



(51) Classification internationale des brevets :
C12Q 1/68 (2006.01)

(21) Numéro de la demande internationale :
PCT/FR2010/000323

(22) Date de dépôt international :
21 avril 2010 (21.04.2010)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
09/01936 22 avril 2009 (22.04.2009) FR
61/171,495 22 avril 2009 (22.04.2009) US

(71) Déposants (pour tous les États désignés sauf US) :
UNIVERSITÉ VICTOR SEGALEN - BORDEAUX 2 [FR/FR]; 146 rue Léo Saignât, F-33076 Bordeaux Cedex (FR). **INSTITUT BERGONIÉ** [FR/FR]; 229 cours de l'Argonne, F-33076 Bordeaux Cedex (FR). **INSTITUT CURIE** [FR/FR]; 26 rue d'Ulm, F-75005 Paris (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (pour US seulement) : **CHIBON, Frédéric** [FR/FR]; 9 route de Rabey, F-33640 Isle-Saint-Georges (FR). **COINDRE, Jean-Michel** [FR/FR]; 19 rue de Soulac, F-33200 Bordeaux, France (FR). **AURIAS, Alain** [FR/FR]; 20 avenue Beauséjour, F-91440 Bures-sur-Yvette (FR).

(74) Mandataire : **NOVAGRAAF TECHNOLOGIES**; 122 rue Edouard Vaillant, F-92593 Levallois-Perret Cedex (FR).

(81) États désignés (sauf indication contraire, pour tout titre de protection nationale disponible) : AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD,

[Suite sur la page suivante]

(54) Title : PROGNOSTIC MOLECULAR SIGNATURE OF SARCOMAS, AND USES THEREOF

(54) Titre : SIGNATURE MOLÉCULAIRE PRONOSTIQUE DES SARCOMES ET UTILISATIONS

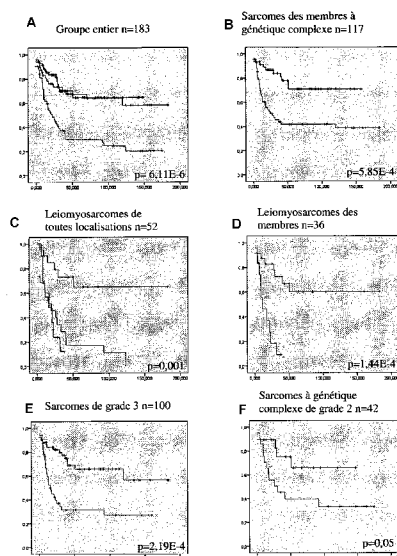


FIGURE 2

- A Entire group n = 183
B Genetically complex sarcomas on limbs n = 117
C Leiomyosarcomas in all locations n = 52
D Leiomyosarcomas on limbs n = 36
E Grade 3 sarcomas n = 100
F Genetically complex Grade 2 sarcomas n = 42

(57) Abstract : The present invention relates to a prognostic molecular signature for sarcomas, and in particular for genetically complex sarcomas, and to the use thereof for predicting the overall survival of a patient, particularly for predicting the appearance of metastases but also for differentiating prognostically separate sub-groups within a group of tumors having the same histological grade. Said molecular signature can also be used for testing the effectiveness of a treatment or for producing novel therapeutic agents.

(57) Abrégé : La présente invention est relative à une signature moléculaire pronostique des sarcomes, et notamment des sarcomes à génétique complexe, et son utilisation pour prédire la survie globale d'un patient, en particulier de prédire la survenue de métastases, mais également pour distinguer des sous-groupes de pronostic distinct au sein d'un groupe de tumeurs de même grade histologique. Cette signature moléculaire peut également être utilisée pour tester l'efficacité d'un traitement, ou pour obtenir de nouveaux agents thérapeutiques.



SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR,
TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
ML, MR, NE, SN, TD, TG).

(84) États désignés (*sauf indication contraire, pour tout titre de protection régionale disponible*) : ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasién (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM,

Publiée :

- avec rapport de recherche internationale (*Art. 21(3)*)
- avec la partie de la description réservée au listage des séquences (*règle 5.2.a*)

SIGNATURE MOLÉCULAIRE PRONOSTIQUE DES SARCOMES ET UTILISATIONS

DOMAINE DE L'INVENTION

5 La présente invention est relative à une signature moléculaire pronostique des sarcomes, en particulier des sarcomes à génétique complexe, et à son utilisation pour prédire la survie sans métastases et la survie globale de patients atteints d'un sarcome.

 Elle trouve de nombreuses applications, en particulier dans le domaine
10 du pronostic ou du diagnostic des sarcomes ou pour surveiller le traitement de patients atteints d'un sarcome.

ART ANTÉRIEUR

 Les sarcomes des tissus mous (STS) de l'adulte sont rares et hétérogènes
15 en termes de localisation, d'histologie, d'anomalies moléculaires et de pronostic. Les STS peu différenciés sont les tumeurs malignes les plus fréquentes chez les adultes, représentant environ 50% des diagnostics pathologiques, et comprennent principalement les sarcomes avec un caryotype complexe, à savoir les leiomyosarcomes (LMS), les sarcomes indifférenciés (US) ou histiocytofibromes malins (MFH), et les liposarcomes
20 dédifférenciés (DD-LPS) (FLETCHER et al., World Health Organisation (WHO) classification of tumours. Pathology and genetics of tumours of soft tissue and bone. Lyon, IARC Press, 2002). Au niveau génétique, les STS peu différenciés peuvent être divisés en deux groupes principaux, un groupe avec un profil génomique complexe (80%) regroupant essentiellement les US, les LMS, les rhabdomyosarcomes pléomorphes et liposarcomes
25 pléomorphes, associés avec des profils très complexes, mais récurrents, de déséquilibres génomiques (IDBAIH et al., Lab. Invest., 85(2) : 176-181, 2005 ; CHIBON et al., Cancer Genet. Cytogenet., 141(1) : 75-78, 2003 ; DERRE et al., Lab. Invest., 81(2) : 211-215, 2001), et un second groupe avec un profil génétique simple (20%) basé sur un niveau élevé d'amplifications limitées et composé exclusivement des DD-LPS (CHIBON et al., Cancer
30 Genet. Cytogenet., 139(1) : 24-29, 2002 ; COINDRE et al., Mod. Pathol., 16(3) : 256-262, 2003). Les STS sont des tumeurs agressives capables de rechute locale et métastatique. Les patients avec de telles tumeurs ont habituellement un mauvais pronostic, 40 à 50% développent éventuellement des métastases distantes, principalement dans les poumons, généralement dans les 5 ans suivant le diagnostic (WEITZ et al., J. Clin. Onc., 21(14) :
35 2719-2725, 2003 ; ZAGARS et al., Cancer, 97(10) : 2530-2543, 2003).

 Le traitement clinique des STS consiste principalement en la résection chirurgicale, avec des thérapies d'appoint dont la durée et la nature dépendent des marges chirurgicales, de l'histotype tumoral et du grade histologique. Cependant, le bénéfice des

thérapies d'appoint comme la chimiothérapie est actuellement contesté bien que des études récentes tendent à démontrer un effet sur les rechutes locales et distantes (SMAC, Lancet, 350(9092) : 1647-1654, 1997 ; FRUSTACI et al., J. Clin. Oncol., 19(5) : 1238-1247, 2001 ; PERVAIZ et al., Cancer, 113(3) : 573-581, 2008). Néanmoins l'efficacité de la chimiothérapie est marginale (de 3 à 10% selon le critère d'évaluation, PERVAIZ et al., 2008, précité) ; ce qui pourrait résulter de la sélection des patients pour lesquels la malignité tumorale est évaluée par le grade histologique. Par ailleurs la prise en charge des patients dépend essentiellement du stade de la maladie. Bien qu'il fournisse une information valable quant à l'évolution clinique de certains types de sarcomes, le typeage histologique a une valeur prédictive limitée pour d'autres types de sarcomes, notamment les sarcomes non classés, peu différenciés et à non-translocation. Pour augmenter la valeur prédictive de l'histologie en termes de pronostic, plusieurs systèmes de grade ont été générés (BRODERS et al., Surg. Gynecol., Obstet., 69 : 267-280, 1939 ; RUSSELL et al., Cancer, 40(4) : 1562-1570, 1977 ; MARKHEDE et al., Cancer, 49(8) : 1721-1733, 1982 ; TROJANI et al., Int. J. Cancer, 33(1) : 37-42, 1984 ; COSTA et al., Cancer, 53(3) : 530-541, 1984). Parmi ceux-ci, les systèmes de l'Institut National du Cancer (NCI) (COSTA et al., 1984, précité) et de la Fédération Nationale de Centre de Lutte Contre le Cancer (FNCLCC) (TROJANI et al., 1984, précité) ont été largement utilisés même si le second système permet d'augmenter légèrement la capacité de prédiction des métastases distantes et a été considéré de ce fait comme « l'étalon-or » (GUILLOU et al., J. Clin. Onc., 15(1) : 350-362, 1997).

Jusqu'à présent, le grade histologique est le meilleur critère prédictif de la survie sans métastase et de la survie globale. Le grade de la FNCLCC, le plus efficace, a été établi depuis plus de 20 ans et est toujours le système le plus communément utilisé. Il est basé sur l'évaluation semi-quantitative de la différenciation tumorale, la nécrose, et l'indice mitotique. Toutefois ce système présente plusieurs limitations : sa reproductibilité d'un pathologiste à l'autre n'est pas parfaite, il ne s'applique pas à tous les types de sarcomes (COINDRE et al., Cancer, 91(10) : 1914-1926, 2001) et n'est pas informatif pour les cas classés grade 2 (qui représentent environ 40% des cas). Cependant malgré ces limitations, depuis plus de 20 ans, aucune étude n'a apporté de critères pronostiques susceptibles de remplacer ce système de grade histologique.

Ces dix dernières années ont vu l'émergence de signatures moléculaires pronostiques dans un nombre grandissant de pathologies. À ce jour, le meilleur exemple de signature moléculaire est certainement celui du cancer du sein dans lequel une signature d'expression prédictive de la rechute métastatique a été établi en 2002 puis validée la même année par la même équipe sur un groupe indépendant de 295 tumeurs (VAN'T VEER et al., Nature, 415(6871) : 530-536, 2002; VAN de VIJVER et al., N. Engl. J. Med., 347(25) : 1999-2009, 2002).

Jusqu'à présent dans le domaine des sarcomes, les profils d'expression ont surtout été établis dans le but d'identifier de nouveaux marqueurs diagnostiques ou de mieux comprendre l'oncogenèse de ces tumeurs en relation avec la différenciation tumorale (NIELSEN et al., Lancet, 359(9314) : 1301-1307, 2002 ; BAIRD et al., Cancer Res., 65(20) : 9226-9235, 2005 ; FRITZ et al., Cancer Res., 62(11) : 2993-2998, 2002 ; MATUSHANSKY et al., Am. J. Pathol., 172(4) : 1069-1080, 2008 ; SEGAL et al., Am. J. Pathol., 163(2) : 691-700, 2003 ; LEE et al., J. Cancer, 88(4) : 510-515, 2003 ; NAKAYAMA et al., Mod. Pathol., 20(7) : 749-759, 2007 ; SINGER et al., Cancer Res., 67(14) : 6626-6636, 2007). Seulement deux études, portant sur 30 leiomyosarcomes (LEE et al., Cancer Res., 64(20) : 7201-7204, 2004) et sur 89 sarcomes pléomorphes (FRANCIS et al., BMC Genomics, 8 : 73, 2007) proposent une signature moléculaire pronostique. Mais ces deux signatures sont composées de nombreux gènes (respectivement 335 et 244, respectivement) sans lien biologique clair entre eux. Par ailleurs ces deux signatures ont été établies sur un nombre relativement restreint de sous-types de sarcomes spécifiques donnant des significations relativement faibles. Enfin, à ce jour, ces deux signatures n'ont pas été comparées au système de grade FNCLCC et n'ont pas encore été validées sur un groupe indépendant, limitant de ce fait leur utilité clinique.

Dans le domaine des sarcomes, il est à noter que le nombre d'études cherchant à corrélér les altérations moléculaires au pronostic est nécessairement limité du fait de la difficulté d'obtenir un groupe d'étude homogène de tumeurs complètement annotées. Aussi à ce jour, aucune corrélation claire et avérée n'a pu être établie entre le profil génétique et la survie sans métastases.

De ce fait, la progression tumorale reste difficile à prédire à l'intérieur d'un groupe de sarcomes, et les traitements ne sont pas aussi adaptés qu'ils pourraient l'être. Par conséquent, il existe un besoin certain pour améliorer le pronostic et le diagnostic des sarcomes et par conséquent assurer un meilleur suivi clinique des patients.

C'est donc un but de la présente invention de fournir un système de grade plus efficace, fiable et reproductible permettant de pallier les inconvénients de l'art antérieur. C'est également un autre but de la présente invention de fournir les moyens et kits permettant de mettre en œuvre un tel système de grade.

EXPOSE DE L'INVENTION

Selon l'hypothèse que le système de grade de la FNCLCC pourrait représenter un résumé phénotypique d'altérations génomiques, les inventeurs de la présente invention ont découvert de manière toute à fait inattendue que l'établissement d'un profil moléculaire à l'aide de technologies émergentes, telles que les puces à ADN, pourrait permettre l'identification d'altérations/de gènes à l'origine de l'agressivité tumorale, avec de ce fait, une possibilité de définir un système de grade plus efficace basé

sur les altérations moléculaires ; ce qui conduit à une avancée majeure dans le domaine de l'analyse des sarcomes.

Alors que le nombre d'études cherchant à corréler les altérations moléculaires au pronostic est limitée du fait de la difficulté d'obtenir un groupe d'étude homogène de tumeurs complètement annotées, les inventeurs ont initié un projet original visant à déterminer les profils génomiques et d'expression à partir de 183 tumeurs primaires à génétique complexe, non traitées et complètement annotées, référencées dans la base de données du GSF (Groupe Sarcome Français), partie intégrante de la base Européenne Conticabase (www.conticabase.org). L'analyse par groupement a été utilisée pour identifier les altérations moléculaires associées au résultat clinique du patient.

Cette démarche illustrée dans la partie expérimentale ci-après, a permis d'identifier un ensemble particulier de gènes, appelé « pool » ou « signature moléculaire », associés à la complexité du génome, l'agressivité tumorale, et dont le profil d'expression a permis d'établir un pronostic fiable de patients atteints d'un sarcome, en particulier de prédire la survenue de métastases, mais également de distinguer au sein d'un groupe de patients atteints d'un sarcome de même grade histologique, des sous-groupes présentant des pronostics significativement différents.

La présente invention se rapporte donc, en premier lieu, à un pool de polynucléotides comprenant au moins deux polynucléotides choisis parmi les séquences de polynucléotides SEQ ID NO : 1 à SEQ ID NO : 67. En d'autres termes, le pool de polynucléotides de l'invention peut comprendre 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, ou 67 polynucléotides choisis parmi les séquences de polynucléotides SEQ ID NO : 1 à SEQ ID NO : 67.

L'analyse des 67 gènes identifiés (SEQ ID NO : 1 à SEQ ID NO : 67) par la base de données Gene Ontology (GO) a montré en outre que tous étaient impliqués dans le même processus biologique, c'est-à-dire le contrôle de l'intégrité du chromosome.

Les inventeurs ont en outre démontré que ces gènes peuvent être répartis en 5 groupes principaux selon leur rôle dans la mitose : point de contrôle de la mitose et du cycle cellulaire (12 gènes, SEQ ID NOs : 1-12) ; biogenèse des chromosomes, condensation, alignement et ségrégation (26 gènes, SEQ ID NOs : 13-38) ; fuseau mitotique et centrosome (12 gènes, SEQ ID NOs : 39-50) ; moteur des microtubules, complexe kinésine (8 gènes, SEQ ID NOs : 51-58), et cytokinèse (4 gènes, SEQ ID NOs : 59-62) ; parmi les 5 derniers gènes regroupés en fonction des résultats expérimentaux (SEQ ID NOs : 63-67), 3 sont connus comme étant impliqués dans l'instabilité chromosomique (SEQ ID NOs : 63-65) et 2 sont associés au grade histologique selon l'étude (SEQ ID NOs : 66 et 67).

Le tableau 1 ci-dessous indique le nom de chacun des gènes, leur répartition en cinq groupes principaux, et leurs séquences respectives (références GenBank et SEQ ID NO :).

Tableau 1

SET	Nom du gène	Réf. ProbeSet Affymétrie™	Réf. GenBank	SEQ ID NO :
Set 1 : mitose et cycle cellulaire	ASPM	219918_a_at	NM_018136	1
	FOXM1	202580_x_at	NM_021953	2
	PAK3 / UBE2C	202954_at	NM_002578	3
	CDC7	204510_at	NM_003503	4
	CDC20	202870_s_at	NM_001255	5
	CDC45L	204126_s_at	NM_003504	6
	CCNA2	203418_at	NM_001237	7
	CCNB1	214710_s_at	NM_031966	8
	CCNB2	202705_at	NM_004701	9
	CKS2	204170_s_at	NM_001827	10
	MELK	204825_at	NM_014791	11
	CDCA3	223307_at	NM_031299	12
Set 2 : biogenèse des chromosomes, alignement, ségrégation	NCAPH	212949_at	NM_015341	13
	HP1BP3	1554251_at	NM_016287	14
	CENPA	204962_s_at	NM_001042426	15
	KIAA1794	213007_at	NM_018193	16
	SMC2	204240_s_at	NM_001042550	17
	CHEK1	205394_at	NM_001274	18
	H2AFX	205436_s_at	NM_002105	19
	OIP5	213599_at	NM_007280	20
	MCM2	202107_s_at	NM_004526	21
	MCM7	210983_s_at	NM_005916	22
	BIRC5	202095_s_at	NM_001012270	23
	AURKA	204092_s_at	NM_003600	24
	MAD2L1	1554768_a_at	NM_002358	25
	BUB1	215509_s_at	NM_004336	26
	AURKB	239219_at	NM_004217	27
	BUB1B	203755_at	NM_001211	28
	SGOL2	230165_at	NM_152524	29
	PTTG1	203554_x_at	NM_004219	30
	CENPE	205046_at	NM_001813	31
	NUF2	223381_at	NM_031423	32
	CDCA8	221520_s_at	NM_018101	33
	CENPL	1554271_a_at	NM_033319	34
	ZWINT	204026_s_at	NM_001005413	35
	SPBC25	209891_at	NM_020675	36
	TOP2A	201291_s_at	NM_001067	37
	ESPL1	38158_at	NM_012291	38
Set 3 : fuseau mitotique et centrosome	CDC2	203213_at	NM_001786	39
	TTK	204822_at	NM_003318	40
	RRM2	201890_at	NM_001034	41
	SPAG5	203145_at	NM_006461	42
	FBXO5	234863_x_at	NM_012177	43
	NDE1	222625_s_at	NM_017668	44
	CDC6	203967_at	NM_001254	45
	PLK4	204886_at	NM_014264	46
	NEK2	204641_at	NM_002497	47
	TPX2	210052_s_at	NM_012112	48
	CEP55	218542_at	NM_018131	49

(suite set 3)	CKAP5	1555278_a_at	NM_001008938	50
Set 4 : Microtubules	KIF11	204444_at	NM_004523	51
	KIF15	219306_at	NM_020242	52
	KIF23	244427_at	NM_004856	53
	KIF4A	218355_at	NM_012310	54
	KIF14	236641_at	NM_014875	55
	KIF18A	221258_s_at	NM_031217	56
	KIF20A	218755_at	NM_005733	57
	KIF2C	209408_at	NM_006845	58
Cytokinèse	ECT2	219787_s_at	NM_018098	59
	ANLN	1552619_a_at	NM_018685	60
	PBK	219148_at	NM_018492	61
	PRC1	218009_s_at	NM_003981	62
Set 5 : Instabilité selon Carter	RAD51AP1	204146_at	NM_006479	63
	RNASEH2A	203022_at	NM_006397	64
	TRIP13	204033_at	NM_004237	65
Grade dans l'étude	CDCA2	236957_at	NM_152562	66
	C13orf34	219544_at	NM_024808	67

Avantageusement, le pool de polynucléotides peut comprendre les polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58, SEQ ID NO : 24.

- Avantageusement, le pool de polynucléotides peut comprendre les polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58, SEQ ID NO : 24 et au moins un gène dont la séquence est choisie parmi les 62 autres séquences de gènes identifiés dans le cadre de l'invention. En d'autres termes, le pool de polynucléotides peut comprendre les polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58, SEQ ID NO : 24 et au moins un polynucléotide dont la séquence est choisie parmi les séquences SEQ ID NO : 1, SEQ ID NO : 2, SEQ ID NO : 4 à SEQ ID NO : 9, SEQ ID NO : 11 à SEQ ID NO : 23, SEQ ID NO : 25 à SEQ ID NO : 46, SEQ ID NO : 48 à SEQ ID NO : 57, SEQ ID NO : 59 à SEQ ID NO : 67.

- Alternativement, le pool de polynucléotides peut être constitué des polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58 et SEQ ID NO : 24. En d'autres termes, le pool de polynucléotides peut comprendre uniquement les polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58 et SEQ ID NO : 24.

- Alternativement, le pool de polynucléotides de l'invention peut être constitué des polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58, SEQ ID NO : 24 et d'au moins un gène dont la séquence est choisie parmi les 62 autres séquences de gènes identifiés dans le cadre de l'invention. En d'autres termes, le pool de polynucléotides peut être constitué uniquement des

polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58 et SEQ ID NO : 24 et d'au moins un polynucléotide dont la séquence est choisie parmi les séquences SEQ ID NO : 1, SEQ ID NO : 2, SEQ ID NO : 4 à SEQ ID NO : 9, SEQ ID NO : 11 à SEQ ID NO : 23, SEQ ID NO : 25 à SEQ ID NO : 46, SEQ ID NO : 48 à SEQ ID NO : 57, SEQ ID NO : 59 à SEQ ID NO : 67.

Selon un autre mode de réalisation de la présente invention, le pool de polynucléotides de l'invention peut comprendre au moins un polynucléotide choisi dans chacun des sets de polynucléotides suivants :

- Set 1 : SEQ ID NO : 1 à SEQ ID NO : 12 ;
- Set 2 : SEQ ID NO : 13 à SEQ ID NO : 38 ;
- Set 3 : SEQ ID NO : 39 à SEQ ID NO : 50 ;
- Set 4 : SEQ ID NO : 51 à SEQ ID NO : 58, et SEQ ID NO : 59 à SEQ ID NO : 62 ;
- Set 5 : SEQ ID NO : 63 à SEQ ID NO : 65, et SEQ ID NO : 66 à SEQ ID NO : 67.

Selon un autre mode de réalisation de la présente invention, le pool de polynucléotides de la présente invention peut être choisi parmi les Sets 1 à 5. En d'autres termes, le pool d'au moins deux polynucléotides peut être constitué en totalité ou en partie du Set 1, du Set 2, du Set 3, du Set 4 ou du Set 5. En d'autres termes, le pool de la présente invention peut être constitué en totalité ou en partie du Set 1, ou en totalité ou en partie du set 2, ou en totalité ou en partie du set 3, ou en totalité ou en partie du set 4, ou en totalité ou en partie du set 5.

Selon un autre mode de réalisation de la présente invention, le pool de polynucléotides peut comprendre les polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58 et SEQ ID NO : 24 et au moins un polynucléotide choisi dans le set 5. Ce pool de polynucléotides peut comprendre en outre au moins un des autres gènes identifiés dans le cadre de l'invention.

Selon un autre mode de réalisation de la présente invention, le pool de polynucléotides peut être constitué des polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58 et SEQ ID NO : 24 et d'au moins un polynucléotide choisi dans le set 5.

Selon un autre mode de réalisation de la présente invention, le pool de polynucléotides de l'invention comprend les polynucléotides de séquences SEQ ID NO : 1 à SEQ ID NO : 67. Il peut s'agir par exemple d'un pool constitué des séquences SEQ ID NO : 1 à SEQ ID NO : 67.

Quel que soit le mode de réalisation de l'invention, avantageusement, le pool de polynucléotides peut comprendre avantageusement au plus 10 polynucléotides. Il peut s'agir par exemple d'un pool comprenant au plus 10 polynucléotides, comprenant les polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58 et SEQ ID NO : 24 et au moins un polynucléotide de séquence choisie parmi SEQ ID NO : 1, SEQ ID NO : 2, SEQ ID NO : 4 à SEQ ID NO : 9, SEQ ID NO : 11 à SEQ ID NO : 23, SEQ ID NO : 25 à SEQ ID NO : 46, SEQ ID NO : 48 à SEQ ID NO : 57, SEQ ID NO : 59 à SEQ ID NO : 67.

Quel que soit le mode de réalisation de la présente invention, avantageusement le pool de polynucléotides de l'invention est immobilisé sur un support, par exemple un support solide ou un support liquide. Dans le cas où le support est un support liquide, il peut comprendre des billes sur lesquelles sont fixés les acides nucléiques. Le milieu liquide peut être un surnageant de culture cellulaire, du sérum, du plasma, cette liste n'étant pas limitative. Il peut s'agir par exemple du support mis en œuvre dans la technologie Luminex®. Dans le cas où le support est un support solide, il est de préférence choisi dans le groupe comprenant une membrane de nylon, une membrane de nitrocellulose, une lame de verre, des billes de verre, une membrane sur un support de verre ou une puce de silicium, un support plastique. De manière particulièrement préférée, le support solide peut être une puce à acides nucléiques, par exemple une puce ADN, (aussi appelée puce à gènes, biopuce, puce d'expression). De telles puces permettent de mesurer de manière quantitative une variation d'expression (expression différentielle) de deux ou plusieurs polynucléotides du pool de polynucléotides de l'invention entre (i) 2 conditions expérimentales : généralement une condition de référence et une condition pathologique ou (ii) plusieurs tumeurs afin de déterminer une moyenne d'expression, en fonction de laquelle les tumeurs peuvent être classées les unes par rapport aux autres. À titre d'exemple non limitatif, il peut s'agir d'une puce ADN Affimétrie®, ou d'une puce ADN de la société Agilent Technologies.

Les gènes identifiés par les présents inventeurs, qui sont tous impliqués dans le même processus biologique, peuvent en outre être des cibles potentielles de nouvelles approches thérapeutiques ciblant l'étape précoce d'acquisition du potentiel métastatique. En outre, un pronostic vital des patients sur base du profil d'expression de ces gènes peut être réalisé très tôt, voire lors du diagnostic initial.

Ainsi selon un mode particulier de mise en œuvre de la présente invention, le pool de polynucléotides de l'invention peut être utilisé pour détecter, pronostiquer, diagnostiquer un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST), ou pour surveiller le traitement d'un patient atteint d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST).

Selon un autre mode particulier de mise en œuvre de la présente invention, le pool de polynucléotides de l'invention peut être utilisé pour l'obtention d'un composé destiné au traitement d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST).

5 Afin d'identifier, généralement à partir des données de puces à ADN d'expression, le profil d'expression associé à un groupe de pronostic, deux approches principales peuvent être utilisées, l'approche descendante ou « top-down » supervisée destinée à sélectionner les gènes directement corrélés à un mauvais pronostic (VAN'T VEER et al., 2002, précité ; SOTIRIOU et al., J. Natl. Cancer Inst., 98(4) : 262-272, 2006)
10 et l'approche « bottom-up » supervisée par laquelle les profils d'expression associés à un phénotype biologique particulier sont tout d'abord identifiés puis ultérieurement corrélés à un résultat clinique (SOTIRIOU et al., N. Engl. J. Med., 360(8) : 790-800, 2009). Dans le cadre de la présente invention, la seconde approche « bottom-up » (lit. de bas en haut) a été appliquée au sens où les profils d'expression tumoraux ont été comparés en fonction des
15 phénotypes biologiques (instabilité chromosomique, complexité génomique et grades histologiques) mais au lieu d'une sélection directe des gènes, les voies biologiques particulièrement pertinentes pour les phénotypes testés ont tout d'abord été identifiées puis les gènes significativement impliqués dans ces voies ont été identifiés. Cette sélection de
20 voie biologique (et non de gènes) est l'étape importante ayant conduit aux résultats heureux de la présente invention dans un groupe hétérogène tel que celui des sarcomes à non-translocation, et par ailleurs dans différents types de tumeurs telles que les GISTs (tumeurs stromales gastro-intestinales) et les cancers du sein.

La présente invention se rapporte donc également à un procédé *in vitro* de sélection d'un pool de polynucléotides, par exemple ceux de l'invention, comprenant
25 les étapes suivantes :

- a) fournir des échantillons biologiques tumoraux provenant de patients atteints d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST) ;
- b) détecter et/ou quantifier chacun des polynucléotides, séparément dans
30 chacun des échantillons biologiques tumoraux ;
- c) comparer le profil d'expression des pools de polynucléotides obtenus à l'étape c) par rapport à un phénotype biologique, de préférence d'instabilité chromosomique, de complexité génomique ou de grade histologique, pour chacun des échantillons biologiques tumoraux ;
- 35 d) sélectionner la voie biologique statistiquement significative ($p < 10^{-5}$) pour le phénotype testé ;

e) sélectionner les polynucléotides significativement impliqués dans cette voie biologique, et dont l'expression est indicative de la probabilité de survenue de métastases.

On entend par « profil d'expression », la totalité des résultats obtenus quand l'expression d'un ensemble de polynucléotides est déterminée. Un tel profil facilite l'utilisation de techniques analytiques statistiques quantitatives et permet une comparaison visuelle rapide des résultats. De préférence, un tel profil est obtenu à sur un support solide, tel qu'une puce à ADN.

Par « phénotype biologique », au sens de la présente invention, on entend la manifestation d'un statut génétique, soit l'ensemble des caractéristiques observables caractérisant un échantillon issu d'un patient atteint d'un STS ou d'un GIST, qui reflètent l'expression de l'information portée par les chromosomes (le génotype).

Par « instabilité chromosomique », au sens de la présente invention, on entend des réarrangements clonaux ou non clonaux. Cette instabilité conduit à des pertes et des gains de bras chromosomiques et à des réarrangements chromosomiques non équilibrés. L'instabilité des chromosomes à l'intérieur du noyau des cellules d'un individu rend celles-ci plus vulnérables en terme de néoplasie (survenue de cancer). Ce sont dans les cellules tumorales que l'on retrouve cette instabilité.

Par « complexité génomique », au sens de la présente invention, on entend la détermination du nombre de déséquilibres et de la nature des fragments chromosomiques impliqués.

Par « grade histologique », au sens de la présente invention, on entend un indicateur consensuel de la prolifération tumorale, du risque de métastases et de la réponse à une thérapie adjuvante (chimiothérapie). Le grade histologique ou tumoral est un facteur décisionnel pour le traitement d'une tumeur. Il est déterminé par l'analyse histologique de la tumeur et le système de grading utilisé est par exemple celui de la FNCLCC. Ce système adopté par la Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC) repose sur les 3 caractéristiques suivantes :

Différenciation tumorale	Index mitotique	Nécrose tumorale
Score 1 Sarcome qui ressemble à un tissu adulte normal Exemple : liposarcome bien différencié	Score 1 0-9 mitoses pour 10 champs Score 2 10-19 mitoses pour 10 champs Score 3 > 19 mitoses pour 10 champs	Score 0 pas de nécrose Score 1 < 50 % de nécrose tumorale Score 2 > 50 % de nécrose tumorale
Score 2 Sarcome pour lequel le diagnostic de type histologique est certain Exemple : liposarcome myxoïde	un champ mesure 0,1734 mm ²	

Score 3 Sarcome embryonnaire, sarcome épithélioïde, synoviosarcome, sarcome à cellules claires, sarcome alvéolaire des parties molles, sarcome indifférencié et sarcome pour lequel le type histologique est incertain		
--	--	--

Le grade histologique des tumeurs des tissus mous de la FNCLCC est la somme des 3 scores "Différenciation", "Index mitotique" et "Nécrose tumorale" : Grade 1 (score total de 2 ou 3), Grade 2 (score total de 4 ou 5), et Grade 3 (score total de 6 à 8).

La présente invention se rapporte également à un procédé *in vitro* d'analyse d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST), ledit procédé comprenant la détermination du niveau d'expression d'un pool de polynucléotides selon l'invention dans un échantillon biologique tumoral.

Par « échantillon biologique tumoral », au sens de la présente invention, on entend un échantillon tissulaire issu au choix (i) d'une tumeur primaire (ii) du centre d'une tumeur (iii) d'un site dans la tumeur autre que le centre et (iv) de toute tumeur localisée à l'extérieur du tissu tumoral *per se* d'un patient atteint d'un STS. Ledit échantillon biologique tumoral peut provenir par exemple d'un acte chirurgical ou d'une résection tumorale réalisée sur le STS du patient, d'une biopsie où une partie du tissu tumoral est collecté du STS du patient pour une analyse ultérieure ; d'un échantillon sanguin, par exemple de sang entier, plasma ou sérum, contenant des cellules tumorales issues de la tumeur primaire ou des protéines tumorales produites par les cellules tumorales issues de la tumeur primaire.

Le niveau d'expression d'un pool de polynucléotides de la présente invention peut être déterminée par n'importe quelle méthode connue de l'art. Par exemple, le niveau d'expression d'au moins deux polynucléotides impliqués dans la signature moléculaire de l'invention dans les échantillons obtenus de patients atteints d'un STS peut être déterminé en mesurant le niveau d'ARNm correspondant au polynucléotide et/ou la protéine codée par le polynucléotide. L'ARN peut être isolé des échantillons par des méthodes bien connues de l'homme du métier, par exemple par celle décrite dans AUSUBEL et al. (Curr. Protocols Mol. Biol., 1 : 4.1.1-4.2.9 et 4.5.1-4.5.3, John Wiley & Sons, Inc., 1996). Les méthodes pour la détection du niveau d'expression de l'ARNm utilisables pour mettre en œuvre la présente invention sont bien connues de l'art et comprennent, sans s'y limiter, les puces d'expression, le northern blotting, la PCR quantitative en temps réel, la RT-PCR, la RT-PCT avec sondes Taqman ou les cartes micro fluides, et de manière générale, des techniques d'hybridation (à savoir d'association par

des liaisons non covalentes de deux polynucléotides simple brin, totalement complémentaires ou suffisamment complémentaires pour s'hybrider entre elles, et former une structure double-brin).

Avantageusement, lorsque le pool de polynucléotides comprend au plus
5 10 polynucléotides, le niveau d'expression d'un pool de polynucléotides de la présente invention peut être déterminée par PCR quantitative en routine. Il peut en outre être possible d'utiliser des ARN issus de blocs de paraffine contenant des prélèvements de tissus ou d'organes, ou d'échantillons biologiques.

Selon l'invention, une méthode particulièrement efficace pour détecter le
10 niveau des transcrits d'ARNm exprimés à partir d'une pluralité des polynucléotides décrits implique l'hybridation d'ARNm marqué à une puce d'oligonucléotides (aussi appelée puce à ADN, puce à gènes, puces d'expression). Une telle méthode permet de déterminer simultanément le niveau de transcription d'une pluralité de polynucléotides pour générer des profils d'expression des polynucléotides.

15 Les oligonucléotides utilisés dans cette méthode d'hybridation sont généralement fixés sur un support, par exemple un support solide ou un support liquide. Dans le cas où le support est un support liquide, il peut comprendre des billes sur lesquelles sont fixés les acides nucléiques. Le milieu liquide peut être un surnageant de culture cellulaire, du sérum, du plasma, cette liste n'étant pas limitative. Il peut s'agir par
20 exemple du support mis en œuvre dans la technologie Luminex®. Des exemples de supports solides comprennent, sans s'y limiter, des membranes, des filtres, des lames, du papier, du nylon, des fibres, des billes magnétiques ou non, des gels, des polymères et tout support solide connu de l'homme du métier. N'importe quel support solide sur lequel des oligonucléotides peuvent être immobilisés, soit directement soit indirectement, soit de
25 manière covalente soit de manière non-covalente, peut être utilisé. Un support solide particulièrement avantageux consiste en une puce à acides nucléiques, en particulier une puce à ADN. Ces puces contiennent une sonde oligonucléotidique particulière en une localisation prédéfinie de la puce. Chaque localisation prédéfinie peut contenir plus d'une molécule de la sonde particulière. Du fait que les oligonucléotides se situent à des endroits
30 spécifiques du support, les profils d'hybridation et les intensités (qui forment ensemble un profil d'expression unique) peuvent être interprétés en termes de niveau d'expression de polynucléotides particuliers.

Les sondes oligonucléotidiques sont de préférence de longueur suffisante pour s'hybrider spécifiquement, seulement aux transcrits complémentaires des
35 polynucléotides de l'invention. Par « oligonucléotides », au sens de la présente invention, on entend un acide nucléique simple brin. Généralement les sondes oligonucléotidiques sont constituées de 16-20 nucléotides, et dans certains cas jusqu'à 25 nucléotides, voire jusqu'à 500 nucléotides ou plus.

Une fois que les sondes sont mises en contact avec l'ARNm ou une copie de l'ADNc, la présence de l'ARNm ou de l'ADNc hybridé est détectée par des méthodes connues de l'art. Par exemple, les sondes oligonucléotidiques sont marquées par un ou plusieurs marqueurs pour permettre la détection des complexes sonde
5 hybridée/polynucléotide cible. Les marqueurs peuvent comprendre des compositions pouvant être détectées par des moyens spectroscopiques, biochimiques, photochimiques, bioélectroniques, immunochimiques, électriques, optiques ou chimiques. Des exemples de marqueurs comprennent, sans s'y limiter, des radioisotopes, des composés chimioluminescents, des protéines de liaison marquées, des atomes de métaux lourds, des
10 marqueurs spectroscopiques, tels que les marqueurs fluorescents et colorants, les enzymes liées, des étiquettes de spectrométrie de masse et des marquages magnétiques. Par exemple, il s'agit de marquage Cy³/Cy⁵ ou Alexa pour les biopuces, de marquage FAM (6-carboxyfluorescéine) ou TAMRA (6-carboxy-tétraméthyl-rhodamine) pour des sondes Taqman.

15 Les puces de sonde oligonucléotidique pour la surveillance de l'expression peuvent être préparées et utilisées selon des méthodes bien connues de l'art, comme décrit par exemple dans LOCKHART et al. (Nature Biotechnol., 14 : 1675-1680, 1996 ; McGALL et al., Proc. Natl. Acad. Sci. USA., 93 : 13555-13460, 1996 ; US 6,040,138. De telles biopuces sont disponibles dans le commerce auprès, par exemple,
20 d'Affimérix (Santa Clara, Californie).

Il est également possible de détecter l'expression d'une protéine codée par deux ou plusieurs des polynucléotides impliqués dans la signature moléculaire de l'invention. Ceci peut être réalisé par des méthodes bien connues de l'art, tel que, par exemple, l'utilisation d'une sonde qui est marquée de manière détectable, ou qui peut être
25 marquée ultérieurement. Généralement, la sonde est un anticorps qui reconnaît la protéine exprimée. Le niveau d'expression de la protéine dans l'échantillon est alors déterminé par une méthode d'immunodosage utilisant les anticorps, par exemple dot blotting, western blotting, ELISA, immunohistochimie, FACS, etc.....

Selon un mode particulier de mise en œuvre, le procédé de l'invention
30 permet d'établir le pronostic d'un patient atteint d'un STS ou d'un GIST, en particulier permet de déterminer le risque de/prédire la - survenue de métastases.

Par « prédire la survenue de métastases », au sens de la présente invention, on entend déterminer une valeur relative permettant de quantifier la probabilité de survenue de métastases d'un ou plusieurs tissus ou organes, chez un patient atteint d'un
35 STS ou d'un GIST. De préférence, la prédiction de la survenue de métastases est exprimée par une valeur statistique, incluant une valeur p, calculée à partir des valeurs d'expression obtenues pour chacun des polynucléotides testés.

Selon un autre mode particulier de mise en œuvre, le procédé de l'invention permet d'établir le pronostic d'un patient atteint d'un STS ou d'un GIST, en particulier de distinguer des sous-groupes de bon ou mauvais pronostic parmi un groupe de sarcomes des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST) initialement considérés comme appartenant au même grade histologique.

Par « bon pronostic », au sens de la présente invention, on entend l'indication de patients non susceptibles de présenter une rechute, c'est-à-dire la survenue de métastases, pendant leur traitement ou dans les 5 à 6 ans suivant leur traitement, soit une survie sans métastases à long terme significativement différente. Ainsi dans le cadre de la présente invention, on peut considérer que des patients atteints d'un STS ou d'un GIST appartiennent à un sous-groupe de « bon pronostic » lorsque qu'ils sous-expriment les gènes de la signature moléculaire de l'invention et sont sujet à développer des métastases dans moins de 20% des cas tout type de sarcomes, et en particulier dans aucun des cas de GIST. *A contrario* on entend par « mauvais pronostic », l'indication de patients susceptibles de présenter une rechute (survenue de métastases) pendant leur traitement ou dans les 5 à 6 ans suivant leur traitement. Ainsi dans le cadre de la présente invention, on peut considérer que des patients atteints d'un STS ou d'un GIST appartiennent à un sous-groupe de mauvais pronostic lorsqu'ils sur-expriment les gènes de la signature moléculaire et sont sujet à développer des métastases dans au moins 50% des cas.

Avantageusement, la détermination du niveau d'expression du pool de polynucléotides dans le procédé de l'invention est réalisée sur une puce à acides nucléiques, aussi appelée biopuce, puces à ADN, puce à gènes, puce d'expression. De telles puces permettent de mesurer de manière quantitative et de visualiser rapidement une variation du niveau d'expression, ou expression différentielle, de deux ou plusieurs polynucléotides entre (i) 2 conditions expérimentales, par exemple une condition expérimentale de référence et une pathologique, à partir d'un échantillon biologique d'un patient ou (ii) plusieurs tumeurs afin de déterminer une moyenne d'expression, en fonction de laquelle les tumeurs peuvent être classées les unes par rapport aux autres. À titre d'exemple non limitatif, on peut utiliser des puces ADN Affimétrie™, ou des puces ADN de la société Agilent Technologies.

Selon un mode particulier de mise en œuvre, le procédé de l'invention peut être utilisé pour détecter, pronostiquer, diagnostiquer un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST), ou pour surveiller le traitement d'un patient atteint d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST), comprenant la mise en œuvre d'un procédé de l'invention sur les acides nucléiques d'un échantillon biologique dudit patient.

La présente invention se rapporte également à un procédé *in vitro* de prédiction de la survenue de métastases chez un patient atteint d'un sarcome des tissus

mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST) comprenant les étapes suivantes :

a) fournir un échantillon biologique tumoral préalablement collecté dudit patient à tester ;

5 b) déterminer dans ledit échantillon biologique tumoral le niveau d'expression d'un pool de polynucléotides de l'invention ;

 c) comparer le niveau d'expression obtenu à l'étape b) au niveau d'expression du même pool de polynucléotides mesuré dans un échantillon biologique contrôle ; une dérégulation du niveau d'expression du pool d'oligonucléotides par rapport
10 à son niveau d'expression correspondant mesuré dans un échantillon biologique contrôle étant prédictive de la survenue de métastase.

On entend par « dérégulation du niveau d'expression », la sur-expression ou la sous-expression de deux ou plusieurs polynucléotides d'un pool de polynucléotides selon l'invention mesurée dans un échantillon biologique tumoral d'un patient atteint d'un
15 STS ou d'un GIST à tester, par rapport à l'expression correspondante mesurée dans un échantillon biologique contrôle tel que défini ci-après. En particulier, un niveau d'expression supérieur dans l'échantillon biologique tumoral d'un patient atteint d'un STS ou d'un GIST à tester par rapport à celui d'un échantillon biologique contrôle est l'indication d'un patient susceptible de développer des métastases, ce qui est assimilé à
20 l'indication d'un mauvais pronostic. *A contrario*, un niveau d'expression inférieur dans l'échantillon biologique tumoral d'un patient atteint d'un STS ou d'un GIST à tester par rapport à celui d'un échantillon biologique contrôle est l'indication d'un patient non susceptible de développer des métastases, c'est-à-dire assimilé à l'indication d'un bon pronostic.

25 Par « échantillon biologique contrôle », au sens de la présente invention, on entend (i) un échantillon tissulaire issu d'une tumeur d'un autre patient atteint d'un STS ou d'un GIST que celui à tester ou (ii) un échantillon tissulaire d'un sujet sain à savoir un individu ne présentant aucune pathologie ou symptômes pathologiques diagnostiqués par un médecin. Ainsi les tumeurs peuvent être classées les unes par rapport aux autres, en
30 fonction du niveau d'expression des gènes de la signature moléculaire de l'invention dans chaque cas.

La présente invention se rapporte également à un procédé *in vitro* d'évaluation du pronostic d'un patient atteint d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST), comprenant les étapes suivantes :

35 a) fournir un échantillon biologique tumoral préalablement collecté du patient atteint d'un STS ou d'une tumeur stromale gastro-intestinale (GIST) à tester ;

 b) déterminer dans ledit échantillon biologique tumoral le niveau d'expression d'un pool de polynucléotides de l'invention ;

c) comparer le niveau d'expression obtenu à l'étape b) au niveau d'expression du même pool de polynucléotides mesuré dans un échantillon biologique contrôle,

5 où une dérégulation du niveau d'expression du pool d'oligonucléotides par rapport à son niveau d'expression correspondant mesuré dans un échantillon biologique contrôle permet d'identifier un sous-groupe de bon pronostic ou un sous-groupe de mauvais pronostic.

La présente invention rapporte également à un procédé *in vitro* de criblage de composés candidats pour le traitement d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST) comprenant les étapes suivantes :

a) mettre en contact un échantillon biologique tumoral préalablement collecté avec un composé test ;

b) déterminer dans ledit échantillon biologique tumoral le niveau d'expression d'un pool de polynucléotides de l'invention ;

15 c) comparer ledit niveau d'expression obtenue dans l'étape b) avec celui du même échantillon biologique tumoral n'ayant pas été mis en contact avec le composé test,

20 où une diminution du niveau d'expression dans l'échantillon biologique tumoral en présence du composé test par rapport à celui de l'échantillon biologique tumoral en absence du composé test est l'indication d'un composé candidat pour le traitement d'un STS ou d'un GIST.

La présente invention se rapporte également à un procédé *in vitro* de surveillance de l'efficacité anti-métastase d'un traitement d'un patient atteint d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST), comprenant les

25 étapes suivantes :

a) fournir un échantillon biologique tumoral préalablement collecté dudit patient traité à tester ;

b) déterminer dans ledit échantillon biologique tumoral le niveau d'expression d'un pool de polynucléotides de l'invention ;

30 c) comparer ledit niveau d'expression obtenu à l'étape b) avec celui d'un échantillon biologique contrôle ou d'un échantillon biologique tumoral dudit patient avant traitement,

35 où une diminution du niveau d'expression de l'échantillon biologique tumoral après traitement par rapport à celui de l'échantillon biologique contrôle ou de l'échantillon biologique tumoral avant traitement est l'indication d'une efficacité anti-métastase du traitement thérapeutique.

La présente invention se rapporte, en quatrième lieu, à une trousse (« kit ») comprenant un pool de polynucléotides de l'invention.

Selon l'invention, cette trousse ou kit peut être utilisée par exemple pour la prédiction *in vitro* de la survenue de métastases et/ou pour l'évaluation du pronostic d'un patient atteint d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST) et/ou pour la surveillance de l'efficacité anti-métastase d'un traitement thérapeutique d'un patient atteint d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST).

Selon l'invention, cette trousse ou kit peut comprendre en outre les moyens de détection et/ou quantification de l'expression d'un pool de nucléotides de l'invention. Ces moyens peuvent être par exemple l'un de ceux définis ci-dessus ou exemplifiés ci-dessous.

La présente invention se rapporte, en cinquième lieu, à une puce à acides nucléiques, en particulier à une puce à ADN, comprenant ou étant constituée d'un pool de polynucléotides de l'invention. Cette puce à ADN peut être par exemple telle que définie ci-dessus, notamment concernant le support.

Avantageusement une puce à acides nucléiques de l'invention peut comprendre des « sondes », par exemple des fragments d'ADNc ou oligonucléotides (par exemple de 60 à 80 bases, ou plus), etc..., fixées sur un support solide. Ces « sondes » fixent de façon spécifique par hybridation les « cibles », par exemple les gènes complémentaires, présents dans les échantillons biologiques à tester. Cette hybridation nécessite l'association par des liaisons non covalentes des séquences d'acides nucléiques simple brin, totalement complémentaires ou suffisamment complémentaires pour s'hybrider entre elles, et former une structure double-brin.

BREVE DESCRIPTION DES FIGURES

- La Figure 1 représente 3 types de profil génomique (a) amplifié (16%) (a) bras (23%) et (d) réarrangé (61%).

- La Figure 2 représente les courbes de survie sans métastases de Kaplan-Meier de différents groupes de sarcomes selon la signature CINSARC.

- La Figure 3 représente les courbes de survie sans progression/métastases de Kaplan-Meier de trois groupes de tumeurs selon la signature CINSARC.

- La Figure 4 représente les courbes de survie sans progression/métastases (% de cas sans métastases en fonction des années après le traitement) de Kaplan-Meier d'un groupe de sarcomes (groupe de tumeurs dans lequel a été définie la signature) selon la signature au moyen du pool de nucléotides consistant en les polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58 et SEQ ID NO : 24. La courbe A montre une courbe de survie de patients à bon pronostique, présentant environ 80% de cas sans métastases à 5 ans. La courbe B montre

une courbe de survie de patients à un mauvais pronostique, présentant environ 50% de cas sans métastases à 5 ans.

- La Figure 5 représente les courbes de survie sans progression/métastases (% de cas sans métastases en fonction des années après le traitement) de Kaplan-Meier d'un groupe de sarcomes (groupe de tumeurs indépendant de l'identification de la signature) selon la signature au moyen du pool de nucléotides consistant en les polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58 et SEQ ID NO : 24. La courbe A montre une courbe de survie de patients à bon pronostique, présentant environ 90% de cas sans métastases à 5 ans. La courbe B montre une courbe de survie de patients à un mauvais pronostique, présentant environ 50% de cas sans métastases à 5 ans.

EXEMPLES

Exemple 1 : pool de la présente invention

Patients et échantillons

- La base de données du groupe sarcome Français (GSF) en tant qu'entité de la Conticabase (www.conticabase.org) contient les données de sarcomes des tissus mous d'adulte traités dans 11 centres avec la description des patients, des tumeurs primaires, des traitements, du suivi et des échantillons tumoraux. Cette base de données contenait environ 3800 cas au moment de l'étude. Tous les cas ont été revus par le sous-groupe de pathologistes et classés selon la classification WHO 2002 en utilisant l'histologie, l'immunohistochimie et la cytogénétique et génétique moléculaire lorsque cela a été nécessaire. Pour cette étude, des sarcomes des tissus mous sans translocations chromosomiques récurrentes ont été sélectionnés et pour lesquels un échantillon tissulaire congelé de la tumeur primaire non traitée était disponible. Au final les échantillons biologiques provenant de 183 patients décrits dans le tableau 2 ci-dessous ont été étudiés.

Tableau 2

Caractéristique	Groupe des tumeurs (n=183)	Valeur p
Suivi moyen (mois)	64	
Age moyen (années)	63	0,07
Ecart-type	15	
Sexe masculin (%)	98(53)	0,75
Grade FNCLCC (%)		0,43
1	7 (4)	

2	62 (34)	
3	102 (56)	
nd	12(6)	
Histotype (%)		0,01
Sarcomes indifférenciés	71 (39)	
Leiomyosarcomes	52 (28)	
Liposarcomes dédifférenciés	44 (24)	
Autres	16 (9)	
Localisation (%)		0,23
Paroi du tronc	28 (16)	
Membres	113 (62)	
Tête et cou	2 (1)	
Tronc interne	39 (21)	
Taille moyenne (cm)	10	0,33
Tumeur profonde (%)		nd
Oui	173 (95)	
Non	10 (5)	
Envahissement vasculo-nerveux ou osseux (%)		<0,001
Oui	27 (15)	
Non	156 (85)	
Événements de rechute (%)		
Métastases	78 (43)	
Rechutes locales	72 (39)	
Type de traitement (%)		
Chirurgie	50 (27)	
Chirurgie+radiothérapie	80 (44)	
Chirurgie+chimiothérapie	10 (5)	
Chirurgie+radiothérapie+chimiothérapie	40 (22)	
Données manquantes	3 (2)	

Extraction ADN et analyse par CGH (Hybridation Génomique Comparative) sur puce ADN

- 5 L'ADN génomique de tissus tumoraux congelés a été isolé en utilisant un protocole d'extraction au phénol-chloroforme standard et a été analysé sur un spectrophotomètre (Nanodrop). Ainsi après digestion avec DpnII (Ozyme, Saint-Quentin en Yvelines, France) et purification sur colonne (Qiagen PCR Purification Kit, Qiagen), 1,5 µg d'ADN tumoral et 1,5 µg d'ADN normal ont été marqués en utilisant le BioPrime

DNA labeling System Kit (Invitrogen, Cergy Pontoise, France) avec Cy5-dCTP ou Cy3-dCTP (Perkin Elmer), respectivement. Les ADNs normaux et tumoraux marqués ont été mélangés et précipités ensemble avec 100 µg d'ADN Cot-1 humain (Invitrogen), re-suspendus dans 72 µl de tampon d'hybridation (50% formamide, 40 mM NaH₂PO₄, 0,1% SDS, 10% sulfate de dextran, 2X SSC). Des sondes préhybridées ont été déposées sur des lames et introduites dans des chambres humides (Corning) et l'hybridation a eu lieu à 37°C pendant 48 h.

Afin d'établir les profils génomiques, des puces BAC (Bacterial Artificial Chromosome) composées de 3803 clones BAC ont été réalisées avec une moyenne de 1 Mb entre les clones. Les clones BAC ont été déposés en trois exemplaires.

Les lavages après l'hybridation ont été réalisés comme suit : un lavage à 65°C dans du 0,5X SSC, 0,03% SDS, suivi d'un lavage à 45°C dans la même solution.

Les lames ont été scannées (Scanarray 4000XL, Packard Bioscience) et analysées avec le logiciel d'analyse d'images GenePix Pro 5.1. La normalisation, la filtration subdivisée, l'analyse par groupement et la représentation graphique ont été réalisées en utilisant la plateforme d'analyse par CGH sur puce ADN (CAPWeb). Les clones avec plus de 50% de valeurs manquantes ont été éliminés. Les ratios Cy5-Cy3 supérieurs à 2 ont été considérés comme des amplifications, des ratios supérieurs à 1,2 et inférieurs à 0,8 ont été considérés comme des gains et des pertes, respectivement.

L'analyse par CGH sur puce ADN (calcul des altérations génomiques) a été réalisée par l'interface VAMP (LA ROSA et al., Bioinformatics, 22(17) : 2066-2073, 2006).

Extraction ARN et analyse d'expression

L'ARN total a été extrait à partir d'échantillon tumoraux congelés avec le réactif TRIzol (Life technologies, Inc.). L'ARN a alors été purifié en utilisant le RNeasy® Min Elute™ Cleanup Kit (Qiagen), selon les instructions du fabricant. La qualité de l'ARN a été vérifiée sur le bioanalyseur Agilent 2100 (Agilent Technologies).

Les échantillons ont ensuite été analysés sur puce Human genome U133 Plus 2.0 (Affimetrix®), selon les instructions du fabricant. Toutes les données des puces à ADN ont été normalisées simultanément en utilisant l'algorithme GCRMA (WU et al., J. Am. Stat. Assoc., 99 : 909-917, 2004). Les analyses de groupement hiérarchiques ont été réalisées en utilisant le logiciel dChip (<http://biosun1.harvard.edu/complab/dchip/>). Pour les tests de Welch, Willcoxon et SAM, les valeurs p ont été ajustées en utilisant la procédure de Benjamini-Hochberg (R-multitest package).

L'analyse dans la base de données Gene Ontology (GO ; <http://www.geneontology.org/>) a été réalisée pour trouver un enrichissement statistique aux limites de la GO.

Analyse statistique

Des tests de Chi-deux (test χ^2) ont été réalisés pour évaluer le lien entre les différentes caractéristiques tumorales, altérations génomiques, profils d'expression et résultat clinique. L'influence réciproque parmi les différents facteurs prédictifs a été déterminée par une analyse multivariée en utilisant un test de régression logistique amont. Tous les facteurs ont été inclus dans les analyses de régression logistique, sans tenir compte de leurs valeurs P obtenues par analyse univariée, mais seulement ceux avec une valeur $P \leq 5\%$ ont été retenus dans les modèles finaux. Les survies sans métastases ont été obtenues par la méthode de Kaplan-Meier et comparées avec le test logarithmique par rangs. Tous les tests statistiques ont été à deux faces et le seuil de signification a été de $p = 0,05$. Toutes les analyses statistiques (modèle de régression logistique) ont été réalisées en utilisant la version 8 du logiciel SAS.

RÉSULTATS

Profil génomique des 183 sarcomes mal différenciés

L'établissement du profil génomique des 183 sarcomes peu différenciés a été réalisé par analyse CGH sur une puce BAC contenant 3803 clones. Trois profils récurrents principaux ont été identifiés, selon à la fois le nombre et le type d'altérations identifiés, parmi 174 profils génomiques interprétables *in fine* (Figure 1). Un premier groupe de 28 tumeurs (16%) à génétique simple, dénommé profil « amplifié », basé sur les co-amplifications et correspondant presque exclusivement à des liposarcomes dédifférenciés. Un second groupe de 40 tumeurs (23%), dénommé profil « bras », avec quelques altérations (moins de 30), impliquant principalement une altération du chromosome entier ou de la totalité d'un bras du chromosome. Un troisième groupe de 106 tumeurs (61%), dénommé profil « réarrangé », caractérisé par un niveau élevé de complexité chromosomique avec plus de 30 à 85 altérations.

Il reste à démontrer si le profil génomique est associé au résultat clinique.

L'analyse de groupement supervisée selon le profil génomique (profil « bras » vs profil « réarrangé ») n'a pas permis de prédire de manière significative la survenue de métastases ($p = 0,17$). De manière intéressante, une corrélation positive a été trouvée entre le profil « réarrangé » et le grade 3 histologique, dans les 183 sarcomes de l'étude ($p = 0,001$), et dans le sous-groupe des 117 sarcomes des membres à génétique complexe avec les profils « bras » et « réarrangé » ($p = 2,2 \times 10^{-4}$). Le grade histologique étant une évaluation indirecte de l'agressivité tumorale, il a été montré que, même si aucune corrélation avec un mauvais résultat clinique n'a été obtenue, la complexité génomique est associée à l'agressivité tumorale.

Il reste à démontrer si l'expression des gènes associée à la complexité génomique et/ou le grade tumoral pourrait être prédictif de la survenue de métastases.

Profils d'expression et établissement de la signature moléculaire pronostique

5 Les profils d'expression génique des 183 sarcomes de l'étude ont été reconsidérés afin de tester l'hypothèse d'une corrélation entre l'expression spécifique des gènes dans les tumeurs à génome complexe et la survenue de métastases.

Pour ce faire, les 183 échantillons ont été groupés dans un premier temps en fonction d'une signature précédemment établie et composée de 70 gènes sélectionnés
10 comme étant liés à l'instabilité chromosomique (CARTER et al., Nat. Genet., 38(9) : 1043-1048, 2006). Mais cela a conduit à une prédiction de tendance mais non significative de la survie sans métastases.

Aussi dans un deuxième temps, le but a été d'établir un jeu de gènes spécifiques des sarcomes associé au niveau des déséquilibres et capable de prédire le
15 devenir d'un patient. Dans trois analyses supervisées, ont été analysés les profils d'expression de tumeurs classées en deux groupes selon i) le nombre de déséquilibres CGH, moins de 20 déséquilibres vs plus de 35 déséquilibres, ii) le grade histologique FNCLCC 3 vs le grade tumoral 2, et iii) la signature de Carter. À partir des deux premières comparaisons, 118 clones correspondant à 86 gènes et 92 clones correspondant à 73 gènes
20 ont été significativement exprimés de manière différentielle entre les tumeurs stratifiées soit par les déséquilibres CGH (facteur d'expression différentielle (= nombre de fois où le gène est plus exprimé) >3 ; ou non, $p < 0,01$) ou par grade (facteur d'expression différentielle >2 ; $p < 0,01$), respectivement. Ces gènes ont ensuite été soumis à l'analyse par la base de données Gene Ontology dans le but de déterminer les voies associées aux
25 déséquilibres CGH et au grade histologique. De manière intéressante, ces voies sont extrêmement similaires dans les groupes déterminés selon les déséquilibres CGH et ceux déterminés par les comparaisons de grade histologique, et sont principalement impliquées dans l'intégrité des chromosomes et le contrôle de la mitose (Tableau 3). Parmi les gènes de la signature de Carter, 22 gènes, qui n'ont pas encore été identifiés dans les deux
30 premières comparaisons, ont été significativement exprimés ($p < 10^{-5}$) de manière différentielle entre les deux groupes de sarcomes.

D'après ces résultats, ont été sélectionnés tous les gènes significatifs appartenant aux voies sur-représentées de manière significative à partir des deux premières comparaisons ($p < 10^{-5}$; tableau 3) et les 22 gènes de la signature de Carter définis ci-
35 dessus.

Tableau 3

a) d'après le test de Welch

Nombre de sondes/clones d'entrée : 92

Nombre d'identifications trouvées : 73				
Identification GO	Observé dans la sélection	Observé sur puce	Valeur p exacte de Fisher	Limite GO
GO:0007067	19	122	1,46E-24	Mitose
GO:0051301	19	174	7,10E-22	Division cellulaire
GO:0007049	21	422	1,36E-17	Cycle cellulaire
GO:0000775	6	37	1,78E-08	Chromosome, région péricentrique
GO:0000074	8	181	1,02E-06	Régulation progression à travers cycle cellulaire
GO:0005694	6	118	1,02E-05	Chromosome
GO:0004674	8	338	7,96E-05	Activité sérine/thréonine kinase
GO:0008283	7	248	8,41E-05	Prolifération cellulaire
GO:0006270	3	19	1,00E-04	Initiation réplication ADN
GO:0000776	3	21	1,20E-04	Centromère
GO:0003777	4	62	1,54E-04	Activité moteur des microtubules
GO:0007018	4	75	3,22E-04	Mouvement basé sur microtubules
GO:0000079	3	35	5,23E-04	Régulation activité kinase cycline dépendante
GO:0005813	3	48	1,14E-03	Centrosome
GO:0005875	3	54	1,58E-03	Complexe associé aux microtubules
GO:0006468	7	475	3,67E-03	Phosphorylation d'acides aminés protéiques
GO:0046982	3	80	4,83E-03	Activité d'hétérodimérisation protéique
GO:0005874	4	178	6,25E-03	Microtubules
GO:0006260	3	96	8,13E-03	Réplication ADN
GO:0016301	3	184	4,18E-02	Activité kinase

b) d'après le test de Welch

Nombre de sondes/clones d'entrée : 118				
Nombre d'identifications trouvées : 86				
Identification GO	Observé dans la sélection	Observé sur puce	Valeur p exacte de Fisher	Limite GO
GO:0007067	23	122	4,50E-28	Mitose
GO:0051301	23	174	7,15E-25	Division cellulaire
GO:0007049	27	422	1,48E-21	Cycle cellulaire
GO:0000775	8	37	9,04E-11	Chromosome, région péricentrique
GO:0005819	6	14	7,45E-10	Fuseau
GO:0007018	9	75	8,12E-10	Mouvement basé sur microtubules

GO:0003777	8	62	1,51E-09	Activité moteur des microtubules
GO:0005876	5	12	2,39E-08	Microtubule du fuseau
GO:0000074	10	181	9,46E-08	Régulation progression à travers cycle cellulaire
GO:0008283	11	248	1,72E-07	Prolifération cellulaire
GO:0005874	8	178	7,01E-06	Microtubules
GO:0007089	3	5	9,18E-06	Traversée point contrôle démarrage cycle cellulaire mitotique
GO:0005875	5	54	1,62E-05	Complexe associé aux microtubules
GO:0005694	6	118	5,52E-05	Chromosome
GO:0005871	3	16	1,39E-04	Complexe kinésine
GO:0000079	3	35	1,23E-03	Régulation activité kinase cycline dépendante
GO:0004674	7	338	1,40E-03	Activité sérine/thréonine kinase
GO:0006468	8	475	5,32E-03	Phosphorylation aminoacide protéique
GO:0006260	3	96	1,79E-02	Réplication ADN
GO:0008284	3	145	4,99E-02	Régulation positive prolifération cellulaire

Ce set final de gènes, dénommé par les inventeurs CINSARC (Complexity INdex SARComas), est constitué de 67 gènes tous impliqués dans le contrôle du génome.

5 **Exemple 2 : Prédiction de la survenue de métastases dans les sarcomes à l'aide du CINSARC**

La corrélation de la signature d'expression CINSARC avec la survenue de métastases a été évaluée dans le groupe entier d'étude (183 sarcomes). L'analyse par groupement a permis la classification des tumeurs en trois sous-groupes (sous-groupes 1, 2, 3), avec une différence significative dans la survenue des métastases (Figure 2). L'analyse multivariée a montré que les tumeurs du sous-groupe 3 ont un risque de métastases triple en comparaison des tumeurs du sous-groupe 1 (analyse de Kaplan-Meier ; HR = 3,01 ; 95% CI [1,8-5,2] ; $p < 10^{-3}$). Une analyse multivariée prenant en compte d'autres facteurs pronostiques standards, tel que le type histologique, le grade tumoral FNCLCC, la taille des tumeurs, la localisation, l'envahissement vasculo-nerveux ou osseux, le sexe et l'âge, a montré également un risque de métastases trois fois supérieur pour la sous-groupe 3 en comparaison du sous-groupe 1 (modèle de Cox ; HR = 3,1 ; 95% CI [1,8-5,4], $p < 10^{-3}$). Ces résultats ont montré que la signature CINSARC est un facteur pronostique indépendant associé fortement au développement de métastases.

Après cette validation de la signature CINSARC en tant que facteur pronostique indépendant, 6 sous-groupes spécifiques de sarcomes ont également été testés par une analyse par groupement non-supervisée (Figure 2). Parmi les 117 sarcomes des

membres à génétique complexe, l'analyse univariée a réparti les tumeurs en deux sous-groupes et démontré un risque de métastases triple pour le sous-groupe 2 vs le sous-groupe 1 (analyse de Kaplan-Meier ; HR = 3,1 ; 95% CI [1,6-6,0] ; $p < 10^{-3}$). De manière similaire, parmi les 52 leiomyosarcomes, trois sous-groupes de résultat clinique significatif différent (p=0,001) ont été trouvés (il est intéressant de noter que le sous-groupe 2 consiste presque exclusivement en des LMS développés dans le tronc interne au lieu du tronc externe pour les deux autres sous-groupes). Aussi lorsque seuls les LMS du tronc externe sont pris en considérations par une analyse par groupement non-supervisée, les 36 patients sont répartis en deux sous-groupes avec une différence sextuple de risque métastatique (analyse de Kaplan-Meier ; HR = 6 ; 95% CI [2,1-16,9] ; $p < 10^{-3}$).

La performance de la signature CINSARC a également été analysée pour des patients du même grade histologique (Figure 2). Au sein des tumeurs de grade 3 (100 cas), un risque métastatique triple a été observé dans les tumeurs du sous-groupe 2 vs les tumeurs du sous-groupe 1 (analyse de Kaplan-Meier ; HR = 3 ; 95% CI [1,6-5,6] ; $p < 10^{-3}$) et au sein des tumeurs de grade 2 (40 cas) avec des profils bras ou réarrangé (à savoir toutes sauf les DD-LPS), les patients ont été également répartis en deux groupes de résultat clinique différent (analyse de Kaplan-Meier ; HR = 2,6 ; 95% CI [1-7,5] ; $p = 0,05$). La survie sans métastases n'est pas significativement différente dans les deux groupes de liposarcomes différenciés regroupés selon la signature CINSARC.

Ainsi la signature CINSARC de la présente invention a permis de séparer en deux groupes ayant une probabilité de survenue de métastases différente, des tumeurs considérées avec le même potentiel métastatique selon le système de grade FNCLLC (Figure 2). Ce résultat est peut-être le plus important, dans la mesure où il démontre clairement que la signature CINSARC peut être un système plus efficace que celui actuellement utilisé pour déterminer des stratégies thérapeutiques.

En outre, pour la première fois dans le domaine des sarcomes, un profil d'expression de gènes attribue un pronostic clinique meilleur que celui obtenu avec le système de grade FNCLLC. Ainsi, dans le groupe entier réunissant différents histotypes, la signature CINSARC a permis l'identification d'un sous-groupe de tumeurs avec un mauvais pronostic alors que le système de grade FNCLLC n'a pas réussi à séparer ces tumeurs de pronostics distincts (données non représentées).

Exemple 3 : Prédiction de la survenue de métastases dans d'autres cancers à l'aide du CINSARC

La valeur prédictive du CINSARC dans d'autres sarcomes a été testée et une série de 32 GIST a été analysée (YAMAGUCHI et al., J ; Clin. Oncol., 26(25) : 4100-4108, 2008). Comme représenté dans la Figure 3, la signature CINSARC a permis une analyse par groupement non-supervisée hiérarchique conduisant à deux groupes de GIST

avec un pronostic différent ($p < 10^{-3}$). De manière intéressante, cette classification est indépendante de la localisation même si les GIST de l'intestin grêle et celles de l'estomac forment deux groupes séparés dans chaque groupe de pronostic différent.

La signature CINSARC étant exclusivement composée de gènes impliqués dans l'intégrité des chromosomes et l'expression étant associée aux déséquilibres chromosomiques, la signature CINSARC pourrait également avoir une valeur pronostique pour des tumeurs fortement réarrangées, tel que les carcinomes du sein. Par conséquent, deux séries de cancer du sein (78 et 295 cas) de l'Institut du Cancer des Pays-Bas (VAN'T VEER et al., 2002, précité ; VAN de VIJVER et al., 2002, précité) ont été rassemblées selon la signature CINSARC, et une fois encore deux groupes de patients avec un résultat clinique différent très significatif ($p < 10^{-3}$) ont été obtenus (Figure 3).

Comme démontré dans l'étude, la signature CINSARC est un outil prédictif indépendant puissant permettant une meilleure évaluation de la survenue des métastases ainsi que l'attribution aux patients d'un meilleur pronostic clinique par rapport au système de grade FNCLCC. Ce nouveau système de grade moléculaire devrait ainsi permettre d'améliorer le suivi clinique des patients. En outre cette signification biologique des gènes de la signature CINSARC les définit comme des cibles potentielles de nouvelles approches thérapeutiques ciblant l'étape précoce d'acquisition du potentiel métastatique.

Le fait que la signature CINSARC soit associée à la survenue de métastases à travers de tels groupes hétérogènes de tumeurs (des sarcomes aux carcinomes) est suffisamment encourageant pour envisager, à la place du système de grade histologique actuel, l'utilisation de ce profil d'expression pour identifier des patients à risque élevé de métastases et cibler des stratégies chimio-thérapeutiques complémentaires.

Les stratégies thérapeutiques courantes associent la résection chirurgicale et la chimiothérapie/radiothérapie dans des situations adjuvante ou néo-adjuvante. Cependant seuls les sarcomes ayant un potentiel métastatique élevé devraient bénéficier d'un tel traitement. C'est actuellement le cas pour les GIST dont le traitement adjuvant par l'imatinib est en cours de validation pour les tumeurs à haut risque de récurrence. Cependant les systèmes actuellement utilisés sont imparfaits. L'utilisation de la signature CINSARC pourrait améliorer la sélection de ces patients et ainsi augmenter le bénéfice des thérapies adjuvantes.

Il existe donc un intérêt important dans l'utilisation de la signature CINSARC en tant que critère majeur de décision pour l'admissibilité d'une thérapie adjuvante, en particulier concernant les GIST (tumeurs stromales gastro-intestinales) pour lesquelles une thérapie ciblée existe déjà (Glivec®).

Exemple 4 : Prédiction de la survenue de métastases dans les sarcomes à l'aide d'un pool de 5 polynucléotides du CINSARC

La corrélation de la signature d'expression des 5 polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58 et SEQ ID NO : 24 du CINSARC avec la survenue de métastases a été évaluée sur deux séries de sarcomes (Figures 4 et 5). L'analyse par groupement a permis la classification des tumeurs en deux sous-groupes (sous-groupes A et B), avec une différence significative dans la survenue des métastases. L'analyse par la méthode des centres les plus proches a montré que les tumeurs du sous-groupe B ont un risque de métastases supérieur en comparaison des tumeurs du sous-groupe A. Ces résultats ont montré que la signature CINSARC à 5 gènes est un facteur pronostique indépendant associé fortement au développement de métastases.

Ce résultat est important, dans la mesure où il démontre clairement que la signature de cinq gènes du CINSARC peut être un système plus efficace que celui actuellement utilisé pour déterminer des stratégies thérapeutiques.

En outre, pour la première fois dans le domaine des sarcomes, un profil d'expression de gènes attribue un pronostic clinique meilleur que celui obtenu avec le système de grade FNCLLC. Ainsi, dans le groupe entier réunissant différents histotypes, la signature CINSARC a permis l'identification d'un sous-groupe de tumeurs avec un mauvais pronostic alors que le système de grade FNCLLC n'a pas réussi à séparer ces tumeurs de pronostics distincts (données non représentées).

REVENDICATIONS

1. Pool de polynucléotides comprenant au moins deux polynucléotides choisis parmi les séquences de polynucléotides SEQ ID NO : 1 à SEQ ID NO : 67.

5

2. Pool de polynucléotides selon la revendication 1, ledit pool de polynucléotides comprenant les polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58 et SEQ ID NO : 24.

10

3. Pool de polynucléotides selon la revendication 1, ledit pool de polynucléotides comprenant au moins un polynucléotide choisi dans chacun des sets de polynucléotides suivants :

Set 1 : SEQ ID NO : 1 à SEQ ID NO : 12 ;

Set 2 : SEQ ID NO : 13 à SEQ ID NO : 38 ;

15

Set 3 : SEQ ID NO : 39 à SEQ ID NO : 50 ;

Set 4 : SEQ ID NO : 51 à SEQ ID NO : 58, et SEQ ID NO : 59 à SEQ ID NO : 62 ;

Set 5 : SEQ ID NO : 63 à SEQ ID NO : 65, et SEQ ID NO : 66 à SEQ ID NO : 67.

20

4. Pool de polynucléotides selon la revendication 3, ledit pool comprenant les polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58 et SEQ ID NO : 24 et au moins un polynucléotide du Set 5.

25

5. Pool de polynucléotides selon la revendication 1, ledit pool comprenant au moins deux polynucléotides choisis dans:

Set 1 : SEQ ID NO : 1 à SEQ ID NO : 12 ; ou

Set 2 : SEQ ID NO : 13 à SEQ ID NO : 38 ; ou

Set 3 : SEQ ID NO : 39 à SEQ ID NO : 50 ; ou

30

Set 4 : SEQ ID NO : 51 à SEQ ID NO : 58, et SEQ ID NO : 59 à SEQ ID NO : 62 ; ou

Set 5 : SEQ ID NO : 63 à SEQ ID NO : 65, et SEQ ID NO : 66 à SEQ ID NO : 67.

35

6. Pool de polynucléotides selon l'une quelconque des revendications précédentes, ledit pool de polynucléotides comprenant au plus dix polynucléotides.

7. Pool de polynucléotides selon la revendication 1, ledit pool de polynucléotides étant constitué des polynucléotides de séquences SEQ ID NO : 10, SEQ ID NO : 3, SEQ ID NO : 47, SEQ ID NO : 58, SEQ ID NO : 24.

5 8. Pool de polynucléotides selon la revendication 1, ledit pool de polynucléotides comprenant les polynucléotides de séquences SEQ ID NO : 1 à SEQ ID NO : 67.

9. Pool de polynucléotides selon l'une quelconque des revendications 1 à 3, immobilisé sur un support solide ou liquide.

10. Utilisation d'un pool de polynucléotides selon l'une quelconque des revendications 1 à 5, pour l'obtention d'un composé destiné au traitement d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST).

15 11. Procédé *in vitro* d'analyse d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST), ledit procédé comprenant une détermination du niveau d'expression d'un pool de polynucléotides selon l'une quelconque des revendications 1 à 5 dans un échantillon biologique.

20 12. Procédé selon la revendication 11, dans lequel ladite détermination du niveau d'expression permet de prédire la survenue de métastases.

25 13. Procédé selon la revendication 11, dans lequel ladite détermination du niveau d'expression permet de distinguer des sous-groupes de bon ou mauvais pronostic parmi un groupe de sarcomes des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST) de même grade histologique.

30 14. Procédé selon l'une quelconque des revendications 11 à 13, dans lequel ladite détermination du niveau d'expression est réalisée sur une puce à acides nucléiques.

35 15. Utilisation d'un procédé selon la revendication 11, pour détecter et/ou pronostiquer et/ou diagnostiquer un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST), ou pour surveiller le traitement d'un patient atteint d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST).

16. Procédé *in vitro* de prédiction de la survenue de métastases chez un patient atteint d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST) comprenant les étapes suivantes :

5 a) fournir un échantillon biologique tumoral préalablement collecté dudit patient à tester ;

b) déterminer dans ledit échantillon biologique tumoral le niveau d'expression d'un pool de polynucléotides selon l'une quelconque des revendications 1 à 4 ;

10 c) c) comparer le niveau d'expression obtenu à l'étape b) au niveau d'expression du même pool de polynucléotides mesuré dans un échantillon biologique contrôle, une dérégulation du niveau d'expression du pool d'oligonucléotides par rapport à son niveau d'expression correspondant mesuré dans un échantillon biologique contrôle étant prédictive de la survenue de métastase.

15 17. Procédé *in vitro* d'évaluation du pronostic d'un patient atteint d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST), comprenant les étapes suivantes :

a) fournir un échantillon biologique tumoral préalablement collecté du patient atteint d'un STS à tester ;

20 b) déterminer dans ledit échantillon biologique tumoral le niveau d'expression d'un pool de polynucléotides selon l'une quelconque des revendications 1 à 4 ;

25 c) comparer le niveau d'expression obtenu à l'étape b) au niveau d'expression du même pool de polynucléotides mesuré dans un échantillon biologique contrôle,

où une dérégulation du niveau d'expression du pool d'oligonucléotides par rapport à son niveau d'expression correspondant mesuré dans un échantillon biologique contrôle permet d'identifier un sous-groupe de bon pronostic ou un sous-groupe de mauvais pronostic.

30

18. Procédé *in vitro* de criblage de composés candidats pour le traitement d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST) comprenant les étapes suivantes :

35 a) mettre en contact un échantillon biologique tumoral avec un composé test ;

b) déterminer dans ledit échantillon biologique tumoral le niveau d'expression d'un pool de polynucléotides selon l'une quelconque des revendications 1 à 4 ;

c) comparer ledit niveau d'expression obtenue dans l'étape b) avec celui du même échantillon biologique tumoral n'ayant pas été mis en contact avec le composé test,

5 où une diminution du niveau d'expression dans l'échantillon biologique tumoral en présence du composé test par rapport à celui de l'échantillon biologique tumoral en absence du composé test est l'indication d'un composé utile dans le traitement d'un STS ou d'un GIST.

10 19. Procédé *in vitro* de surveillance de l'efficacité anti-métastase d'un traitement d'un patient atteint d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST), comprenant les étapes suivantes :

a) fournir un échantillon biologique tumoral préalablement collecté dudit patient traité à tester ;

15 b) déterminer dans ledit échantillon biologique tumoral le niveau d'expression d'un pool de polynucléotides selon l'une quelconque des revendications 1 à 4 ;

c) comparer ledit niveau d'expression obtenu à l'étape b) avec celui d'un échantillon biologique contrôle ou d'un échantillon biologique tumoral dudit patient avant traitement,

20 où une diminution du niveau d'expression de l'échantillon biologique tumoral après traitement par rapport à celui de l'échantillon biologique contrôle ou de l'échantillon biologique tumoral avant traitement est l'indication d'une efficacité anti-métastase du traitement thérapeutique.

25 20. Procédé *in vitro* de sélection d'un pool de polynucléotides selon l'une quelconque des revendications 1 à 4 comprenant les étapes suivantes :

a) fournir des échantillons biologiques tumoraux provenant de patients atteints d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST) ;

30 b) détecter et/ou quantifier chacun des polynucléotides, individuellement dans chacun des échantillons biologiques tumoraux ;

c) comparer le profil d'expression des pools de polynucléotides obtenus à l'étape c) par rapport à un phénotype biologique, de préférence d'instabilité chromosomique, de complexité génomique ou de grade histologique, pour chacun des échantillons biologiques tumoraux ;

35 d) sélectionner la voie statistiquement significative ($p < 10^{-5}$) pour le phénotype testé ;

e) sélectionner les polynucléotides significativement impliqués dans cette voie biologique, et dont l'expression est indicative de la probabilité de survenue de métastases.

5 21. Kit comprenant un pool de polynucléotides selon l'une quelconque des revendications 1 à 5.

10 22. Kit pour la prédiction *in vitro* de la survenue de métastases, l'évaluation du pronostic d'un patient atteint d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST) et/ou la surveillance de l'efficacité anti-métastase d'un traitement thérapeutique d'un patient atteint d'un sarcome des tissus mous (STS) ou d'une tumeur stromale gastro-intestinale (GIST), ledit kit comprenant les moyens pour détecter et/ou quantifier l'expression d'un pool de nucléotides selon l'une quelconque des revendications 1 à 5.

15 23. Puce à acides nucléiques comprenant un pool de polynucléotides selon l'une quelconque des revendications 1 à 4, immobilisé sur un support solide ou liquide.

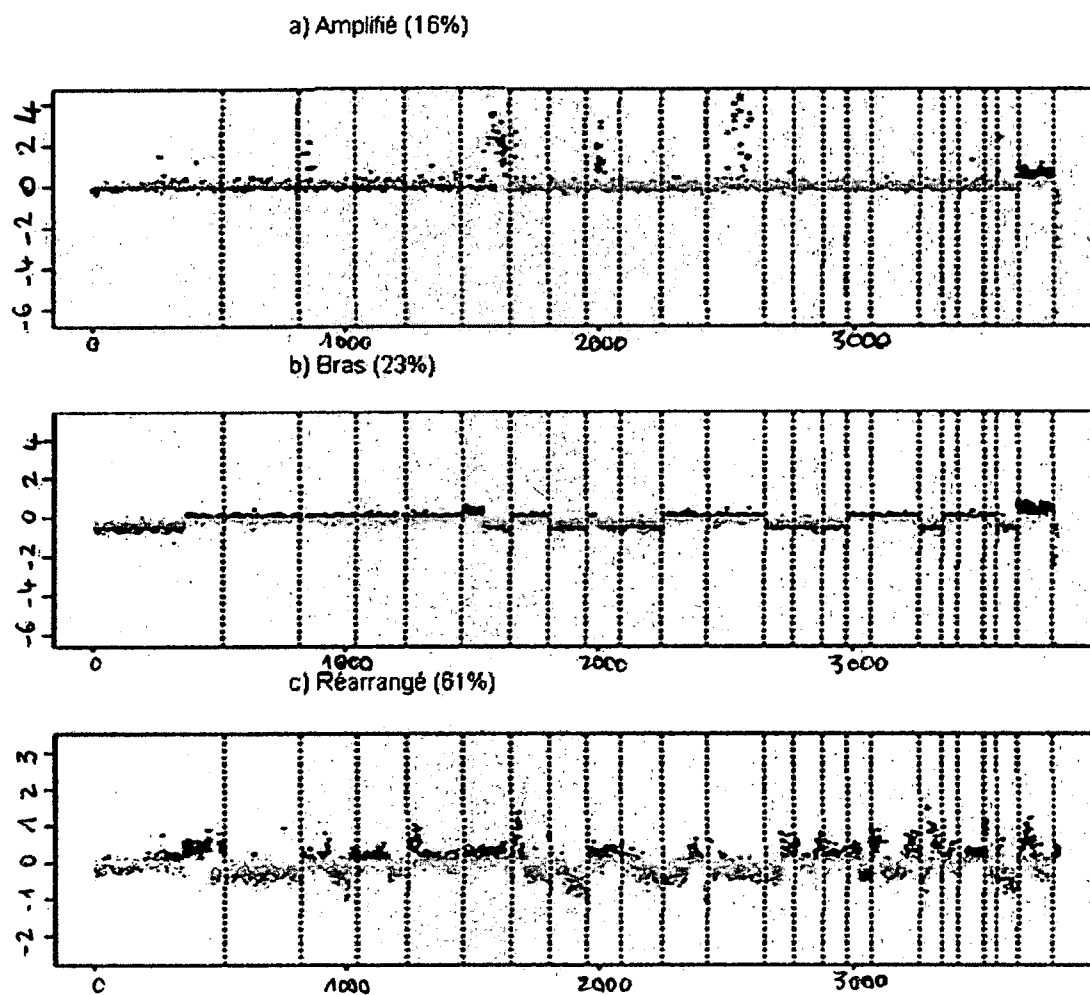
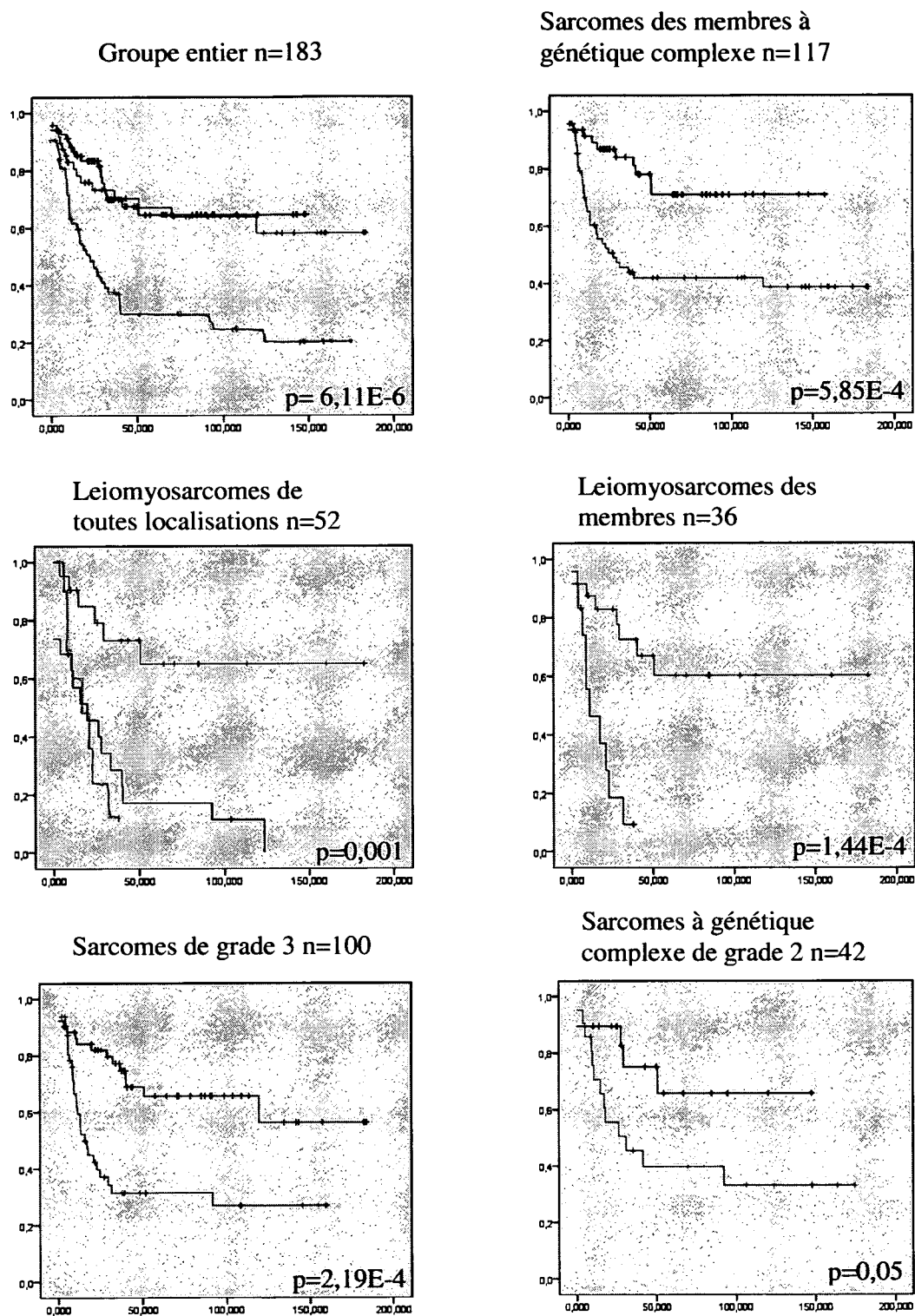


FIGURE 1



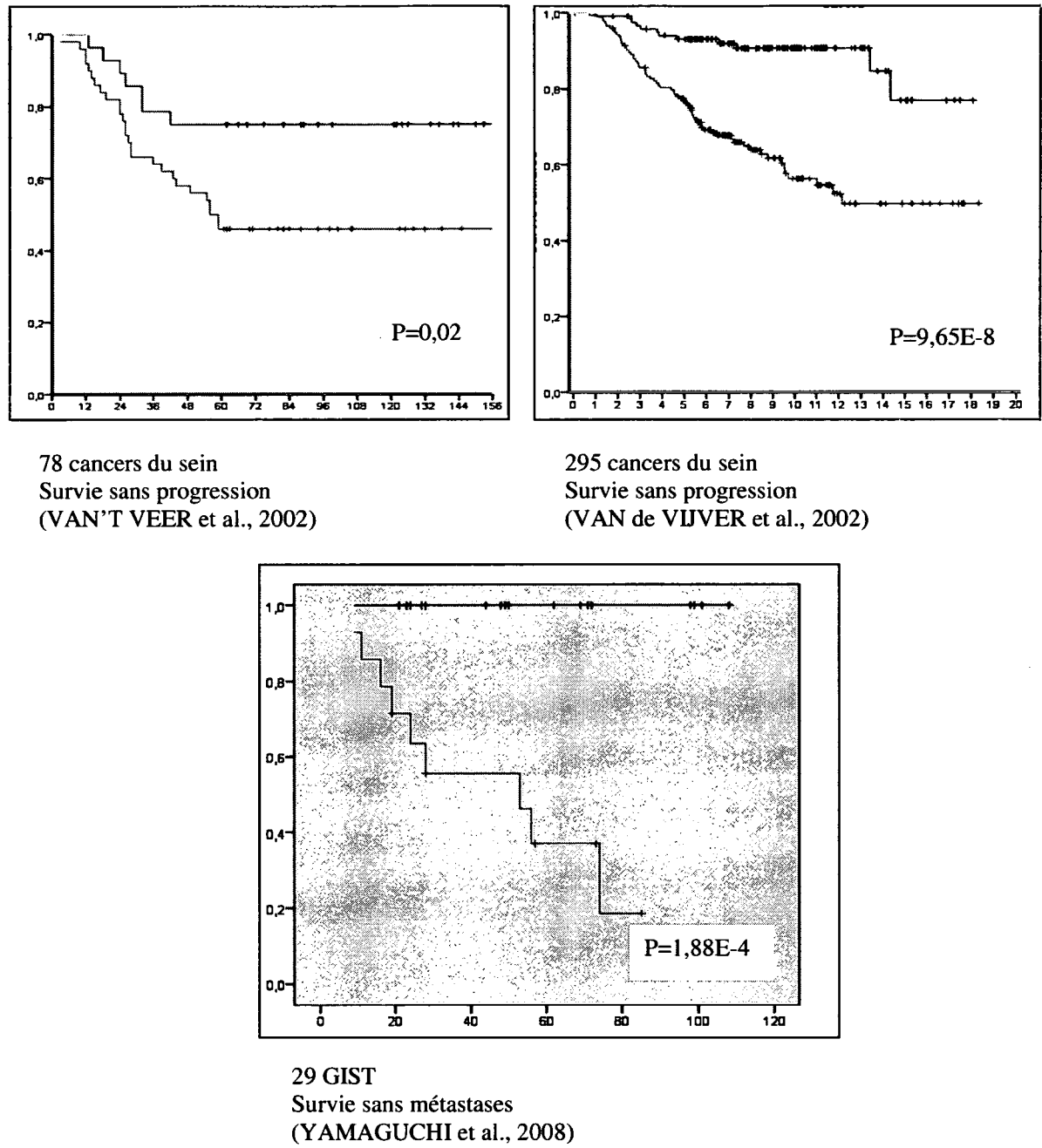


FIGURE 3

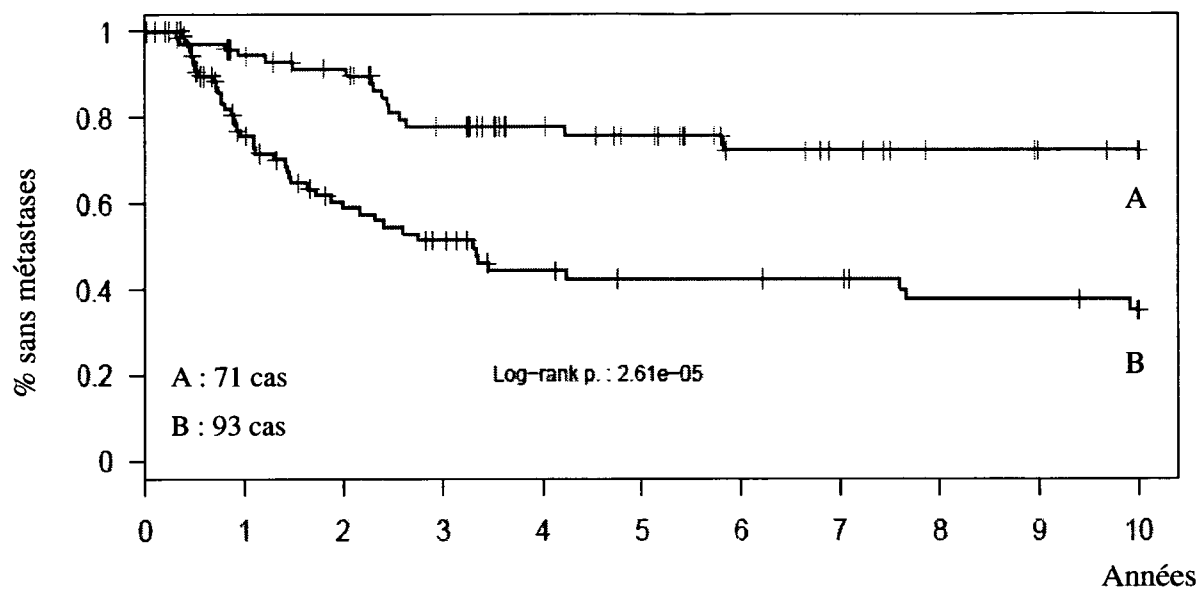


FIGURE 4

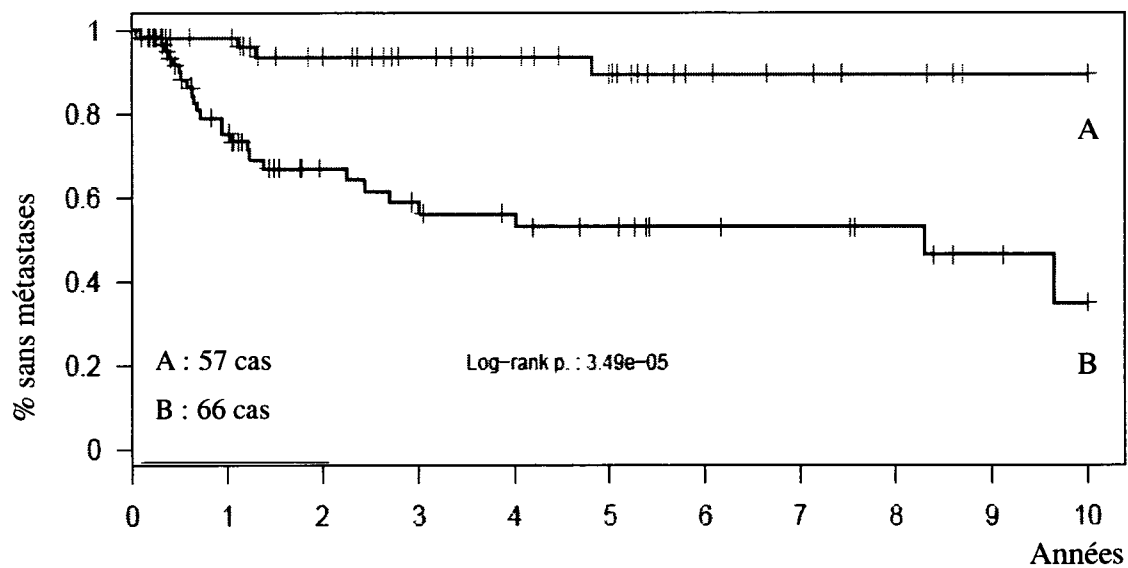


FIGURE 5

INTERNATIONAL SEARCH REPORT

International application No
PCT/FR2010/000323

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	"Affymetrix GeneChip Human Genome U133 Array Set HG-U133A" GEO,, 11 March 2002 (2002-03-11), XP002254749 the whole document	22
A	DAVICIONI ELAI ET AL: "Identification of a PAX-FKHR gene expression signature that defines molecular classes and determines the prognosis of alveolar rhabdomyosarcomas" CANCER RESEARCH, vol. 66, no. 14, July 2006 (2006-07), pages 6936-6946, XP002577781 ISSN: 0008-5472 the whole document	1-23

-/--

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

25 June 2010

Date of mailing of the international search report

01/07/2010

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Cornelis, Karen

INTERNATIONAL SEARCH REPORT

International application No

PCT/FR2010/000323

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	FRANCIS PRINCY ET AL: "Diagnostic and prognostic gene expression signatures in 177 soft tissue sarcomas: hypoxia-induced transcription profile signifies metastatic potential" BMC GENOMICS, BIOMED CENTRAL, LONDON, GB LNKD- DOI:10.1186/1471-2164-8-73, vol. 8, no. 1, 14 March 2007 (2007-03-14), page 73, XP021022381 ISSN: 1471-2164 the whole document	1-23
X	WO 2004/048938 A2 (PROTEIN DESIGN LABS INC [US]; AZIZ NATASHA [US]; GINSBURG WENDY M [US]) 10 June 2004 (2004-06-10) Revendication 16	1,3,5,9, 21-23
A	TAUBERT H ET AL: "Stem cell-associated genes are extremely poor prognostic factors for soft-tissue sarcoma patients" ONCOGENE, vol. 26, no. 50, November 2007 (2007-11), pages 7170-7174, XP002577782 ISSN: 0950-9232 the whole document	1-23
A	LINTON K M ET AL: "Acquisition of biologically relevant gene expression data by Affymetrix microarray analysis of archival formalin-fixed paraffin-embedded tumours." BRITISH JOURNAL OF CANCER 22 APR 2008 LNKD- PUBMED:18382428, vol. 98, no. 8, 22 April 2008 (2008-04-22) , pages 1403-1414, XP002577786 ISSN: 1532-1827 the whole document	1-23
X	WO 2005/117943 A2 (VERTEX PHARMA [US]; JAIN-PANDEY JUGNU [US]; HARDING MATTHEW W [US]; RA) 15 December 2005 (2005-12-15)	10-20
A	the whole document	1-9

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/FR2010/000323

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2004048938 A2	10-06-2004	AU 2003298786 A1	18-06-2004
WO 2005117943 A2	15-12-2005	AU 2005249446 A1	15-12-2005
		CA 2568405 A1	15-12-2005
		CN 101023185 A	22-08-2007
		EP 1766091 A2	28-03-2007
		JP 2008500057 T	10-01-2008
		US 2008318803 A1	25-12-2008

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale n°

PCT/FR2010/000323

A. CLASSEMENT DE L'OBJET DE LA DEMANDE

INV. C12Q1/68

ADD.

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)

C12Q

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)

EPO-Internal, BIOSIS, WPI Data

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	"Affymetrix GeneChip Human Genome U133 Array Set HG-U133A" GEO,, 11 mars 2002 (2002-03-11), XP002254749 le document en entier	22
A	DAVICIONI ELAI ET AL: "Identification of a PAX-FKHR gene expression signature that defines molecular classes and determines the prognosis of alveolar rhabdomyosarcomas" CANCER RESEARCH, vol. 66, no. 14, juillet 2006 (2006-07), pages 6936-6946, XP002577781 ISSN: 0008-5472 le document en entier	1-23
-/--		



Voir la suite du cadre C pour la fin de la liste des documents



Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent

"E" document antérieur, mais publié à la date de dépôt international ou après cette date

"L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)

"O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens

"P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

"&" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

25 juin 2010

Date d'expédition du présent rapport de recherche internationale

01/07/2010

Nom et adresse postale de l'administration chargée de la recherche internationale

Office Européen des Brevets, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Cornelis, Karen

C(suite). DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	FRANCIS PRINCY ET AL: "Diagnostic and prognostic gene expression signatures in 177 soft tissue sarcomas: hypoxia-induced transcription profile signifies metastatic potential" BMC GENOMICS, BIOMED CENTRAL, LONDON, GB LNKD- DOI:10.1186/1471-2164-8-73, vol. 8, no. 1, 14 mars 2007 (2007-03-14), page 73, XP021022381 ISSN: 1471-2164 le document en entier	1-23
X	WO 2004/048938 A2 (PROTEIN DESIGN LABS INC [US]; AZIZ NATASHA [US]; GINSBURG WENDY M [US]) 10 juin 2004 (2004-06-10) Revendication 16	1,3,5,9, 21-23
A	TAUBERT H ET AL: "Stem cell-associated genes are extremely poor prognostic factors for soft-tissue sarcoma patients" ONCOGENE, vol. 26, no. 50, novembre 2007 (2007-11), pages 7170-7174, XP002577782 ISSN: 0950-9232 le document en entier	1-23
A	LINTON K M ET AL: "Acquisition of biologically relevant gene expression data by Affymetrix microarray analysis of archival formalin-fixed paraffin-embedded tumours." BRITISH JOURNAL OF CANCER 22 APR 2008 LNKD- PUBMED:18382428, vol. 98, no. 8, 22 avril 2008 (2008-04-22) , pages 1403-1414, XP002577786 ISSN: 1532-1827 le document en entier	1-23
X	WO 2005/117943 A2 (VERTEX PHARMA [US]; JAIN-PANDEY JUGNU [US]; HARDING MATTHEW W [US]; RA) 15 décembre 2005 (2005-12-15)	10-20
A	le document en entier	1-9

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande internationale n°

PCT/FR2010/000323

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 2004048938 A2	10-06-2004	AU 2003298786 A1	18-06-2004
WO 2005117943 A2	15-12-2005	AU 2005249446 A1	15-12-2005
		CA 2568405 A1	15-12-2005
		CN 101023185 A	22-08-2007
		EP 1766091 A2	28-03-2007
		JP 2008500057 T	10-01-2008
		US 2008318803 A1	25-12-2008