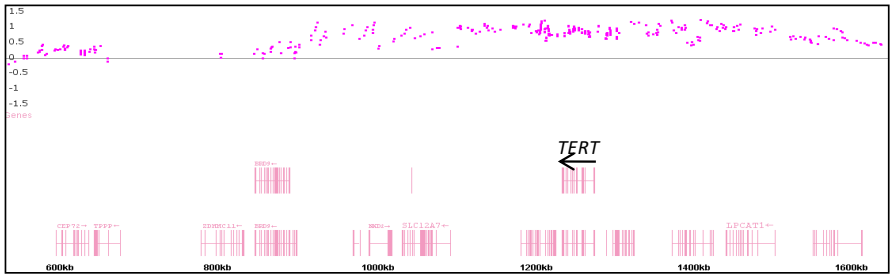


Supplementary Figure 2: aCGH (SNP-array) profile of CL1 which present a genomic rearrangement of *TRIO* and *TERT* genes. Horizontal black line represent chromosome 5 from 11500 kb to 16000 kb (A) and 600 kb to 1600kb (B). The log ratios of tumor versus reference are indicated on the *y* axis. Each purple point is a probe representing a short genomic region status. Above the aCGH plot, genomic location of genes are shown in pink, exons are represented in vertical lines and introns in horizontal. *TRIO* (A) and *TERT* (B) genes are indicated by arrows showing the strand orientations. For both *TRIO* and *TERT* genes genomic regions are amplified. Legends: *kb* is kilo base.

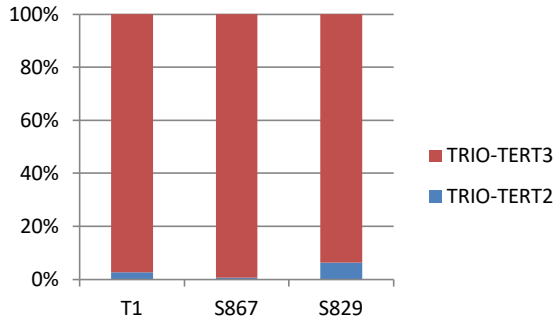
A



B



Supplementary Figure 3: Expression percentage of the two *TRIO-TERT* isoforms in three cases. *TRIO-TERT*$_2$ expression percentage is indicated in red and *TRIO-TERT*$_3$ expression percentage in blue.



Supplementary Figure 4: Predicted acceptor splicing regions in intron 2 and 3 of *LINC01504*. *LINC01504* intronic sequences (intron 2 and 3, respectively) with validated fusion point represented by the "|" character. Nucleic acids in red are acceptor splicing site predicted by Human Splicing Finder.

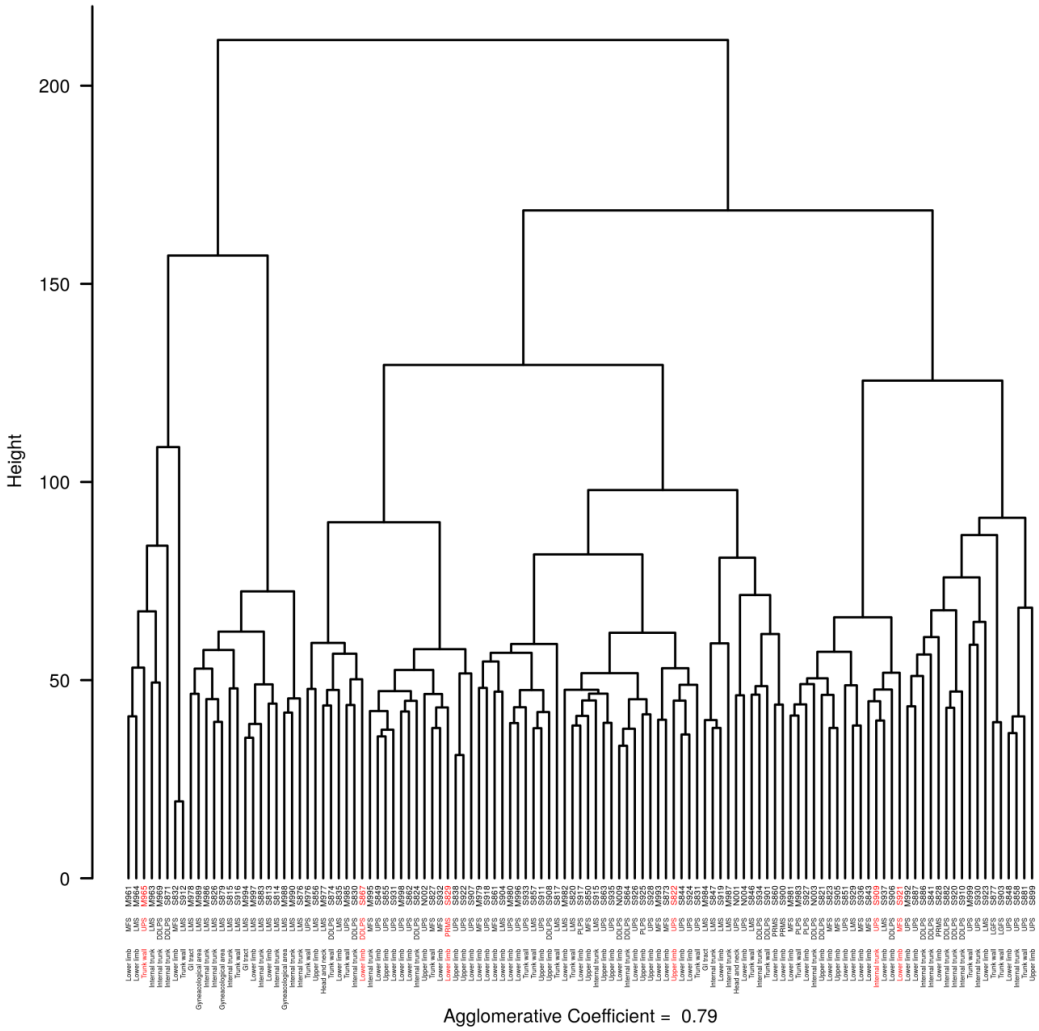
LINC01504 intron 2

ctctgagacatagataatattgttattcccattttacag|atgaaggaacgagggcacagagagattcgagttgtcacccaagctcacatagcacca

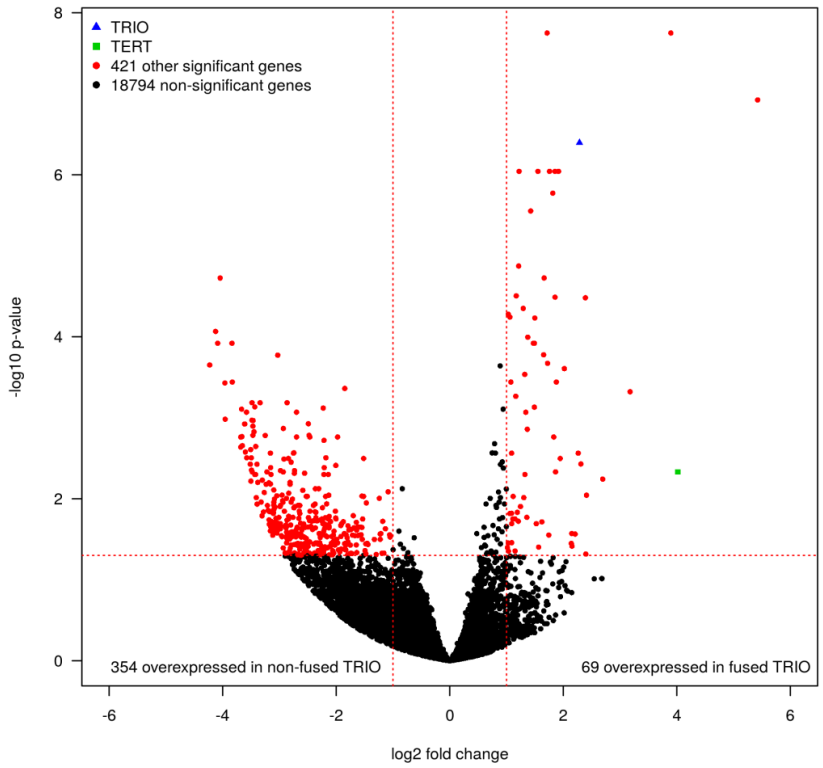
LINC01504 intron 3

ctctttcctctccgcctcagagtttacctatttattccag|aaacaaaggcacagacgtcattgtcacgttgtgcacttgaatttcaactccgaactgacc

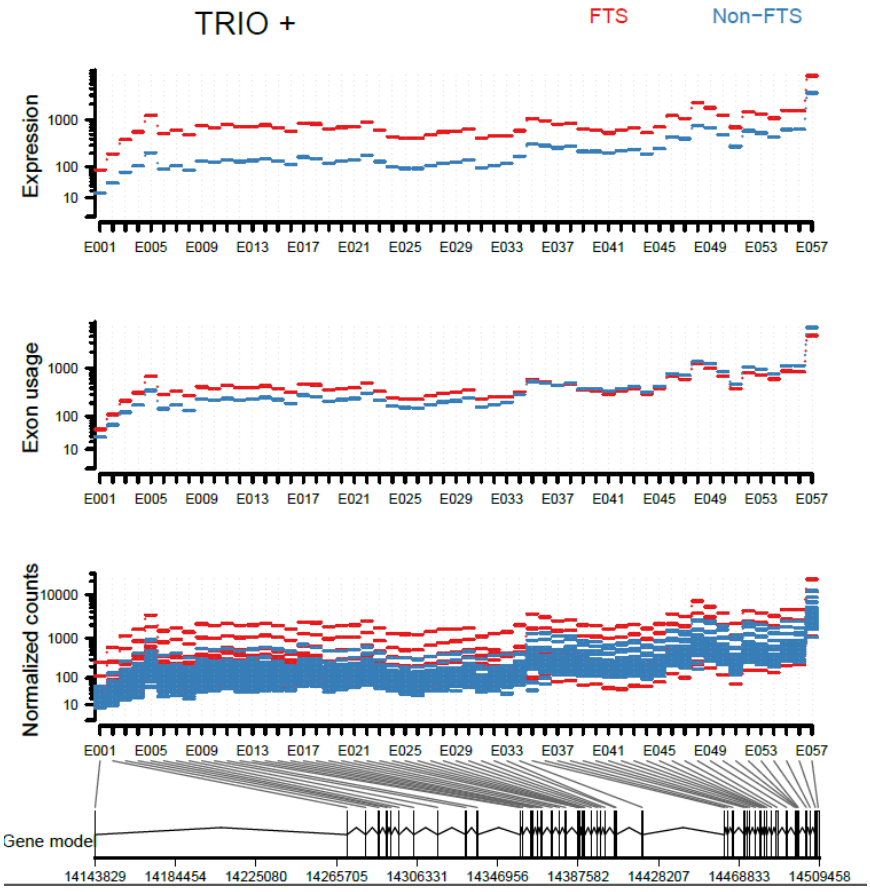
Supplementary Figure 5: Transcriptomic profile clustering of the 117 non-translocation-related sarcomas. Additional information were added above sample names (from top to bottom): red refers to fused TRIO samples, *UPS*, *LMS*, *DDLPS*, *MFS*, *PLPS*, *PRMS* and *Other* refer to sarcoma histotypes. This indicates fused *TRIO* samples do not share a similar transcriptomic profile. Legend: CI : UPS: Undifferentiated Pleomorphic Sarcoma ; LMS: leiomyosarcoma ; DD LPS : Dedifferentiated Liposarcoma; MFS : Myxofirosarcoma ; PLPS : pleomorphic liposarcoma ; PRMS : pleomorphic Rhabdomyosarcoma.



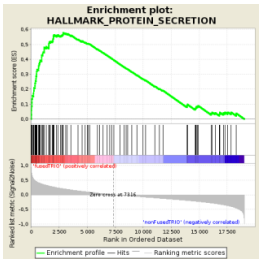
Supplementary Figure 6: Volcano plot of the 117 sarcomas with six fused *TRIO* cases versus 111 non-fused *TRIO* cases. Horizontal and vertical dashed red lines are log2 fold change and –log10 adjusted p-value thresholds (1/-1 and log2(0.05) respectively). Red points are significant differently expressed genes according to the defined thresholds. *TRIO* gene is highlighted in blue triangular point and *TERT* gene in green square. *TRIO* and *TERT* genes are identified among the most differentially expressed genes with the higher expression in fused *TRIO* samples versus non-fused *TRIO* samples.



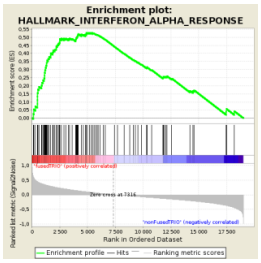
Supplementary Figure 7: Exonic expression of fused *TRIO* fused and non-fused *TRIO* cases. *TRIO* differential exonic expression of fused *TRIO* samples versus non-fused *TRIO* samples using DEXSeq. Since the analysis could not be done on the whole non-fused *TRIO* samples cohort (n=111), we selected the 40 (maximum analysis capacity) samples that highly expressed *TRIO*. Layer "Expression" represents the average exonic expression. Layer "Exon usage" represents the exon usage considering similar expression between fused *TRIO* samples and non-fused *TRIO* samples. Layer "Normalized counts" represents expression values for all tumors. Legends: FTS, fused *TRIO* samples; non-FTS, non-fused *TRIO* samples.



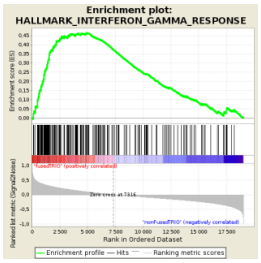
Supplementary Figure 8: Enrichment plot for significant *Hallmarks* in six fused *TRIO* samples included in the MSig5.0 database according to GSEA. Plots represent the distribution of differentially genes between fused *TRIO* versus non-fused *TRIO* samples (from left to right). Genes involved in the gene set are shown with vertical black lines. The green line represents the enrichment profile depending on gene distribution and tested gene set. Adjusted *P*-value for gene sets enrichments are indicated below each plot (FDR *Q*-value).



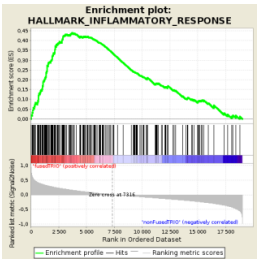
FDR *Q*-value : 0.000



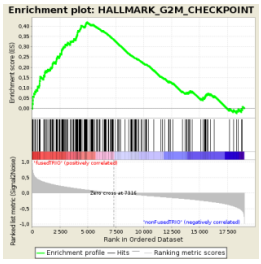
FDR *Q*-value : 0.000



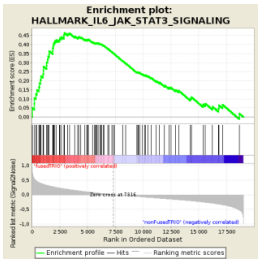
FDR *Q*-value : 0.000



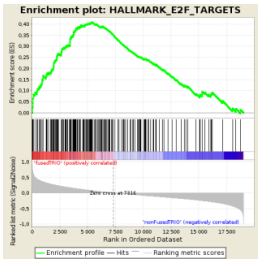
FDR *Q*-value : 0.000



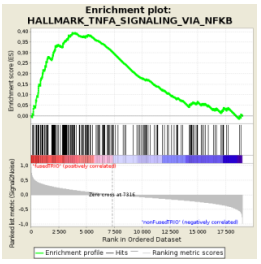
FDR *Q*-value : 0.000



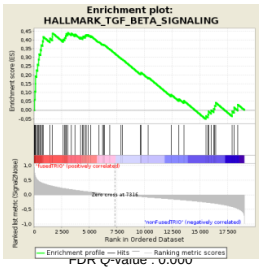
FDR *Q* -value : 0.000



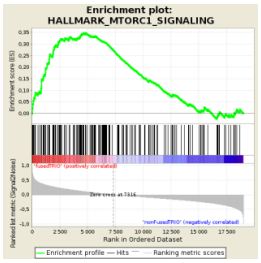
FDR *Q*-value : 0.000



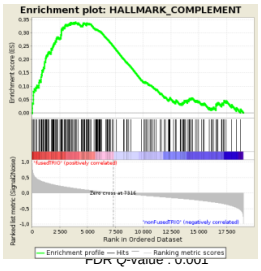
FDR *Q*-value : 0.000



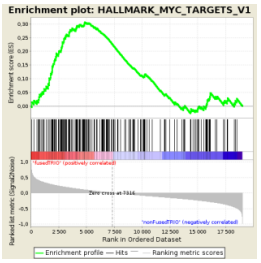
FDR *Q*-value : 0.000



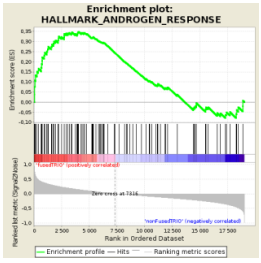
FDR *Q*-value : 0.000



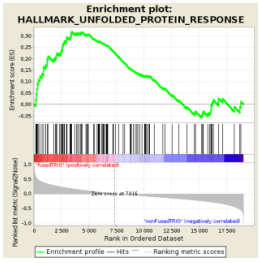
FDR *Q*-value : 0.001



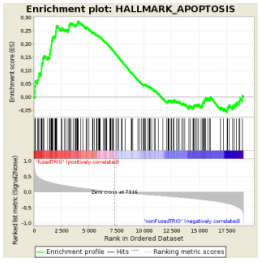
FDR *Q*-value : 0.005



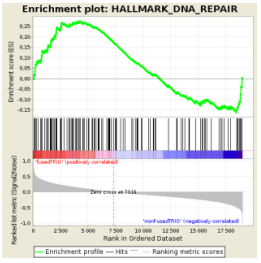
FDR *Q*-value : 0.004



FDR *Q*-value : 0.010

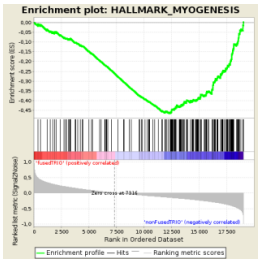


FDR *Q*-value : 0.021

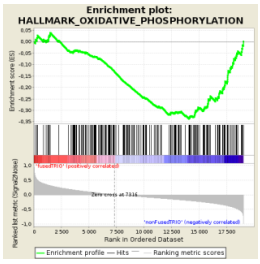


FDR *Q*-value : 0.031

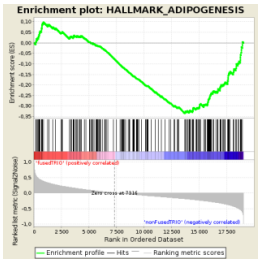
Supplementary Figure 9: Enrichment plot for significant *Hallmarks* in 111 non-fused *TRIO* samples included in the MSig5.0 database according to GSEA. Plots represent the distribution of differentially genes between fused *TRIO* versus non-fused *TRIO* samples (from left to right). Genes involved in the gene set are shown with vertical black lines. The green line represents the enrichment profile depending on gene distribution and tested gene set. Adjusted *P*-value for gene sets enrichments are indicated below each plot (FDR *Q*-value).



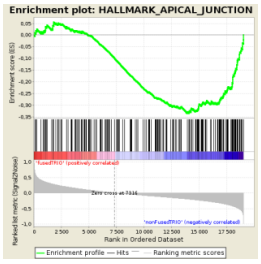
FDR *Q*-value : 0.000



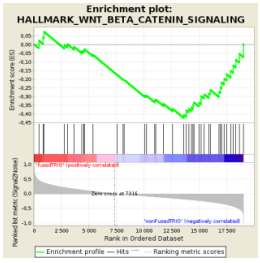
FDR *Q*-value : 0.046



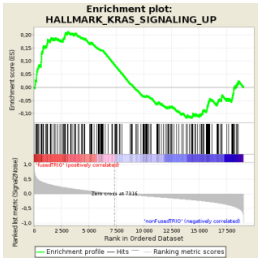
FDR *Q*-value : 0.036



FDR *Q*-value : 0.030

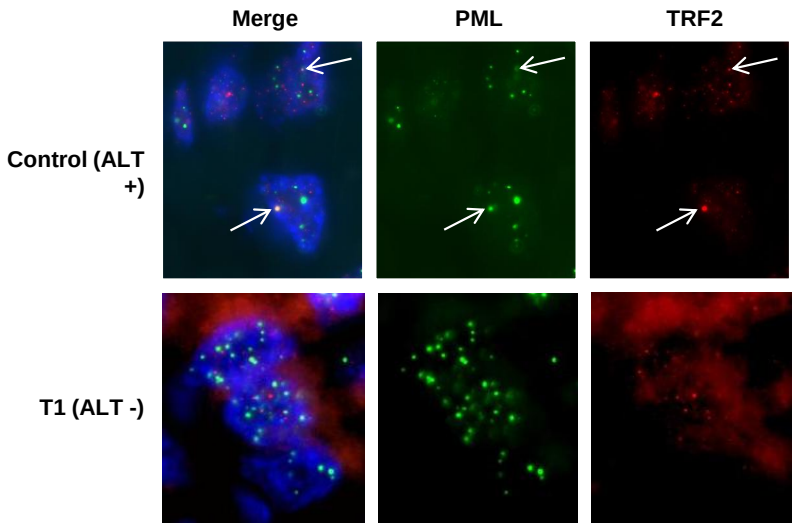


FDR *Q*-value : 0.024

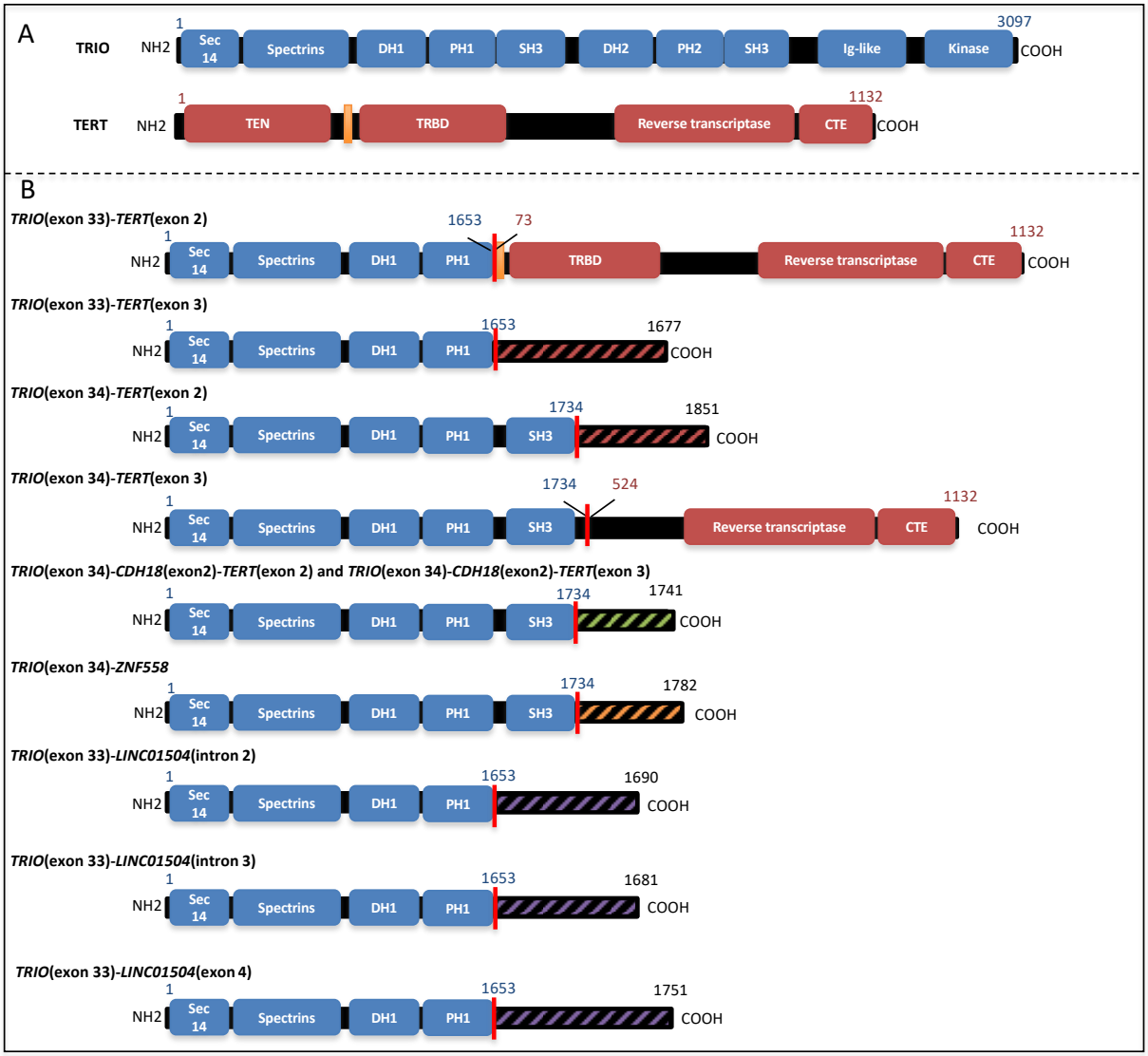


FDR *Q*-value : 0.035

Supplementary Figure 10: ALT mechanism status determined by PML/TERF2 immunofluorescence analysis. IF showing an of an example of PML (green) and TERF2 signals (red) colocalization in a tumor sample (Control, ALT+) and an example of absence of PML and TERF2 proteins colocalization (T1, ALT-) indicating that the ALT mechanism should be inactive in this tumor (ALT-). Magnification: X1000.



Supplementary Figure 11: Schematics of predicted TRIO fusions proteins. A) is the representation of wild type TRIO (blue) and TERT (red) respectively. Orange line in wild type TERT indicates the nuclear location signal domain. B) are the fused TRIO combinations. New supplementary C-terminal domains issued of out-of frame chimeric transcripts are indicated in black hatches. Fusion point is represented by vertical red line. Legends: *DH* : Dbl Homology domain; *PH* : Pleckstrin Homology domain ; *SH3* : SRC Homology 3 domain ; *Ig-like* : Imunoglobulin-like domain; *TEN* : TERT N-terminal domain; *TRBD* : Telomerase RNA-binding domain ; *CTE* : C-terminal domain.



Supplementary Tables

	Forward	Reverse
TRIO exon 33	ccaaggagacggcagcag	atcaccactgtcagctcaca
TRIO exon 34	gtggttcactgtgcatcgc	gatcatgtggggctggagg
TERT exon 3	gagctgacgtggaagatgag	acgtacacactcatcagcca
TERT exon 2	cagtgcctggtgtgcgtg	aaggccagcacgttcttc
ZNF558	gtgatcttatttccctcaatctccag	tcctgagagttttaagggaaagcc
AK095210 intron 2	ggtcacacaagagtttctccaa	ggctgcctgattcaaatctc
AK095210 intron 3	aagaaccactccctctttcctc	catgtcaccatgaggcattc
AK095210 exon 4	ggaagccgaggagaagtcat	tggaggttgcagtgaggc
FARSA	gtcttccgtccagagatgct	ggggactgtcatacaccatct
SYCE2	gaagagaactgcgaggagga	tgaagttggtcatgagtgca
GSE1	cccgggtgagataagcagtt	gttttggttggagcgatggt
RP11-980G10.1	tgtggatctctggaagcatct	atctgggccttgttctgctc

Supplementary Table 1: RT-PCR primers used for fusion transcripts detection. Forward and reverse primers used to control PCR before proceeding to fusion transcript detection.

Transcript	Forward	Reverse	Probe
TRIO(exon 33)-TERT(exon 2)	ctcacggacgtctcagaaca	aaggccagcacgttcttc	taaggtgtcctgcctgaagg
TRIO(exon 33)-TERT(exon 3)	ccagcctgatacgatttcca	acacagccaacccttatcg	cgcctcacggacgtctcagaac
TRIO(exon 34)-TERT(exon 2)	tccagaagtagcatggaaatgga	cgggccaccagctcctt	catcttcaaccacaaaggtgtcctgcc
TRIO(exon 34)-TERT(exon 3)	ccactccagaagtagcatggaa	tgcggccggaacaca	tggagggcatcttcaaccacaaagg
TRIO (exon 33-34)	tcacggacgtctcagaacac	gatcaccactgtcgctcaca	tggacagcgataagctctctggtgg

Supplementary Table 2: Primers used for fusion transcript quantification. Forward, reverse and probe primers used to proceed to fusion transcript quantification by qPCR .

ShRNA Name	Sequences
shRNA NT F	**5'-CCGGCAACAAGATGAAGAGCACCAACTCGAGTTGGTGCTCTTCATCTTGTTGTTTTT-3'**
shRNA NT R	**5'-AATTAAAAACAACAAGATGAAGAGCACCAACTCGAGTTGGTGCTCTTCATCTTGTTG-3'**
shRNA TRIO33TERT2 F	**5'-CCGGGCGATAAGGTGTCCTGCCCTCGAGGGCAGGACACCTTATCGCTTTTT-3'**
shRNA TRIO33TERT2 R	**5'-AATTAAAAAGCGATAAGGTGTCCTGCCCTCGAGGGCAGGACACCTTATCGC-3'**

Supplementary Table 3: shRNA sequences. Forward and reverse sequences of *TRIO33-TERT2* shRNAs and control Non Targeting shRNA (shRNA NT).

Characteristics	Cohort (n=24)
Median follow-up (years)	2.41 (CI95% 1.44-6.88)
Median Age (years)	63 (CI95% 56-66)
Sex	
Female	9 (37.5%)
Male	15 (62.5%)
FNCLCC grade (%)	
1	2 (8.33%)
2	4 (16.66%)
3	15 (62.5%)
NA	3 (12.5%)
Histological type (%)	
UPS	8 (33.33%)
LMS	7 (29.16%)
MFS	3 (12.5%)
PLPS	3 (12.5%)
DDLPS	2 (8.33%)
PRMS	1 (4.16%)
Location	
Lower limb	13 (54.16%)
Trunk wall	6 (25%)
Internal trunk	3 (12.5%)
Upper limb	1 (4.16%)
GI tract	1 (4.16%)
Relapse events (%)	
Metastasis	13 (54.16%)
Local recurrences	7 (29.16%)
Median Size of tumour (mm) (n=21)	80 (CI95% 75-120)

Supplementary Table 4: Sarcomas cohort analysed in aCGH (n=24).

Legend: CI : confidence interval ; UPS: Undifferentiated Pleomorphic Sarcoma ; LMS: leiomyosarcoma ; DD LPS : Dedifferentiated Liposarcoma; MFS : Myxofirosarcoma ; PLPS : pleomorphic liposarcoma ; PRMS : pleomorphic Rhabdomyosarcoma ; *GI*: Gastro intestinal

3.2.3 – Conclusion

À sa publication, cette étude représentait la plus vaste exploration de transcrits chimériques récurrents dans les SGC. Par la suite, ces résultats ont été confirmés par d'autres groupes qui retrouvent ces mêmes types de réarrangements de *TRIO*. Une première étude rapporta trois cas exprimant *TRIO-TERT* (deux DDLPS et un UPS) et deux cas *TRIO* réarrangés (*TRIO-CTNNA2* dans un UPS et *TRIO-CDH18-TERT* dans un DDLPS ; The Cancer Genome Atlas Research Network, 2017). Deux études supplémentaires rapportèrent l'expression de *TRIO-TERT* dans un myxofibrosarcome (Ogura *et al.*, 2018) et dans un liposarcome à cellules fusiformes (Suster *et al.*, 2019). Cette récurrence est également retrouvée au niveau des points de fusion de *TRIO* (exons 33 ou 34).

En analyse pan-cancer sur 33 types de cancers explorés dans le cadre du consortium TCGA, de nouveaux gènes de fusion impliquant *TRIO* ont pu être mis en évidence (**table 7** ; Gao *et al.*, 2018). Bien que de nombreux partenaires de *TRIO* soient retrouvés dans l'ensemble des tumeurs, l'expression de *TRIO-TERT* est spécifique aux sarcomes ainsi que la récurrence de(s) point(s) de fusion impliquant les exons 33 et 34 de *TRIO*. Ces données nous laissent alors suspecter l'existence d'une séquence et/ou structure d'ADN particulière qui rendrait cette région sensible à des cassures dans les sarcomes.

Dans une lignée cellulaire d'UPS, nous avons montré que l'inhibition de l'expression de $TRIO_{\text{exon 33}}$-$TERT_{\text{exon 2}}$ diminue la prolifération cellulaire (Delespaul *et al.*, 2017). La perte des exons 35 à 57, dans tous ces transcrits chimériques validés, pourrait ainsi conférer de nouvelles propriétés à ces sarcomes, participant à la biologie de la tumeur par une sélection clonale positive. Ceci serait en accord avec la conservation du domaine GEF1 de *TRIO*, permettant l'activation des GTPases Rac1 et RhoG ; Rac1 étant impliquée dans la mobilité et la prolifération (Jaffe & Hall, 2002 ; Schmidt & Debant, 2014).

Plusieurs informations nous laissent suspecter que l'expression d'un *TRIO* tronqué ne délivre pas un message oncogène fort. Tout d'abord, il n'existe pas de lien entre l'altération de *TRIO* et les données cliniques associées à la tumeur (différences en matière de sous-types histologique, localisation tumorale et agressivité). Ceci contraste avec les gènes de fusion agissant

Localisation tumorale	Gène(s) de fusion	Région de *TRIO*
Sein	*TRIO-LRFN5*	Exon 46
	TRIO-MARCH11	Exon 48
	TRIO-IFT52	Exon 54
	TRIO-FAM53B	Exon 11
Tête et cou (CS)	*TRIO-PPP2R5E & ARIH2-TRIO*	Exon 3 & exon 4
Foie	*TRIO-ZNF131*	Exon 1
	TRIO-SNHG18	**Exon 34**
Poumon	*TRIO-DNAH5*	Exon 27
	TRIO-UHRF2	Exon 1
Poumon (CS)	*TRIO-FBXL7*	Exon 27
Prostate	*TRIO-IQUB*	Exon 8
Tissus mous	*TRIO-TERT*	**Exon 34**
	TRIO-TERT	**Exon 33**
	TRIO-TERT	**Exon 34**
	TRIO-CTNNA2	Exon 11
	TRIO-CDH18 & CDH18-TERT	**Exon 34**
Peau	*TRIO-CLNK*	Exon 31
	TRIO-ADAMTS12	Exon 21
	TRIO-SLC9A3	Exon 28

TABLE 7 – Réarrangements de *TRIO* identifiés en analyse pan-cancer sur 9624 tumeurs (Gao *et al.*, 2018). Dans ce tableau, une ligne correspond à l'expression d'un transcrit chimérique impliquant *TRIO* dans une seule tumeur. CS : cellules squameuses.

comme des oncogènes forts (tels que *BCR-ABL1* et *EWSR1-FLI1*) qui sont généralement associés à une différenciation et/ou un phénotype agressif distincts. Ensuite, si l'expression d'un transcrit chimérique oncogène gouverne le profil transcriptomique des sarcomes à translocation (**partie 1.2.3.2 page 52**), ici l'expression de ces différents transcrits ne conduit pas à une quelconque spécificité transcriptomique.

TRIO-TERT étant le seul gène de fusion récurrent exprimé dans les SGC, l'oncogenèse de ces tumeurs n'est pas permise par l'expression de transcrits chimériques *drivers*. Nous avons par la suite continué d'explorer ces données en mettant l'accent sur les nombreux transcrits chimériques tronquants exprimés. La poursuite de ce travail a conduit à identifier des inactivations d'*ATRX* dans 16% de notre cohorte, par altérations structurales (délétions, translocations et inversions) et ponctuelles (mutations faux-sens et non-sens). Les conséquences fonctionnelles de l'inactivation d'*ATRX* font aujourd'hui l'objet d'un sujet de thèse au sein du laboratoire.

4 | Discussion et perspectives

L'objectif de ma thèse était d'étudier les conséquences transcriptomiques du remodelage génomique des sarcomes pléomorphes. Cette étude s'est appuyée sur deux axes, tenant compte de l'important remaniement du génome de ces tumeurs et de leur fort potentiel métastatique. La complétion de ces travaux nous a amené à valider l'application de la signature CINSARC sur le système nCounter (NanoString). Ceci permet d'obtenir la prédiction du risque métastatique, donné par la signature, en adéquation avec les pratiques cliniques courantes. Par ailleurs, l'analyse des transcrits chimériques exprimés dans ces tumeurs nous a conduit à identifier le seul gène de fusion exprimé de façon récurrent : *TRIO-TERT*.

4.1 – La signature CINSARC

Les travaux que j'ai eu l'occasion de réaliser durant ma thèse ont principalement porté sur la signature CINSARC. Un long chemin a été parcouru depuis sa définition il y a bientôt 10 ans (Chibon *et al.*, 2010) et les différentes études réalisées au cours de ma thèse ont contribué à mieux comprendre ses conséquences biologiques et à répondre aux enjeux cliniques de cette signature transcriptomique. Une revue sur CINSARC est présentée **annexe IV page 263** (Chibon *et al.*, 2019).

4.1.1 – Sa régulation

La complétion du projet TCGA sarcomes (The Cancer Genome Atlas Research Network, 2017) nous a donné l'opportunité d'étudier les possibles origines de l'expression de CINSARC au travers de puissants régulateurs de l'expression génique : la méthylation de l'ADN et l'expression de miRNA. Nous avons ainsi identifié les tumeurs à haut et bas risque métastatique selon la signature et réalisé des analyses différentielles entre ces groupes.

4.1.1.1 – Via la méthylation

Nous avons commencé par mesurer d'importantes modifications de méthylation de l'ADN entre groupes de pronostics différents, dont le plus agressif (CINSARC C2) est associé à une hypométhylation globale du génome. Ceci concorde avec les données de la littérature puisqu'une diminution de la méthylation est observée aux stades précoces de l'oncogenèse et est encore plus prononcée chez les tumeurs les plus agressives en contribuant activement au processus d'instabilité génétique (Ehrlich, 2002, 2009 ; Kulis & Esteller, 2010). Cela est cohérent avec le statut instable des tumeurs CINSARC C2 puisqu'elles présentent un plus grand nombre d'altérations quantitatives ainsi qu'une charge mutationnelle, une hétérogénéité et une ploïdie plus importantes (**figure 30 page 153**).

De façon plus spécifique, l'hypométhylation des régions péri-centromériques a pu être associée à des réarrangements génomiques de bras chromosomiques tels que le gain du 1q dans les hépatocarcinomes (Wong *et al.*, 2001) ou le gain du 8q dans les tumeurs de la prostate (Schulz *et al.*, 2002). De plus, les régions génomiques répétées sont normalement hautement méthylées afin de maintenir l'intégrité du génome et de prévenir les recombinaisons anormales (Kulis & Esteller, 2010). Puisque les SGC sont très réarrangées et plus particulièrement les CINSARC C2, il pourrait être informatif de comparer le niveau de méthylation autour des cassures génomiques par rapport aux régions non réarrangées.

Si le génome des tumeurs CINSARC C2 est globalement hypométhylé, nous n'avons pas identifié de méthylation préférentielle des régions promotrices des 67 gènes de la signature. Cela suggèrerait que la méthylation, bien qu'importante pour la progression tumorale, ne serait pas directement motrice de l'expression de CINSARC. Afin de préciser le rôle de la méthylation de l'ADN sur l'agressivité des SGC au travers de la signature CINSARC, il serait intéressant de mener des expériences supplémentaires. Nous pourrions mener, sur des lignées cellulaires établies au laboratoire, des expériences d'hypométhylation sur des lignées C1 (avec de l'azacitidine ou de la décitabine ; Stresemann & Lyko, 2008) et d'hyperméthylation sur des lignées C2 (avec du méthotrexate ; Nyce, 1989) afin de mesurer un potentiel impact sur l'expression de la signature. Le cas échéant, des expériences de caractérisation phénotypique (migration, invasion, prolifération et étude du potentiel métastatique *in vivo*) pourraient être envisagées.

4.1.1.2 – Via l'expression de miRNA

Les données de miRNA-seq nous ont permis de mesurer d'importantes variations entre les groupes CINSARC C1 et CINSARC C2. De nombreux oncomirs sont exprimés dans le groupe hautement métastatique (**figure 28 page 148**) et leur niveau d'expression corrèle avec celui de la signature CINSARC (**figure 29 page 151**). Dans le groupe CINSARC C1, nous avons identifié trois candidats susceptibles de réguler négativement l'expression de la signature. Des analyses additionnelles nous ont permis de retenir mir-29c, dont l'expression est significativement anti-corrélée à celle de 54 gènes de CINSARC (soit 81% de la signature).

Ce miRNA a d'importantes fonctions biologiques et son expression a déjà pu être corrélée à un potentiel anti-métastatique dans les MPNST, en régulant négativement l'expression de *MMP2* (Presneau *et al.*, 2013). De plus, il agit comme un régulateur négatif de la progression du cycle cellulaire via la sous-expression de *CCNE1* dans les carcinomes à cellules squameuses de l'œsophage (Ding *et al.*, 2011). Il régule négativement la prolifération en ciblant *RCC2* dans les carcinomes gastriques (Matsuo *et al.*, 2013). Il influe sur la méthylation de l'ADN en régulant *DNMT3A* et *DNMT3B*, deux méthyltransférases d'ADN, dans les carcinomes pulmonaires (Fabbri *et al.*, 2007). Il a un effet pro-apoptotique de concert avec p53 car il régule négativement l'expression de *CDC42* et *PIK3R1* (deux régulateurs négatifs de p53) dans le cancer du côlon (Park *et al.*, 2009). Finalement, sa diminution d'expression est corrélée à un mauvais pronostic dans plusieurs types de cancers comme le lymphome à cellule du manteau (Zhao *et al.*, 2010), l'hépatocarcinome (Dong *et al.*, 2016), le mésothéliome pleural malin (Pass *et al.*, 2010) ou encore le carcinome gastrique (Wang *et al.*, 2015).

Collectivement, ces données suggèrent que mir-29c agirait comme un suppresseur de tumeurs au niveau transcriptomique. Ce miRNA n'a qu'une seule cible de CINSARC connue (*CCNA2* ; Hafner *et al.*, 2010 ; Kishore *et al.*, 2011 ; Whisnant *et al.*, 2013 ; Memczak *et al.*, 2013) et sept cibles potentielles identifiées grâce à des algorithmes bioinformatiques (Andrés-León *et al.*, 2015) : *CDC7*, *ECT2*, *H2AFX*, *KIF4A*, *NDE1*, *TPX2* et *TRIP13* (*CCNA2* étant à la fois une cible prédite et validée). En conséquence, une plus grande exploration de ce miRNA pourra statuer sur sa relation avec l'expression de CINSARC. Nous pourrions envisager de le sur-exprimer dans des lignées C2 et d'inhiber son expression dans des lignées C1 afin de me-

surer la potentielle évolution de l'expression de la signature CINSARC et la modification de caractères phénotypiques (migration, invasion, prolifération et étude du potentiel métastatique *in vivo*). Des expériences pourraient aussi être réalisées afin de valider les interactions prédites entre mir-29c et gènes de CINSARC.

4.1.1.3 – Analyse critique des résultats obtenus

Ces données (méthylation et miRNA) constituent la première exploration bioinformatique d'une possible origine d'expression de la signature CINSARC. Il est certain que d'autres analyses devront être réalisées pour affiner ces premières observations. Tout d'abord, nous pourrions confirmer (voire compléter) ces résultats sur les autres cohortes des projets TCGA puisque nous avons démontré que CINSARC est universellement pronostique (Lesluyes *et al.*, 2017). Ceci d'autant que la cohorte TCGA sarcomes dispose d'un suivi des patients assez court (médiane de 2,29 ans [1,75-3,22]), impliquant qu'un certain nombre de tumeurs non métastatiques avec un suivi d'un an (seuil utilisé pour l'analyse) seront reclassées comme métastatiques avec un suivi plus important. La mise à jour de ces données dans quelques années permettra alors de mieux préciser le devenir métastatique, les analyses seront ainsi plus proches de la réalité biologique des tumeurs.

Les analyses réalisées ne statuent que sur une association directe de l'expression de la signature avec une possible voie de régulation. Cette approche est abordée en première intention afin d'obtenir un premier regard non supervisé sur ces données mais la biologie des tumeurs ne se résume pas uniquement à des associations aussi simples. Nous avons testé une association directe de ces facteurs avec les 67 gènes de CINSARC mais ces gènes agissent de concert avec d'autres au sein d'un réseau extrêmement bien régulé (Su *et al.*, 2004 ; Subramanian *et al.*, 2005 ; Szklarczyk *et al.*, 2015), aussi bien dans les tumeurs C1 que dans les C2. Il est donc tout à fait possible que d'autres membres de ce réseau soient ciblés par une méthylation particulière et/ou un miRNA et que la modification d'expression de CINSARC pourrait être indirecte, en passant par d'autres membres du réseau de co-expression. Il faudrait en ce sens définir le réseau complet de co-régulation autour de CINSARC pour préciser les analyses et peut être inclure les miRNA car l'expression de certains d'entre eux corrèle avec la signature. Ce genre d'approche,

notamment via la méthode *weighted gene co-expression network analysis* (Zhang & Horvath, 2005 ; Langfelder & Horvath, 2008), a permis d'identifier des réseaux transcriptomiques associés à l'évolution clinique des plusieurs cancers, comme celui du sein, de la glande surrénale ou encore de la vessie (Clarke *et al.*, 2013 ; Yuan *et al.*, 2018 ; Di *et al.*, 2019). Il serait aussi très intéressant de faire de la comparaison de réseaux entre celui identifié dans les sarcomes et ceux que nous pourrions identifier dans d'autres tumeurs via les données du TCGA. Ceci permettrait d'en mesurer la spécificité dans chaque type de cancer ou, au contraire, l'universalité.

Nous nous sommes ici restreints à l'exploration de données concernant la méthylation et l'expression de miRNA mais d'autres pistes encore non explorées pourraient être à l'origine de l'expression de la signature. Il faudrait réaliser une analyse intégrative, combinant tous les paramètres disponibles afin d'identifier une possible combinaison de facteurs (cliniques, génomiques, transcriptomiques, épigénétiques, *etc.*) expliquant au mieux la définition des groupes C1 et C2. Plusieurs algorithmes proposent ce genre d'analyse, comme iCluster ou PARADIGM qui ont été utilisés dans différents projets du TCGA (Shen *et al.*, 2009 ; Vaske *et al.*, 2010).

4.1.1.4 – Autres facteurs possibles

Les facteurs de transcription sont de puissants régulateurs du transcriptome et nombre d'entre eux sont impliqués dans le processus d'oncogenèse en tant qu'oncogène (*EWSR1-FLI1*, *JUN*, *MYC*, *STAT3*, *etc.* ; Lambert *et al.*, 2018). À ce titre, deux acteurs majeurs impliqués dans la régulation du cycle cellulaire pourraient concerner CINSARC. Premièrement, la famille des *E2F* est une famille de facteurs de transcription dont le rôle dans le cycle cellulaire est bien décrit (Thurlings & de Bruin, 2016). Ces protéines permettent, entre autre, la transition G1/S par expression de leurs gènes cibles sous contrôle de pRb. Deuxièmement, *FOXM1* (oncogène et lui-même membre de CINSARC), est également impliqué dans le cycle cellulaire, notamment pour la transition G2/M par expression de ses gènes cibles tout en étant négativement régulé par p53 (Liao *et al.*, 2018). Il est intéressant de noter que ces deux acteurs soient sous le contrôle de *RB1* et *TP53*, les deux gènes les plus altérés dans les SGC (à hauteur de respectivement 80% et 50% ; Chibon *et al.*, 2000 ; Pérot *et al.*, 2010). Il est tout à fait envisageable que ces facteurs

de transcription soient impliqués dans l'expression conditionnelle de CINSARC mais doivent s'accompagner d'autres facteurs car l'importante proportion de tumeurs ayant perdu *TP53* et/ou *RB1* ferait, sans ces facteurs additionnels, que la quasi-totalité des SGC seraient CINSARC C2, ce qui n'est pas le cas.

Une autre piste pourrait passer par la conformation de l'ADN dont l'architecture 3D peut être associée à la formation d'usines de transcription (Jackson *et al.*, 1993). De tels sites pourraient rassembler plusieurs régions génomiques contenant les gènes de CINSARC et délivrer un message transcriptomique coordonné. Un effort de caractérisation de l'architecture de l'ADN commence à émerger, notamment dans des lignées cellulaires (Dixon *et al.*, 2012 ; Schmitt *et al.*, 2016). Des expériences d'Hi-C (Lieberman-Aiden *et al.*, 2009) sont justement planifiées dans l'équipe et seront réalisées d'ici quelques mois, afin de mieux comprendre l'impact des altérations structurales sur l'architecture du génome des SGC. Il sera alors possible de tester si l'expression de CINSARC pourrait être permise via des usines de transcription particulières. Peut-il exister des différences spatiales de la répartition des gènes de CINSARC entre les cellules saines et les sarcomes, les sarcomes C1 et C2, ou même entre les sarcomes est les autres tumeurs ? Quelle peut être la contribution des réarrangements génomiques, plus importants dans les C2 que dans les C1, vis-à-vis de l'architecture du génome ? L'une des explications de la valeur pronostique de CINSARC au-delà des sarcomes pourrait précisément être la présence d'une usine de transcription qui soit universellement stable dans les cancers.

4.1.2 – N'est-ce "que" de l'instabilité ?

Comparativement aux autres types de cancers, les sarcomes pléomorphes présentent de hauts niveaux d'instabilité chromosomique (The Cancer Genome Atlas Research Network, 2017). Parmi cette instabilité, les SGC possédant le plus d'altérations structurales sont significativement associées à un statut CINSARC C2. Il a également été démontré que la sur-expression de plusieurs membres de CINSARC (tels que *MAD2L1*, *BUB1*, *CCNB1*, *CCNB2* et *ESPL1*) induit de l'instabilité chromosomique et permet le développement tumoral chez la souris (Sotillo *et al.*, 2007 ; Ricke *et al.*, 2011 ; Nam & van Deursen, 2014 ; Mukherjee *et al.*, 2014).

De récents travaux ont pu établir un lien direct entre cette instabilité génétique et l'acquisi-

tion d'un potentiel métastatique (Bakhoum *et al.*, 2018). Les cellules génétiquement instables, suite à des erreurs de ségrégation chromosomique, peuvent exclure un ou plusieurs chromosomes dans un micro-noyau. La présence d'ADN cytoplasmique, par rupture de la membrane nucléaire du micro-noyau, est alors détectée par la voie cGAS-STING capable d'activer la voie NF-κB de façon alternative et d'accroitre les capacités migratoires et invasives des cellules. Pour autant, l'acquisition d'un potentiel métastatique par l'expression de la signature ne serait-elle permise que par un phénotype d'instabilité génétique ?

Nous avons montré, dans deux lignées d'UPS CINSARC C2, que des sous-clones tétraploïdes étaient plus mobiles que les diploïdes (**annexe III page 242** ; Jemaà *et al.*, 2017). Si l'expression transcriptomique de la signature CINSARC ne varie pas entre clones diploïdes et clones tétraploïdes, des différences majeures sont néanmoins mesurées au niveau protéique. Les protéines codées par les gènes de CINSARC sont en effet retrouvées en quantité nettement supérieure dans les cellules tétraploïdes. De façon intéressante, la kinase aurora A (*AURKA*) ne semble pas dégradée en fin de mitose, ce qui laisse suspecter une défaillance du complexe de promotion de l'anaphase. Ce complexe permet la transition de la métaphase à l'anaphase en dégradant des protéines cibles (dont les kinases aurora), permettant, entre autres, le détachement des chromatides sœurs par dégradation de la sécurine puis la sortie de la phase mitotique via la dégradation de cyclines (Peters, 2006). Cette persistance protéique pourrait induire un phénotype de non-adhérence par modification du cytosquelette comme chez une cellule mitotique, permettant une mobilité accrue des cellules ce qui augmenterait leur potentiel métastatique. Il est dans ce contexte intéressant de questionner l'apport du doublement du génome sur le niveau d'expression protéique de CINSARC et sur ses conséquences, sachant que les tumeurs C2 ont des ploïdies et des fréquences de doublements génomiques supérieurs aux tumeurs C1.

Parmi la signature CINSARC, on dénombre huit kinésines (*KIF2C*, *KIF4A*, *KIF11*, *KIF14*, *KIF15*, *KIF18A*, *KIF20A* et *KIF23*). Les kinésines sont une famille de protéines motrices qui se déplacent le long des microtubules. Elles participent au processus de mitose et à l'architecture du cytosquelette en permettant le transport de complexes protéiques et d'organites cellulaires (Hirokawa *et al.*, 2009). Il est rapporté que deux de ces kinésines (*KIF11* et *KIF20A*) sont impliquées dans le processus d'angiogenèse (Exertier *et al.*, 2013 ; Robitaille *et al.*, 2014). Une conséquence fonctionnelle de la sur-expression de la signature pourrait alors être une aug-

mentation de la vascularisation tumorale, pouvant accroitre son potentiel métastatique.

Plusieurs moyens d'acquisition de capacités métastatiques seraient ainsi permis par l'expression de la signature CINSARC. Cela pourrait tout aussi bien passer de l'instabilité génétique que par un autre mécanisme, voire une combinaison de facteurs. L'hypothèse des conséquences fonctionnelles multiples serait en accord avec la valeur pronostique universelle de la signature car de grandes variations d'instabilité génétique sont observées entre les sarcomes et les autres types de tumeurs. Il est certain que cette instabilité joue un rôle crucial dans les SGC puisqu'elle est leur signature génétique, mais CINSARC pourrait jouer sur d'autres voies biologiques pour moduler le phénotype métastatique.

4.1.3 – Application clinique

Les différents travaux que nous avons réalisés ont permis de rendre CINSARC applicable en conditions cliniques. Le premier transfert technologie effectué nous a permis de démontrer que le RNA-seq est une technologique adaptée à l'obtention du pronostic sur des tissus congelés mais constitue une limitation majeure pour l'analyse de blocs FFPE avec près de 60% de cas exclus car trop dégradés (Lesluyes *et al.*, 2016). Avec le second transfert, nous avons démontré que le système nCounter est une solution éprouvée pour l'analyse de blocs FFPE (Le Guellec *et al.*, 2018).

Nous avons pu valider un système de quantification d'expression qui puisse être quotidiennement applicable dans un contexte clinique. Cette méthode est ainsi tout à fait comparable avec les autres techniques de biologie moléculaire (telles que la FISH ou la CGH-*array*) en matière de quantité de matériel requis, coût d'analyse et de temps de manipulation effectif (Baffert *et al.*, 2013 ; Perrier *et al.*, 2018). La phase finale de son application consiste en sa validation au travers d'essais cliniques, où l'impact de l'utilisation de la signature sur le pronostic des patients sera évalué.

4.1.3.1 – En route vers les essais cliniques

Plusieurs essais cliniques intégrant CINSARC sont en cours et d'autres sont sur le point de débuter. Ces essais testeront la valeur pronostique de CINSARC de façon rétrospective mais également prospective, ce qui devrait être la dernière étape de l'intégration de la signature dans les pratiques cliniques.

Études incluant une validation rétrospective de CINSARC :

ISG-STS 1001 est une étude de phase 3 du groupe EuroSarc ayant démontré qu'une chimiothérapie néoadjuvante guidée par le sous-type histologique n'apportait aucun bénéfice par rapport à une chimiothérapie conventionnelle (Gronchi *et al.*, 2017).

STRASS est une étude de phase 3 de l'*European Organisation for Research and Treatment of Cancer* qui évalue l'impact de la radiothérapie pré-opératoire pour les STM du rétropéritoine (NCT01344018).

Études prospectives basées sur CINSARC :

NEOSarcomics est une étude du GSF sur 200 patients qui évalue la valeur prédictive de réponse à une chimiothérapie néoadjuvante par rapport au pronostic donné par la signature CINSARC.

CIRSARC est une étude du GSF (financement PHRCK-2016) de phase 3 sur 300 patients atteints de STM non métastatiques et résécables de grade 3. Le groupe CINSARC C1 (estimé à 100 patients) sera traité de façon conventionnelle. Le groupe CINSARC C2 (estimé à 200 patients) sera randomisé (1:1) en deux groupes : traitement de trois cycles de chimiothérapie néoadjuvante comparativement à six cycles de chimiothérapie néoadjuvante.

CHIC-STS est une étude du GSF (financement PHRCK-2018) de phase 3 sur 500 patients atteints de STM non métastatiques et résécables de grades 1 et 2. Le groupe CINSARC C1 (estimé à 250 patients) sera traité de façon conventionnelle. Le groupe CINSARC C2 (estimé à 250 patients) sera randomisé (1:1) en deux groupes : sans et avec (quatre cycles) de chimiothérapie péri-opératoire (pouvant être adjuvante ou néoadjuvante).

Les trois études cliniques prospectives basées sur la valeur pronostique de la signature

CINSARC permettront de répondre à des enjeux médicaux cruciaux pour la prise en charge des patients. NEOSarcomics permettra d'évaluer la valeur prédictive de la réponse au traitement de la signature. CIRSARC permettra de statuer sur l'apport d'une chimiothérapie intensifiée en situation néoadjuvante. CHIC-STS permettra d'évaluer le bénéfice de la chimiothérapie périopératoire pour les tumeurs de bas grades (1 et 2) dont le consensus actuel est de ne pas les traiter par chimiothérapie, malgré un haut potentiel métastatique évalué par CINSARC. La validation au sein de ces essais cliniques permettra alors de proposer CINSARC comme la nouvelle référence pronostique pour la prise en charge des STM non métastatiques et résécables.

4.1.3.2 – Au-delà des sarcomes ?

Le premier rapport de la valeur pronostique de CINSARC (Chibon *et al.*, 2010) s'est enrichi avec une étude pan-cancer où nous démontrons sa capacité à prédire l'agressivité tumorale dans 17 types de cancers (Lesluyes *et al.*, 2017). La transversalité de la signature couplée avec une méthode d'analyse compatible avec les pratiques cliniques (Le Guellec *et al.*, 2018) nous donne l'opportunité de tester sa valeur pronostique dans d'autres types de cancers de façon rétrospective : les tumeurs musculaires lisses utérines, de l'endomètre et de l'ovaire.

Au-delà de sa valeur pronostique, des tests sont en cours pour définir sa potentielle utilité à visée diagnostique. Dans ce contexte, les léiomyomes dits "à noyaux bizarres" (LM-BN pour *leiomyoma with bizarre nuclei*) sont des tumeurs utérines bénignes mais dont l'aspect morphologique et l'analyse des altérations génomiques ne permettent pas toujours de les identifier comme tels et peuvent donc être confondus avec les LMS utérins qui sont des tumeurs très agressives (Croce *et al.*, 2014). En collaboration avec Sabrina Croce (pathologiste à l'Institut Bergonié, Bordeaux), nous testons l'apport de CINSARC pour discriminer les LM-BN parmi les léiomyomes conventionnels et les LMS utérins. Les premiers résultats obtenus sont encourageants en ce sens où, malgré leurs remaniements génomiques, ces tumeurs sont toutes classées CINSARC C1. Ces résultats préliminaires obtenus sur 10 cas seront confirmés par la constitution d'une cohorte plus importante. Ceci souligne aussi que CINSARC, au moins dans les LM-BN, n'est pas le simple reflet de l'instabilité génétique puisque ces tumeurs peuvent être aussi remaniées que les LMS utérins.

4.1.3.3 – Futurs challenges

De nouveaux défis attendent CINSARC sur le terrain clinique. L'une des limitations du grade FNCLCC est sa difficulté d'application sur micro-biopsies car ces fragments tumoraux peuvent ne pas être représentatifs du phénotype global de la tumeur. Sur le même principe, nous n'avons encore jamais mesuré comment l'hétérogénéité intra-tumorale peut influencer l'expression de la signature, ce d'autant que les tumeurs CINSARC C2 sont plus hétérogènes que les CINSARC C1 (**figure 25D page 142**).

L'hétérogénéité que nous avons mesurée dans deux lignées de sarcomes montre des phénotypes cellulaires distincts entre cellules diploïdes et cellules tétraploïdes (**annexe III page 242** ; Jemaà *et al.*, 2017). L'expression de CINSARC ne semble pas différente entre ces sous-clones mais nous n'avons pas la puissance statistique pour le démontrer (deux clones diploïdes et deux clones tétraploïdes). La valeur pronostique donnée par CINSARC est basée sur le message transcriptomique d'une partie de la tumeur. Ceci est d'autant plus important que nous avons démontré que CINSARC, via nCounter (Le Guellec *et al.*, 2018), peut être appliqué sur micro-biopsies. Quelle peut alors être la variation d'expression de la signature dans différents fragments tumoraux ?

Des analyses transcriptomiques de cellules uniques (type *single-cell* RNA-seq ; Tang *et al.*, 2009) pourraient mettre en lumière un lignage de sous-clones qui possèderait un haut potentiel métastatique au sein de la tumeur selon CINSARC. Si un tel lignage peut exister, alors en quelle proportion cette population de cellules doit être présente pour que le risque métastatique soit détecté comme élevé ? De façon alternative, l'expression de la signature peut-elle être précoce lors de l'initiation tumorale et stable dans le temps de telle sorte que ce profil transcriptomique ne soit que peu variable entre les différents sous-clones ? Ces questions sont d'intérêts majeurs car la dynamique de l'hétérogénéité intra-tumorale pourrait permettre d'expliquer une partie des différences mesurées en matière de pronostic et de devenir clinique.

La détection des ARN circulants de CINSARC pourrait être pertinente car ces gènes sont très exprimés pour les tumeurs C2 et pourraient ainsi être détectés (malgré leur instabilité et leur dégradation). La présence de certains ARN circulants a pu être corrélée à de mauvais

pronostics dans les cancers du sein (*BMI1* et *CCND1* ; Silva *et al.*, 2007 ; García *et al.*, 2008), les lymphomes (*MYC* ; Garcia *et al.*, 2009), les tumeurs de la prostate (*TERT* ; March-Villalba *et al.*, 2012), *etc.* Ceci pourrait constituer un test non invasif de choix dans le suivi des patients.

D'un point de vue thérapeutique et selon la composition de CINSARC, nous nous demandons si la diminution de son expression pourrait être envisagée comme une solution thérapeutique. Les kinases aurora (dont *AURKA* et *AURKB* qui font partie de CINSARC) sont des sérines/thréonines kinases qui jouent un rôle important dans la ségrégation des chromosomes. L'expression de gènes est augmentée dans de nombreux cancers (Fu *et al.*, 2007) et de nombreux inhibiteurs sont actuellement testés dans des essais de phases 1/2 sur différents types de cancers (Tang *et al.*, 2017 ; Borisa & Bhatt, 2017) et pourraient être testés dans les sarcomes (Nair & Schwartz, 2016 ; Toulmonde, 2016).

4.2 – Où en sommes nous avec la génétique des sarcomes ?

4.2.1 – L'expression de gènes de fusion dans les sarcomes

Le travail que nous avons effectué sur les gènes de fusion nous a permis d'identifier *TRIO-TERT* comme le seul de transcrit chimérique récurrent dans les SGC (Delespaul *et al.*, 2017). Ces résultats ont par la suite été validés dans d'autres études indépendantes (The Cancer Genome Atlas Research Network, 2017 ; Ogura *et al.*, 2018 ; Suster *et al.*, 2019). Nous avons ainsi caractérisé les gènes de fusion exprimés dans ces tumeurs et, en dépit des nombreux remaniements génomiques (incluant les translocations) observés, nous pouvons en conclure que l'oncogenèse des SGC ne passe pas par une protéine chimérique oncogène.

4.2.1.1 – Identification de nouveaux gènes de fusion dans les SGS

L'exploration de ces données nous a permis d'acquérir une solide expérience dans la détection et la validation des transcrits chimériques. Nous avons eu l'occasion de mettre cette expérience en pratique dans d'autres cohortes de sarcomes atypiques, ne présentant pas leurs altérations génétiques caractéristiques.

Nous avons débuté avec un cas de dermatofibrosarcome de Darier-Ferrand négatif pour le transcrit chimérique classique *COL1A1-PDGFB*, habituellement retrouvé dans 96% des cas. Nous avons pu identifier et valider l'expression d'un transcrit chimérique alternatif : *COL6A3-PDGFD*. L'analyse d'une cohorte de 19 cas supplémentaires confirma la présence récurrente de *COL6A3-PDGFD* (9 cas sur 20) mais également l'expression de *EMILIN2-PDGFD* (2 cas sur 20). Sur les neufs derniers cas, huit exprimaient bien le transcrit chimérique classique *COL1A1-PDGFB* non retrouvé avec les outils classiques de diagnostic : FISH et PCR. Ceci pouvant s'expliquer par la grande variabilité de la localisation du point de fusion à l'origine de l'expression du transcrit chimérique. Si ces résultats apportent de nouveaux éléments diagnostiques concrets, des analyses supplémentaires seront à réaliser pour comprendre les effets des alté-

rations de *PDGFD* comparativement à *PDGFB* sur la biologie des dermatofibrosarcomes de Darier-Ferrand. L'article est présenté **annexe V page 269** (Dadone-Montaudié *et al.*, 2018) et les résultats furent confirmés moins d'un mois après par une étude américaine (Dickson *et al.*, 2018).

La détection de transcrits chimériques s'est poursuivie dans un cas particulièrement rare : une tumeur utérine ressemblant à une tumeur ovarienne des cordons sexuels (UTROSCT pour *uterine tumor resembling ovarian sex cord tumor*). Ces tumeurs ne présentent pas d'altération génétique particulière et sont généralement peu agressives (Kurman *et al.*, 2014). Toutefois, une tumeur montra un phénotype agressif avec une rechute locale 17 mois après le diagnostic puis des métastases pulmonaires et péritonéales encore un an après. La CGH-*array* de la tumeur primaire ne montra aucune altération mais la rechute présenta deux gains spécifiques qui pouvaient laisser suspecter la présence d'une translocation. Le RNA-seq confirma cette suspicion en identifiant l'expression de *GREB1-CTNNB1*. Ce réarrangement produit un codon stop dans l'intron 8 de *GREB1* et un codon d'initiation de *CTNNB1* à partir de son exon 4, ne codant pas le domaine de dégradation de la β-caténine. L'immunohistochimie confirme ces observations avec une forte positivité nucléaire de cette protéine et le transcriptome montra aussi une activation de la voie Wnt/β-caténine. *GREB1* jouerait alors le rôle d'un promoteur fort, ce gène étant fortement exprimé dans les tissus en réponse aux œstrogènes. Même si cette altération semble privée, n'ayant pas été retrouvée dans 11 UTROSCT additionnelles, nous avons rapporté la première hyper-activation de la β-caténine induite par une translocation car le gène n'était jusqu'ici décrit que par des mutations, exclusivement au niveau de l'exon 3. L'article est présenté **annexe VI page 280** (Croce *et al.*, 2019).

Nous avons finalement caractérisé une entité de sarcomes rares : les sarcomes nasosinusiens biphénotypiques. Ces tumeurs se développent dans la sphère nasale, sont localement agressives et expriment le gène de fusion récurrent *PAX3-MAML3* (Lewis *et al.*, 2012 ; Wang *et al.*, 2014). Nous avons constitué une cohorte de 41 cas dont 37 (90%) exprimaient bien *PAX3-MAML3*. Les quatre derniers cas, négatifs pour le gène de fusion classique par les techniques de biologie moléculaire (PCR et FISH) ont été passés en RNA-seq. Ceci nous a permis d'identifier l'expression de transcrits chimériques impliquant *PAX3* avec trois partenaires : *FOXO1* (un cas), *NCOA2* (un cas) et *WWTR1* (deux cas). L'existence de *PAX3-FOXO1* est déjà rapportée dans

la littérature pour ces tumeurs (Wong *et al.*, 2016) et l'existence de *PAX3-NCOA2* était suspectée car *NCOA2* partage des similarités structurelles et fonctionnelles avec *MAML3*, *FOXO1* et *NCOA1*, partenaires de *PAX3* précédemment identifiés (Fritchie *et al.*, 2016). L'implication de *WWTR1* était jusqu'ici non évoquée, il est avec *YAP1* l'un des régulateurs maitres de la voie Hippo impliquée dans la prolifération et la différenciation. L'article est présenté **annexe VII page 289** (Le Loarer *et al.*, 2019).

Le RNA-seq est bien la meilleure méthode pour identifier de nouveaux transcrits de fusion sans *a priori* (Nakano & Takahashi, 2018b). Ceci est particulièrement vrai pour les sarcomes "à translocation" atypiques, sans expression du transcrit de fusion classique, où des partenaires alternatifs peuvent être retrouvés (*PAX3* ou *PAX7* avec *FOXO1* pour les RMS alvéolaires, *SS18* avec *SSX1* ou *SSX2* dans les synovialosarcomes, *etc.* ; **partie 1.2.2.2 page 37**). Ceci souligne la spécificité de l'oncogenèse de ces tumeurs où quelques variations génétiques sont tolérées mais doivent aboutir à un message oncogène attendu.

4.2.1.2 – Transcrits chimériques en l'absence d'altérations génétiques

Dans notre analyse (Delespaul *et al.*, 2017), nous avons exclu les *read-throughs* qui sont des transcrits chimériques issus de la transcription de deux gènes adjacents (Akiva *et al.*, 2006) et qui ne sont donc pas formés par un réarrangement génomique. Toutefois, quel peut être l'impact transcriptomique de l'expression de ces *read-throughs* dans les cancers ? Si cette question reste en suspens dans les sarcomes, pour lesquels des efforts de caractérisation sont à planifier, l'implication de ces transcrits particuliers dans d'autres pathologies a été rapportée. On note par exemple l'expression de *CTSD-IFITM10* dans les cancers du sein (lignées cellulaires et tumeurs primaires) mais pas dans les tissus normaux (mammaires ou d'autres origines tissulaires ; Varley *et al.*, 2014). Ce *read-through* permettrait la conservation du cadre de lecture et la synthèse d'une protéine chimérique. Il a expérimentalement été démontré que la diminution de son expression diminue la prolifération cellulaire. D'autres *read-throughs* ont été caractérisés dans les tumeurs de la prostate, rénales, gastriques, *etc.* (Kumar-Sinha *et al.*, 2012 ; Grosso *et al.*, 2015 ; Choi *et al.*, 2016) et sont associés à une augmentation de l'agressivité tumorale. Le rôle de l'expression des *read-throughs* dans les sarcomes doit donc encore être précisé.

Une autre catégorie de transcrits chimériques existe et sont issus de la fusion de deux ARN pré-messagers via un mécanisme de *trans-splicing* (Lei *et al.*, 2016). S'il est estimé que 4 à 5% des couples de gènes adjacents dans le génome humain seraient capables de transcrire un *read-through* (Parra *et al.*, 2006), les transcrits chimériques issus d'un *trans-splicing* chez l'humain sont très rares. Deux cas notables ont toutefois été rapportés et concernent l'expression de deux transcrits chimériques oncogènes retrouvés dans des cellules saines, sans remaniement génomique. Le premier est *JAZF1-JJAZ1*, retrouvé dans des cellules de l'endomètre (Li *et al.*, 2008) mais dont l'expression est normalement permise par la translocation récurrente t(7;17)(p15;q21) dans 50% des tumeurs stromales de l'endomètre (Koontz *et al.*, 2001). Le second est *PAX3-FOXO1*, retrouvé dans des cellules en cours de différenciation vers du muscle squelettique (Yuan *et al.*, 2013) et qui lui aussi est normalement exprimé suite à la translocation récurrente t(2;13)(q35;q14) dans 65% des RMS alvéolaires (Barr *et al.*, 1993). Une caractérisation supplémentaire de ces transcrits chimériques sera nécessaire pour déterminer si leur expression peut favoriser l'initiation tumorale par pression d'acquisition des translocations, ou s'il existe une cause commune entre le *trans-splicing* et la formation de translocations (comme une conformation 3D particulière de l'ADN ; Jividen & Li, 2014).

4.2.2 – Le remaniement du génome

Dans les SGC, la très grande majorité des transcrits chimériques que nous avons identifié sont tronquants (Delespaul *et al.*, 2017), ne permettant pas la conservation du cadre de lecture du second partenaire. Une analyse sommaire sur les gènes retrouvés dans ces transcrits nous informe qu'il s'agit de gènes bien connus pour leur rôle de suppresseur de tumeurs et souvent ciblés par des altérations génomiques qui peuvent conduire à l'expression de transcrits chimériques : *TP53*, *ATRX* et *RB1* (Kanezaki *et al.*, 2006 ; Lee *et al.*, 2015 ; Hu *et al.*, 2018).

Ceci est en accord avec les observations du TCGA sarcomes où il est conclu que les SGC, hormis les DDLPS avec leurs amplifications spécifiques, présentent plus de pertes génomiques que d'amplifications et plus de mutations d'intérêt dans des gènes suppresseurs de tumeurs que dans des oncogènes (The Cancer Genome Atlas Research Network, 2017). Malgré d'importants efforts de caractérisation génétique, nous faisons toujours face à une absence flagrante

d'altération motrice dans ces tumeurs.

Nous sommes maintenant à un tournant génétique concernant les SGC car la grande majorité des altérations sont privées et les quelques altérations récurrentes identifiées ne sont pas spécifiques aux différents sous-types histologiques. De plus, ces dernières concernent principalement des inactivations de gènes suppresseurs de tumeurs (tels que *RB1*, *TP53*, *ATRX*, *PTEN*, *CDKN2A*, *NF1*, *etc.* ; The Cancer Genome Atlas Research Network, 2017). En l'absence d'altérations spécifiques et récurrentes, nous n'avons pas encore déterminé quelle est la voie d'oncogenèse entreprise par ces tumeurs, ni même s'il peut y en avoir plusieurs.

Les SGC présentent des altérations génétiques très privées mais il est tout à fait envisageable que ces différentes altérations impactent des voies de signalisation communes (Sanchez-Vega *et al.*, 2018). Des analyses intégrant des données multidimensionnelles (niveaux d'expression, variations du nombre de copies, points de cassures génomiques, altérations ponctuelles, méthylation, *etc.*) pourraient ainsi permettre d'identifier ces voies en l'absence de gène récurrent.

Puisque l'instabilité est au cœur de la génétique des sarcomes pléomorphes, il est tentant de postuler qu'un évènement de remaniement génomique puisse être en lui-même une altération oncogène motrice (Delespaul, 2018). Dans ce modèle, les inactivations récurrentes de gènes suppresseurs de tumeurs pourraient faciliter la survenue de cette catastrophe chromosomique et/ou permettraient aux cellules d'y résister. De plus, le remaniement global du génome pourrait activer un message oncogène par altérations (génétiques, épigénétiques, transcriptomiques, *etc.*) de gènes ayant une faible capacité oncogénique mais dont l'effet cumulé pourrait égaler celui d'un oncogène fort (Davoli *et al.*, 2013). Ceci expliquerait pourquoi les altérations identifiées dans ces tumeurs sont majoritairement privées et pourquoi aucune spécificité transcriptomique n'est identifiée.

L'hypothèse d'un remaniement génomique comme événement moteur de l'oncogenèse n'explique pas la diversité des sous-types histologiques des SGC car ces différents sous-types partagent une grande instabilité et un transcriptome similaire (excepté les LMS présentant une amplification de *MYOCD* qui ont une différenciation particulière ; Segal *et al.*, 2003 ; Baird *et al.*, 2005 ; Pérot *et al.*, 2009 ; Gibault *et al.*, 2011). Le contexte cellulaire dans lequel cette

instabilité s'opère pourrait alors être à l'origine d'une différenciation particulière. Ainsi, une catastrophe chromosomique dans une cellule musculaire lisse pourrait par exemple aboutir à l'initiation d'un LMS alors que le même événement dans un fibroblaste pourrait initier un UPS. Il serait alors très intéressant de se pencher sur l'apport du contexte mésenchymateux concernant l'acquisition et la résistance aux altérations génomiques.

5 | Conclusion

Les sarcomes pléomorphes, dits "à génétique complexe" (SGC), sont un groupe de tumeurs d'origine mésenchymateuse présentant une grande variabilité en matière de caractéristiques cliniques. Les précédentes analyses de cytogénétique ont permis de mettre en évidence de nombreux réarrangements génomiques dans ces tumeurs. Toutefois, leur processus d'initiation reste mal compris en raison de l'absence d'altération récurrente et spécifique. Ces tumeurs sont également agressives, avec des taux de survie sans métastase de 50% à 5 ans et pour lesquelles la chirurgie reste le traitement ayant le plus d'impact sur le devenir clinique des patients.

La signature CINSARC

De précédents travaux de l'équipe ont établi un lien entre l'instabilité génétique des SGC et leur agressivité au travers de l'expression d'une signature transcriptomique nommée CINSARC. Cette signature a une meilleure valeur pronostique que le standard d'évaluation international, le grade FNCLCC, mais son application était jusqu'alors restreinte aux ARN extraits de tissus congelés et analysés sur puces d'expression. Nous avons ainsi mené différents travaux permettant de rendre CINSARC applicable aux centres de soins via deux technologies : le séquençage des ARN (RNA-seq) et le système nCounter, ce dernier pouvant conjointement quantifier des ARN extraits de tissus congelés ou bien fixés. Le potentiel médical de cette signature sera prochainement évalué au travers de deux essais cliniques de phase 3. Nous avons également montré que l'expression de CINSARC est un facteur pronostique au-delà des sarcomes, soulignant son universalité et sa possible application dans d'autres pathologies. En conséquence, des tests sont en cours pour évaluer sa pertinence vis-à-vis d'autres cancers.

Nous avons finalement recherché ses possibles mécanismes de régulation grâce aux données rendues publiques par la complétion du projet TCGA sarcomes : la méthylation de l'ADN et l'expression de micro-ARN. Si d'importantes modifications du méthylome sont observées entre groupes de haut et de bas risque métastatique, les régions promotrices des gènes composant CINSARC ne sont pas différentiellement méthylées. Parmi les micro-ARN différen-

tiellement exprimés entre ces groupes, mir-29c serait un bon candidat à la régulation négative d'expression de la signature. En effet, il est sur-exprimé dans le groupe de bon pronostic et son expression est significativement anti-corrélée à une bonne partie (81%) de la signature. Ces premières observations nécessiteront de plus amples investigations pour statuer sur la contribution de ces deux facteurs à la régulation de CINSARC. Des analyses additionnelles sur d'autres acteurs potentiels, comme les facteurs de transcription ou la formation d'une usine de transcription, pourraient également mettre en évidence de nouvelles relations avec l'expression de la signature.

L'expression de gènes de fusion

Si de nombreuses translocations ont été mises en évidence dans les SGC, l'expression de gènes de fusion qui y serait associée restait inexplorée car la nature et la résolution des précédentes analyses de cytogénétique ne permettaient pas leur détection. Nous avons séquencé le transcriptome de plus d'une centaine de sarcomes pléomorphes afin de déterminer si l'initiation de ces tumeurs pouvait être permise par la formation d'une protéine chimérique motrice de l'oncogenèse, comme cela est le cas dans plus d'une dizaine de sous-types de sarcomes dits "à translocation". Cette étude nous a conduit à identifier *TRIO-TERT* comme le seul gène de fusion récurrent (5% des cas) et conservant le cadre de lecture des deux partenaires. L'expression de *TRIO-TERT* fut par la suite confirmée dans d'autres études indépendantes. Toutefois, ce gène de fusion n'est pas spécifique d'un sous-type histologique et n'induit pas un profil transcriptomique spécifique. Ces résultats démontrent ainsi que l'oncogenèse des SGC ne passe pas par l'expression d'un gène de fusion de forte capacité oncogénique.

L'expérience acquise lors de l'étude menée sur *TRIO-TERT* nous a amené à poursuivre la détection de transcrits chimériques dans des sarcomes dits "à génétique simple" mais laissant suspecter la présence de translocations ayant un impact fonctionnel majeur. Nous avons identifié plusieurs gènes de fusion qui n'étaient jusque-là pas rapportés dans la littérature : *COL6A3-PDGFD* et *EMILIN2-PDGFD* dans des dermatofibrosarcomes de Darier-Ferrand, *GREB1-CTNNB1* dans une tumeur utérine ressemblant à une tumeur ovarienne des cordons sexuels, ainsi que *PAX3-NCOA2* et *PAX3-WWTR1* dans des sarcomes nasosinusiens biphénotypiques.

Bibliographie

Adamowicz M., Radlwimmer B., Rieker R. J., Mertens D., *et al.* (2006). *Frequent amplifications and abundant expression of TRIO, NKD2, and IRX2 in soft tissue sarcomas*. *Genes Chromosomes Cancer* 45:829–838.

Akiva P., Toporik A., Edelheit S., Peretz Y., *et al.* (2006). *Transcription-mediated gene fusion in the human genome*. *Genome Res.* 16:30–36.

Alexandrov L. B., Nik-Zainal S., Wedge D. C., Aparicio S. A. J. R., *et al.* (2013). *Signatures of mutational processes in human cancer*. *Nature* 500:415–421.

Allis C. D. & Jenuwein T. (2016). *The molecular hallmarks of epigenetic control*. *Nat. Rev. Genet.* 17:487–500.

Alvarado M. D., Prasad C., Rothney M., Cherbavaz D. B., *et al.* (2015). *A Prospective Comparison of the 21-Gene Recurrence Score and the PAM50-Based Prosigna in Estrogen Receptor-Positive Early-Stage Breast Cancer*. *Adv Ther* 32:1237–1247.

Anderson N. D., de Borja R., Young M. D., Fuligni F., *et al.* (2018). *Rearrangement bursts generate canonical gene fusions in bone and soft tissue tumors*. *Science* 361.

Andrés-León E., González Peña D., Gómez-López G. & Pisano D. G. (2015). *miRGate: a curated database of human, mouse and rat miRNA-mRNA targets*. *Database (Oxford)* 2015:bav035.

Annala M. J., Parker B. C., Zhang W. & Nykter M. (2013). *Fusion genes and their discovery using high throughput sequencing*. *Cancer Lett.* 340:192–200.

Arbiser Z. K., Folpe A. L. & Weiss S. W. (2001). *Consultative (expert) second opinions in soft tissue pathology. Analysis of problem-prone diagnostic situations*. *Am. J. Clin. Pathol.* 116:473–476.

Aurias A., Rimbaut C., Buffe D., Zucker J. M. & Mazabraud A. (1984). *Translocation involving chromosome 22 in Ewing's sarcoma. A cytogenetic study of four fresh tumors*. *Cancer Genet. Cytogenet.* 12:21–25.

Baca S. C., Prandi D., Lawrence M. S., Mosquera J. M., *et al.* (2013). *Punctuated evolution of prostate cancer genomes*. *Cell* 153:666–677.

Baffert S., Italiano A., Pierron G., Traoré M.-A., *et al.* (2013). *Comparative cost analysis of molecular biology methods in the diagnosis of sarcomas*. *Bull Cancer* 100:963–971.

Bailey M. H., Tokheim C., Porta-Pardo E., Sengupta S., *et al.* (2018). *Comprehensive Characterization of Cancer Driver Genes and Mutations*. *Cell* 174:1034–1035.

Bainbridge M. N., Warren R. L., Hirst M., Romanuik T., *et al.* (2006). *Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach*. *BMC Ge-*

nomics 7:246.

Baird K., Davis S., Antonescu C. R., Harper U. L., *et al.* (2005). *Gene expression profiling of human sarcomas: insights into sarcoma biology*. *Cancer Res.* 65:9226–9235.

Bakhoum S. F., Ngo B., Laughney A. M., Cavallo J.-A., *et al.* (2018). *Chromosomal instability drives metastasis through a cytosolic DNA response*. *Nature* 553:467–472.

Baldini E. H., Le Cesne A. & Trent J. C. (2018). *Neoadjuvant Chemotherapy, Concurrent Chemoradiation, and Adjuvant Chemotherapy for High-Risk Extremity Soft Tissue Sarcoma*. *Am Soc Clin Oncol Educ Book* pp. 910–915.

Ballinger M. L., Goode D. L., Ray-Coquard I., James P. A., *et al.* (2016). *Monogenic and polygenic determinants of sarcoma risk: an international genetic study*. *Lancet Oncol.* 17:1261–1271.

Balmain A. (2001). *Cancer genetics: from Boveri and Mendel to microarrays*. *Nat. Rev. Cancer* 1:77–82.

Banito A., Li X., Laporte A. N., Roe J.-S., *et al.* (2018). *The SS18-SSX Oncoprotein Hijacks KDM2B-PRC1.1 to Drive Synovial Sarcoma*. *Cancer Cell* 33:527–541.e8.

Barr F. G., Galili N., Holick J., Biegel J. A., Rovera G. & Emanuel B. S. (1993). *Rearrangement of the PAX3 paired box gene in the paediatric solid tumour alveolar rhabdomyosarcoma*. *Nat. Genet.* 3:113–117.

Barretina J., Taylor B. S., Banerji S., Ramos A. H., *et al.* (2010). *Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy*. *Nat. Genet.* 42:715–721.

Bartel D. P. (2009). *MicroRNAs: target recognition and regulatory functions*. *Cell* 136:215–233.

Bartlett J. M. S., Bayani J., Marshall A., Dunn J. A., *et al.* (2016). *Comparing Breast Cancer Multiparameter Tests in the OPTIMA Prelim Trial: No Test Is More Equal Than the Others*. *J. Natl. Cancer Inst.* 108.

Baylin S. B. & Ohm J. E. (2006). *Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nat. Rev. Cancer* 6:107–116.

Beck A. H., Lee C.-H., Witten D. M., Gleason B. C., *et al.* (2010). *Discovery of molecular subtypes in leiomyosarcoma through integrative molecular profiling*. *Oncogene* 29:845–854.

Bell D. W. (2010). *Our changing view of the genomic landscape of cancer*. *J. Pathol.* 220:231–243.

Ben Ami E., Hemming M. L., Nathenson M., Thornton K. A., *et al.* (2018). *Targeted tumor profiling and actionable somatic variants in sarcoma.* Présenté à : ASCO 2018, Chicago.

Berger S. L., Kouzarides T., Shiekhattar R. & Shilatifard A. (2009). *An operational definition of epigenetics*. *Genes Dev.* 23:781–783.

Betel D., Koppal A., Agius P., Sander C. & Leslie C. (2010). *Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites*. *Genome Biol.* 11:R90.

Bignold L. P., Coghlan B. L. D. & Jersmann H. P. A. (2006). *Hansemann, Boveri, chromosomes and the gametogenesis-related theories of tumours*. *Cell Biol. Int.* 30:640–644.

Bird A. (2002). *DNA methylation patterns and epigenetic memory*. *Genes Dev.* 16:6–21.

Blair L. P. & Yan Q. (2012). *Epigenetic mechanisms in commonly occurring cancers*. *DNA Cell Biol.* 31 Suppl 1:S49–61.

Blauvelt A. (1999). *The role of human herpesvirus 8 in the pathogenesis of Kaposi's sarcoma*. *Adv Dermatol* 14:167–206; discussion 207.

Borisa A. C. & Bhatt H. G. (2017). *A comprehensive review on Aurora kinase: Small molecule inhibitors and clinical trial studies*. *Eur J Med Chem* 140:1–19.

Boveri T. H. (1904). *Ergebnisse über die Konstitution der chromatischen Substanz des Zellkerns*. *1st edition*, *Gustav Fischer*, *Jena*.

Boveri T. H. (1914). *Zur Frage der Entstehung maligner Tumouren*. *1st edition*, *Gustav Fischer*, *Jena*.

Brady M. S., Gaynor J. J. & Brennan M. F. (1992). *Radiation-associated sarcoma of bone and soft tissue*. *Arch. Surg.* 127:1379–1385.

Brulard C. & Chibon F. (2013). *Robust gene expression signature is not merely a significant P value*. *Eur. J. Cancer* 49:2771–2773.

Butcher L. M. & Beck S. (2015). *Probe Lasso: a novel method to rope in differentially methylated regions with 450K DNA methylation data*. *Methods* 72:21–28.

Byron S. A., Van Keuren-Jensen K. R., Engelthaler D. M., Carpten J. D. & Craig D. W. (2016). *Translating RNA sequencing into clinical diagnostics: opportunities and challenges*. *Nat. Rev. Genet.* 17:257–271.

Campbell P. J., Yachida S., Mudie L. J., Stephens P. J., *et al.* (2010). *The patterns and dynamics of genomic instability in metastatic pancreatic cancer*. *Nature* 467:1109–1113.

Cardoso F., van't Veer L. J., Bogaerts J., Slaets L., *et al.* (2016). *70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer*. *N. Engl. J. Med.* 375:717–729.

Carter S. L., Eklund A. C., Kohane I. S., Harris L. N. & Szallasi Z. (2006). *A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers*. *Nat. Genet.* 38:1043–1048.

Carter S. L., Cibulskis K., Helman E., McKenna A., *et al.* (2012). *Absolute quantification of somatic DNA alterations in human cancer*. *Nat. Biotechnol.* 30:413–421.

Casali P. G., Abecassis N., Bauer S., Biagini R., *et al.* (2018). *Soft tissue and visceral sarcomas: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up*. *Ann. Oncol.*

Casamassimi A., Federico A., Rienzo M., Esposito S. & Ciccodicola A. (2017). *Transcriptome Profiling in Human Diseases: New Advances and Perspectives*. *Int J Mol Sci* 18.

Chang M. C., Souter L. H., Kamel-Reid S., Rutherford M., *et al.* (2017). *Clinical utility of multigene profiling assays in early-stage breast cancer*. *Curr Oncol* 24:e403–e422.

Chang K. T. E., Goytain A., Tucker T., Karsan A., *et al.* (2018). *Development and Evaluation of a Pan-Sarcoma Fusion Gene Detection Assay Using the NanoString nCounter Platform*. *J Mol Diagn* 20:63–77.

Charville G. W. & Lazar A. J. (2018). *Prognostic Gene Expression Signatures in Sarcoma: Finding Clarity in Complexity*. *Ann. Oncol.*

Chen J.-J., Knudsen S., Mazin W., Dahlgaard J. & Zhang B. (2012). *A 71-gene signature of TRAIL sensitivity in cancer cells*. *Mol. Cancer Ther.* 11:34–44.

Chepelev I., Wei G., Tang Q. & Zhao K. (2009). *Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq*. *Nucleic Acids Res.* 37:e106.

Chi P., Chen Y., Zhang L., Guo X., *et al.* (2010). *ETV1 is a lineage survival factor that cooperates with KIT in gastrointestinal stromal tumours*. *Nature* 467:849–853.

Chibon F., Mairal A., Fréneaux P., Terrier P., *et al.* (2000). *The RB1 gene is the target of chromosome 13 deletions in malignant fibrous histiocytoma*. *Cancer Res.* 60:6339–6345.

Chibon F., Mariani O., Derré J., Malinge S., *et al.* (2002). *A subgroup of malignant fibrous histiocytomas is associated with genetic changes similar to those of well-differentiated liposarcomas*. *Cancer Genet. Cytogenet.* 139:24–29.

Chibon F., Mariani O., Mairal A., Derré J., *et al.* (2003). *The use of clustering software for the classification of comparative genomic hybridization data. an analysis of 109 malignant fibrous histiocytomas*. *Cancer Genet. Cytogenet.* 141:75–78.

Chibon F., Mariani O., Derré J., Mairal A., *et al.* (2004). *ASK1 (MAP3K5) as a potential therapeutic target in malignant fibrous histiocytomas with 12q14-q15 and 6q23 amplifications*. *Genes Chromosomes Cancer* 40:32–37.

Chibon F., Lagarde P., Salas S., Pérot G., *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*. *Nat. Med.* 16:781–787.

Chibon F., Lesluyes T., Valentin T. & Le Guellec S. (2019). *CINSARC signature as a prognostic marker for clinical outcome in sarcomas and beyond*. *Genes Chromosomes Cancer* 58:124–129.

Cho Y. J., Kim S. H., Kim E. K., Han J. W., *et al.* (2018). *Prognostic implications of polycomb proteins ezh2, suz12, and eed1 and histone modification by H3K27me3 in sarcoma*. *BMC Cancer* 18:158.

Choi E.-S., Lee H., Lee C.-H. & Goh S.-H. (2016). *Overexpression of KLHL23 protein from read-through transcription of PHOSPHO2-KLHL23 in gastric cancer increases cell proliferation*. *FEBS Open Bio* 6:1155–1164.

Chou C.-H., Shrestha S., Yang C.-D., Chang N.-W., *et al.* (2018). *miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions*. *Nucleic Acids Res.* 46:D296–D302.

Chudasama P., Mughal S. S., Sanders M. A., Hübschmann D., *et al.* (2018). *Integrative genomic and transcriptomic analysis of leiomyosarcoma*. *Nat Commun* 9:144.

Cieślik M. & Chinnaiyan A. M. (2018). *Cancer transcriptome profiling at the juncture of clinical translation*. *Nat. Rev. Genet.* 19:93–109.

Ciriello G., Miller M. L., Aksoy B. A., Senbabaoglu Y., Schultz N. & Sander C. (2013). *Emerg-*

ing landscape of oncogenic signatures across human cancers. *Nat. Genet.* 45:1127–1133.

Clark J., Rocques P. J., Crew A. J., Gill S., *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*. *Nat. Genet.* 7:502–508.

Clarke C., Madden S. F., Doolan P., Aherne S. T., *et al.* (2013). *Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis*. *Carcinogenesis* 34:2300–2308.

Coindre J.-M. (2006). *Grading of soft tissue sarcomas: review and update*. *Arch. Pathol. Lab. Med.* 130:1448–1453.

Coindre J.-M., Trojani M., Contesso G., David M., *et al.* (1986). *Reproducibility of a histopathologic grading system for adult soft tissue sarcoma*. *Cancer* 58:306–309.

Coindre J.-M., Terrier P., Guillou L., Le Doussal V., *et al.* (2001). *Predictive value of grade for metastasis development in the main histologic types of adult soft tissue sarcomas: a study of 1240 patients from the French Federation of Cancer Centers Sarcoma Group*. *Cancer* 91:1914–1926.

Corbett A. H. (2018). *Post-transcriptional regulation of gene expression and human disease*. *Curr. Opin. Cell Biol.* 52:96–104.

Cordes K. R., Sheehy N. T., White M. P., Berry E. C., *et al.* (2009). *miR-145 and miR-143 regulate smooth muscle cell fate and plasticity*. *Nature* 460:705–710.

Costa V., Aprile M., Esposito R. & Ciccodicola A. (2013). *RNA-Seq and human complex diseases: recent accomplishments and future perspectives*. *Eur. J. Hum. Genet.* 21:134–142.

Croce S., Young R. H. & Oliva E. (2014). *Uterine leiomyomas with bizarre nuclei: a clinicopathologic study of 59 cases*. *Am. J. Surg. Pathol.* 38:1330–1339.

Croce S., Ducoulombier A., Ribeiro A., Lesluyes T., *et al.* (2018). *Genome profiling is an efficient tool to avoid the STUMP classification of uterine smooth muscle lesions: a comprehensive array-genomic hybridization analysis of 77 tumors*. *Mod. Pathol.* 31:816–828.

Croce S., Lesluyes T., Delespaul L., Bonhomme B., *et al.* (2019). *GREB1-CTNNB1 fusion transcript detected by RNA-sequencing in a uterine tumor resembling ovarian sex cord tumor (UTROSCT): A novel CTNNB1 rearrangement*. *Genes Chromosomes Cancer* 58:155–163.

Culhane A. C., Schwarzl T., Sultana R., Picard K. C., *et al.* (2010). *GeneSigDB–a curated database of gene expression signatures*. *Nucleic Acids Res.* 38:D716–725.

Cuzick J., Swanson G. P., Fisher G., Brothman A. R., *et al.* (2011). *Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study*. *Lancet Oncol.* 12:245–255.

Czernilofsky A. P., Levinson A. D., Varmus H. E., Bishop J. M., Tischer E. & Goodman H. M. (1980). *Nucleotide sequence of an avian sarcoma virus oncogene (src) and proposed amino acid sequence for gene product*. *Nature* 287:198–203.

Dadone-Montaudié B., Alberti L., Duc A., Delespaul L., *et al.* (2018). *Alternative PDGFD rearrangements in dermatofibrosarcomas protuberans without PDGFB fusions*. *Mod. Pathol.*

31:1683–1693.

Davicioni E., Finckenstein F. G., Shahbazian V., Buckley J. D., Triche T. J. & Anderson M. J. (2006). *Identification of a PAX-FKHR gene expression signature that defines molecular classes and determines the prognosis of alveolar rhabdomyosarcomas*. *Cancer Res.* 66:6936–6946.

Davicioni E., Anderson M. J., Finckenstein F. G., Lynch J. C., *et al.* (2009). *Molecular classification of rhabdomyosarcoma–genotypic and phenotypic determinants of diagnosis: a report from the Children's Oncology Group*. *Am. J. Pathol.* 174:550–564.

Davis R. J., D'Cruz C. M., Lovell M. A., Biegel J. A. & Barr F. G. (1994). *Fusion of PAX7 to FKHR by the variant t(1;13)(p36;q14) translocation in alveolar rhabdomyosarcoma*. *Cancer Res.* 54:2869–2872.

Davoli T., Xu A. W., Mengwasser K. E., Sack L. M., *et al.* (2013). *Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome*. *Cell* 155:948–962.

De S. & Michor F. (2011). *DNA secondary structures and epigenetic determinants of cancer genome evolution*. *Nat. Struct. Mol. Biol.* 18:950–955.

Deaton A. M. & Bird A. (2011). *CpG islands and the regulation of transcription*. *Genes Dev.* 25:1010–1022.

Debiec-Rychter M., Sciot R., Le Cesne A., Schlemmer M., *et al.* (2006). *KIT mutations and dose selection for imatinib in patients with advanced gastrointestinal stromal tumours*. *Eur. J. Cancer* 42:1093–1103.

Delattre O., Zucman J., Plougastel B., Desmaze C., *et al.* (1992). *Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours*. *Nature* 359:162–165.

Delespaul L. (2018). *Le génome remanié comme oncogène des sarcomes pléomorphes ? [PhD defense]*. Présenté à : Toulouse.

Delespaul L., Lesluyes T., Pérot G., Brulard C., *et al.* (2017). *Recurrent TRIO Fusion in Nontranslocation-Related Sarcomas*. *Clin. Cancer Res.* 23:857–867.

Demicco E. G., Maki R. G., Lev D. C. & Lazar A. J. (2012). *New therapeutic targets in soft tissue sarcoma*. *Adv Anat Pathol* 19:170–180.

Der S., Zhu C.-Q., Brower S. & Uihlein A. (2015). *Predicting Prognosis of Early-Stage Non-Small Cell Lung Cancer Using the GeneFx® Lung Signature*. *PLoS Curr* 7.

Derrien T., Johnson R., Bussotti G., Tanzer A., *et al.* (2012). *The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression*. *Genome Res.* 22:1775–1789.

Dewhurst S. M., McGranahan N., Burrell R. A., Rowan A. J., *et al.* (2014). *Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution*. *Cancer Discov* 4:175–185.

Deyrup A. T., Lee V. K., Hill C. E., Cheuk W., *et al.* (2006). *Epstein-Barr virus-associated*

smooth muscle tumors are distinctive mesenchymal tumors reflecting multiple infection events: a clinicopathologic and molecular analysis of 29 tumors from 19 patients. *Am. J. Surg. Pathol.* 30:75–82.

Di Y., Chen D., Yu W. & Yan L. (2019). *Bladder cancer stage-associated hub genes revealed by WGCNA co-expression network analysis*. *Hereditas* 156:7.

Dickson M. A., Schwartz G. K., Keohan M. L., D'Angelo S. P., *et al.* (2016). *Progression-Free Survival Among Patients With Well-Differentiated or Dedifferentiated Liposarcoma Treated With CDK4 Inhibitor Palbociclib: A Phase 2 Clinical Trial*. *JAMA Oncol* 2:937–940.

Dickson B. C., Hornick J. L., Fletcher C. D. M., Demicco E. G., *et al.* (2018). *Dermatofibrosarcoma protuberans with a novel COL6A3-PDGFD fusion gene and apparent predilection for breast*. *Genes Chromosomes Cancer*.

Ding D.-P., Chen Z.-L., Zhao X.-H., Wang J.-W., *et al.* (2011). *miR-29c induces cell cycle arrest in esophageal squamous cell carcinoma by modulating cyclin E expression*. *Carcinogenesis* 32:1025–1032.

Dixon J. R., Selvaraj S., Yue F., Kim A., *et al.* (2012). *Topological domains in mammalian genomes identified by analysis of chromatin interactions*. *Nature* 485:376–380.

Dodd R. D., Mito J. K. & Kirsch D. G. (2010). *Animal models of soft-tissue sarcoma*. *Dis Model Mech* 3:557–566.

Dong C. W., Wang Y. X., Du F. T., Ding W. & Hu S. Y. (2016). *Low miR-29c expression is a prognostic marker in hepatocellular carcinoma*. *Genet. Mol. Res.* 15.

Dowsett M., Sestak I., Lopez-Knowles E., Sidhu K., *et al.* (2013). *Comparison of PAM50 risk of recurrence score with oncotype DX and IHC4 for predicting risk of distant recurrence after endocrine therapy*. *J. Clin. Oncol.* 31:2783–2790.

Du P., Zhang X., Huang C.-C., Jafari N., *et al.* (2010). *Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis*. *BMC Bioinformatics* 11:587.

Ducimetière F., Lurkin A., Ranchère-Vince D., Decouvelaere A.-V., *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:e20294.

Dulbecco R. (1986). *A turning point in cancer research: sequencing the human genome*. *Science* 231:1055–1056.

Eden A., Gaudet F., Waghmare A. & Jaenisch R. (2003). *Chromosomal instability and tumors promoted by DNA hypomethylation*. *Science* 300:455.

Ehrlich M. (2002). *DNA methylation in cancer: too much, but also too little*. *Oncogene* 21:5400–5413.

Ehrlich M. (2009). *DNA hypomethylation in cancer cells*. *Epigenomics* 1:239–259.

ENCODE Project Consortium, Birney E., Stamatoyannopoulos J. A., Dutta A., *et al.* (2007). *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project*. *Nature* 447:799–816.

Erho N., Crisan A., Vergara I. A., Mitra A. P., *et al.* (2013). *Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy*. *PLoS ONE* 8:e66855.

Eriksson M., Hardell L. & Adami H. O. (1990). *Exposure to dioxins as a risk factor for soft tissue sarcoma: a population-based case-control study*. *J. Natl. Cancer Inst.* 82:486–490.

Esquela-Kerscher A. & Slack F. J. (2006). *Oncomirs - microRNAs with a role in cancer*. *Nat. Rev. Cancer* 6:259–269.

Evans D. G. R., Baser M. E., McGaughran J., Sharif S., Howard E. & Moran A. (2002). *Malignant peripheral nerve sheath tumours in neurofibromatosis 1*. *J. Med. Genet.* 39:311–314.

Exertier P., Javerzat S., Wang B., Franco M., *et al.* (2013). *Impaired angiogenesis and tumor development by inhibition of the mitotic kinesin Eg5*. *Oncotarget* 4:2302–2316.

Fabbri M., Garzon R., Cimmino A., Liu Z., *et al.* (2007). *MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B*. *Proc. Natl. Acad. Sci. U.S.A.* 104:15805–15810.

Fan C., Oh D. S., Wessels L., Weigelt B., *et al.* (2006). *Concordance among gene-expression-based predictors for breast cancer*. *N. Engl. J. Med.* 355:560–569.

Fang Y. & Fullwood M. J. (2016). *Roles, Functions, and Mechanisms of Long Non-coding RNAs in Cancer*. *Genomics Proteomics Bioinformatics* 14:42–54.

Farid M. & Ngeow J. (2016). *Sarcomas Associated With Genetic Cancer Predisposition Syndromes: A Review*. *Oncologist* 21:1002–1013.

Fedorowicz G., Guerrero S., Wu T. D. & Modrusan Z. (2009). *Microarray analysis of RNA extracted from formalin-fixed, paraffin-embedded and matched fresh-frozen ovarian adenocarcinomas*. *BMC Med Genomics* 2:23.

Feinberg A. P. & Vogelstein B. (1983). *Hypomethylation distinguishes genes of some human cancers from their normal counterparts*. *Nature* 301:89–92.

Feinberg A. P., Ohlsson R. & Henikoff S. (2006). *The epigenetic progenitor origin of human cancer*. *Nat. Rev. Genet.* 7:21–33.

Ferrari A., Gronchi A., Casanova M., Meazza C., *et al.* (2004). *Synovial sarcoma: a retrospective analysis of 271 patients of all ages treated at a single institution*. *Cancer* 101:627–634.

Ferrari A., De Salvo G. L., Brennan B., van Noesel M. M., *et al.* (2015). *Synovial sarcoma in children and adolescents: the European Pediatric Soft Tissue Sarcoma Study Group prospective trial (EpSSG NRSTS 2005)*. *Ann. Oncol.* 26:567–572.

Fletcher C. D. M., Bridge J. A., Hogendoorn P. C. & Mertens F. (2013). *WHO Classification of Tumours of Soft Tissue and Bone*. *4th edition*, *IARC Press*, *Lyon*.

Fligman I., Lonardo F., Jhanwar S. C., Gerald W. L., Woodruff J. & Ladanyi M. (1995). *Molecular diagnosis of synovial sarcoma and characterization of a variant SYT-SSX2 fusion transcript*. *Am. J. Pathol.* 147:1592–1599.

Füllgrabe J., Kavanagh E. & Joseph B. (2011). *Histone onco-modifications*. *Oncogene* 30:3391–3403.

Francis P., Namløs H. M., Müller C., Edén P., *et al.* (2007). *Diagnostic and prognostic gene expression signatures in 177 soft tissue sarcomas: hypoxia-induced transcription profile signifies metastatic potential*. *BMC Genomics* 8:73.

Frankish A., Diekhans M., Ferreira A.-M., Johnson R., *et al.* (2019). *GENCODE reference annotation for the human and mouse genomes*. *Nucleic Acids Res.* 47:D766–D773.

Franklin R. E. & Gosling R. G. (1953). *Molecular configuration in sodium thymonucleate*. *Nature* 171:740–741.

Friedman R. C., Farh K. K.-H., Burge C. B. & Bartel D. P. (2009). *Most mammalian mRNAs are conserved targets of microRNAs*. *Genome Res.* 19:92–105.

Friend S. H., Bernards R., Rogelj S., Weinberg R. A., *et al.* (1986). *A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma*. *Nature* 323:643–646.

Fritchie K. J., Jin L., Wang X., Graham R. P., *et al.* (2016). *Fusion gene profile of biphenotypic sinonasal sarcoma: an analysis of 44 cases*. *Histopathology* 69:930–936.

Fu J., Bian M., Jiang Q. & Zhang C. (2007). *Roles of Aurora kinases in mitosis and tumorigenesis*. *Mol. Cancer Res.* 5:1–10.

Fujiwara T., Bandi M., Nitta M., Ivanova E. V., Bronson R. T. & Pellman D. (2005). *Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells*. *Nature* 437:1043–1047.

Gao Q., Liang W.-W., Foltz S. M., Mutharasu G., *et al.* (2018). *Driver Fusions and Their Implications in the Development and Treatment of Human Cancers*. *Cell Rep* 23:227–238.e3.

García V., García J. M., Peña C., Silva J., *et al.* (2008). *Free circulating mRNA in plasma from breast cancer patients and clinical outcome*. *Cancer Lett.* 263:312–320.

Garcia V., Garcia J. M., Silva J., Martin P., *et al.* (2009). *Extracellular tumor-related mRNA in plasma of lymphoma patients and survival implications*. *PLoS ONE* 4:e8173.

Geiss G. K., Bumgarner R. E., Birditt B., Dahl T., *et al.* (2008). *Direct multiplexed measurement of gene expression with color-coded probe pairs*. *Nat. Biotechnol.* 26:317–325.

Gentles A. J., Newman A. M., Liu C. L., Bratman S. V., *et al.* (2015). *The prognostic landscape of genes and infiltrating immune cells across human cancers*. *Nat. Med.* 21:938–945.

Gerlinger M., McGranahan N., Dewhurst S. M., Burrell R. A., Tomlinson I. & Swanton C. (2014). *Cancer: evolution within a lifetime*. *Annu. Rev. Genet.* 48:215–236.

Gibault L., Pérot G., Chibon F., Bonnin S., *et al.* (2011). *New insights in sarcoma oncogenesis: a comprehensive analysis of a large series of 160 soft tissue sarcomas with complex genomics*. *J. Pathol.* 223:64–71.

Gibault L., Ferreira C., Pérot G., Audebourg A., *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*. *Mod. Pathol.* 25:197–211.

Golub T. R., Slonim D. K., Tamayo P., Huard C., *et al.* (1999). *Molecular classification of cancer: class discovery and class prediction by gene expression monitoring*. *Science* 286:531–

537.

Gong J., Chehrazi-Raffle A., Reddi S. & Salgia R. (2018). *Development of PD-1 and PD-L1 inhibitors as a form of cancer immunotherapy: a comprehensive review of registration trials and future considerations*. *J Immunother Cancer* 6:8.

Gonzalez K. D., Noltner K. A., Buzin C. H., Gu D., *et al.* (2009). *Beyond Li Fraumeni Syndrome: clinical characteristics of families with p53 germline mutations*. *J. Clin. Oncol.* 27:1250–1256.

Gordon D. J., Resio B. & Pellman D. (2012). *Causes and consequences of aneuploidy in cancer*. *Nat. Rev. Genet.* 13:189–203.

Gounder M. M., Ali S. M., Robinson V., Bailey M., *et al.* (2017). *Impact of next-generation sequencing (NGS) on diagnostic and therapeutic options in soft-tissue and bone sarcoma.* Présenté à : ASCO 2017, Chicago.

Groffen J., Stephenson J. R., Heisterkamp N., de Klein A., Bartram C. R. & Grosveld G. (1984). *Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22*. *Cell* 36:93–99.

Gronchi A., Ferrari S., Quagliuolo V., Broto J. M., *et al.* (2017). *Histotype-tailored neoadjuvant chemotherapy versus standard chemotherapy in patients with high-risk soft-tissue sarcomas (ISG-STS 1001): an international, open-label, randomised, controlled, phase 3, multicentre trial*. *Lancet Oncol.* 18:812–822.

Grosso A. R., Leite A. P., Carvalho S., Matos M. R., *et al.* (2015). *Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma*. *Elife* 4.

Guillou L. & Aurias A. (2010). *Soft tissue sarcomas with complex genomic profiles*. *Virchows Arch.* 456:201–217.

Gurbuz A. K., Giardiello F. M., Petersen G. M., Krush A. J., *et al.* (1994). *Desmoid tumours in familial adenomatous polyposis*. *Gut* 35:377–381.

Hafner M., Landgraf P., Ludwig J., Rice A., *et al.* (2008). *Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing*. *Methods* 44:3–12.

Hafner M., Landthaler M., Burger L., Khorshid M., *et al.* (2010). *Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP*. *Cell* 141:129–141.

Hajjari M. & Salavaty A. (2015). *HOTAIR: an oncogenic long non-coding RNA in different cancers*. *Cancer Biol. Med.* 12:1–9.

Haldar M., Hancock J. D., Coffin C. M., Lessnick S. L. & Capecchi M. R. (2007). *A conditional mouse model of synovial sarcoma: insights into a myogenic origin*. *Cancer Cell* 11:375–388.

Hanahan D. & Weinberg R. A. (2000). *The hallmarks of cancer*. *Cell* 100:57–70.

Hanahan D. & Weinberg R. A. (2011). *Hallmarks of cancer: the next generation*. *Cell* 144:646–674.

Heinrich M. C., Corless C. L., Demetri G. D., Blanke C. D., *et al.* (2003). *Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor*. *J. Clin. Oncol.*

21:4342–4349.

Heinrich M. C., Maki R. G., Corless C. L., Antonescu C. R., *et al.* (2008a). *Primary and secondary kinase genotypes correlate with the biological and clinical activity of sunitinib in imatinib-resistant gastrointestinal stromal tumor*. *J. Clin. Oncol.* 26:5352–5359.

Heinrich M. C., Owzar K., Corless C. L., Hollis D., *et al.* (2008b). *Correlation of kinase genotype and clinical outcome in the North American Intergroup Phase III Trial of imatinib mesylate for treatment of advanced gastrointestinal stromal tumor: CALGB 150105 Study by Cancer and Leukemia Group B and Southwest Oncology Group*. *J. Clin. Oncol.* 26:5360–5367.

Henson J. D., Hannay J. A., McCarthy S. W., Royds J. A., *et al.* (2005). *A robust assay for alternative lengthening of telomeres in tumors shows the significance of alternative lengthening of telomeres in sarcomas and astrocytomas*. *Clin. Cancer Res.* 11:217–225.

Hirokawa N., Noda Y., Tanaka Y. & Niwa S. (2009). *Kinesin superfamily motor proteins and intracellular transport*. *Nat. Rev. Mol. Cell Biol.* 10:682–696.

Hirota S., Isozaki K., Moriyama Y., Hashimoto K., *et al.* (1998). *Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors*. *Science* 279:577–580.

Hofvander J., Tayebwa J., Nilsson J., Magnusson L., *et al.* (2015). *Recurrent PRDM10 gene fusions in undifferentiated pleomorphic sarcoma*. *Clin. Cancer Res.* 21:864–869.

Honoré C., Méeus P., Stoeckle E. & Bonvalot S. (2015). *Soft tissue sarcoma in France in 2015: Epidemiology, classification and organization of clinical care*. *J Visc Surg* 152:223–230.

Hu X., Wang Q., Tang M., Barthel F., *et al.* (2018). *TumorFusions: an integrative resource for cancer-associated transcript fusions*. *Nucleic Acids Res.* 46:D1144–D1149.

International Human Genome Sequencing Consortium (2001). *Initial sequencing and analysis of the human genome*. *Nature* 409:860–921.

International Human Genome Sequencing Consortium (2004). *Finishing the euchromatic sequence of the human genome*. *Nature* 431:931–945.

Italiano A., Lagarde P., Brulard C., Terrier P., *et al.* (2013). *Genetic profiling identifies two classes of soft-tissue leiomyosarcomas with distinct clinical characteristics*. *Clin. Cancer Res.* 19:1190–1196.

Iyer M. K., Chinnaiyan A. M. & Maher C. A. (2011). *ChimeraScan: a tool for identifying chimeric transcription in sequencing data*. *Bioinformatics* 27:2903–2904.

Jackson D. A., Hassan A. B., Errington R. J. & Cook P. R. (1993). *Visualization of focal sites of transcription within human nuclei*. *EMBO J.* 12:1059–1065.

Jaffe A. B. & Hall A. (2002). *Rho GTPases in transformation and metastasis*. *Adv. Cancer Res.* 84:57–80.

Jaffe A. E., Murakami P., Lee H., Leek J. T., *et al.* (2012). *Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies*. *Int J Epidemiol* 41:200–209.

Jemaà M., Abdallah S., Lledo G., Perot G., *et al.* (2017). *Heterogeneity in sarcoma cell lines reveals enhanced motility of tetraploid versus diploid cells*. *Oncotarget* 8.

Jiang L., Schlesinger F., Davis C. A., Zhang Y., *et al.* (2011). *Synthetic spike-in standards for RNA-seq experiments*. *Genome Res.* 21:1543–1551.

Jividen K. & Li H. (2014). *Chimeric RNAs generated by intergenic splicing in normal and cancer cells*. *Genes Chromosomes Cancer* 53:963–971.

Kadoch C., Williams R. T., Calarco J. P., Miller E. L., *et al.* (2017). *Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states*. *Nat. Genet.* 49:213–222.

Kanai Y. (2010). *Genome-wide DNA methylation profiles in precancerous conditions and cancers*. *Cancer Sci.* 101:36–45.

Kanezaki R., Toki T., Xu G., Narayanan R. & Ito E. (2006). *Cloning and characterization of the novel chimeric gene p53/FXR2 in the acute megakaryoblastic leukemia cell line CMK11-5*. *Tohoku J. Exp. Med.* 209:169–180.

Kawaguchi K.-i., Oda Y., Saito T., Yamamoto H., *et al.* (2003). *Mechanisms of inactivation of the p16INK4a gene in leiomyosarcoma of soft tissue: decreased p16 expression correlates with promoter methylation and poor prognosis*. *J. Pathol.* 201:487–495.

Kawaguchi K.-i., Oda Y., Saito T., Takahira T., *et al.* (2005). *Genetic and epigenetic alterations of the PTEN gene in soft tissue sarcomas*. *Hum. Pathol.* 36:357–363.

Kawaguchi K.-i., Oda Y., Saito T., Yamamoto H., *et al.* (2006). *DNA hypermethylation status of multiple genes in soft tissue sarcomas*. *Mod. Pathol.* 19:106–114.

Kelly C. M., Bernard P. S., Krishnamurthy S., Wang B., *et al.* (2012). *Agreement in risk prediction between the 21-gene recurrence score assay (Oncotype DX®) and the PAM50 breast cancer intrinsic Classifier™ in early-stage estrogen receptor-positive breast cancer*. *Oncologist* 17:492–498.

Kennedy R. D., Bylesjo M., Kerr P., Davison T., *et al.* (2011). *Development and independent validation of a prognostic assay for stage II colon cancer using formalin-fixed paraffin-embedded tissue*. *J. Clin. Oncol.* 29:4620–4626.

Kertesz M., Iovino N., Unnerstall U., Gaul U. & Segal E. (2007). *The role of site accessibility in microRNA target recognition*. *Nat. Genet.* 39:1278–1284.

Khorasanizadeh S. (2004). *The nucleosome: from genomic organization to genomic regulation*. *Cell* 116:259–272.

Kim D. & Salzberg S. L. (2011). *TopHat-Fusion: an algorithm for discovery of novel fusion transcripts*. *Genome Biol.* 12:R72.

Kim J. I., Suh J. T., Choi K. U., Kang H. J., *et al.* (2009). *Inactivation of O6-methylguanine-DNA methyltransferase in soft tissue sarcomas: association with K-ras mutations*. *Hum. Pathol.* 40:934–941.

Kishore S., Jaskiewicz L., Burger L., Hausser J., Khorshid M. & Zavolan M. (2011). *A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins*. *Nat. Methods* 8:559–564.

Kleinerman R. A., Schonfeld S. J. & Tucker M. A. (2012). *Sarcomas in hereditary retinoblas-*

toma. *Clin. Sarcoma Res.* 2:15.

Knezevich S. R., McFadden D. E., Tao W., Lim J. F. & Sorensen P. H. (1998). *A novel ETV6-NTRK3 gene fusion in congenital fibrosarcoma*. *Nat. Genet.* 18:184–187.

Koontz J. I., Soreng A. L., Nucci M., Kuo F. C., *et al.* (2001). *Frequent fusion of the JAZF1 and JJAZ1 genes in endometrial stromal tumors*. *Proc. Natl. Acad. Sci. U.S.A.* 98:6348–6353.

Koulaxouzidis G., Schwarzkopf E., Bannasch H. & Stark G. B. (2015). *Is revisional surgery mandatory when an unexpected sarcoma diagnosis is made following primary surgery? World J Surg Oncol* 13:306.

Kozomara A., Birgaoanu M. & Griffiths-Jones S. (2019). *miRBase: from microRNA sequences to function*. *Nucleic Acids Res.* 47:D155–D162.

Krüger J. & Rehmsmeier M. (2006). *RNAhybrid: microRNA target prediction easy, fast and flexible*. *Nucleic Acids Res.* 34:W451–454.

Kulis M. & Esteller M. (2010). *DNA methylation and cancer*. *Adv. Genet.* 70:27–56.

Kumar-Sinha C., Kalyana-Sundaram S. & Chinnaiyan A. M. (2012). *SLC45A3-ELK4 chimera in prostate cancer: spotlight on cis-splicing*. *Cancer Discov* 2:582–585.

Kurman R. J., for Research on Cancer I. A. & Organization W. H., eds (2014). *WHO classification of tumours of female reproductive organs*. World Health Organization classification of tumours, *4th edition*, *IARC Press*, *Lyon*.

Kwon J. G., Hwang S. J., Hennig G. W., Bayguinov Y., *et al.* (2009). *Changes in the structure and function of ICC networks in ICC hyperplasia and gastrointestinal stromal tumors*. *Gastroenterology* 136:630–639.

Ladanyi M., Lui M. Y., Antonescu C. R., Krause-Boehm A., *et al.* (2001). *The der(17)t(X;17)(p11;q25) of human alveolar soft part sarcoma fuses the TFE3 transcription factor gene to ASPL, a novel gene at 17q25*. *Oncogene* 20:48–57.

Lagarde P., Pérot G., Kauffmann A., Brulard C., *et al.* (2012). *Mitotic checkpoints and chromosome instability are strong predictors of clinical outcome in gastrointestinal stromal tumors*. *Clin. Cancer Res.* 18:826–838.

Lagarde P., Przybyl J., Brulard C., Pérot G., *et al.* (2013). *Chromosome instability accounts for reverse metastatic outcomes of pediatric and adult synovial sarcomas*. *J. Clin. Oncol.* 31:608–615.

Lagrange J. L., Ramaioli A., Chateau M. C., Marchal C., *et al.* (2000). *Sarcoma after radiation therapy: retrospective multiinstitutional study of 80 histologically confirmed cases. Radiation Therapist and Pathologist Groups of the Fédération Nationale des Centres de Lutte Contre le Cancer*. *Radiology* 216:197–205.

Lambert M., Jambon S., Depauw S. & David-Cordonnier M.-H. (2018). *Targeting Transcription Factors for Cancer Treatment*. *Molecules* 23.

Langfelder P. & Horvath S. (2008). *WGCNA: an R package for weighted correlation network analysis*. *BMC Bioinformatics* 9:559.

Laporte A. N., Barrott J. J., Yao R. J., Poulin N. M., *et al.* (2017). *HDAC and Proteasome*

Inhibitors Synergize to Activate Pro-Apoptotic Factors in Synovial Sarcoma. *PLoS ONE* 12:e0169407.

Lasota J. & Miettinen M. (2008). *Clinical significance of oncogenic KIT and PDGFRA mutations in gastrointestinal stromal tumours*. *Histopathology* 53:245–266.

Le Guellec S., Lesluyes T., Sarot E., Valle C., *et al.* (2018). *Validation of the Complexity INdex in SARComas prognostic signature on formalin-fixed, paraffin-embedded, soft-tissue sarcomas*. *Ann. Oncol.* 29:1828–1835.

Le Loarer F., Watson S., Pierron G., de Montpreville V. T., *et al.* (2015). *SMARCA4 inactivation defines a group of undifferentiated thoracic malignancies transcriptionally related to BAF-deficient sarcomas*. *Nat. Genet.* 47:1200–1205.

Le Loarer F., Laffont S., Lesluyes T., Tirode F., *et al.* (2019). *Clinicopathologic and Molecular Features of a Series of 41 Biphenotypic Sinonasal Sarcomas Expanding Their Molecular Spectrum*. *Am. J. Surg. Pathol.*

Lee Y.-F., John M., Falconer A., Edwards S., *et al.* (2004). *A gene expression signature associated with metastatic outcome in human leiomyosarcomas*. *Cancer Res.* 64:7201–7204.

Lee I., Ajay S. S., Yook J. I., Kim H. S., *et al.* (2009). *New class of microRNA targets containing simultaneous 5'-UTR and 3'-UTR interaction sites*. *Genome Res.* 19:1175–1183.

Lee J.-C., Jeng Y.-M., Liau J.-Y., Tsai J.-H., Hsu H.-H. & Yang C.-Y. (2015). *Alternative lengthening of telomeres and loss of ATRX are frequent events in pleomorphic and dedifferentiated liposarcomas*. *Mod. Pathol.* 28:1064–1073.

Lei Q., Li C., Zuo Z., Huang C., Cheng H. & Zhou R. (2016). *Evolutionary Insights into RNA trans-Splicing in Vertebrates*. *Genome Biol Evol* 8:562–577.

Lesluyes T., Pérot G., Largeau M. R., Brulard C., *et al.* (2016). *RNA sequencing validation of the Complexity INdex in SARComas prognostic signature*. *Eur J Cancer* 57:104–111.

Lesluyes T., Delespaul L., Coindre J.-M. & Chibon F. (2017). *The CINSARC signature as a prognostic marker for clinical outcome in multiple neoplasms*. *Sci Rep* 7.

Levy S., Sutton G., Ng P. C., Feuk L., *et al.* (2007). *The diploid genome sequence of an individual human*. *PLoS Biol.* 5:e254.

Lewis J. T., Oliveira A. M., Nascimento A. G., Schembri-Wismayer D., *et al.* (2012). *Low-grade sinonasal sarcoma with neural and myogenic features: a clinicopathologic analysis of 28 cases*. *Am. J. Surg. Pathol.* 36:517–525.

Li H., Wang J., Mor G. & Sklar J. (2008). *A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells*. *Science* 321:1357–1361.

Liao G.-B., Li X.-Z., Zeng S., Liu C., *et al.* (2018). *Regulation of the master regulator FOXM1 in cancer*. *Cell Commun. Signal* 16:57.

Liau J.-Y., Tsai J.-H., Jeng Y.-M., Lee J.-C., Hsu H.-H. & Yang C.-Y. (2015). *Leiomyosarcoma with alternative lengthening of telomeres is associated with aggressive histologic features, loss of ATRX expression, and poor clinical outcome*. *Am. J. Surg. Pathol.* 39:236–244.

Liberzon A., Subramanian A., Pinchback R., Thorvaldsdóttir H., Tamayo P. & Mesirov J. P.

(2011). *Molecular signatures database (MSigDB) 3.0*. *Bioinformatics* 27:1739–1740.

Liberzon A., Birger C., Thorvaldsdóttir H., Ghandi M., Mesirov J. P. & Tamayo P. (2015). *The Molecular Signatures Database (MSigDB) hallmark gene set collection*. *Cell Syst.* 1:417–425.

Lieberman-Aiden E., van Berkum N. L., Williams L., Imakaev M., *et al.* (2009). *Comprehensive mapping of long-range interactions reveals folding principles of the human genome*. *Science* 326:289–293.

Lin P. P., Pandey M. K., Jin F., Raymond A. K., Akiyama H. & Lozano G. (2009). *Targeted mutation of p53 and Rb in mesenchymal cells of the limb bud produces sarcomas in mice*. *Carcinogenesis* 30:1789–1795.

Love M. I., Huber W. & Anders S. (2014). *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*. *Genome Biol.* 15:550.

Ly P. & Cleveland D. W. (2017). *Rebuilding Chromosomes After Catastrophe: Emerging Mechanisms of Chromothripsis*. *Trends Cell Biol.* 27:917–930.

Ma X.-J., Wang Z., Ryan P. D., Isakoff S. J., *et al.* (2004). *A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen*. *Cancer Cell* 5:607–616.

Maher C. A., Palanisamy N., Brenner J. C., Cao X., *et al.* (2009). *Chimeric transcript discovery by paired-end transcriptome sequencing*. *Proc. Natl. Acad. Sci. U.S.A.* 106:12353–12358.

Mairal A., Terrier P., Chibon F., Sastre X., Lecesne A. & Aurias A. (1999). *Loss of chromosome 13 is the most frequent genomic imbalance in malignant fibrous histiocytomas. A comparative genomic hybridization analysis of a series of 30 cases*. *Cancer Genet. Cytogenet.* 111:134–138.

Mairal A., Chibon F., Rousselet A., Couturier J., Terrier P. & Aurias A. (2000). *Establishment of a human malignant fibrous histiocytoma cell line, COMA. Characterization By conventional cytogenetics, comparative genomic hybridization, and multiplex fluorescence In situ hybridization*. *Cancer Genet. Cytogenet.* 121:117–123.

March-Villalba J. A., Martínez-Jabaloyas J. M., Herrero M. J., Santamaria J., Aliño S. F. & Dasí F. (2012). *Cell-free circulating plasma hTERT mRNA is a useful marker for prostate cancer diagnosis and is associated with poor prognosis tumor characteristics*. *PLoS ONE* 7:e43470.

Mardis E. R. (2006). *Anticipating the 1,000 dollar genome*. *Genome Biol.* 7:112.

Mardis E. R. (2010). *The $1,000 genome, the $100,000 analysis? Genome Med* 2:84.

Margulies M., Egholm M., Altman W. E., Attiya S., *et al.* (2005). *Genome sequencing in microfabricated high-density picolitre reactors*. *Nature* 437:376–380.

Mariani O., Brennetot C., Coindre J.-M., Gruel N., *et al.* (2007). *JUN oncogene amplification and overexpression block adipocytic differentiation in highly aggressive sarcomas*. *Cancer Cell* 11:361–374.

Marzec P., Armenise C., Pérot G., Roumelioti F.-M., *et al.* (2015). *Nuclear-receptor-mediated telomere insertion leads to genome instability in ALT cancers*. *Cell* 160:913–927.

Mathoulin-Pélissier S., Chevreau C., Bellera C., Bauvin E., *et al.* (2014). *Adherence to*

consensus-based diagnosis and treatment guidelines in adult soft-tissue sarcoma patients: a French prospective population-based study. *Ann. Oncol.* 25:225–231.

Matsuo M., Nakada C., Tsukamoto Y., Noguchi T., *et al.* (2013). *MiR-29c is downregulated in gastric carcinomas and regulates cell proliferation by targeting RCC2*. *Mol. Cancer* 12:15.

McBride M. J., Pulice J. L., Beird H. C., Ingram D. R., *et al.* (2018). *The SS18-SSX Fusion Oncoprotein Hijacks BAF Complex Targeting and Function to Drive Synovial Sarcoma*. *Cancer Cell* 33:1128–1141.e7.

McGranahan N. & Swanton C. (2017). *Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future*. *Cell* 168:613–628.

McPherson A., Hormozdiari F., Zayed A., Giuliany R., *et al.* (2011). *deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data*. *PLoS Comput. Biol.* 7:e1001138.

Memczak S., Jens M., Elefsinioti A., Torti F., *et al.* (2013). *Circular RNAs are a large class of animal RNAs with regulatory potency*. *Nature* 495:333–338.

Mertens F., Johansson B., Fioretos T. & Mitelman F. (2015). *The emerging complexity of gene fusions in cancer*. *Nat. Rev. Cancer* 15:371–381.

Mertens F., Antonescu C. R. & Mitelman F. (2016). *Gene fusions in soft tissue tumors: Recurrent and overlapping pathogenetic themes*. *Genes Chromosomes Cancer* 55:291–310.

Michaud J., Simpson K. M., Escher R., Buchet-Poyau K., *et al.* (2008). *Integrative analysis of RUNX1 downstream pathways and target genes*. *BMC Genomics* 9:363.

Michiels S., Ternès N. & Rotolo F. (2016). *Statistical controversies in clinical research: prognostic gene signatures are not (yet) useful in clinical practice*. *Ann. Oncol.* 27:2160–2167.

Miettinen M., Makhlouf H., Sobin L. H. & Lasota J. (2006). *Gastrointestinal stromal tumors of the jejunum and ileum: a clinicopathologic, immunohistochemical, and molecular genetic study of 906 cases before imatinib with long-term follow-up*. *Am. J. Surg. Pathol.* 30:477–489.

Milhem M. M., Knutson T., Yang S., Zhu D., *et al.* (2011). *Correlation of MTDH/AEG-1 and HOTAIR Expression with Metastasis and Response to Treatment in Sarcoma Patients*. *J Cancer Sci Ther* S5.

Mir O., Honoré C. & Adam J. (2017). *PD-1 inhibition in bone sarcoma and soft-tissue sarcoma*. *Lancet Oncol.* 18:1430–1431.

Mitelman F., Johansson B. & Mertens F. (2007). *The impact of translocations and gene fusions on cancer causation*. *Nat. Rev. Cancer* 7:233–245.

Mito J. K., Qian X., Doyle L. A., Hornick J. L. & Jo V. Y. (2017). *Role of Histone H3K27 Trimethylation Loss as a Marker for Malignant Peripheral Nerve Sheath Tumor in Fine-Needle Aspiration and Small Biopsy Specimens*. *Am. J. Clin. Pathol.* 148:179–189.

Miyaki M., Konishi M., Kikuchi-Yanoshita R., Enomoto M., *et al.* (1993). *Coexistence of somatic and germ-line mutations of APC gene in desmoid tumors from patients with familial adenomatous polyposis*. *Cancer Res.* 53:5079–5082.

Miyata T., Sonoda K., Tomikawa J., Tayama C., *et al.* (2015). *Genomic, Epigenomic, and Transcriptomic Profiling towards Identifying Omics Features and Specific Biomarkers*

That Distinguish Uterine Leiomyosarcoma and Leiomyoma at Molecular Levels. *Sarcoma* 2015:412068.

Moore L. D., Le T. & Fan G. (2013). *DNA methylation and its basic function*. *Neuropsychopharmacology* 38:23–38.

Morris T. J., Butcher L. M., Feber A., Teschendorff A. E., *et al.* (2014). *ChAMP: 450k Chip Analysis Methylation Pipeline*. *Bioinformatics* 30:428–430.

Mortazavi A., Williams B. A., McCue K., Schaeffer L. & Wold B. (2008). *Mapping and quantifying mammalian transcriptomes by RNA-Seq*. *Nat. Methods* 5:621–628.

Mukherjee M., Ge G., Zhang N., Edwards D. G., *et al.* (2014). *MMTV-Espl1 transgenic mice develop aneuploid, estrogen receptor alpha (ERα)-positive mammary adenocarcinomas*. *Oncogene* 33:5511–5522.

Nair J. S. & Schwartz G. K. (2016). *MLN-8237: A dual inhibitor of aurora A and B in soft tissue sarcomas*. *Oncotarget* 7:12893–12903.

Nakano K. & Takahashi S. (2018a). *Current Molecular Targeted Therapies for Bone and Soft Tissue Sarcomas*. *Int J Mol Sci* 19.

Nakano K. & Takahashi S. (2018b). *Translocation-Related Sarcomas*. *Int J Mol Sci* 19.

Nakayama R., Nemoto T., Takahashi H., Ohta T., *et al.* (2007). *Gene expression analysis of soft tissue sarcomas: characterization and reclassification of malignant fibrous histiocytoma*. *Mod. Pathol.* 20:749–759.

Nam H.-J. & van Deursen J. M. (2014). *Cyclin B2 and p53 control proper timing of centrosome separation*. *Nat. Cell Biol.* 16:538–549.

Neuville A., Chibon F. & Coindre J.-M. (2014). *Grading of soft tissue sarcomas: from histological to molecular assessment*. *Pathology* 46:113–120.

Nielsen T. O., West R. B., Linn S. C., Alter O., *et al.* (2002). *Molecular characterisation of soft tissue tumours: a gene expression study*. *Lancet* 359:1301–1307.

Nielsen T., Wallden B., Schaper C., Ferree S., *et al.* (2014). *Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens*. *BMC Cancer* 14:177.

Nowell P. C. & Hungerford D. A. (1960). *Chromosome studies on normal and leukemic human leukocytes*. *J. Natl. Cancer Inst.* 25:85–109.

Nyce J. (1989). *Drug-induced DNA hypermethylation and drug resistance in human tumors*. *Cancer Res.* 49:5829–5836.

Ogura K., Hosoda F., Arai Y., Nakamura H., *et al.* (2018). *Integrated genetic and epigenetic analysis of myxofibrosarcoma*. *Nat Commun* 9:2765.

Oppermann H., Levinson A. D., Varmus H. E., Levintow L. & Bishop J. M. (1979). *Uninfected vertebrate cells contain a protein that is closely related to the product of the avian sarcoma virus transforming gene (src)*. *Proc. Natl. Acad. Sci. U.S.A.* 76:1804–1808.

Paik S., Shak S., Tang G., Kim C., *et al.* (2004). *A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer*. *N. Engl. J. Med.* 351:2817–2826.

Park S.-Y., Lee J. H., Ha M., Nam J.-W. & Kim V. N. (2009). *miR-29 miRNAs activate p53 by targeting p85 alpha and CDC42*. *Nat. Struct. Mol. Biol.* 16:23–29.

Parker J. S., Mullins M., Cheang M. C. U., Leung S., *et al.* (2009). *Supervised risk predictor of breast cancer based on intrinsic subtypes*. *J. Clin. Oncol.* 27:1160–1167.

Parkinson D. R., McCormack R. T., Keating S. M., Gutman S. I., *et al.* (2014). *Evidence of clinical utility: an unmet need in molecular diagnostics for patients with cancer*. *Clin. Cancer Res.* 20:1428–1444.

Parra G., Reymond A., Dabbouseh N., Dermitzakis E. T., *et al.* (2006). *Tandem chimerism as a means to increase protein complexity in the human genome*. *Genome Res.* 16:37–44.

Pass H. I., Goparaju C., Ivanov S., Donington J., *et al.* (2010). *hsa-miR-29c* is linked to the prognosis of malignant pleural mesothelioma*. *Cancer Res.* 70:1916–1924.

Pedeutour F., Simon M. P., Minoletti F., Barcelo G., *et al.* (1996). *Translocation, t(17;22)(q22;q13), in dermatofibrosarcoma protuberans: a new tumor-associated chromosome rearrangement*. *Cytogenet. Cell Genet.* 72:171–174.

Pekmezci M., Cuevas-Ocampo A. K., Perry A. & Horvai A. E. (2017). *Significance of H3K27me3 loss in the diagnosis of malignant peripheral nerve sheath tumors*. *Mod. Pathol.* 30:1710–1719.

Pennisi E. (1999). *Academic sequencers challenge Celera in a sprint to the finish*. *Science* 283:1822–1823.

Perrier L., Rascle P., Morelle M., Toulmonde M., *et al.* (2018). *The cost-saving effect of centralized histological reviews with soft tissue and visceral sarcomas, GIST, and desmoid tumors: The experiences of the pathologists of the French Sarcoma Group*. *PLoS ONE* 13:e0193330.

Peters J.-M. (2006). *The anaphase promoting complex/cyclosome: a machine designed to destroy*. *Nat. Rev. Mol. Cell Biol.* 7:644–656.

Peters T. J., Buckley M. J., Statham A. L., Pidsley R., *et al.* (2015). *De novo identification of differentially methylated regions in the human genome*. *Epigenetics Chromatin* 8:6.

Pfeifer G. P. (2006). *Mutagenesis at methylated CpG sequences*. *Curr. Top. Microbiol. Immunol.* 301:259–281.

Pierron G., Tirode F., Lucchesi C., Reynaud S., *et al.* (2012). *A new subtype of bone sarcoma defined by BCOR-CCNB3 gene fusion*. *Nat. Genet.* 44:461–466.

Podlaha O., Riester M., De S. & Michor F. (2012). *Evolution of the cancer genome*. *Trends Genet.* 28:155–163.

Potapova T. A., Zhu J. & Li R. (2013). *Aneuploidy and chromosomal instability: a vicious cycle driving cellular evolution and cancer genome chaos*. *Cancer Metastasis Rev.* 32:377–389.

Pratt A. J. & MacRae I. J. (2009). *The RNA-induced silencing complex: a versatile gene-silencing machine*. *J. Biol. Chem.* 284:17897–17901.

Presneau N., Eskandarpour M., Shemais T., Henderson S., *et al.* (2013). *MicroRNA profiling of peripheral nerve sheath tumours identifies miR-29c as a tumour suppressor gene involved in tumour progression*. *Br. J. Cancer* 108:964–972.

Pérot G., Derré J., Coindre J.-M., Tirode F., *et al.* (2009). *Strong smooth muscle differentiation is dependent on myocardin gene amplification in most human retroperitoneal leiomyosarcomas*. *Cancer Res.* 69:2269–2278.

Pérot G., Chibon F., Montero A., Lagarde P., *et al.* (2010). *Constant p53 pathway inactivation in a large series of soft tissue sarcomas with complex genetics*. *Am. J. Pathol.* 177:2080–2090.

Pérot G., Mendiboure J., Brouste V., Velasco V., *et al.* (2014). *Smooth muscle differentiation identifies two classes of poorly differentiated pleomorphic sarcomas with distinct outcome*. *Mod. Pathol.* 27:840–850.

Rabbitts T. H., Forster A., Larson R. & Nathan P. (1993). *Fusion of the dominant negative transcription regulator CHOP with a novel gene FUS by translocation t(12;16) in malignant liposarcoma*. *Nat. Genet.* 4:175–180.

Rajagopalan H. & Lengauer C. (2004). *Aneuploidy and cancer*. *Nature* 432:338–341.

Ray-Coquard I., Montesco M. C., Coindre J.-M., Dei Tos A. P., *et al.* (2012). *Sarcoma: concordance between initial diagnosis and centralized expert review in a population-based study within three European regions*. *Ann. Oncol.* 23:2442–2449.

Reddy E. P., Reynolds R. K., Santos E. & Barbacid M. (1982). *A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene*. *Nature* 300:149–152.

Reuter J. A., Spacek D. V. & Snyder M. P. (2015). *High-throughput sequencing technologies*. *Mol. Cell* 58:586–597.

Rhomberg W. (2006). *The radiation response of sarcomas by histologic subtypes: a review with special emphasis given to results achieved with razoxane*. *Sarcoma* 2006:87367.

Ricke R. M., Jeganathan K. B. & van Deursen J. M. (2011). *Bub1 overexpression induces aneuploidy and tumor formation through Aurora B kinase hyperactivation*. *J. Cell Biol.* 193:1049–1064.

Robinson D. R., Wu Y.-M., Kalyana-Sundaram S., Cao X., *et al.* (2013). *Identification of recurrent NAB2-STAT6 gene fusions in solitary fibrous tumor by integrative sequencing*. *Nat. Genet.* 45:180–185.

Robitaille J. M., Gillett R. M., LeBlanc M. A., Gaston D., *et al.* (2014). *Phenotypic overlap between familial exudative vitreoretinopathy and microcephaly, lymphedema, and chorioretinal dysplasia caused by KIF11 mutations*. *JAMA Ophthalmol* 132:1393–1399.

Rode A., Maass K. K., Willmund K. V., Lichter P. & Ernst A. (2016). *Chromothripsis in cancer cells: An update*. *Int. J. Cancer* 138:2322–2333.

Ørom U. A., Derrien T., Beringer M., Gumireddy K., *et al.* (2010). *Long noncoding RNAs with enhancer-like function in human cells*. *Cell* 143:46–58.

Rosenwald A., Wright G., Wiestner A., Chan W. C., *et al.* (2003). *The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma*. *Cancer Cell* 3:185–197.

Rota R., Ciarapica R., Giordano A., Miele L. & Locatelli F. (2011). *MicroRNAs in rhab-*

domyosarcoma: pathogenetic implications and translational potentiality. *Mol. Cancer* 10:120.

Rous P. (1910). *A transmissible avian neoplasm (sarcoma of the common fowl). J. Exp. Med.* 12:696–705.

Rous P. (1911). *A sarcoma of the fowl transmissible by an agent separable from the tumor cells*. *J. Exp. Med.* 13:397–411.

Rowley J. D. (1973). *Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining*. *Nature* 243:290–293.

Roychowdhury S. & Chinnaiyan A. M. (2016). *Translating cancer genomes and transcriptomes for precision oncology*. *CA Cancer J. Clin.* 66:75–88.

Rutledge S. D., Douglas T. A., Nicholson J. M., Vila-Casadesús M., *et al.* (2016). *Selective advantage of trisomic human cells cultured in non-standard conditions*. *Sci Rep* 6:22828.

Saiag P., Grob J.-J., Lebbe C., Malvehy J., *et al.* (2015). *Diagnosis and treatment of dermatofibrosarcoma protuberans. European consensus-based interdisciplinary guideline*. *Eur. J. Cancer* 51:2604–2608.

Salazar R., Roepman P., Capella G., Moreno V., *et al.* (2011). *Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer*. *J. Clin. Oncol.* 29:17–24.

Sanchez-Vega F., Mina M., Armenia J., Chatila W. K., *et al.* (2018). *Oncogenic Signaling Pathways in The Cancer Genome Atlas*. *Cell* 173:321–337.e10.

Sansregret L. & Swanton C. (2017). *The Role of Aneuploidy in Cancer Evolution*. *Cold Spring Harb Perspect Med* 7.

Sansregret L., Vanhaesebroeck B. & Swanton C. (2018). *Determinants and clinical implications of chromosomal instability in cancer*. *Nat Rev Clin Oncol* 15:139–150.

Saxonov S., Berg P. & Brutlag D. L. (2006). *A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters*. *Proc. Natl. Acad. Sci. U.S.A.* 103:1412–1417.

Schena M., Shalon D., Davis R. W. & Brown P. O. (1995). *Quantitative monitoring of gene expression patterns with a complementary DNA microarray*. *Science* 270:467–470.

Schmidt S. & Debant A. (2014). *Function and regulation of the Rho guanine nucleotide exchange factor Trio*. *Small GTPases* 5:e29769.

Schmitt A. D., Hu M., Jung I., Xu Z., *et al.* (2016). *A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome*. *Cell Rep* 17:2042–2059.

Schneider D., Bianchini G., Horgan D., Michiels S., *et al.* (2015). *Establishing the Evidence Bar for Molecular Diagnostics in Personalised Cancer Care*. *Public Health Genomics* 18:349–358.

Schulz W. A., Elo J. P., Florl A. R., Pennanen S., *et al.* (2002). *Genomewide DNA hypomethylation is associated with alterations on chromosome 8 in prostate carcinoma*. *Genes Chromosomes Cancer* 35:58–65.

Segal N. H., Pavlidis P., Antonescu C. R., Maki R. G., *et al.* (2003). *Classification and subtype prediction of adult soft tissue sarcoma by functional genomics*. *Am. J. Pathol.* 163:691–700.

Serratì S., De Summa S., Pilato B., Petriella D., *et al.* (2016). *Next-generation sequencing: advances and applications in cancer diagnosis*. *Onco Targets Ther* 9:7355–7365.

Sestak I., Buus R., Cuzick J., Dubsky P., *et al.* (2018). *Comparison of the Performance of 6 Prognostic Signatures for Estrogen Receptor-Positive Breast Cancer: A Secondary Analysis of a Randomized Clinical Trial*. *JAMA Oncol* 4:545–553.

Shen R., Olshen A. B. & Ladanyi M. (2009). *Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis*. *Bioinformatics* 25:2906–2912.

Sheth J., Arnoldo A., Zhong Y., Marrano P., *et al.* (2018). *Sarcoma Subgrouping by Detection of Fusion Transcripts Using NanoString nCounter Technology*. *Pediatr. Dev. Pathol.* p. 1093526618790747.

Shimizu A., O'Brien K. P., Sjöblom T., Pietras K., *et al.* (1999). *The dermatofibrosarcoma protuberans-associated collagen type Ialpha1/platelet-derived growth factor (PDGF) B-chain fusion gene generates a transforming protein that is processed to functional PDGF-BB*. *Cancer Res.* 59:3719–3723.

Shitoh K., Konishi F., Iijima T., Ohdaira T., *et al.* (1999). *A novel case of a sporadic desmoid tumour with mutation of the beta catenin gene*. *J. Clin. Pathol.* 52:695–696.

Silva J., García V., García J. M., Peña C., *et al.* (2007). *Circulating Bmi-1 mRNA as a possible prognostic factor for advanced breast cancer patients*. *Breast Cancer Res.* 9:R55.

Skytting B., Nilsson G., Brodin B., Xie Y., *et al.* (1999). *A novel fusion gene, SYT-SSX4, in synovial sarcoma*. *J. Natl. Cancer Inst.* 91:974–975.

Smart J. E., Oppermann H., Czernilofsky A. P., Purchio A. F., Erikson R. L. & Bishop J. M. (1981). *Characterization of sites for tyrosine phosphorylation in the transforming protein of Rous sarcoma virus (pp60v-src) and its normal cellular homologue (pp60c-src)*. *Proc. Natl. Acad. Sci. U.S.A.* 78:6013–6017.

Smolle M. A., Leithner A., Posch F., Szkandera J., Liegl-Atzwanger B. & Pichler M. (2017). *MicroRNAs in Different Histologies of Soft Tissue Sarcoma: A Comprehensive Review*. *Int J Mol Sci* 18.

Snyder E. L., Sandstrom D. J., Law K., Fiore C., *et al.* (2009). *c-Jun amplification and overexpression are oncogenic in liposarcoma but not always sufficient to inhibit the adipocytic differentiation programme*. *J. Pathol.* 218:292–300.

Solomon D. A., Kim T., Diaz-Martinez L. A., Fair J., *et al.* (2011). *Mutational inactivation of STAG2 causes aneuploidy in human cancer*. *Science* 333:1039–1043.

Sotillo R., Hernando E., Díaz-Rodríguez E., Teruya-Feldstein J., *et al.* (2007). *Mad2 overexpression promotes aneuploidy and tumorigenesis in mice*. *Cancer Cell* 11:9–23.

Sparano J. A., Gray R. J., Makower D. F., Pritchard K. I., *et al.* (2018). *Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer*. *N. Engl. J. Med.* 379:111–121.

Steele C. D., Tarabichi M., Oukrif D., Webster A. P., *et al.* (2019). *Undifferentiated Sarcomas Develop through Distinct Evolutionary Pathways*. *Cancer Cell* 35:441–456.e8.

Stehelin D., Varmus H. E., Bishop J. M. & Vogt P. K. (1976). *DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA*. *Nature* 260:170–173.

Stephens P. J., Greenman C. D., Fu B., Yang F., *et al.* (2011). *Massive genomic rearrangement acquired in a single catastrophic event during cancer development*. *Cell* 144:27–40.

Storlazzi C. T., Mertens F., Nascimento A., Isaksson M., *et al.* (2003). *Fusion of the FUS and BBF2H7 genes in low grade fibromyxoid sarcoma*. *Hum. Mol. Genet.* 12:2349–2358.

Strahl B. D. & Allis C. D. (2000). *The language of covalent histone modifications*. *Nature* 403:41–45.

Stresemann C. & Lyko F. (2008). *Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine*. *Int. J. Cancer* 123:8–13.

Su A. I., Wiltshire T., Batalov S., Lapp H., *et al.* (2004). *A gene atlas of the mouse and human protein-encoding transcriptomes*. *Proc. Natl. Acad. Sci. U.S.A.* 101:6062–6067.

Subramanian A., Tamayo P., Mootha V. K., Mukherjee S., *et al.* (2005). *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles*. *Proc. Natl. Acad. Sci. U.S.A.* 102:15545–15550.

Sultan M., Schulz M. H., Richard H., Magen A., *et al.* (2008). *A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome*. *Science* 321:956–960.

Sun Y., Yu X. & Bai Q. (2015). *miR-204 inhibits invasion and epithelial-mesenchymal transition by targeting FOXM1 in esophageal cancer*. *Int J Clin Exp Pathol* 8:12775–12783.

Suster D. I., Deshpande V., Chebib I., Taylor M. S., *et al.* (2019). *Spindle cell liposarcoma with a TRIO-TERT fusion transcript*. *Virchows Arch.*

Suzuki H., Maruyama R., Yamamoto E. & Kai M. (2013). *Epigenetic alteration and microRNA dysregulation in cancer*. *Front Genet* 4:258.

Szklarczyk D., Franceschini A., Wyder S., Forslund K., *et al.* (2015). *STRING v10: protein-protein interaction networks, integrated over the tree of life*. *Nucleic Acids Res.* 43:D447–452.

Tabin C. J., Bradley S. M., Bargmann C. I., Weinberg R. A., *et al.* (1982). *Mechanism of activation of a human oncogene*. *Nature* 300:143–149.

Takai D. & Jones P. A. (2002). *Comprehensive analysis of CpG islands in human chromosomes 21 and 22*. *Proc. Natl. Acad. Sci. U.S.A.* 99:3740–3745.

Tang F., Barbacioru C., Wang Y., Nordman E., *et al.* (2009). *mRNA-Seq whole-transcriptome analysis of a single cell*. *Nat. Methods* 6:377–382.

Tang A., Gao K., Chu L., Zhang R., Yang J. & Zheng J. (2017). *Aurora kinases: novel therapy targets in cancers*. *Oncotarget* 8:23937–23954.

Tap W. D., Jones R. L., Van Tine B. A., Chmielowski B., *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*. *Lancet* 388:488–497.

Taparowsky E., Suard Y., Fasano O., Shimizu K., Goldfarb M. & Wigler M. (1982). *Activation of the T24 bladder carcinoma transforming gene is linked to a single amino acid change*. *Nature* 300:762–765.

Taylor B. S., Barretina J., Maki R. G., Antonescu C. R., Singer S. & Ladanyi M. (2011). *Advances in sarcoma genomics and new therapeutic targets*. *Nat. Rev. Cancer* 11:541–557.

Thadani R. & Tammi M. T. (2006). *MicroTar: predicting microRNA targets from RNA duplexes*. *BMC Bioinformatics* 7 Suppl 5:S20.

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

Thurlings I. & de Bruin A. (2016). *E2F Transcription Factors Control the Roller Coaster Ride of Cell Cycle Gene Expression*. *Methods Mol. Biol.* 1342:71–88.

Tian Y., Morris T. J., Webster A. P., Yang Z., *et al.* (2017). *ChAMP: updated methylation analysis pipeline for Illumina BeadChips*. *Bioinformatics* 33:3982–3984.

Tirode F., Surdez D., Ma X., Parker M., *et al.* (2014). *Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations*. *Cancer Discov* 4:1342–1353.

Tlemsani C., Leroy K., Gimenez-Roqueplo A.-P., Mansuet-Lupo A., *et al.* (2018a). *Chemoresistant pleomorphic rhabdomyosarcoma: whole exome sequencing reveals underlying cancer predisposition and therapeutic options*. *J. Med. Genet.*

Tlemsani C., Pasmant E., Boudou-Rouquette P., Bellesoeur A., *et al.* (2018b). *BRCA2 Loss-of-Function and High Sensitivity to Cisplatin-Based Chemotherapy in a Patient With a Pleomorphic Soft Tissue Sarcoma: Effect of Genomic Medicine*. *Am. J. Med. Sci.*

Torres E. M., Williams B. R. & Amon A. (2008). *Aneuploidy: cells losing their balance*. *Genetics* 179:737–746.

Toulmonde M. (2016). *Searching for aurora in the night of sarcoma phase II trials: isn't it time to move to second gear? Ann. Oncol.* 27:1815–1817.

Toulmonde M., Bonvalot S., Méeus P., Stoeckle E., *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*. *Ann. Oncol.* 25:735–742.

Trojani M., Contesso G., Coindre J.-M., Rouesse J., *et al.* (1984). *Soft-tissue sarcomas of adults; study of pathological prognostic variables and definition of a histopathological grading system*. *Int. J. Cancer* 33:37–42.

Tsai M.-C., Manor O., Wan Y., Mosammaparast N., *et al.* (2010). *Long noncoding RNA as modular scaffold of histone modification complexes*. *Science* 329:689–693.

Turc-Carel C., Philip I., Berger M. P., Philip T. & Lenoir G. (1983). *Chromosomal translocation (11; 22) in cell lines of Ewing's sarcoma*. *C R Seances Acad Sci III* 296:1101–1103.

Valentin T., Lesluyes T., Le Guellec S. & Chibon F. (2019). *Chemotherapy in localized soft tissue sarcoma: will we soon have to treat grade 1 tumors? Update on CINSARC performances*. *Ann. Oncol.* 30:153–155.

van de Vijver M. J., He Y. D., van't Veer L. J., Dai H., *et al.* (2002). *A gene-expression signature as a predictor of survival in breast cancer*. *N. Engl. J. Med.* 347:1999–2009.

Varley K. E., Gertz J., Roberts B. S., Davis N. S., *et al.* (2014). *Recurrent read-through fusion transcripts in breast cancer*. *Breast Cancer Res. Treat.* 146:287–297.

Vaske C. J., Benz S. C., Sanborn J. Z., Earl D., *et al.* (2010). *Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM*. *Bioinformatics* 26:i237–245.

Venet D., Dumont J. E. & Detours V. (2011). *Most random gene expression signatures are significantly associated with breast cancer outcome*. *PLoS Comput. Biol.* 7:e1002240.

Venkatesan S. & Swanton C. (2016). *Tumor Evolutionary Principles: How Intratumor Heterogeneity Influences Cancer Treatment and Outcome*. *Am Soc Clin Oncol Educ Book* 35:e141–149.

Venter J. C., Adams M. D., Myers E. W., Li P. W., *et al.* (2001). *The sequence of the human genome*. *Science* 291:1304–1351.

Versteege I., Sévenet N., Lange J., Rousseau-Merck M. F., *et al.* (1998). *Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer*. *Nature* 394:203–206.

Vogelstein B., Papadopoulos N., Velculescu V. E., Zhou S., Diaz L. A. & Kinzler K. W. (2013). *Cancer genome landscapes*. *Science* 339:1546–1558.

von Ahlfen S., Missel A., Bendrat K. & Schlumpberger M. (2007). *Determinants of RNA quality from FFPE samples*. *PLoS ONE* 2:e1261.

von Hansemann D. (1890). *Ueber asymmetrische Zelltheilung in Epithelkrebsen und deren biologische Bedeutung*. *Archiv für Pathologische Anatomie und Physiologie und für Klinische Medicin* 119:299–326.

Wang K. C. & Chang H. Y. (2011). *Molecular mechanisms of long noncoding RNAs*. *Mol. Cell* 43:904–914.

Wang Z., Gerstein M. & Snyder M. (2009). *RNA-Seq: a revolutionary tool for transcriptomics*. *Nat. Rev. Genet.* 10:57–63.

Wang X., Bledsoe K. L., Graham R. P., Asmann Y. W., *et al.* (2014). *Recurrent PAX3-MAML3 fusion in biphenotypic sinonasal sarcoma*. *Nat. Genet.* 46:666–668.

Wang Y., Liu C., Luo M., Zhang Z., *et al.* (2015). *Chemotherapy-Induced miRNA-29c/Catenin-δ Signaling Suppresses Metastasis in Gastric Cancer*. *Cancer Res.* 75:1332–1344.

Watson J. D. & Crick F. H. (1953). *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid*. *Nature* 171:737–738.

Watson S., Perrin V., Guillemot D., Reynaud S., *et al.* (2018). *Transcriptomic definition of molecular subgroups of small round cell sarcomas*. *J. Pathol.* 245:29–40.

Weaver B. A. A. & Cleveland D. W. (2006). *Does aneuploidy cause cancer? Curr. Opin. Cell Biol.* 18:658–667.

Weigelt B., Mackay A., A'hern R., Natrajan R., *et al.* (2010). *Breast cancer molecular profiling with single sample predictors: a retrospective analysis*. *Lancet Oncol.* 11:339–349.

Wheeler D. A. & Wang L. (2013). *From human genome to cancer genome: the first decade*. *Genome Res.* 23:1054–1062.

Wheeler D. A., Srinivasan M., Egholm M., Shen Y., *et al.* (2008). *The complete genome of an individual by massively parallel DNA sequencing*. *Nature* 452:872–876.

Whisnant A. W., Bogerd H. P., Flores O., Ho P., *et al.* (2013). *In-depth analysis of the interaction of HIV-1 with cellular microRNA biogenesis and effector mechanisms*. *MBio* 4:e000193.

Wiklund T. A., Blomqvist C. P., Räty J., Elomaa I., Rissanen P. & Miettinen M. (1991). *Postirradiation sarcoma. Analysis of a nationwide cancer registry material*. *Cancer* 68:524–531.

Wilkins M. H. F., Stokes A. R. & Wilson H. R. (1953). *Molecular structure of deoxypentose nucleic acids*. *Nature* 171:738–740.

Willems S. M., Debiec-Rychter M., Szuhai K., Hogendoorn P. C. W. & Sciot R. (2006). *Local recurrence of myxofibrosarcoma is associated with increase in tumour grade and cytogenetic aberrations, suggesting a multistep tumour progression model*. *Mod. Pathol.* 19:407–416.

Wimmer I., Tröscher A. R., Brunner F., Rubino S. J., *et al.* (2018). *Systematic evaluation of RNA quality, microarray data reliability and pathway analysis in fresh, fresh frozen and formalin-fixed paraffin-embedded tissue samples*. *Sci Rep* 8:6351.

Wong N., Lam W. C., Lai P. B., Pang E., Lau W. Y. & Johnson P. J. (2001). *Hypomethylation of chromosome 1 heterochromatin DNA correlates with q-arm copy gain in human hepatocellular carcinoma*. *Am. J. Pathol.* 159:465–471.

Wong W. J., Lauria A., Hornick J. L., Xiao S., Fletcher J. A. & Marino-Enriquez A. (2016). *Alternate PAX3-FOXO1 oncogenic fusion in biphenotypic sinonasal sarcoma*. *Genes Chromosomes Cancer* 55:25–29.

Wu S. P., Cooper B. T., Bu F., Bowman C. J., *et al.* (2017). *DNA Methylation-Based Classifier for Accurate Molecular Diagnosis of Bone Sarcomas*. *JCO Precis Oncol* 2017.

Xie B., Ding Q., Han H. & Wu D. (2013). *miRCancer: a microRNA-cancer association database constructed by text mining on literature*. *Bioinformatics* 29:638–644.

Xie C.-H., Cao Y.-M., Huang Y., Shi Q.-W., *et al.* (2016). *Long non-coding RNA TUG1 contributes to tumorigenesis of human osteosarcoma by sponging miR-9-5p and regulating POU2F1 expression*. *Tumour Biol.* 37:15031–15041.

Yates L. R. & Campbell P. J. (2012). *Evolution of the cancer genome*. *Nat. Rev. Genet.* 13:795–806.

Young M. D., Wakefield M. J., Smyth G. K. & Oshlack A. (2010). *Gene ontology analysis for RNA-seq: accounting for selection bias*. *Genome Biol.* 11:R14.

Yuan H., Qin F., Movassagh M., Park H., *et al.* (2013). *A chimeric RNA characteristic of rhabdomyosarcoma in normal myogenesis process*. *Cancer Discov* 3:1394–1403.

Yuan L., Qian G., Chen L., Wu C.-L., *et al.* (2018). *Co-expression Network Analysis of Biomarkers for Adrenocortical Carcinoma*. *Front Genet* 9:328.

Zack T. I., Schumacher S. E., Carter S. L., Cherniack A. D., *et al.* (2013). *Pan-cancer patterns of somatic copy number alteration*. *Nat. Genet.* 45:1134–1140.

Zampetaki A., Albrecht A. & Steinhofel K. (2018). *Long Non-coding RNA Structure and Function: Is There a Link? Front Physiol* 9:1201.

Zhang B. & Horvath S. (2005). *A general framework for weighted gene co-expression network analysis*. *Stat Appl Genet Mol Biol* 4:Article17.

Zhao J.-J., Lin J., Lwin T., Yang H., *et al.* (2010). *microRNA expression profile and identification of miR-29 as a prognostic marker and pathogenetic factor by targeting CDK6 in mantle cell lymphoma*. *Blood* 115:2630–2639.

Annexes

I – Le profilage génomique est un outil permettant d'éviter la classification en STUMP des tumeurs musculaires lisses de l'utérus

Modern Pathology (2018) 31, 816–828

Genome profiling is an efficient tool to avoid the STUMP classification of uterine smooth muscle lesions: a comprehensive array-genomic hybridization analysis of 77 tumors

Sabrina Croce[1,2,22], Agnès Ducoulombier[3,4,22], Agnès Ribeiro[1], Tom Lesluyes[2,5], Jean-Christophe Noel[6], Frédéric Amant[7,8], Louis Guillou[9,10], Eberhard Stoeckle[11], Mojgan Devouassoux-Shisheboran[12], Nicolas Penel[3], Anne Floquet[13], Laurent Arnould[14], Frédéric Guyon[11], Florence Mishellany[15], Camille Chakiba[13], Tine Cuppens[7], Michal Zikan[16], Agnès Leroux[17], Eric Frouin[18], Isabelle Farre[19], Catherine Genestie[20], Isabelle Valo[21], Gaëtan MacGrogan[1] and Frédéric Chibon[1,2]

[1]Department of Biopathology, Institut Bergonié, Comprehensive Cancer Centre, Bordeaux, France; [2]Institut National de la Santé et de la Recherche Medicale (INSERM) U1218, Bordeaux, France; [3]Oncology Department, Centre Oscar Lambret, Comprehensive Cancer Centre, Lille, France; [4]Oncology Department, Centre Antoine Lacassagne, Comprehensive Cancer Centre, Nice, France; [5]University of Bordeaux, Bordeaux, France; [6]Department of Pathology, Clinic of Gynecopathology and Senology, Erasme University Hospital, Brussels, Belgium; [7]KU Leuven - University of Leuven, Department of Oncology, Gynaecologic Oncology; University Hospitals Leuven, Department of Obstetrics and Gynaecology, Leuven, Belgium; [8]Centre for Gynecologic Oncology Amsterdam (CGOA), Antoni Van Leeuwenhoek - Netherlands Cancer Institute, Amsterdam, The Netherlands; [9]Argot-Lab, Lausanne, Switzerland; [10]Institut Universitaire de Pathologie, Lausanne, Switzerland; [11]Department of Surgery, Institut Bergonié, Comprehensive Cancer Centre, Bordeaux, France; [12]Department of Pathology, Hôpital Universitaire Lyon Sud, Pierre Benite, France; [13]Department of Medical Oncology, Institut Bergonié, Comprehensive Cancer Centre, Bordeaux, France; [14]Department of Pathology, Centre JF Leclerc, Comprehensive Cancer Centre, Dijon, France; [15]Department of Pathology, Centre Jean Perrin, Comprehensive Cancer Centre, Clermont-Ferrand, France; [16]Gynaecological Oncology Center, Department of Obstetrics and Gynaecology, Charles University in Prague – First Faculty of Medicine and General University Hospital, Prague, Czech Republic; [17]Department of Pathology, Centre Alexis Vautrin, Comprehensive Cancer Centre, Vandoeuvre-les Nancy, France; [18]Department of Pathology, University Hospital, Poitiers, France; [19]Department of Pathology, Centre Oscar Lambret, Comprehensive Cancer Centre, Lille, France; [20]Department of Pathology, Institut Gustave Roussy, Comprehensive Cancer Centre, Villejuif, France and [21]Department of Pathology, ICO Site Paul Papin, Comprehensive Cancer Centre, Angers, France

The diagnosis of a uterine smooth muscle lesion is, in the majority of cases, straightforward. However, in a small number of cases, the morphological criteria used in such lesions cannot differentiate with certainty a benign from a malignant lesion and a diagnosis of smooth muscle tumor with uncertain malignant potential (STUMP) is made. Uterine leiomyosarcomas are often easy to diagnose but it is difficult or even impossible to identify a prognostic factor at the moment of the diagnosis with the exception of the stage. We hypothesize, for uterine smooth muscle lesions, that there is a gradient of genomic complexity that correlates to outcome. We first tested this hypothesis on STUMP lesions in a previous study and demonstrated that this 'gray category' could be split according to genomic index into two groups. A benign group, with a low to moderate alteration rate without recurrence and a malignant group, with a highly rearranged profile akin to uterine leiomyosarcomas. Here, we analyzed a large series of 77 uterine smooth muscle lesions (from 76 patients) morphologically classified as 19 leiomyomas, 14 STUMP and 44 leiomyosarcomas with clinicopathological and genomic correlations. We

Correspondence: Dr S Croce, MD, Department of Pathology, Institut Bergonie, 229 cours de l'Argonne, Bordeaux F-33000, France.
E-mail: s.croce@bordeaux.unicancer.fr
[22]These authors contributed equally to this work.
Received 13 July 2017; revised 10 November 2017; accepted 12 November 2017; published online 12 January 2018

confirmed that genomic index with a cut-off = 10 is a predictor of recurrence ($P < 0.0001$) and with a cut-off = 35 is a marker for poor overall survival ($P = 0.035$). For the tumors confined to the uterus, stage as a prognostic factor was not useful in survival prediction. At stage I, among the tumors reclassified as molecular leiomyosarcomas (ie, genomic index ≥ 10), the poor prognostic markers were: 5p gain (overall survival $P = 0.0008$), genomic index at cut-off = 35 (overall survival $P = 0.0193$), 13p loss including *RB1* (overall survival $P = 0.0096$) and 17p gain including *MYOCD* gain (overall survival $P = 0.0425$). Based on these findings (and the feasibility of genomic profiling by array-comparative genomic hybridization), genomic index, 5p and 17p gains prognostic value could be evaluated in future prospective chemotherapy trials.
Modern Pathology (2018) **31,** 816–828; doi:10.1038/modpathol.2017.185; published online 12 January 2018

Uterine sarcomas are rare neoplasms accounting for 4% of all uterine tumors with an estimated incidence of 0.86 per 100 000 women.[1] Uterine smooth muscle lesions include very common benign tumors such as leiomyomas and malignant as leiomyosarcomas.[2] The overall incidence of leiomyomas is 70–80% by 50 years of age.[3] Leiomyomas have a benign behavior despite the benign metastasizing leiomyoma and leiomyomatoses, whereas leiomyosarcomas eventually lead to death through recurrences in up to 75% of cases.[4,5]

The diagnosis of uterine smooth muscle lesions, based on the Stanford criteria (three morphological features: presence of cytologic atypia, mitotic count and tumor cell necrosis),[6] is straightforward in the majority of the cases. However, sometimes morphology is confusing and introduces a degree of subjectivity in the interpretation of these criteria. External factors such as prior treatments (hormonal therapies) or a non-optimal fixation of the sample may pose diagnostic challenges. Hence, such lesions are usually classified as smooth muscle tumors of uncertain malignant potential (STUMP).[2] This classification could result in a risk of under or over diagnosis with clinical consequences for the treatment that could impact the patient.

The FIGO stage is still the most important uterine leiomyosarcomas prognostic factor and forms the basis of the therapeutic strategy.[4,7–9] However, FIGO staging fails to identify patients with high risk of death who could potentially be eligible for chemotherapy.[10–13] Even if this disease is clinically very aggressive,[5] it is difficult to predict the outcome especially when the diagnosis is made at stage I (confined to the uterus). Recently, a specific uterine leiomyosarcomas nomogram for predicting post-resection 5-year overall survival using seven clinicopathological items (age, tumor size, tumor grade, cervical involvement, loco-regional metastases, distant metastases and mitotic index) was published[14] and was subsequently validated on an independent series.[15] In this nomogram, the mitotic index and the tumor grade were taken into account as biological parameters. Nevertheless, the tumor grading in uterine leiomyosarcomas is controversial because by definition leiomyosarcomas diagnosed on the basis of Stanford criteria[6] are of high grade. Contrary to other soft tissue sarcomas, the tumor grade does not show a prognostic value in uterine leiomyosarcomas.[16] Among the seven parameters, three (presence of regional metastasis, distant metastasis and size) are strictly linked to the FIGO stage.[14]

Hence, there is a need to clarify these prognostic strategies. A few years ago, we published a new classification method based on genomic profiling complementary to the morphology able to distinguish within the STUMP category those uterine smooth muscle lesions with a risk of recurrence and poor outcomes from benign lesions.[17]

In this study, we analyzed the genomic profile by array-comparative genomic hybridization in a series of 77 uterine smooth muscle lesions (44 leiomyosarcomas, 14 STUMP and 19 leiomyomas) to validate the power of genomic index as a recurrence predictor. Furthermore, we set out to improve our previous results by studying a larger series with a broader follow-up and to identify overall survival prognostic factors in uterine leiomyosarcomas.

Materials and methods

Tumor Samples

Seventy-seven formalin-fixed and paraffin-embedded uterine smooth muscle lesions from 76 patients (for one patient, both the primary tumor and the recurrence were examined) were collected in France through the French sarcoma network (RRePS and GYN RRePS) (Institut Bergonié of Bordeaux, Centre Oscar Lambret of Lille, Hôpitaux Universitaires Lyon Sud, Centre JF LeClerc, Dijon, Hôpital Universitaire of Poitiers, Centre Jean Perrin, Clermont-Ferrand, Centre Alexis Vautrin, Vandouevre-les Nancy, Institut Gustave Roussy, ICO, Paul Papin, Angers), from Belgium (Hopitaux Universitaires de Bruxelles, Catholic University of Leuven), Czech Republic (University Hospital from Prague) and Switzerland (Hôpitaux Universitaires and Argot-Lab of Lausanne). Fourteen uterine tumors diagnosed as STUMP along with 19 uterine leiomyomas previously published[17] were included in the series. The tumors were diagnosed between 1977 and 2013. For each patient, 1–8 slides were available (mean: 2 slides). All cases were centrally reviewed by one of the authors (SC) and classified according to Stanford

criteria[6] and 2014 WHO guidelines for female reproductive organs.[2] Cytologic atypia was evaluated at medium power magnification (objective ×10) according to the presence of high nuclear size or nuclear pleomorphism and hyperchromatism,[18] the mitotic count was evaluated on 10 high-power fields (objective ×40, field diameter 0.53 mm) and the tumor cell necrosis defined by an abrupt transition from necrotic to non-necrotic areas without interposed granulation tissue.[18] The mitotic cut-off was 10 mitoses/10 power fields for spindle cell smooth muscle tumors, ≥4 mitoses for epithelioid and ≥2 mitoses for myxoid tumors.[2,18]

Frozen material was available for five samples. The samples from the tumor archives of each participating department were centralized in the Biological Resources Center of Institut Bergonié, which the French authorities authorized for scientific research (AC-2008-812).

DNA and RNA Extraction

Genomic DNA was extracted from formalin-fixed and paraffin-embedded tissues according to the protocol for DNA isolation from formalin-fixed and paraffin-embedded tissues (http://www.chem-agilent.com/pdf/G441090020v3_1_CGH_ULS_Protocol.pdf) (Agilent Technologies, Santa Clara, CA, USA). A cut-off of 50% of cellularity in tumor samples was set for the analysis.

Array-Comparative Genomic Hybridization Analysis

DNA was hybridized onto 8×60 K whole-genome arrays (G4450A; Agilent Technologies) according to the manufacturer's protocol. Microarray slides were scanned using a DNA Microarray Scanner, images were analyzed by Feature Extraction V10.1.1.1 followed by Agilent Cytogenomic software 4.0. The ADM-2 algorithm of the Comparative Genomic Hybridization Analytics v4.0.76 software (Agilent Technologies) was used to identify the DNA copy number anomalies at the probe level. A low-level copy number gain was defined as a log 2 ratio >0.25 and a copy number loss was defined as a log 2 ratio <0.25. A high-level gain or amplification was defined as a log 2 ratio >1.5 and a homozygous deletion was suspected when the ratio was <-1. The range for derivative log ratio spread cut-off was fixed to 0.50. Genomic index was calculated for each profile as follows: genomic index = A^2/C, where A is the total number of alterations (segmental gains and losses) and C is the number of involved chromosomes.[19,17]

Statistical Analysis

Metastasis-free survival was calculated by the Kaplan–Meier method from the date of initial diagnosis to the date of first metastasis or last follow-up. Overall survival, using the Kaplan–Meier method, was calculated from the date of diagnosis to death or last follow-up. Survival curves were compared with the log-rank test. Univariate survival analyses were performed by using the R software version 3.4.0 (R Development Core Team, Vienna, Austria, 2009) and the 'survival' package (A Package for Survival Analysis in S; Terry Therneau, February 2002; R package, version 2.40-1). To test whether gene alterations (losses or gains) are enriched in groups of tumors classified by the genomic index, Fisher's exact tests were performed. Multivariate survival analyses were performed using Cox regression with Firth's correction with the 'coxphf' package (Georg Heinze and Meinhard Ploner, 2016, version 1.12).

Results

Pathologic Features

After centralized pathological review, the series comprised 19 leiomyomas, 14 STUMP and 44 leiomyosarcomas. Morphologically, all leiomyomas but one case of bizarre nuclei leiomyoma, showed conventional spindle cell morphology. Spindle cell features were observed in the STUMP group (all 14 tumors).

Among the 44 leiomyosarcomas (43 patients), 29/44 tumors showed a spindle cell morphology, 10/44 epithelioid, 4/44 pleomorphic with giant osteoclastic cells (with negative stains for melanocytic markers) and 1/44 myxoid (Figure 1). The clinicopathological data are summarized in Table 1.

Genomic Data and Clinical Correlations

Genome complexity evaluation by genomic index assessment (quantitative approach) and prognostic value. Array-comparative genomic hybridization was analyzable in all 77 tumors. Follow-up data were available for all but two patients with leiomyomas (mean follow-up: 63.6 months, range: 9–232 months). The genomic profiling split the present series of uterine smooth muscle lesions in two groups according to the cut-off defined in our previous paper (genomic index = 10):[17] a group with genomic index <10 (19/74) and a second group with genomic index ≥10 (55/74; Figure 2a).

The first group (genomic index <10) is characterized by a low level of chromosomal rearrangements (Figure 3a; mean genomic index: 2.3, range: 0–9.14) in contrast with the second group (genomic index ≥10) harboring complex genomic profiles (mean genomic index: 51.8, range: 11–180; Figure 3b).

The Kaplan–Meier analysis demonstrated a significant difference in clinical outcome with no recurrence in the group with genomic index <10

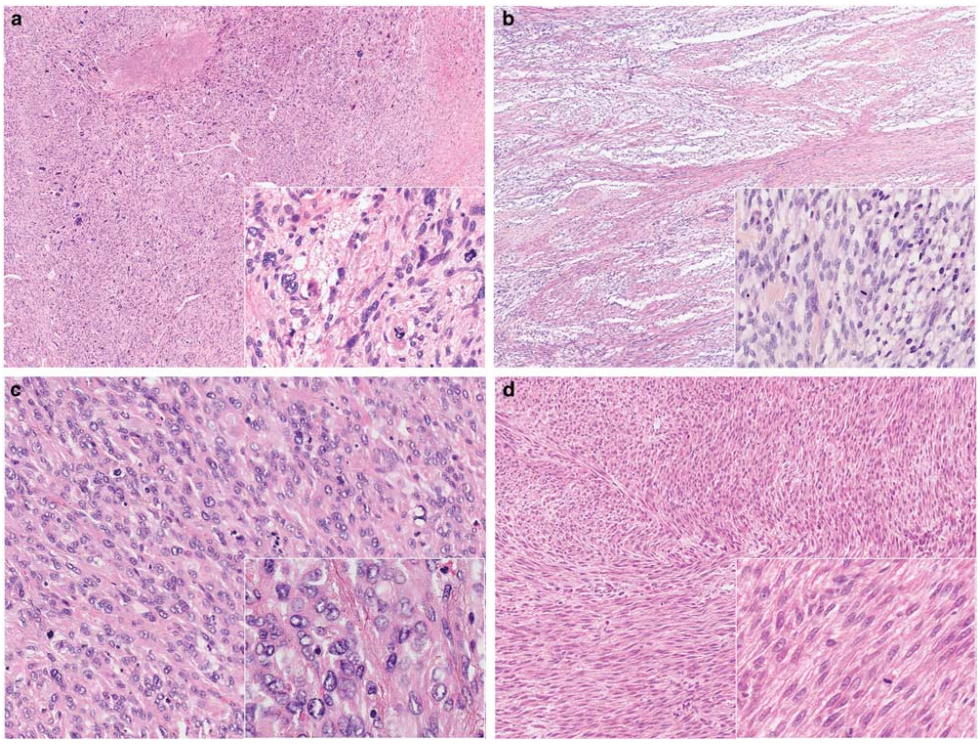


Figure 1 A representation of the morphological features of uterine smooth muscle lesions. (**a**) Uterine leiomyosarcoma with spindle cells, pleomorphic and osteoclastic features and atypical mitoses, stage FIGO IB and genomic index = 84. The patient developed lung metastases and died 32.4 months after the diagnosis (patient 68). (**b**) Myxoid leiomyosarcoma with infiltration of myometrial wall. Some areas are highly cellular with four mitoses, stage FIGO 2B and genomic index = 13.5 (patient 47). (**c**) Epithelioid leiomyosarcoma with genomic index = 27. The patient died of disease 26 months after the diagnosis with peritoneal metastases (patient 48). (**d**) Smooth muscle lesion with mild atypia, 20 mitoses (STUMP) with genomic index = 14.3. The patient had multiple peritoneal and retroperitoneal recurrences and she is alive after 103.2 months (patient 33).

and recurrences and deaths in the second group (Figure 2b), confirming our previous results.[17]

Morphologically, the first group included all leiomyomas, two STUMPs and no leiomyosarcomas. The second group included all leiomyosarcomas and 12 STUMPs. Only four patients with leiomyosarcomas did not recur (of note, one patient died of pulmonary embolism 9 days after surgery) and among the 12 STUMPs with genomic index ≥ 10, seven recurred (Table 1).

All leiomyomas showed a flat or very simple profile (mean genomic index: 1.9, range: 0–9.1). The leiomyosarcomas group showed a complex, rearranged chromosome profile with numerous intrachromosomal breaks (gains and losses). The mean genomic index in the leiomyosarcomas group was 55 (range: 13.5–180). The STUMP group was then split into two groups according to genomic index. Owing to the clinical aggressiveness and outcome of the tumors with genomic index ≥ 10, they were thereafter considered as leiomyosarcomas.

Therefore, the second step was to identify a death predictor in this newly defined group of leiomyosarcomas. Different genomic index cut-offs (five for each step) were tested and a cut-off of genomic index = 35 was identified as an efficient predictor of death ($P = 0.035$, HR = 2.18 (1.04–4.58); Figure 2c).

In order to identify specific chromosomal alterations (qualitative approach) related to prognostic value, we analyzed differential penetrance plots among the patients alive, whose tumor had genomic index ≥ 10 (19/55) and patients who died of the disease (36/55). We identified five frequent alterations, in addition to genomic index ≥ 35 (Figure 2c), associated with overall survival: 5p gain (Figure 2d, Table 2a), 1p gain, 13q loss (including *RB1*) and 17p gain including *MYOCD* (Table 2a). Staging being the gold standard for clinical/pathological prognosis, all

Table 1

Clinical data										*Genomic data*									
Patient	*FU status*	*Age*	*RFS months*	*OS months*	*Recurrence*	*Type of recurrence*	*FIGO stage*	*Size cm*	*Location*	*Type of surgery*	*Centralized diagnosis*	*Morphology*	*FH features*	*5p gain*	*1p gain*	*17p gain*	*17p loss*	*13 loss*	*GI*
1	A-NOS	40	53.6	53.6	No			4	Uterus	Hysterectomy R0	LM	Spindle	No	0	0	0	0	0	0
2	A-NOS	49	75.3	75.3	No			3.7	Uterus	Hysterectomy R0	LM	Spindle	No	0	0	0	0	0	1
3	A-NED	50	68.6	68.6	No			5.5	Uterus	Hysterectomy R0	LM	Spindle	Yes	0	0	0	0	0	0
4	A-NOS	63	37	37	No			1.4	Uterus	Hysterectomy R0	LM	Spindle	No	0	0	0	0	0	0
5	Lost	45						3.5	Pelvis/ peritoneum	Total resection R0	LM	Spindle	No	0	0	0	0	0	4
6	A-NED	46	58.4	58.4	No			4	Uterus	Total hysterectomy R0	LM	Spindle	Yes	0	0	0	0	0	2
7	A-NED	48	53	53	No			15	Uterus	Total hysterectomy R0	LM	Spindle	Yes	0	0	0	0	0	2
8	A-NED	49	72.9	72.9	No			6.5	Uterus	Total hysterectomy R0	LM	Spindle	Yes	0	0	0	0	0	0
9	DOC	67	21.3	21.3	No			7	Large ligament	Total resection R0	LM	Spindle	No	0	0	0	0	0	2
10	Lost	48						5.5	Uterus	Total hysterectomy R0	LM	Spindle cellular	Yes	0	0	0	0	0	5
11	A-NED	32	101.5	101.5	No			6.5	Uterus	Myomectomy R0	LM	Spindle	Yes	0	0	0	0	0	0
12	A-NED	44	102	102	No			4.5	Uterus	Total hysterectomy R0	LM	Spindle	No	0	0	0	0	0	0
13	A-NED	67	102.1	102.1	No			1.5	Uterus	Total hysterectomy R0	LM	Spindle	No	0	0	0	0	0	0
14	A-NED	68	102.2	102.2	No			1.2	Uterus	Total hysterectomy R0	LM	Spindle	Yes	0	0	0	0	0	0
15	A-NED	40	137.8	137.8	No			4.5	Uterus	Total hysterectomy R0	LM	Spindle	Yes	0	0	0	0	0	1
16	A-NED	36	137.2	137.2	No			15	Uterus	Myomectomy R0	LM	Spindle	Yes	0	0	0	0	0	1
17	A-NED	44	175.2	175.2	No			4.5	Uterus	Total hysterectomy R0	LM	Spindle	Yes	0	0	0	0	0	1
19	A-NED	80	93.8	93.8	No	Residual disease	ND	15	Pelvis/ uterus	Total hysterectomy R1	STUMP	Spindle	Yes	0	0	0	0	0	3
20	A-NED	36	84.2	84.2	No			7	Uterus	Total hysterectomy R0	LM	BN-LM	Yes	0	0	0	-2 TP53	-2 RB1	9
21	A-NED	34	128.9	128.9	No		ND	NA	Uterus	Myomectomy R0	STUMP	Spindle	Yes	0	0	0	0	0	8.3
22	A-NED	47	99.9	99.9	No			6	Uterus	Total hysterectomy R0	LM	Spindle	No	0	0	0	0	0	9.14
23	A-NED	63	51.6	51.6	No		IB	7	Uterus	Total hysterectomy	STUMP	Spindle	Yes	0	0	0	0	0	56.81
24	A-NED	42	NA	61.2	No		IB	8	Uterus	Subtotal hysterectomy	STUMP	Spindle	No	0	0	0	0	0	32.6
25	A-NED	47	NA	99.6	No		IB	7	Uterus	Total hysterectomy	STUMP	Spindle	Yes	0	0	0	0	0	14.28
26	AWD	48	36	162.0	Yes	Bladder, rectum, omentum, para-aortic LN, lung	IB	7	Uterus	Total hysterectomy	STUMP	Spindle	Yes	0	0	0	0	-1RB1	16.9

Table 1 (Continued)

Clinical data										*Genomic data*									
Patient	*FU status*	*Age*	*RFS months*	*OS months*	*Recurrence*	*Type of recurrence*	*FIGO stage*	*Size cm*	*Location*	*Type of surgery*	*Centralized diagnosis*	*Morphology*	*FH features*	*5p gain*	*1p gain*	*17p gain*	*17p loss*	*13 loss*	*GI*
27	AWD	50	57	104.4	Yes	Paravertebral soft tissue, bone	I	NA	Uterus	Total hysterectomy	STUMP	Epithelioid	No	0	0	0	-1 no TP53	-1 RB1	22.23
28	A-NOS	77	17	25.2	Yes	Peritoneum, ileum, vagina, pelvis	IB	11	Uterus	Total hysterectomy	STUMP	Spindle	No	0	1	0	0	-2 RB1	48
29	A-NED	60	NA	42.0	No		IA	2	Uterus	Myomectomy	STUMP	Spindle	No	0	1	2 ampl MYOC	0	-2 RB1	52
30	DOD	66	55	55.2	Yes	Vagina	IB	8	Uterus	Total hysterectomy	STUMP	Spindle	Yes	0	0	0	0	-1 RB1	21.33
31	A-NED	85	NA	106.8	No		IB	11	Uterus	Total hysterectomy	STUMP	Epithelioid	No	0	1	0	0	-1 RB1	94.1
32	DOD	43	10	15.6	Yes	Bone	IB	20	Uterus	Total hysterectomy	STUMP	Spindle	Yes	0	1	2 ampl MYOC	-1 TP53	-1 RB1	100
33	A-NED	46	35	103.2	Yes	Peritoneum, right ovary, retroperitoneum	IA	5	Uterus	Total hysterectomy	STUMP	Spindle	Yes	0	0	0	0	0	14.3
34	A-NED	48	12	147.6	Yes	Uterine cervix, peritoneum, soft tissue (leg and arm)	I	NA	Uterus	Total hysterectomy	STUMP	Spindle	Yes	0	0	0	0	0	11
35	DOD	55	26.4	56.4	Yes	Pelvis, peritoneum	III	NA	Uterus	Total hysterectomy	LMS	Spindle	Yes	0	0	0	0	0	44.64
36	DOD	51	26.4	61.2	Yes	Soft tissue, bone, lung	IB	9	Uterus	Total hysterectomy	LMS	Epithelioid	Yes	0	0	0	0	-2 RB1	156
37	DOD	56	136.8	171.6	Yes	Lung	IA	1.8	Uterus	Total hysterectomy	LMS	Epithelioid	No	0	1	0	-1 TP53	-2 RB1	18
38	DOD	50	19.2	57.6	Yes	Lung	IB	7	Uterus	Total hysterectomy	LMS	Spindle	Yes	0	0	0	-2TP53	-2 RB1	22
39	DOD	59	21.6	48	Yes	Lung	IA	3	Uterus	Total hysterectomy	LMS	Spindle	Yes	0	0	1 MYOCD	-1 TP53	-2RB1	52.26
40	AWD	40	72	110.4	Yes	Clavicular LN, lung, pancreas	IB	19	Uterus	Total hysterectomy	LMS	Spindle	Yes	0	0	1 MYOCD	-2TP53	-1 RB1	25
40	AWD	45	NA	NA	—		NA	NA	Uterus	Total hysterectomy	LMS	Spindle	No	0	0	1 MYOCD	-2TP53	-1 RB1	69
41	DOD	40	0	13.2	Yes	Lung	IVB	8	Uterus	Total hysterectomy	LMS	Spindle	Yes	0	1	0	-1 TP53	-1 RB1	28.9
42	DOD	63	8.5	27.6	Yes	Lung	IB	7	Uterus	Total hysterectomy	LMS	Spindle	No	0	1	0	0	-2 RB1	19.6
43	DOD	53	9	27.6	Yes	Lung	III	2.8	Uterus	Total hysterectomy	LMS	Spindle	No	1	1	1 MYOCD	- 1 TP53	-1 RB1	69
44	DOD	62	1.6	34.8	Yes		IA	4.5	Uterus	Total hysterectomy	LMS	Spindle	Yes	1	0	2MYOCD	-1 TP53	-2 RB1	40
45	AWD	63	1.2	84	Yes	Lung	IB	10	Uterus	Total hysterectomy	LMS	Spindle	No	0	0	0	0	-2 RB1	15.12
46	DOD	68	ND	6	ND	ND	IIB	12	Uterus	Total hysterectomy	LMS	Epithelioid	Yes	0	1	1 MYOCD	0	-1RB1	46
47	DOC pulmonary embolism	62	NA	0 9 days	No		IIB	8	Uterus	Total hysterectomy	LMS	Myxoid	No	0	0	0	0	0	13.5
48	DOD	71	13.2	26	Yes	Peritoneum	III	12	Uterus	Total hysterectomy	LMS	Epithelioid	No	0	0	2 ampl MYOC	-2 TP53	-1RB1	27

Modern Pathology (2018) 31, 816–828

Table 1 (Continued)

Clinical data										*Genomic data*									
Patient	*FU status*	*Age*	*RFS months*	*OS months*	*Recurrence*	*Type of recurrence*	*FIGO stage*	*Size cm*	*Location*	*Type of surgery*	*Centralized diagnosis*	*Morphology*	*FH features*	*5p gain*	*1p gain*	*17p gain*	*17p loss*	*13 loss*	*GI*
49	DOD	57	19.2	23	Yes	Lung	IB	7.5	Uterus	Total hysterectomy	LMS	Spindle	No	0	0	0	0	-1RB1	36
50	DOD	38	27.6	ND	Yes	Lung	IB	7.5	Uterus	Total hysterectomy	LMS	Spindle	No	0	0	1MYOCD	0	-1RB1	41.66
51	DOD	47	12	32	Yes	Vagina, bladder, rectum	I	NA	Uterus	Total hysterectomy	LMS	Spindle	No	0	0	1MYOCD	0	0	44
52	DOD	59	0	3	Yes	Lung	IVB	23	Uterus	Total hysterectomy	LMS	Spindle	Yes	0	0	0	0	0	33
53	DOD	87	1.6	3.2	Yes	Peritoneum, ileum	IIB	11	Uterus	Total hysterectomy	LMS	Epithelioid	No	1	1	1MYOCD	0	-2 RB1	38.4
54	DOD	58	0	2.9	Yes	Lung	IVB	6	Uterus	Total hysterectomy	LMS	Spindle	No	1	1	1 no MYOCD	0	-1RB1	24.9
55	DOD	26	25.2	48	Yes	Soft tissue, lung	IA	4	Uterus	Total hysterectomy	LMS	Spindle	No	1	1	1MYOCD	- 1 TP53	-2 RB1	141
56	DOD	60	0	ND	Yes	Peritoneum, omentum	III	15	Uterus	Total hysterectomy	LMS	Pleomorphic osteoclastic	No	1	1	0	0	-1RB1	35
57	DOD	63	2.8	8.2	Yes	Lung, peritoneum, brain	IB	10	Uterus	Total hysterectomy	LMS	Pleomorphic osteoclastic	Yes	0	1	0	0	-1RB1	24.9
58	A-NED	76	NA	104.4	No		IA	2	Uterus	Total hysterectomy	LMS	Epithelioid	No	0	0	0	0	0	18.7
59	DOD	67	7	12	Yes	Peritoneum	III	20	Uterus	Total hysterectomy	LMS	Spindle	No	0	0	0	0	-1 RB1	17.28
60	DOD	27	12	19.2	Yes	Peritoneum	IIB	8	Uterus	Total hysterectomy	LMS	Spindle	Yes	0	0	1MYOCD	-2 TP53	-2RB	39.7
61	DOD	44	0	13.2	Yes	Lung, peritoneum	IVB	16	Uterus	Total hysterectomy	LMS	Spindle	No	1	1	1 no MYOCD	- 1 TP53	- 1 RB1	88.47
62	DOD	54	73.2	102	Yes	Lung	IB	7.5	Uterus	Total hysterectomy	LMS	Spindle	No	0	0	0	- 1 TP53	- 1 RB1	60
63	AWD	59	14.4	75.6	Yes	Lung, liver, peritoneum	IVA	6	Uterus	Total hysterectomy	LMS	Spindle	Yes	0	0	1 no MYOCD	-2 TP53	-2 RB1	115
64	DOD	58	50.4	232.8	Yes	Lung	I	NA	Uterus	Total hysterectomy	LMS	Spindle	Yes	0	0	1 no MYOCD	0	0	48
65	DOD	36	2.9	9.4	Yes	Vagina, peritoneum	IB	6	Uterus	Total hysterectomy	LMS	Spindle	No	1	0	2 MYOCD	-2 TP53	-2RB	82
66	A-NED	53	NA	57.6	No		IB	6	Uterus	Total hysterectomy	LMS	Spindle	No	0	0	0	- 1 TP53	-2RB	45
67	DOD	63	10.6	18	Yes	Lung, bone, pelvis	IB	10	Uterus	Total hysterectomy	LMS	Spindle	No	1	0	1MYOCD	0	-2 RB1	101
68	DOD	60	1.1	32.4	Yes	Lung	IB	9	Uterus	Total hysterectomy	LMS	Pleomorphic osteoclastic	No	1	1	1MYOCD	0	- 1 RB1	84
69	A-NED	43	NA	42	No		IB	6.5	Uterus	Total hysterectomy	LMS	Epithelioid	Yes	0	0	0	0	0	32
70	DOD	58	1.1	3.4	Yes	Peritoneum	IVA	14	Uterus	Total hysterectomy	LMS	Epithelioid	Yes	1	1	1MYOCD	0	- 1 RB1	72.2
71	DOD	76	2.3	2.6	Yes	Lung, peritoneum	III	7	Uterus	Total hysterectomy	LMS	Spindle	Yes	1	1	1MYOCD	- 1 TP53	-2RB	64.69
72	A-NED	69	4.5	96	Yes	Lung	IB	30	Uterus	Total hysterectomy	LMS	Epithelioid	Yes	0	0	1MYOCD	0	- 1 RB1	180
73	AWD	69	21.6	58.8	Yes	Lung	III	13	Uterus	Total hysterectomy	LMS	Spindle	No	0	0	0	0	- 1 RB1	33.9

Table 1 (Continued)

Clinical data													*Genomic data*								
Patient	*FU status*	*Age*	*RFS months*	*OS months*	*Recurrence*	*Type of recurrence*	*FIGO stage*	*Size cm*	*Location*	*Type of surgery*	*Centralized diagnosis*	*Morphology*	*FH features*	*5p gain*	*1p gain*	*17p gain*	*17p loss*	*13 loss*	*GI*
74	DOD	80	1.3	2.6	Yes	Peritoneum, bone, lung	III	14	Uterus	Total hysterectomy	LMS	Epithelioid	Yes	1	1	1 no MYOCD	- 1 TP53	- 1 RB1	114.33
75	DOD	55	NA	39.6	ND	ND	IB	10	Uterus	Total hysterectomy	LMS	Spindle	Yes	0	0	0	0	-2 RB1	56.81
76	DOD	57	48	78.0	Yes	Pelvis, peritoneum, ureter	IB	7	Uterus	Total hysterectomy	LMS	Spindle	Yes	0	0	1MYOCD	0	- 1 RB1	42.88
77	DOD	53	8.1	57.6	Yes	Peritoneum, liver, lung	IB	13	Uterus	Total hysterectomy	LMS	Pleomorphic osteoclastic	Yes	1	0	1MYOCD	0	- 1 RB1	60.84

Abbreviations: A-NED: alive not evidence of disease; A-NOS: alive not otherwise specified; AWD: alive with disease; DOD: dead of disease; DOC: dead of other disease; GI: genomic index; MYOCD: myocardin gene; OS: overall survival; RB1: retinoblastoma gene; RFS: relapse-free survival; -1 heterozygous loss; -2 homozygous loss. 1 gain; +2 amplification.

markers (including genomic index ≥ 35) were tested for independency against the stage in a multivariate analysis. Stage and 5p gain were shown to be statistically independent prognostic factors at multivariate analysis (stage: $P < 0.001$, HR = 4.36 (1.82–10.42), 5p gain: $P = 0.04$, HR = 3.24 (1.04–10.04); Table 2a). Given that stage I tumors can have a risk of metastasis, we then further refined staging by testing molecular prognostic factors within stage. Among the tumors with genomic index ≥ 10 and stage I (37/55 patients), four overall survival prognostic factors were still significant: 5p gain ($P < 0.001$, HR = 4.88 (1.74–13.7); Figure 2e), genomic index at the cut-off of 35 ($P = 0.0193$, HR = 3.2 (1.15–8.92); Figure 2f), 13 chromosome loss including *RB1* ($P = 0.0096$, HR = 9.04 (1.2–67.81)) (Supplementary Figure 1A) and 17p gain including *MYOCD* ($P = 0.0425$, HR = 2.45 (1–5.97)) (Supplementary Figure 1B).

Among the other clinicomorphological parameters tested, the presence of moderate and marked atypia (Fisher's $P = 0.043$), the presence of tumor necrosis (Fisher's $P = 0{,}001$) and a mitotic index (cut-off ≥ 20; *t*-test $P < 0.001$) were poor prognostic markers for overall survival (Table 2b,Supplementary Figure 1C). For these parameters, multivariate analysis showed that genomic index ≥ 35 remained significantly independent ($P = 0.0333$; Table 2c).

Correlation Between Chromosomal Alterations and Morphology

No correlation was found between genomic index and morphology of the tumor (spindle, epithelioid or myxoid). No correlation was observed between any specific genomic alteration (5p gain, 17p loss, 13 loss, 17p gain) and tumor morphology (spindle *vs* epithelioid).

Discussion

In the last 5 years, many analyses based on whole-genome approaches have improved our knowledge of uterine smooth muscle lesion biology.[20–22] Nevertheless, the routine diagnostic practice lacks complementary diagnostic tools. For uterine leiomyosarcomas, despite very aggressive clinical features,[5] it is difficult to predict the outcome, especially when the diagnosis is made at stage I (tumor confined to the uterus).

According to the literature,[8,10,23] adjuvant treatment of a stage I uterine leiomyosarcomas is an option without evidence of benefit. There is a need for clinical trials to highlight the benefit of chemotherapy in terms of overall survival and relapse-free survival. In absence of clinical prognostic markers, the identification of new genomic prognostic markers appears critical for setting up clinical trials aiming to evaluate new treatments in uterine leiomyosarcomas. Genomic prognostic markers

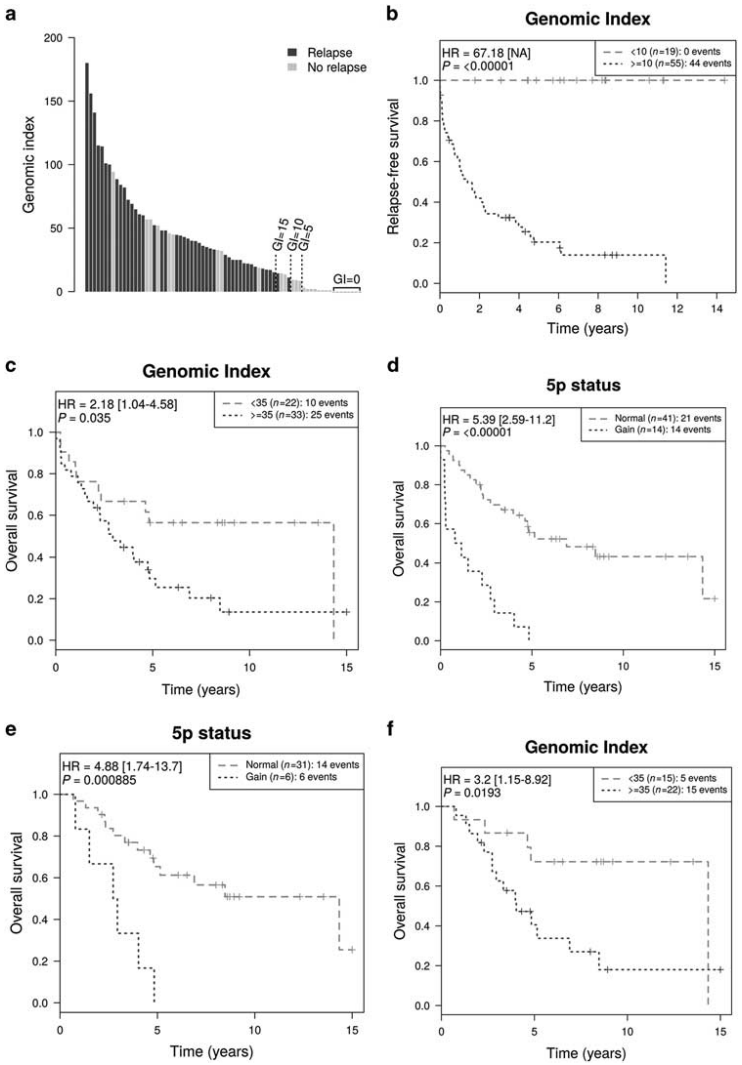


Figure 2 Genomic index and genomic alterations and clinical outcome. (**a**) Histogram. The number of patients with relapse (black bar) and without relapse (gray bar) shown in *x* axis; the genomic index level in *y* axis. The genomic index at the cut-off of 10 splits the population of 74 patients with uterine smooth muscle lesion in a group of 19 patients without metastases and a group of 55 patients with 44 metastatic events. (**b**) Kaplan-Meier relapse-free survival analysis of 74 uterine smooth muscle tumors according the genomic index at the cut-off of 10. (**c**) Kaplan-Meier for overall survival according to genomic index at the cut-off of 35 in the subgroup with genomic index ≥ 10 (genomic leiomyosarcomas). (**d**) Kaplan-Meier analysis of overall survival in the group with genomic index ≥ 10 (genomic leiomyosarcomas) with 5p gain. (**e**) Kaplan-Meier analysis of overall survival in the group with genomic index ≥ 10 (genomic leiomyosarcomas) with 5p gain with tumors limited to the uterus (stage I). (**f**) Kaplan-Meier analysis of overall survival in the group with genomic index ≥ 10 (genomic leiomyosarcomas) with tumors limited to the uterus (stage I) according to genomic index at the cut-off of 35.

could have an essential role in deciding on adjuvant chemotherapy for stage I uterine leiomyosarcomas. In our series of 37 molecular leiomyosarcomas (genomic index ≥10) at stage I, four genomic prognostic markers correctly separated the outcomes for patients alive or dead: 5p gain, chromosome 13 loss including *RB1*, proximal 17p gain with *MYOCD* and genomic index ≥ 35.

Figure 3 Penetrance plots of the different subgroups classified according to genomic index. (**a**) Penetrance plot of the tumors with genomic index < 10 (benign tumors: all leiomyomas and two STUMPs). (**b**) Penetrance plot of tumors with genomic index > 10 (malignant tumors: all leiomyosarcomas and 12 STUMPs). (**c**) Penetrance plot of tumors with genomic index > 10 (malignant tumors: all leiomyosarcomas and 12 STUMPs) of patients alive. (**d**) Penetrance plot of tumors with genomic index > 10 (malignant tumors: all leiomyosarcomas and 12 STUMPs) of patients dead of disease.

Further analyses are required to understand whether this is due to a chromosomal mechanism (specific or general) or due to genes located in these regions that are specifically overexpressed as a consequence of a chromosomal gain. The 5p gain was previously reported in extra-uterine[24] and uterine leiomyosarcomas[25] but no association with outcome was observed. The 17p proximal gain, including the *MYOCD* gene, was previously reported in literature in soft tissue leiomyosarcomas[24,26–28] and in uterine leiomyosarcomas.[25] *MYOCD* gene induces smooth muscle differentiation and promotes cell migration.[27] In a human uterine leiomyosarcomas cell line, *MYOCD* induced a phenotypic cell switch from a dedifferentiated to a differentiated smooth muscle phenotype.[29] MYOCD expression level controls smooth muscle differentiation protein expression and has an impact on cell migration in soft tissue leiomyosarcomas.[27] Furthermore, it confers aggressive outcome in soft tissue sarcomas.[30]

Hu et al[25] found a gain of 17p in 38% of the uterine leiomyosarcomas and interestingly, no 17p gain was found in alive patients (4/19). Chromosome 13 was lost in 80% of the leiomyosarcomas in our

Table 2 Statistical data: univariate and multivariate analyses for OS of the subgroup with GI >10 (malignant tumors)

	Univariate		*Multivariate*	
OS	**P-*value***	***HR***	**P-*value***	***HR***
Stage	**< 0.001**	**4.32 (2.03–9.2)**	**< 0.001**	**4.36 (1.82–10.42)**
5p gain	**< 0.001**	**5.39 (2.59–11.2)**	**0.0042**	**3.24 (1.04–10.04)**
1p gain	**< 0.001**	**3.3 (1.66–6.56)**	0.2989	1.61 (0.66–3.96)
GI ≥ 35	**0.0349**	**2.18 (1.04–4.58)**	0.8419	1.10 (0.44–2.71)
17p loss TP53	0.1119	**1.72 (0.87–3.36)**	0.4807	0.75 (0.33–1.68)
13 loss RB1	**0.0103**	**4.19 (1.28–13.77)**	0.1860	2.44 (0.65–9.19)
17p gain MYOCD	**0.0054**	**2.57 (1.29–5.13)**	0.6188	1.29 (0.48–3.47)

Abbreviations: GI: genomic index; HR: hazard ratio; OS: overall survival.
Bold values signify the significant results.

Table 2B Univariate analysis for RFS and OS of the subgroup with GI >10 (malignant tumors)

	RFS	*OS*
Atypia	***P* = 0.04918**	***P* = 0.04345**
Mitoses cut-off 20	***P* < 0.001**	***P* < 0.001**
Necrosis	*P* = 0.0963	*P* = 0.001

Abbreviations: GI: genomic index; OS: overall survival; RFS: relapse-free survival.
Bold values signify the significant results.

Table 2C Multivariate analysis for OS for stage I LMS

OS	P-*value*	*HR*
GI ≥ 35	**0.0333**	**3.10 (1.09–10.78)**
Atypia	0.0837	10.89 (0.77–>100)
Mitoses cut-off 20	0.128	0.48 (0.09–1.63)
Necrosis	0.263	0.48 (0.09–1.63)

Abbreviation: OS: overall survival.
Bold values signify the significant results.

series and is the most common genomic event in uterine (76%) [25] and extra-uterine leiomyosarcomas (ranging from 54% [28,24] to 71% [26]) and in the majority of the cases, correlation between this event and follow-up was not established.

The morphological analysis based on the presence of atypia, mitotic count and tumor cell necrosis correlated to a poor outcome in our series. The prognostic value of cytological atypia,[31,32] as well the mitotic count[5,33,34] was reported in previous publications. However, atypia and mitoses could be difficult to assess and there is only a moderate inter-observer agreement on tumor cell necrosis in uterine smooth muscle lesions among gynecological pathologists.[35] Genomic index assessment could be a useful tool, as highlighted in our multivariate analysis (Table 2c), to avoid such diagnostic discrepancies.

In our series, the genomic index cut-off of 10 splits the STUMP group into two: a flat or very simple genomic profile group akin to leiomyomas and a group of tumors with complex genomic profile similar to leiomyosarcomas (with recurrences and deaths) thereby erasing the STUMP category. In the benign lesions group with genomic index < 10 (all leiomyomas and two STUMP), there were no chromosomal alterations such as *RB1* and *TP53* loss (Table 1). One exception is the bizarre nuclei leiomyoma case, with a borderline genomic index = 9, which showed chromosome 13 loss including the *RB1* gene and chromosome 17p loss including the *TP53* gene. These alterations have already been reported in these benign lesions.[36–39] In fact, some bizarre nuclei leiomyomas inexplicably show rearranged profiles (in our experience lower than leiomyosarcomas) and a good outcome. Furthermore, the origin of this subtype of leiomyoma is not clear either.

Genomic profiling by array-comparative genomic hybridization on formalin-fixed and paraffin-embedded samples is a useful, easy and accessible tool complementary to the morphological approach. It is a diagnostic tool that splits the STUMP category into benign (leiomyoma) and malignant (leiomyosarcomas) tumors. It is a prognostic marker and a predictor of overall survival in stage I uterine leiomyosarcomas. Indeed, all the comparative genomic hybridization analyses on our series were performed on formalin-fixed and paraffin-embedded and -extracted DNA, with 100% feasibility. Genomic profiling could be used even on a limited amount of material such as pre-operative biopsies in order to guide surgical intervention (hysterectomy *vs* myomectomy or a minimally invasive surgery).

In conclusion, we have demonstrated that STUMP classification could be overcome by utilizing genomic index at the cut-off of 10. The 5p gain as genomic event and the stage as clinical parameter are poor overall survival prognostic factors in uterine leiomyosarcomas. In stage I tumors, the 5p gain, 17p gain, chromosome 13 loss and genomic index at the cut-off of 35 are poor prognostic factors of overall survival and therefore they could be potential parameters for randomization in prospective clinical trials. This approach opens the way to new insights into uterine and other gynecological smooth muscle lesions and would allow reclassification of lesions

according to genomic complexity. In fact, there is a continuum gradient of genomic complexity and instability correlating with tumor aggressiveness. As genomic profiling by array-comparative genomic hybridization on formalin fixed and paraffin embedded is feasible and accessible in hospital laboratories, this approach could be used in routine practice as a complement to histology.

Acknowledgments

This study was supported by the grants from ARC and la Ligue Régionale contre le cancer. We thank Dr Ravi Nookala of Institut Bergonié for the medical writing service and Magali Philip, Quitterie Fontanges, Patrick Murat and Virginie Duvignau for photographs. FA is senior researcher for the Research Fund Flanders (F.W.O.).

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

1 Skorstad M, Kent A, Lieng M. Uterine leiomyosarcoma - incidence, treatment, and the impact of morcellation. A nationwide cohort study. Acta Obstet Gynecol Scand 2016;95:984–990.
2 Oliva E, Carcangiu ML, Carinelli SG, *et al.* Mesenchymal tumours. Smooth muscle tumour of uncertain malignant potential. In: Kurman RJ CM, Herrington CS, Young RH (eds). WHO Classification of Tumours of Female Reproductive Organs, 2014, pp 135–147.
3 Baird DD, Dunson DB, Hill MC, *et al.* High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence. Am J Obstet Gynecol 2003;188:100–107.
4 Ricci S, Stone RL, Fader AN. Uterine leiomyosarcoma: epidemiology, contemporary treatment strategies and the impact of uterine morcellation. Gynecol Oncol 2017;145:208–216.
5 Garcia C, Kubat JS, Fulton RS, *et al.* Clinical outcomes and prognostic markers in uterine leiomyosarcoma: a population-based cohort. Int J Gynecol Cancer 2015;25:622–628.
6 Bell SW, Kempson RL, Hendrickson MR. Problematic uterine smooth muscle neoplasms. A clinicopathologic study of 213 cases. Am J Surg Pathol 1994;18:535–558.
7 Amant F, Coosemans A, Debiec-Rychter M, *et al.* Clinical management of uterine sarcomas. Lancet Oncol 2009;10:1188–1198.
8 Hensley ML, Ishill N, Soslow R, *et al.* Adjuvant gemcitabine plus docetaxel for completely resected stages I-IV high grade uterine leiomyosarcoma: results of a prospective study. Gynecol Oncol 2009;112:563–567.
9 Amant F, Lorusso D, Mustea A, *et al.* Management strategies in advanced uterine leiomyosarcoma: focus on trabectedin. Sarcoma 2015;2015:704124.
10 Omura GA, Blessing JA, Major F, *et al.* A randomized clinical trial of adjuvant adriamycin in uterine sarcomas: a Gynecologic Oncology Group Study. J Clin Oncol 1985;3:1240–1245.
11 Reed NS, Mangioni C, Malmstrom H, *et al.* Phase III randomised study to evaluate the role of adjuvant pelvic radiotherapy in the treatment of uterine sarcomas stages I and II: an European Organisation for Research and Treatment of Cancer Gynaecological Cancer Group Study (protocol 55874). Eur J Cancer 2008;44:808–818.
12 Rose PG, Boutselis JG, Sachs L. Adjuvant therapy for stage I uterine sarcoma. Am J Obstet Gynecol 1987;156:660–662.
13 Sorbe B, Johansson B. Prophylactic pelvic irradiation as part of primary therapy in uterine sarcomas. Int J Oncol 2008;32:1111–1117.
14 Zivanovic O, Jacks LM, Iasonos A, *et al.* A nomogram to predict postresection 5-year overall survival for patients with uterine leiomyosarcoma. Cancer 2012;118:660–669.
15 Iasonos A, Keung EZ, Zivanovic O, *et al.* External validation of a prognostic nomogram for overall survival in women with uterine leiomyosarcoma. Cancer 2013;119:1816–1822.
16 Pautier P, Genestie C, Rey A, *et al.* Analysis of clinicopathologic prognostic factors for 157 uterine sarcomas and evaluation of a grading score validated for soft tissue sarcoma. Cancer 2000;88:1425–1431.
17 Croce S, Ribeiro A, Brulard C, *et al.* Uterine smooth muscle tumor analysis by comparative genomic hybridization: a useful diagnostic tool in challenging lesions. Mod Pathol 2015;28:1001–1010.
18 Oliva E. Practical issues in uterine pathology from banal to bewildering: the remarkable spectrum of smooth muscle neoplasia. Mod Pathol 2016;29(Suppl 1):S104–S120.
19 Lagarde P, Perot G, Kauffmann A, *et al.* Mitotic checkpoints and chromosome instability are strong predictors of clinical outcome in gastrointestinal stromal tumors. Clin Cancer Res 2012;18:826–838.
20 Mehine M, Kaasinen E, Makinen N, *et al.* Characterization of uterine leiomyomas by whole-genome sequencing. N Engl J Med 2013;369:43–53.
21 Mehine M, Kaasinen E, Aaltonen LA. Chromothripsis in uterine leiomyomas. N Engl J Med 2013;369:2160–2161.
22 Makinen N, Aavikko M, Heikkinen T, *et al.* Exome sequencing of uterine leiomyosarcomas identifies frequent mutations in TP53, ATRX, and MED12. PLoS Genet 2016;12:e1005850.
23 Hempling RE, Piver MS, Baker TR. Impact on progression-free survival of adjuvant cyclophosphamide, vincristine, doxorubicin (adriamycin), and dacarbazine (CYVADIC) chemotherapy for stage I uterine sarcoma. A prospective trial. Am J Clin Oncol 1995;18:282–286.
24 El-Rifai W, Sarlomo-Rikala M, Knuutila S, *et al.* DNA copy number changes in development and progression in leiomyosarcomas of soft tissues. Am J Pathol 1998;153:985–990.
25 Hu J, Khanna V, Jones M, *et al.* Genomic alterations in uterine leiomyosarcomas: potential markers for clinical diagnosis and prognosis. Genes Chromosomes Cancer 2001;31:117–124.
26 Larramendy ML, Kaur S, Svarvar C, *et al.* Gene copy number profiling of soft-tissue leiomyosarcomas by array-comparative genomic hybridization. Cancer Genet Cytogenet 2006;169:94–101.
27 Perot G, Derre J, Coindre JM, *et al.* Strong smooth muscle differentiation is dependent on myocardin gene amplification in most human retroperitoneal leiomyosarcomas. Cancer Res 2009;69:2269–2278.

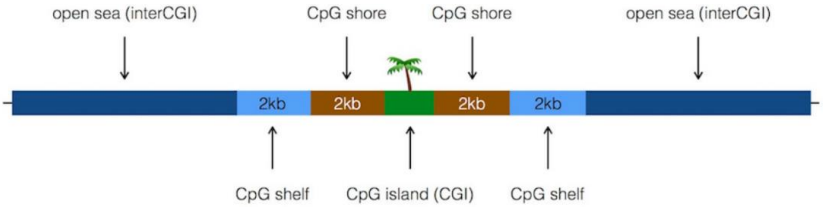
28 Agaram NP, Zhang L, LeLoarer F, *et al.* Targeted exome sequencing profiles genetic alterations in leiomyosarcoma. Genes Chromosomes Cancer 2016;55:124–130.
29 Kimura Y, Morita T, Hayashi K, *et al.* Myocardin functions as an effective inducer of growth arrest and differentiation in human uterine leiomyosarcoma cells. Cancer Res 2010;70:501–511.
30 Perot G, Mendiboure J, Brouste V, *et al.* Smooth muscle differentiation identifies two classes of poorly differentiated pleomorphic sarcomas with distinct outcome. Mod Pathol 2014;27:840–850.
31 Nordal RR, Kristensen GB, Kaern J, *et al.* The prognostic significance of stage, tumor size, cellular atypia and DNA ploidy in uterine leiomyosarcoma. Acta Oncol 1995;34:797–802.
32 Wang WL, Soslow R, Hensley M, *et al.* Histopathologic prognostic factors in stage I leiomyosarcoma of the uterus: a detailed analysis of 27 cases. Am J Surg Pathol 2011;35:522–529.
33 Pelmus M, Penault-Llorca F, Guillou L, *et al.* Prognostic factors in early-stage leiomyosarcoma of the uterus. Int J Gynecol Cancer 2009;19:385–390.
34 Davidson B, Kjaereng ML, Forsund M, *et al.* Progesterone receptor expression is an independent prognosticator in FIGO stage I uterine leiomyosarcoma. Am J Clin Pathol 2016;145:449–458.
35 Lim D, Alvarez T, Nucci MR, *et al.* Interobserver variability in the interpretation of tumor cell necrosis in uterine leiomyosarcoma. Am J Surg Pathol 2013;37:650–658.
36 Downes KA, Hart WR. Bizarre leiomyomas of the uterus: a comprehensive pathologic study of 24 cases with long-term follow-up. Am J Surg Pathol 1997;21:1261–1270.
37 Croce S, Young RH, Oliva E. Uterine leiomyomas with bizarre nuclei: a clinicopathologic study of 59 cases. Am J Surg Pathol 2014;38:1330–1339.
38 Liegl-Atzwanger B, Heitzer E, Flicker K, *et al.* Exploring chromosomal abnormalities and genetic changes in uterine smooth muscle tumors. Mod Pathol 2016;29:1262–1277.
39 Zhang Q, Ubago J, Li L, *et al.* Molecular analyses of 6 different types of uterine smooth muscle tumors: emphasis in atypical leiomyoma. Cancer 2014;120:3165–3177.

Supplementary Information accompanies the paper on Modern Pathology website (http://www.nature.com/modpathol)

II – Matériel et méthodes de la partie 3.1.4.2 (Exploration des données de méthylation)

- **Présentation des données et sélection des sondes**

Les données de méthylation provenant du projet TCGA sarcomes ont été produites en utilisant les puces Infinium 450K (Affymetrix) selon le protocole décrit dans l'article correspondant (The Cancer Genome Atlas Research Network, 2017). Les données regroupent 374931 sondes associées à des îlots CpG répartis sur l'ensemble du génome humain. Ces sondes peuvent être localisées à l'intérieur des îlots CpG ou en dehors. En dehors des îlots, la distance de la sonde par rapport son îlot définit plusieurs régions : à moins de 2k (régions *shore*), entre 2kb et 4kb (régions *shelf*) ou au-delà de 4kb (régions *open sea* ; figure ci-dessous reprise de Cavalcante & Sartor 2017).



Il a été démontré que la présence de polymorphismes (SNP) à proximité des sondes peut artificiellement modifier le signal capté par la puce et donc biaiser le niveau de méthylation qui y est associé (Chen *et al.*, 2013). Nous avons donc conservé, parmi l'ensemble des sondes initiales, 359279 sondes (95,83%) pour la suite des analyses. Cette sélection a été effectuée par la fonction *rmSNPandCH* implémentée dans le paquet DMRcate (Peters *et al.*, 2015).

- **Description et conversion entre B-*values* et M-*values***

Les données de méthylation du TCGA sarcomes sont fournies sous forme de β-*values* qui représentent le niveau de méthylation pour chaque sonde pour un échantillon. Une β-*value* varie de 0 à 1 où 0 indique qu'aucune copie d'ADN n'est sous forme méthylée tandis que 1 indique que toutes les copies d'ADN sont sous forme méthylée. Bien que l'interprétation de cette valeur soit simple, elle ne convient pas pour réaliser des analyses statistiques où les données doivent être converties en M-*value* (Du *et al.*, 2010). Les formules permettant de passer des β-*values* aux M-*values* sont indiquées ci-dessous.

$$\text{M-}value_i = \log_2\left(\frac{\beta\text{-}value_i}{1-\beta\text{-}value_i}\right) \; ; \; \beta\text{-}value_i = \frac{2^{\text{M-}value_i}}{2^{\text{M-}value_i}+1}$$

Cavalcante R. G. & Sartor M. A. (2017). *annotatr: genomic regions in context*. *Bioinformatics* 33:2381–2383.

Chen Y.-a., Lemire M., Choufani S., Butcher D. T., et al. (2013). *Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray*. *Epigenetics* 8:203–209.

Du P., Zhang X., Huang C.-C., Jafari N., et al. (2010). *Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis*. *BMC Bioinformatics* 11:587.

Peters T. J., Buckley M. J., Statham A. L., Pidsley R., et al. (2015). *De novo identification of differentially methylated regions in the human genome*. *Epigenetics Chromatin* 8:6.

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

1/1

III – L'hétérogénéité dans des lignées de sarcomes révèle une motilité accrue des cellules tétraploïdes par rapport aux diploïdes

www.impactjournals.com/oncotarget/ Oncotarget, 2017, Vol. 8, (No. 10), pp: 16669-16689

Research Paper

Heterogeneity in sarcoma cell lines reveals enhanced motility of tetraploid versus diploid cells

Mohamed Jemaà[1,2], Samer Abdallah[1,2], Gwendaline Lledo[1,2], Gaelle Perrot[3], Tom Lesluyes[3], Catherine Teyssier[4], Pierre Roux[1,2], Juliette van Dijk[1,2], Frederic Chibon[3,5], Ariane Abrieu[1,2], Nathalie Morin[1,2]

[1]Universités de Montpellier, 34293 Montpellier, France

[2]CRBM, CNRS, UMR 5237, 34293 Montpellier, France

[3]INSERM U1218, Bergonié Cancer Institute, F-33076 Bordeaux, France

[4]U1194 INSERM, IRCM, 34298 Montpellier, France

[5]Department of Pathology, Bergonié Cancer Institute, F-33076 Bordeaux, France

Correspondence to: *Abrieu Ariane,* ***email:*** *ariane.abrieu@crbm.cnrs.fr*
Morin Nathalie, ***email:*** *nathalie.morin@crbm.cnrs.fr*

Keywords: *sarcoma, CINSARC, mitosis, motility, diploid/tetraploid*

Received: *May 13, 2016* ***Accepted:*** *November 30, 2016* ***Published:*** *December 27, 2016*

ABSTRACT

Soft tissue sarcomas with complex genomics are very heterogeneous tumors lacking simple prognosis markers or targeted therapies. Overexpression of a subset of mitotic genes from a signature called CINSARC is of bad prognosis, but the significance of this signature remains elusive. Here we precisely measure the cell cycle and mitosis duration of sarcoma cell lines and we found that the mitotic gene products overexpression does not reflect variation in the time spent during mitosis or G2/M. We also found that the CINSARC cell lines, we studied, are composed of a mixture of aneuploid, diploid, and tetraploid cells that are highly motile *in vitro*. After sorting diploid and tetraploid cells, we showed that the tetraploid cell clones do not possess a proliferative advantage, but are strikingly more motile and invasive than their diploid counterparts. This is correlated with higher levels of mitotic proteins overexpression. Owing that mitotic proteins are almost systematically degraded at the end of mitosis, we propose that it is the abnormal activity of the mitotic proteins during interphase that boosts the sarcoma cells migratory properties by affecting their cytoskeleton. To test this hypothesis, we designed a screen for mitotic or cytoskeleton protein inhibitors affecting the sarcoma cell migration potential independently of cytotoxic activities. We found that inhibition of several mitotic kinases drastically impairs the CINSARC cell invasive and migratory properties. This finding could provide a handle by which to selectively inhibit the most invasive cells.

INTRODUCTION

Soft tissue sarcomas (STS) [1] are rare mesenchymal malignant tumors that can derive from different cell lineage and from almost any soft tissue in the body [2]. The genome of STS cells may present simple translocations or harbor complex chromosome rearrangements that add another level of complexity to the lineage heterogeneity. As a result, the classification of STS, based on tumor cell differentiation and histological grading, is very challenging. Indeed, although these tumors account for less than 1% of human cancers, there are over 75 STS groups [3, 4]. Tumor resection followed by radiotherapy is widely used for low graded and localized tumors, but 5 years overall survival remains very poor. Patients that develop local relapse or distant metastasis, benefit of palliative chemotherapy treatments with a very low response rate [5].

Improved classification and identification of markers started with the development of genomic analyses and genetic characterization of STS. As a result, a list of specific gene mutations, amplifications and chromosome rearrangements is rapidly growing for many types of soft tissue sarcoma tumors [6]. Functional studies aimed at

confirming altered pathways are developed with generation of sarcoma derived cell lines and *in vivo* models [7] and should allow the design of targeted therapies for patients developing STS tumors.

In this context, genomic and expression profiling of STS with highly rearranged genomes led to the identification of a validated 67 overexpressed genes signature called CINSARC for genome Complexity Index in SARComas. In STS and several other cancers, CINSARC predicts metastasis outcome with better confidence than the histological grading system developed by the French Federation of Cancer Centers Sarcoma Group (FNCLCC) [8, 9]. CINSARC is a molecular signature mostly composed of genes that encode cell cycle proteins controlling chromosome integrity and mitotic progression. Strikingly, most of the mitotic kinases and kinesins regulating the establishment and the evolution of the microtubule spindle during mitotic progression, required to faithfully segregate sister chromatids in daughter cells, are found among these 67 gene products.

Cell tetraploidization often occurs early in tumorigenesis [10] and may result from different insults to the cells such as telomere attrition [11, 12] or mitotic defects. For example, centriole amplification or overexpression of spindle assembly checkpoint proteins, but also cytokinesis failure or loss of AURKB dependent abscission checkpoint can all induce tetraploidization [12–20]. In normal cells, cytokinesis failure triggers activation of Hippo pathway, leading to TP53 dependent G1 arrest [1, 21–23]. However, when the *TP53*-dependent surveillance mechanism is lost, tetraploid cells will keep cycling, further promoting chromosome instability (CIN), aneuploidy and tumorigenesis [14, 18, 24]. To keep their proliferative advantage, derived cancer cells may adapt by inducing centrosomes clustering and/or silencing during mitosis [25–27].

Thus deregulation of mitotic genes observed in CINSARC positive tumors likely participates to the increasing complexity in chromosomes aberrations, rearrangements and instability observed in metastatic STS tumors.

Most intriguing, however, is the link that CINSARC signature makes between deregulation of mitosis and increased metastatic potential. As yet, only a few studies suggested that mitotic gene products could be linked to increased cell motility and invasive properties. Among them mitotic checkpoint proteins BUBR1 and Mad1 were proposed to regulate cellular motility through the regulation of MMP-2/9 metalloproteases and integrin secretion respectively [28, 29]. Most interestingly, identification of novel partners also directly linked mitotic CCNA2 and AURKA Kinase to the promotion of cell migration through RhoA and SSH1/CFL1 mediated regulation of actin dynamics [30, 31].

Here, we show that cell lines derived from CINSARC tumors acquired highly motile and invasive phenotypes. From these heterogeneous cell lines, we derived diploid and tetraploid clonal cell lines. RNA sequencing showed that clonal populations were positive for CINSARC signature and likely shared a same origin.

Relative to control fibroblast, overexpression of CINSARC transcripts was similar in all clones. In contrast, increase of most candidate mitotic gene products steady state levels was more elevated in tetraploid versus diploid cells**.** This did not confer significant proliferation advantage to tetraploid cells but strikingly increased their motility potential. In addition, we found that drug mediated inhibition of major mitotic kinases AURKA/B and TTK considerably reduced the migratory potential of all CINSARC clones.

RESULTS

CINSARC positive MFH137 and MFH152 sarcoma cell lines are hyper motile

We investigated the motile and invasive properties of CINSARC positive MFH152 and MFH137 sarcoma cell lines derived from metastatic tumors. Using a wound-healing assay to monitor cell migration, we found that MFH152 and MFH137 cells are extremely motile compared to non-transformed IMR90 fibroblast cells (Figure 1A and Supplementary Video 1). To avoid putative caveats from the wound procedure, we used size-controlled cell seeding stoppers from Oris™ and confirmed the high migratory potential of MFH137 and MFH152 cells (Figure 1B and 1C). Since these assays measure migration of grouped cells, we used fibronectin-coated narrow lines to monitor individual cells migration (Supplementary Video 2). As shown in Figure 1D, MFH137 and MFH152 still moved very fast, when forced to migrate individually. Finally, using a collagen layer-based 3D migration assay we found that MFH137 and MFH152 are highly invasive compared to untransformed fibroblasts (Figure 1E and 1F). In conclusion and as expected from the metastatic potential of the tumors they derive from, we found that MFH152 and MFH137 cells are highly motile and invasive *in vitro*.

Cell cycle analysis reveals the heterogeneity of MFH152 and MFH137 sarcoma cell lines

Since one striking feature of the MFH152 and MFH137 cells is to overexpress mRNAs involved in cell cycle regulation, we wondered whether it is correlated with cell cycle perturbations. To address this question, we constrained cells within fibronectin-coated circular areas and followed them by time-lapse microscopy over several divisions (72 hours, Supplementary Videos 3-5). The average cell cycle length was almost twice as long for MFH137 and MFH152 cells (32 hours) compared to IMR90 fibroblast cells (Figure 2A). In addition, cell cycle

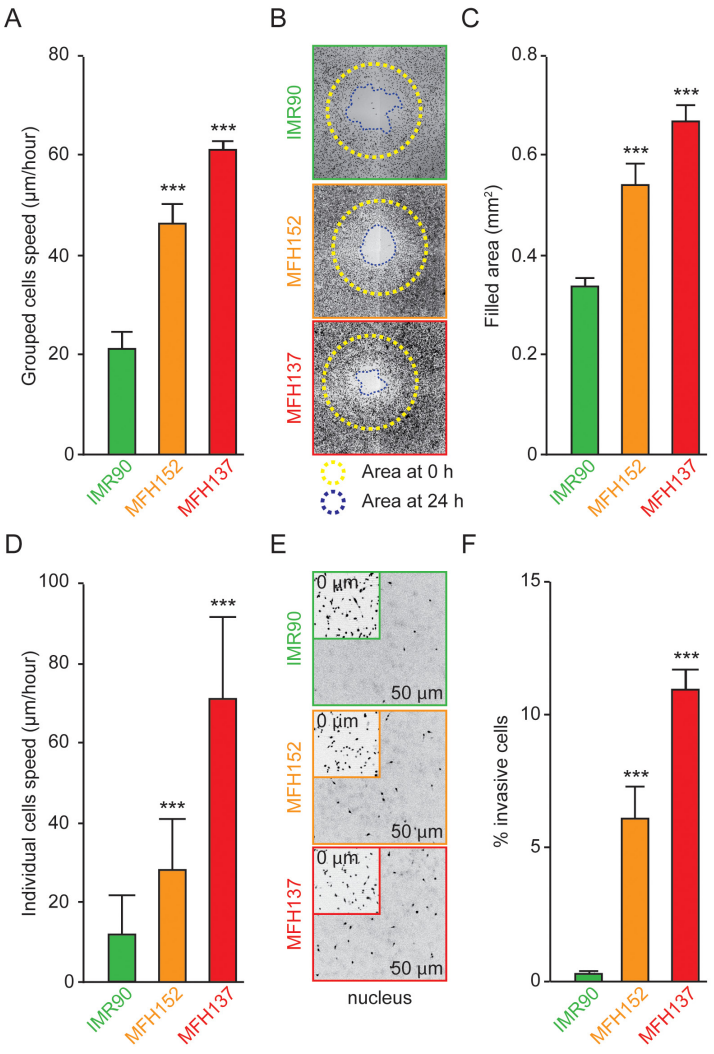


Figure 1: MFH137 and MFH152 sarcoma cell lines have a great invasive and motile potential. **(A)** Confluent IMR90 fibroblast or MFH152 or MFH137 sarcoma cell monolayers were scratched with pipet tips and imaged every hour to study collective migration. Quantitative histogram shows the average speed, calculated from the wound closure rate, using image J software. **(B)** and **(C)** two-dimensional migration assay using the Oris™ cell assay. Cells were allowed to migrate for 24h after the removal of cell seeding stoppers, fixed and stained with phalloidin to evaluate their motile potential. Representative photomicrographs are shown in panel **(B)**, the yellow broken lines show the empty migration zone at the start of the experiment while the blue lines delimit the cell-free area after 24h of migration. Quantitative data are presented in panel **(C)**. **(D)** Speed of individual IMR90, MFH152 and MFH137 cells was evaluated using the CytooChips Motility system to follow individual cell migration. **(E)** and **(F)** Invasive potential of IMR90, MFH152 and MFH137 cell lines measured after 24h. Representative micrographs show starting cell density at 0µm (**E,** insets) and invasive cells at 50 µm **(E)**. Quantitative data are represented in panel **(F)**. For all experiments, data are reported as means ± SEM ($n \geq 3$). ***$p < 0.001$ (ANOVA test), as compared with the human fibroblast IMR90 cell lines.

length was very heterogeneous between individual sarcoma cells lasting from 21 to 50 hours and from 22 to 46 hours respectively for MFH152 and MFH137 cells (Figure 2A). The cell heterogeneity was even more striking with mitosis duration that ranged from 24 min to 432 min (average 93 min) and from 16 to 368 minutes (average 70 min) for respectively MFH152 and MFH137 cells. By comparison, IMR90 spent 27 min average in mitosis (Figure 2B). Analyses of DNA content by flow cytometry showed that both MFH152 and MFH137 cell lines display an unusual proportion of cells containing at least 4n DNA content (Figure 2C–2D). To analyze whether the "shoulder" of the 4n DNA peak in MFH137 and MFH152 cells could represent G1 phase of a tetraploid subpopulation and/or aneuploid cell subpopulation, we performed metaphase chromosome spreads of these cells (Figure 2E). MFH152 and MFH137 cell lines contain three distinct cell populations with respectively 46, 92 or intermediate chromosomes number indicative of diploid, tetraploid and aneuploid cells (Figure 2F). Altogether our results suggest that the heterogeneity of the cell cycle in both sarcoma cell lines likely reflects the mixture of these cell lines subpopulations.

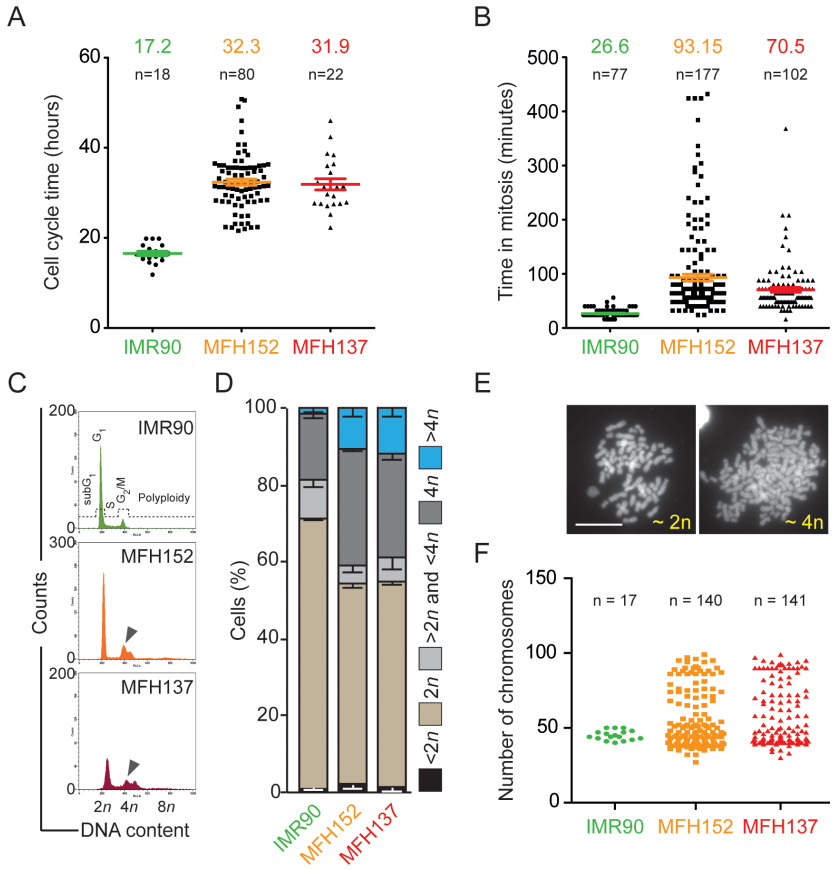


Figure 2: MFH137 and MFH152 sarcoma are a heterogeneous population. **(A)** and **(B)** quantitative analyses of cell cycle and mitosis length of IMR90, MFH152 and MFH137 cell lines following image acquisition of the cell lines for up to 72h. Panel **(A)** displays the quantitative data of cell cycle duration, while panel **(B)** shows the length of mitosis. **(C)** and **(D)** IMR90, MFH152 and MFH137 cells lines were fixed and stained with propidium iodide for cytofluorometric assessment of cell cycle progression. Cell cycle distribution analysed by flow cytometry panel **(C)** and quantitative data (means ± SEM; n = 3) of corresponding flow cytometry profiles panel **(D)** are shown. **(E)** and **(F)** Metaphase spread of IMR90, MFH152 and MFH137 cells lines. Panel **(E)** shows examples of DAPI-stained chromosomes in the metaphase spreads with the corresponding ploidy while panel **(F)** reports quantitative data.

Cell cycle profiles of diploid and tetraploid clones are similar but spindle assembly checkpoint is weakened in tetraploid clones

To be able to compare the behavior of diploid and tetraploid cells in the MFH137 and MFH152 sarcoma cell lines, we sorted them by flow cytometry using limiting dilution subcloning [32] and according to their size (Supplementary Figure 1A). A number of clones derived from single cells had a G1 profile with a 2n-4n intermediate content in DNA indicating strong chromosome instability (Supplementary Figure 1B). This aneuploid subpopulation was previously identified in the parental cell lines by metaphase chromosome spread (Figure 2F)

We chose to further analyze diploid and tetraploid clones that had DNA content profiles, in agreement with respective 2n/4n and 4n/8n populations containing less than 20% of cells in G2/M phase (Supplementary Figure 1B, Figure 3A–3B for MFH152, Supplementary Figure 2A-2B for MFH137).

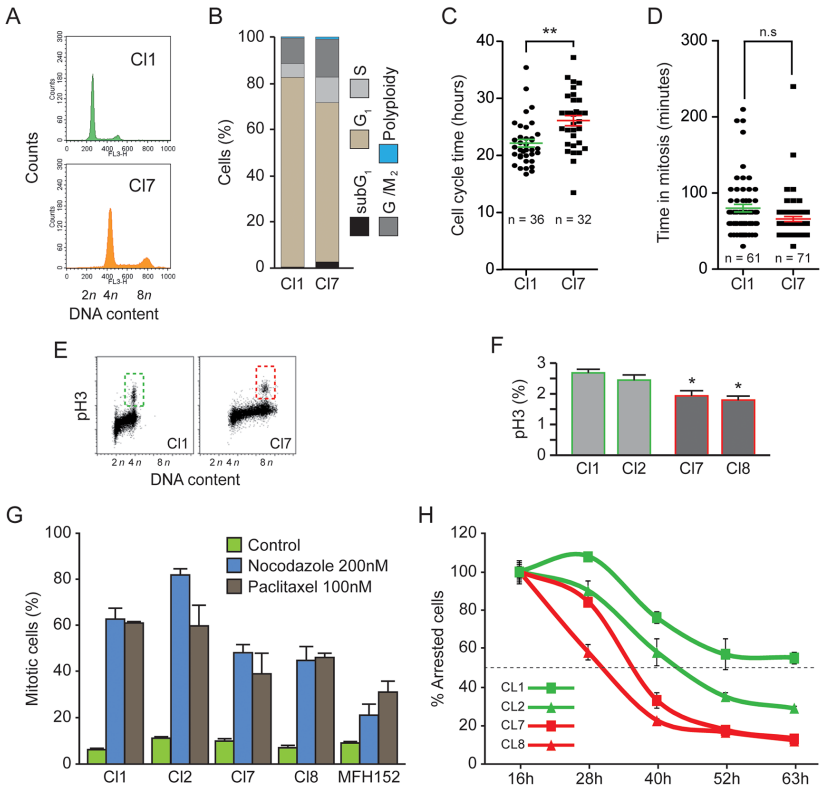


Figure 3: Cell cycle of diploid and tetraploid clones is similar but tetraploid clones have a weak spindle assembly checkpoint. **(A)** and **(B)** Diploid (Cl1) and tetraploid (Cl7) MFH152 clones were fixed and stained with propidium iodide for the cytofluorometric assessment of cell cycle progression. Cell cycle distribution analysed by flow cytometry is displayed in panel **(A).** Quantitative data of corresponding flow cytometry profiles are plotted in panel **(B)**. **(C)** and **(D)** Diploid (Cl1) and tetraploid (Cl7) MFH152 clones were imaged for up to 72h, by time lapse microscopy, to evaluate their respective cell cycle length and the time spent in mitosis. Panel **(C)** displays the quantitative data of cell cycle duration, while panel **(D)** shows the length of mitosis. **p < 0.05 (ANOVA test), n.s., not significant. **(E–F)** Fraction of G2/M cells present in two diploid and two tetraploid clones were analyzed by Flow cytometry profile **(E)** and quantified **(F)** following staining with antibodies directed against Ser10 Phospho-histone H3 (means ± SEM; n = 3). **(G–H)** Robustness of SAC in two diploid and two tetraploid clones. Percentage of the cell population arrested in mitosis following 16 hours exposure to carrier (control), nocodazole (200nM) or paclitaxel (100nM) **(G)** and percentage of the total number of arrested mitotic cells at 16 hours (100%) sustaining SAC over prolonged exposure to nocodazole (means ± SEM; n = 3).

Flow cytometry analyses at different cell passages revealed that the DNA content of some of these clones remained stable for three months (data not shown). We thus characterized their cell cycle and motility profiles at early passages. A more homogenous cell cycle length was observed in subclones compared to parental cell line and averaged 22/26 hours respectively for diploid/tetraploid clones derived from MFH152 (Figure 3C). A similar behavior was observed for MFH137 cells (Supplementary Figure 2C). The statistically significant small variation of cell cycle length between diploid and tetraploid cells could reflect a required lengthening of G1 to G2 phases to prepare larger cells for cell division.

Duration of mitosis was similar in diploid and tetraploid clones derived from MFH137 and MFH152 cell lines (Figure 3D and Supplementary Figure 2D). To better compare the G2/M cell populations in diploid and tetraploid clones derived from MFH152 cells, we next analyzed by flow cytometry, their Ser10 Phospho-histone H3 content (Figure 3E). Quantification of the flow cytometry profiles shows that the fraction of Ser10 Phospho-histone H3 positive cells is largely similar in tetraploid and diploid clones (Figure 3F).

Cell tetraploidization is often linked to the weakening of the mitotic spindle assembly checkpoint (SAC). We thus investigated the strength of SAC in diploid and tetraploid clones derived from MFH152 cells. Nocodazole and taxol drugs interfere with microtubule dynamic and induce, in exposed cells, activation of the SAC and prometaphase arrest. Over a large range of drug concentrations, parental MFH152 cell line could not efficiently be synchronized likely reflecting the lack of SAC in the aneuploid cell subpopulation (data not shown). After 16 hours nocodazole or taxol drug exposure, diploid clones were more efficiently arrested (60-80%) than tetraploid clones (40-50%) and parental MFH152 cells (20-30%) suggesting that diploid cells respond better than tetraploid cells to checkpoint conditions (Figure 3G).

Next, to analyze, how spindle assembly checkpoint is sustained in arrested cells. The decay of all nocodazole arrested mitotic cells measured after 16 hours exposure (Figure 3G) was monitored over longer time of drug exposure. As observed (Figure 3H), tetraploid cells escaped faster (50% escape at respectively 30 and 36 hours for clone 8 and 7) than diploid cells (50% escape at respectively 46 and 63 hours for clone 2 and1) confirming the weakness of the SAC in tetraploid clones.

We conclude that the dispersion, we observed for mitosis length and cell cycle duration in parental cell lines, likely reflected the behavior of their aneuploid cell subpopulations. In contrast, MFH152 derived tetraploid and diploid clones have homogeneous and comparable cell cycle profiles with similar cell fraction in G2/M phases. Although the spindle assembly checkpoint is more robust in diploid than in tetraploid clones, the latter do not have significant proliferative advantage over the diploid cells.

Sorting of diploid and tetraploid clones reveals that the tetraploid cells are more motile than their diploid counterparts

We next compared motile properties of diploid and tetraploid cells and found that tetraploid MFH152 cells migrate more efficiently than diploid cells (Figure 4A-4D).

First, motility assays were performed with confluent cells using cells seeding stoppers, as in Figure 1B. To avoid artifactual quantification related to diploid and tetraploid cell size difference, we counted the number of cells that penetrated inside the migration zone after 24 hours and found that roughly twice as many tetraploid cells colonized the empty space compared to diploid cells (Figure 4A-4B). The same assays performed on 2n and 4n MFH137 clones (Supplementary Figure 2E-2F) confirmed that motility is greatly enhanced in tetraploid compared to diploid sarcoma subclones.

Migrating cells may behave differently during collective cell migration, when cell/cell junctions play important functions, or as individualized cells. We therefore followed individual cells motility, of diploid and tetraploid clones derived from MFH152 cells, over a 20 hours time range. As for collective cell migration, we found that individualized tetraploid cells were significantly more motile than diploid cells (Figure 4C–4D).

We next sought to investigate and compare the invasive properties of MFH152 derived diploid and tetraploid clones. Cell Index Invasion was determined between 10 and 20 hours after seeding the cells. We found that tetraploid/diploid clones respectively display the most/the less invasive behavior while the parental MFH152 cells have an intermediate invasive behavior (Figure 4E).

Altogether, our results show that compared to diploid, tetraploid clones developed a more aggressive behavior characterized by increased motility and invasiveness.

Diploid and tetraploid subclones are closely related and overexpress mitotic gene products

Many mitotic gene transcripts are deregulated in CINSARC positive cells [8] and abnormal mitotic figures are observed in these cells. As an example, strong overexpression of kinesin KIF11 is detected on metaphase spindle poles and in membrane blebs in MFH137 and MFH152 parental cell lines (Figure 5A).

Since diploid and tetraploid cell subclones derive from the same CINSARC positive tumor, we wondered whether their different motile behavior solely results from the specific properties of the cells they derive from inside the heterogeneous resected tumor.

To test this hypothesis, we performed transcriptome analyses of two diploid and two tetraploid MFH152 clones, together with IMR90 control fibroblast by RNA

sequencing. Clustering results show that the two diploid clones together and the two tetraploid clones together were very similar with r=0.9901 and r=0.9939 respectively. A strong similarity was again observed between 2n and 4n clones (r=0.9701) while respectively more distant from the IMR90 (r=0.9114 and r=0.9046). In summary, transcriptomic profiles of 2n and 4n subclones are very closely related, as shown by the agglomerative coefficient that measures the clustering structure of the data set. Thus, they could be of same origin although this analysis does not definitely prove that they derive from each other (Figure 5B).

Since, mRNAs deregulation of CINSARC mitotic genes is conserved in derived diploid and tetraploid MFH152 subclones (Table 1 and Supplementary Figure 3) that are closely related but have different motility potential, we wondered whether representativity of mitotic gene products might be different in diploid and tetraploid subclones.

We analyzed, by western blot, the steady state level of several candidate mitotic and non-mitotic gene products in diploid and tetraploid MFH152 clones. We found that most mitotic proteins tested were only modestly overexpressed in diploid MFH152 clones compared to

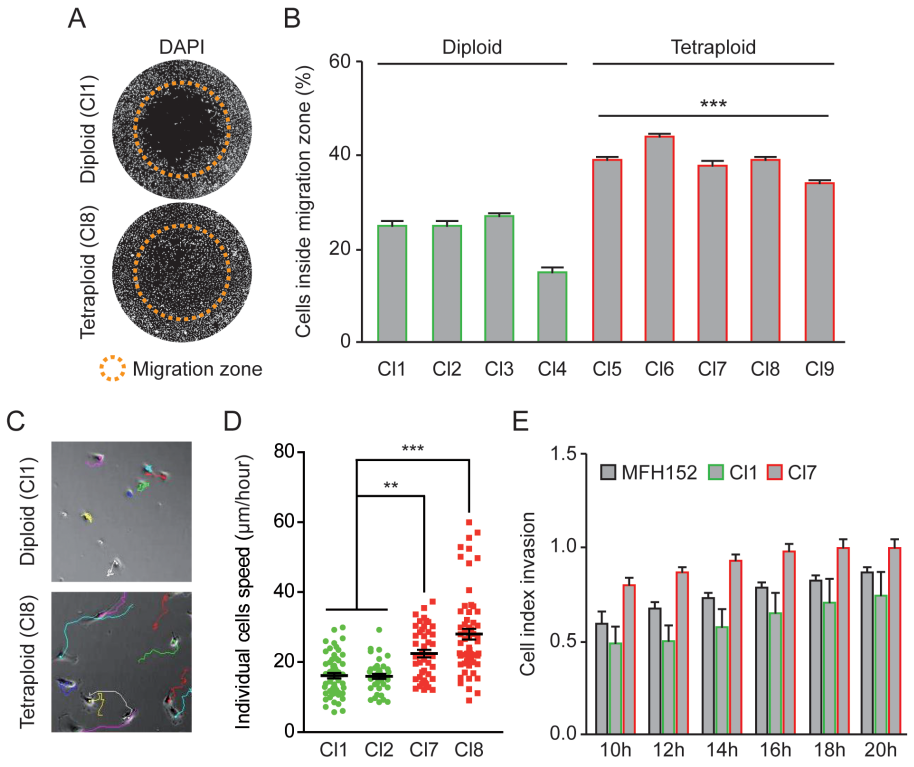


Figure 4: Tetraploid clones are more motile and invasive than diploid clones. (A-B) Four diploid (Cl1-4) and five tetraploid (Cl5-9) MFH152 clones were plated using the Oris™ cell migration assay and grown for 24h before fixation. DAPI staining was used to evaluate the cells migratory potential. Representative micrographs are shown in panel **(A)** while panel **(B)** shows the percentage of nuclei inside the migration zone (means ± SEM; n = 3), ***p < 0.001 (ANOVA test, each tetraploid over each diploid). In panel **(A)**, orange circles show the migration zone at the start of the experiment (0h). **(C-D)** Two diploid (Cl 1-2) and two tetraploid (Cl 7-8) MFH152 clones were grown in non-confluent conditions and imaged for 24 hours, using time-lapse microscopy, to evaluate respective migration potential of individual cells. Panel **(C)** shows representative micrographs of the trajectories reconstituted using imageJ software of some cells from diploid (Cl 1) and tetraploid (Cl 8) clones, while panel **(D)** shows the individual cells speed (means ± SEM; n = 3), ***p < 0.001, **p<0.05 (ANOVA test). **(E)** Invasive behavior of cells from a diploid (Cl 1), a tetraploid (Cl 7) clones and MFH152 mother cell line was followed between 10 and 20 hours after seeding using the *xCELLigence* RTCA DP instrument (Ozyme, France) (means ± SEM; n = 4).

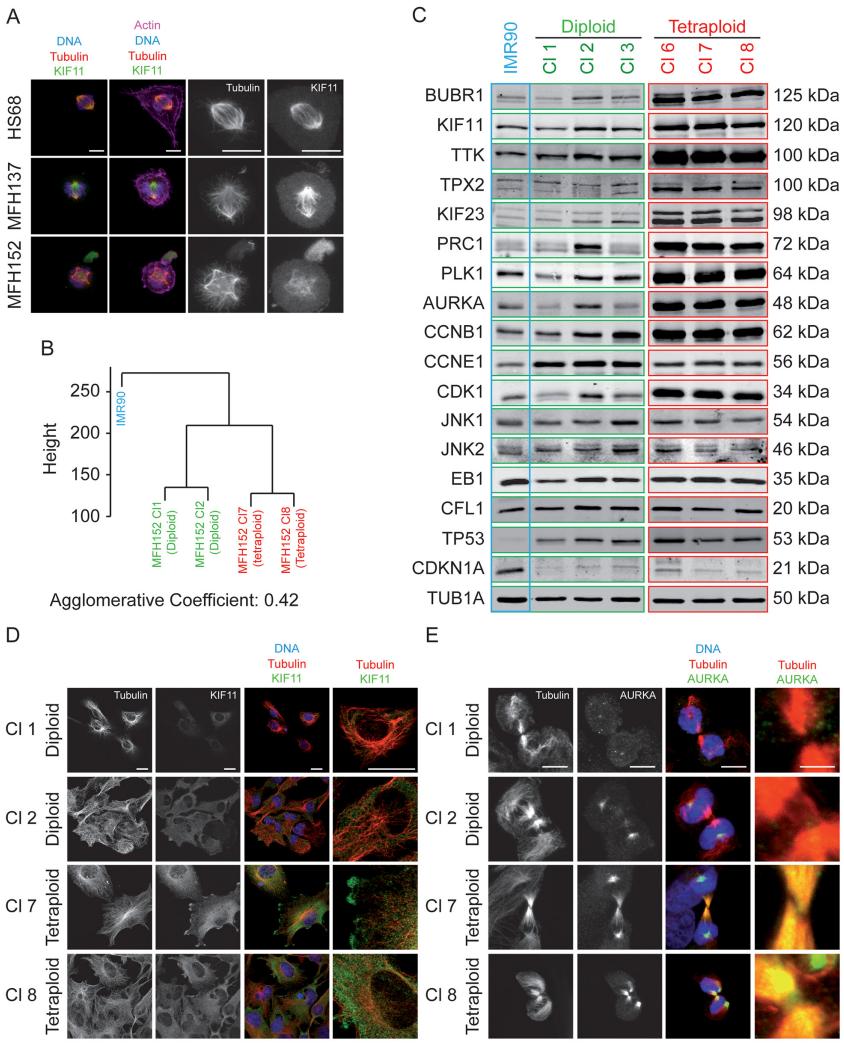


Figure 5: Tetraploid clones overexpress numerous proteins implicated in mitotic control and chromosome integrity. (A) Metaphases of HS68 fibroblast, MFH137 and MFH152 parental sarcoma cells plated on L shape CYTOOchips™ coated with fibronectin and stained for indicated antibodies and for actin (phalloidin). Phalloidin staining shows that mitotic sarcoma cells have multiple blebs and do not adhere properly to the fibronectin coated L shape. Kinesin KIF11 is overexpressed and mislocalized in blebs. These mitotic figures are representative examples of the numerous mitotic defects observed in sarcoma cell lines (n>5). Detail of images of tubulin and KIF11 stainings are shown on the right. Bar is 10 µm. **(B)** Clustering results of IMR90 together with MFH152 diploid clones 1 and 2 together and tetraploid clones 7 and 8. **(C)** Three diploid (Cl 1-3) and three tetraploid (Cl 6-8) MFH152 clones (framed in green and red, respectively) and the IMR90 cell line (framed in blue) were collected for western-blot analysis of the indicated proteins involved in cytoskeleton, cell cycle and mitotic regulation. **(D-E)** Immunofluorescence of interphasic **(D)** or mitotic **(E)** cells from diploid (Cl 1-2) and tetraploid (Cl 7-8) MFH152 subclones were stained for tubulin together with KIF11 **(D)** or AURKA **(E)**. KIF11 is overexpressed in tetraploid clones. Detail of merge images between tubulin and KIF11 is shown on the right, Bars: 20 µm **(D)**. AURKA is overexpressed in midzone region of tetraploid clones. Detail of merge images between tubulin and AURKA is shown on the right, Bars: 15 µm **(E)**.

Table 1: CINSARC genes RNA Seq data for MFH152 diploid and tetraploid clones normalized to IMR90 expression levels

		MFH152			
		Diploid		Tetraploid	
GeneSymbol	IMR90	Cl 1	Cl 2	Cl 7	Cl 8
ANLN	1.0	3.0	2.9	3.1	3.2
ASPM	1.0	2.4	2.5	2.1	2.0
AURKA	1.0	1.8	2.1	1.6	1.5
AURKB	1.0	3.7	4.4	3.0	3.1
BIRC5	1.0	3.0	3.1	2.0	1.7
BUB1	1.0	3.6	3.2	2.9	3.0
BUB1B	1.0	1.8	2.2	1.7	1.7
CCNA2	1.0	1.5	1.8	1.2	1.0
CCNB1	1.0	1.9	2.4	1.8	1.3
CCNB2	1.0	2.2	3.0	2.3	1.9
CDC20	1.0	1.9	2.1	2.7	2.2
CDC6	1.0	1.7	1.6	1.5	1.7
CDC7	1.0	4.2	5.3	3.2	3.8
CDCA2	1.0	1.5	1.8	2.0	1.8
CDCA3	1.0	2.4	2.5	2.0	1.4
CDCA8	1.0	1.8	2.1	2.6	2.5
CENPA	1.0	2.9	3.5	2.4	1.8
CENPE/KIF10	1.0	2.2	2.2	1.2	1.0
CENPL	1.0	2.1	1.9	1.8	2.1
CEP55	1.0	2.9	2.6	2.2	2.0
CHEK1	1.0	1.4	1.5	1.3	0.8
CKS2	1.0	1.6	1.9	1.5	1.2
ECT2	1.0	2.6	2.9	2.7	2.3
ESPL1	1.0	2.3	2.4	2.4	2.3
FBXO5	1.0	2.3	2.5	2.5	2.4
FOXM1	1.0	2.7	3.2	2.8	2.2
H2AFX	1.0	1.7	2.0	1.2	0.8
HP1BP3	1.0	0.6	0.6	0.9	0.7
FANCI	1.0	3.0	3.4	4.2	4.6
Eg5/KIF11	1.0	2.2	2.1	1.9	1.7
KIF14	1.0	2.3	2.3	2.6	2.5
KIF15	1.0	1.6	2.0	1.6	1.7
KIF18A	1.0	2.7	2.8	1.7	1.9
KIF20A	1.0	1.0	1.4	1.5	1.1
KIF23	1.0	2.6	2.8	3.1	3.0

(*Continued*)

		MFH152			
		Diploid		Tetraploid	
GeneSymbol	IMR90	Cl 1	Cl 2	Cl 7	Cl 8
KIF2C	1.0	2.3	2.9	3.9	3.7
KIF4A	1.0	3.2	4.0	3.5	2.4
KIFC1	1.0	1.8	2.0	2.4	2.2
MAD2L1	1.0	2.0	2.2	1.4	1.2
MCM2	1.0	3.1	3.4	2.8	2.6
MCM7	1.0	2.7	2.7	3.0	3.0
MELK	1.0	3.1	3.7	3.2	3.2
NCAPH	1.0	4.6	3.5	4.2	4.8
NDE1	1.0	0.9	1.1	1.1	0.7
NEK2	1.0	2.6	3.2	2.8	2.3
NUF2	1.0	2.2	2.8	2.4	2.6
OIP5	1.0	3.9	4.4	3.5	2.8
PBK	1.0	2.4	2.9	3.3	3.1
PLK4	1.0	1.8	1.9	1.4	1.4
PRC1	1.0	2.0	2.1	2.2	1.9
PTTG1	1.0	6.4	7.5	7.3	5.1
RAD51AP1	1.0	3.4	3.7	3.3	3.6
RNASEH2A	1.0	2.9	4.5	3.1	3.0
RRM2	1.0	1.3	1.4	1.5	1.4
SGOL2	1.0	3.0	2.9	2.8	2.8
SMC2	1.0	2.0	2.0	1.9	2.1
SPAG5	1.0	1.9	2.0	1.4	1.5
SPC25	1.0	3.0	2.6	2.6	2.2
TOP2A	1.0	1.4	1.6	1.0	0.9
TPX2	1.0	4.2	4.5	4.0	3.5
TRIO	1.0	2.2	1.6	2.0	1.8
TRIP13	1.0	1.7	1.7	1.4	1.7
TTK	1.0	2.7	3.0	3.3	3.2
UBE2C	1.0	3.0	3.7	3.1	3.0
ZWINT	1.0	3.4	3.0	3.3	3.3

IMR90 cells. In contrast, overexpression of CCNB1, along with its associated kinase CDK1, other major mitotic kinases AURKA, BUBR1, PLK1, and TTK, the MAPs and kinesins PRC1, TPX2, KIF11, KIF23, were much increased in tetraploid than in diploid cell lines while non-CINSARC protein EB1, CFL1, JNK1, and JNK2 steady state levels were similar (Figure 5C). On the contrary, CCNE1, a G1/S gene product was mainly overexpressed in diploid cells. We also noted a strong overexpression of TP53 in both diploid and tetraploid clones. This is consistent with the fact that TP53 is mutated in MFH152 and MFH137 cell lines. Although stabilized, the mutant TP53 protein remains unable to promote CDKN1A expression, in these cells. In contrast to IMR90 fibroblast cells which have a low level of TP53 and high CDKN1A steady state level.

The differences in the steady state level of mitotic proteins, in diploid and tetraploid clones, unlikely results of cell cycle difference since diploid and tetraploid cells mitotic index are similar.

Interestingly, at the cellular level, we also found that kinesin KIF11 is more overexpressed in the cytoplasm of interphase tetraploid than in diploid MFH152 subclones (Figure 5D). Similarly, the interphase AURKA kinase centrosomal staining is more prominent in tetraploid than diploid cells (not shown), but more strikingly AURKA strongly stained the midzone and midbody region of mitotic tetraploid cells only (Figure 5E). Such a staining is similar to the one expected for Aurora B kinase and suggest that APC mediated proteolysis of AURKA may be partially defective in tetraploid clones.

Misexpression and overexpression of microtubule binding proteins together with regulatory kinases can result in profound microtubule and actin cytoskeleton rearrangements. Since actin and microtubule networks regulate the migratory capacities of cells, we asked whether overexpression of mitotic gene products confers to CINSARC positive cells their increased motility potential.

Reversine and SP600125 inhibit sarcoma cell migration

To address this question, we developed a screen using 22 mitosis and cytoskeleton inhibitors at 3 different doses (0.1; 1 and 10 uM) for 24 hours (Supplementary Table 1). The initial screen was based on our assay of inhibition of cell migration using cell seeding stoppers. The first readout was the consequence of these treatments on cell death and the second was on cell migration. Cytotoxic effect inducing cell death was assessed by the ratio of counted nuclei (inside + outside migration zone) in the treated versus non-treated cells. We did not consider further, drugs that display a cytotoxic effect higher than 30% (Supplementary Table 2). Following the first round of analyses several drugs, with a too high cytotoxic effect, such as the PLK1 and CDK1 inhibitors, the CENPE and KIF11 kinesin inhibitors, the proteasome inhibitors and the cytoskeleton poisons were eliminated.

We then narrowed the screen to the drugs that passed the cytotoxicity test. We further tested the efficacy of these drugs at their non-toxic concentration to inhibit cell motility in chosen diploid and tetraploid MH152 subclones. We used the cell migration assay we developed, and counted the number of cells entering the migration zone. Table 2 summarizes the results, we obtained. We decided to consider 40% of migration inhibition, as a cut off, to consider the drug efficient in inhibiting the cell motility. Most of the kinesins, mitotic kinases and phosphatase inhibitors that proved to be very toxic at high doses, were only used at low concentrations with no noticeable effects on cell migration capacity (GSK 923295, Dimethylenastron, ZM 447439, RO3306, CDKi, Roscovitin, CDC25).

Finally only the two drugs, Reversine and SP600125, met the criteria of conjugating low toxicity with significant inhibition of diploid and tetraploid sarcoma cells migration. These drugs target major mitotic kinases. Specificity of Reversine is best toward TTK (IC_{50} 6 nM) and good toward AURKB (IC_{50} of 98.5 nM) while SP600125 inhibits AURKA and AURKB (IC_{50} 60nM and 190nM) but also JNK (IC_{50} 100-200nM) and TTK (IC_{50} 1,9 μM).

We focused our study on these promising two molecules, in order to better characterize their efficacy toward inhibition of migration. Assays were enlarged to 3 diploid sub-clones and 3 tetraploid sub-clones. We first and again evaluated the toxicity of the drugs using cytofluorometry-based assays. Measures of the cell death–associated parameters, vital dye propidium iodide (PI) and mitochondrial membrane potential (Δψm)-sensing dye $DiOC_6$ [16]) confirmed the non toxic effect of both molecules at 24h on all sub-clones (Figure 6A). We next repeated the two-dimensional cell migration assays.

Both Reversine and SP600125 inhibited the migration of all sub-clones. Strikingly, the residual migration of the highly motile tetraploid subclones was similar to the diploid clones. These results show that tetraploid cells, that overexpress CINSARC mitotic gene products, were more sensitive to Reversine and SP600125 treatments (Figure 6B).

Next, to further study the effect of reversine and SP600125 on cell invasion, we used 3D MultiCellular Tumor Spheroid Assay (MCTS) that mimics several features of tumor microenvironment and privilege cell-cell adhesion over cell/extracellular matrix interactions [33]. Diploid cells efficiently formed regularly shaped spheroid after 24 hours culture. In contrast, even after 48-72 hours of culture, tetraploid cells only formed small cell clumps that broke easily and did not properly aggregate (data not shown), behaving as a number of tumor cell lines described in an extensive survey [33].

Since tetraploid MFH152 cells were unable to form spheroids, we concentrated on diploid MFH152 and initiated the 3D tumor invasion assay by transferring the spheroids in collagen-based medium and at low serum concentration (0.1%), in order to minimize cell proliferation. Using time-lapse imaging, we observed a remarkable induction of cell migration from the tight untreated MFH152 diploid spheroids embedded in the collagen matrix. Indeed, the spheroid volume rapidly expanded in collagen matrix, as invading cells that induced matrix proteolysis/degradation colonize the newly emptied space (Figure 7A and Supplementary Video 6). In contrast, when spheroids were incubated in collagen-based medium containing either reversine or SP600125, a complete inhibition of invasion was observed (Figure 7A and Supplementary Video 7-8) and quantified (Figure 7B).

Table 2: Migration inhibition ratio for MFH152 diploid and tetraploid clones upon incubation with three doses of the indicated drugs

		Migration inhibition					
		Cell counts					
		0.1 µM		1 µM		10 µM	
		Diploid	Tetraploid	Diploid	Tetraploid	Diploid	Tetraploid
1	**AZ 3146**	0.00	0.00	0.12	0.00	N.R.	0.25
2	**Reversine**	0.04	0.00	0.08	0.00	0.40	0.47
3	**SP600125**	0.00	0.00	0.12	0.00	0.44	0.47
4	**ZM 447439**	0.00	0.10	N.R.	0.01	N.R.	N.R.
5	**Dimethylenastron**	0.00	0.03	N.R.	N.R.	N.R.	N.R.
6	**GSK 923295**	0.00	0.05	N.R.	N.R.	N.R.	N.R.
7	**SB203508**	0.00	0.03	0.00	0.01	0.01	0.12
8	**RO 3306**	0.18	0.00	0.01	0.00	N.R	N.R
9	**Cdk1 Inhibitor III**	0.00	0.05	0.00	0.00	N.R	N.R
10	**Roscovitine**	0.00	0.13	0.00	N.R	N.R	N.R
11	**NSC 95397**	0.12	0.08	0.15	0.11	N.R	N.R
12	**IPA3**	0.00	0.00	0.00	0.00	0.23	0.02
13	**Y-27632**	0.00	0.01	0.04	0.09	0.04	0.12
14	**ITX-3**	0.08	0.01	0.00	0.01	0.09	0.11
15	**Blebbistatin**	0.06	0.06	0.06	0.00	0.00	0.00

Green labeling indicates a migration inhibition ratio above 0.40.

This drug-mediated inhibition of invasion cannot be attributed to mitotic arrest since we did not observe any spheroids shape changes/expansion following drug treatment, even at very early time points (Supplementary Videos 6-8).

The combination of several CINSARC mitotic gene products is likely required to upregulate tetraploid sarcoma cells migration

Reversine and SP600125 target several major kinases. In order to identify whether one of these kinases could specifically regulate sarcoma cells migration. We focused on AURKA, AURKB and TTK kinases that are targeted by reversine and SP600125 and also chose KIF11 that strongly stains the lamellipodia of interphase tetraploid cells (Figure 5D). We used siRNAs directed against AURKA, AURKB, TTK and KIF11 and highly specific drugs commercially available (Supplementary Table 2). MLN8237 (Alisertib), a selective AURKA inhibitor (IC50 of 1.2nM in cell free assay, with >200 fold higher selectivity for AURKA than AURKB); AZD1152-HQPA (Baraserib) that is highly selective against AURKB (IC 0.37nM is cell free assay, with >3700 fold more selective for AURKB than AURKA); SB743921, a KIF11 inhibitor (K_i of 0.1 nM, with almost no affinity to other kinesins); Mps-BAY2a, a TTK inhibitor (IC50 of 1nM).

Consequences of siRNAs and drug treatments on cell migration potential were tested with the cell seeding stoppers assay (Oris ™). Almost complete protein depletion was observed with 5nM siRNA treatment, none of the siRNAs displayed a cytoxic effect higher than 30% when used at 100nM (Supplementary Figure 4A). However, when tested at 5 and 100nM, none of the siRNA treatment induced significant effect on either diploid or tetraploid cell migration capacity (Supplementary Figure 4B).

To confirm these results, we used the specific inhibitors described above. First, diploid (clone 1) and tetraploid cells (Cl7) were exposed for 24 hours to a range of drug concentration to determine by flow cytometry their cell cycle profile (Supplementary Figure 5A-5B). Deregulation of AURKA kinase induces centrosomal abnormalities, defective chromosome segregation followed by aneuploidy and/or cell death. MLN8237 (AURKAi) inhibits AURKA at low nanomolar range (10nM). But, as reported [34], MLN8237 concentration above 100nM induced AURKB inhibition dependent cell polyploidization. KIF11 is required for bipolar spindle assembly and KIF11 defects result in monopolar

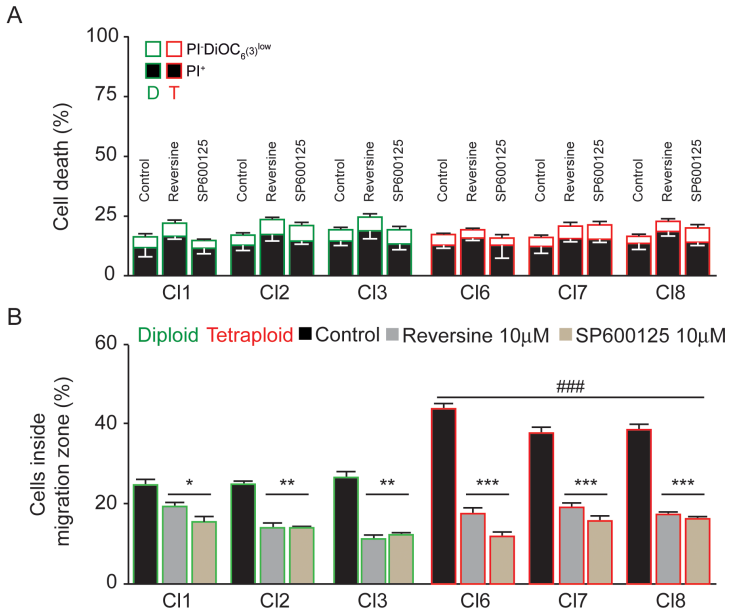


Figure 6: Reversine and SP600125 inhibit cell migration. (A) Toxicity induced in diploid and tetraploid subclones following 24h treatment with either reversine or SP600125 was assessed by flow cytometry upon co-staining with the cell death–associated parameters dyes, propidium iodide (PI) and DiOC6 [72] (means ± SEM; n = 3). **(B)** Three diploid (Cl 1-3) and three tetraploid (Cl 6-8) MFH152 were plated using the Oris™ cell migration assay. Cells were grown with or without Reversine or SP600125 at 10 µM for 24h, fixed and stained with DAPI to measure their migratory potential. Quantitative data shows the percentage of cells (DAPI staining) inside the migration zone (means ± SEM; n = 3). *($p<0.05$), **($p<0.01$),***($p<0.001$) indicates significant difference from the absence of both Reversine or SP600125 treatment while ### ($p<0.001$) indicates significant difference of the migration inhibition between all tetraploid clones taken one by one *vs* all diploid clones taken one by one (ANOVA test).

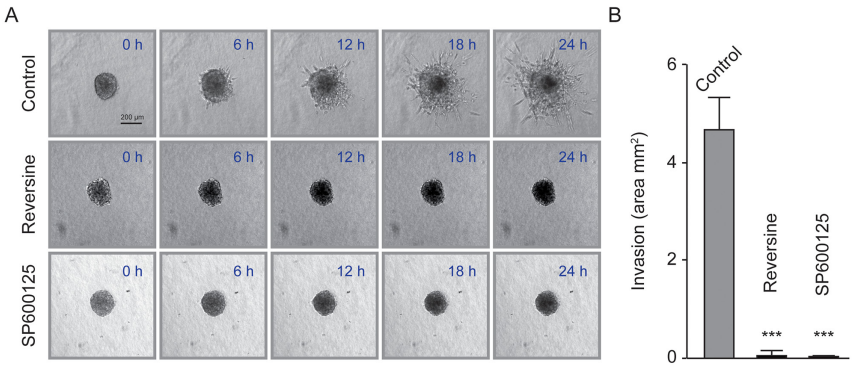


Figure 7: Reversine and SP600125 inhibit diploid MFH152 clone invasion in 3D. (A) and **(B)** A diploid MFH152 clone was used to generate spheroids. Spheroids were transferred in collagen matrix (Panel **A**, 0 h) to allow invasion. Drug dilution buffer (Control) or Reversine or SP600125 were added to the collagen matrix and spheroids were imaged using time-lapse microscopy for up to 24 h to evaluate the drug effects on invasion. Panel **(A)** displays the fate of a single spheroid in the control and treatment conditions, at 0, 6, 12, 18 h while quantitative data of invasion are represented in panel **(B)**.

spindle formation and mitotic arrest [35], a phenotype we observed with SB743921 concentrations, as low as 10nM. Cytokinesis defect and polyploidization resulting from AURKB dysregulation [36] was induced by 20nM AZD1152-HQPA treatment. We previously reported that Mps-BAY2a mediated TTK inhibition prevents SAC activation and ultimately results in aneuploidization/polyploidization [37]. Here, cell cycle perturbations were induced with Mps-BAY2a concentration as low as 100nM.

Cell cytotoxicity remained below 30% both for diploid and tetraploid clones using these defined low but effective inhibitor concentrations (Supplementary Figure 5C-5F). Nevertheless, when used at similar concentrations, none of these drugs had a significant inhibitory effect on either diploid or tetraploid cell motility (Supplementary Figure 5G-5H).

In summary, interference with chosen single mitotic gene products is not sufficient to inhibit cell migratory activity. We suggest that it is the combination of several deregulated mitotic gene products of CINSARC signature that is responsible for the increased migratory capacity of tetraploid clones. Inhibition of cell invasiveness and motility mediated by SP60025 and reversine that have broader effects on several mitotic gene products further support this hypothesis.

DISCUSSION

Mitosis is a highly regulated process that leads, in most cells, to the equal partitioning of chromosomes and cytoplasmic material in the two daughter cells. In normal tissues, protective mechanisms insure that tetraploidization does not happen [1, 19, 38] and if an accident occurs the resulting tetraploid cells are rapidly eliminated [39, 40]. Nonetheless, cycling tetraploid cells with supernumerary centrosomes and aberrant mitoses are observed at very early stages of tumorigenesis [10, 41–43] and surviving tetraploid cells were proposed to trigger chromosome instability (CIN) and aneuploidy in absence of *TP53*-mediated G1 checkpoint [12, 18, 24, 44, 45]. Indeed, tetraploidy associates with increased DNA damage, abnormal bipolar/multipolar mitoses, chromosome breaks and/or missegregration [10, 12, 20, 46, 47].

But, in contrast to functions in tumorigenesis, few data exist concerning a potential involvement of tetraploidization in the promotion of cell migration and/or invasion [18].

In tumors, cells that become invasive must detach from other cells and degrade the extracellular matrix through which they will migrate. Invasive behavior requires profound changes of cell shape, adhesion and mechanical properties that depend upon dynamic changes of actin, microtubule but also intermediate filament networks [48–51].

In this study, we showed that cell lines derived from CINSARC positive soft tissue sarcoma have, as expected, highly motile profiles and enhanced capacities to migrate compared to non-transformed fibroblasts. Because of the heterogeneity of the cell lines derived from these tumors, we isolated cellular clones that proved to be aneuploid, for most. We were, however, capable to isolate and grow diploid and tetraploid clones that remain stable over a three months period. Unexpectedly, the motile and invasive behavior of tetraploid clones was greatly enhanced compared to diploid cell clones. RNA sequencing indicated that they likely share the same origin and kept the CINSARC signature with similar levels of deregulated mitotic gene transcripts. Nonetheless, the overexpression of candidate mitotic gene products, i.e. proteins, was greatly enhanced in the tetraploid population. Since mitotic indexes were similar in diploid and tetraploid cells, we assume that tetraploid dependent increased steady state level of mitotic proteins depends upon a regulation that is beyond the transcriptional level, and acquired with tetraploidization. Degradation of many mitotic proteins occurs at the end of mitosis and is required either to exit mitosis or to prevent inappropriate functioning in interphase [30, 52–58]. Although, we do not know how mitotic proteins are stabilized in tetraploids [59], it is tempting to speculate that ubiquitin-mediated proteasome degradation of proteins at the end of mitosis could be partly defective. Indeed, we observed by immunofluorescence that a number of mitotic proteins, such as KIF11, KIF23 and PLK1 were overexpressed in interphase in sarcoma cell lines (data not shown). KIF11 kinesin, which is overexpressed in tetraploid sarcoma cells, was shown to be involved in regulation of cell migration [16, 60]. In addition, AURKA accumulation, at the midzone and midbody region, in late mitosis of tetraploid sarcoma cells also indicates that its Apc/C-cdh1 mediated proteolysis is deficient in these cells. While AURKA degradation is important for the organization of the spindle during anaphase [61] [62], a link was also proposed between stabilization of AURKA and tumorigenesis [63].

Overexpression of mitotic gene products after the completion of mitosis would also be consistent with the fact that invading cancer cells are mostly in G0/G1 [64].

We identified that inhibitors targeting the major mitotic kinases AURKA/B and TTK inhibited cell migration of diploid and tetraploid sarcoma cell subclones and invasion from diploid spheroids. Thus, misexpression of these mitotic kinases regulates the motile behavior of sarcoma cells. However, using highly specific drugs and targeted siRNA approaches, we failed to pinpoint novel functions of a specific mitotic gene product in upregulating motility.

This further suggests that it is a combination of several deregulated mitotic gene products that participate to the cytoskeleton remodeling responsible for the

increased migratory capacity observed in tetraploid clones. Inhibition of cell invasiveness and motility mediated by SP60025 and reversine that affect several mitotic gene products further support this hypothesis.

In the future, it will be of interest to study the motile behavior of *TP53*$^{-/-}$ tetraploids derived from non-CINSARC diploid cells together with the diploid parental cells. This will allow to address whether tetraploidization by itself can promote increased cell motility and eventually whether such a phenotype will be linked to the stabilization of mitotic gene products.

Deregulation of many mitotic gene product observed in CINSARC signature correlates with high metastatic potential [8]. We show that the significance of CINSARC signature, over motile and invading cell behavior, derives from over/misexpression of mitotic kinesins and kinases. Our study further highlights novel functions of mitotic proteins in regulation of cell motility [28–31]. Validation of CINSARC signature in different tumor types [8, 65] increases its importance for prognosis and potential targeted therapies. In this context, it would be of interest to focus on the understanding of the origin of deregulation of so many mitotic gene transcripts. In that regard, an interesting candidate will be the FoxM1 transcription factor that regulates, as a bulk, most kinetochores genes [66] and which is highly expressed in sarcoma [67].

MATERIALS AND METHODS

Cell lines and culture conditions

MFH137 and MFH152 cell lines derive from two distinct tumors of mesenchymal origin that were initially characterized as malignant fibrous histiocytoma (MFH) and were subsequently reclassified as undifferentiated pleomorphic sarcoma (UPS). Isolation of MFH137 cell line was previously described [68]. Briefly, sterile fragment from resected tumor was minced in culture medium and then disaggregated by overnight incubation in collagenase (100 units/mL) at 37°C. Long-term culture (more than 70 passages were done) and standard harvesting procedures were used. We followed the same procedure to isolate the MFH152 cell line. The consistency between the cell lines and the original tumors was done by comparison of the genomic profiles.

IMR90 human fibroblast cell line was a gift of Dr. DA Skoufias (IBS, Grenoble). All cell lines were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% foetal calf serum and antibiotics.

Reagents and antibodies

The drugs used for screening in this study are listed in Supplementary Table 1.

Reversine, SP600125, ZM 447439, BI 2536, STLC, Dimethylenastron, SB203508, RO 3306, Cdk1 Inhibitor III, Roscovitine, NSC 95397, IPA3, Y-27632, ITX-3, Nocodazole, Paclitaxel, Blebbistatin, Cytochalasin B, MG 132 and Velcade were purchased from Sigma–Aldrich (St. Louis, MO, USA). AZ 3146 and Mps-BAY2a were obtained from Tocris (Bristol, United Kingdom). MLN8237 (Alisertib), AZD1152-HQPA (Baraserib) and SB743921 were purchased from Selleckchem. GSK923295 is from MedChem express.

Monoclonal antibodies against α-tubulin, CDK1, AURKA, PLK1, KIF11 (Sigma-Aldrich); CCNB1, CCNE1, TP53, CDKN1A, JNK1, JNK2 (Cell signaling); EB1, COFILIN (Santa Cruz), TTK, BUBR1 (Abcam). Polyclonal antibodies against PRC1 (Santa Cruz), Tpx2 [69]; Kif23 [70] were used.

Migration assays

For wound-healing assay, scratches were performed, with sterile 1000 µl tips, on confluent cell monolayers and monitored every hour by time-lapse microscopy. The two dimensional Oris cell migration assays were performed according to the manufacturer's instructions (Platypus Technologies, Madison, USA). Briefly, cells were seeded (2×10^4 cells per well) into 96-well plates with "silicone stopper" and grown overnight. Then stoppers were removed and cells were incubated for an additional 24h to allow their migration into the empty zone. Cells were fixed with 4% PFA in PBS, stained with DAPI (for DNA detection) and FITC-phalloidin (for actin staining). Data acquisitions were performed using Cellomics Arrayscan (Arrayscan VTI Live; Carl Zeiss, Inc). The number of cells in the migration zone and the total number of cells outside and inside the migration zone (same area for both zones) were counted, by thresholding the DAPI stained nuclei, using image J software. The ratio between these two numbers was used as the percentage of migrating cells. For individual cell migration (Figure 1), cells were plated on glass coverslip coated with fibronectin on linear tracks of different widths (CYTOO™Chips Motility, CYTOO). Cells were seeded at low concentration (approximately 20 cells per mm^2), to avoid cell-cell contacts that would affect analyses. For individual cell migration (Figure 4C–4D), cells were plated on 6 well plates at very low density (30000 cells per well). Cells were monitored by time-lapse microscopy. Image acquisitions were performed every 15 or 30 min (Figure 1D/Figure 4C–4D) for 24h using a Leica DMIRE2 inverted cell microscope equipped with a LMC 20 × 0.4 lens (Leica, Wetzlar, Germany). Time lapse were analyzed using ImageJ software (freely available from the National Institute of Health, http://rsb.info.nih.gov/ij/).

Invasion assays

3D invasion assays (Figure 1E–1F) were performed in 96-well plates (PerkinElmer) coated with 0.2% BSA (Sigma-Aldrich) and 10^4 red fluorescent polystyrene microspheres (FluoSpheres; Invitrogen). In brief, cells were suspended in 2.3 mg/ml serum-free liquid bovine

collagen at 10^5 cells/ml. 100 µl aliquots were dispensed into the plates. Plates were centrifuged at 1,000 rpm and further incubated at 37°C. Once collagen had polymerized, serum was added on top of the collagen to a final concentration of 5% (final volume of 30 µl). After 24 h, cells were fixed with 4% PFA and stained with Hoechst 33342 (Invitrogen). Images were acquired from each well at 50 and 0 µm from the bottom of the well using a Cellomics Arrayscan. The invasion index was calculated as the ratio of the cell number at 50 µm versus 0 µm.

Invasion experiments (Figure 4E) were performed with the *xCELLigence* RTCA DP instrument (Ozyme, France) according to the manufacturer's guidelines. 24 hours prior experiment, cells were deprived of fœtal calf serum. A layer of Matrigel (300µg/ml, BD Biosciences) was applied on the CIM-Plate 16 upper chamber membranes as described [71]. Subsequently, the coated upper chambers were incubated at 37°C to homogenously gelify during a minimum of four hours, followed by addition of 160 µL media to the lower and 30 µL serum free media to the upper chambers. 20 000 cells were seeded in every well of the upper chambers. Cell Index (CI) of each well was automatically monitored with the *xCELLigence* system every hour during a 24 hour period. Each condition was performed in quadruplicate. Cell Index Invasion represents the ratio of Cell Index of Matrigel-coated wells (invasion) to Cell Index of uncoated wells (migration) at specific time points.

In 3D MultiCellular Tumor Spheroids assay MCTS, spheroid formation was performed by incubating the cells (1000) in presence of 2.4% methylcellulose in U-shaped bottom wells of 96-wells plate. After 24hours, formation of multicellular spheroids was observed for diploid clones. Then, using 100 µl pipet tips diploid spheroids were transferred to wells of 96-wells plate and embedded into a collagen matrix for the invasion assay (Invasion matrix contains collagen (PureCol, Sigma) mixed with culture medium at 0.1% FCS with/without drugs). Spheroids behavior was monitored every hour by time-lapse microscopy. Z stack images that spanned the entire size of the spheroids were acquired at every time point. Relative cell invasion/expansion was quantified using ImageJ software. The area colonized by cells, away from the spheroids, was measured in several microscopic slices in the middle of the spheres and compared to the original spheroid diameter. The largest surface was arbitrarily chosen as the representative value of invasive potential for the specific condition.

Measures of cell cycle length and mitosis duration

Cells were seeded on glass coverslip coated with adhesive discs of fibronectin (CYTOOchips™ Motility, CYTOO, France) and imaged every 15 minute for 72h with a Leica DMIRE2 inverted microscope with a LMC 20 × 0.4 lens and appropriate filters (Leica, Wetzlar, Germany). Images were analyzed with Image J software. Cell cycle length was estimated by measuring the time interval between the rounding of the same cell at mitotic entry. Length of mitosis was measured, as the time interval between cell rounding (mitotic entry) and start of spreading after mitotic exit.

Cytofluorometric studies

- For the assessment of cell cycle distribution, cells were collected, washed once with 0.1% (w/v) D-glucose (Sigma-Aldrich) in PBS and then fixed by gentle vortexing in ice-cold 75% (v/v) ethanol for 30 sec. After overnight incubation at -20°C, samples were centrifuged, PBS washed and stained with 50 µg/mL PI in 0.1% (w/v) D-glucose in PBS supplemented with 1 µg/mL (w/v) RNase A (Sigma-Aldrich) for 30 min at 37°C. Afterwards, samples were incubated for at least 2 h at 4°C before cytofluorometric analysis.

- Flow cytometry acquisition and analyses to determine the plasma membrane integrity and mitochondrial transmembrane potential ($\Delta\psi m$) were performed as described previously [72]. Briefly, cells were collected and stained with 1 µg/ mL propidium iodide (PI), which only incorporates into dead cells, and 40 nM 3,3'-dihexyloxacarbocyanine iodide ($DiOC_6(3)$, a $\Delta\psi m$-sensitive dye for 30 min at 37°C before FACs acquisition.

- Cytofluorometric sorting of diploid and tetraploid cells was performed with a FACSAria cell sorter and the gating of small and big cell was based on the size and granularity parameters using the normal light scattering parameters Forward Scatter (FSC) *vs* Side Scatter (SSC). Single cells, sorted from mother cell line, were seeded in 96-well plates. 1000 "small" plus 1000 "big" cells for each MFH152 and MFH137 cell lines were sorted. After 15 days, surviving clones were cultured in 6-well plates. We succeeded to isolate stable tetraploid clones (3 from MFH137 and 5 from MFH152) and stable diploid clones (more than 20 for each cell line) for up to 3 month. The majority of surviving clones were aneuploid.

- To measure histone H3 phosphorylation on serine 10, cells were fixed with 75% (v/v) ethanol in PBS, permeabilized with 0.25% (v/v) Tween 20 in PBS and stained with a rabbit antiserum specific for phosphorylated histone H3 (rabbit polyclonal IgG1 #06–570, Millipore-Chemicon International), as previously described [73]. Cytofluorometric acquisitions were performed by means of a FACSCalibur (BD Biosciences).

Measures of spindle assembly checkpoint robustness

A range of concentration of microtubule dynamic interfering drugs nocodazole and taxol was first assayed on sarcoma cells. Mitotic arrest, indicative of activation of the spindle assembly checkpoint, was quantified by counting the fraction of round cells in the total cell population,

following 16 hours drug exposure using optimized drug concentration (200nM nocodazole and 100nM taxol). To evaluate SAC robustness, diploid (Cl1, Cl2), tetraploid (cl7, cl8) and MFH152 cells were seeded on six well plates, 200nM nocodazole was added after 2 hours after cell adhesion. Cells were imaged every hour for 64h with a Leica DMIRE2 inverted microscope with a LMC 20 × 0.4 lens and appropriate filters (Leica, Wetzlar, Germany). All mitotic cells at 16 hours were assigned a 100% for every condition. The loss of spindle assembly checkpoint, characterized by mitotic exit of these cells was determined at 28, 40, 52 and 63hours in presence of the drugs. Images were analysed with Image J software.

Chromosome spreads

Cells were treated with 100nM nocodazole for 16h to enrich the percentage of mitotic cells, then collected and subjected to hypotonic lysis by incubation in 75mM KCl for 10 min at 37°C. After removal of hypotonic solution, cells were fixed in freshly prepared Carnoy solution (3/1 methanol/acetic acid) and stored at -20°C. Fixed cells were dropped onto pre-cooled glass microscope slides and dried at room temperature. Chromosomes were stained with 100ng/ml DAPI (Molecular Probes–Invitrogen) and mounted in Vectashield H-1000 mounting medium (Vector Laboratories, Burlingame, CA, USA). Fluorescent images were acquired using a Zeiss AxioimagerZ1 motorized microscope (Zeiss, Oberkochen, Germany) driven by MetaMorph (Molecular Devices).

Immunofluorescence microscopy

Immunofluorescence was performed as described [69]. Briefly, cells were either fixed in methanol for 5min at −20°C or in 4% PFA (in PBS with 0.2% Triton X-100), immunostained with indicated primary antibodies Secondary antibodies conjugated to Alexa fluorochromes (Thermo Fisher Scientific-Invitrogen) were used. F-actin and DNA were, respectively, stained with phalloidin-Atto647 and DAPI (Sigma-Aldrich). For analyses of mitotic parental cells, cells were seeded on L shape CYTOOchips™, as recommended by the manufacturer. Confocal microscopy was performed using a Leica SP8-UV microscope equipped with Leica 63x HCX PL APO 1.4 oil CS2 objective. For quantitative comparison, fluorescent staining of respectively KIF11 or AURKA were acquired using same laser power and conditions. Stack images were acquired and maximum intensity projection (MIP) of images (keeping same number of grey levels between images) were analyzed with Image J software.

RNA sequencing

Total RNA from IMR90 and from two diploid and two tetraploid MFH152 subclones were prepared with a TRIzol® Plus RNA Purification Kit (Thermofisher) according to the manufacturer's instructions. The library from total RNA was prepared using the TruSeq®Stranded Total Sample Preparation kit (Illumina Inc.) according to 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 the major peak at 160nt. First strand cDNA synthesis by random hexamers and reverse transcriptase was followed by second strand cDNA synthesis, performed in the presence of dUTP instead of dTTP. The blunt-ended double stranded cDNA was 3´adenylated and Illumina indexed adapters were ligated. The resulting library was enriched with 15 PCR cycles. The libraries were sequenced on HiSeq2000 (Illumina, Inc) in paired-end mode with a read length of 2x76bp using TruSeq SBS Kit v3-HS. Primary data analysis was carried out with the standard Illumina software Real Time Analysis (RTA 1.13.48) and followed by generation of FASTQ files.

RNA seq data analyses and clustering

We estimated general read quality using FastQC (v0.10.1), trimmed low-quality 5' and 3' ends (Phred score <20) using Sickle (v1.2), merged overlapping reads (at least 10 bp) using SeqPrep (v1.1) and split merged reads with an Awk routine. Finally, paired-ends with one read shorter than 30 bp were discarded. Transcriptomic (coding RefSeq (Pruitt, 2012 #106) genes from Hg19 UCSC Table Browser [74] fixed on 2014/01) alignments were performed by TopHat2 (v2.0.1) [75]. Once aligned, we removed paired-end reads that mapped at multiple locations on transcriptome, low-quality (score <20) and discordant paired-end reads using SAMtools (v0.1.19) [76]. MarkDuplicates (PicardTools v1.99) removed PCR duplicates and Cufflinks (v2.1.1) [77] estimated gene and transcript abundances. Miscellaneous filters, statistics and plots were performed by R (v.3.1.1).

Clustering results of IMR90 together with MFH152 diploid clones 1 and 2 together and tetraploid clones 7 and 8. "Height" represents euclidean distances computed with Ward's method between whole transcriptomic profiles (in log2[FPKM]+1 unit). The more height, the more distances between profiles. "Agglomerative coefficient", ranging from 0 to 1, represents the strength of clustering structure. The highest value, the less number of different clusters.

Statistical procedures

Data are expressed as arithmetic means ± SEM. As indicated in the figure legends, statistical analysis was made with Graph pad Prism using ANOVA with Tukey's test as post-test. n denotes the number of different experiments or counted cell.

ACKNOWLEDGMENTS AND FUNDING

We thank Virginie Georget and Julien Bellis for help with Cellomics Arrayscan and data quantification. Cédric Hassen-Khodja for help with HEAT MAP. All imaging was performed at the Montpellier RIO Imaging facility (MRI), Montpellier. The live imaging reported in this paper was performed within the France- BioImaging national research infrastructure, at MRI, Montpellier. France-BioImaging is supported by the French National Research Agency through the ''Investments for the Future'' program (ANR-10-INSB-04). MJ was a recipient from a post-doctoral fellowship from la Ligue Nationale contre le Cancer and GL was funded by an engineer grant from the "Fondation de la Recherche Médicale". Salary to AA and NM is provided by "Institut National de la Santé et de la Recherche Médicale" (INSERM). This project was funded by grants from the "Association pour la Recherche contre le Cancer", "la ligue régionale de l'Hérault" and "la Fondation de France". We thank D. Fesquet, J. Espeut, K. Hached for critical discussions and comments.

CONFLICTS OF INTEREST

No potential conflicts of interest were disclosed.

REFERENCES

1. Andreassen PR, Lohez OD, Lacroix FB and Margolis RL. Tetraploid state induces p53-dependent arrest of nontransformed mammalian cells in G1. Mol Biol Cell. 2001; 12:1315-1328.
2. Clark MA, Fisher C, Judson I and Thomas JM. Soft-tissue sarcomas in adults. N Engl J Med. 2005; 353:701-711.
3. Borden EC, Baker LH, Bell RS, Bramwell V, Demetri GD, Eisenberg BL, Fletcher CD, Fletcher JA, Ladanyi M, Meltzer P, O'Sullivan B, Parkinson DR, Pisters PW, et al. Soft tissue sarcomas of adults: state of the translational science. Clin Cancer Res. 2003; 9:1941-1956.
4. Helman LJ and Meltzer P. Mechanisms of sarcoma development. Nat Rev Cancer. 2003; 3:685-694.
5. Schoffski P, Cornillie J, Wozniak A, Li H and Hompes D. Soft tissue sarcoma: an update on systemic treatment options for patients with advanced disease. Oncol Res Treat. 2014; 37:355-362.
6. Taylor BS, Barretina J, Maki RG, Antonescu CR, Singer S and Ladanyi M. Advances in sarcoma genomics and new therapeutic targets. Nat Rev Cancer. 2011; 11:541-557.
7. Dodd RD, Mito JK and Kirsch DG. Animal models of soft-tissue sarcoma. Dis Model Mech. 2010; 3:557-566.
8. Chibon F, Lagarde P, Salas S, Perot G, Brouste V, Tirode F, Lucchesi C, de Reynies A, Kauffmann A, Bui B, Terrier P, Bonvalot S, Le Cesne A, 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; 16:781-787.
9. Neuville A, Chibon F and Coindre JM. Grading of soft tissue sarcomas: from histological to molecular assessment. Pathology. 2014; 46:113-120.
10. Galipeau PC, Cowan DS, Sanchez CA, Barrett MT, Emond MJ, Levine DS, Rabinovitch PS and Reid BJ. 17p (p53) allelic losses, 4N (G2/tetraploid) populations, and progression to aneuploidy in Barrett's esophagus. Proc Natl Acad Sci U S A. 1996; 93:7081-7084.
11. Meeker AK and Argani P. Telomere shortening occurs early during breast tumorigenesis: a cause of chromosome destabilization underlying malignant transformation? J Mammary Gland Biol Neoplasia. 2004; 9:285-296.
12. Davoli T and de Lange T. Telomere-driven tetraploidization occurs in human cells undergoing crisis and promotes transformation of mouse cells. Cancer Cell. 2012; 21:765-776.
13. Meraldi P, Honda R and Nigg EA. Aurora-A overexpression reveals tetraploidization as a major route to centrosome amplification in p53-/- cells. Embo J. 2002; 21:483-492.
14. Davoli T and de Lange T. The causes and consequences of polyploidy in normal development and cancer. Annu Rev Cell Dev Biol. 2011; 27:585-610.
15. Coelho PA, Bury L, Shahbazi MN, Liakath-Ali K, Tate PH, Wormald S, Hindley CJ, Huch M, Archer J, Skarnes WC, Zernicka-Goetz M and Glover DM. Over-expression of Plk4 induces centrosome amplification, loss of primary cilia and associated tissue hyperplasia in the mouse. Open Biol. 2015; 5:150209.
16. Castillo A, Morse HC, 3rd, Godfrey VL, Naeem R and Justice MJ. Overexpression of Eg5 causes genomic instability and tumor formation in mice. Cancer Res. 2007; 67:10138-10147.
17. Sotillo R, Hernando E, Diaz-Rodriguez E, Teruya-Feldstein J, Cordon-Cardo C, Lowe SW and Benezra R. Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. Cancer Cell. 2007; 11:9-23.
18. Fujiwara T, Bandi M, Nitta M, Ivanova EV, Bronson RT and Pellman D. Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. Nature. 2005; 437:1043-1047.
19. Steigemann P, Wurzenberger C, Schmitz MH, Held M, Guizetti J, Maar S and Gerlich DW. Aurora B-mediated abscission checkpoint protects against tetraploidization. Cell. 2009; 136:473-484.
20. Valente D, Bossi G, Moncada A, Tornincasa M, Indelicato S, Piscuoglio S, Karamitopoulou ED, Bartolazzi A, Pierantoni GM, Fusco A, Soddu S and Rinaldo C. HIPK2 deficiency causes chromosomal instability by cytokinesis failure and increases tumorigenicity. Oncotarget. 2015; 6:10320-10334. doi: 10.18632/oncotarget.3583.
21. Ganem NJ, Cornils H, Chiu SY, O'Rourke KP, Arnaud J, Yimlamai D, Thery M, Camargo FD and Pellman D. Cytokinesis failure triggers hippo tumor suppressor pathway activation. Cell. 2014; 158:833-848.

22. Margolis RL, Lohez OD and Andreassen PR. G1 tetraploidy checkpoint and the suppression of tumorigenesis. J Cell Biochem. 2003; 88:673-683.
23. Lehman NL, Verschuren EW, Hsu JY, Cherry AM and Jackson PK. Overexpression of the anaphase promoting complex/cyclosome inhibitor Emi1 leads to tetraploidy and genomic instability of p53-deficient cells. Cell Cycle. 2006; 5:1569-1573.
24. Ganem NJ, Storchova Z and Pellman D. Tetraploidy, aneuploidy and cancer. Curr Opin Genet Dev. 2007; 17:157-162.
25. Quintyne NJ, Reing JE, Hoffelder DR, Gollin SM and Saunders WS. Spindle multipolarity is prevented by centrosomal clustering. Science. 2005; 307:127-129.
26. Sabino D, Gogendeau D, Gambarotto D, Nano M, Pennetier C, Dingli F, Arras G, Loew D and Basto R. Moesin is a major regulator of centrosome behavior in epithelial cells with extra centrosomes. Curr Biol. 2015; 25:879-889.
27. Nano M and Basto R. The Janus soul of centrosomes: a paradoxical role in disease? Chromosome Res. 2016; 24:127-144.
28. Chou CK, Wu CY, Chen JY, Ng MC, Wang HM, Chen JH, Yuan SS, Tsai EM, Chang JG and Chiu CC. BubR1 Acts as a Promoter in Cellular Motility of Human Oral Squamous Cancer Cells through Regulating MMP-2 and MMP-9. Int J Mol Sci. 2015; 16:15104-15117.
29. Wan J, Zhu F, Zasadil LM, Yu J, Wang L, Johnson A, Berthier E, Beebe DJ, Audhya A and Weaver BA. A Golgi-localized pool of the mitotic checkpoint component Mad1 controls integrin secretion and cell migration. Curr Biol. 2014; 24:2687-2692.
30. Arsic N, Bendris N, Peter M, Begon-Pescia C, Rebouissou C, Gadea G, Bouquier N, Bibeau F, Lemmers B and Blanchard JM. A novel function for Cyclin A2: control of cell invasion via RhoA signaling. J Cell Biol. 2012; 196:147-162.
31. Wang LH, Xiang J, Yan M, Zhang Y, Zhao Y, Yue CF, Xu J, Zheng FM, Chen JN, Kang Z, Chen TS, Xing D and Liu Q. The mitotic kinase Aurora-A induces mammary cell migration and breast cancer metastasis by activating the Cofilin-F-actin pathway. Cancer Res. 2010; 70:9118-9128.
32. Coller HA and Coller BS. Poisson statistical analysis of repetitive subcloning by the limiting dilution technique as a way of assessing hybridoma monoclonality. Methods Enzymol. 1986; 121:412-417.
33. Vinci M, Gowan S, Boxall F, Patterson L, Zimmermann M, Court W, Lomas C, Mendiola M, Hardisson D and Eccles SA. Advances in establishment and analysis of three-dimensional tumor spheroid-based functional assays for target validation and drug evaluation. BMC Biol. 2012; 10:29.
34. Nair JS and Schwartz GK. MLN-8237: A dual inhibitor of aurora A and B in soft tissue sarcomas. Oncotarget. 2016; 7:12893-12903. doi: 10.18632/oncotarget.7335.
35. Mayer TU, Kapoor TM, Haggarty SJ, King RW, Schreiber SL and Mitchison TJ. Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. Science. 1999; 286:971-974.
36. Wilkinson RW, Odedra R, Heaton SP, Wedge SR, Keen NJ, Crafter C, Foster JR, Brady MC, Bigley A, Brown E, Byth KF, Barrass NC, Mundt KE, et al. AZD1152, a selective inhibitor of Aurora B kinase, inhibits human tumor xenograft growth by inducing apoptosis. Clinical Cancer Research. 2007; 13:3682-3688.
37. Jemaa M, Galluzzi L, Kepp O, Senovilla L, Brands M, Boemer U, Koppitz M, Lienau P, Prechtl S, Schulze V, Siemeister G, Wengner AM, Mumberg D, et al. Characterization of novel MPS1 inhibitors with preclinical anticancer activity. Cell Death Differ. 2013; 20:1532-1545.
38. Stukenberg PT. Triggering p53 after cytokinesis failure. J Cell Biol. 2004; 165:607-608.
39. Rieder CL and Maiato H. Stuck in division or passing through: what happens when cells cannot satisfy the spindle assembly checkpoint. Dev Cell. 2004; 7:637-651.
40. De Santis Puzzonia M, Gonzalez L, Ascenzi S, Cundari E and Degrassi F. Tetraploid cells produced by absence of substrate adhesion during cytokinesis are limited in their proliferation and enter senescence after DNA replication. Cell Cycle. 2016; 15:274-282.
41. Danes BS. The Gardner syndrome: increased tetraploidy in cultured skin fibroblast. J Med Genet. 1976; 13:52-56.
42. Lothschutz D, Jennewein M, Pahl S, Lausberg HF, Eichler A, Mutschler W, Hanselmann RG and Oberringer M. Polyploidization and centrosome hyperamplification in inflammatory bronchi. Inflamm Res. 2002; 51:416-422.
43. Olaharski AJ, Sotelo R, Solorza-Luna G, Gonsebatt ME, Guzman P, Mohar A and Eastmond DA. Tetraploidy and chromosomal instability are early events during cervical carcinogenesis. Carcinogenesis. 2006; 27:337-343.
44. Tanaka H, Goto H, Inoko A, Makihara H, Enomoto A, Horimoto K, Matsuyama M, Kurita K, Izawa I and Inagaki M. Cytokinetic Failure-induced Tetraploidy Develops into Aneuploidy, Triggering Skin Aging in Phosphovimentin-deficient Mice. J Biol Chem. 2015; 290:12984-12998.
45. Duelli DM, Padilla-Nash HM, Berman D, Murphy KM, Ried T and Lazebnik Y. A virus causes cancer by inducing massive chromosomal instability through cell fusion. Curr Biol. 2007; 17:431-437.
46. Lv L, Zhang T, Yi Q, Huang Y, Wang Z, Hou H, Zhang H, Zheng W, Hao Q, Guo Z, Cooke HJ and Shi Q. Tetraploid cells from cytokinesis failure induce aneuploidy and spontaneous transformation of mouse ovarian surface epithelial cells. Cell Cycle. 2012; 11:2864-2875.
47. Dewhurst SM, McGranahan N, Burrell RA, Rowan AJ, Gronroos E, Endesfelder D, Joshi T, Mouradov D, Gibbs P, Ward RL, Hawkins NJ, Szallasi Z, Sieber OM and Swanton C. Tolerance of whole-genome doubling

propagates chromosomal instability and accelerates cancer genome evolution. Cancer Discov. 2014; 4:175-185.

48. Woodham EF and Machesky LM. Polarised cell migration: intrinsic and extrinsic drivers. Curr Opin Cell Biol. 2014; 30:25-32.
49. Parsons JT, Horwitz AR and Schwartz MA. Cell adhesion: integrating cytoskeletal dynamics and cellular tension. Nat Rev Mol Cell Biol. 2010; 11:633-643.
50. Leduc C and Etienne-Manneville S. Intermediate filaments in cell migration and invasion: the unusual suspects. Curr Opin Cell Biol. 2015; 32:102-112.
51. Katsetos CD, Reginato MJ, Baas PW, D'Agostino L, Legido A, Tuszyn Ski JA, Draberova E and Draber P. Emerging microtubule targets in glioma therapy. Semin Pediatr Neurol. 2015; 22:49-72.
52. Pines J. Cubism and the cell cycle: the many faces of the APC/C. Nat Rev Mol Cell Biol. 2011; 12:427-438.
53. Lindon C, Grant R and Min M. Ubiquitin-Mediated Degradation of Aurora Kinases. Front Oncol. 2015; 5:307.
54. Sivakumar S and Gorbsky GJ. Spatiotemporal regulation of the anaphase-promoting complex in mitosis. Nat Rev Mol Cell Biol. 2015; 16:82-94.
55. Brown KD, Coulson RM, Yen TJ and Cleveland DW. Cyclin-like accumulation and loss of the putative kinetochore motor CENP-E results from coupling continuous synthesis with specific degradation at the end of mitosis. J Cell Biol. 1994; 125:1303-1312.
56. Jiang W, Jimenez G, Wells NJ, Hope TJ, Wahl GM, Hunter T and Fukunaga R. PRC1: a human mitotic spindle-associated CDK substrate protein required for cytokinesis. Mol Cell. 1998; 2:877-885.
57. Uzbekov R, Prigent C and Arlot-Bonnemains Y. Cell cycle analysis and synchronization of the Xenopus laevis XL2 cell line: study of the kinesin related protein XlEg5. Microsc Res Tech. 1999; 45:31-42.
58. Fischer M, Grundke I, Sohr S, Quaas M, Hoffmann S, Knorck A, Gumhold C and Rother K. p53 and cell cycle dependent transcription of kinesin family member 23 (KIF23) is controlled via a CHR promoter element bound by DREAM and MMB complexes. PLoS One. 2013; 8:e63187.
59. Liot C, Seguin L, Siret A, Crouin C, Schmidt S and Bertoglio J. APC(cdh1) mediates degradation of the oncogenic Rho-GEF Ect2 after mitosis. PLoS One. 2011; 6:e23676.
60. Falnikar A, Tole S and Baas PW. Kinesin-5, a mitotic microtubule-associated motor protein, modulates neuronal migration. Mol Biol Cell. 2011; 22:1561-1574.
61. Lindon C and Pines J. Ordered proteolysis in anaphase inactivates Plk1 to contribute to proper mitotic exit in human cells. J Cell Biol. 2004; 164:233-241.
62. Floyd S, Pines J and Lindon C. APC/C Cdh1 targets aurora kinase to control reorganization of the mitotic spindle at anaphase. Curr Biol. 2008; 18:1649-1658.
63. Kitajima S, Kudo Y, Ogawa I, Tatsuka M, Kawai H, Pagano M and Takata T. Constitutive phosphorylation of aurora-a on ser51 induces its stabilization and consequent overexpression in cancer. PLoS One. 2007; 2:e944.
64. Yano S, Miwa S, Mii S, Hiroshima Y, Uehara F, Yamamoto M, Kishimoto H, Tazawa H, Bouvet M, Fujiwara T and Hoffman RM. Invading cancer cells are predominantly in G0/G1 resulting in chemoresistance demonstrated by real-time FUCCI imaging. Cell Cycle. 2014; 13:953-960.
65. Lagarde P, Perot G, Kauffmann A, Brulard C, Dapremont V, Hostein I, Neuville A, Wozniak A, Sciot R, Schoffski P, Aurias A, Coindre JM, Debiec-Rychter M and Chibon F. Mitotic checkpoints and chromosome instability are strong predictors of clinical outcome in gastrointestinal stromal tumors. Clin Cancer Res. 2012; 18:826-838.
66. Thiru P, Kern DM, McKinley KL, Monda JK, Rago F, Su KC, Tsinman T, Yarar D, Bell GW and Cheeseman IM. Kinetochore genes are coordinately up-regulated in human tumors as part of a FoxM1-related cell division program. Mol Biol Cell. 2014; 25:1983-1994.
67. Kelleher FC and O'Sullivan H. FOXM1 in sarcoma: role in cell cycle, pluripotency genes and stem cell pathways. Oncotarget. 2016; 7:42792-42804. doi: 10.18632/oncotarget.8669.
68. Perot G, Derre J, Coindre JM, Tirode F, Lucchesi C, Mariani O, Gibault L, Guillou L, Terrier P and Aurias A. Strong smooth muscle differentiation is dependent on myocardin gene amplification in most human retroperitoneal leiomyosarcomas. Cancer Res. 2009; 69:2269-2278.
69. Bompard G, Rabeharivelo G, Cau J, Abrieu A, Delsert C and Morin N. P21-activated kinase 4 (PAK4) is required for metaphase spindle positioning and anchoring. Oncogene. 2013; 32:910-919.
70. Fesquet D, De Bettignies G, Bellis M, Espeut J and Devault A. Binding of Kif23-iso1/CHO1 to 14-3-3 is regulated by sequential phosphorylations at two LATS kinase consensus sites. PLoS One. 2015; 10:e0117857.
71. Limame R, Wouters A, Pauwels B, Fransen E, Peeters M, Lardon F, De Wever O and Pauwels P. Comparative analysis of dynamic cell viability, migration and invasion assessments by novel real-time technology and classic endpoint assays. PLoS One. 2012; 7:e46536.
72. Jemaa M, Vitale I, Kepp O, Berardinelli F, Galluzzi L, Senovilla L, Marino G, Malik SA, Rello-Varona S, Lissa D, Antoccia A, Tailler M, Schlemmer F, et al. Selective killing of p53-deficient cancer cells by SP600125. EMBO Mol Med. 2012; 4:500-514.
73. Vitale I, Jemaa M, Galluzzi L, Metivier D, Castedo M and Kroemer G. Cytofluorometric assessment of cell cycle progression. Methods Mol Biol. 2013; 965:93-120.
74. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D and Kent WJ. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 2004; 32:D493-496.

75. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R and Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14:R36.

76. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078-2079.

77. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ and Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010; 28:511-515.

IV – La signature CINSARC comme marqueur pronostique du devenir clinique dans les sarcomes et au-delà

Received: 3 October 2018 | Accepted: 15 October 2018
DOI: 10.1002/gcc.22703

WILEY

REVIEW ARTICLE

CINSARC signature as a prognostic marker for clinical outcome in sarcomas and beyond

Frederic Chibon[1,2] | Tom Lesluyes[1,3,4] | Thibaud Valentin[1,5] | Sophie Le Guellec[1,2]

[1]INSERM U1037, Cancer Research Center of Toulouse (CRCT), Toulouse, France
[2]Department of Pathology, Institut Claudius Regaud, Toulouse, France
[3]University of Bordeaux, Bordeaux, France
[4]Institut Claudius Regaud, Toulouse, France
[5]Department of Medical Oncology, Institut Claudius Regaud, Toulouse, France

Correspondence
Frederic Chibon, DR INSERM-Eq-19: ONCOSARC, CRCT/IUCT-Oncopole, UMR1037 Inserm-Université Toulouse 3 - ERL5294 CNRS, 2 avenue Hubert Curien, CS 53717, 31037 Toulouse CEDEX 1 – France.
Email: frederic.chibon@inserm.fr

Abstract

Prognostication is a key issue for sarcoma patients' care as it triggers the therapeutic approach including chemotherapy, which is still not standard for localized patients. Current prognostic evaluation, based on the FNCLCC grading system, has recently been improved by the CINSARC signature outperforming histology-based grading system by identifying high-risk patients in every grade, even in those considered as low. CINSARC is an expression-based signature related to mitosis and chromosome integrity with prognostic value in a wide range of cancers additional to sarcoma. First developed with frozen material, CINSARC is now coupled with NanoString technology allowing evaluation from FFPE blocks used in clinical practice. Consequently, CINSARC is currently evaluated in clinical trials with a dual objective of demonstrating the benefit of chemotherapy in sarcoma patients and testing its response prediction. Considering its overarching value in oncology, its development is welcome in any cancers where the prognostication needs to be improved.

KEYWORDS

chromosome instability, CINSARC, prognosis, sarcoma

1 | INTRODUCTION

Whether chemotherapy should be standard care in localized soft tissue sarcoma (STS) is still a debated question. It is obviously a striking situation in the modern area of personalized medicine.

The clinical evaluation of any therapeutic strategy will be favorable only if the latter is given to patients who will benefit from it. Sarcomas exhibit a wide range of clinical behavior from locally aggressive proliferation, which is cured by local excision, to widely metastatic and even fatal tumors that may benefit from systemic treatment. Therefore, predicting the likelihood of progression at the time of initial biopsy or resection is essential.

Currently, prognostication in sarcoma is performed the world over by morphologic assessment of tumor grade using the *Fédération Nationale des Centres de Lutte Contre le Cancer* (FNCLCC) criteria.[1] The FNCLCC system includes differentiation, mitotic index, and tumor necrosis to split sarcomas into three groups, each carrying a distinct risk of metastatic spread and death. Histologic type, resection margin status, tumor size, and site are also informative. Although the FNCLCC system remains the gold standard for the prognosis of adult STS more than three decades after its development, it has shown its limitations. Because the parameters are somewhat subjective, its reproducibility is imperfect. Moreover, FNCLCC grading is of greater predictive value in some histologic subtypes than in others.[2] However, maybe the most important limitation is that small biopsies may not adequately represent necrosis, information that is critical to accurate grading.

The concept of histological grade was introduced by Broders in 1920 in his princeps work on squamous cell carcinoma of the lip and was based on the percentage of the well-differentiated component which correlated with mortality.[3] Since the pioneering work of Broders, several grading systems have been set up and used on many tumor types such as the breast, prostate, endometrium, ovary, bladder, kidney, colon/rectum, brain, lymphoma, and sarcoma. In every type, grade is based on the histological evaluation of the primary tumor to predict aggressiveness. Histological grade should be considered as a morphological translation of molecular events that determine tumor aggressiveness. One of the main molecular events now clearly associated with metastatic spread is chromosome instability.[4] In an attempt to decipher this mechanism by applying both genomic and transcriptomic characterizations on a sarcoma cohort, the French Sarcoma Group identified a link between chromosomal structural alterations and metastatic outcome in sarcoma.[5] The pivotal study

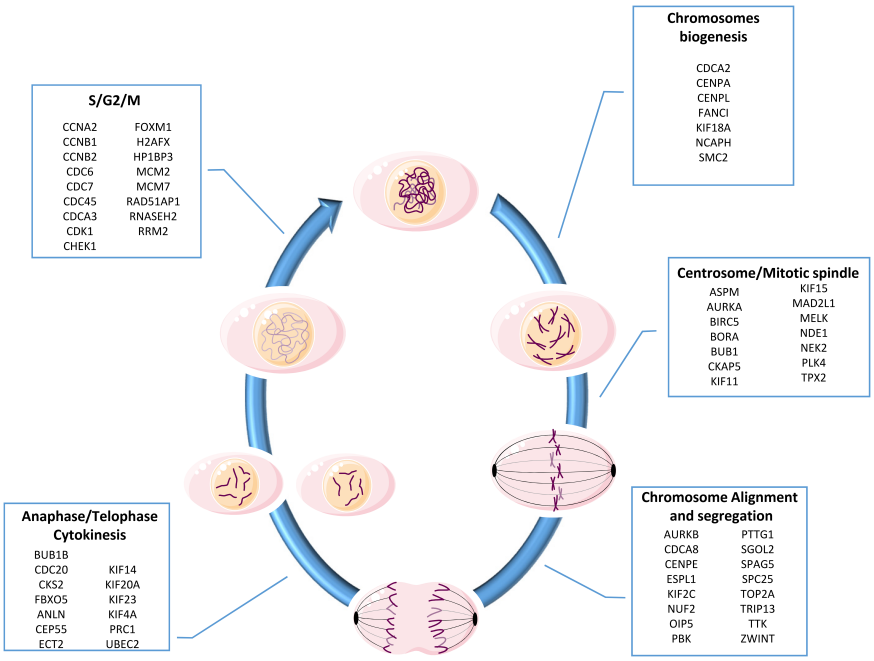


FIGURE 1 Schematic representation of the main role of each of the Complexity INdex in SARComas (CINSARC) genes

found that FNCLCC grade was associated with genomic instability: the more rearranged the genome is, the higher the grade. Thereafter, a transcriptomic signature named CINSARC, which stands for Complexity INdex in SARComas, was derived combining differentially expressed genes in FNCLCC grade 3 versus grades 1 and 2 tumors and in highly versus lowly rearranged tumors. By outperforming the FNCLCC grading system for the prognosis of metastatic outcome, CINSARC was seen as the tool that had been missing to select patients who will mostly benefit from chemotherapy.

This review focuses on the biological definition of CINSARC, its robustness and wide-ranging domain of application, and finally on its technological development leading to clinical application in sarcomas and, maybe, beyond.

1.1 | What is CINSARC?

CINSARC genes were initially selected for their statistical significance. They all belong to the significantly enriched pathways in highly versus lowly rearranged sarcomas and in FNCLCC G3 versus G2 sarcomas.[5] Despite this objective selection, they are all related to chromosome biogenesis, mitosis control, and chromosome segregation (Figure 1 and Supporting Information Table S1).

Gene ontology enrichment analysis indicates (Supporting Information Table S1) that mitosis, spindle microtubules, and kinase activity are the most enriched "process," "component," and "function" in this gene set. Obviously, such a highly homogeneous biological meaning suggests that the mechanism highlighted should be a putative driver of metastatic spread and of poor outcome. Is CINSARC really driver or just permissive to? This is the main question we are still facing.

One might argue that the CINSARC signature is primarily indicative of proliferation, but CINSARC-positive cells proliferate slowly and undergo a longer period of mitosis than negative cells.[6] Overexpression of several individual CINSARC genes disrupts chromosomal segregation and generates genomic alterations and tumors in mice, that is, *AURKA*, *MAD2L1*, *BUB1*, *CCNB1*, *CCNB2*, and *ESPL1*,[7–11] which argue for a driver impact. Of note, strong associations have been reported between CINSARC expression and the level of genome complexity/instability.[12,13] CINSARC is an orchestrated gene expression signature in which genes are overexpressed without gain, amplification, or structural variation of their matching locus. Consequently, their expression may be induced by deregulated transcription factors. Motif analysis shows that CINSARC genes are mainly under transcriptional control of NFY (A, B, or C) and E2F transcription factors (Supporting Information Table S2), which are both known to regulate cell cycle genes.[14] *E2F* refers to the *RB1* gene family, which is one of the most frequently altered pathways in sarcomas.[15] However, no association between *RB1* pathway alteration and CINSARC expression has been identified and reported so far, with the exception of GIST, in which CINSARC expression, genomic instability, and metastatic spread are associated with *CDKN2A* deletion.[12] The role of E2F transcription factors in controlling the gene expression central to the cell cycle and cell fate decisions has been well documented,[16] and there is evidence of a complex combinatorial mechanism of gene regulation involving E2F proteins together with

other cooperating transcriptional regulators.[17–19] This includes regulatory cascades in which E2Fs activate the expression of genes encoding other transcription factors and then cooperate with these induced factors to regulate additional target genes, a relationship known as a feed-forward regulatory loop.[20,21] This has been reported for E2F1 and NFYB[22] as well as for DREAM and FOXM1,[23,24] the latter belonging to the CINSARC gene set. Such transcriptional regulation could indeed be more complex and even related to host, metabolism, or epigenetics (eg, chromatin methylation and 3D-genome structure), fields still underexplored in this context.

Another mechanistic issue is the link between genomic complexity, as measured by CINSARC, and metastatic spread. This point has been recently clarified by an elegant study demonstrating that genomic instability triggers DNA fragment spilling into the cytosol. This leads to the activation of the cGAS-STING cytosolic DNA-sensing pathway and downstream noncanonical NF-κB signaling, which in turn co-opt chronic activation of innate immune pathways to spread to distant organs.[25]

The prognostic value of CINSARC is thus supported mechanistically by (1) the pathway enrichment of the gene set restricted to mitosis and chromosome integrity and (2) the driver role of chromosomal instability in metastatic spread. Moreover, it is now clear that metastatic competence is afforded by chromosome complexity in other cancers.[26] Therefore, the CINSARC signature could serve as an important biomarker for stratifying patients for systemic therapy in almost every type of malignant tumor.

1.2 | CINSARC to universally predict outcome?

Initially developed for pleomorphic sarcomas, CINSARC was then tested in sarcomas with simple genetics, either associated with a translocation, for example, synovial sarcoma, or a point mutation, for example, GIST (*SS18-SSX1/2* fusion genes and *KIT* or *PDGFRA* mutations, respectively).[12,13] Its prognostic value was validated for both entities, outperforming the "gold standard" histological classifications (FNCLCC and AFIP grading systems, respectively).

Then, Lesluyes et al reported a pan-cancer study demonstrating that CINSARC was significantly enriched with top-ranked poor prognosis genes in 21 out of 39 cancer types tested, including carcinomas (eg, breast, liver, prostate, lung, kidney, and so forth), hematological (eg, multiple myeloma, Burkitt's, and Mantle cell lymphomas), and brain-related (eg, astrocytoma, medulloblastoma, neuroblastoma, and glioma) cancers.[27] With 21 different cancer types, it covered the widest spectrum of neoplasm families compared to all other signatures (15 499 other signatures included in the GeneSigDB[28] and MSigDB[29] databases).

At cancer-specific level, the prognostic value of CINSARC was tested with overall and disease-specific survival analyses. Accordingly, 83 datasets representing 27 cancer types were evaluated. CINSARC was a significant predictor of 33 datasets (40%) representing 17 (63%) cancer types, demonstrating its possible application as a global marker for tumor aggressiveness (metastasis and/or overall survival) in a wide range of tumor families.

Therefore, the next step to make CINSARC a new standard for patients' care, at least in sarcomas, is to establish its applicability in daily practice and its prospective validation. For the former, its use should be feasible with RNA extracted from FFPE blocks as it is the material pathologists handle for diagnosis. Thereafter, patients could be selected for clinical trials on the basis of this signature.

The rise in high-throughput sequencing has led to better understanding of numerous, if not all, biological mechanisms.[30,31] RNA sequencing (RNA-seq) aims to achieve the global characterization of transcriptomic patterns at a single base-pair resolution.[32] It also makes it possible to identify novel alternative splicing,[33] including the expression of transcription-induced chimeras (also termed readthroughs).[34] Furthermore, RNA-seq has uncovered expressed genomic alterations such as point mutations and insertions/deletions,[35] translocations leading to fusion genes,[36] and viral genome integrations.[37]

Consequently, the routine application of RNA-seq in clinical centers would lead to better management of patients. In this context, RNA sequencing on different cohorts of soft-tissue sarcomas was performed to assess whether CINSARC can be applied with FFPE samples. First, using 95 fresh-frozen cases, CINSARC still predicted outcome using both microarrays and RNA-seq with metastasis-free survival analyses ($P = 2.77 \times 10^{-2}$ and $P = 1.34 \times 10^{-2}$ for microarrays and RNA-seq, respectively). In an independent cohort of sarcomas analyzed by RNA-seq,[38] it also remained a significant outcome predictor at gene ($P = 4.14 \times 10^{-3}$) and transcript levels ($P = 7.9 \times 10^{-4}$).[39] Second, to test the validity of FFPE material support, 41 paired fresh-frozen and FFPE cases were processed (polyA-selected and ribo-depleted libraries, respectively).[39] Unfortunately, most of the tested samples (24 out of 41, 59%) had too degraded RNA for correct prognostication. Among the other 17 cases, 15 (88%) remained in the same prognostic groups compared to paired fresh-frozen tissues.

The use of RNA-seq for CINSARC evaluation thus appears to be limited to RNA extracted from fresh-frozen tissues, being unable to correctly process more than half FFPE cases. A precise evaluation of RNA degradation in FFPE samples regarding specific factors such as sunlight, temperature, and fixation time should therefore be undertaken to try to optimize RNA-seq processing.[40]

1.3 | NanoCind: The daily practice application of CINSARC for pathologists and physicians

NanoString, a probe-based technique using direct digital measurement of transcript abundance with multiplexed color-coded probe pairs, has been shown to quantify mRNA expression accurately in FFPE and fragmented material.[41,42] Several publications have demonstrated the accuracy and precision of NanoString with FFPE material.[41,43,44] NanoString is hybridization-based and does not require reverse transcription of mRNA and subsequent cDNA amplification. This feature offers advantages compared to PCR-based methods, including the absence of amplification bias, which is higher when using fragmented RNA isolated from FFPE blocks (eg, RNA-seq).[39]

In this context, the nCounter code set NanoCind (patent number EP18305190.3) was developed and comprises a panel of 75 probes, including 67 distinct test probes derived from the 67 CINSARC genes and 8 from housekeeping genes for normalization purposes.[45] It strongly predicted metastatic outcome on frozen samples ($n = 124$; HR = 2.9, 95% CI 1.23-6.82) with similar risk-group classifications (86%) compared to RNA-seq technology. However, unlike RNA-seq,

all (n = 67) FFPE blocks (comprising 20 core-needle biopsies) were analyzable with NanoString, despite their poor RNA quality. Both the value and the superiority of NanoCind were confirmed by three results: first, the significant prognostic value was more discriminating than that established by RNA sequencing ($P = 2.45 \times 10^{-3}$ vs $P = 3.42 \times 10^{-2}$); second, it outperformed the current histological grade (FNCLCC; $P = 9.39 \times 10^{-3}$ vs $P = 6.58 \times 10^{-1}$); and third, in grade 2 (FNCLCC) tumors, it clearly and significantly differentiated two groups of different outcomes ($P = 5.16 \times 10^{-4}$). CINSARC's advantages over histologic grading could be in part related to the stratification of patients into two groups instead of three, which is essential for facilitating clinical management. Of note, Le Guellec et al[45] showed that NanoCind is applicable to microbiopsies, which are currently the gold standard in terms of STS management.[46]

1.4 | CINSARC and patient care

The aim of this ongoing work on the technical optimization of CINSARC is to allow it to be used in routine settings. The final step of this optimization[45] overcomes all the former limitations to its clinical use, with a technique now perfectly feasible in FFPE samples and requiring only low RNA inputs, making it applicable in core-needle biopsies. This is very timely, considering that the treatments of patients with localized STS has hardly evolved in recent decades, making their prognoses desperately stagnating. In particular, by selecting patients with a high-risk of metastatic relapse in a different way compared to classical risk factors such as tumor size/situation/FNCLCC grade, CINSARC will shed considerable light on the most debated question concerning the treatment of localized STS: the role of perioperative chemotherapy.[47] Until now, the main pitfall has been selecting which patients are most likely to benefit from chemotherapy. A meta-analysis including all sarcoma patients analyzed with the CINSARC signature (605 patients, fig. 2 from Valentin et al)[48] demonstrated that it identifies a subgroup of high-risk patients in each FNCLCC grade. This feature has already been proposed to overcome the classical limitation of FNCLCC grade in grade 2 patients. However, the meta-analysis showed that in patients with grade 1 STS, who are typically considered as low risk, CINSARC identifies high-risk patients carrying a poor prognosis (Figure 2).

Several retrospective and prospective projects are ongoing to confirm these findings, establish its value, and use it in therapeutic settings. CINSARC status is currently being retrospectively analyzed in patients included in two major recent studies: the first evaluates the effect of histotype-tailored neoadjuvant chemotherapy in localized STS (ISG-STS 1001),[49] and the second the role of preoperative radiotherapy in retroperitoneal STS (STRASS trial NCT01344018). The NEOSARCOMICS study (NCT02789384) is underway, aiming to assess CINSARC's value in predicting response to neoadjuvant chemotherapy in STS. In addition, two French prospective clinical trials are aiming to base the treatment of patients with localized STS on the CINSARC risk. Indeed, the underlying questions may not be the same for all STS patients: in patients previously considered as high-risk using conventional factors (G3 FNCLCC, deep, and large tumors) and in whom chemotherapy was thought to be beneficial, one question could be the most effective number of perioperative chemotherapy (dose intensity notion), whereas patients previously considered conventionally as low-risk (G1 or G2 FNCLCC), but with a high-risk CINSARC signature, might benefit from chemotherapy that is currently not part of their treatment.

If the value of the CINSARC signature is confirmed by these studies, it would represent nothing less than the most significant advance in the field of STS treatment in recent decades.

2 | DISCUSSION

Comprising genes involved in genomic instability, CINSARC is a powerful tool for identifying tumors in which this mechanism acts and triggers metastatic spread. As genomic instability is likely universal, CINSARC has been demonstrated to be efficient for detecting poor outcome in every cancer category, that is, carcinomas, lymphomas, leukemias, and sarcomas. Obviously, for the latest, much work has been done or is ongoing with multivariate validation, technical transfer, and prospective validation to make CINSARC the biomarker for patients' eligibility for perioperative chemotherapy. Considering the performance of CINSARC in other cancers, work now needs to be done in each individual malignancy where the first question is how CINSARC performs as compared to the classical tools in a given pathology. The more widespread use of CINSARC will be greatly facilitated by NanoCind, which allows CINSARC analysis to be performed with FFPE blocks and even with fine-needle biopsies. The movement has already started with gynecological malignancies where NanoCind is being tested for its ability to detect the rare patients with a poor prognosis in endometrial carcinoma, or its ability to detect patients with advanced ovarian carcinoma who will benefit from surgery and adjuvant chemotherapy.

In clinical terms, the question remains about its value in predicting response to chemotherapy. Ongoing clinical trials should answer this question. Beyond its clinical validation, there are still technical improvements to be made, first of all being a noninvasive procedure to evaluate CINSARC, from circulating tumor cells in a blood sampling, which would thus be helping following up response to treatment.

In mechanistic terms, the remaining question is what drives CINSARC expression? Large sequencing projects either published or ongoing (TCGA, ICGC) and integrating multiscale genomic and transcriptomic data are currently aiming to resolve this issue and should provide insight for functional testing in the coming months. Depending on the findings, a question will be whether CINSARC is finally targetable in a therapeutic setting? Some of the individual CINSARC genes have already been targeted, that is, *AURKA*,[50] but what would be the effect of targeting the CINSARC driver? Despite these missing answers, it has now started its clinical evaluation to become the selection criterion for (neo-)adjuvant therapy in sarcoma and maybe beyond. Using it to answer the question of the benefit of chemotherapy in sarcoma would thus be a meaningful breakthrough in the field.

ORCID

Frederic Chibon https://orcid.org/0000-0002-6548-2181
Thibaud Valentin https://orcid.org/0000-0002-3189-8445
Sophie Le Guellec https://orcid.org/0000-0003-3207-0081

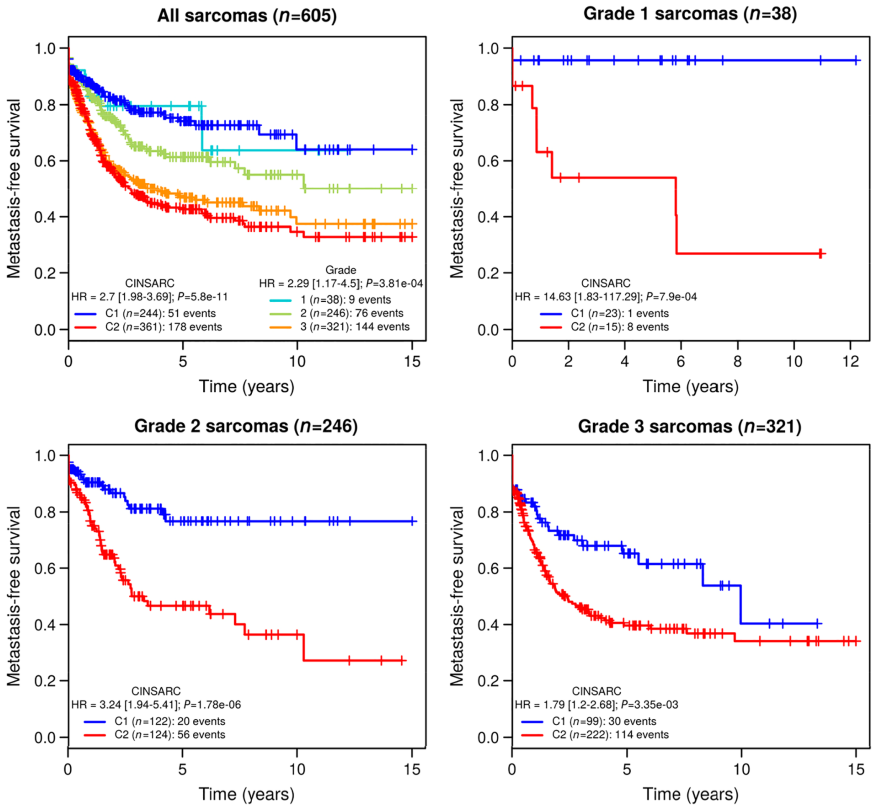


FIGURE 2 A, Metastasis-free survival of patients with localized soft tissue sarcoma (STS) according to *Fédération Nationale des Centres de Lutte Contre le Cancer* (FNCLCC) grade and Complexity INdex in SARComas (CINSARC) signature. Metastasis-free survival of patients with localized FNCLCC grade 1 (B), grade 2 (C), or grade 3 (D) STS, according to CINSARC signature (from Valentin et al)[48]

REFERENCES

1. Trojani M, Contesso G, Coindre JM, et al. Soft-tissue sarcomas of adults; study of pathological prognostic variables and definition of a histopathological grading system. *Int J Cancer*. 1984;33(1):37-42.
2. Coindre J-M. Grading of soft tissue sarcomas: review and update. *Arch Pathol Lab Med*. 2006;130(10):1448-1453. https://doi.org/10.1043/1543-2165(2006)130[1448:GOSTSR]2.0.CO;2.
3. Wright JR, Albert C. Broders' paradigm shifts involving the prognostication and definition of cancer. *Arch Pathol Lab Med*. 2012;136(11):1437-1446. https://doi.org/10.5858/arpa.2011-0567-HP.
4. Bakhoum SF, Cantley LC. The multifaceted role of chromosomal instability in cancer and its microenvironment. *Cell*. 2018;174(6):1347-1360. https://doi.org/10.1016/j.cell.2018.08.027.
5. Chibon F, Lagarde P, Salas S, et al. Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. *Nat Med*. 2010;16(7):781-787. https://doi.org/10.1038/nm.2174.
6. Jemaà M, Abdallah S, Lledo G, et al. Heterogeneity in sarcoma cell lines reveals enhanced motility of tetraploid versus diploid cells. *Oncotarget*. 2017;8(10):16669-16689. https://doi.org/10.18632/oncotarget.14291.
7. Zhou H, Kuang J, Zhong L, et al. Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat Genet*. 1998;20(2):189-193. https://doi.org/10.1038/2496.
8. Sotillo R, Hernando E, Díaz-Rodríguez E, et al. Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. *Cancer Cell*. 2007;11(1):9-23. https://doi.org/10.1016/j.ccr.2006.10.019.
9. Ricke RM, Jeganathan KB, van Deursen JM. Bub1 overexpression induces aneuploidy and tumor formation through Aurora B kinase hyperactivation. *J Cell Biol*. 2011;193(6):1049-1064. https://doi.org/10.1083/jcb.201012035.
10. Nam H-J, van Deursen JM. Cyclin B2 and p53 control proper timing of centrosome separation. *Nat Cell Biol*. 2014;16(6):538-549. https://doi.org/10.1038/ncb2952.
11. Mukherjee M, Ge G, Zhang N, et al. MMTV-Espl1 transgenic mice develop aneuploid, estrogen receptor alpha (ERα)-positive mammary adenocarcinomas. *Oncogene*. 2014;33(48):5511-5522. https://doi.org/10.1038/onc.2013.493.
12. Lagarde P, Pérot G, Kauffmann A, et al. Mitotic checkpoints and chromosome instability are strong predictors of clinical outcome in gastrointestinal stromal tumors. *Clin Cancer Res*. 2012;18(3):826-838. https://doi.org/10.1158/1078-0432.CCR-11-1610.
13. Lagarde P, Przybyl J, Brulard C, et al. Chromosome instability accounts for reverse metastatic outcomes of pediatric and adult synovial sarcomas. *J Clin Oncol*. 2013;31(5):608-615. https://doi.org/10.1200/JCO.2012.46.0147.
14. Gatta R, Dolfini D, Mantovani R. NF-Y joins E2Fs, p53 and other stress transcription factors at the apoptosis table. *Cell Death Dis*. 2011;2(5):e162. https://doi.org/10.1038/cddis.2011.45.

15. Chibon F, Mairal A, Fréneaux P, et al. The RB1 gene is the target of chromosome 13 deletions in malignant fibrous histiocytoma. *Cancer Res*. 2000;60(22):6339-6345.
16. Polager S, Ginsberg D. p53 and E2f: partners in life and death. *Nat Rev Cancer*. 2009;9(10):738-748. https://doi.org/10.1038/nrc2718.
17. Giangrande PH, Hallstrom TC, Tunyaplin C, Calame K, Nevins JR. Identification of E-box factor TFE3 as a functional partner for the E2F3 transcription factor. *Mol Cell Biol*. 2003;23(11):3707-3720.
18. Hallstrom TC, Nevins JR. Jab1 is a specificity factor for E2F1-induced apoptosis. *Genes Dev*. 2006;20(5):613-623. https://doi.org/10.1101/gad.1345006.
19. Schlisio S, Halperin T, Vidal M, Nevins JR. Interaction of YY1 with E2Fs, mediated by RYBP, provides a mechanism for specificity of E2F function. *EMBO J*. 2002;21(21):5775-5786.
20. Shats I, Gatza ML, Liu B, Angus SP, You L, Nevins JR. FOXO transcription factors control E2F1 transcriptional specificity and apoptotic function. *Cancer Res*. 2013;73(19):6056-6067. https://doi.org/10.1158/0008-5472.CAN-13-0453.
21. Lu H, Hallstrom TC. The nuclear protein UHRF2 is a direct target of the transcription factor E2F1 in the induction of apoptosis. *J Biol Chem*. 2013;288(33):23833-23843. https://doi.org/10.1074/jbc.M112.447276.
22. Jiang X, Nevins JR, Shats I, Chi J-T. E2F1-mediated induction of NFYB attenuates apoptosis via joint regulation of a pro-survival transcriptional program. *PLoS One*. 2015;10(6):e0127951. https://doi.org/10.1371/journal.pone.0127951.
23. Grant GD, Brooks L, Zhang X, et al. Identification of cell cycle–regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors. *Mol Biol Cell*. 2013;24(23):3634-3650. https://doi.org/10.1091/mbc.E13-05-0264.
24. Fischer M, Grossmann P, Padi M, DeCaprio JA. Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. *Nucleic Acids Res*. 2016;44(13):6070-6086. https://doi.org/10.1093/nar/gkw523.
25. Bakhoum SF, Ngo B, Laughney AM, et al. Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature*. 2018;553(7689):467-472. https://doi.org/10.1038/nature25432.
26. Turajlic S, Xu H, Litchfield K, et al. Tracking cancer evolution reveals constrained routes to metastases: TRACERx renal. *Cell*. 2018;173(3):581-594.e12. https://doi.org/10.1016/j.cell.2018.03.057.
27. Lesluyes T, Delespaul L, Coindre J-M, Chibon F. The CINSARC signature as a prognostic marker for clinical outcome in multiple neoplasms. *Sci Rep*. 2017;7(1):5480. https://doi.org/10.1038/s41598-017-05726-x.
28. Culhane AC, Schwarzl T, Sultana R, et al. GeneSigDB--a curated database of gene expression signatures. *Nucleic Acids Res*. 2010;38:D716-D725. https://doi.org/10.1093/nar/gkp1015.
29. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinforma (Oxf Engl)*. 2011;27(12):1739-1740. https://doi.org/10.1093/bioinformatics/btr260.
30. Bainbridge MN, Warren RL, Hirst M, et al. Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. *BMC Genomics*. 2006;7:246. https://doi.org/10.1186/1471-2164-7-246.
31. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5(7):621-628. https://doi.org/10.1038/nmeth.1226.
32. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57-63. https://doi.org/10.1038/nrg2484.
33. Sultan M, Schulz MH, Richard H, et al. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*. 2008;321(5891):956-960. https://doi.org/10.1126/science.1160342.
34. Akiva P, Toporik A, Edelheit S, et al. Transcription-mediated gene fusion in the human genome. *Genome Res*. 2006;16(1):30-36. https://doi.org/10.1101/gr.4137606.
35. Chepelev I, Wei G, Tang Q, Zhao K. Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. *Nucleic Acids Res*. 2009;37(16):e106. https://doi.org/10.1093/nar/gkp507.
36. Maher CA, Palanisamy N, Brenner JC, et al. Chimeric transcript discovery by paired-end transcriptome sequencing. *Proc Natl Acad Sci U S A*. 2009;106(30):12353-12358. https://doi.org/10.1073/pnas.0904720106.
37. Sung W-K, Zheng H, Li S, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet*. 2012;44(7):765-769. https://doi.org/10.1038/ng.2295.
38. Cancer Genome Atlas Research Network. Comprehensive and integrated genomic characterization of adult soft tissue sarcomas. *Cell*. 2017;171(4):950-965.e28. https://doi.org/10.1016/j.cell.2017.10.014.
39. Lesluyes T, Pérot G, Largeau MR, et al. RNA sequencing validation of the Complexity INdex in SARComas prognostic signature. *Eur J Cancer*. 2016;57:104-111. https://doi.org/10.1016/j.ejca.2015.12.027.
40. von Ahlfen S, Missel A, Bendrat K, Schlumpberger M. Determinants of RNA quality from FFPE samples. *PLoS One*. 2007;2(12):e1261. https://doi.org/10.1371/journal.pone.0001261.
41. Geiss GK, Bumgarner RE, Birditt B, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol*. 2008;26(3):317-325. https://doi.org/10.1038/nbt1385.
42. Malkov VA, Serikawa KA, Balantac N, et al. Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter Assay system. *BMC Res Notes*. 2009;2:80. https://doi.org/10.1186/1756-0500-2-80.
43. Northcott PA, Shih DJH, Remke M, et al. Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. *Acta Neuropathol (Berl)*. 2012;123(4):615-626. https://doi.org/10.1007/s00401-011-0899-7.
44. Reis PP, Waldron L, Goswami RS, et al. mRNA transcript quantification in archival samples using multiplexed, color-coded probes. *BMC Biotechnol*. 2011;11:46. https://doi.org/10.1186/1472-6750-11-46.
45. Le Guellec S, Lesluyes T, Sarot E, et al. Validation of the Complexity INdex in SARComas prognostic signature on formalin-fixed, paraffin-embedded, soft-tissue sarcomas. *Ann Oncol*. 2018;29(8):1828-1835. https://doi.org/10.1093/annonc/mdy194.
46. Fletcher CD, Bridge JA, Hogendoorm PCW. *World Health Organization Classification of Tumours: Pathology and Genetics of Tumours of Soft Tissue and Bone*. Lyon: IARC Press; 2012.
47. Casali PG, Abecassis N, Bauer S, et al. Soft tissue and visceral sarcomas: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2018;29:iv51-iv67. https://doi.org/10.1093/annonc/mdy096.
48. Valentin T, Lesluyes T, Le Guellec S, Chibon F. Chemotherapy in localized soft tissue sarcoma: will we soon have to treat grade 1 tumours ? An update of CINSARC performances. *Ann Oncol*. https://doi.org/10.1093/annonc/mdy465.
49. Gronchi A, Ferrari S, Quagliuolo V, et al. Histotype-tailored neoadjuvant chemotherapy versus standard chemotherapy in patients with high-risk soft-tissue sarcomas (ISG-STS 1001): an international, open-label, randomised, controlled, phase 3, multicentre trial. *Lancet Oncol*. 2017;18(6):812-822. https://doi.org/10.1016/S1470-2045(17)30334-0.
50. Dauch D, Rudalska R, Cossa G, et al. A MYC-aurora kinase A protein complex represents an actionable drug target in p53-altered liver cancer. *Nat Med*. 2016;22(7):744-753. https://doi.org/10.1038/nm.4107.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Chibon F, Lesluyes T, Valentin T, Le Guellec S. CINSARC signature as a prognostic marker for clinical outcome in sarcomas and beyond. *Genes Chromosomes Cancer*. 2019;58:124–129. https://doi.org/10.1002/gcc.22703

V – Réarrangements alternatifs de *PDGFD* dans les dermatofibrosarcomes de Darier-Ferrand sans les fusions de *PDGFB*

Modern Pathology (2018) 31:1683–1693
https://doi.org/10.1038/s41379-018-0089-4

ARTICLE

Alternative *PDGFD* rearrangements in dermatofibrosarcomas protuberans without *PDGFB* fusions

Bérengère Dadone-Montaudié[1] · Laurent Alberti[2,3] · Adeline Duc[3] · Lucile Delespaul[4,5,11] · Tom Lesluyes[4,5,11] · Gaëlle Pérot[6] · Agnès Lançon[3] · Sandrine Paindavoine[3] · Ilaria Di Mauro[1] · Jean-Yves Blay[2,7] · Arnaud de la Fouchardière[3] · Frédéric Chibon[4,6,11] · Marie Karanian[3] · Gaëtan MacGrogan[6] · Valérie Kubiniek[1] · Frédérique Keslair[1] · Nathalie Cardot-Leccia[8] · Audrey Michot[9] · Virginie Perrin[10] · Yanis Zekri[10] · Jean-Michel Coindre[5,6] · Franck Tirode[2,10] · Florence Pedeutour[1] · Dominique Ranchère-Vince[3] · François Le Loarer[5,6] · Daniel Pissaloux[2,3]

Received: 13 February 2018 / Revised: 7 May 2018 / Accepted: 7 May 2018 / Published online: 28 June 2018

Abstract

Dermatofibrosarcoma protuberans is underlined by recurrent *collagen type I alpha 1 chain-platelet-derived growth factor B chain (COL1A1-PDGFB)* fusions but ~4% of typical dermatofibrosarcoma protuberans remain negative for this translocation in routine molecular screening. We investigated a series of 21 cases not associated with the pathognomonic *COL1A1-PDGFB* fusion on routine fluorescence in situ hybridization (FISH) testing. All cases displayed morphological and clinical features consistent with the diagnosis of dermatofibrosarcoma protuberans. RNA-sequencing analysis was successful in 20 cases. The classical *COL1A1-PDGFB* fusion was present in 40% of cases ($n = 8/20$), and subsequently confirmed with a *COL1A1* break-apart FISH probe in all but one case ($n = 7/8$). 55% of cases ($n = 11/20$) displayed novel *PDGFD* rearrangements; *PDGFD* being fused either to the 5′ part of *COL6A3* (2q37.3) ($n = 9/11$) or *EMILIN2* (18p11) ($n = 2/11$). All rearrangements led to in-frame fusion transcripts and were confirmed at genomic level by FISH and/or array-comparative genomic hybridization. *PDGFD*-rearranged dermatofibrosarcoma protuberans presented clinical outcomes similar to typical dermatofibrosarcoma protuberans. Notably, the two *EMILIN2-PDGFD* cases displayed fibrosarcomatous transformation and homozygous deletions of *CDKN2A* at genomic level. We report the first recurrent molecular variant of dermatofibrosarcoma protuberans involving *PDGFD*, which functionally mimic bona fide COL1A1-PDGFB fusions, leading presumably to a similar autocrine loop-stimulating PDGFRB. This study also emphasizes that *COL1A1-PDGFB* fusions can be cytogenetically cryptic on FISH testing in a subset of cases, thereby representing a diagnostic pitfall that pathologists should be aware of.

Introduction

Dermatofibrosarcoma protuberans is a superficial and locally invasive mesenchymal tumor, which undergoes fibrosarcomatous transformation, portending a metastatic potential in 10% of cases [1]. Giant cell fibroblastoma represents a morphological variant of dermatofibrosarcoma protuberans occurring in children. Both dermatofibrosarcoma protuberans and giant cell fibroblastoma are molecularly defined by the presence of translocation (17;22)(q21.3;q13.1), which fuses *collagen type I alpha 1 chain* (*COL1A1*) with *platelet-derived growth factor B chain* (*PDGFB*) [2]. Most dermatofibrosarcomas protuberans display typical morphological features including a dense storiform proliferation of cells

These authors contributed equally: Bérengère Dadone-Montaudié, Laurent Alberti

These authors jointly supervised this work: Francois Le Loarer, Daniel Pissaloux

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41379-018-0089-4) contains supplementary material, which is available to authorized users.

✉ Daniel Pissaloux
daniel.pissaloux@lyon.unicancer.fr

Extended author information available on the last page of the article

expressing diffusely and strongly CD34 [1]. Fibrous morphological variants of dermatofibrosarcoma protuberans may be confused with benign cellular fibrous histiocytomas or dermatofibromas [1]. A diagnosis of dermatofibrosarcoma protuberans warrants complete surgical excision that may require large resection owing to the propensity of dermatofibrosarcoma protuberans to infiltrate adjacent tissues [3]. Imatinib and pazopanib therapies, which target PDGF downstream signaling, may be used in unresectable or metastatic cases [3, 4]. Molecular confirmation of dermatofibrosarcoma protuberans is therefore strongly recommended. *COL1A1-PDGFB* fusion gene is mostly screened with dual fusion or break-apart fluorescence in situ hybridization (FISH) in routine practice [5]. Reverse transcription polymerase chain reaction (RT-PCR) assays are slightly less sensitive than FISH owing to the variability of *COL1A1* breakpoints and require non-degraded RNA, although some studies report comparable sensitivity [6, 7]. Cytogenetic studies have shown that t(17;22) is mostly unbalanced, associated with supernumerary ring chromosomes, and may therefore be evidenced by array-comparative genomic hybridization [8, 9]. Altogether, routine molecular screening remains negative in ~4% of otherwise typical dermatofibrosarcomas protuberans, raising clinical uncertainties and leading to render a diagnosis of dermatofibrosarcoma protuberans lacking the canonical translocation and therefore "molecularly unconfirmed" [5]. Single reports of molecular variants of dermatofibrosarcoma protuberans have been reported but no systematic study has been performed so far [10–13].

In order to identify novel fusion genes in dermatofibrosarcoma protuberans, we investigated by RNA-sequencing and array-comparative genomic hybridization a series of 21 morphologically typical dermatofibrosarcomas protuberans, negative for *COL1A1-PDGFB* fusion on routine molecular testing.

Materials and methods

Sample selection

We identified 25 cases in our records diagnosed as dermatofibrosarcoma protuberans between 2006 and 2017 but negative on molecular testing. Cases were retrieved from the archives of the department of Genetics of University Hospital of Nice (Nice, France), and departments of Pathology of Centre Léon Bérard (Lyon, France) and Institut Bergonié (Bordeaux, France). Notably, all three participating institutions used FISH for molecular testing, but they applied two different screening strategies. First, *COL1A1-PDGFB* fusion probe followed by *PDGFB* break-apart probe in case of negative or equivocal result. Second, a screening with *PDGFB* break-apart probe in the first place, followed if negative or ambiguous by a screening with a *COL1A1* break-apart probe. Owing to the evolution of molecular testing strategies over time, seven cases of our series had not been previously tested with a *PDGFB* break-apart probe. Hence, FISH analysis was completed in these cases, rendering positive results in four cases, which were therefore removed from the final cohort ($n = 21$).

All cases from the series were registered in the RREPS sarcoma database, a national clinical database approved by the Commission Nationale Informatique et Libertés (CNIL) and the Comité national d'éthique.

The material used in this publication was provided by the SarcomaBCB, the Conticanet database (https://auth.sarcomabcb.org).

Clinical review

Clinical follow-up was obtained from the medical records of patients either available at our institutions or through corresponding clinicians. Follow-up duration was calculated from the date of the clinical detection of the lesion. Clinical data are provided in Supplementary Table 1.

Histopathological analyses

All cases of the final series were reviewed blindly by soft tissue expert pathologists (FLL, JMC, DRV). Fibrosarcomatous transformation was defined by the presence of a fascicular component with increased mitotic activity contrasting with the storiform pattern of the DFSP component [1]. All samples were studied with a panel of antibodies available in routine including AE1/E3, CD34, S100 protein, EMA, and Ki67. Pathological data are summarized in Supplementary Table 1.

FISH analyses

FISH analyses were performed on formalin-fixed paraffin-embedded tissue sections using commercial and custom bacterial artificial chromosome (BAC) probes. Commercial FISH probes used in the study included dual fusion probe *COL1A1/PDGFB* (#Z-2116-200, Zytovision, Bremerhaven, Germany), *PDGFB* Break-Apart probe (#Z-2119-200, Zytovision), and *COL1A1* Break Apart Probe (#Z-2121-200, ZytoVision). *PDGFD*, *EMILIN2*, and *COL6A3* break-apart probes were prepared with BAC listed in Supplementary Table 2. FISH analysis was performed by assessing at least 100 non-overlapping intact nuclei by two independent operators. The positive threshold to call the FISH assay positive was 15%. BACs were cultured and labeled as previously described [14].

Table 1 Clinical and molecular features of the cohort of dermatofibrosarcoma protuberans variants

Case number	Age (years)/ gender	Location	Diagnosis	RNA Sequencing	Array-comparative genomic hybridization	FISH validation	RT-PCR validation
1	33/M	Abdominal wall	Dermatofibrosarcoma protuberans	*COL6A3-PDGFD* fusion	*COL6A3* breakpoint: 3′*COL6A3* heterozygous deletion	Split *PDGFD:* +; Split *COL6A3:* +	*COL6A3-PDGFD* fusion
2	38/F	Breast	Dermatofibrosarcoma protuberans	*COL6A3-PDGFD* fusion	*PDGFD* breakpoint: *PDGFD* intragenic deletion	Split *PDGFD:* +; Split *COL6A3:* +	*COL6A3-PDGFD* fusion
3	38/F	Chest wall	Dermatofibrosarcoma protuberans	*COL6A3-PDGFD* fusion	*PDGFD* breakpoint: *PDGFD* intragenic deletion	Split *PDGFD:* + ; Split *COL6A3:* +	*COL6A3-PDGFD* fusion
4	48/F	Chest wall	Dermatofibrosarcoma protuberans	*COL6A3-PDGFD* fusion	*COL6A3* breakpoint: 3′*COL6A3* heterozygous deletion ; *PDGFD* breakpoint: *PDGFD* intragenic deletion	Split *PDGFD*: Failure ; Split *COL6A3:* Failure	*COL6A3-PDGFD* fusion
5	33/F	Shoulder	Dermatofibrosarcoma protuberans	*COL6A3-PDGFD* fusion	*COL6A3* breakpoint: 3′*COL6A3* heterozygous deletion ; *PDGFD* breakpoint: 5′*PDGFD* heterozygous deletion	ND	*COL6A3-PDGFD* fusion
6	39/F	Back	Dermatofibrosarcoma protuberans	*COL6A3-PDGFD* fusion	*COL6A3* breakpoint: 3′*COL6A3* heterozygous deletion ; *PDGFD* breakpoint: 5′*PDGFD* heterozygous deletion	Split *PDGFD:* +; Split *COL6A3:* +	*COL6A3-PDGFD* fusion
7	29/F	Breast	Dermatofibrosarcoma protuberans	*COL6A3-PDGFD* fusion	Failure	Split *PDGFD:* +; Split *COL6A3:* +	*COL6A3-PDGFD* fusion
8	41/F	Chest wall	Dermatofibrosarcoma protuberans	*COL6A3-PDGFD* fusion	ND	Split *PDGFD:* +; Split *COL6A3:* Failure	*COL6A3-PDGFD* fusion
9	21/F	Leg	Fibrosarcoma	*EMILIN2-PDGFD* fusion *COL5A1-UBAP1L*fusion	No breakpoint within *PDGFD* or *EMILIN2*	Split *PDGFD:* +; Split *EMILIN2:* +	*EMILIN2-PDGFD* fusion
10	50/M	Leg	Fibrosarcoma	*EMILIN2-PDGFD*fusion	*PDGFD* breakpoint: *PDGFD* intragenic deletion	Split *PDGFD:* +; Split *EMILIN2:* +	*EMILIN2-PDGFD* fusion
11	64/M	Back	Atypical dermatofibrosarcoma protuberans	*COL1A1-PDGFB* fusion	*COL1A1* breakpoint: 5′*COL1A1* gain	*Split COL1A1:* +	
12	24/M	Thigh	Fibrosarcoma	*COL1A1-PDGFB* fusion	*COL1A1* breakpoint: 5′*COL1A1* gain	Split *COL1A1:* +	
13	43/F	Thigh	Dermatofibrosarcoma protuberans	*COL1A1-PDGFB* fusion	*COL1A1* breakpoint: 5′*COL1A1* gain	Split *COL1A1:* +	
14	40/M	Forehead	Dermatofibrosarcoma protuberans	*COL1A1-PDGFB* fusion	*COL1A1* breakpoint: 5′*COL1A1* gain	Split *COL1A1:* +	
15	37/F	Cheek	Dermatofibrosarcoma protuberans	*COL1A1-PDGFB* fusion	*COL1A1* breakpoint: 5′*COL1A1* gain	Split *COL1A1:* +	
16	38/M	Neck	Dermatofibrosarcoma protuberans	*COL1A1-PDGFB* fusion	ND	Split *COL1A1:* +	
17	73/F	Arm	Dermatofibrosarcoma protuberans	*COL1A1-PDGFB* fusion	ND	Split *COL1A1:* +	
18	29/F	Shoulder	Fibrosarcoma	*COL1A1-PDGFB* fusion	ND	Split *COL1A1:* -	
19	29/F	Breast	Dermatofibrosarcoma protuberans	*COL6A3-PDGFD* fusion (frozen)	*COL6A3* breakpoint: 5′*COL6A3* gain	Split *PDGFD:* + ; Split *COL6A3:* +	*COL6A3-PDGFD* fusion
20	4/M	Calf	Pigmented dermatofibrosarcoma protuberans	No fusion transcript	No breakpoint	Split *PDGFD:* - ; Split *COL6A3:* - ; Split *EMILIN2:* -	

M male, *F* female, *RNA sequencing* ribonucleic acid sequencing, *FISH* fluorescence in situ hybridization, *ND* not done

Array-comparative genomic hybridization analyses

Genomic DNA was extracted from formalin-fixed paraffin-embedded tissue using QIAamp, DNA micro kit (Qiagen, Hilden, Germany). Genomic DNA and human reference DNA (Promega) were labeled with cyanin 5 (Cy5) and cyanine 3 (Cy3), respectively, using the Genomic DNA High-Throughput ULS Labeling Kit (Agilent Technologies, Santa Clara, California) and co-hybridized onto a Sureprint G3 Human CGH microarray 4 × 180K (Agilent) following manufacturer's recommendations. Data were analyzed by Agilent Genomic Workbench software v7.0 or by Cytogenomics software (v2.9.2.4, Agilent) and expressed according to the human reference hg19 (GRCh37, Genome Reference Consortium Human Reference 36). The identification of aberrant copy number segments was based on ADM-2 segmentation algorithm with a threshold of 6.0.

RNA-sequencing analyses

RNA sequencing was performed with formalin-fixed paraffin-embedded material in all cases. Total RNA was extracted from formalin-fixed paraffin-embedded tissue section using Trizol reagent (Thermo Fisher Scientific, Courtaboeuf, France) following manufacturers' recommendations. DNA was removed using RNase-free DNase set (Qiagen) followed by second Trizol extraction. Quantity and quality of total RNA were evaluated using NanoDrop (Thermo Fisher Scientific) and Tape Station with Hs RNA Screen Tape (Agilent) using a cutoff of DV_{200} (defined as the percentage of RNA fragments above 200 nucleotides) above 13%. All samples passed quality criteria. Libraries were prepared with 100 ng of total RNA using TruSeq RNA Access Library Prep Kit (Illumina, San Diego, USA). Libraries were pooled by group of 14 samples at 4 nM with 1% PhiX. Sequencing was performed (75 cycles paired end) using a NextSeq 500/550 High Output V2 kit on Illumina NextSeq 500 platform (Illumina, San Diego, CA). RNA-sequencing was successful in 20 out of 21 cases. The case with failure was ruled out from the final series as no material was available to complete its characterization by array-comparative genomic hybridization and FISH (final series $n = 20$, Table 1).

Sequencing data (up to 15 million reads per sample) were analyzed with BaseSpace sequence Hub (Illumina). Reads were aligned with STAR and TopHat2 on GRCh38 reference genome. The fusion transcripts were called with Manta, STAR-Fusion, FusionMap and TopHat2 fusion and validated if the Manta score was up to 0.7 and present in fusion list with TopHat2 fusion and/or STAR-Fusion and/or FusionMap [15–17].

To perform the clustering analysis, gene expression values were extracted using Kallisto v0.42.5 tool [18] with GENECODE release 23 genome annotation based on GRCh38 genome reference. Kallisto TPM expression values were transformed in log2(TPM + 2) and all samples were normalized together using the quantile method from the R limma package within R (version 3.1.1) environment. Clustering was performed with the R package Cluster v2.0.3 using Pearson correlation distance and Ward's clustering method. Significance of clusters was assessed using SigClust as previously described [19].

RT-PCR and Sanger sequencing

In 11 cases, we performed reverse transcriptase-polymerase chain reaction (RT-PCR) to confirm either the *COL6A3-PDGFD* fusion transcript or the *EMILIN2-PDGFD* fusion transcript, using the following custom primers (designed using Primer 3 program): *COL6A3* forward primer 5′-GCAAGGTCAGCTTCTAGTTCA-3', *PDGFD* reverse primer 5′-TGGCCAACTTCAGCTCTTCT-3′ (in case number 1, 2, 4, 5, 6, 7, 8, 19), *COL6A3* forward primer 5′-CAGCAAGGTCAGCTTCTAGTT-3′ and *PDGFD* reverse primer 5'-CAGTTCCACAGCCACAATTTC-3′ (in case number 3) and *EMILIN2* forward primer 5′-GCCACGTCTTCCAGATTTCTA-3′ and *PDGFD* reverse primer 5′-CAGTTCCACAGCCACAATTTC-3′ (in case number 9, 10). Fusion transcripts were sequenced by Sanger after extraction and enzymatic purification of RT-PCR product using High pure PCR product purification kit (Roche Diagnostics, Meylan, France).

Results

Clinicopathological features

All cases were located in classical anatomical sites including trunk ($n = 11/20$), limbs ($n = 6/20$), and head and neck ($n = 3/20$) (Table 1). Patients' age ranged from 4 to 73 years old. Clinically, lesions presented with a classical scar-like appearance ($n = 2$), slow-growing cutaneous plaques ($n = 1$) or cutaneous plaques with red discoloration ($n = 2$). In addition, one case had a multinodular presentation. All tumors displayed outcomes in line with the local malignant potential of dermatofibrosarcoma protuberans with late local recurrences occurring up to 12 years after initial resection (Supplementary Table 1). All cases presented morphological features either typical of dermatofibrosarcoma protuberans ($n = 12/20$) (Figs. 1 and 2) or consistent with a morphological variant of dermatofibrosarcoma protuberans including fibrous pattern ($n = 5/20$), pigmented pattern

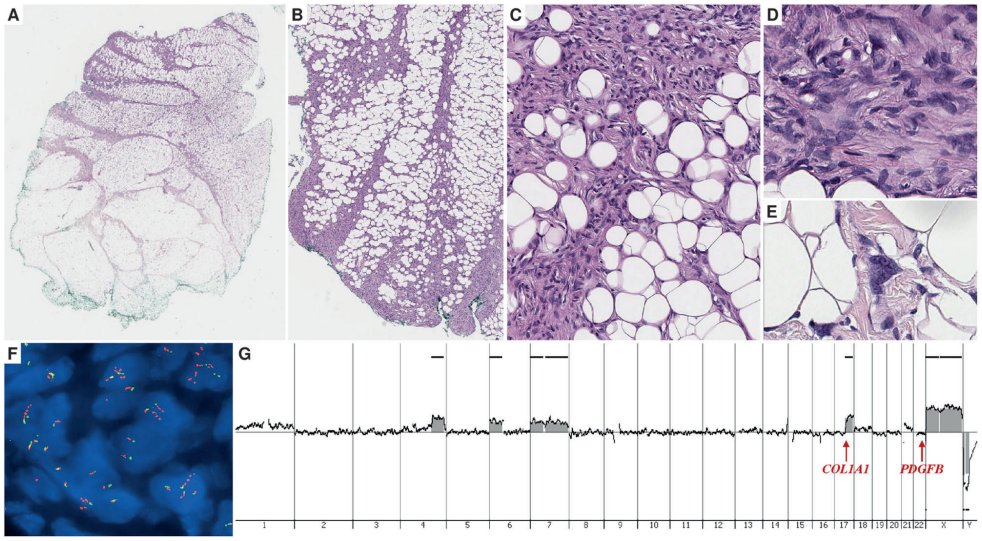


Fig. 1 Cryptic *COL1A1-PDGFB* fusion in dermatofibrosarcoma protuberans. **a–b** Band-like infiltration of hypodermis and subcutaneous adipose tissue (case number 15, HES staining, × 2.5 and × 6.5 magnifications, respectively) **c** Storiform proliferation infiltrating hypodermis with a honeycomb pattern (HES, × 200 magnification). **d–e** High-power field view of tumor cell nuclei displaying ovoid shape and monomorphism **d** and focal interspersed multinucleated cells **e** (HES, × 200 magnification) **f** Fluorescence in situ hybridization (FISH) using *COL1A1* break-apart probe. FISH analysis shows an unbalanced translocation with a gain of the 5′ part of *COL1A1* (probe labeled in red). **g** Array-comparative genomic hybridization profile harbors a genomic breakpoint at *COL1A1* locus with a 5′*COL1A1* gain but no breakpoint at *PDGFB* locus (chromosome 22 view not shown). Additional copy number alterations include gains of 4q28-q35, 6p, and whole chromosome 7

(also referred to as Bednar tumor) ($n = 1/20$) or hybrid Bednar/dermatofibrosarcoma protuberans variant with focal presence of pigments or giant cells ($n = 2/20$) (Supplementary Table 1). Four cases displayed fibrosarcomatous transformation (Fig. 3). All but two cases showed classical honeycomb infiltrative borders. Five cases were deeply located in hypodermis without dermal infiltration.

RNA sequencing

Classical *COL1A1-PDGFB* fusions were found in eight cases ($n = 8/20$). *PDGFB* breakpoint was constantly found in exon 2, whereas breakpoints in *COL1A1* varied from exons 11 to 45 (Supplementary Table 3).

PDGFD rearrangements were found in 11 cases ($n = 11/20$), *PDGFD* being fused either to *collagen type VI alpha 3 chain* (*COL6A3)* at 2q37 ($n = 9/11$) or *elastin microfibril Interface 2* (*EMILIN2)* at 18p11 ($n = 2/11$). Breakpoints in *COL6A3* involved either exon 42 ($n = 8/9$) or exon 43 ($n = 1/9$), whereas breakpoint in *PDGFD* constantly involved exon 6 (Fig. 2) and breakpoints in *EMILIN2* were located within exon 4 (Fig. 3).

One case (case number 20) failed to show a fusion transcript by RNA-sequencing with the different algorithms.

FISH and RT-PCR

COL6A3-PDGFD transcript and *EMILIN2-PDGFD* transcript were confirmed by RT-PCR in all cases (case number 1–10 and 19). Regarding cases with *COL1A1-PDGFB* fusions, FISH analyses were completed with a *COL1A1* break-apart probe, which was positive in 7/8 cases (Fig. 1). Regarding cases with *PDGFD* fusions, FISH analyses were carried out with *PDGFD* break-apart probe, which was positive in nine cases out of 11 ($n = 9/11$). Gene partners fused to *PDGFD* were also assessed with *COL6A3* and *EMILIN2* break-apart probes (Figs. 2 and 3, respectively), showing rearrangements in 5/7 and 2/2 cases, respectively (Supplementary Table 3).

Array-comparative genomic hybridization

On profiles, all cases with cryptic *COL1A1-PDGFB* highlighted genomic breakpoints within *COL1A1* with a genomic gain in flanking region 17q21.33-qter ($n = 5/5$). Additional non-recurrent copy number alterations were present in all cases. Genomic gains were seen in 1q21-q25.3, 4q28-q35, 6p, 7pter, 22q11 in one case each. Whole-chromosome gains involved chromosomes 6, 7, 10, and

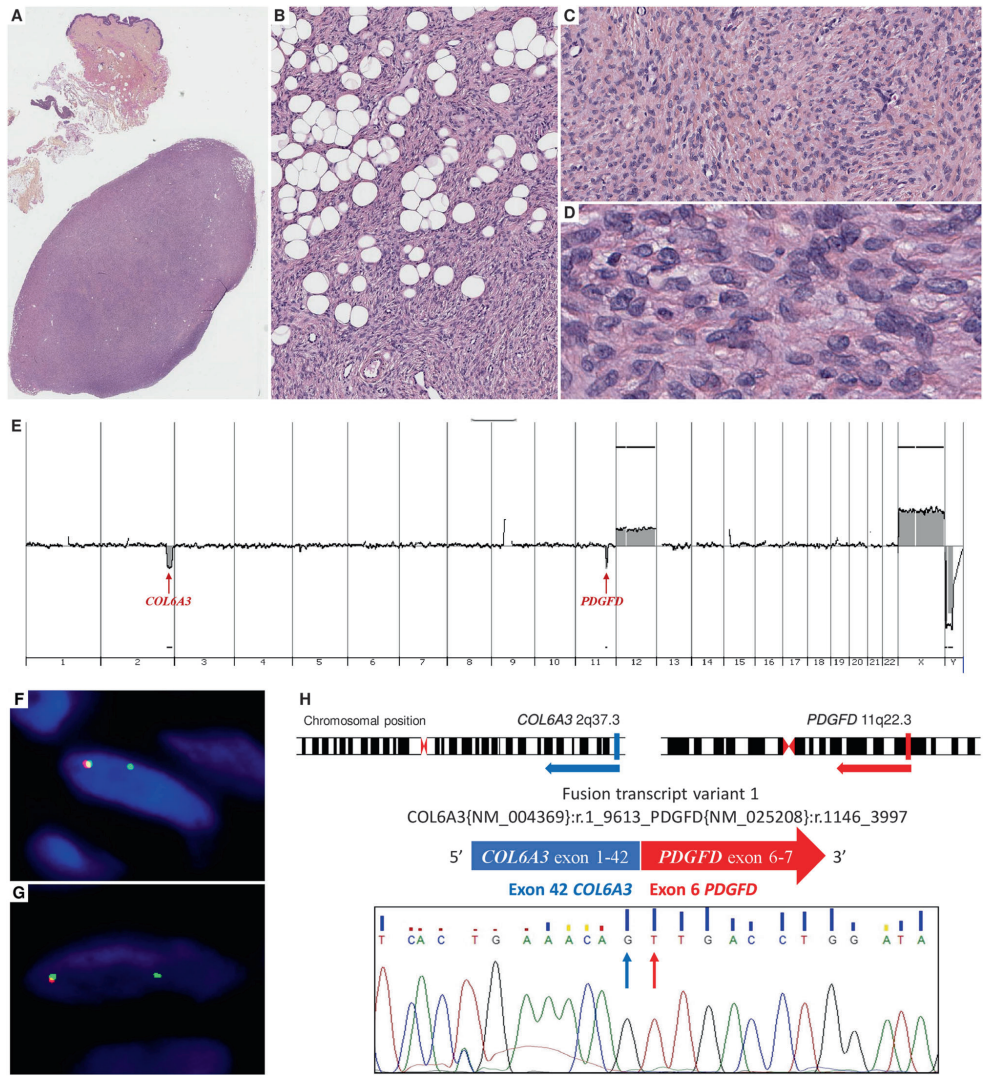


Fig. 2 *COL6A3-PDGFD* dermatofibrosarcoma protuberans. **a** Biopsy specimen highlighting deep-seated dense nodular proliferation (case number 6, HES staining, × 20 magnification). **b** Storiform proliferation with honeycomb infiltrative pattern typical of dermatofibrosarcoma protuberans (HES, × 200). **c** Fibrous stromal changes focally interposed between tumor cells (HES, × 250) **d** Tumor cells display ovoid monomorphic nuclei (HES, × 400). **e** Array-comparative genomic hybridization profile showing deletions and breakpoints involving *COL6A3* and *PDGFD*. Additional copy number alteration includes gain of chromosome 12. **f–g** Fluorescence in situ hybridization (FISH) using *COL6A3* break-apart probe **f** and *PDGFD* break-apart probe **g** showing an unbalanced rearrangement with loss of the 3′*COL6A3* and the 5′*PDGFD*, respectively. **h** Structure of *COL6A3-PDGFD* fusion transcript. From top to bottom: scheme representing locus and chromosomal position of *COL6A3* and *PDGFD*; schematic of breakpoints positions involving *COL6A3* exon 42 and *PDGFD* exon 6; nucleotide sequence of adjoined sequences

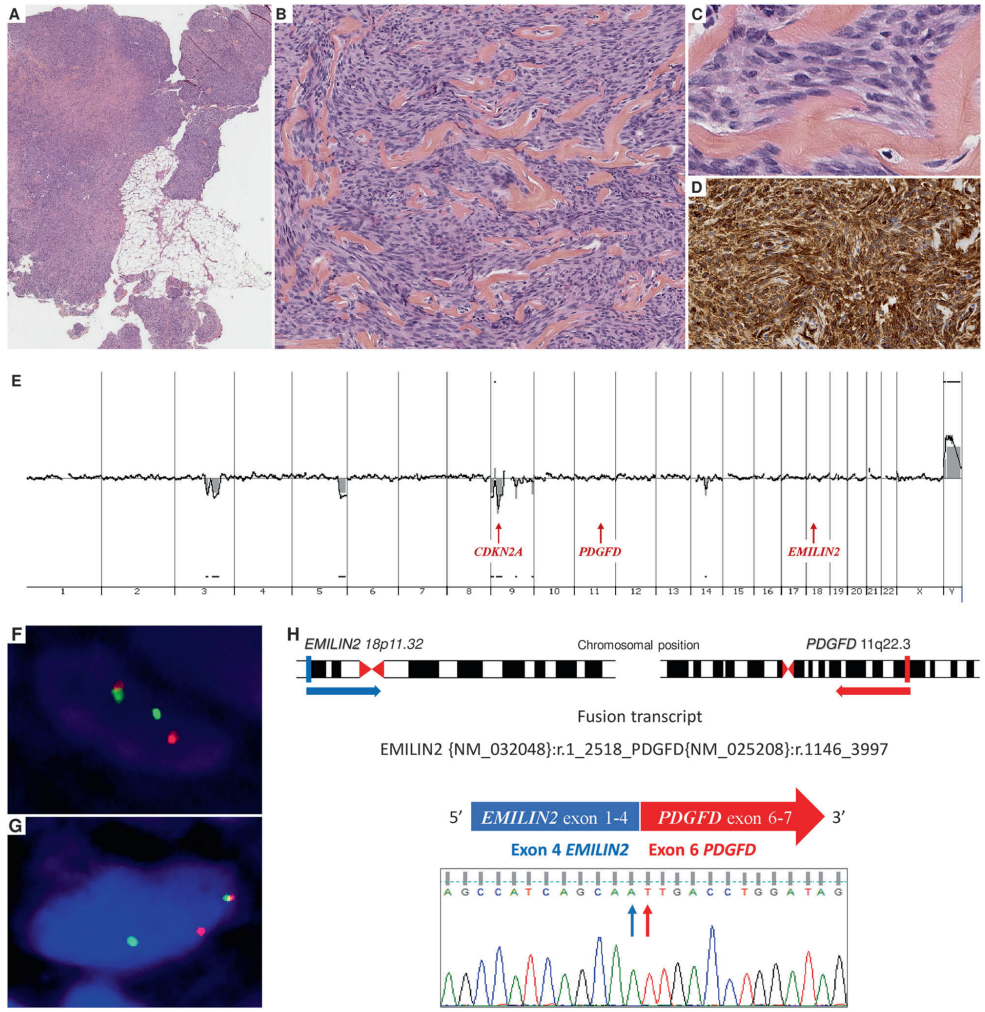


Fig. 3 *EMILIN2-PDGFD* dermatofibrosarcoma protuberans. **a** Biopsy specimen highlighting densely cellular tumor infiltrating hypodermis (case number 9, HES staining, ×2 magnification). **b** Focal collagenized stroma arranged in elongated fascicles suspicious for fibrosarcomatous transformation (×200 magnification). **c** Tumor cells display ovoid monomorphic nuclei embedded in a collagenous stroma (×400 magnification). **d** CD34 staining in this case (×250 magnification). **e** Array-comparative genomic hybridization profile showing a homozygous deletion of *CDKN2A* (P16) but no breakpoint within *EMILIN2* or *PDGFD* loci. **f–g** Fluorescence in situ hybridization (FISH) using *EMILIN2* break-apart probe **f** and *PDGFD* break-apart probe **g** showing a rearrangement of the *EMILIN2* and *PDGFD loci*, respectively. **h** Structure of *EMILIN2-PDGFD* fusion. From top to bottom scheme representing locus and chromosomal position of *EMILIN2* and *PDGFD*; schematic of breakpoints positions which constantly occurred in exon 4 of *EMILIN2* and in exon 6 of *PDGFD*; nucleotide sequence of adjoined sequences

genomic loss were seen in 1q22–23 6q12-q26 and 8p in one case each (Supplementary Table 3).

Array-comparative genomic hybridization analysis of the dermatofibrosarcomas protuberans with *COL6A3-PDGFD* fusion highlighted no breakpoint within *PDGFB* nor *COL1A1* but within *COL6A3* and *PDGFD* in 4/7 and 6/8 cases, respectively. Copy number alterations were present in the regions flanking the breakpoints in all but one case ($n =$

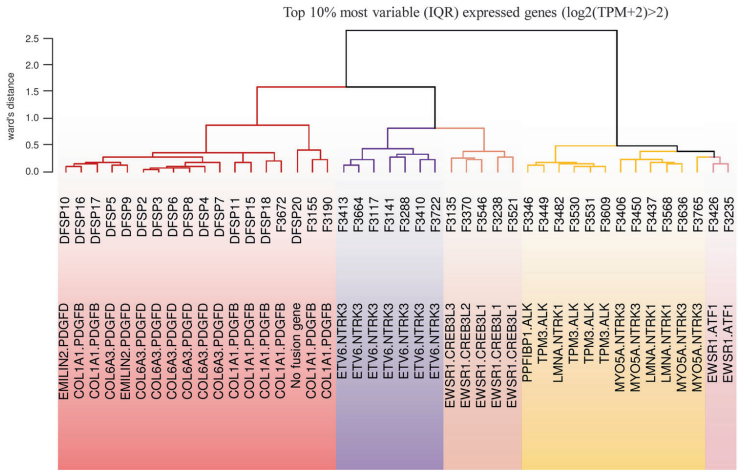


Fig. 4 Transcriptomic classification of dermatofibrosarcoma protuberans. Unsupervised hierarchical clustering, using the top 10% most variant genes based on interquantile range, comparing *PDGFB*-rearranged dermatofibrosarcoma protuberans ($n = 8$, including 3 "conventional" dermatofibrosarcomas protuberans positive for *PDGFB* rearrangement using FISH analysis), *PDGFD*-rearranged dermatofibrosarcoma protuberans ($n = 9$), infantile fibrosarcomas ($n = 7$), clear cell sarcomas of soft tissue (superficially seated) ($n = 2$), *ALK*-rearranged Spitzoid neoplasms ($n = 5$), and *NTRK1/3*-rearranged Spitzoid neoplasms (n = 7). Branches depicted with colors within the dendrogram were found significant (p value < 10e-5) by SigClust clustering significance assessment. Samples are listed in Table 1 and Supplementary Table 4. Infantile fibrosarcomas are highlighted in purple, low grade fibromyxoid sarcomas in orange, spitzoid neoplasms in yellow, clear cell sarcomas in light red, and dermatofibrosarcoma protuberans in red

6/7). Additional non-recurrent copy number alterations included gain of chromosome 12 ($n = 1/7$), loss of 3q ($n = 1/7$), gains of chromosomes 1, 7, and 17 ($n = 1/7$). The genomic profiles of the two cases of *EMILIN2-PDGFD* dermatofibrosarcoma protuberans displayed breakpoints within *PDGFD* ($n = 1/2$), without copy number alteration involving *EMILIN2*. Interestingly, both profiles presented homozygous deletion of *CDKN2A* locus ($n = 2/2$). None of these cases displayed copy number alterations in the genomic regions flanking the breakpoints.

Clustering analysis

To assess whether *PDGFD*-rearranged dermatofibrosarcoma protuberans were biologically similar to *PDGFB*-rearranged dermatofibrosarcoma protuberans, we compared the RNA-seq expression profiles of samples of our cohort (including five *PDGFB*- and nine *PDGFD*-rearranged dermatofibrosarcoma protuberans) to other spindled cutaneous neoplasms including infantile fibrosarcomas ($n = 7$), clear cell sarcomas of soft tissue ($n = 2$), *ALK*-rearranged Spitzoid neoplasms ($n = 5$), *NTRK1/3*-rearranged Spitzoid neoplasms ($n = 7$) and conventional *PDGFB*-rearranged dermatofibrosarcomas protuberans positive for gene rearrangement upon routine molecular screening (n = 3) (Supplementary Table 4). All cases clustered together with transcriptional profiles distinct from *ALK*- and *NTRK1/3*-rearranged Spitzoid proliferations and infantile fibrosarcomas. The dermatofibrosarcoma protuberans cluster lumped together all dermatofibrosarcoma protuberans including those with *PDGFB* fusions (both cryptic and overt rearrangements) and *PDGFD* fusions (Fig. 4).

Discussion

Dermatofibrosarcoma protuberans has historically been defined by the presence of recurrent *COL1A1-PDGFB* fusions, which lead to the upregulation of PDGFRB signaling through an autocrine activating loop [2,20–22]. However, ~4% of dermatofibrosarcoma protuberans prove negative for the translocation upon routine FISH testing [5].

We evidenced by RNA-sequencing that molecularly unconfirmed dermatofibrosarcoma protuberans displayed "cryptic" *COL1A1-PDGFB* fusions in 40% of cases ($n = 8/20$). All dermatofibrosarcomas protuberans associated with "cryptic" fusions involved previously reported *COL1A1* breakpoints, which position have been shown to vary considerably from exon 7 to exon 49. Interestingly, all cases displayed visible genomic breakpoints within *COL1A1* on quantitative genomic profiles along with and/or gains of the

17q locus flanking the breakpoint ($n = 5/5$) (Fig. 1). Notably, subsequent FISH analyses with *COL1A1* break-apart probe were positive in all but one case ($n = 7/8$). Altogether, our findings support that array-comparative genomic hybridization or screening with *COL1A1* break-apart probe may support the diagnosis of cryptic dermatofibrosarcoma protuberans. This complementary screening strategy remains cost-effective as compared with the cost of RNA-sequencing performed in all cases suspicious of dermatofibrosarcoma protuberans. Cryptic rearrangements have also been reported in other translocation-related sarcomas and are thought to be related to structural variations of the translocation or induction of neo-exons secondary to splicing variations [23–27].

Second, we report herein that 55% ($n = 11/20$) of cytogenetically negative- dermatofibrosarcoma protuberans are associated with alternative rearrangements involving *PDGFD*. Two different types of fusion transcripts were evidenced with the 3′ part of *PDGFD* being fused to the 5′ part either of *COL6A3* ($n = 9/11$) or *EMILIN2* ($n = 2/11$). *PDGFD*, located at 11q22.3, encodes a protein belonging to the same family of *platelet-derived growth factor* than PDGFB [28]. PDGFD is able to bind to PDGF receptor B (PDGFRB) [29, 30]. *PDGFD* has oncogenic properties through promotion of cell proliferation and angiogenesis [28] and its overexpression has been linked to a variety of malignancies [31–34]. The breakpoint was constantly located within exon 6, preserving the binding domain to PDGF receptors in a similar manner than translocations involving *PDGFB*. Interestingly, the loss of exon 6 have been shown to induce truncation owing to a premature stop codon in mice, therefore a breakpoint located downstream exon 6 might jeopardize PDGFD signaling [35]. *COL6A3*, located at 2q37.3, encodes a protein of extracellular matrix involved in cellular adhesion belonging to the same superfamily as *COL1A1* [36]. COL6A3 contains 3152 amino acids-long alpha 3 chains, which have the same sequence as alpha 1 chains and harbor the same triple-helical and von Willebrand factor A domains [36]. Notably, somatic rearrangements of *COL6A3* have been previously reported in tenosynovial giant cell tumors in which *COL6A3* is fused to colony stimulating factor 1 (CSF1) [37, 38]. In addition, overexpression of *COL6A3* has been described in a wide array of malignancies including gastric [39], ovarian [40], colorectal [41, 42], and pancreatic cancers [43]. It has also been correlated to resistance to chemotherapy [40, 44]. The breakpoint occurred in exon 42 of *COL6A3* in all but one case ($n = 8/9$), in contrast to the frequent variations seen in *COL1A1*. All *COL6A3-PDGFD* cases displayed genomic imbalances in regions flanking the breakpoints of *COL6A3* and/or *PDGFD* and three cases displayed additional copy number alterations ($n = 3/7$). The morphological and clinical features of *COL6A3*-rearranged dermatofibrosarcoma protuberans did not differ from those seen in bona fide dermatofibrosarcoma protuberans. The second fusion partner *EMILIN2* was involved in 2/11 cases a. *EMILIN2*, located at *18p11.32–p11.31*, encodes a glycoprotein related to the superfamily of collagen, which contains collagen-like domains [36]. EMILIN2 assembles into multimers, involved in the composition of extracellular matrix [45, 46]. The genomic breakpoints were located constantly in exons 4 and 6 of *EMILIN2* and *PDGFD*, respectively. The fusion preserved preserves key structural domains of *EMILIN2*, including EMI domain, coiled-coil structures and leucine zippers and gC1q domain [45, 46]. *EMILIN2-PDGFD* dermatofibrosarcoma protuberans displayed homozygous deletion of *CDKN2A* in both cases ($n = 2/2$). Interestingly, both cases displayed fibrosarcomatous transformation and were deep-seated within hypodermis without dermal connection. It is notable that homozygous deletions of *CDKN2A* are linked to malignancy in varied tumor types including mesenchymal, melanocytic, and epithelial neoplasms [47–49]. Therefore, this alteration may portend an increased malignant potential in this subset of dermatofibrosarcoma protuberans, which warrants further studies on larger cohorts with longer follow-up to assess whether the behavior of cases of *EMILIN2-PDGFD* dermatofibrosarcoma protuberans is similar or not to classical and *COL6A3-PDGFD*-associated cases.

Altogether, *PDGFD* rearrangements may functionally mimic the biology of *COL1A1-PDGFB* fusions as it involves functionally related genes and display a similar structure. Rearrangements might therefore be targeted by imatinib as classical dermatofibrosarcoma protuberans.

In conclusion, we report herein new and recurrent molecular variants of dermatofibrosarcoma protuberans underlined by *PDGFD* rearrangements, which are clinically, morphologically and molecularly indistinguishable from *COL1A1-PDGFB* dermatofibrosarcoma protuberans. Furthermore, we showed that roughly half of cases of dermatofibrosarcoma protuberans negative upon routine FISH screening are associated with cryptic *COL1A1-PDGFB* rearrangements that may be identified by array-comparative genomic hybridization or FISH with a *COL1A1* break-apart probe. The functional similarities between the classical *COL1A1-PDGFB* and the alternative *COL6A3-PDGFD* or *EMILIN2-PDGFD* transcripts suggest they activate an analogous oncogenic autocrine loop involving PDGFRB signaling.

Acknowledgements We wish to thank pathologists and clinicians who provided samples and clinical follow-up: M Chargeboeuf (Lons le Saunier), C Charon-Barra (Dijon), C Delfour (Montpellier), S Depardieu (Metz), A Leroux (Vandoeuvre les Nancy), A Lipandy (Nimes), E Maubec (Paris), C de Mauroy (Rouen), YM Robin (Lille), JF Titton (Besancon), M Battistella (Paris), and MP Wissler (Villeurbanne). We are grateful to A Bazin, AC Peyron, and A Ribeiro for

their insights and participation in this study. This work was supported by grants from La Ligue contre le Cancer de l'Ain et de Savoie, Institut National du Cancer (NETSARC, RREPS, RESOS).

Compliance with ethical standards

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

References

1. Mentzel T, Pedeutour F, Lazar A and Coindre JM. Dermatofibrosarcoma protuberans. In: Fletcher CDM, Bridge JA, Hogendoorn P and Mertens F, editors. WHO classification of tumours of soft tissue and bone. 4th ed. Lyon: IARC; 2013. p. 77-79
2. Simon MP, Pedeutour F, Sirvent N, et al. Deregulation of the platelet-derived growth factor B-chain gene via fusion with collagen gene COL1A1 in dermatofibrosarcoma protuberans and giant-cell fibroblastoma. Nat Genet. 1997;15:95–98.
3. Saiag P, Grob J-J, Lebbe C, et al. Diagnosis and treatment of dermatofibrosarcoma protuberans. European consensus-based interdisciplinary guideline. Eur J Cancer. 2015;51:2604–8.
4. Stacchiotti S, Pantaleo MA, Negri T, et al. Efficacyand biological activity of imatinib in metastatic dermatofibrosarcoma protuberans (DFSP). Clin Cancer Res. 2016;22:837–46.
5. Karanian M, Pérot G, Coindre J-M, et al. Fluorescence in situ hybridization analysis is a helpful test for the diagnosis of dermatofibrosarcoma protuberans. Mod Pathol. 2015;28:230–7.
6. Salgado R, Llombart B, M Pujol R, et al. Molecular diagnosis of dermatofibrosarcoma protuberans: a comparison between reverse transcriptase-polymerase chain reaction and fluorescence in situ hybridization methodologies. Genes Chromosomes Cancer. 2011;50:510–7.
7. Patel KU, Szabo SS, Hernandez VS, et al. Dermatofibrosarcoma protuberans COL1A1-PDGFB fusion is identified in virtually all dermatofibrosarcoma protuberans cases when investigated by newly developed multiplex reverse transcription polymerase chain reaction and fluorescence in situ hybridization assays. Hum Pathol. 2008;39:184–93.
8. Linn SC, West RB, Pollack JR, et al. Gene expression patterns and gene copy number changes in dermatofibrosarcoma protuberans. Am J Pathol. 2003;163:2383–95.
9. Sirvent N, Maire G, Pedeutour F. Genetics of dermatofibrosarcoma protuberans family of tumors: from ring chromosomes to tyrosine kinase inhibitor treatment. Genes Chromosomes Cancer. 2003;37:1–19.
10. Nakamura I, Kariya Y, Okada E, et al. A novel chromosomal translocation associated With COL1A2-PDGFB gene fusion in dermatofibrosarcoma protuberans: PDGF expression as a new diagnostic tool. JAMA Dermatol. 2015;151:1330–7.
11. Bianchini L, Maire G, Guillot B, et al. Complex t(5;8) involving the CSPG2 and PTK2B genes in a case of dermatofibrosarcoma protuberans without the COL1A1-PDGFB fusion. Virchows Arch Int J Pathol. 2008;452:689–96.
12. Saab J, Rosenthal IM, Wang L, et al. Dermatofibrosarcoma protuberans-like tumor with COL1A1 copy number gain in the absence of t(17;22). Am J Dermatopathol. 2017;39:304–9.
13. Sinovic J, Bridge JA. Translocation (2;17) in recurrent dermatofibrosarcoma protuberans. Cancer Genet Cytogenet. 1994;75:156–7.
14. Sirvent N, Coindre J-M, Maire G, et al. Detection of MDM2-CDK4 amplification by fluorescence in situ hybridization in 200 paraffin-embedded tumor samples: utility in diagnosing adipocytic lesions and comparison with immunohistochemistry and real-time PCR. Am J Surg Pathol. 2007;31:1476–89.
15. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012;7:562–78.
16. Ge H, Liu K, Juan T, et al. FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. Bioinformatics. 2011;27:1922–8.
17. Lågstad S, Zhao S, Hoff AM, et al. chimeraviz: a tool for visualizing chimeric RNA. Bioinformatics. 2017;33:2954–6.
18. Bray NL, Pimentel H, Melsted P, et al. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34:525–7.
19. Le Loarer F, Watson S, Pierron G, et al. SMARCA4 inactivation defines a group of undifferentiated thoracic malignancies transcriptionally related to BAF-deficient sarcomas. Nat Genet. 2015;47:1200–5.
20. Greco A, Fusetti L, Villa R, et al. Transforming activity of the chimeric sequence formed by the fusion of collagen gene COL1A1 and the platelet derived growth factor b-chain gene in dermatofibrosarcoma protuberans. Oncogene. 1998;17:1313–9.
21. Shimizu A, O'Brien KP, Sjöblom T, et al. The dermatofibrosarcoma protuberans-associated collagen type Ialpha1/platelet-derived growth factor (PDGF) B-chain fusion gene generates a transforming protein that is processed to functional PDGF-BB. Cancer Res. 1999;59:3719–23.
22. Sjöblom T, Shimizu A, O'Brien KP, et al. Growth inhibition of dermatofibrosarcoma protuberans tumors by the platelet-derived growth factor receptor antagonist STI571 through induction of apoptosis. Cancer Res. 2001;61:5778–83.
23. Brassesco MS, Cortez MA, Valera ET, et al. Cryptic SYT/SXX1 fusion gene in high-grade biphasic synovial sarcoma with unique complex rearrangement and extensive BCL2 overexpression. Cancer Genet Cytogenet. 2010;196:189–93.
24. Torres L, Lisboa S, Cerveira N, et al. Cryptic chromosome rearrangement resulting in SYT-SSX2 fusion gene in a monophasic synovial sarcoma. Cancer Genet Cytogenet. 2008;187:45–49.
25. Lestou VS, O'Connell JX, Robichaud M, et al. Cryptic t(X;18), ins(6;18), and SYT-SSX2 gene fusion in a case of intraneural monophasic synovial sarcoma. Cancer Genet Cytogenet. 2002;138:153–6.
26. Kovar H, Jugovic D, Melot T, et al. Cryptic exons as a source of increased diversity of Ewing tumor-associated EWS-FLI1 chimeric products. Genomics. 1999;60:371–4.
27. Chen S, Deniz K, Sung Y-S, et al. Ewing sarcoma with ERG gene rearrangements: a molecular study focusing on the prevalence of FUS-ERG and common pitfalls in detecting EWSR1-ERG fusions by FISH. Genes Chromosomes Cancer. 2016;55:340–9.
28. Wang Z, Kong D, Li Y, et al. PDGF-D signaling: a novel target in cancer therapy. Curr Drug Targets. 2009;10:38–41.
29. Bergsten E, Uutela M, Li X, et al. PDGF-D is a specific, protease-activated ligand for the PDGF beta-receptor. Nat Cell Biol. 2001;3:512–6.
30. LaRochelle WJ, Jeffers M, McDonald WF, et al. PDGF-D, a new protease-activated growth factor. Nat Cell Biol. 2001;3:517–21.
31. Torres-Martin M, Lassaletta L, Isla A, et al. Global expression profile in low grade meningiomas and schwannomas shows upregulation of PDGFD, CDH1 and SLIT2 compared to their healthy tissue. Oncol Rep. 2014;32:2327–34.
32. Ustach CV, Kim H-RC. Platelet-derived growth factor D is activated by urokinase plasminogen activator in prostate carcinoma cells. Mol Cell Biol. 2005;25:6279–88.
33. Wang Z, Kong D, Banerjee S, et al. Down-regulation of platelet-derived growth factor-D inhibits cell growth and angiogenesis through inactivation of Notch-1 and nuclear factor-kappaB signaling. Cancer Res. 2007;67:11377–85.

34. Xu L, Tong R, Cochran DM, et al. Blocking platelet-derived growth factor-D/platelet-derived growth factor receptor beta signaling inhibits human renal cell carcinoma progression in an orthotopic mouse model. Cancer Res. 2005;65:5711–9.
35. Reigstad LJ, Varhaug JE, Lillehaug JR. Structural and functional specificities of PDGF-C and PDGF-D, the novel members of the platelet-derived growth factors family. FEBS J. 2005;272:5723–41.
36. Ricard-Blum S. The collagen family. Cold Spring Harb Perspect Biol. 2011;3:a004978.
37. West RB, Rubin BP, Miller MA, et al. A landscape effect in tenosynovial giant-cell tumor from activation of CSF1 expression by a translocation in a minority of tumor cells. Proc Natl Acad Sci USA. 2006;103:690–5.
38. Möller E, Mandahl N, Mertens F, et al. Molecular identification of COL6A3-CSF1 fusion transcripts in tenosynovial giant cell tumors. Genes Chromosomes Cancer. 2008;47:21–25.
39. Xie X, Liu X, Zhang Q, et al. Overexpression of collagen VI α3 in gastric cancer. Oncol Lett. 2014;7:1537–43.
40. Sherman-Baust CA, Weeraratna AT, Rangel LBA, et al. Remodeling of the extracellular matrix through overexpression of collagen VI contributes to cisplatin resistance in ovarian cancer cells. Cancer Cell. 2003;3:377–86.
41. Qiao J, Fang C-Y, Chen S-X, et al. Stroma derived COL6A3 is a potential prognosis marker of colorectal carcinoma revealed by quantitative proteomics. Oncotarget. 2015;6:29929–46.
42. Yu J, Wu WKK, Li X, et al. Novel recurrently mutated genes and a prognostic mutation signature in colorectal cancer. Gut. 2015;64:636–45.
43. Arafat H, Lazar M, Salem K, et al. Tumor-specific expression and alternative splicing of the COL6A3 gene in pancreatic cancer. Surgery. 2011;150:306–15.
44. Park J, Morley TS, Scherer PE. Inhibition of endotrophin, a cleavage product of collagen VI, confers cisplatin sensitivity to tumours. EMBO Mol Med. 2013;5:935–48.
45. Mongiat M, Mungiguerra G, Bot S, et al. Self-assembly and supramolecular organization of EMILIN. J Biol Chem. 2000;275:25471–80.
46. Colombatti A, Spessotto P, Doliana R, et al. The EMILIN/Multimerin family. Front Immunol. 2011;2:93.
47. Hwang HC, Sheffield BS, Rodriguez S, et al. Utility of BAP1 immunohistochemistry andp16 (CDKN2A) FISH in the diagnosis of malignant mesothelioma in effusion cytology specimens. Am J Surg Pathol. 2016;40:120–6.
48. Li Z, Gonzalez CL, Wang B, et al. Cdkn2a suppresses metastasis in squamous cell carcinomas induced by the gain-of-function mutantp53(R172H). J Pathol. 2016;240:224–34.
49. Rebouissou S, Hérault A, Letouzé E, et al. CDKN2A homozygous deletion is associated with muscle invasion in FGFR3-mutated urothelial bladder carcinoma. J Pathol. 2012;227:315–24.

Affiliations

Bérengère Dadone-Montaudié[1] · Laurent Alberti[2,3] · Adeline Duc[3] · Lucile Delespaul[4,5,11,11] · Tom Lesluyes[4,5,11,11] · Gaëlle Pérot[6] · Agnès Lançon[3] · Sandrine Paindavoine[3] · Ilaria Di Mauro[1] · Jean-Yves Blay[2,7] · Arnaud de la Fouchardière[3] · Frédéric Chibon[4,6,11,11] · Marie Karanian[3] · Gaëtan MacGrogan[6] · Valérie Kubiniek[1] · Frédérique Keslair[1] · Nathalie Cardot-Leccia[8] · Audrey Michot[9] · Virginie Perrin[10] · Yanis Zekri[10] · Jean-Michel Coindre[5,6] · Franck Tirode[2,10] · Florence Pedeutour[1] · Dominique Ranchère-Vince[3] · François Le Loarer[5,6] · Daniel Pissaloux[2,3]

1 Laboratory of Solid Tumors Genetics, Institute for Research on Cancer and Aging of Nice (IRCAN) CNRS UMR 7284/INSERM U1081, Université Côte d'Azur, Centre Hospitalier Universitaire de Nice, Nice, France

2 Cancer Research Center of Lyon, Université Claude Bernard Lyon 1, CNRS 5286, INSERM U1052, Lyon, France

3 Department of Biopathology, Centre Léon Bérard, Lyon, France

4 INSERM U1218 ACTION, Bordeaux, France

5 University of Bordeaux, Bordeaux, France

6 Department of Pathology, Institut Bergonié, Bordeaux, France

7 Department of Oncology, Centre Léon Bérard, Lyon, France

8 Central Laboratory of Pathology, Centre Hospitalier Universitaire de Nice, Nice, France

9 Department of Surgery, Institut Bergonié, Bordeaux, France

10 Department of Translational Research and Innovation, Centre Léon Bérard, Lyon, France

11 INSERM U1037, Cancer Research Center of Toulouse (CRCT), Institut Claudius Regaud, IUCT-Oncopole, Toulouse, France

VI – Le transcrit de fusion *GREB1-CTNNB1* détecté par RNA-seq dans une UTROSCT : un nouveau réarrangement de *CTNNB1*

Received: 18 July 2018 | Revised: 14 October 2018 | Accepted: 16 October 2018
DOI: 10.1002/gcc.22694

WILEY

RESEARCH ARTICLE

GREB1-CTNNB1 fusion transcript detected by RNA-sequencing in a uterine tumor resembling ovarian sex cord tumor (UTROSCT): A novel *CTNNB1* rearrangement

Sabrina Croce[1] | Tom Lesluyes[2,3,4,5] | Lucile Delespaul[3,4] | Benjamin Bonhomme[1] | Gaëlle Pérot[1] | Valérie Velasco[1] | Laetitia Mayeur[1] | Flora Rebier[1] | Houda Ben Rejeb[1] | Frédéric Guyon[6] | W Glenn McCluggage[7] | Anne Floquet[8] | Denis Querleu[3,6] | Camille Chakiba[8] | Mojgan Devouassoux-Shisheboran[9] | Eliane Mery[5] | Laurent Arnould[10] | Gerlinde Averous[11] | Isabelle Soubeyran[1] | Sophie Le Guellec[4,5] | Frédéric Chibon[4,5]

[1]Department of Biopathology, Institut Bergonié, Comprehensive Cancer Center, Bordeaux, France
[2]Comprehensive Cancer Center, INSERM U1218, Institut Bergonié, Bordeaux, France
[3]University of Bordeaux, Bordeaux, France
[4]Cancer Research Center of Toulouse, Oncosarc, INSERM UMR1037, Toulouse, France
[5]Department of Pathology, Institut Claudius Regaud, IUCT-Oncopole, Toulouse, France
[6]Department of Surgery, Institut Bergonié, Comprehensive Cancer Center, Bordeaux, France
[7]Department of Pathology, Belfast Health and Social Care Trust, Belfast, United Kingdom
[8]Department of Oncology, Institut Bergonié, Comprehensive Cancer Center, Bordeaux, France
[9]Department of Pathology, CHU Lyon Sud, Pierrebenite, France
[10]Department of Pathology, Centre JF Leclerc, Comprehensive Cancer Center, Dijon, France
[11]Department of Pathology, CHU Strasbourg, France

Correspondence
Sabrina Croce, Department of Biopathology, Institut Bergonié, Comprehensive anticancer center, 229 cours de l'Argonne, Bordeaux F-33000, France.
Email: s.croce@bordeaux.unicancer.fr

Mutations of *CTNNB1* have been implicated in tumorigenesis in many organs. However, tumors harboring a *CTNNB1* translocation are extremely rare and this translocation has never been reported in a uterine mesenchymal neoplasm. We report a novel translocation t(2;3)(p25;p22) involving the *GREB1* (intron 8) and *CTNNB1* (exon 3) in a uterine tumor resembling ovarian sex cord tumor (UTROSCT), which exhibited extrauterine metastasis. The translocation detected by RNA-sequencing was validated by RT-PCR, and resulted in nuclear expression of β-catenin. Juxta-positioning with *GREB1*, which is overexpressed in response to estrogens, resulted in overexpression of a truncated and hypophosphorylated nuclear β-catenin in the primary and recurrent tumors. This accumulation of nuclear β-catenin results in a constitutive activation of the Wnt/β-catenin signaling pathway with a major oncogenic effect. The *CTNNB1* gene fusion, promoted by an estrogen-responsive gene (*GREB1*), could be a potential driver of tumorigenesis in this case and a therapeutic target with adapted inhibitors. RT-PCR and immunohistochemistry performed on 11 additional UTROSCTs showed no CTNNB1 fusion transcript or nuclear β-catenin immunoreactivity.

KEYWORDS
cytogenetic, molecular biology, pathology, translocation, UTROSCT

1 | INTRODUCTION

Uterine tumor resembling ovarian sex cord tumor (UTROSCT) is a rare mesenchymal neoplasm predominantly arising in perimenopausal and postmenopausal women. The 2014 World Health Organization (WHO) classification of tumors of the female reproductive organs includes these tumors in the category of endometrial stromal and related neoplasms and defines these as "neoplasms that resemble ovarian sex cord tumors without a component of recognizable endometrial stroma".[1] First described in 1976 by Clement and Scully,[2] they

can exhibit a wide array of morphologic features (including diffuse, trabecular, tubular, nested, and retiform architecture) and typically have a polyphenotypic immunophenotype with variable expression of epithelial, smooth muscle, endometrial stromal and sex cord markers and hormone receptors.[2–8] Clinically, these are regarded as tumors of uncertain but low malignant potential.[9] In a recently reported series of 34 cases, 23.4% developed extrauterine recurrence and 8.8% died of disease, although the authors noted that the high percentage of cases exhibiting malignant behavior may reflect a degree of referral bias as some of the cases were referred for specialist review because of malignant behavior.[10]

Only a few studies have investigated UTROSCT at the molecular level.[11–14] UTROSCT does not harbor the characteristic rearrangements seen in low-grade and high-grade endometrial stromal sarcomas (ESSs)[14,15] and does not contain *DICER1* or *FOXL2* mutations, these mutations being characteristic of ovarian Sertoli-Leydig cell tumor and adult granulosa cell tumor, respectively.[12] To the best of our knowledge, no rearrangement has been reported in UTROSCT with the exception of one case harboring t(X;6)(p22.3;q23.1) and t(4;18)(q21.1;q21.3) translocations without identified genes.[13]

Herein, we report the first case of UTROSCT with an identified specific fusion transcript involving Growth Regulation by Estrogen in Breast cancer 1 (*GREB1*) and catenin (cadherin-associated protein), beta 1 (*CTNNB1*) genes; the transcript was identified by high-throughput paired-end RNA sequencing.

2 | MATERIAL AND METHODS

2.1 | Independent series of 11 UTROSCTs

Eleven additional cases of UTROSCT from the Biopathology Department of Institut Bergonié, CHU Lyon Sud, IUCT Oncopole Toulouse, CHU Strasbourg and Anticancer Research Center of Dijon were collected. The cases were reviewed by one of the authors (SC) and the diagnoses confirmed.

2.2 | RNA extraction and high-throughput paired-end RNA-sequencing

Total RNA was extracted from the primary tumor and the adjacent normal myometrium (frozen material). RNA extractions (around 50 mg of tumor) were performed using a standard TRIzol (Life Technologies, Carlsbad, CA)/chloroform extraction followed by 70% ethanol precipitation and RNA purification using the RNeasy Mini Kit (Qiagen, Hilden, Germany) with a DNase treatment (RNase-Free DNase Set, Qiagen, Hilden, Germany). RNA was quantified using a Nanodrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA) and quality checked with the Agilent 2100 Bioanalyzer (Agilent, Technologies, Santa Clara, CA) using the Agilent RNA 6000 Nano Kit (Agilent) according to the manufacturer's instructions. An ERCC RNA Spike-In Mix (Life technologies) was added to RNA as recommended by the manufacturer.

Each library was sequenced using TruSeq Stranded mRNA prep kit (Illumina Inc., San Diego, CA) in paired-end mode with a read length of 76 bp. A minimum of 20 million paired-end reads were produced using a HiSeq2000 (Illumina Inc.) following the manufacturer's protocol. Image analysis, base calling and base quality scoring of the run were processed by integrated primary analysis software—Real Time Analysis and followed by generation of FASTQ sequence files by CASAVA. We used the deFuse algorithm (v0.6.1; Ref. 16) to look for gene fusions on FastQ files with ENSEMBL GRCh37.74 annotations, with breakpoint homology <30 and maximum identity of fusion sequence alignments to EST, EST island, and cDNA <10%. Gene fusions fulfilling the following conditions were retained: probability >50%, non-read-through, non-read-through-like (same chromosome, indicated as deletion, coding/coding gene region, and genomic distance <20 000) and breakpoint homology <30. Expression quantification was performed as previously described.[17] Gene set enrichment analysis (GSEA) was performed on expressed genes (FPKM > 0; n = 17 178) ranked by decreasing expression difference between tumor and surrounding normal tissue (normal myometrium).[18] We ran a GSEA for each different signature database in MSigDB version 6.1.[19,20]

2.3 | Reverse transcriptase PCR and Sanger sequencing for chimeric transcript validation

Reverse transcriptase PCR (RT-PCR) was performed on the primary tumor and its recurrence as well as on the 11 additional UTROSCTs. Total RNA was reverse transcribed using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions.

For PCR, primers were designed using the Primer 3 program (http://frodo.wi.mit.edu/primer3/) and are presented in Supporting Information Table S1. Touch-down 60°C program was used (TD 60°C; 2 cycles at a 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). PCR was performed on 50 ng of cDNA using AmpliTaqGold DNA polymerase (Applied Biosystems). PCR products were then purified using ExoSAP-IT PCR Purification Kit (GE Healthcare, Chicago, IL) and sequencing reactions were performed with the Big Dye Terminator V1.1 Kit (Applied Biosystems) according to the manufacturer's recommendations. Samples were then purified using the Big Dye XTerminator Purification kit (Applied Biosystems) according to the manufacturer's instructions and sequencing was performed on a 3130xl Genetic Analyzer (Applied Biosystems). Sequence analyses were performed with SeqScape software v2.5 (Applied Biosystems).

2.4 | DNA extraction and array-comparative genomic hybridization analysis

Genomic DNA was extracted from frozen tissues using QIAmp DNA Mini kit according to the manufacturer's protocol for DNA isolation (Agilent Technologies). A cut-off of 50% of cellularity in tumor samples was set for the analysis. DNA was hybridized onto 8 × 60 K whole-genome arrays (G4450A; Agilent Technologies) according to the manufacturer's protocol. Microarray slides were scanned using a DNA Microarray Scanner and images were analyzed by Feature

Extraction V10.1.1.1 followed by Agilent Cytogenomic software 4.0. The ADM-2 algorithm of the comparative genomic hybridization (CGH) Analytics v4.0.76 software (Agilent Technologies) was used to identify the DNA copy number anomalies at the probe level. A low-level copy number gain was defined as a log 2 ratio > 0.25 and a copy number loss was defined as a log 2 ratio < −0.25. A high-level gain or amplification was defined as a log 2 ratio > 1.5 and a homozygous deletion was suspected when the ratio was <−1. The range for derivative log ratio spread cut-off was fixed to 0.50.

2.5 | *CTNNB1* exon 3 sequencing

Exon 3 of *CTNNB1* was amplified by PCR; details of the primers are presented in Supporting Information Table S2. The PCR condition included an initial denaturation step at 95°C for 10 minutes, 40 cycles of 95°C denaturation for 30 seconds, 62°C annealing for 45 seconds, 72°C elongation for 45 seconds, and a final elongation step at 72°C for 20 minutes. The quality of the PCR products was analyzed by 2% agarose gel electrophoresis. The PCR products were sequenced using a Big Dye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) on a 3130XL Genetic Analyzer (Applied Biosystems) to identify the mutation.

2.6 | β-Catenin immunohistochemistry

Immunohistochemistry was performed on 4 μm thick sections from a representative paraffin block of both the primary and recurrent tumors and on whole tissue sections from the 11 UTROSCTs. Staining was also performed on a tissue microarray (TMA) containing 103 uterine mesenchymal tumors which comprised 31 low-grade ESS, one endometrial stromal nodule, 14 high-grade ESS, 24 undifferentiated uterine sarcomas, 12 leiomyomas, 11 smooth muscle tumors of uncertain malignant potential (STUMP), and 10 leiomyosarcomas. The primary antibody against mouse monoclonal anti-β-catenin (14/β-Catenin clone, from Cell Marque, 760-4242, prediluted) directed against the C-terminus of human β-catenin was incubated for 56 minutes on a Benchmark ULTRA (Roche-Ventana, Tucson, AZ) with detection Kit Optiview DAB IHC (reference: 760-500). Heat-induced antigen retrieval was performed using Cell Ventana Conditioning buffer (CC1 standard), for 64 minutes.

UltraView Ventana was used as revelation system. A desmoid tumor served as a positive external control (nuclear immunoreactivity).

3 | RESULTS

3.1 | Case history and pathological findings

A 70-year-old woman underwent a total hysterectomy with bilateral salpingo-oophorectomy because of a myometrial mass. Gross examination showed a 10 cm yellow, soft and fleshy well-circumscribed myometrial-based mass, which had ruptured the uterine serosa (Figure 1A). Both ovaries and fallopian tubes were grossly normal.

Histologically the myometrial mass comprised a relatively well-circumscribed tumor with a diffuse, tubular, nested, and trabecular growth pattern (Figure 1B). The tumor was composed of epithelioid cells with round to ovoid nuclei and prominent nucleoli. Focally, there was a rhabdoid appearance with eccentric nuclei and abundant eosinophilic cytoplasm (Figure 1C). Elsewhere the tumor cells had clear foamy cytoplasm (Figure 1D). There was little in the way of nuclear atypia and the mitotic rate was low (approximately one mitosis/10 HPF). The tumor cells were set in a fibrous stroma.

The tumor diffusely expressed estrogen receptor (ER), progesterone receptor (PR), and CD10. Desmin, AE1–AE3, epithelial membrane antigen (EMA), CK8/18, calretinin, WT-1, and Melan A were focally positive. H-caldesmon, HMB45, S100, FOXL2, inhibin, myogenin, SALL4, cyclin D1, and BCOR were negative. There was retention of nuclear immunoreactivity with SMARCA4 (BRG1).

The break apart FISH for *JAZF1*, *PHF1*, or *YWHAE* were negative.

The endometrium was atrophic with no hyperplasia or malignancy and the cervix, both ovaries and fallopian tubes were histologically normal.

Based on the morphological features and immunophenotype, a diagnosis of UTROSCT was made.

Seventeen months later, the tumor recurred with widespread pelvic nodules. The patient underwent a posterior exenteration. Morphologically, the tumor was identical to that seen in the original specimen with epithelioid cells in diffuse, corded, and nested arrangements (Figure 1E). The immunophenotype was virtually identical to the original neoplasm. Following the posterior exenteration, the patient was treated with aromatase inhibitors but the tumor did not respond. She developed lung metastases and abdominal peritoneal recurrence a year later.

The original and recurrent neoplasms were reviewed by two gynecological pathologists (SC, WGM) and the diagnosis of UTROSCT confirmed.

The samples from the tumor archives have been declared in the Biological Resources Center of Institut Bergonié, for which the French authorities authorized for scientific research (AC-2008-812) and the patient signed an informed consent for research.

3.2 | RNA-sequencing, CGH analysis and immunohistochemical results

RNA-sequencing on the initial index case identified a fusion transcript *GREB1-CTNNB1* as a product of the translocation t(2;3)(p25;p22). No other in frame rearrangements was detected. This fusion transcript was validated by RT-PCR with specific primers (Table S1, Supporting Information) and was detected in the primary tumor and in the recurrence (Figure 2A). No reciprocal transcript was detected in both (data not shown). The Sanger sequencing analysis of the amplified fusion transcript showed a chimeric transcript consisting of the first eight exons (of 32) plus a small part of intron 8 of *GREB1* gene (NM_014668) and the end of exon 3 of *CTNNB1* gene until its 3′ end. (NM_001098209) (Figure 2B and C). This produces a stop codon in intron 8 of *GREB1*, leading to no or truncated GREB1 protein. It also produces a start codon at the beginning of the exon 4 of *CTNNB1*.The *GREB1-CTNNB1* fusion transcript codes for a truncated form of β-catenin without the 87 N-terminal amino acids that are involved in the protein degradation upon phosphorylation (Refs. 20,21; Figure 2D).

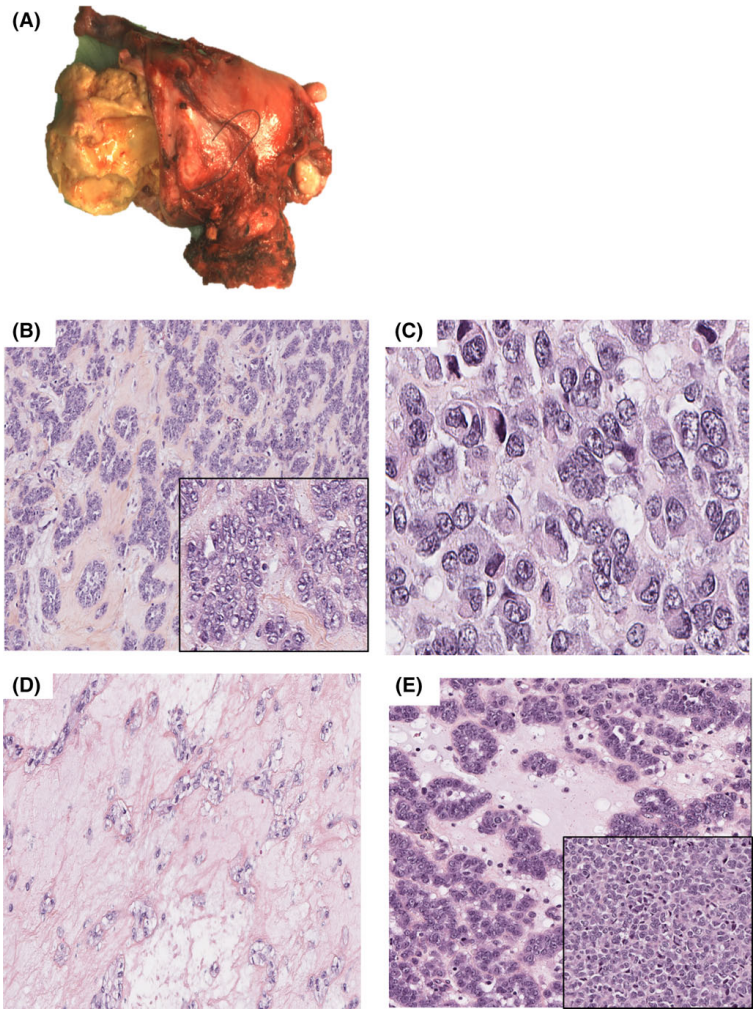


FIGURE 1 A, The UTROSCT comprises a large fleshy soft yellow myometrial mass. B, The primary tumor is composed of epithelioid cells with round nuclei and prominent nucleoli arranged in nested, trabecular, and tubular arrangements. C, Rhabdoid appearance with eccentric nuclei and abundant eosinophilic cytoplasm. D, Occasional cells have foamy cytoplasm. E, Morphologic features of the recurrent tumor that are similar to the primary neoplasm with diffuse, nested, and corded arrangements [Color figure can be viewed at wileyonlinelibrary.com]

Array-CGH performed on the primary tumor and the recurrence showed a very simple profile for the primary tumor (Figure 2E) without evidence of break in chromosome band 2p25 (*GREB1* gene) or in chromosome band 3p22 (*CTNNB1* gene). A more complex genomic profile was observed in the recurrence (Figure 2F) with evidence of breakpoints in chromosome band 3p22 and 2p25 involving both rearranged genes.

Immunohistochemistry was performed to evaluate whether this translocation resulted in overexpression of nuclear β-catenin. This showed strong nuclear positivity in the primary tumor (Figure 3A), as well as in the recurrence (Figure 3B). To rule out the possibility that the nuclear accumulation was caused by *CTNNB1* mutation, we performed Sanger sequencing of exon 3 and observed no β-catenin mutation.

We then performed GSEA between the primary tumor and surrounding normal tissue (adjacent myometrium) (see Section 2). We observed a higher expression of genes belonging to the Wnt/β-catenin signaling pathway in the tumor compared to the surrounding myometrium (FDR q-value = 1.79e-2; Figure 4A). In addition, genes whose proteins interact (selectively and non covalently) with Wnt-protein were also overexpressed in the tumor (FDR q-value = 2.34e-2; Figure 4B). We also observed, using the same method, a lower level

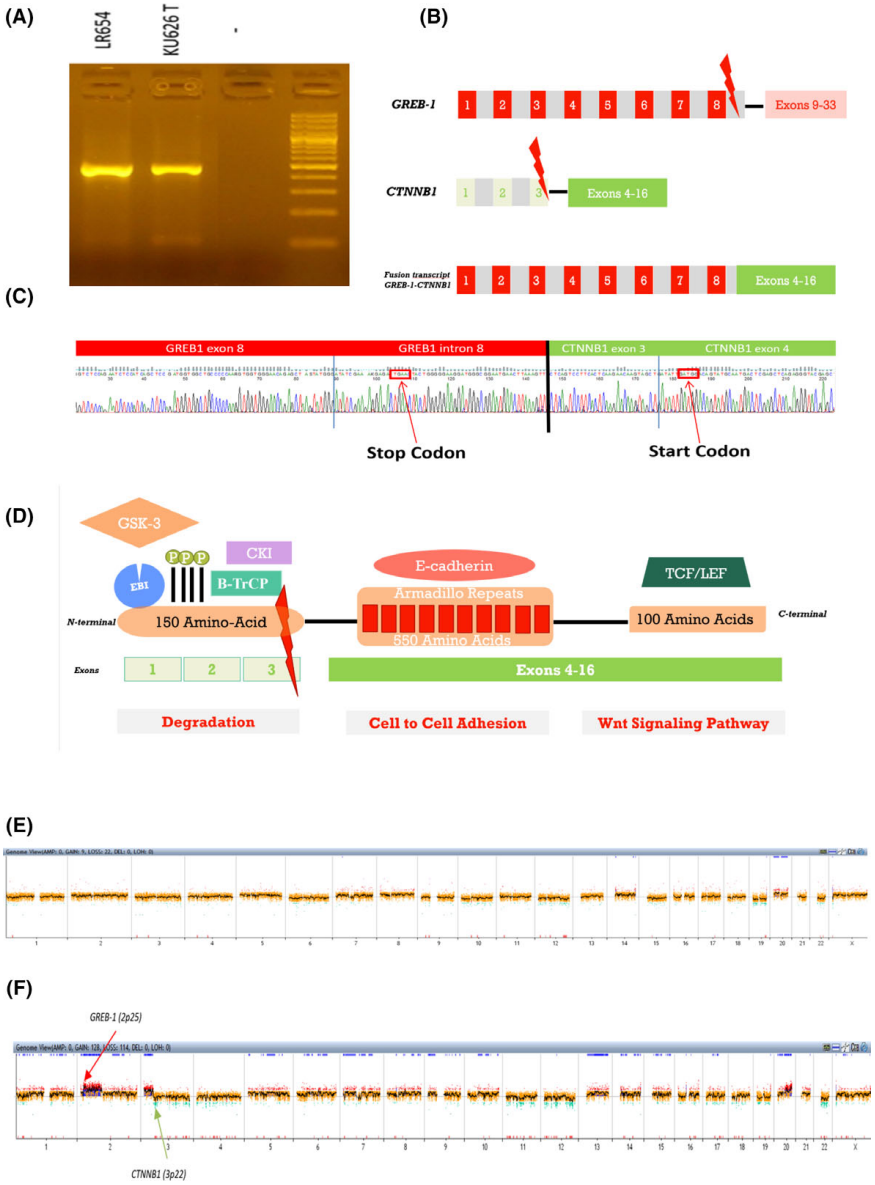


FIGURE 2 A, RT-PCR obtained for GREB1-CTNNB1 chimeric transcript. KU626T: primary tumor; LR 654: recurrence; -: lymphocyte negative control. B, Schematic representation of *GREB1* and *CTNNB1* genes and fusion transcript. *GREB1* and *CTNNB1* exons implicated in translocation are indicated in red and green, respectively. Exons implicated in translocation are shown in light color. Fusion point is represented with a red lightning. C, Partial chromatography electropherogram showing the junction point between intron 8 of *GREB1* gene and exon 4 of *CTNNB1* gene. *GREB1* and *CTNNB1* exons are indicated in red and green, respectively. Point fusion is indicated with black line. D, Primary structure of chimeric protein (*modified from* Ref. 41). Break point is represented with a red lightning. Different partners of *CTNNB1* are indicated. The corresponding exons to different *CTNNB1* domains are represented at bottom of *CTNNB1* protein domains. E, Array-CGH of the primary tumor showing a flat profile. Absence of evidence of breakpoint on *CTNNB1* (Ch 3p22) or *GREB1* (Ch2 p25) genes. F, Array-CGH of recurrence showing a breakpoint on chr 3p22 with *CTNNB1* gene [Color figure can be viewed at wileyonlinelibrary.com]

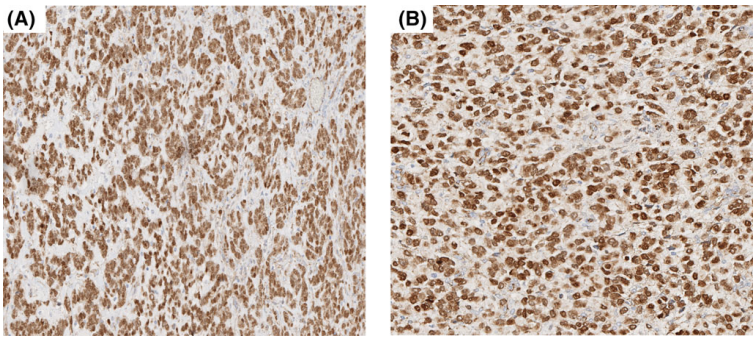


FIGURE 3 A, β catenin nuclear staining in the primary tumor. B, β catenin nuclear staining in the recurrent tumor [Color figure can be viewed at wileyonlinelibrary.com]

of early (Figure 4C) and late (Figure 4D) estrogen response genes (*GREB-1* gene being one of those genes) in the tumor sample as compared with the normal adjacent myometrium.

We performed specific RT-PCR and β-catenin immunohistochemistry in the additional 11 UTROSCTs and observed no fusion transcript or nuclear immunoreactivity (data not shown). In the TMA of 103 uterine mesenchymal tumors, there was no nuclear expression of β-catenin except for a single undifferentiated uterine sarcoma which exhibited nuclear and cytoplasmic immunoreactivity (Supporting Information Figure S1).

4 | DISCUSSION

We report the first case of specific gene fusion in a UTROSCT, namely the *GREB1-CTNNB1* fusion transcript as a product of the translocation t(2;3)(p25;p22). The diagnosis of UTROSCT was established on the basis of the characteristic morphology and a supportive polyphenotypic immunophenotype with expression of CD10, hormone receptors (ER, PR), desmin, epithelial markers and markers which are commonly positive in ovarian sex cord-stromal tumors, including calretinin, WT-1, and Melan A. Inhibin and FOXL2 were negative but these markers are not positive in all UTROSCTs.[10,12]

UTROSCT was originally described as a neoplasm of uncertain but low malignant potential which usually exhibits a benign behavior.[2] In a recent series, one of us (WGM) coauthored the largest published series of cases with follow-up[10]; in that series, 8 of 34 (23.5%) cases, including the index case in this article, exhibited malignant behavior with extrauterine spread. It is possible that the *GREB1-CTNNB1* fusion transcript in this case contributed to the aggressive behavior. Array-CGH showed a very simple genomic profile of the primary tumor and, consonant with the behavior, a more complex profile for the recurrence with evidence of breakpoints in chromosome band 3p22 and 2p25 involving both rearranged genes. This fusion was not found in the series of 11 additional UTROSCTs tested by RT-PCR. There has been limited study of the molecular abnormalities in UTROSCT but in a prior study coauthored by two of us (SC, WGM), mutations in *FOXL2* and *DICER1* were not identified[12]; these mutations are characteristic of ovarian adult granulosa cell tumor and Sertoli-Leydig cell tumor respectively, two neoplasms with morphological and immunophenotypic overlap with UTROSCT.

β-Catenin protein, encoded by the *CTNNB1* gene, is implicated in the canonical Wnt signaling pathway which modulates cell proliferation, cell polarity and differentiation during embryonic development, and regulates the cell-cell adhesion, extra-cellular signals and gene transcription.[21] Wnt-induced β-catenin stabilization and nuclear shuttling results in the T-cell factor/lymphoid enhancer factor (TCF) family of proteins forming a complex with β-catenin, which recruits other coactivators for gene activation via displacement of Groucho.[22,23] β-Catenin acts as a coactivator for TCF and with its transactivation domain activates transcription initiation, histone methyltransferases, histone modification, chromatin modification, and facilitates transcription.[23]

Mutations (in particular in exon 3) are the most frequent mechanisms of alteration in *CTNNB1*, affecting the ubiquitination and resulting in hypophosphorylated β-catenin accumulation and translocation in the nucleus. The resultant effect is the constitutive activation of the Wnt/β-catenin signaling pathway and reprogramming of downstream nuclear transcriptional networks.[24–27] Although some tumors have been reported to harbor 3p21 breakpoints, according to the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (http://cgap.nci.nih.gov/Chromosomes/Mitelman), translocations involving *CTNNB1* are extremely rare[28–30]; the only well-characterized translocation in the literature with *CTNNB1* as a specific partner until now is the t(3;8)(p21;q12) involving *CTNNB1* and *PLAG1* genes in pleomorphic salivary adenoma[31,32]; this translocation results in the activation of *PLAG1* and reduced expression of β-catenin. To our knowledge, t(2;3)(p25;p22) involving *GREB1* and *CTNNB1* has never been reported. Unlike the *CTNNB1-PLAG1* fusion, the chimeric transcript results in the overexpression of β-catenin via excision of the first three exons of *CTNNB1*, wherein the serine phosphorylation sites and the GSK3, CK1 binding sites (Figure 2D), responsible for the degradation of β-catenin, are located.[33] *GREB1* gene, the partner of *CTNNB1* in this fusion, is rearranged in some other types of neoplasia such as T-cell acute lymphoblastic leukemia[34] and prostate carcinoma,[30] and recently rearrangement has been demonstrated in an undifferentiated uterine sarcoma.[35] In our case, given the modality

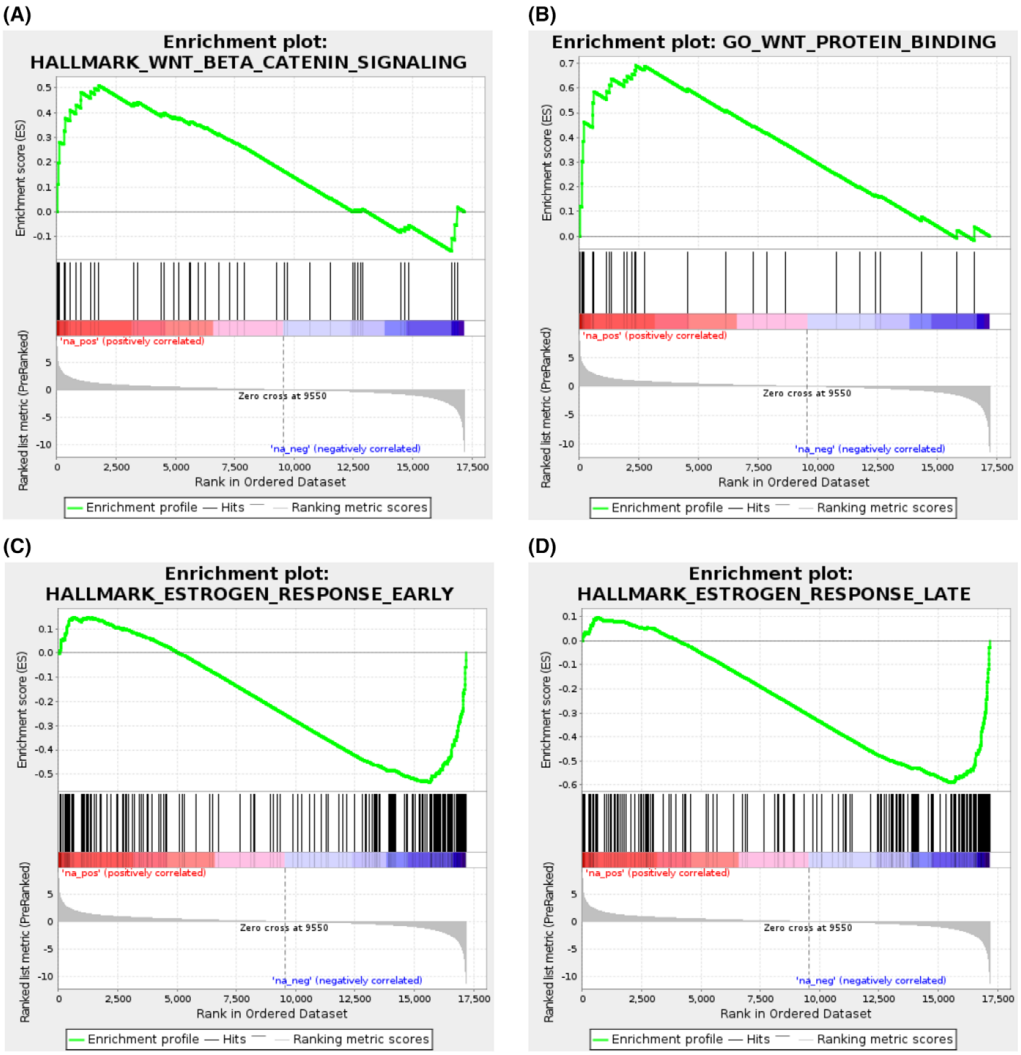


FIGURE 4 A, Gene set enrichment analysis of tumor ("na_pos" on the left side) versus normal adjacent myometrium ("na_neg" on the right side) transcriptomic profiles. This identified an overexpression of genes involved in the Wnt/β-catenin signaling pathway (FDR q-value = 1.79e-2) in the tumor compared to peripheral normal tissues. B, Gene set enrichment analysis of tumor ("na_pos" on the left side) versus normal adjacent myometrium ("na_neg" on the right side) transcriptomic profiles. This indicates an overexpression of genes whose proteins interact (selectively and non covalently) with Wnt-protein (FDR q-value = 2.34e-2) in the tumor compared to peripheral normal tissues. C, Gene set enrichment analysis of normal adjacent myometrium ("na_pos" on the left side) versus tumor ("na_neg" on the right side) transcriptomic profiles. The graph indicates underexpression of genes related to the early response to estrogens in the tumor compared to the peripheral normal tissue. D, Gene set enrichment analysis of normal adjacent myometrium ("na_pos" on the left side) versus tumor ("na_neg" on the right side) transcriptomic profiles. The graph indicates underexpression of genes related to the late response to estrogens in the tumor compared to the peripheral normal tissue [Color figure can be viewed at wileyonlinelibrary.com]

of the rearrangement resulting in a truncated form of *GREB1* in the transcript fusion (Figure 2B), we can assume that GREB1 protein is not synthesized and that the oncogenic effect is driven by truncated β-catenin expression. Nevertheless, this gene has an important role in promoting mRNA expression as it is highly expressed in estrogen-related tissues.[36,37] Furthermore, in our case, we observed that the normal myometrium expressed more of the early (Figure 4C) and late (Figure 4D) estrogen-response genes than the tumor suggesting a

diminution of GREB1 expression. This observation could explain the poor response of the tumor to antihormone therapy (aromatase inhibitors) and the high expression of the translocated β-catenin, under the transcriptomic control of GREB1 promoter.

The nuclear localization of β-catenin as result of *GREB1-CTNNB1* rearrangement was confirmed by immunohistochemistry in the primary tumor and in the recurrence. There was no nuclear expression in the additional 11 UTROSCTs and in the TMA series of 103 uterine mesenchymal tumors with the exception of one undifferentiated uterine sarcoma which exhibited nuclear and cytoplasmic immunoreactivity.

The main limitation of this study is the absence of a cellular model derived from the tumor. Such a model would allow demonstration of the effects of *GREB1-CTNNB1* gene fusion on cellular proliferation and investigation of its driver role in the genesis of this tumor. A possible significance of the *CTNNB1* rearrangement is that β-catenin could be a possible therapeutic target in this tumor. Prior studies have demonstrated inhibition of tumor proliferation in pancreatic neuroendocrine tumors,[38] acute myeloid leukemia,[39] breast and prostate cancer with β-catenin inhibitors.[40]

In summary, we report a novel translocation t(2;3)(p25;p22) involving *GREB1* (intron 8) and *CTNNB1* (exon 3) genes in a UTROSCT. This is the first report of this specific gene fusion in a mesenchymal tumor and in a UTROSCT.

ACKNOWLEDGMENTS

We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for providing help and/or computing and/or storage resources and the "Comité Prévention et Dépistage des Cancers" for founding the purchase of TMA arrayer.

The authors would like to thank Dr Ravi Nookala of Institut Bergonié for the medical writing service.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interests with the contents of this article.

ORCID

Sabrina Croce https://orcid.org/0000-0002-6487-5418
Sophie Le Guellec https://orcid.org/0000-0003-3207-0081
Frédéric Chibon https://orcid.org/0000-0002-6548-2181

REFERENCES

1. Oliva ECM, Carinelli SG, et al. Mesenchymal tumors. Uterine tumour resembling ovarian sex cord tumour. In: CM KRJ, Herrington CS, Young RH, eds. Lyon, France: *WHO Classification of Tumours of Female Reproductive Organs*. 2014:141-147.
2. Clement PB, Scully RE. Uterine tumors resembling ovarian sex-cord tumors. A clinicopathologic analysis of fourteen cases. *Am J Clin Pathol*. 1976;66(3):512-525.
3. Hurrell DP, McCluggage WG. Uterine tumour resembling ovarian sex cord tumour is an immunohistochemically polyphenotypic neoplasm which exhibits coexpression of epithelial, myoid and sex cord markers. *J Clin Pathol*. 2007;60(10):1148-1154.
4. Irving JA, Carinelli S, Prat J. Uterine tumors resembling ovarian sex cord tumors are polyphenotypic neoplasms with true sex cord differentiation. *Mod Pathol*. 2006;19(1):17-24.
5. de Leval L, Lim GS, Waltregny D, Oliva E. Diverse phenotypic profile of uterine tumors resembling ovarian sex cord tumors: an immunohistochemical study of 12 cases. *Am J Surg Pathol*. 2010;34(12):1749-1761.
6. Czernobilsky B. Uterine tumors resembling ovarian sex cord tumors: an update. *Int J Gynecol Pathol*. 2008;27(2):229-235.
7. Krishnamurthy S, Jungbluth AA, Busam KJ, Rosai J. Uterine tumors resembling ovarian sex-cord tumors have an immunophenotype consistent with true sex-cord differentiation. *Am J Surg Pathol*. 1998;22(9):1078-1082.
8. Baker RJ, Hildebrandt RH, Rouse RV, Hendrickson MR, Longacre TA. Inhibin and CD99 (MIC2) expression in uterine stromal neoplasms with sex-cord-like elements. *Hum Pathol*. 1999;30(6):671-679.
9. Bakula-Zalewska E, Danska-Bidzinska A, Kowalewska M, Piascik A, Nasierowska-Guttmejer A, Bidzinski M. Uterine tumors resembling ovarian sex cord tumors, a clinicopathologic study of six cases. *Ann Diagn Pathol*. 2014;18(6):329-332.
10. Moore M, McCluggage WG. Uterine tumour resembling ovarian sex cord tumour (UTROSCT): first report of a large series with follow-up. *Histopathology*. 2017;71:751-759.
11. Chiang S, Staats PN, Senz J, et al. FOXL2 mutation is absent in uterine tumors resembling ovarian sex cord tumors. *Am J Surg Pathol*. 2015;39(5):618-623.
12. Croce S, de Kock L, Boshari T, et al. Uterine tumor resembling ovarian sex cord tumor (UTROSCT) commonly exhibits positivity with sex cord markers FOXL2 and SF-1 but lacks FOXL2 and DICER1 mutations. *Int J Gynecol Pathol*. 2016;35(4):301-308.
13. Wang J, Blakey GL, Zhang L, Bane B, Torbenson M, Li S. Uterine tumor resembling ovarian sex cord tumor: report of a case with t(X;6)(p22.3;q23.1) and t(4;18)(q21.1;q21.3). *Diagn Mol Pathol*. 2003;12(3):174-180.
14. Staats PN, Garcia JJ, Dias-Santagata DC, et al. Uterine tumors resembling ovarian sex cord tumors (UTROSCT) lack the JAZF1-JJAZ1 translocation frequently seen in endometrial stromal tumors. *Am J Surg Pathol*. 2009;33(8):1206-1212.
15. D'Angelo E, Ali RH, Espinosa I, et al. Endometrial stromal sarcomas with sex cord differentiation are associated with PHF1 rearrangement. *Am J Surg Pathol*. 2013;37(4):514-521.
16. McPherson A, Hormozdiari F, Zayed A, et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput Biol*. 2011;7(5):e1001138.
17. Lesluyes T, Perot G, Largeau MR, et al. RNA sequencing validation of the complexity INdex in SARComas prognostic signature. *Eur J Cancer*. 2016;57:104-111.
18. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102(43):15545-15550.
19. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-1740.
20. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417-425.
21. Logan CY, Nusse R. The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol*. 2004;20:781-810.
22. Daniels DL, Weis WI. Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nat Struct Mol Biol*. 2005;12(4):364-371.
23. MacDonald BT, Tamai K, He X. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev Cell*. 2009;17(1):9-26.
24. Korinek V, Barker N, Morin PJ, et al. Constitutive transcriptional activation by a beta-catenin-Tcf complex in $APC^{-/-}$ colon carcinoma. *Science*. 1997;275(5307):1784-1787.
25. Morin PJ, Sparks AB, Korinek V, et al. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science*. 1997;275(5307):1787-1790.

26. Xia J, Urabe K, Moroi Y, et al. beta-Catenin mutation and its nuclear localization are confirmed to be frequent causes of Wnt signaling pathway activation in pilomatricomas. *J Dermatol Sci*. 2006;41(1):67-75.
27. Lax SF. Molecular genetic pathways in various types of endometrial carcinoma: from a phenotypical to a molecular-based classification. *Virchows Arch*. 2004;444(3):213-223.
28. Guseva NV, Jaber O, Tanas MR, et al. Anchored multiplex PCR for targeted next-generation sequencing reveals recurrent and novel USP6 fusions and upregulation of USP6 expression in aneurysmal bone cyst. *Genes Chromosomes Cancer*. 2017;56(4):266-277.
29. Liu Y, Easton J, Shao Y, et al. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat Genet*. 2017;49(8):1211-1218.
30. Yoshihara K, Wang Q, Torres-Garcia W, et al. The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene*. 2015;34(37):4845-4854.
31. Kas K, Voz ML, Roijer E, et al. Promoter swapping between the genes for a novel zinc finger protein and beta-catenin in pleiomorphic adenomas with t(3;8)(p21;q12) translocations. *Nat Genet*. 1997;15(2):170-174.
32. Kas K, Voz ML, Hensen K, Meyen E, Van de Ven WJ. Transcriptional activation capacity of the novel PLAG family of zinc finger proteins. *J Biol Chem*. 1998;273(36):23026-23032.
33. Fukuchi T, Sakamoto M, Tsuda H, Maruyama K, Nozawa S, Hirohashi S. Beta-catenin mutation in carcinoma of the uterine endometrium. *Cancer Res*. 1998;58(16):3526-3528.
34. Atak ZK, Gianfelici V, Hulselmans G, et al. Comprehensive analysis of transcriptome variation uncovers known and novel driver events in T-cell acute lymphoblastic leukemia. *PLoS Genet*. 2013;9(12):e1003997.
35. Brunetti M, Panagopoulos I, Gorunova L, Davidson B, Heim S, Micci F. RNA-sequencing identifies novel GREB1-NCOA2 fusion gene in a uterine sarcoma with the chromosomal translocation t(2;8)(p25;q13). *Genes Chromosomes Cancer*. 2018;57(4):176-181.
36. Camden AJ, Szwarc MM, Chadchan SB, et al. Growth regulation by estrogen in breast cancer 1 (GREB1) is a novel progesterone-responsive gene required for human endometrial stromal decidualization. *Mol Hum Reprod*. 2017;23(9):646-653.
37. Pellegrini C, Gori I, Achtari C, et al. The expression of estrogen receptors as well as GREB1, c-MYC, and cyclin D1, estrogen-regulated genes implicated in proliferation, is increased in peritoneal endometriosis. *Fertil Steril*. 2012;98(5):1200-1208.
38. Jiang X, Cao Y, Li F, et al. Targeting beta-catenin signaling for therapeutic intervention in MEN1-deficient pancreatic neuroendocrine tumours. *Nat Commun*. 2014;5:5809.
39. Fiskus W, Sharma S, Saha S, et al. Pre-clinical efficacy of combined therapy with novel beta-catenin antagonist BC2059 and histone deacetylase inhibitor against AML cells. *Leukemia*. 2015;29(6):1267-1278.
40. Lu W, Li Y. Salinomycin suppresses LRP6 expression and inhibits both Wnt/beta-catenin and mTORC1 signaling in breast and prostate cancer cells. *J Cell Biochem*. 2014;115(10):1799-1807.
41. Gao C, Wang Y, Broaddus R, Sun L, Xue F, Zhang W. Exon 3 mutations of CTNNB1 drive tumorigenesis: a review. *Oncotarget*. 2018;9(4):5492-5508.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Croce S, Lesluyes T, Delespaul L, et al. *GREB1-CTNNB1* fusion transcript detected by RNA-sequencing in a uterine tumor resembling ovarian sex cord tumor (UTROSCT): A novel *CTNNB1* rearrangement. *Genes Chromosomes Cancer*. 2019;58:155–163. https://doi.org/10.1002/gcc.22694

VII – Caractéristiques clinico-pathologiques et moléculaires d'une série de 41 BSNS élargissant leur spectre moléculaire

ORIGINAL ARTICLE

Clinicopathologic and Molecular Features of a Series of 41 Biphenotypic Sinonasal Sarcomas Expanding Their Molecular Spectrum

François Le Loarer, MD, PhD,†‡ Sophie Laffont, MD,* Tom Lesluyes, MSc,†‡§ Franck Tirode, PhD,‖¶ Cristina Antonescu, MD,# Anne-Catherine Baglin, MD,** Lucile Delespaul, MSc,†‡§ Isabelle Soubeyran, MD, PhD,*‡ Isabelle Hostein, PhD,* Gaëlle Pérot, PhD,* Frédéric Chibon, PhD,‡§ Jessica Baud, PhD,‡ Sophie Le Guellec, MD,†† Marie Karanian, MD,‖¶‡‡ Valérie Costes-Martineau, MD,§§ Claire Castain, MD,‖‖ Sandrine Eimer, MD,‖‖ Brigitte Le Bail, MD,†‖‖ Michel Wassef, MD,** and Jean-Michel Coindre, MD*†‡*

Abstract: Biphenotypic sinonasal sarcoma (BSNS) is a locally aggressive tumor occurring in the sinonasal region. It harbors both myogenic and neural differentiation and is characterized by *PAX3* rearrangement with *MAML3* as the most frequent fusion partner, but the partner of *PAX3* remains unidentified in a subset of cases. About 70 cases have been reported so far. In this study, we report a series of 41 cases with clinical, pathologic, and molecular description. Twenty-five (61%) patients were female individuals, and the median age was 49 years. Tumors arose predominantly in the nasal cavity and ethmoidal sinuses. Local recurrences occurred in 8 cases of the 25 (32%). Histologic features were characteristic of BSNS, with 5 cases showing focal rhabdomyoblastic differentiation. Immunohistochemistry showed a constant positivity of S100 protein and PAX3 and negativity of SOX10. MyoD1 was focally positive in 91% of cases, whereas only 20% were positive for myogenin. Molecular analysis showed a *PAX3-MAML3* transcript in 37 cases (90%). RNA sequencing was performed in the 4 negative cases for *PAX3-MAML3* fusion, and it showed that 1 case harbored a *PAX3-FOXO1* fusion, as previously described in the literature, and 2 novel fusions: *PAX3-WWTR1* fusion in 2 cases and *PAX3-NCOA2* fusion in 1 case. RNA sequencing results were confirmed by fluorescence in situ hybridization, reverse transcription-polymerase chain reaction, and Sanger sequencing. The *PAX3-NCOA2*-positive case showed focal rhabdomyoblastic differentiation. In conclusion, we report 2 novel fusions (*PAX3-WWTR1* and *PAX3-NCOA2*) in BSNS and show that MyoD1 is more sensitive than myogenin for demonstrating myogenic differentiation in this tumor.

Key Words: biphenotypic sinonasal sarcoma, PAX3, MAML3, FOXO1, WWTR1, NCOA2

(*Am J Surg Pathol* 2019;00:000–000)

Biphenotypic sinonasal sarcoma (BSNS), also known as low-grade sinonasal sarcoma with neural and myogenic features, is a locally aggressive tumor occurring in the sinonasal region, first delineated at the morphological level by Lewis et al in 2012.[1] In 2014, these tumors were shown to harbor a recurrent *PAX3-MAML3* fusion.[2] In 2016, Huang et al[3] and Wong et al[4] reported, respectively, 2 cases with a *PAX3-NCOA1* and 1 case with a *PAX3-FOXO1* fusion. The same year, Fritchie et al[5] reassessed the Mayo Clinic series initially studied by Lewis and colleagues and Wang and colleagues. In this larger series of 44 SNS, they reported 24 cases with a *PAX3-MAML3* fusion, 3 with a *PAX3-FOXO1* fusion, and 1 with a *PAX3-NCOA1* fusion, whereas 11 cases showed a *PAX3* rearrangement with no defined partner, one case showed a *MAML3* rearrangement with no defined partner, and 4 were negative for *PAX3*, *MAML3*, *FOXO1*, *NCOA1*, and *NCOA2* genes.

We investigated a retrospective and prospective series of 41 cases at the clinical, histologic, immunohistochemical, and molecular levels. Our findings expand the molecular

From the *Department of Pathology, Institut Bergonie; ‡INSERM U1218 ACTION; ‖‖Department of Pathology, CHU de Bordeaux, Bordeaux; †Université de Bordeaux, Institut Bergonié, Talence; §Cancer Research Center of Toulouse; ††Department of Pathology, Institut Claudius Regaud, Institut Universitaire du Cancer Toulouse-Oncopole, Toulouse; ‖INSERM 1052, CNRS 5286, Cancer Research Center of Lyon, University of Lyon, Université Claude Bernard Lyon 1; ¶Department of Translational Research and Innovation, Centre Leon Berard; ‡‡Department of Pathology, Centre Leon Berard, Lyon; **Department of Pathology, Hôpital Lariboisiere, Paris; §§Department of Pathology, CHU de Montpellier, Montpellier, France; and #Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY.

F.L.L. and S.L. contributed equally

M.W. and J.-M.C.: supervised the work.

Conflicts of Interest and Source of Funding: The authors have disclosed that they have no significant relationships with, or financial interest in, any commercial companies pertaining to this article.

Correspondence: Jean-Michel Coindre, MD, Institut Bergonié, 276 cours de l'Argonne, 33000 Bordeaux, France (e-mail: j.coindre@bordeaux.unicancer.fr).

spectrum of these lesions with the description of 2 previously unreported fusion variants, *PAX3-NCOA2* and *PAX3-WWTR1*.

MATERIALS AND METHODS

Selection of Cases

Ethical approval from the appropriate committees was obtained. All sarcoma cases are recorded in the national sarcoma pathology RREPS database, approved by the National Committee for Protection of Personal Data (CNIL, no. 910390), in compliance with the ethical principles of the Helsinki Declaration.

Formalin-fixed paraffin-embedded specimens diagnosed between January 2000 and June 2018 were retrieved from the archives of the pathology departments involved in the French soft tissue (RRePS) and head and neck (REFCOR) pathology networks. Forty-four cases were identified, but 3 cases were excluded, as the available material was not suitable for molecular analysis. The following clinical data were collected: sex, age at diagnosis, location and size of tumor, initial treatment, and follow-up. All cases were microscopically reviewed by 2 pathologists (S.L., J.-M.C.).

Immunohistochemistry

The tissue slides were deparaffinized in xylene, hydrated in alcohol, and baked in a microwave (30 min in Trisbuffer, pH 9). Endogenous peroxidase was blocked. Staining was performed on the Benchmark ultra-automated stainer (Ventana) using diamino-benzidine as chromogen (Dako, Glostrup, Denmark). The following antibodies were used: Pankeratin AE1/AE3 (clone PCK26, prediluted; Ventana Media Systems, Tucson, AZ); EMA (clone E29, prediluted; Ventana Media Systems); S100 protein (clone poly Z311, dilution 1:500; Dako); smooth muscle actin (clone 1A4, dilution 1:12000; Sigma Aldrich, Saint Louis, MO); desmin (clone DE-R-11, prediluted; Ventana Media Systems); myogenin (clone LO 26, dilution 1:20; Leica Biosystems, Buffalo Grove, IL); MyoD1 (clone EP212, prediluted; Cell Marque, Rocklin, CA); CD34 (clone QBEnd10, prediluted; Ventana Media Systems); beta-catenin (clone 14, prediluted; Cell Marque, Rocklin, CA); SOX10 (clone EP268, dilution 1:100; BioSB, Santa Barbara, CA); H3K27Me3 (clone C36B11, dilution 1:200; Cell Signaling Technology, Danvers, MA); and PAX3 (clone 274212, dilution 1:100; RD Systems Europe, Lille, France).

For myogenin, MyoD1, SOX10, H3K27me3 beta-catenin, and PAX3 only a nuclear staining was considered as positive.

Fluorescence In Situ Hybridization

Fluorescence in situ hybridization (FISH) analyses on interphase nuclei from paraffin-embedded 4-μm-thick sections were performed by applying custom probes using bacterial artificial chromosomes (BACs) flanking *PAX3*, *MAML3*, and *NCOA1* genes according to the procedure previously described.[3]

FISH analysis for *FOXO1*, *NCOA2*, and *WWTR1* was performed with 4-μm sections of formalin-fixed paraffin-embedded tissue and the Histology FISH Accessory Kit (Agilent K5799; Dako) using commercially available break-apart probes covering *FOXO1* (Zytovision, Zytolight spec Z-2208-200), *NCOA2* (Empire genomics, EG-NCOA2BA-20-ORGR) and *WWTR1* (Zytovision, Zytolight spec Z-2212-50).

RNA Extraction for Real-Time Reverse Transcription-Polymerase Chain Reaction and Paired-end RNA sequencing

Total RNA was extracted from the formalin-fixed paraffin-embedded tissue section using Trizol reagent (Thermo Fisher Scientific, Courtaboeuf, France) according to the manufacturers' recommendations. DNA was removed using RNase-free DNase set (Qiagen) followed by a second Trizol extraction. The yield of total RNA obtained was evaluated using NanoDrop (Thermo Fisher Scientific).

Reverse Transcription-Polymerase Chain Reaction Analysis for *PAX3-MAML3*, *PAX3-FOXO1*, *PAX3-NCOA2*, and *PAX3-WWTR1* Fusion Genes

Real-time reverse transcription-polymerase chain reaction (RT-PCR), using Taqman technology according to the technique previously described by Hostein et al,[6] was performed for *PAX3-MAML3* and *PAX3-FOXO1* fusion genes.

For *PAX3-MAML3* gene fusion, the following primers and probe were used:

PAX3 forward primer: 5′-TTT CCA GCT ATA CAG ACA GCT TTG-3′

MAML3 reverse primer: 5′-TCC TTC CAA CTT CCT TTT CAC AGT-3′

Probe: 5′-FAM-AACCCCACCATTGGCAATGG CCT-TAMRA-3′

For *PAX3-FOXO1* gene fusion, the following primers and probe were used:

PAX3 forward primer: 5′-TTG GCA ATG GCC TCT CAC C-3′

FOXO1 reverse primer: 5′-ATC CAC CAA GAA CTT TTT CCA G-3′

Probe: 5′-TET-CCCTAC ACA GCA AGT TCA TTC GTG TGC AG-TAMRA-3′

Conventional RT-PCR was performed for *PAX3-NCOA2* and *PAX3-WWTR1* fusion genes.

An aliquot of the RNA extracted from FFPE tissue was used to confirm the novel fusion transcripts identified. One microgram of total RNA was reverse-transcribed in cDNA with the High Capacity cDNA Reverse Transcription Kit with RNase Inhibitor (Invitrogen, cat. No. 4374966). PCR was performed using the AmpliTaq Gold DNA Polymerase kit (Applied Biosystems, cat. No. 4311806) on 50 ng of cDNA with the following primers: PAX3_FWD: 5′ GACCCTGT-CACAGGCTAC 3′ and WWTR1_REV: 5′ TCTGCTGG CTCAGGGTACT 3′ and for the reciprocal transcript: WWTR1_FWD: 5′ CACACCAGTGCCTCAGAG 3′ and PAX3_REV: 5′ CGTGTTCAAAAGGATTTGAAACC 3′. For PAX3-NCOA2 validation, the following primers were used: PAX3_FWD: 5′ CTTTGTGCCTCCGTCGG G 3′ and NCOA2_REV: 5′ CTCGTGTCTGGGAAAA

GCTG 3′. The Touchdown 60°C program was used (TD 60°C; 2 cycles at 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). PCR products were then purified using the Illustra ExoProStar PCR Purification Kit (GE Healthcare, cat. no. US77702), and sequencing reactions were performed with the Big Dye Terminator V1.1 Kit (Applied Biosystems, cat. no. 4337450). After purification with the Big Dye XTerminator Purification Kit (Applied Biosystems, cat. no. 4376486), the samples were sequenced on a 3130xl Genetic Analyzer (Applied Biosystems).

Paired-end RNA Sequencing

Four cases were studied by RNA sequencing with formalin-fixed paraffin-embedded material (cases 5, 8, 29, and 40). All samples had a percentage of RNA fragments above 200 nucleotides (DV200) above 13% upon Tape Station analysis using the Hs RNA Screen Tape (Agilent).

Libraries were prepared with 100 ng of total RNA using the TruSeq RNA Access Library Prep Kit (Illumina, San Diego). Libraries were pooled by groups of 12 samples. Paired-end sequencing was performed using the NextSeq. 500/550 High Output V2 kit (150 cycles) on an Illumina NextSeq. 500 platform (Illumina, San Diego, CA).

Sequencing data (average of 65 million reads per sample) were aligned with STAR on GRCh 38 reference genome. The fusion transcripts were called with STAR-Fusion, FusionMap, FusionCatcher, ERICSCRIPT, and TopHat-fusion and validated if present in the fusion list of at least 2 algorithms.[7–11]

RESULTS

Clinical Features

A total of 41 cases were included in this study. Patient characteristics and clinical follow-up are presented in Table 1. Sixteen patients were male individuals (39%) and 25 were female individuals (61%). The mean and median ages at diagnosis were 52.2 and 49 years, respectively (range, 25 to 84 y). Tumors arose predominantly in the nasal cavity (28 cases, 68%) and ethmoid sinuses (20 cases, 49%), with 11 cases in both the nasal cavity and ethmoid sinuses (27%). Seven cases (17%) arose in the facial sinuses NOS, and 6 cases (15%) showed an extensive tumor with involvement of the sinuses and adjacent bones. Tumor size was known in 17 cases and ranged from 10 to 90 mm (median size 35 mm).

Treatment was known in 33 patients and consisted in surgery for 32, with radiotherapy in 9 patients, chemotherapy in 2, and radiotherapy and chemotherapy in 2 patients.

Clinical follow-up varied from 11 to 185 months (median, 45 mo) and was available for 25 patients of the cohort with 8 recent cases. Local recurrences occurred in 8 cases (32%) at 9 to 95 months of follow-up. No patient showed evidence of distant metastasis.

Pathologic Features

Initial diagnosis was schwannoma (n = 6), neurofibroma (n = 1), MPNST (n = 13, with rhabdomyoblastic differentiation in 5 cases), synovial sarcoma (n = 1), fibrosarcoma (n = 3), and low-grade sinonasal sarcoma (n = 17). Of 20 cases seen after 2014, 17 were properly classified by the initial pathologist. Final diagnosis of BSNS was based on histology and immunohistochemistry according to the original description.[1] The histologic characteristics are highly reproducible, and diagnosis is easy in most cases (Fig. 1).

Tumors were poorly circumscribed with infiltrative involvement of surrounding tissues, particularly the sinonasal bones (19/33 samples containing bone tissues). Twenty-five cases showed hyperplasia of the overlying respiratory epithelium with entrapment of benign glands in the tumor. The tumors consisted of hypercellular proliferation of monotonous spindle cells arranged in fascicles, often with a herringbone pattern. The cellularity was typically high with usually scanty collagen and focally myxoid changes in 7 cases. A hemangiopericytoma-like pattern was focally present in 33 cases. Tumor cells were uniform with a monotonous elongated hyperchromatic nucleus with fine granular chromatin and a small amount of cytoplasm with indistinct borders. Pleomorphic cells were visible focally in only 1 case (Fig. 2). Cells with abundant eosinophilic cytoplasm in favor of rhabdomyoblastic differentiation were present in 5 cases (cases 5, 6, 7, 14, and 16). Mitotic activity was low in most cases (from 0 to 4 mitoses/10 high-power fields, median count, 1), but 2 cases showed 9 and 12 mitoses per 10 high-power fields. Necrosis was always absent.

Immunohistochemical Features

Results are summarized in Table 2. All 41 cases (100%) showed focal (56%) or diffuse (44%) positivity for S100 protein, whereas SOX10 was negative in all cases tested (Fig. 3). H3K27me3 was retained in 33% of cases, and there was a partial loss in 67% of cases (median, 65% of positive tumor cells). Smooth muscle actin was positive in 90% of cases, desmin in 66%, myogenin in 20%, and MyoD1 in 91%. Desmin, myogenin, and MyoD1 were positive only focally. PAX3 was positive in 29 cases that were tested, with 21 cases showing a diffuse positivity (strong positivity in 14 and weak in 7 cases), and 8 cases with a focal positivity (strong positivity in 2 cases and weak in 6 cases). Nuclear beta-catenin was focally positive in 26% of cases. Pankeratin AE1/AE3, EMA, and CD34 were focally positive in 11, 11, and 10% of cases, respectively.

Molecular Biological Features

Initially, 44 cases were screened with RT-PCR for *PAX3-MAML3*: 35 cases were positive, 5 negative, and 4 noninterpretable because of nucleic acid degradation. Theereafter, 8 of these negative or noninterpretable cases were screened with FISH for probes for *PAX3*: 5 were positive and 3 noninterpretable. These 3 noninterpretable cases for RT-PCR and FISH were excluded from the study because of nucleic acid degradation. Among the 5 positive cases with FISH for *PAX3*, 2 were positive with FISH for *MAML3* (cases 9 and 15) and 3 were negative. These 3 negative cases (cases 5, 8, and 29) were also negative with FISH for *NCOA1* and *FOXO1*. These 3 cases as well as case 40 (recent case negative for initial

TABLE 1. Clinical and Follow-up Data in a Series of 41 Biphenotypic Sinonasal Sarcomas

Case	Sex	Age (y)	Site of Tumor	Treatment Modalities	Local Recurrence (mo)	Follow-up (mo)	Status
1	F	69	Sinus (ethmoid)	Surgery	Yes (36 m)	185	AWD
2	F	25	Nasal cavity	Surgery	No	175	NED
3	M	35	Nasal cavity	NA	NA	NA	NA
4	F	49	Sinus (ethmoid)	Surgery	No	145	NED
5	F	49	Nasal cavity/sinus (ethmoid)	Chemotherapy, surgery, radiotherapy	No	83	NED
6	F	40	Nasal cavity/sinus (ethmoid)/cribriform plate/skull base	Surgery	NA	NA	NA
7	F	38	Sinus (NOS)	NA	NA	NA	NA
8	M	76	Nasal cavity	Surgery	No	119	NED
9	F	63	Sinus (ethmoid)/skull base/orbit	Chemotherapy, radiotherapy	Yes (91 m)	102	AWD
10	F	37	Nasal cavity	Surgery, chemotherapy	No	107	NED
11	M	63	Nasal cavity	Radiotherapy, surgery	Yes (95 m)	96	NED
12	F	43	Sinus (ethmoid)/skull base	Surgery, radiotherapy, chemotherapy	No	69	NED
13	M	28	Nasal cavity	Surgery	Yes (24 m)	33	NED
14	F	62	Nasal cavity/sinus (ethmoid)	Surgery, chemotherapy	No	55	NED
15	F	81	Nasal cavity/sinus (ethmoid)	Surgery	NA	NA	NA
16	F	25	Nasal cavity	NA	NA	NA	NA
17	M	28	Nasal cavity	Surgery	Yes (27 m)	34	NED
18	M	38	Nasal cavity	Surgery	No	54	NED
19	M	50	Nasal cavity/sinus (ethmoid/frontal)	Surgery	Yes (15 m)	49	NED
20	F	84	Sinus (ethmoid/frontal)/orbit	Surgery	No	45	NED
21	M	77	Nasal cavity/sinus (ethmoid/frontal)	NA	NA	NA	NA
22	F	71	Sinus (etmoid/frontal)	Surgery	No	23	NED
23	F	77	Nasal cavity/sinus (ethmoid)/orbit/skull base	Surgery	No	17	NED
24	M	32	Nasal cavity/sinus (ethmoid/frontal)	Surgery	No	27	NED
25	M	59	Sinus (ethmoid/frontal)/orbit	Surgery, radiotherapy	No	12	NED
26	M	65	Nasal cavity/sinus (ethmoid)	Surgery, radiotherapy	Yes (11 m)	23	NED
27	F	50	Nasal cavity	Surgery, radiotherapy	No	25	NED
28	F	70	Nasal cavity	Surgery	No	25	NED
29	F	65	Nasal cavity	Surgery	NA	NA	NA
30	F	49	Sinus (NOS)	NA	NA	NA	NA
31	M	29	Sinus (ethmoid)	Surgery, radiotherapy	No	20	NED
32	M	41	Nasal cavity/sinus (ethmoid)	Surgery, radiotherapy	No	18	NED
33	F	64	Nasal cavity/sinus (NOS)	Surgery	Yes (9 m)	11	AWD
34	F	77	Nasal cavity/sinus (ethmoid)	Surgery	NA	Recent case	NA
35	M	38	Sinus (ethmoid)	Surgery, radiotherapy	NA	Recent case	NA
36	F	41	Sinus (NOS)	NA	NA	Recent case	NA
37	F	60	Nasal cavity/sinus (NOS)	NA	NA	Recent case	NA
38	M	65	Nasal cavity/sinus (NOS)	Surgery	NA	Recent case	NA
39	F	31	Nasal cavity	Surgery	NA	Recent case	NA
40	M	39	Nasal cavity	Surgery, radiotherapy	NA	Recent case	NA
41	F	48	Sinus (NOS)	NA	NA	Recent case	NA

AWD indicates alive with disease; F, female; M, male; NA, not available; NED, no evolutive disease.

RT-PCR) were analyzed by RNA sequencing. This revealed a fusion transcript, which was confirmed by both RT-PCR and FISH: *PAX3-NCOA2* (case 5), *PAX3-WWTR1* (cases 8 and 29), and *PAX3-FOXO1* (case 40). Fusion transcripts *PAX3-NCOA2* and *PAX3-WWTR1* were also confirmed by Sanger sequencing (Figs. 4, 5). They were both in frame fusions involving *PAX3* exon 7 with *NCOA2* exon 12 and *PAX3* exon 8 with *WWTR1* exon 5, respectively.

DISCUSSION

To our knowledge, 3 series of BSNS have been published with 44, 11, and 15 cases, respectively.[5,12,13] We report here the second largest series with 41 cases. We confirm that this tumor is locally aggressive with frequent bone destruction and local recurrence but with no distant metastasis. It occurs in the nasal cavity and/or paranasal sinuses, predominantly in middle-aged women. Histologically, BSNS is a poorly circumscribed and hypercellular proliferation of monotonous spindle cells with a low mitotic rate and, frequently, entrapped hyperplastic surface epithelium. In our series, only 1 case showed focal pleomorphic cells, 2 cases showed some mitotic activity, and focal myxoid changes were present in 7 cases. We also confirm previously published data, with consistent, at least, focal S100 protein positivity, frequent positivity with muscle markers, and negativity for SOX10, epithelial markers, and CD34. As recently reported by Jo et al,[13] all our tested tumors were PAX3 positive but with weak and/or focal positivity in about half of them.

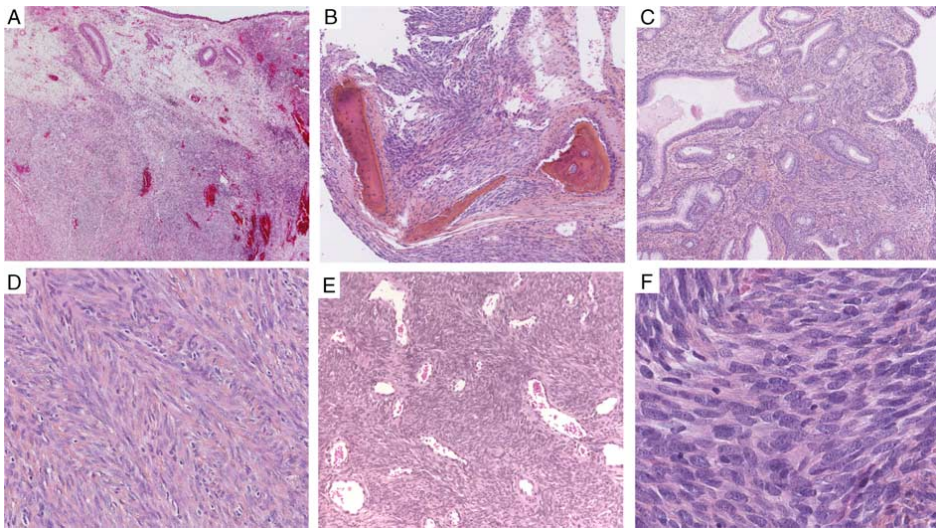


FIGURE 1. Main histologic features of biphenotypic sinonasal sarcoma. The tumor is poorly circumscribed (A), with frequent involvement of the sinonasal bones (B). Hyperplasia of the overlying respiratory epithelium with entrapment of benign glands by tumor cells is a typical feature (C). The tumor is composed of hypercellular proliferation of monomorphic spindle cells arranged in medium-to-long fascicles, often with a herringbone pattern (D). Hemangiopericytoma-like pattern is common (E). Tumor cells are uniform with monotonous spindle nucleus and low mitotic activity (F).

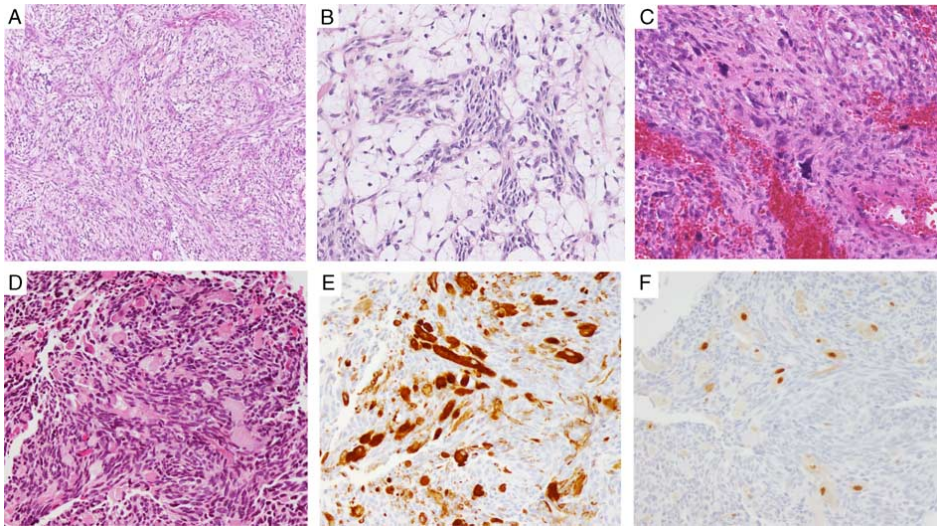


FIGURE 2. Rare morphologic patterns of biphenotypic sinonasal sarcoma. Myxoid changes may be focally present (A). Focal chronic inflammatory infiltrate is rare. One case (#5) showed a histiocytic infiltrate (B). Only one case (#19) showed focal nuclear atypia (C). Focal rhabdomyoblastic differentiation was present in 5 cases with large round or elongated cells with abundant eosinophilic cytoplasm (D) and focal cross-striation. Desmin (E) and myogenin (F) were always positive in these areas.

TABLE 2. Immunohistochemical Results in a Series of 41 Biphenotypic Sinonasal Sarcomas

Antibody	Negativity (n)	Focal Positivity (n)	Diffuse Positivity (n)	Total Positivity (%)
S100 protein	0	23	18	100
SOX10	34	0	0	0
H3K27me3	0	10	20	100
Desmin	14	27	0	66
MyoD1	3	32	0	91
Myogenin	33	8	0	20
Smooth muscle actin	4	28	7	90
PAX3	0	8	21	100
Beta-catenin	16	5	1	27
AE1/AE3	34	4	0	11
EMA	32	4	0	11
CD34	36	4	0	10

n indicates number of cases.

With regard to muscle markers, we found frequent focal positivity for MyoD1 (91% of cases) but a much lower rate of positivity for myogenin (20%). This is in agreement with the expression profiling reported by Wang et al[2] in 8 cases of BSNS. In that study, MyoD1 was one of the top 150 overexpressed genes. However, they observed a focal expression of MyoD1 in only 4 of 25 cases, whereas we found positivity in 91% of our cases. This difference may be due to the different antibodies used in the studies (clone 5.8 A by Wang and colleagues and clone EP212 in our study). Unlike Rooper et al[12] who reported an almost constant nuclear positivity for beta-catenin, only 26% of our cases showed focal nuclear positivity. Therefore, other studies are necessary to evaluate the usefulness of this marker in BSNS.

A recurrent *PAX3-MAML3* fusion event is present in most BSNS cases, with a few cases showing alternative fusion of *PAX3* with *NCOA1* and *FOXO1*. In their series of 44 cases, Fritchie et al[5] reported 24 cases with a *PAX3-MAML3*, 3 cases with a *PAX3-FOXO1*, and 1 case with a *PAX3-NCOA1* fusion gene, whereas 11 cases showed a *PAX3* rearrangement, but with no rearrangement of *MAML3*, *FOXO1*, *NCOA1*, and *NCOA2* genes. In the present study, 37 cases showed a *PAX3-MAML3* fusion, 2 cases a *PAX3-WWTR1* fusion, 1 case a *PAX3-FOXO1* fusion, and 1 case a *PAX3-NCOA2* fusion. Given the structural and functional similarity of the MAML3, FOXO1, NCOA1, and NCOA2 proteins, the presence of a *PAX3-NCOA2* fusion transcript in BSNS was expected.[5] Like the 2 cases of BSNS with an *NCOA1* rearrangement reported by Huang et al,[3] our case with an *NCOA2* rearrangement showed focal rhabdomyoblastic differentiation. *PAX3-WWTR1* is a new fusion in BSNS. WWTR1, also known as TAZ, is a transcriptional coactivator with a PDZ binding motif. In mammals, *YAP1* and *WWTR1* are downstream effectors of the Hippo signaling pathway, which is an evolutionarily conserved network that plays a

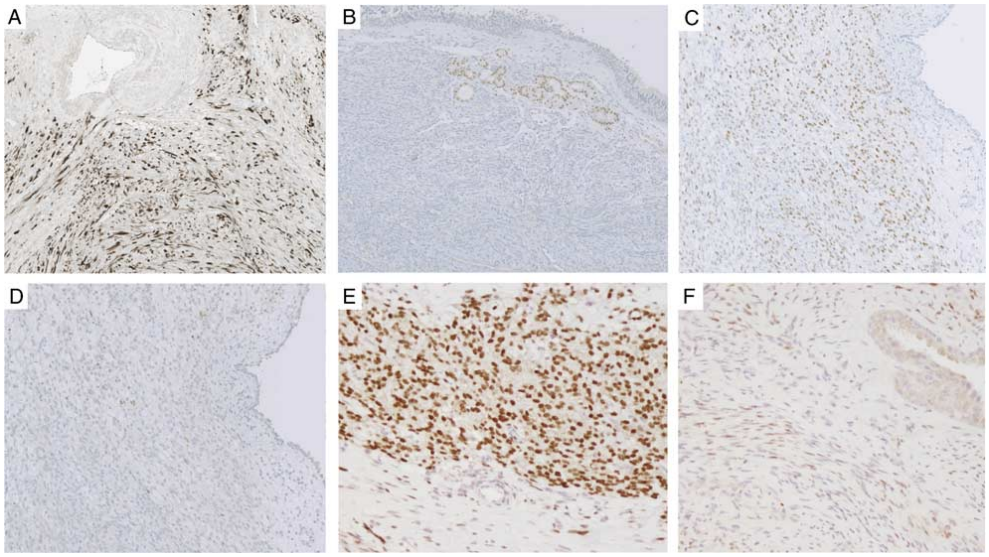


FIGURE 3. Immunohistochemical profile of biphenotypic sinonasal sarcoma. S100 protein was always positive focally or diffusely (A), whereas SOX 10 was always negative (B). MyoD1 was positive in about 90% of cases (C). Myogenin was focally positive in only 20% of cases (D). PAX3 was positive in all tested cases, strongly and diffusely in about half of them (E) but weakly and/or focally in the other cases with a background in some cases (F).

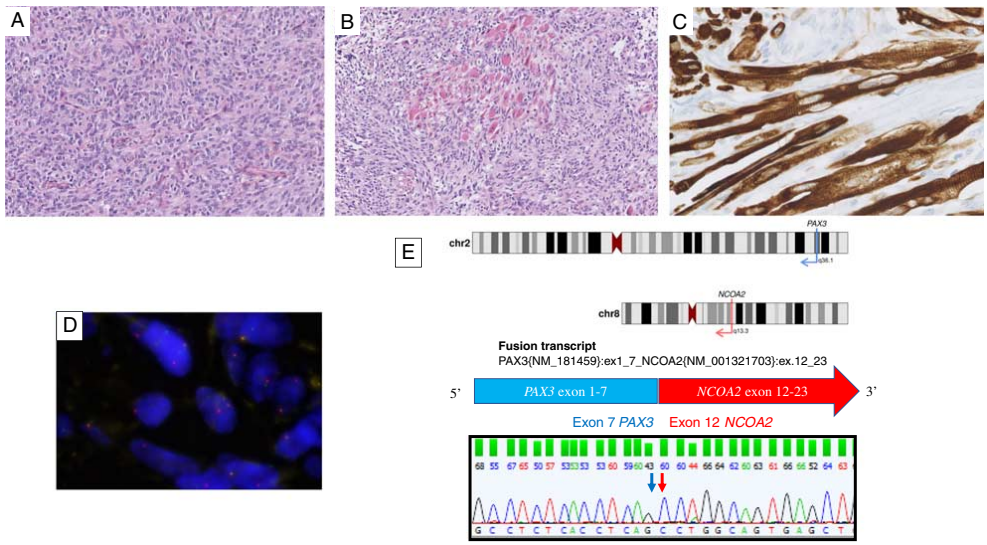


FIGURE 4. Novel *PAX3-NCOA2* fusion in biphenotypic sinonasal sarcoma (case 5). Typical aspect of BSNS (A) with areas showing rhabdomyoblastic differentiation (B). Immunohistochemistry with desmin highlights cross-striations (C). FISH using a break-apart probe shows a rearrangement of *NCOA2* (D). E, Structure of *PAX3-NCOA2* fusion transcript. From up to bottom: schematic representing locus and chromosomal positions of *PAX3* and *NCOA2*; schematic of breakpoint positions involving *PAX3* exon 7 and *NCOA2* exon 12; nucleotidic sequence of adjoined sequences.

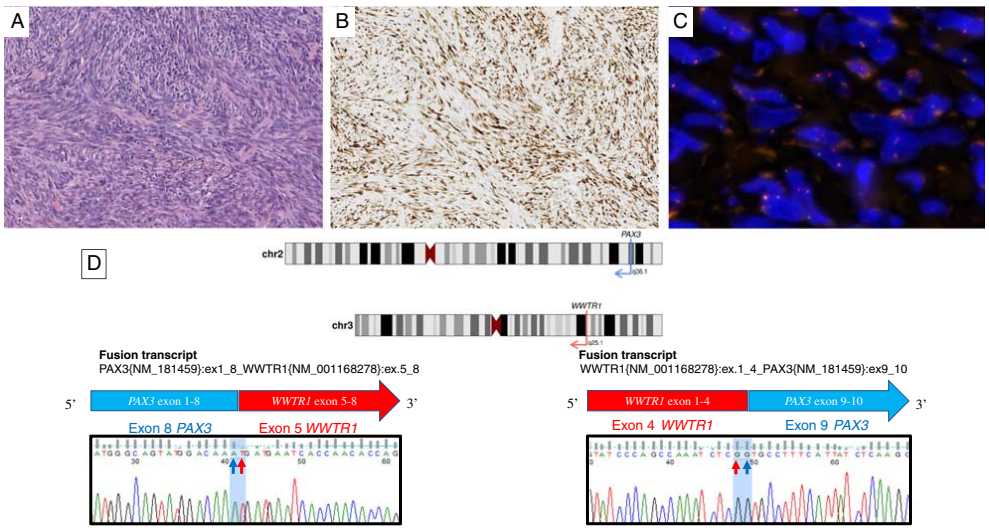


FIGURE 5. Novel *PAX3-WWTR1* fusion in biphenotypic sinonasal sarcoma (cases 8 and 29). Typical aspect of BSNS (A: case 8) with diffuse positivity of S100 protein (B: case 8). FISH using a break-apart probe shows a rearrangement of *WWTR1* (C: case 29). D, From up to bottom: schematic representing locus and chromosomal positions of *PAX3* and *WWTR1*; schematic of breakpoint positions involving *PAX3* exon 8 and *WWTR1* exon 5 for both cases, and associated reciprocal fusion transcript *WWTR1* exon 4 and *PAX3* exon 10 for case 29; nucleotidic sequence of adjoined sequences for cases 8 and 29.

central role in regulating cell proliferation and cell fate to control organ growth and regeneration. The Hippo pathway controls gene expression by inhibiting the activity of *YAP1* and *WWTR1*. Hyperactivity of these 2 genes promotes uncontrolled cell proliferation, impairs differentiation, and is associated with cancer.[14] A *WWTR1-CAMTA1* fusion has been shown in epithelioid hemangioendothelioma and is now a key diagnostic tool for this rare tumor.[15] BSNS is a second example of the direct role of *WWTR1* in the development of a cancer. Recently, Sun et al[16] reported the important role of *YAP1* and *WWTR1* in skeletal muscle stem cell function. They showed that *YAP1* and *WWTR1* play a similar role in promoting myoblastic proliferation and that, during the later stage of myogenesis, *WWTR1* switches toward influencing satellite cell fate by promoting myogenic differentiation. *YAP1* and *WWTR1* have many common target genes, but *WWTR1* regulates some genes independently of *YAP1*, including myogenic genes such as *PAX7*, *MYF5*, and *MYOD1*.

In conclusion, this is the second largest series of BSNS. We report 2 new fusion transcripts, *PAX3-NCOA2* and *PAX3-WWTR1*, and confirm the major value of immunohistochemistry for the diagnosis, with constant positivity of S100 protein and PAX3 associated with negativity of SOX10. Moreover, we found that MyoD1 is more sensitive than myogenin for demonstrating myogenic differentiation in this tumor.

ACKNOWLEDGMENTS

The authors are grateful to the pathologists who provided material and clinical follow-up: G. Aillet (Nantes), E. Ancelin-Malbreil (Mont de Marsan), C. Bouvier (CHU Marseille), P. Camparo (Amiens), E. Cassagnau (CHU Nantes), Croue (CHU Angers), M. Desrousseau (Biarritz), François (CHU Rouen), Gauchez (CHU Marseille), Jouannelle-Sulpicy (Fort de France Martinique), E. Lavoine (Angers), Lhermitte (La Reunion), A. Lobna (Sfax, Tunisia), N. Macagno (CHU Marseille), Milin (CHU Poitiers), Poizat (Montpellier), P. Reau (Bordeaux), C. Sader Ghorra (Beyrouth, Lebanon), Savel (La Roche sur Yon), E. Uro-Coste (CHU Toulouse), S. Valmary (CHU Besançon), I. Valo (Angers), and N. Weingertner (CHU Strasbourg).

REFERENCES

1. Lewis JT, Oliveira AM, Nascimento AG, et al. Low-grade sinonasal sarcoma with neural and myogenic features: a clinicopathologic analysis of 28 cases. *Am J Surg Pathol*. 2012;36:517–525.
2. Wang X, Bledsoe KL, Graham RP, et al. Recurrent PAX3-MAML3 fusion in biphenotypic sinonasal sarcoma. *Nat Genet*. 2014;46:666–668.
3. Huang S-C, Ghossein RA, Bishop JA, et al. Novel PAX3-NCOA1 fusions in biphenotypic sinonasal sarcoma with focal rhabdomyoblastic differentiation. *Am J Surg Pathol*. 2016;40:51–59.
4. Wong WJ, Lauria A, Hornick JL, et al. Alternate PAX3-FOXO1 oncogenic fusion in biphenotypic sinonasal sarcoma. *Genes Chromosomes Cancer*. 2016;55:25–29.
5. Fritchie KJ, Jin L, Wang X, et al. Fusion gene profile of biphenotypic sinonasal sarcoma: an analysis of 44 cases. *Histopathology*. 2016;69:930–936.
6. Hostein I, Andraud-Fregeville M, Guillou L, et al. Rhabdomyosarcoma: value of myogenin expression analysis and molecular testing in diagnosing the alveolar subtype. An analysis of 109 paraffin-embedded specimen. *Cancer*. 2004;101:2817–2824.
7. Ge H, Liu K, Juan T, et al. FusionMap: detecting fusion genes from next-generation sequencing data at base-pair resolution. *Bioinformatics*. 2011;27:1922–1928.
8. Trapnell C, Roberts A, Goff L, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;7:562–578.
9. Lagstad S, Zhao S, Hoff AM, et al. Chimeraviz: a tool for visualizing chimeric RNA. *Bioinformatics*. 2017;33:2954–2956.
10. Nicorici DSM, Edgren H, Kangaspeska S, et al. FusionCatcher—a tool for finding somatic fusion genes in paired-end RNA-sequencing data. *bioRxiv*. 2014. Doi: https://doi.org/10.1101/011650.
11. Benelli M, Pescucci C, Marseglia G, et al. Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. *Bioinformatics*. 2012;28:3232–3239.
12. Rooper LM, Huang S-C, Antonescu CR, et al. Biphenotypic sinonasal sarcoma: an expanded immunoprofile including consistent nuclear β-catenin positivity and absence of SOX10 expression. *Hum Pathol*. 2016;55:44–50.
13. Jo VY, Mariño-Enriquez A, Fletcher CDM, et al. Expression of PAX3 distinguishes biphenotypic sinonasal sarcoma from histologic mimics. *Am J Surg Pathol*. 2018;42:1275–1285.
14. Misra JR, Irvine KD. The Hippo signaling network and its biological functions. *Annu Rev Genet*. 2018;52:3.1–3.23.
15. Errani C, Zhang L, Sung YS, et al. A novel WWTR1-CAMTA1 gene fusion is a consistent abnormality in epithelioid hemangioendothelioma of different anatomic sites. *Genes Chromosomes Cancer*. 2011;50:644–653.
16. Sun C, De Mello V, Mohamed A, et al. Common and distinctive functions of the Hippo effectors Taz and Yap in skeletal muscle stem cell function. *Stem Cells*. 2017;35:1958–1972.