

Université de Sherbrooke

**Détection, annotation fonctionnelle et régulation des isoformes de
l'épissage alternatif associées au cancer de l'ovaire**

Par

Jean-Philippe Brosseau
Département de biochimie

Thèse présentée à la Faculté de médecine et des sciences de la santé
en vue de l'obtention du grade de philosophiae doctor (Ph.D.) en biochimie

Sherbrooke, Québec, Canada
décembre, 2012

Membres du jury d'évaluation

Jean-Pierre Perreault, biochimie
Sherif Abou Elela, microbiologie
Nathalie Rivard, anatomie et biologie cellulaire
Rodney Ouellette, biochimie
Anne-Marie Mes Masson, Centre de recherche CHUM et Institut du cancer de Montréal

Jean-Philippe Brosseau, 2012

RÉSUMÉ

Détection, annotation fonctionnelle et régulation des isoformes de l'épissage alternatif impliquées dans le cancer de l'ovaire

Par

Jean-Philippe Brosseau
Département de biochimie

Thèse présentée à la Faculté de médecine et des sciences de la santé en vue de l'obtention du diplôme de philosophiae doctor (Ph.D.) en biochimie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec, Canada, J1H 5N4

L'expression des gènes modifiés contribue à l'initiation et la progression du cancer. Malheureusement, les recherches actuelles s'attardent essentiellement à l'expression globale des gènes sans tenir compte des différentes isoformes des ARNm résultant de l'épissage alternatif, codant potentiellement pour des protéines de fonctions différentes. En effet, la haute similitude et la complexité des isoformes ARNm rendent la discrimination des isoformes laborieuse, raisons pour lesquelles l'épissage alternatif est historiquement étudié gène par gène. C'est la raison pour laquelle nous avons développé une méthodologie à haut débit pour la quantification des isoformes d'ARNm. Basé sur cette technique, cette thèse présente une série d'évidences qui appuie l'hypothèse selon laquelle certaines isoformes d'ARNm sont nécessaires à la tumorigénèse. Précédemment, des études bioinformatiques à grande échelle ainsi que des validations expérimentales à petite échelle ont révélé la présence d'isoformes spécifiques à certains types de cancers. Toutefois, ces conclusions ont été tirées à partir de tumeurs dont le contenu cellulaire est hétérogène. Or, il est bien connu que les niveaux des isoformes d'ARNm diffèrent largement en fonction du type cellulaire et il est possible que les différences observées dans des tumeurs entières soient des artéfacts résultant de la complexité cellulaire des tissus comparés. En nous basant sur le cancer de l'ovaire comme modèle, nos préparations homogènes de différents types cellulaires nous ont permis de révéler la présence d'isoformes associées au cancer indépendamment de l'hétérogénéité des tumeurs. Étonnamment, les changements les plus intéressants ne se retrouvent pas dans les cellules cancéreuses à proprement dites, mais dans les cellules « normales » en bordure de la tumeur. Notre analyse des facteurs de régulation de ces isoformes nous a permis d'affirmer que ces changements ne sont pas dûs à un dérèglement anarchique des cellules cancéreuses, mais découlent plutôt d'un programme destiné à avantager les cellules cancéreuses. Pour évaluer la fonction des isoformes d'ARNm, nous avons généralisé des outils moléculaires permettant de reprogrammer spécifiquement l'expression d'isoformes de gènes cible dans des lignées cellulaires cancéreuses. Ainsi, le criblage systématique d'isoformes d'ARNm de gènes cibles associés au cancer a permis de révéler leur caractère pro-cancer dans des essais *in vitro*. L'ensemble de ces travaux démontrent la contribution des isoformes d'épissage dans le développement des tumeurs solides.

Mots-clés: Cancer de l'ovaire, PCR en temps réel à haut débit, analyse de l'ARN, siRNA, microdissection au laser, épissage alternatif, microenvironnement.

SUMMARY

Detection, functional annotation and regulation of splicing isoforms in ovarian cancer

By

Jean-Philippe Brosseau
Département de biochimie

Thesis presented to the Faculté de médecine et des sciences de la santé for the obtention of the title of philosophiae doctor (Ph.D.) in biochemistry, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec, Canada, J1H 5N4

Cancer cells modify their gene expression pattern to promote tumor growth. Unfortunately, current research focuses almost exclusively on global gene expression changes, without taking into account the fact that the majority of genes produce multiple alternative splicing variants, which potentially code for proteins of different functions. In fact, the main reason splicing isoforms are still studied on a gene by gene basis is because the high sequence similarity and complexity of mRNA isoforms make it very difficult to detect and quantify them. To overcome this difficulty, we developed a high-throughput methodology to accurately quantify splicing isoforms. This methodology has been applied in this thesis to generate evidence supporting a causal role for splicing isoforms in the tumorigenesis of solid tumors. A few years ago, genome-wide bioinformatics predictions, and some experimental validations in human tissues, suggested the existence of cancer-specific splicing isoforms. However, these conclusions may be simply explained by the cellular heterogeneity of tumor. In fact, it is well established that alternative splicing is a cell type-specific process. To address this, we microdissected cancer tissues and demonstrated that some splicing isoforms are truly cancer-associated and independent of tumor cell type content. Surprisingly, the most interesting changes were found in the “normal” stromal cells surrounding the tumor and not within the cancer cells. Our analysis of splicing-factor expression changes using the same tissue set allowed us to confirm that these splicing shifts were not a random process, but rather part of a defined splicing program promoting tumor development. To evaluate the functional contribution of splicing isoforms, we developed molecular tools that modulate splicing of endogenous targets in cancer cell lines. We used these tools to decipher the pro-cancer properties of cancer-associated splicing isoforms using *in vitro* assays. Overall, this work demonstrated the contribution of splicing isoforms in solid tumor development.

Keywords: Ovarian cancer, high-throughput real-time PCR, TOSS, siRNA, laser microdissection, alternative splicing, microenvironment.

TABLE DES MATIÈRES

RÉSUMÉ.....	ii
SUMMARY	iii
TABLE DES MATIÈRES.....	iv
LISTE DES TABLEAUX	vii
LISTE DES FIGURES.....	ix
LISTE DES ABRÉVIATIONS	xii
INTRODUCTION	16
<i>1 Le Cancer</i>	16
<i>1.1 Généralités</i>	16
<i>1.1.1 Définitions</i>	16
<i>1.1.2 Statistiques</i>	17
<i>1.1.3 Facteurs de risque</i>	17
<i>1.2 Modélisation de la cancérogenèse</i>	18
<i>1.2.1 Cancérogènes</i>	18
<i>1.2.2 Oncogènes</i>	18
<i>1.2.4 Progression séquentielle de la cancérogenèse</i>	20
<i>1.2.3 Gènes suppresseurs de tumeurs</i>	23
<i>1.2.5 Rôle du microenvironnement de la tumeur</i>	25
<i>1.2.6 La transition épithéliale-mésenchymale et mésenchymale-épithéliale</i>	27
<i>1.2.7 Origines des CAFs</i>	32
<i>1.3 Cancer de l’ovaire</i>	33
<i>2 Épissage des ARNm</i>	39
<i>2.1 Épissage constitutif</i>	39
<i>2.2 Épissage alternatif</i>	44
<i>2.2.1 Type d’évènements d’épissage alternatif</i>	44
<i>2.2.2 Mécanisme de régulation</i>	45
<i>2.2.2.1 Mécanisme de régulation classique</i>	45
<i>2.2.2.2 Expression tissu-spécifique de facteurs d’épissage</i>	48
<i>2.2.3 Rôle dans la cancérogenèse</i>	49

2.2.3.1 Dérégulation des isoformes	49
2.2.3.2 Dérégulation des facteurs d'épissage.....	50
2.2.3.3 Relation de cause à effet	50
3 Quantification de l'ARNm	54
3.1 Méthodologies de quantification des isoformes alternativement épissés	54
3.2 Utilisation d'échantillon clinique fixé	56
3.3 Microdissection au laser et amplification d'ARN	58
4 Méthodes pour reprogrammer l'épissage alternatif des cellules cancéreuses.....	59
4.1 Antisens	59
4.2 Modulateur artificiel de l'épissage alternatif.....	62
4.3 siRNAs isoforme-spécifique	66
OBJECTIFS DE CETTE THÈSE DE DOCTORAT	70
MÉTHODOLOGIES ET RÉSULTATS	71
Chapitre 1	71
Avant-propos	71
Mise en contexte	72
Résumé	72
Article 1: High-throughput quantification of splicing isoforms	74
Chapitre 2.....	97
Avant-propos	97
Mise en contexte	98
Méthodologies	99
Résultats : Développement d'outils moléculaires isoforme-spécifiques	103
Chapitre 3	142
Avant-propos	142
Mise en contexte	143
Résumé	148
Article 2: Alternative splicing of SYK regulates mitosis and cell survival.....	149
Chapitre 4	187
Avant-propos	187
Mise en contexte	187
Résumé	188

<i>Article 3: Programmed modulation of alternative splicing in the tumor microenvironment</i>	189
DISCUSSION.....	240
<i>1 Quantification de séquences hautement similaires</i>	241
<i>1.1 Quantification de l'expression globale et isoforme-spécifique sur une même plateforme</i>	241
<i>1.2 Quantification transcrit-spécifique</i>	243
<i>1.3 Application des règles de dessin d'amorce-jonction pour d'autres cas de séquences très semblables</i>	244
<i>2 Nouvelles stratégies pour la modulation de l'AS de cibles endogènes</i>	247
<i>2.1 Modulation d'ASEs simples par les TOSS</i>	247
<i>2.2 Modulation d'ASEs complexes par les TOSS</i>	248
<i>2.3 Nouvelles stratégies pour stimuler l'inclusion d'exon</i>	252
<i>3 Impact fonctionnel des isoformes de l'épissage alternatif dans la cancérogenèse</i>	253
<i>3.1 Impact fonctionnel des isoformes de l'épissage alternatif de gènes associés au cancer</i>	253
<i>3.2 Cibles potentielles pour le TOSS</i>	260
<i>3.3 Impact fonctionnel du programme RBFOX2 et QKI dans la cancérogenèse</i>	261
<i>4 Potentiel clinique des signatures moléculaires obtenues par microdissection</i>	265
CONCLUSIONS	267
REMERCIEMENTS.....	269
LISTE DES PUBLICATIONS	271
ANNEXES	286
<i>Annexe 1 Base de données de TOSS</i>	286
<i>Annexe 2 Étude de spécificité des siRNAs spécifiques à l'isoforme court</i>	286
<i>Annexe 3 Tableau S1-17 du chapitre 4</i>	293

LISTE DES TABLEAUX

Introduction

Tableau 1 Prédispositions héréditaires liées à des gènes suppresseurs de tumeurs.	25
--	----

Chapitre 1

Tableau 1 Comparison between the specificity of the boundary-spanning primers carrying different 3'end mismatches.....	83
Tableau 2 Summary of the primer design data generated by RASE.....	85
Tableau 3 Experimental validation of RASE designed primers.	87

Chapitre 2

Tableau 1 Effet de la portion « tail » du TOSS.....	121
Tableau 2 Effet de la position du TOSS par rapport au site 5' d'épissage.....	123
Tableau 3 Taux de succès des TOSS.....	141

Chapitre 3

Tableau 1 Relation fonctionnelle entre les isoformes de l'AS et influence théorique des outils moléculaires sur le phénotype.	144
Tableau 2 Comparaison des effets de modulateurs d'AS et de répression isoforme-spécifique du gène BCL2L1 sur l'induction d'apoptose dans des cellules cancéreuses humaines.	146

Chapitre 4

Tableau S1 Description of the cancer epithelial signature (CES) alternative splicing events (ASEs).	235
Tableau S2 Description of the cancer stromal signature (CSS) alternative splicing events (ASEs).	235
Tableau S3 Description of the common RBFOX2 and QKI splicing targets associated with the tumor microenvironment.....	235
Tableau S4 Expression patterns of cancer-associated splicing factors.	236
Tableau S5 Validation of the splicing factors knockdown.....	236
Tableau S6 The expression of splicing factors associated with mesenchymal-to-epithelial transition (MET) are not systematically modified in cancer tissues.	236
Tableau S7 List of EMT associated alternative splicing events that are preferentially expressed in tumor tissues.	236
Tableau S8 Identification of RBFOX2 and QKI common splicing targets.....	237
Tableau S9 RBFOX2 and QKI preferentially regulate the expression of mesenchymal-enriched exons.....	237
Tableau S10 Splicing pattern of RBFOX2 and QKI splicing targets in cancer tissues.	237

Tableau S11 QKI-6 and QKI-7 splicing pattern in undissected tissues.	237
Tableau S12 QKI relative global expression in undissected tissues.	237
Tableau S13 Clinical data of tissues used for laser capture microdissection (LCM) samples.	238
Tableau S14 Global expression levels of established epithelial and stromal markers in normal and cancer tissues.	238
Tableau S15 Clinical data associated with the undissected serous ovarian tumors used in this study.	238
Tableau S16 Clinical data associated with the undissected ductal breast tumors used in this study.	238
Tableau S17 List of siRNAs used for the knockdown of splicing factors.	238

Discussion

Tableau 1 Stratégies de dessin d'amorce allèle-spécifique.	245
Tableau 2 Exemples de gènes associés au cancer ayant été ciblés avec des modulateurs d'AS (SSO, TOSS ou TOES) dans des cellules cancéreuses humaines.	254
Tableau 3 Exemples de gènes associés au cancer et ciblés avec des siRNAs isoforme- spécifiques dans des cellules cancéreuses humaines.	256

LISTE DES FIGURES

Introduction

Figure 1 Mécanismes d'activation des oncogènes.....	20
Figure 2 Progression séquentielle de la cancérogenèse.....	23
Figure 3 Interactions hétérotypiques entre les cellules épithéliales cancéreuses et les cellules stromales du microenvironnement.	27
Figure 4 Rôle de l'EMT/MET dans le développement normal.	28
Figure 5 Schématisation des étapes de l'EMT.....	30
Figure 6 Facteurs de transcription impliqués dans l'EMT/MET.	31
Figure 7 Rôle de l'EMT/MET dans la cancérogenèse.	32
Figure 8 Progression séquentielle du cancer de l'ovaire.....	36
Figure 9 Origine du cancer de l'ovaire.....	38
Figure 10 Unité génique.	39
Figure 11 Réaction d'épissage.	41
Figure 12 Séquences consensus impliquées lors de l'épissage.....	41
Figure 13 Cycle d'assemblage du splicéosome.	43
Figure 14 Type d'évènements d'épissage alternatif.....	45
Figure 15 Mécanismes de régulation de l'épissage alternatif par des facteurs d'épissage. .	47
Figure 16 L'épissage alternatif est omniprésent dans les gènes liés au cancer.	53
Figure 17 Stratégies de détection pour la quantification des isoformes alternativement épissés.....	55
Figure 18 Contraintes de dessin d'amorces en endpoint PCR dues à l'utilisation d'échantillon clinique fixé.	58
Figure 19 Mécanismes d'action des antisens.....	60
Figure 20 Structure des principaux analogues de nucléosides.	62
Figure 21 Mécanismes d'action des SSOs.....	65
Figure 22 Anatomie d'un TOSS.	66
Figure 23 Mécanisme d'action des siRNAs.	67
Figure 24 Utilisation de siRNAs spécifiques à l'isoforme court (siRNA court).	69

Chapitre 1

Figure 1 Problems associated with the detection of alternative splicing isoforms.	77
Figure 2 Systematic tiling of BSPs across an exon-exon junction.....	80
Figure 3 Strategy for automated selection of high fidelity splice sensitive PCR primers....	85
Figure S1 BSP specificity in typical qRT-PCR experiment.....	81
Figure S2 Evaluation of BSP's maximal 3' end T_m	82

Chapitre 2

Figure 1 Effet de A1_Bclx4 sur la modulation de l'épissage de BCL2L1 <i>in cellulo</i>	104
Figure 2 Blocage de la transcription inverse par les oligonucléotides 2'OMe.	105
Figure 3 Le traitement au Trizol® enlève le TOSS d'un extrait d'ARN total.	107
Figure 4 Le Trizol® permet d'enlever la composante artéfactuelle.	109
Figure 5 Estimation de la réelle proportion de la composante artéfactuelle <i>in cellulo</i>	111
Figure 6 Confirmation du réel effet modulateur du TOSS <i>in cellulo</i> par PCR en temps réel.	113
Figure 7 Effets des TOSS sur les isoformes de CASP8 et CASP10 au niveau protéique..	113
Figure 8 Modulation de type A et de type B.....	114
Figure 9 Marche à suivre permettant la validation des TOSS à haut débit.	117
Figure 10 Effet de la modification de la chimie de la portion « tail ».	120
Figure 11 Modulation de l'alt3' de DRF1.	125
Figure 12 Modulation de l'évènement d'épissage alternatif complexe (exon cassette et alt5') de NUP98.....	128
Figure 13 Modulation de l'évènement d'épissage alternatif complexe (cassette et alt3') de TOPBP1.....	129
Figure 14 Modulation de l'évènement d'épissage alternatif complexe (exon cassette, alt3' et rétention d'intron) de FN1.....	131
Figure 15 Impact de la chimie du TOES sur la modulation de l'AS.	134
Figure 16 Autres stratégies favorisant l'inclusion de l'exon alternatif.	136

Chapitre 3

Figure 1 Strategy for functional annotation of cancer-associated alternative splicing events (ASEs).	155
Figure 2 Targeting of SYK, MCL-1 or FN1-EDB alternatively spliced isoforms induces apoptosis.	158
Figure 3 SYK alternative splicing regulates cell cycle progression and mitosis.	164
Figure 4 SYK(L) promotes malignancy and anchorage-independent growth.	168
Figure 5 A model depicting the different functions of the SYK(L) pro-survival isoform.	173
Figure S1 FASE (Platform for High-Content Functional Analysis of Alternative Splicing Events) screen pipeline.	154
Figure S2 ISI design rules.....	156
Figure S3 FASE screen quality control.	157
Figure S4 Western validation for MCL-1 and SYK.	161
Figure S5 SYK phenotype validation in other cell lines.	165
Figure S6 MAPK isoforms essential for cell viability.	170
Figure S7 hnRNP-K controls SYK, MCL-1 and FN1 alternative splicing ratios (a) and is essential for SKOV3ip1 cell survival (b).	176
Figure S8 Alternative splicing regulation of EGF/MAPK signaling.	177
Figure S9 Data analysis flowchart displaying all the data processing, normalization and bioinformatic and statistical analysis steps for the FASE high-content screen.	183

Chapitre 4

Figure 1 Laser capture microdissection (LCM) of normal and cancer ovarian tissues.....	194
Figure 2 Splicing of a subset of mesenchymal-enriched exons is repressed in the ovarian tumor microenvironment.	199
Figure 3 The expression of cancer-associated splicing isoforms is regulated by a small group of splicing factors.....	204
Figure 4 The expression of splicing factors associated with mesenchymal- to- epithelial transition (MET) are not systematically modified in cancer tissues.	208
Figure 5 RBFOX2 and QKI regulate the expression of common epithelial splicing isoforms in the tumor microenvironment.....	211
Figure 6 The common RBFOX2 and QKI splicing targets are deregulated in both ovarian and breast cancer.	215
Figure S1 Comparison between the splicing patterns obtained in undissected and dissected ovarian cancer and Fallopian tube mirror image samples.	196
Figure S2 Mesenchymal-enriched exons are preferentially regulated in the tumor microenvironment.	201
Figure S3 Monitoring the splicing pattern of ASEs associated with whole tumors in dissected tissues.	216
Figure S4 Evaluation of the reliability of the formalin-fixed, paraffin-embedded (FFPE) based RNA expression assays by quantitative RT-PCR.	233
Figure S5 Comparison between quantitative and end-point PCR using amplified and non-amplified RNA extracted from fresh and fixed tissues.	234
Discussion	
Figure 1 Fréquence et mésappariement de l'amorce allèle-spécifique sur l'allèle de type sauvage selon le type de SNP.	246
Figure 2 Mécanismes d'action des ESS selon le modèle de Wang et collaborateurs (2006).	248
Figure 3 Nouveaux dessins de TOSS.....	251
Figure 4 Impact du programme RBFOX2 et QKI sur les fibroblastes normaux.	264

LISTE DES ABRÉVIATIONS

2'Fluoro arabinose.....	FANA
2'Ométhyl.....	2'OMe
3' splice site.....	3' ss
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid.....	HEPES
5' splice site.....	5' ss
Acide désoxyribonucléique.....	ADN
Activateur centromérique.....	CH
Adenomatosis polyposis coli.....	APC
ADN complémentaire.....	cDNA
Alpha smooth muscle actin.....	α SMA
Alternatif 3'.....	alt3'
Alternatif 5'.....	alt5'
Alternative splicing event.....	ASE
Alternative splicing.....	AS
Alternatively included region.....	AIR
American type culture collection.....	ATCC
Antisens.....	ASO
Application programming interface.....	API
Atypical proliferative serous tumor.....	APST
Basal carcinoma cell line.....	BCC
B-cell leukemia/lymphoma.....	BCL
Bone morphogenic 7.....	BMP7
Boundary-spanning primer.....	BSP
Branch point site.....	BPS
Breakpoint cluster region.....	BCR
Breast cancer, early onset.....	BRCA
c-abl oncogene 1.....	ABL1
Cadhérine-E.....	CDH1
Cancer stem cells.....	CSC
Cancer-associated fibroblast.....	CAF
Cleavage and polyadenylation specificity factor.....	CPSF
Comprehensive perl archive network.....	CPAN
CUG-BP and ETR3-like factors.....	Celf
Cycle threshold.....	Ct
Déoxynucléoside triphosphate.....	dNTP
Désoxyribonucléase.....	DNase
Downstream constitutive sequence.....	DCS
Dystrophie musculaire de Duchenne.....	DMD
Epidermal growth factor.....	EGF
Epithelial to mesenchymal transition.....	EMT
Epithelial-specific regulatory protein.....	Esrp
Exonic splicing enhancers.....	ESE
Exonic splicing silencers.....	ESE

Exon-specific splicing enhancement by small chimeric effectors.....	ESSENCE
Expressed sequence tags.....	EST
Fallopian tube epithelial cells.....	FTE
Federal Drug Administration.....	FDA
Fibroblast activation protein.....	FAP
Fibroblast growth factor receptor 2.....	FGFR2
Fibronectin 1.....	FN1
Fluorescence-activating cell sorting.....	FACS
Formalin-fixed, paraffin-embedded.....	FFPE
General transcription factor IIB.....	TFIIB
General transcription factor IID.....	TFIID
Grade 1 tumors.....	G1
Grade 3 tumors.....	G3
Harvey rat sarcoma viral oncogene homolog.....	HRAS
Hématoxylène et éosine.....	H&E
Hepatocyte growth factor.....	HGF
Heterogenous nuclear ribonucleoprotein.....	hnRNP
Heure.....	h
Housekeeping gene.....	HKG
Hydroxyl en 2'.....	2'OH
Induced pluripotent stem cell.....	iPSC
Inhibitory concentration.....	IC
Intronic splicing enhancer.....	ISE
Intronic splicing silencer.....	ISS
Isoform-specific inhibitor.....	ISI
Kirsten rat sarcoma viral oncogene homolog.....	KRAS
Laser capture microdissection.....	LCM
Leucémie myéloïde chronique.....	CML
Lipofectamine 2000.....	LF
Locked nucleic acid.....	LNA
Long-specific primer pair.....	LSPP
Low malignant potential.....	LMP
Matrix metalloproteases.....	MMP
Mesenchymal to epithelial transition.....	MET
Methylphosphonate.....	MP
Micro RNA.....	miRNA
Micropapillary serous carcinoma.....	MPSC
Minute.....	min
Mispriming in 3'.....	MP 3'
Mispriming in 5'.....	MP 5'
Mitogen-activated protein kinase.....	MAPK
murine sarcoma viral oncogene homolog B1.....	BRAF
Muscleblind-like.....	Mbnl
Mutant viable de Quaking.....	qk ^v
Myelocytomatosis viral oncogene homolog.....	MYC
Myeloid cell leukemia sequence 1.....	MCL1
National Center for biotechnology information.....	NCBI

Neuroblastoma viral oncogene homolog.....	..NRAS
Neuro-oncological ventral antigen.....	..Nova
No reverse transcription.....	..no RT
Non transfecté.....	..NT
Non-sense mediated decay.....	..NMD
Nuclear localization signal.....	..NLS
Nucléotides.....	..nt
Ovarian serous high grade.....	..OSHG
Ovarian surface epithelium.....	..OSE
Paire de base.....	..bp
Peptide nucleic acid.....	..PNA
Percent of splicing index.....	.. ψ , psi
phosphatase and tensin homolog.....	..PTEN
Phosphate buffer solution.....	..D-PBS
Phosphorothioate.....	..PS
Platelet-derived growth factor.....	..PDGF
Platform for high-content functional analysis of alternative splicing events.....	..FASE
Polymerase chain reaction.....	..PCR
Polypyrimidine tract binding protein.....	..PTBP
Polypyrimidine tract.....	..PPT
Real-time PCR.....	..Q-PCR
Real-time PCR Annotation of Splicing Events.....	..RASE
Retinoblastoma 1.....	..RB1
Ribonucléase.....	..RNase
Ribonucleic acid.....	..RNA
RNA binding protein, fox homolog.....	..RBFOX
RNA integrity number.....	..RIN
RNA interference.....	..RNAi
RNA recognition motif.....	..RRM
RNA-induced silencing complex.....	..RISC
Rotation par minute.....	..rpm
Seconde.....	..sec
Secreted protein, acidic, cysteine-rich.....	..SPARC
Serial analysis of gene expression.....	..SAGE
Serine/arginine-rich splicing factor kinase 1.....	..SRPK1
Serous tubal intraepithelial carcinoma.....	..STIC
Short-specific primer pair.....	..SSPP
Single nucleotide polymorphism.....	..SNP
siRNA contrôle.....	..ctrl siRNA
Small interfering RNA.....	..siRNA
Small nuclear ribonucleoprotein.....	..snRNP
Small nuclear RNA.....	..snRNA
Sodium dodécyl sulfate.....	..SDS
Sous-unité de 35kDa de U2AF.....	..U2AF35
Sous-unité de 65kDa de U2AF.....	..U2AF65
Spinal muscular atrophy.....	..SMA
Spleen tyrosine kinase.....	..SYK

Splice-switching oligonucleotide.....	..SSO
Splicing factor 1.....	..SF1
Splicing factor 2.....	..SF2
Splicing regulatory element.....	..SRE
Src homology 2.....	..SH2
Substitution nucléophile de type 2.....	..S _N 2
Targeted oligonucleotide enhancer of splicing.....	..TOES
Targeted oligonucleotide silencer of splicing.....	..TOSS
TATA binding protein.....	..TBP
Telomerase reverse transcriptase.....	..TERT
Température de fusion.....	..tm
Température pièce.....	..t.p.
TOSS contrôle.....	..ctrl TOSS
Tris-EDTA.....	..TE
Tumor growth factor beta.....	..TGF β
Tumor protein p53.....	..TP53
Tumor suppressor gene.....	..TSG
U2 auxiliary factor.....	..U2AF
Universal human reference RNA.....	..UHRR
Untranslated region.....	..UTR
Upstream constitutive sequence.....	..UCS
Vascular endothelial growth factor.....	..VEGF
Wingless-type MMTV integration site family, member 4.....	..WNT4

INTRODUCTION

1 Le Cancer

1.1 Généralités

1.1.1 Définitions

Le cancer est défini comme un groupe de maladies diverses, chacune d'elles impliquant une croissance cellulaire non régulée. Les mutations sont des lésions génétiques sous-jacentes qui peuvent contribuer au cancer. La plupart de ces mutations surviennent au cours de la vie d'un individu, c'est-à-dire qu'elles n'étaient pas présentes à la naissance, et qu'elles ne seront pas transmises à la descendance (à moins que celles-ci soient présentes dans une lignée germinale). Une faible proportion de la population possède des défauts génétiques héréditaires qui la prédispose au cancer (Futreal *et al.*, 2004).

Peu importe l'origine de ces mutations, les différents cancers sont d'abord classifiés selon leur type cellulaire d'origine. La plupart des cancers (90%) sont issus de dérèglements des cellules épithéliales (des adénocarcinomes). Selon les statistiques canadiennes sur le cancer (2012), les cancers des poumons et les cancers colorectaux sont les plus fréquemment diagnostiqués, suivis du cancer du sein. Les leucémies (cellules du système sanguin), les lymphomes (cellules du système immunitaire) ou les sarcomes (cellules mésenchymales) regroupent une faible proportion de cas (10%) (Santarius *et al.*, 2010). Ensuite, une classification histologique des cellules est faite selon l'indice de prolifération, la grosseur des noyaux et l'aspect plus ou moins différencié: le grade. Les cellules cancéreuses de haut grade se multiplient rapidement (prolifération active), tendent à ne plus ressembler à leur type cellulaire d'origine (dé-différentiation) et présentent des défauts chromosomiques et des noyaux difformes (aneuploïdes). De plus, une tumeur maligne est composée de cellules cancéreuses qui s'infiltrent dans les tissus avoisinants, alors qu'il n'y a pas de cellules cancéreuses dans une tumeur bénigne et que celles-ci n'infiltrent pas les tissus avoisinants. Ces analyses sont effectuées par un pathologiste et nécessitent un échantillon de la tumeur.

Bien souvent, celui-ci est prélevé sous forme de biopsie, mais il peut arriver qu'il soit récupéré uniquement lors de la chirurgie (cancer de l'ovaire). Une dernière classification se fait en fonction de la migration de la tumeur vers d'autres niches au sein du corps: le stade. Une tumeur métastatique de stade élevé signifie que des cellules cancéreuses primaires se sont implantées et ont formé un nouveau foyer dans au moins un autre organe. Cette conclusion est forgée suite à l'examen de radiographie, scanner et/ou lors de la chirurgie afin d'enlever la masse de cellules cancéreuses. Toutes ces analyses permettent d'orienter grossièrement le traitement et le pronostic du patient, c'est-à-dire le progrès de sa maladie ou de sa survie.

1.1.2 Statistiques

Malgré toutes les récentes percées scientifiques, le cancer demeure une cause de décès majeure. Selon Statistiques Canada, 1 homme sur 3,5 et 1 femme sur 4,2 mourront du cancer. La première cause de décès par cancer tant chez la femme que chez l'homme est le cancer du poumon (26,3% et 28,3%, respectivement). Vient ensuite le cancer du côlon et rectum (12,4%) et de la prostate (11,1%) chez l'homme. Chez la femme, le cancer du sein vient au 2^e rang (15,1%) et celui du côlon et du rectum au 3^e rang (11,8%). L'incidence de décès par cancer n'est pas nécessairement le reflet du nombre de cas diagnostiqués. En fait, les cancers les plus fréquents sont ceux de la prostate (28,6%) chez l'homme et du sein (27,8%) chez la femme. À l'opposé, le cancer de l'ovaire est un de ceux dont l'incidence est faible, mais dont le taux de mortalité est élevé. Pour l'année 2010 au Canada seulement, cela a été estimé à 1750 décès pour 2600 diagnostics.

1.1.3 Facteurs de risque

On entend souvent parler que des aliments ou des habitudes de vie sont susceptibles de donner le cancer ou encore, d'en retarder son apparition. Qu'en est-il vraiment? Les deux facteurs de risque qui sont hors de notre contrôle sont l'âge et les prédispositions héréditaires. En dehors de cela, certains facteurs environnementaux et l'adoption de saines habitudes de vie influencent le risque de cancer. Le tabagisme est responsable de 85% des cancers du poumon, mais également de 30% de tous les décès par cancer.

1.2 Modélisation de la cancérogenèse

1.2.1 Cancérogènes

L'histoire nous a fait connaître son lot de composés réputés comme carcinogéniques. En 1775, un chirurgien de Londres, Percival Pott, a rapporté un nombre élevé de cancers du scrotum chez des hommes de 20 à 50 ans. Après enquête, le dénominateur commun entre ces hommes était leur travail de ramoneur pendant leur jeunesse. L'hypothèse du Dr Pott condamnait le contact de la suie et du goudron avec les plis de peau recouvrant les testicules. Cette hypothèse sera en fait confirmée beaucoup plus tard lorsque K. Yamagisawa en badigeonnera des oreilles de lapins. Le chlorure de vinyle (angiosarcome du foie), le diéthylstilbestrol (carcinome vaginal), et l'amiante (mésothéliome) ne sont que d'autres exemples de composés carcinogènes découverts fortuitement, sans compter pour certains scientifiques ou mineurs, l'exposition à la radioactivité.

En parallèle, des études intensives ont décrit le cancer comme une maladie infectieuse. Vers 1950, une panoplie de virus a été répertoriée comme causant des cancers, surtout des leucémies et des sarcomes, chez le poulet, la souris, le rat et le chat. Tout ceci représente des avancées majeures dans la compréhension du cancer, mais la plupart de ces exemples ne représentent pas les adénocarcinomes habituellement retrouvés chez l'humain. En date de 2008, seulement six virus ont été formellement associés à la formation de cancer humain (Martin and Gutkind, 2008). Quatre d'entre eux causent des cancers de cellules épithéliales (Martin and Gutkind, 2008). Au total, il est estimé que 10 à 15% des cancers humains ont une origine virale (Parkin, 2006).

1.2.2 Oncogènes

Dans les années 1970, l'échec évident d'associer tous les cancers humains à une infection virale poussent les chercheurs à regarder ailleurs, soit le génome des cellules humaines. La nouvelle hypothèse suggérée propose que le contact prolongé avec des composés chimiques ou des radiations entraîne la mutation de gènes impliqués dans le contrôle de la prolifération cellulaire. La course aux oncogènes, gènes qui jouent un rôle dans les cellules

normales, mais participent à la prolifération anormale, lorsque mutés, est alors lancée. Les résultats sont frappants: plusieurs gènes autrefois découverts dans les génomes viraux se retrouvent en fait dans les cellules humaines normales, mais sous une forme mutée. Le gène est alors nommé proto-oncogène avant la mutation, et oncogène une fois la mutation subite (Tabin *et al.*, 1982).

Ces mutations à l'origine du cancer peuvent être aussi simples et localisées qu'un changement impliquant une seule paire de bases (bp) pour une autre. Ceci peut générer un changement d'acide aminé dans la protéine finale comme c'est le cas pour Kirsten rat sarcoma viral oncogene homolog (KRAS), ce qu'on appelle une mutation faux-sens (Tabin *et al.*, 1982) (Fig. 1A). En revanche, une mutation non-sens fait apparaître un codon-stop prématurément, entraînant la production d'une protéine tronquée en C-terminal. À l'extrémité du spectre, un bloc allant de quelques-unes à plusieurs millions de bp (amplification, délétion, translocation) peut être muté. Bien que plusieurs gènes puissent se retrouver dans un bloc de mutation, tous ne sont pas nécessairement impliqués dans la cancérogenèse. La translocation chromosomale est la mutation la plus fréquemment rencontrée parmi les gènes reliés au cancer (Futreal *et al.*, 2004). Celle-ci a pour effet de créer des gènes chimériques ou d'exposer un gène à de nouvelles séquences régulatrices comme c'est le cas pour myelocytomatosis viral oncogene homolog (MYC) dans plusieurs leucémies myéloïdes, altérant ainsi son expression (Fig. 1B). Dans le cas de la leucémie myéloïde chronique (CML), la translocation d'une portion du chromosome 22 au chromosome 9, produit un gène chimérique entre le «break point cluster region (BCR) et c-abl oncogene 1, non-receptor tyrosine kinase (ABL1)», activant ce dernier oncogène (Rowley, 1973, Stam *et al.*, 1985) (Fig. 1C). De plus, cette tyrosine kinase est cruciale pour le développement de la CML puisque l'imatinib, un inhibiteur d'Abl, diminue la progression de la maladie (Druker *et al.*, 2001). Par contre, il est rare qu'un seul évènement soit suffisant pour causer le cancer.

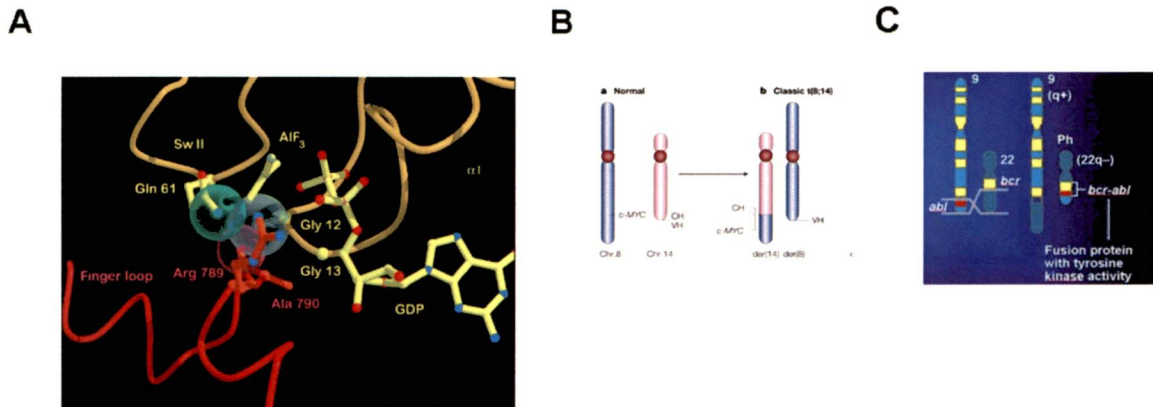


Figure 1 Mécanismes d'activation des oncogènes.

A) Changement structurel. Les mutations du gène KRAS observées dans les cancers humains induisent un changement conformationnel de la protéine, la rendant constitutivement active dû à l'inactivation de l'activité GTPase. Image tirée de Scheffzek (1997). B) Exposition à des séquences régulatrices. Dans les cellules normales, le gène MYC et le locus de l'immunoglobulin heavy chain incluant l'activateur centromérique (CH) sont situés sur les chromosomes 8 et 14, respectivement. La translocation de MYC sur le chromosome 14 juxtapose celui-ci au CH et conséquemment permet sa surexpression. Image tirée de Kuehl and Bergsagel (2002). C) Gène chimérique. Une schématisation montre une portion du chromosome 22 transloquée au chromosome 9, produisant un gène chimérique (BCR-ABL1) qui permet d'activer l'oncogène Abl. Image tirée de www.cancer.gov.

1.2.4 Progression séquentielle de la cancérogenèse

Dès 1954, Armitage et Doll avaient élaboré un modèle mathématique corrélant l'incidence de cancer en fonction de l'âge. De façon indirecte, et sans même discuter du concept de mutation, leurs conclusions supportent la progression du cancer en plusieurs étapes soit environ six (Armitage and Doll, 1954). Comme il est estimé que les mutations surviennent normalement à une fréquence d'une fois toutes les $2,5 \times 10^{-8}$ bp ou 175 mutations par génome diploïde (Nachman and Crowell, 2000), il est par contre difficile d'imaginer que plusieurs mutations puissent s'accumuler à l'intérieur d'une même cellule. De plus, la cellule atteinte doit posséder des capacités de prolifération accrue.

La théorie de l'évolution clonale des cellules cancéreuses soutient qu'il est possible d'arriver à un nombre suffisamment grand de cellules pour qu'il soit probable que la deuxième mutation ait lieu dans une cellule déjà porteuse de la 1re mutation, et ainsi de suite. Sans préciser de mutation en particulier, l'auteur Peter Nowell suggère que celle-ci

doit promouvoir une prolifération intense au point de dominer la population : une expansion clonale (Nowell, 1976). Ceci explique notamment « l'homogénéité » relative des aberrations chromosomiques au sein de la tumeur.

Une autre théorie en vigueur, le « mutator phenotype », fait appel à l'instabilité génomique pour expliquer le fort taux de mutation des cellules cancéreuses. Par le biais de divers mécanismes, Laurence Loeb propose qu'un état d'instabilité génomique doive se produire dans les premiers événements conduisant à la formation multiétapes d'un cancer (Loeb, 1991). Effectivement, de nombreuses mutations ont été répertoriées dans des gènes de réparation de l'ADN pour plusieurs cancers héréditaires. Par contre, peu de mutations ont été retrouvées dans ces mêmes gènes dans les cas de cancer sporadique. Pour expliquer cette situation, Halazonetis et collaborateurs (2008) émettent la théorie selon laquelle certains oncogènes seraient à l'origine de défauts de réplication et des bris subséquents de l'ADN. Les nombreux bris dans l'ADN produits par l'arrêt de la fourche de réplication activeraient p53, signalant ainsi l'apoptose ou la sénescence des cellules, ce qui mettrait un frein temporaire à la progression du cancer. Une mutation du gène TP53 parviendrait définitivement à surmonter cette barrière et permettrait d'expliquer deux caractéristiques propres à pratiquement tous les cancers épithéliaux : l'instabilité génomique et la fréquence de mutation élevée du gène TP53 (Halazonetis *et al.*, 2008). Alternativement, la théorie de la progression prétéumorale mise de l'avant par Calabrese et collaborateurs (2004) énonce que les cellules normales pourraient accumuler des mutations bien avant l'apparition d'un changement phénotypique, ce qui aurait pour effet d'être indépendant d'un taux élevé de mutations.

Outre l'instabilité génomique, Hanahan et Weinberg font le constat de six caractéristiques partagées par la vaste majorité des cancers, les fameux « Hallmarks of Cancer » (Hanahan and Weinberg, 2000). Afin de progresser, toute tumeur doit se rendre insensible aux signaux anti-croissance et parvenir à une prolifération soutenue. Elle doit également résister à l'apoptose et la sénescence, en plus de promouvoir l'angiogenèse et l'invasion. À noter que les auteurs ne mettent pas l'accent sur un ordre à suivre en particulier (Hanahan and Weinberg, 2000). Ces différentes caractéristiques ont récemment été mises au goût du jour

en y incluant notamment la résistance au système immunitaire et la reprogrammation du métabolisme énergétique (Hanahan and Weinberg, 2011).

De par l'abondance et l'accessibilité anatomique qu'offrent les cas de cancer du côlon, ceux-ci en font d'excellents modèles pour les cancers épithéliaux. Au début des années 1990, Fearon et Vogelstein proposent ce qui deviendra le modèle de la séquence adénome à carcinome (Fearon and Vogelstein, 1990). Ils proposent que chacune des mutations favorisant la progression du cancer du côlon (adenomatous polyposis coli (APC), KRAS, TP53) corresponde en fait à une des quatre à six étapes limitantes prédites par la théorie d'Armitage et Doll. Ils insistent toutefois sur le total plutôt que l'ordre des mutations (Fearon and Vogelstein, 1990).

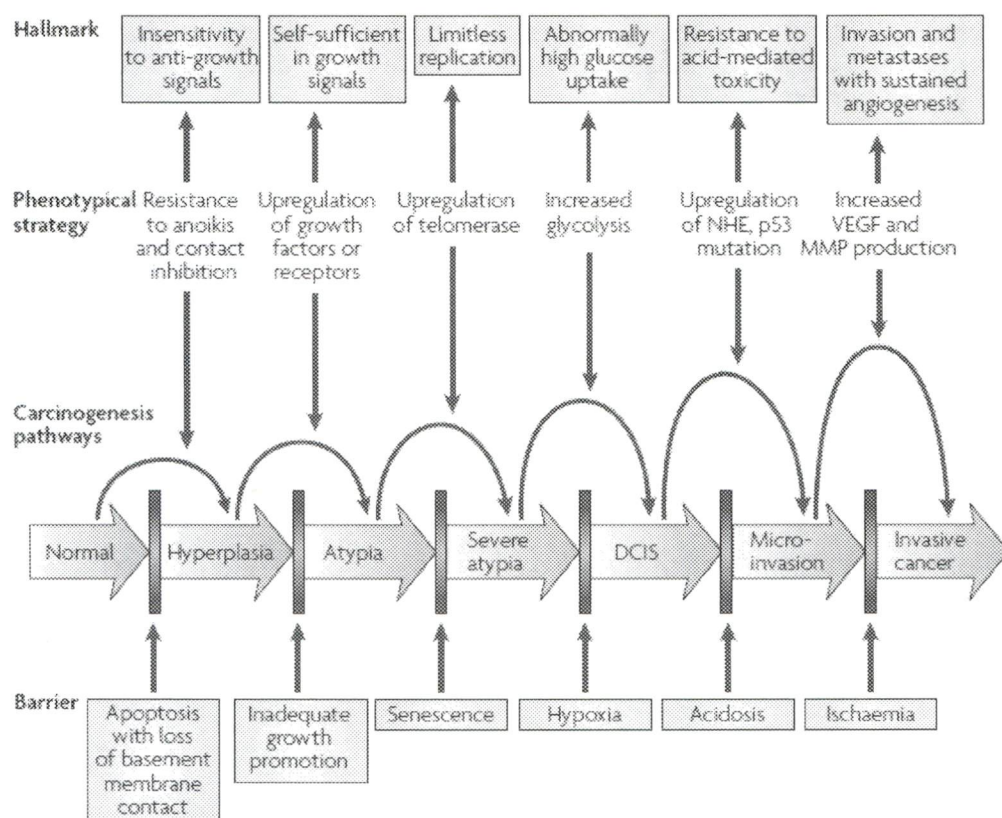


Figure 2 Progression séquentielle de la cancérogenèse.

Les intermédiaires de la séquence adénome à carcinome ainsi que les différentes barrières empêchant leur progression sont représentées. La hauteur des flèches les joignant est proportionnelle à la prolifération croissante des cellules de plus en plus malignes. Les stratégies employées pour contrer les barrières sont le reflet des « Hallmarks of Cancer » proposés par Hanahan et Weinberg (2000). Image tirée de Gatenby and Gillies (2008).

À ce jour, la séquence adénome à carcinome a été décrite pour la plupart des cancers épithéliaux et fait référence à l'acquisition de phénotypes et de changements de morphologie qui transforment peu à peu les cellules épithéliales en tumeur maligne (Fig. 2). L'instabilité génomique est omniprésente et certains traits caractéristiques « Hallmarks of Cancer » y sont systématiquement retrouvés. Est-il possible que les mutations décrites dans la séquence adénome à carcinome de Fearon et Vogelstein (1990) correspondent en fait aux différents « Hallmarks of cancer » décrit par Hanahan et Weinberg (2000) ? Selon Gatenby et Gillies (2008), les différents intermédiaires de la séquence adénome à carcinome sont en fait le reflet de mutation conférant au moins un « Hallmark of cancer ». Plus encore, ces mutations doivent subvenir dans un ordre bien déterminé afin de surmonter les mécanismes physiologiques (barrières) permettant de freiner/éliminer la progression tumorale. En d'autres termes, les cellules possédant des mutations affectant la croissance due à l'inhibition de contact seront beaucoup plus susceptibles de proliférer et produire l'expansion clonale initiale que les mutations affectant l'angiogenèse. De plus, les évènements mutationnels par lesquels s'effectuera la transformation des cellules peuvent être accomplis par un grand nombre de combinaisons relativement équivalentes phénotypiquement parlant. Ce concept d'équivalence fonctionnelle permet de reconnaître que le gain d'une caractéristique propre au cancer « hallmark of cancer » par une cellule peut être attribuable à diverses mutations (Gatenby and Gillies, 2008).

1.2.3 Gènes suppresseurs de tumeurs

Les mutations n'apparaissent pas toutes au cours du temps. Certains individus naissent avec des prédispositions héréditaires et ces mutations se retrouvent dans toutes les cellules de leur corps. L'étude de la génétique des cancers héréditaires a elle aussi largement contribué à l'amélioration de la modélisation de la cancérogenèse. La curiosité scientifique d'Alfred G. Knudson a été piquée lorsqu'il s'est intéressé au rétinoblastome, un cancer des

rétinoblastes pouvant se manifester chez les enfants de moins de deux ans. Bien que 40% des cas soient reliés à une mutation héréditaire, celle-ci n'est pas suffisante pour causer le cancer (Knudson, 1971). Selon ses calculs, il prédit que la formation de rétinoblastome requiert deux mutations, soit deux somatiques dans les cas non héréditaires; soit une somatique et une dans la lignée germinale dans les cas héréditaires (Hethcote and Knudson, 1978). C'est que nos gènes sont en deux exemplaires (allèles), l'un provenant de la contribution paternelle et l'autre maternelle. Dans ces cas, le phénotype serait donc lié à la perte de fonction des deux allèles, et non à un gain de fonction comme les oncogènes, modèle qui est maintenant connu sous le nom de « Two-hit hypothesis » (Nordling, 1953). Ces gènes sont importants pour le maintien de la prolifération normale des cellules et de la réparation de l'acide désoxyribonucléique (ADN) et sont collectivement nommés gènes suppresseurs de tumeurs (tumor suppressor genes, TSG). Dépendamment de leurs fonctions, on peut les sous-diviser en deux catégories: Les « gatekeepers » (ex: tumor protein p53 (TP53), Retinoblastoma (RB1)) inactivent directement la transformation ou promeuvent la mort cellulaire et les « caretakers » (ex: breast cancer 1 and 2, early onset, BRCA1-2) voient à la stabilité du génome (Tableau 1). La plupart des mutations héréditaires n'impliquent donc pas des oncogènes, probablement parce que celles-ci sont létales. En effet, 90% des mutations dans les cellules germinales sont récessives (Futreal *et al.*, 2004) alors que 90% des mutations somatiques sont dominantes, c'est-à-dire que la mutation d'un seul allèle contribue à la tumorigenèse. À ce jour, près de 2% des gènes humains seraient impliqués dans le processus de tumorigenèse (Santarius *et al.*, 2010). Malgré tout, rares sont les cancers ne nécessitant que deux mutations pour générer une masse tumorale importante.

Tableau 1 Prédipositions héréditaires liées à des gènes suppresseurs de tumeurs.

Les principaux TSGs ainsi que la condition pathologique souvent associée et susceptible de se développer pour les individus porteurs d'une mutation héréditaire (www.cancer.gov).

Gène(s) porteur(s) de la mutation héréditaire	Condition pathologique souvent associée
RB1	Rétinoblastome
TP53	Syndrome de Li-Fraumeni (plusieurs types de cancers)
APC	Polypes adénomateuses
MLH1, MSH2, MSH6,	Syndrome de Lynch (Cancer du côlon sans polypes)
WT1	Tumeur de Wilm (Néphroblastome)
BRCA1, BRCA2	Cancer du sein et de l'ovaire
VHL	Maladie de Von Hippel-Lindau (Hémangioblastome)
PTEN	Syndrome de Cowden (Hamartome)

1.2.5 Rôle du microenvironnement de la tumeur

Les recherches centrées sur les cellules cancéreuses et leur génome ont permis de découvrir les oncogènes et les TSGs et jusqu'à un certain point à modéliser la cancérogenèse. En parallèle, une série d'évidences confirme l'apport non négligeable des cellules d'apparence normale avoisinant la tumeur: le microenvironnement de la tumeur. En plus d'une matrice extracellulaire spécifique, ce dernier comprend des cellules du système immunitaire et des cellules mésenchymales (endothéliales et fibroblastes). Elles sont collectivement appelées cellules stromales. On parle de stroma activé, de desmoplasie ou plus souvent de «cancer-associated fibroblasts» (CAFs) lorsque l'on réfère au stroma présentant de l'inflammation aux abords d'une plaie ou d'une tumeur. La matrice extracellulaire est caractérisée par un dépôt important de collagène de type I.

Les CAFs sont très semblables aux myofibroblastes retrouvés au site d'inflammation d'une plaie, à la différence importante que les CAFs n'induisent pas la guérison de la « blessure » et ne se résorbent pas (Dvorak, 1986). Contrairement aux myofibroblastes, les CAFs possèdent une « lamina » et sont dotés de myofilaments. On peut également les identifier grâce à l'expression de certains marqueurs comme l'alpha smooth muscle actin (α -SMA), la vimentin et l'exon EDA de la fibronectin (Eyden, 2009). À noter que les termes myofibroblastes et CAFs sont interchangeables dans la littérature scientifique en désignant de façon erronée la même entité (Eyden, 2009). La présence de certains de ces marqueurs de

CAFs corrèle significativement avec un pronostic défavorable dans les cas de cancer du pancréas et du poumon, suggérant que les CAFs participent à la progression, l'agressivité et/ou la résistance aux traitements des cellules cancéreuses (Infante *et al.*, 2007, Takahashi *et al.*, 2011).

Pour démontrer plus directement les effets pro-tumorigéniques des CAFs, les groupes de Tlsty et Cunha ont administré un mélange de CAFs et de cellules épithéliales de prostate humaine ou un mélange similaire avec des fibroblastes normaux à une souris immunodéficiente. Seule la préparation de CAFs était en mesure de stimuler la formation de tumeur (Olumi *et al.*, 1999). Une expérience semblable avec des CAFs issus de microenvironnements de cancer du sein et des lignées cellulaires épithéliales de cancer du sein a non seulement corroboré l'étude établie sur la prostate, mais a également permis de mettre en valeur le pouvoir pro-angiogénique des CAFs (Orimo *et al.*, 2005). Les CAFs ont des capacités contractiles accrues, promeuvent l'angiogenèse et stimulent la croissance des cellules épithéliales via la sécrétion de cytokines et de facteurs de croissance tels que le «vascular endothelial growth factor» (VEGF). De façon paracrine, les cellules cancéreuses sécrètent des facteurs de croissance comme le «tumor growth factor beta » (TGF- β) et le « platelet-derived growth factor» (PDGF) qui transforment le stroma. À leur tour, les CAFs sécrètent des facteurs de croissance afin de soutenir la prolifération de la tumeur. De plus, les « matrix metalloproteases-2, 3 et 9 » (MMP-2, MMP-3 et MMP-9) sont sécrétées et façonnent la matrice extracellulaire (Kalluri and Zeisberg, 2006) (Fig. 3).

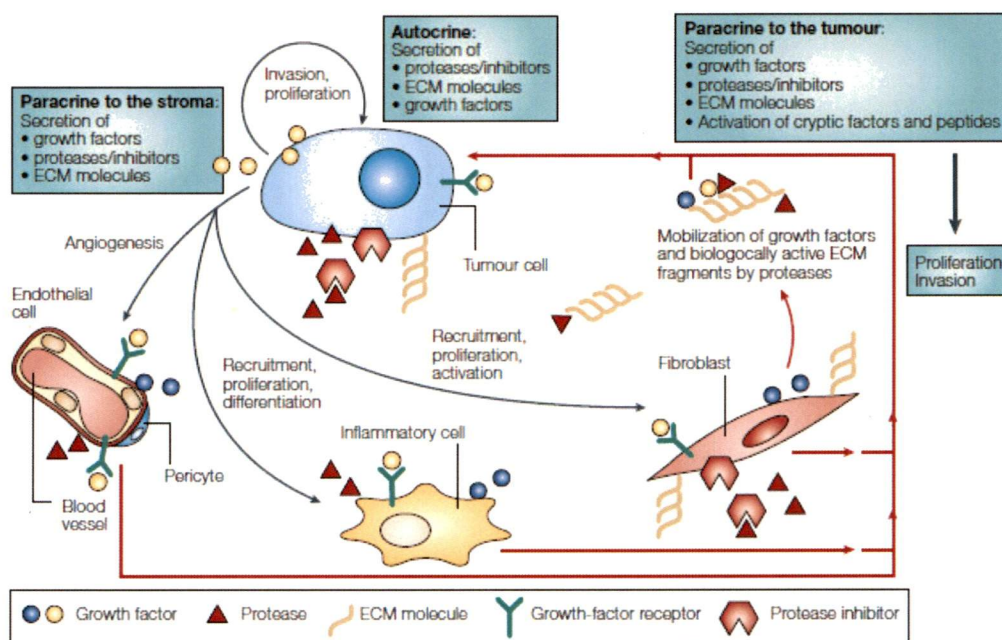


Figure 3 Interactions hétérotypiques entre les cellules épithéliales cancéreuses et les cellules stromales du microenvironnement.

Les cellules cancéreuses activent les fibroblastes par le biais de la sécrétion de facteurs de croissance. À leur tour, ceux-ci supportent la croissance de la tumeur en relâchant des facteurs pro-prolifération et pro-angiogéniques. Ensemble, ils reconditionnent le microenvironnement en remodelant la matrice extracellulaire par le biais de protéases. Image tirée de Mueller et Fusenig (2004).

1.2.6 La transition épithéliale-mésenchymale et mésenchymale-épithéliale

Outre les interactions paracrines entre les cellules stromales et épithéliales, les cellules épithéliales présentent une plasticité leur permettant d'adopter des caractéristiques mésenchymales de façon réversible. C'est ce qu'on appelle la transition épithéliale-mésenchymale (EMT) et son inverse, la transition mésenchymale-épithéliale (MET). En plusieurs occasions au cours du développement normal de l'embryon, une suite finement coordonnée de ces transitions sont nécessaires, notamment pour générer les valves du cœur et les néphrons du rein (Fig. 4). Suivant une première EMT, le mésoderme se condense pour former des structures épithéliales temporaires, dont la splanchnopleure. Ce dernier se différencie en cellules mésenchymales précurseurs des cellules du système hématopoïétique. Une troisième EMT est nécessaire pour former les valves cardiaques (Fig. 4A). La formation du néphron est l'exemple où la MET est la mieux étudiée jusqu'à présent. Les cellules mésenchymales pluripotentes bordant le bourgeon urétéral proviennent

d'une première EMT. Celles-ci s'agrègent et se différencient en cellules épithéliales en formant d'abord une sphère via une signalisation dépendante de « wingless-type MMTV integration site family, member 4 » (WNT4), puis un corps en forme de « S » dont une portion poursuivra la différenciation jusqu'au néphron (Faa *et al.*, 2012) (Fig. 4B).

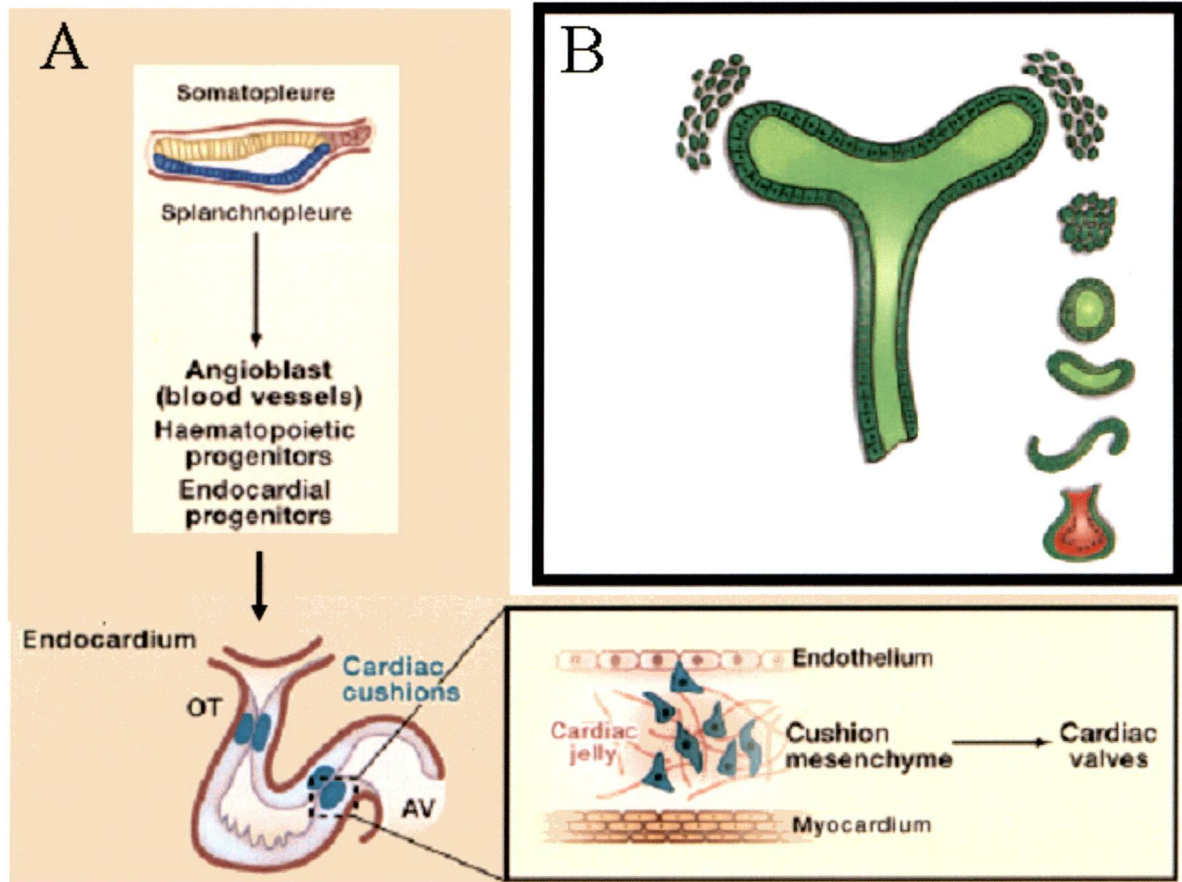


Figure 4 Rôle de l'EMT/MET dans le développement normal.

Développement des valves cardiaques. Les cellules épithéliales du somatopleure et splanchnopleure subissent une EMT générant des cellules mésenchymales qui différencieront en angioblastes. Une seconde EMT est nécessaire afin de générer le tissu mésenchymateux lâche, qui est au cœur de la chambre artérielle (outflow tract (OT)) et du canal atrioventriculaire (AV), et qui servira de précurseur aux valves cardiaques. Image adaptée de Thiery (2009). B) Représentation schématique de la transition mésenchymale-épithéliale lors du développement du néphron. Les cellules du cap mésenchymateux se transforment progressivement en cellules épithéliales (glomérule) en passant par les intermédiaires suivants : la vésicule rénale, le corps en forme de virgule, le corps en forme de « S ». Image tirée de Faa (2012).

Pour compléter l'EMT avec succès, la cellule épithéliale doit subir plusieurs changements importants (Fig 5). Premièrement, les jonctions serrées qui favorisent le rapprochement des cellules épithéliales au niveau apical doivent être démantelées. Deuxièmement, la polarité de la cellule se résout par endocytose des jonctions adhérentes. Troisièmement, il est suggéré que la constriction de la zone apicale, que ce soit au niveau de la circonférence ou directement à travers la cellule tient lieu de force motrice. Quatrièmement, le désencrage de la couche basale s'effectue grâce à des protéases qui façonnent la matrice extracellulaire. Cinquièmement, les nouvelles cellules mésenchymales doivent migrer de la structure épithéliale, se frayer un chemin à travers la matrice extracellulaire et résister à l'apoptose (Shook and Keller, 2003).

Pour orchestrer cette batterie de changements, plusieurs stimuli permettant d'observer l'EMT *in vitro* ont été répertoriés, un des plus utilisés étant le TGF- β . Ce facteur se lie à des récepteurs membranaires de type tyrosine kinase et signale principalement via les effecteurs SMAD (Tian and Schiemann, 2009). Peu importe le type de stimuli, tous semblent ultimement converger vers un petit groupe de facteurs de transcription: Goosecoid, Snail et Twist. La surexpression individuelle de ces trois facteurs récapitule l'EMT dans les cellules humaines mammaires épithéliales du cancer du sein (HMLE). Par ricochet, ils augmentent l'expression d'un deuxième sous-groupe de facteurs de transcription (Zeb1-2 et FOXC2) (Orimo *et al.*, 2005). L'une des cibles communes dont l'expression est régulée négativement par ces facteurs de transcription est la Cadhérine-E (CDH1), une composante typique des jonctions cellulaires épithéliales (Fig. 6).

L'EMT n'intervient pas que dans le développement normal, mais également en condition pathologique. Par exemple, une brèche de l'épiderme stimule également le programme EMT des kératinocytes. Une fois la cascade de coagulation activée, les kératinocytes en bordure de la blessure migrent vers la cavité à ré-épithélialiser. En parallèle, des fibroblastes « s'activent » et exploitent leur capacité de contractions accrues afin de diminuer la taille de la plaie (Fig 7A) (Nakamura and Tokura, 2011).

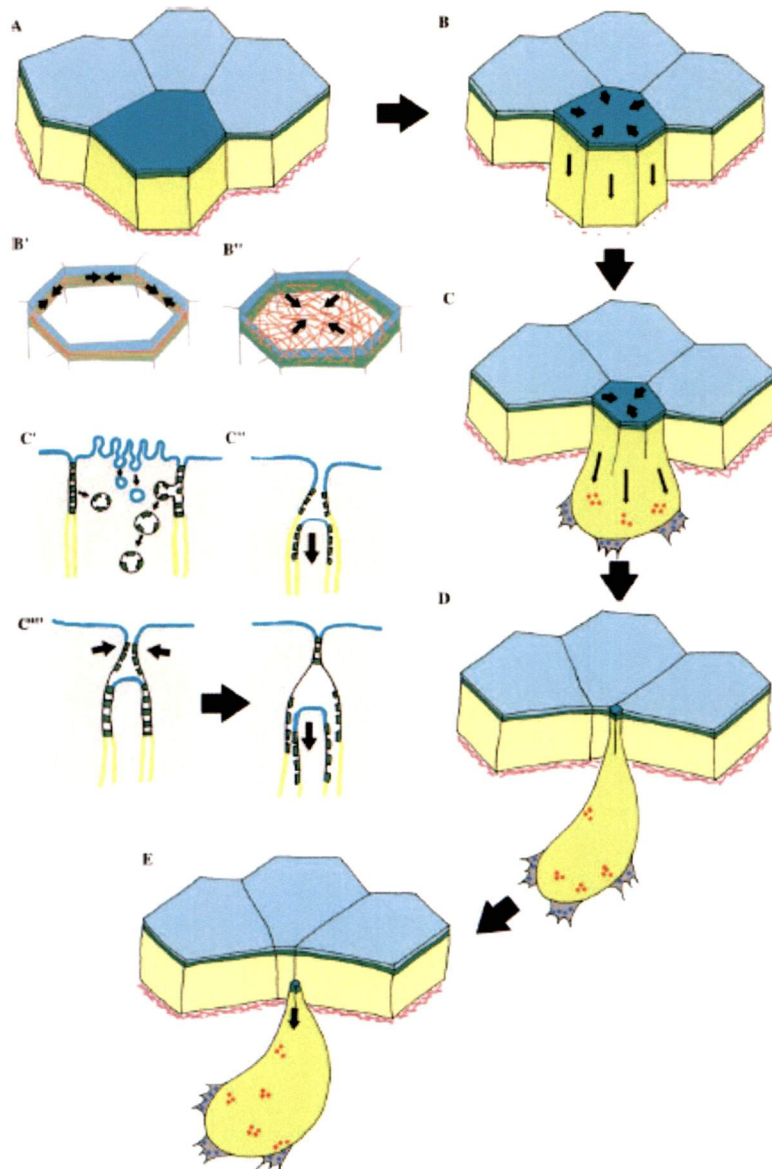


Figure 5 Schématisation des étapes de l'EMT.

Étapes de transition dans la conversion d'une cellule épithéliale et une cellule mésenchymale. A) Les jonctions serrées qui favorisent le rapprochement des cellules épithéliales au niveau apical sont démantelées. B) La polarité de la cellule se résout par endocytose des jonctions adhérentes. C) Il y a constriction de la zone apicale, qui tient lieu de force motrice. D) Le désencrage de la couche basale s'effectue grâce à des protéases qui façonnent la matrice extracellulaire. E) Les nouvelles cellules mésenchymales migrent de la structure épithéliale et se frayent un chemin à travers la matrice extracellulaire. Image adaptée de Shook and Keller (2003).

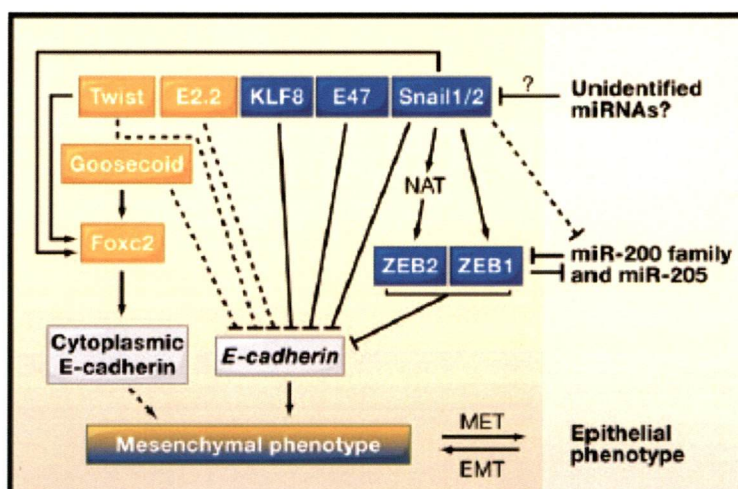


Figure 6 Facteurs de transcription impliqués dans l'EMT/MET.

Les facteurs de transcription comme Snail et Slug (SNAI2), Zeb1 et Zeb2 répriment directement l'activité du promoteur de E-cadherin alors que Twist, Goosecoid et Foxc2 agissent indirectement. Image tirée de (Thiery *et al.*, 2009).

Les facteurs de croissance sécrétés par le microenvironnement de la tumeur permettent d'induire un caractère mésenchymal aux cellules épithéliales. Les jonctions serrées caractéristiques des cellules épithéliales sont relâchées et la polarité apicale-basale est abolie. Ainsi, les récepteurs de la membrane basolatérale autrefois restreints aux signaux provenant du stroma sont maintenant soumis à l'action autocrine de facteurs de croissance (Wells *et al.*, 2011). Ces stimulations entraînent des changements morphologiques permettant aux cellules cancéreuses d'envahir les tissus avoisinants. Les cellules qui réussissent à implanter un nouveau foyer de cellules cancéreuses (métastase) retournent à leur aspect épithélial d'origine en passant par une MET (Fig. 7B). La principale différence entre l'EMT de type I (développement normal) vis-à-vis celle de type II (réparation des tissus) et III (cancérogenèse) concerne le microenvironnement (CAFs et dépôt de collagène de type I).

Les caractéristiques de plasticité des cellules cancéreuses épithéliales ne se limitent pas qu'au caractère pro-migratoire; l'EMT permet également de générer des cellules ayant des propriétés qui s'apparentent aux cellules souches (cancer stem cells, CSC) (Polyak and Weinberg, 2009). Pour ajouter à la complexité, la reprogrammation *in vitro* des fibroblastes en cellules souches induites pluripotentes (induced pluripotent stem cells, iPSC) est

stimulée par le «bone morphogenic 7» (BMP7) et nécessite un intermédiaire épithélial (donc une MET) (Li *et al.*, 2010, Samavarchi-Tehrani *et al.*, 2010). Ces résultats permettent de relancer le débat sur l'origine des CSCs. Il est probable que seule une sous-population de cellules possède les propriétés fonctionnelles des CSCs. Alternativement, toutes les cellules cancéreuses pourraient être fonctionnellement équivalentes, mais présenteraient des comportements différents en fonction du microenvironnement auquel elles sont exposés.

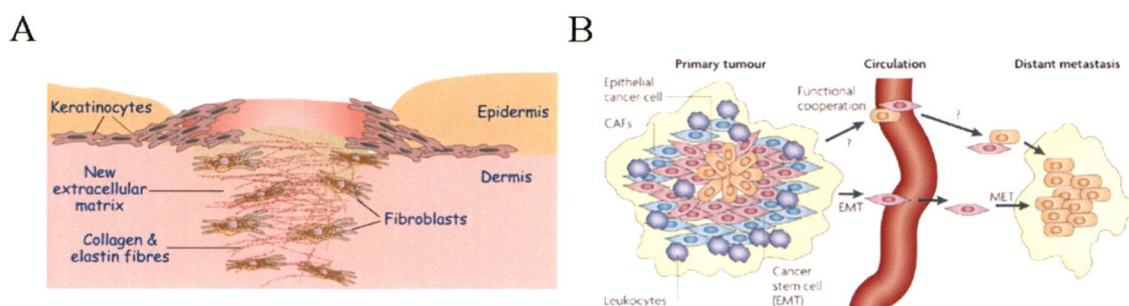


Figure 7 Rôle de l'EMT/MET dans la cancérogenèse.

A) Réparation d'une plaie superficielle de la peau. Des kératinocytes migrent vers la plaie à cicatriser tandis que des fibroblastes activés rapprochent les deux extrémités de la plaie. Image tirée de (<http://www.hcc.uce.ac.uk/physiology/woundhealing.htm>). B) Invasion et métastase. Des cellules épithéliales cancéreuses subissent une EMT et atteignent un site secondaire via la circulation sanguine. Celles-ci subissent une MET afin de retrouver leurs capacités de prolifération et d'adhésion. Image tirée de (Polyak and Weinberg, 2009).

1.2.7 Origines des CAFs

Il est généralement admis que la plupart des myofibroblastes lors de fibroses ou de processus de réparation des tissus sont issus de la prolifération active des fibroblastes résidents ou de cellules souches mésenchymales. Or, des études permettant de suivre l'évolution de type cellulaire choisi a permis de révéler la contribution non négligeable des cellules endothéliales et des cellules épithéliales. Ainsi, une fraction des myofibroblastes proviennent de cellules épithéliales par le biais de l'EMT lors de fibrose des poumons et de l'intestin (Quaggin and Kapus, 2011). Plus précisément, il s'agirait d'une transition épithélial-fibroblaste-myofibroblaste dont seuls quelques fibroblastes seraient en mesure de

compléter en exprimant α -SMA (Quaggin and Kapus, 2011). Comme les myofibroblastes et les CAFs sont très similaires, il est tentant de spéculer sur la contribution des cellules épithéliales cancéreuses comme source de CAFs. Ceci fait actuellement l'objet de débat (Cirri and Chiarugi, 2011).

Concernant la cancérogenèse, les premières études de séquençage ont rapporté des mutations des gènes TP53 et «phosphatase and tensin homolog» (PTEN) dont le patron est mutuellement exclusif dans le stroma et les cellules épithéliales du cancer du sein, ce qui remet en doute l'implication de l'EMT (Kurose *et al.*, 2002). Au contraire, une étude sur le cancer de l'ovaire a révélé un fort pourcentage de mutations dans le stroma activé identique à celles retrouvées dans les cellules épithéliales cancéreuses, appuyant l'implication de l'EMT comme source de CAFs (Tuhkanen *et al.*, 2006). Par contre, une étude plus récente et réalisée sur un plus grand nombre de patientes n'a pas été en mesure de démontrer de mutations tant dans le stroma de cancer du sein que de l'ovaire (Qiu *et al.*, 2008). En fait, l'absence de mutations somatiques dans l'ADN des CAFs concorde bien avec leurs effets non tumorigéniques. Mais alors, comment expliquer que les CAFs, qui se développent sous l'influence de cellules cancéreuses, puissent maintenir indépendamment leur phénotype jusqu'à dix passages cellulaires (Olumi *et al.*, 1999)? Il est possible qu'une partie de la réponse se retrouve au sein de modifications épigénétiques (Hu *et al.*, 2005).

1.3 Cancer de l'ovaire

Typiquement, la patiente atteinte du cancer de l'ovaire est ménopausée et âgée d'environ 60 ans lors du diagnostic initial. Elle présente des symptômes vagues s'apparentant à des troubles digestifs bénins. Il est possible qu'elle se plaigne de constipation ou d'urination fréquente due à une pression de la masse cancéreuse sur le rectum ou la vessie, respectivement. Dans les cas d'un cancer de stade avancé, la cavité abdominale peut se remplir de liquide (ascites) et de masses, et peut aller jusqu'à l'épanchement pleural. Si des ganglions métastatiques sont présents, ceux-ci peuvent être palpables au niveau inguinal, supraclaviculaire ou axillaire. Ces symptômes non spécifiques se manifestent généralement très tard dans la progression de la maladie, d'où le surnom de « tueur silencieux ». Un

échantillon issu de la chirurgie est requis pour poser le diagnostic définitif (www.cancer.ca).

Lors de la progression du cancer, les cellules cancéreuses épithéliales deviennent généralement moins différenciées. On a longtemps cru que le cancer de l'ovaire faisait exception à cette règle puisque les cellules cancéreuses examinées ressemblent en fait à des cellules spécialisées de tissus normaux (Auersperg *et al.*, 2001) (Fig. 8). Plus précisément, il peut s'agir de tissu normal de la trompe de Fallope (séreux), de l'endométrium (endométriïde), du système gastro-intestinal (mucineux), du col de l'utérus ou du vagin (cellule claire) ou finalement de la vessie (Brenner) (Bast *et al.*, 2009, Kurman and Shih, 2010). Chacun de ces histotypes, reconnu par l'Organisation Mondiale de la Santé, est subdivisé en trois catégories soit bénin, malin et à faible taux de malignité (aussi appelé low malignant potential, LMP).

La théorie sur les cellules d'origine du cancer de l'ovaire prônée jusqu'à tout récemment met logiquement de l'avant les cellules épithéliales en surface de l'ovaire (ovarian surface epithelium, OSE). Lors de l'ovulation, les cellules épithéliales en surface sont internalisées et endommagées, provoquant un risque accru de mutations lors de la réparation subséquente. En accord avec cette théorie, les femmes ayant eu de multiples grossesses, des périodes d'allaitement prolongées ou recours aux contraceptifs oraux sont moins à risque d'être diagnostiquées du cancer de l'ovaire (Landen *et al.*, 2008). En ce sens, les femmes ayant subi un plus grand nombre de cycles ovulatoires ont plus de chance de développer un cancer de l'ovaire (Fathalla, 1971). Par contre, ceci n'explique pas pourquoi les contraceptifs de progestérone (qui n'inhibent pas l'ovulation) sont aussi efficaces que les contraceptifs qui inhibent l'ovulation (Risch, 1998). De plus, les femmes ayant le syndrome polykystique (moins de cycle ovulatoire) sont plus à risque de cancer de l'ovaire (Schildkraut *et al.*, 1996). Cette apparente discordance de l'hypothèse de l'ovulation incessante peut être expliquée par une susceptibilité accrue aux mutations lors de processus inflammatoires.

Dans une minorité de cas, un certain nombre de sous-type séreux, mucineux, endométriïde et à cellules claires suivent une progression selon la séquence classique adénome à

carcinome. La progression selon cette séquence proviendrait des OSEs qui s'invagineraient pour former un kyste d'inclusion. Celui-ci donnerait lieu à une lésion précurseur nommée «atypical proliferative serous tumor» (APST) aussi appelée LMP. Celle-ci se développe de façon non-invasive avec une architecture de type micropapillaire pour donner lieu au précurseur «micropapillary serous carcinoma» (MPSC), et finalement aboutir en adénocarcinome séreux de bas grade (Fig. 8). Cette voie de faible grade (ou type I) est caractérisée par une sur-activation de la voie des «mitogen-activated protein kinases» (MAPK) suite à des mutations somatiques mutuellement exclusives dans les oncogènes KRAS, v-raf murine «sarcoma viral oncogene homolog B1 (BRAF) ou v-erb-b2 erythroblastic leukemia viral oncogene homolog 2» (ERBB2) (Kurman and Shih, 2008). Comme ces cellules prolifèrent plutôt lentement, les chimiothérapies conventionnelles sont généralement inefficaces. En revanche, la prise en charge de la patiente par salpingo-oophorectomie (ablation de l'ovaire et de la trompe de Fallope) est une mesure efficace (Trope *et al.*, 2009).

Le type II (voie de haut grade), particulièrement le sous-type séreux (ovarian serous high grade, OSHG), est le plus fréquent (75%). La quasi-totalité des cas possèdent des mutations dans le gène TP53. Contrairement à d'autres cancers épithéliaux, les cellules d'origine de la maladie demeurent à ce jour mal définies. Une certaine proportion de tumeurs pourrait provenir de kyste d'inclusion, mais la majorité serait issue des cellules épithéliales de la trompe de Fallope ou encore du péritoine. Les lésions précurseurs (serous tubal intraepithelial carcinoma, STIC) comporteraient des mutations dans le gène TP53 et mèneraient à la formation de OSHG. En fait, la tumeur de haut grade prolifère très rapidement et semble apparaître *de novo*. Conséquemment, la détection précoce du cancer ovarien de type II est à toute fin pratique inexistante et la majorité des cancers sont diagnostiqués en stage avancé. Dans ces cas, le traitement de première ligne consiste à une cytoréduction de la tumeur suivi d'une chimiothérapie au taxane et platine, bien que la chimiothérapie néoadjuvante soit également pratiquée (Chi and Schwartz, 2008).

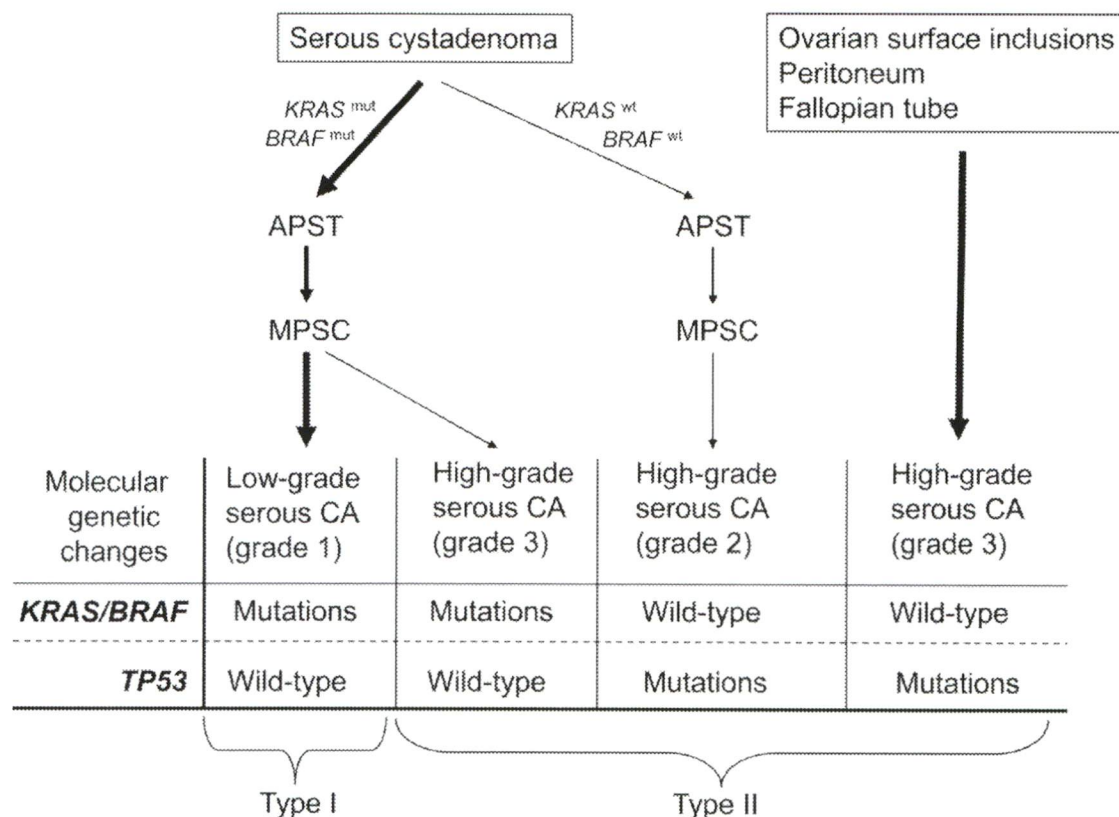


Figure 8 Progression séquentielle du cancer de l'ovaire.

Classification des cancers de l'ovaire en deux sous-types selon Kurman et Shih (2008). Tirée de (Kurman and Shih, 2008).

Des mutations somatiques du “gatekeeper” p53 sont retrouvées dans la quasi-totalité des cas (Ahmed *et al.*, 2010). Malgré tout, la majorité des cancers de l'ovaire ne sont pas reliés à des causes héréditaires bien que les défauts génétiques d'origine familiale dans des gènes de réparation de l'ADN, augmentent considérablement le risque de cancer du sein et de l'ovaire (BRCA1: 40-50%, BRCA2: 20-30%) (Prat *et al.*, 2005). En ce sens, une chirurgie préventive est recommandée afin de diminuer ce risque. C'est d'ailleurs lors de l'inspection soigneuse de ces tissus “normaux” que des lésions ont été observées pour la première fois dans les cellules épithéliales de la trompe de Fallope (Fallopian tube epithelial cells, FTE) (Colgan *et al.*, 2001; Medeiros *et al.*, 2006). En outre, l'examen moléculaire de tissus pairés provenant de FTE et de OSHG révèle des mutations identiques du gène TP53 (Callahan *et al.*, 2007). Il est important de mentionner que l'origine embryonnaire des FTEs est de type Mullerian (contrairement à mésothéliale pour l'OSE), ce qui cadre mieux avec les caractéristiques que présentent les différents sous-types histologiques du cancer de l'ovaire.

Faisant face à la cavité pelvienne, l'ovaire détient une situation anatomique particulière (Fig. 9). Celle-ci est située dans le tiers inférieur du tronc et regroupe les organes logés à l'intérieur des os du pelvis. Elle comprend ceux du système reproducteur tel l'utérus, les trompes de Fallope et les ovaires, ainsi que le petit intestin, la vessie et le rectum. La cavité abdominale est logée entre la cavité thoracique et la cavité pelvienne. La plupart des organes de l'abdomen sont recouverts d'une mince couche de tissu séreux nommé péritoine. À noter que les ovaires sont à découverts vis-à-vis les organes de la cavité pelvienne. Le site métastatique de la plupart des cellules cancéreuses ovariennes est l'omentum; un large coussin adipeux (approximativement 12 × 12 cm) situé entre l'estomac et le petit intestin (Doig and Monaghan, 2006). Alternativement, le péritoine constitue un site d'attachement fréquent (Doig and Monaghan, 2006). Fait intéressant, 50% des cas de cancers séreux d'origine pelvienne de haut stage possèdent également des lésions dans les trompes de Fallope (Levanon *et al.*, 2010). Inversement, de récentes études indiquent que dans certains cas, ce que l'on croyait être un cancer de l'ovaire primaire est en fait originaire du pelvis et atteint l'ovaire de façon secondaire (Kurman and Shih, 2010). L'ensemble de ces résultats ébranlent les fondements et forcent les chercheurs à réévaluer leurs classifications et hypothèses de recherche. En bref, le cancer de l'ovaire est maintenant perçu comme une multitude de type de cancers n'originant pas nécessairement de l'ovaire mais qui y aboutissent ultimement (Vaughan *et al.*, 2011). Conséquemment, ceci compromet un grand nombre d'études où les cellules épithéliales de l'ovaire (ou pire l'ovaire entier) ont été utilisées comme cellules normales de référence (Zorn *et al.*, 2003).

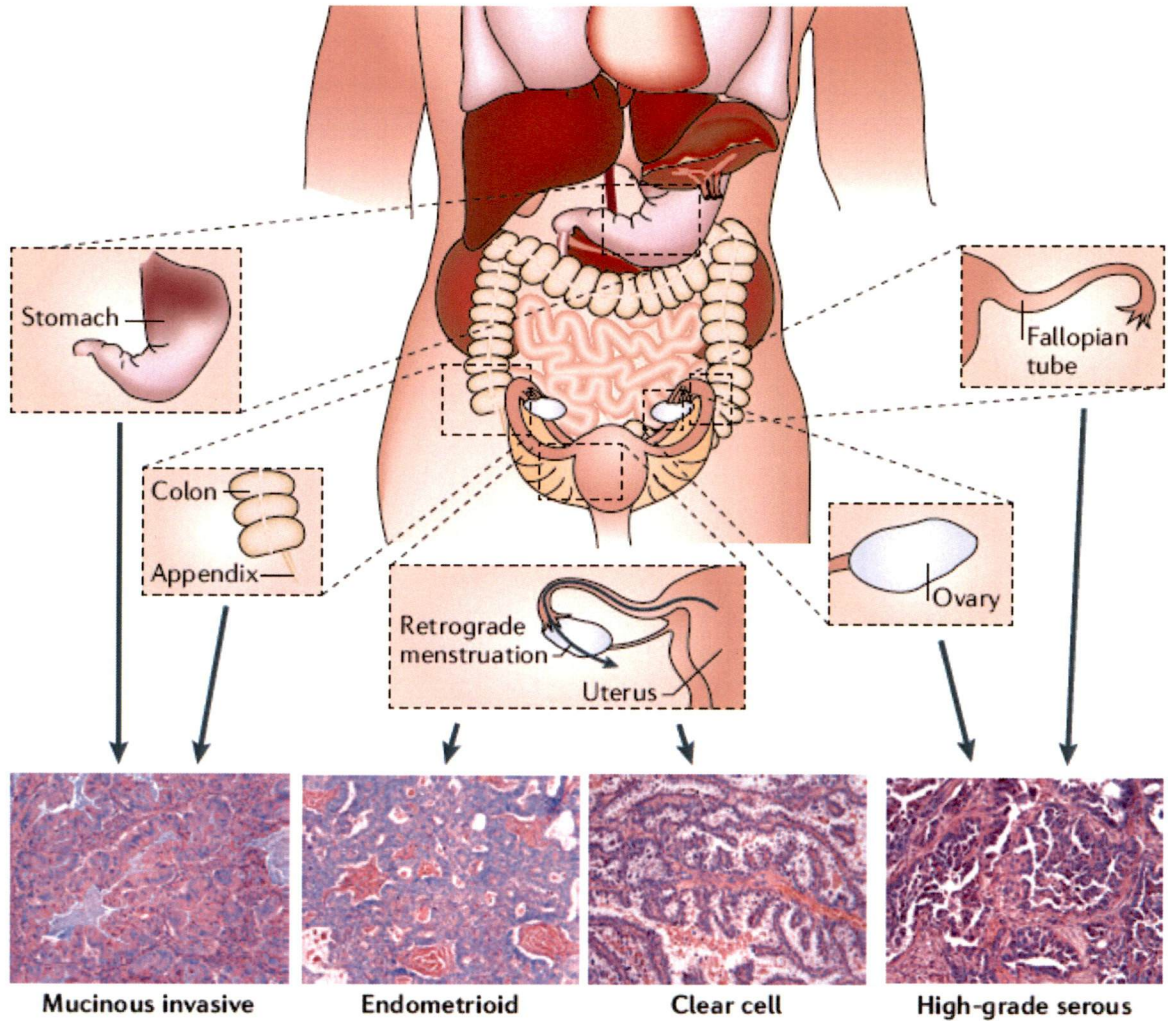


Figure 9 Origine du cancer de l'ovaire.

L'ovaire est perçu comme un site métastatique pour plusieurs cancers du système reproducteur ou gastro-intestinal. Des coupes histologiques colorées à l'hématoxyline et l'éosine (H&E) représentent les principaux sous-types histologiques du cancer de l'ovaire : séreux, endométriode, cellules claires et mucineuses. L'hématoxyline colore en mauve les cellules épithéliales alors que les cellules stromales apparaissent dans des teintes de rose sous l'effet de l'éosine. Tirée de (Vaughan *et al.*, 2011).

2 Épissage des ARNm

2.1 Épissage constitutif

Jusqu'à présent, la notion de gène a été discutée sans introduire et détailler le dogme de la biologie moléculaire moderne qui le sous-tend : Les gènes représentent l'information de référence, l'acide ribonucléique (ARN) le message et la protéine la fonction. Il y a donc transcription de l'ADN en ARNm puis traduction de celui-ci en protéine. La transcription de gènes s'effectue lorsque les facteurs généraux de transcription TBP (TATA binding protein), TFIID et TFIIB (transcription factor IID et IIB, respectivement) reconnaissent une séquence promotrice de la transcription (typiquement une séquence TATA) située 30 nucléotides (nt) en amont du site de début de la transcription sur l'un des deux brins de l'ADN. Ensuite, l'ARN polymérase II et d'autres facteurs de transcription s'assemblent au promoteur et la transcription en ARN débute.

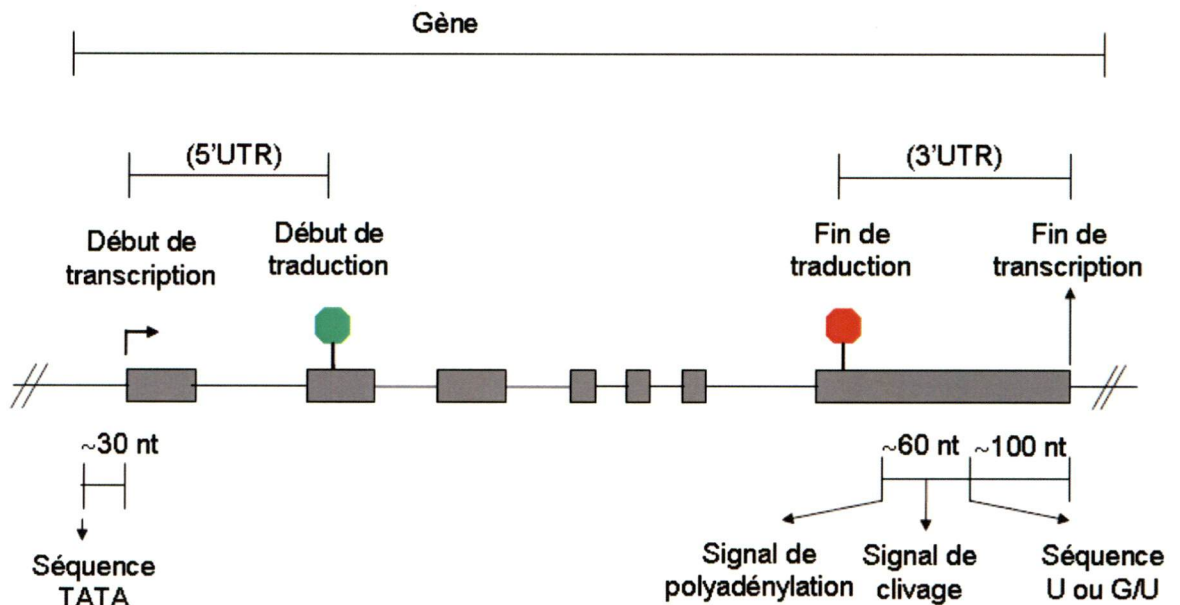


Figure 10 Unité génique.

Le début d'un gène est marqué par la présence d'une séquence promotrice (habituellement une séquence TATA) et autres séquences régulatrices (non-montrées). La portion transcrite inclut le 5'UTR, 3'UTR, les exons et les introns. Le gène moyen possède environ 8 exons dont la plupart sont de moins de 200 nt et sont séparés par des introns 10 fois plus grands. La fin de la transcription coïncide avec une séquence riche en U ou G/U. À noter que la queue polyA en 3' et la coiffe en 5' ne sont pas encodées dans le génome et sont ajoutées lors du processus de maturation du pré-ARNm.

Celle-ci transcrit jusqu'à quelques centaines de nucléotides après avoir atteint une séquence riche en U ou G/U située à 30 nt en aval du signal de clivage et 40 à 60 nt en aval du signal de polyadénylation (Fig. 10).

De façon co-transcriptionnelle, ce pré-ARNm subit plusieurs étapes de maturation. L'extrémité 5' de l'ARN est coiffée suivant trois étapes enzymatiques successives. D'abord, une ARN triphosphatase enlève le phosphate en gamma du premier nucléotide du pré-ARNm. Puis, une ARN guanyltransférase couple un guanosine-5'-monophosphate au pré-ARNm. Finalement, un groupement méthyl est introduit en position 7 de la guanine par une guanine-7-méthyltransférase pour générer une coiffe de type m⁷G(5')ppp(5')N. À l'extrémité 3', la protéine «cleavage and polyadenylation specificity factor» (CPSF) reconnaît le signal de polyadénylation et clive le pré-ARNm 10 à 30 nt plus loin. Ensuite, une polyA polymérase indépendante de la matrice ADN ajoute environ 200 adénines, formant ainsi une queue polyA (Lutz, 2008). Le processus de maturation doit également enlever les séquences ne faisant pas partie de l'ARNm mature (intron, intervening sequences) pour conserver uniquement les portions codant pour une séquence protéique (exons, expressed sequences) et les séquences non-traduites flanquantes en 5' et 3' (untranslated region, UTR). Un gène moyen possède environ 8 exons dont la plupart sont de moins de 200 nt et sont séparés par des introns 10 fois plus grands (Lander *et al.*, 2001).

L'excision d'un intron et le couplage subséquent des deux exons flanquants se nomment épissage. Si on résume cette réaction d'un point de vue chimique, il s'agit d'une double trans-estérification impliquant une adénine du point de branchement (branch point site, BPS) située à environ 30 nt en amont du site d'épissage 3' (3' splice site, 3'ss). Celle-ci procède à une substitution nucléophile de type 2 (S_N2) via l'hydroxyl en 2' (2'OH) sur le phosphate à la jonction du site d'épissage 5' (5' splice site, 5'ss), relâchant ainsi la portion 3' du pré-ARNm. Le groupement hydroxyl du phosphate en 3' attaque le phosphate du 3'ss, produisant un ARNm épissé et libérant l'intron sous forme de lasso (Fig. 11).

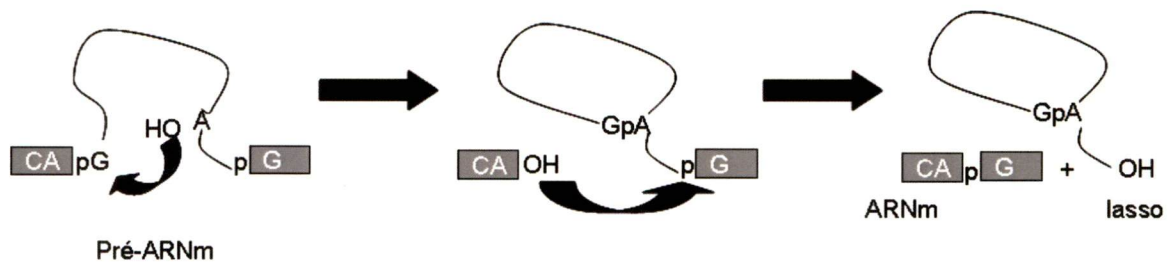


Figure 11 Réaction d'épissage.

L'épissage nécessite deux trans-estérifications afin d'obtenir un ARN épissé et l'intron résiduel sous forme de lasso. Lors de la première étape, l'adénine du BPS procède à une attaque nucléophile via le 2'OH sur le phosphate à la jonction du 5'ss, relâchant ainsi la portion 3' de l'exon en 5'. Lors de la deuxième étape, le groupement 3'OH de celui-ci attaque le phosphate du 3'ss, produisant un ARNm épissé et libérant l'intron sous forme de lasso.

L'épissage est réalisé par le splicéosome, une machinerie complexe regroupant cinq petits ARN nucléaires (snRNA, small nuclear RNA) et une centaine de protéines. Les snRNAs U1, U2, U4 et U5 possèdent une couronne heptamérique de protéines Sm communes (B/B', D1, D2, D3, E, F et G) et sont également liés par une batterie de protéines spécifiques à chacun (Patel and Bellini, 2008). Seul le snRNA U6 est entouré de protéines Sm-like. Les petits ARN nucléaires complexés de protéines Sm ou Sm-like prennent alors le nom de snRNP (small nuclear ribonucleoprotein) et sont des éléments essentiels du splicéosome. Certaines composantes du splicéosome sont en mesure de reconnaître les séquences consensus qui délimitent les jonctions exon/intron, intron/exon ainsi que le BPS (Fig. 12).

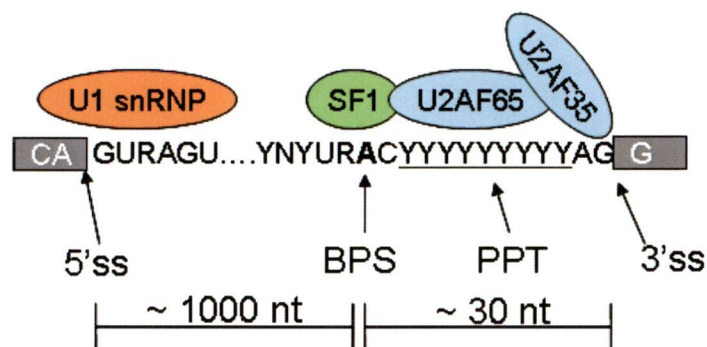


Figure 12 Séquences consensus impliquées lors de l'épissage.

Les séquences consensus du 5'ss, du BPS, du PPT et du 3'ss sont représentées. Le 5'ss est initialement reconnu par U1 snRNP; le BPS par SF1; le PPT par U2AF65 et les deux derniers nucléotides de l'intron par U2AF35.

Initialement, le 5'ss C/AAGGURAGU (la région intronique est soulignée) est lié par la portion 5' du snRNA U1 de façon ATP-indépendante et stabilisé par des protéines membres de la famille SR (S pour sérine, R pour arginine). Subséquemment, le BPS YNYURAC (en gras l'adénine directement impliquée dans l'épissage) est lié par le splicing factor 1 (SF1), la séquence de polypyrimidine (polypyrimidine tract, PPT) par la sous-unité de 65 kDa de U2 auxiliary factor (U2AF65) et les deux derniers nucléotides AG de l'intron par la sous-unité de 35 kDa de U2AF (U2AF35). Ceci forme ce que l'on appelle le complexe E et permet de définir les exons (voir figure 13). Ensuite, le snRNP U2 s'apparie de façon ATP-dépendante avec le BPS et déplace SF1. Cette dernière protéine est remplacée par SF3b14a, et ensemble, mènent au complexe A. Ici, les snRNPs U1 et U2 doivent faire contact « à travers l'intron », passant de la définition de l'exon à la définition de l'intron. Puis, les snRNPs U4, U5 et U6 se pré-assemblent avant de faire contact et créer le complexe B. À ce stade, le splicéosome n'est pas encore catalytiquement actif et doit relarguer les snRNPs U1 et U4 et procéder à des changements conformationnels majeurs pour en arriver au complexe B* (* pour splicéosome activé). La première S_N2 peut alors s'effectuer et mener au complexe C. Suite à quelques réarrangements au niveau du réseau d'interactions ARN-protéine, les deux exons sont joints, les snRNPs U2, U5 et U6 sont dissociés et l'intron est relâché. Les snRNPs U peuvent alors être recyclés pour un prochain cycle d'assemblage. À noter qu'il existe un système parallèle nommé splicéosome mineur qui permet l'épissage des ARNm dont les séquences consensus du 5'ss, BPS et 3'ss sont G/ATATCCTTT, TTTTCTTAACT et CAG, respectivement (Patel and Steitz, 2003). À l'exception du snRNP U5, les autres snRNPs U impliqués (U11, U12, U4atac et U6atac) sont spécifiques à ce système mineur.

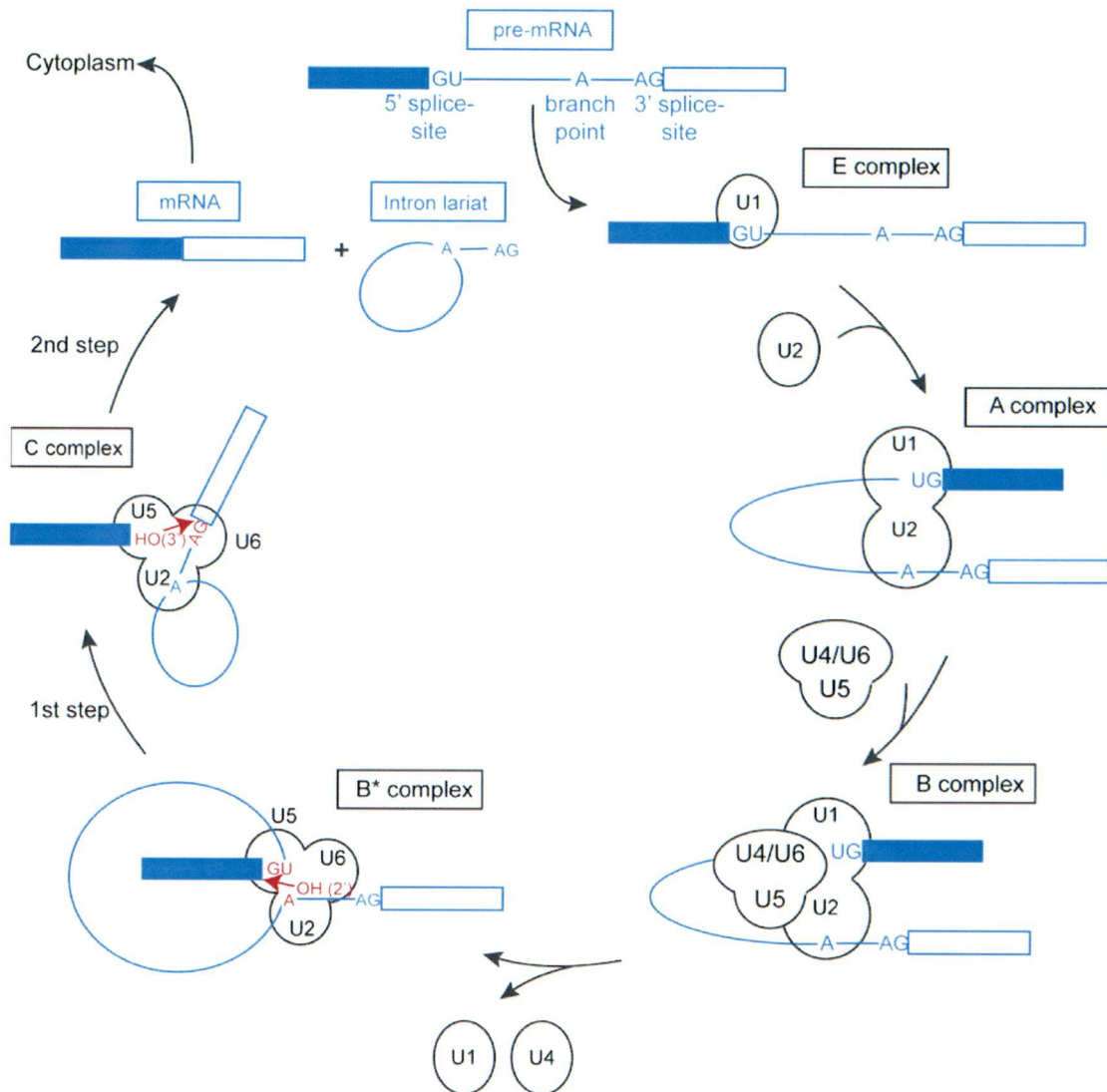


Figure 13 Cycle d'assemblage du spliceosome.

Le complexe E est formé lorsque U1 snRNP reconnaît le 5'ss. Ensuite, U2snRNP se lie au BPS et mène au complexe A. Les tri-snRNP U4/U5/U6 font contact avec U1 et U2, formant le complexe B. À ce stade, le spliceosome n'est pas encore catalytiquement actif et doit relarguer U1 et U4 snRNP pour en arriver au complexe B* (spliceosome activé). La première S_N2 s'effectue menant au complexe C. Suivant quelques réarrangements au niveau du réseau d'interaction RNA-protéine, les deux exons sont joints, les snRNPs U2, U5 et U6 sont dissociés et l'intron est relâché. Les U snRNPs peuvent alors être recyclés pour un prochain cycle d'assemblage. Image tirée de Krummel (2010).

2.2 Épissage alternatif

2.2.1 Type d'évènements d'épissage alternatif

L'épissage d'un pré-ARNm peut également être régulé. C'est-à-dire que sous certaines conditions, la séquence de l'intron à enlever, et donc celles des deux exons à joindre, ne sont pas reconnues fidèlement. Ainsi, il est possible de produire plusieurs ARNm matures distincts à partir d'un même produit de gène. C'est ce qu'on appelle l'épissage alternatif (alternative splicing, AS). En fait, l'AS n'est pas une exception à la règle, mais plutôt la norme chez l'humain: plus de 90% des gènes sont épissés alternativement (Wang *et al.*, 2008). Tel qu'il a été illustré à la figure 14, deux 5'ss peuvent être en compétition pour un même 3'ss ou inversement, deux 3'ss pour un même 5'ss. C'est ce qu'on appelle un évènement d'épissage alternatif (alternative splicing event, ASE) de type alternatif 5' (alt5', Fig. 14A) et alternatif 3' (alt3', Fig. 14B), respectivement. Le ASE le plus fréquent est l'exclusion complète de l'exon (Fig. 14C). Il existe également des cas où une série d'exons est exclue en bloc (exclusion d'exon multiple, Fig. 14D) ou incluse de façon mutuellement exclusive (Fig. 14E).

L'AS peut également être couplé à d'autres mécanismes de régulation de l'ARNm comme la polyadénylation (Lutz, 2008) (Fig. 14F). L'ARNm résultant exhibe une région 3'UTR différente sujette à une régulation via des facteurs de liaison à l'ARN ou de petits ARN non-codants. La dégradation spécifique via la voie «non-sense mediated decay» (NMD) est un mécanisme efficace de contrôle de la quantité d'ARNms que produit le gène (Garneau *et al.*, 2007) (Fig. 14G). Cela se produit lorsqu'un codon stop est introduit trop tôt dans l'ARNm par rapport au dernier exon. Il est prédit qu'environ le tiers des isoformes sont sujet au NMD (Lewis *et al.*, 2003). Le gène moyen possède donc plusieurs ARNm isoformes, dont la majorité sont susceptibles d'être traduites en protéines différentes pouvant résulter en des fonctions différentes. À noter que même si l'ASE se situe à l'extérieur de la portion codante, n'affectant donc en rien la séquence protéique finale, les portions UTR comprennent des séquences régulatrices de l'expression génique (Hughes, 2006). Bien que la notion de promoteur alternatif soit parfois incluse parmi les types possibles d'AS, il ne s'agit pas d'AS à proprement dit. L'AS produit deux ARNm ou plus

à partir du même pré-ARNm alors que l'initiation de transcription alternative produit deux pré-ARNms ou plus (Fig. 14H).

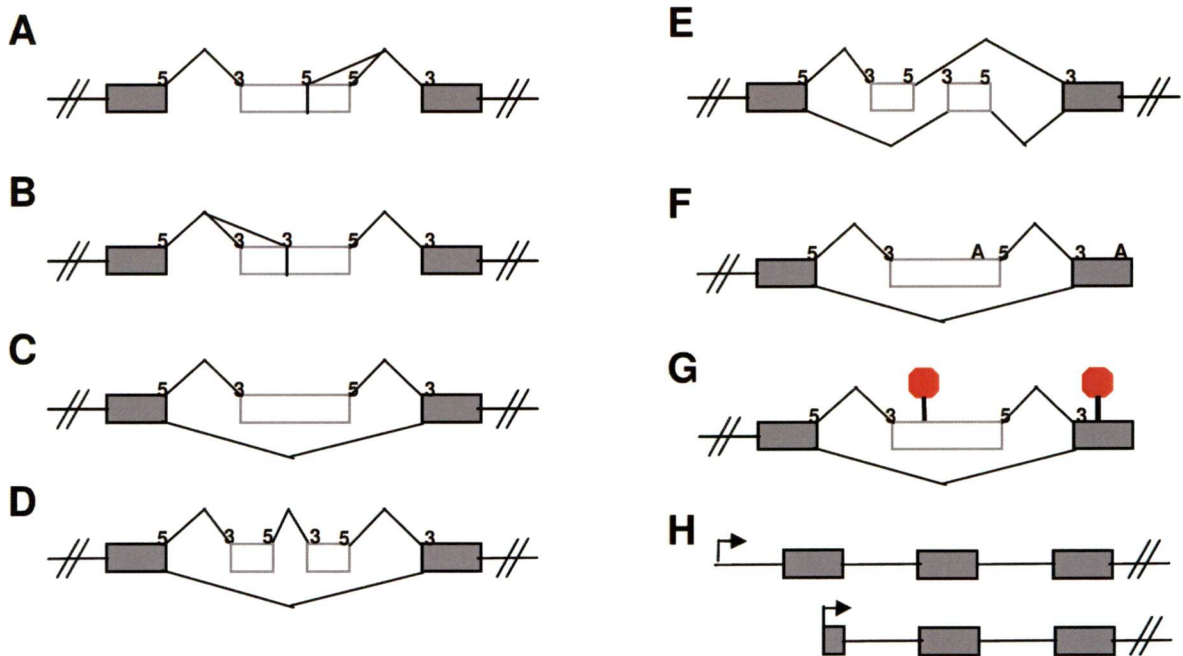


Figure 14 Type d'évènements d'épissage alternatif.

A) Alt5'. B) Alt3'. C) Exon cassette. D) Exons multiples. E) Exons mutuellement exclusifs. F) ASE couplé à la polyadénylation alternative. G) ASE couplé au NMD. H) Transcription alternative.

2.2.2 Mécanisme de régulation

2.2.2.1 Mécanisme de régulation classique

Les séquences consensus décrites à la section 2.1 définissant le 5'ss, le 3'ss, le BPS et la PPT sont communes aux exons constitutifs et alternatifs. Par contre, leurs forces relatives, c'est-à-dire leur fidélité face à la séquence consensus diffèrent et permettent de les classifier: les exons constitutifs possèdent des sites forts et les exons alternatifs des sites faibles. Des analyses bioinformatiques comparant les séquences des exons constitutifs vis-à-vis des introns et des exons avec sites faibles (Fairbrother *et al.*, 2002), ou vis-à-vis des pseudo-exons et des séquences 5'UTR de gènes sans intron (Zhang and Chasin, 2004) ont

révélé des motifs enrichis dans les exons alternatifs. Il apparaît que ceux-ci ont pour effet de promouvoir l'inclusion de l'exon et se nomment «exonic splicing enhancers »(ESEs).

La plupart des ESEs sont liés par des protéines de la famille SR. Ces protéines possèdent des domaines de liaison à l'ARN de type RNA recognition motif (RRM) en N-terminal ainsi que des domaines riches en arginine et sérine (domaine RS) en C-terminal. Ces domaines RS servent de module d'interaction protéine-protéine afin de recruter des composantes du splicéosome aux 5'ss ou 3'ss faibles (Graveley, 2000) (Fig. 15A). À l'opposé, les «heterogenous nuclear ribonucleoproteins» (hnRNPs) représentent une famille de facteurs d'épissage qui favorise généralement l'exclusion de l'exon en liant des «exonic splicing silencers» (ESSs). Ils opèrent en formant des multimères le long des exons (Zhu *et al.*, 2001), en bloquant le recrutement du snRNP U (Tange *et al.*, 2001) ou en détournant l'exon alternatif (looping out) loin du splicéosome (Martinez-Contreras *et al.*, 2006) (Fig 15B-D).

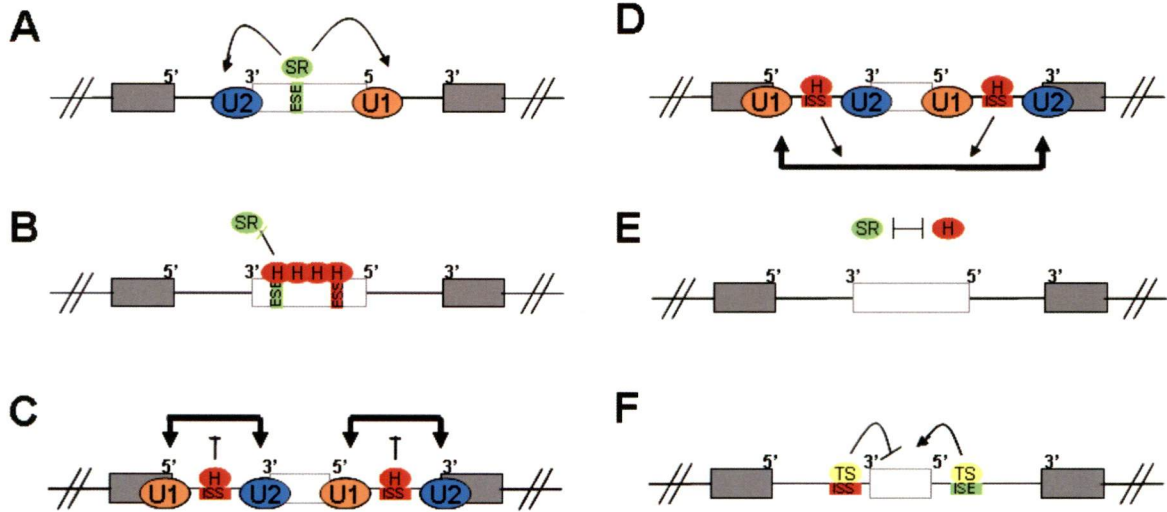


Figure 15 Mécanismes de régulation de l'épissage alternatif par des facteurs d'épissage.

A) Les protéines SR reconnaissent des ESEs et favorisent l'inclusion d'exon en recrutant les snRNPs U1 ou U2 au 5'ss ou 3'ss, respectivement. B) La polymérisation (3' vers 5') des hnRNPs le long de l'exon favorise l'exclusion de l'exon en bloquant l'accès des protéines SR aux ESEs. C) Les hnRNPs se lient à des ISSs et interfèrent avec la définition de l'intron, favorisant l'exclusion de l'exon. D) Les hnRNPs peuvent favoriser l'exclusion en rapprochant les exons éloignés. E) La somme totale des protéines SR vs hnRNPs dicte le sort de l'exon alternatif: Un surplus de hnRNPs favorise l'exclusion et vice versa pour les protéines SR. F) Les facteurs d'épissage exprimés de façon tissu-spécifique comme RBFOX2 et Noval1 favorisent habituellement l'inclusion d'exon lorsqu'ils se lient en aval du 5' ss alors qu'ils favorisent l'exclusion lorsqu'ils se lient en amont du 3' ss alternatif.

Les facteurs d'épissage de ces deux familles peuvent s'antagoniser fonctionnellement (Fig. 15E). Les plus étudiés sont les protéines «splicing factor 2» (SF2) et hnRNPA1 (Mayeda and Krainer, 1992, Eperon *et al.*, 2000). L'affinité de U1 snRNP pour un 5'ss peut être modulée *in vitro* à la hausse ou à la baisse lorsque SF2 ou hnRNPA1, respectivement, est majoritaire. Comme la liaison du snRNP U1 pour le 5'ss en aval est proportionnelle à son épissage, la simple variation du ratio entre les deux protéines suffit à influencer le destin de plusieurs pré-ARNms (Caceres *et al.*, 1994). Il existe également des séquences régulatrices introniques (intronic splicing enhancer, ISE; intronic splicing silencer, ISS) liées par des membres de la famille SR et des hnRNPs. Curieusement, la plupart des hnRNPs se liant dans l'intron en aval de l'ASE favorisent l'inclusion d'exon alors que la liaison de protéines SR à des séquences introniques favorise l'exclusion. Ces fonctions à priori contradictoires peuvent être réconciliées si l'on considère que les hnRNPs permettent de

définir les régions introniques et les protéines SR les régions exoniques. D'autres exemples d'antagonismes fonctionnels ont été répertoriés, notamment mettant en cause les effets répresseurs de la «polypyrimidine tract binding protein 1» (PTBP1) et activateurs de «CUG-BP and ETR3-like factors» (CELF) et pouvant agir avec (Zhang *et al.*, 2002, Gromak *et al.*, 2003) ou sans (Charlet-B *et al.*, 2002) compétition directe pour un site de liaison.

2.2.2.2 Expression tissu-spécifique de facteurs d'épissage

Afin de mieux comprendre les séquences régulatrices de l'épissage (splicing regulatory elements, SREs) introniques, plusieurs groupes de recherche ont examiné les séquences introniques aux abords des ASEs et en arrivent à la même conclusion: la séquence UGCAUG est la plus conservée (Yeo *et al.*, 2007, Voelker and Berglund, 2007). Cette séquence ARN est bien connue pour lier les facteurs d'épissage RNA «binding protein, fox-1,2 homolog» (RBFOX1-2) (Auweter *et al.*, 2006), dont l'expression est restreinte à quelques types tissulaires (RBFOX1: cardiaque, musculaire et neuronale; RBFOX2: cardiaque, musculaire et ovarien) (Kuroyanagi, 2009). Les mécanismes permettant à ces facteurs de réguler l'AS sont encore mal compris. Par contre, de récentes études à grande échelle entreprisent par notre groupe de recherche (Venables *et al.*, 2009) et d'autres groupes (Castle *et al.*, 2008, Zhang *et al.*, 2008) ont permis de mettre en lumière un principe général: la séquence UGCAUG se retrouvant en amont d'un ASE favorise l'exclusion (ISS) alors qu'elle favorise l'inclusion (ISE) lorsque positionnée en aval de l'ASE.

L'expression de plusieurs autres facteurs d'épissage est également limitée à un petit groupe de tissus (Chen and Manley, 2009). C'est le cas notamment de NOVA1-2 (neuro-oncological ventral antigen 1 and 2) qui sont exprimés uniquement dans le système nerveux. De façon intéressante, les séquences de liaison pour les familles muscleblind-like (Mbnl), Celf (Kalsotra *et al.*, 2008), epithelial-specific regulatory protein (Esrp) (Warzecha *et al.*, 2010) et Nova (Zhang *et al.*, 2010) sont également majoritairement introniques. Plus

encore, leur influence sur la décision d'épissage semble suivre la tendance décrite pour RBFOX1-2 (Fig. 15F).

2.2.3 Rôle dans la cancérogenèse

2.2.3.1 Dérégulation des isoformes

Mis à part des exemples ponctuels, le premier indice d'un lien plus généralisé entre l'AS et le cancer provient de l'analyse «d'expressed sequence tags» (ESTs). Les ESTs sont produits à partir du séquençage de bibliothèques d'ADN complémentaire (cDNA). Plusieurs groupes de recherche, dont Wang et collaborateurs (2003) ont classifié les milliers de sources de cDNA comme étant « normal » ou « cancer » pour leur permettre d'associer statistiquement certains isoformes au cancer (revue par Xing 2007). Quelques années plus tard, des micropuces à ADN spécifiques aux isoformes alternativement épissés révèlent la présence d'isoformes associés aux cellules cancéreuses, notamment de celles de lymphome Hodgkinien (Religio *et al.*, 2005), de prostate (Zhang *et al.*, 2006) et du côlon (Gardina *et al.*, 2006). Tout ceci démontre bien le dérèglement de l'AS dans les tissus cancéreux. Une des applications directes est d'exploiter ces isoformes comme biomarqueurs, que ce soit sous forme de test diagnostique ou pronostique (Brinkman, 2004).

Il y a de cela près de 20 ans, une étude a indiqué que 15% des mutations modifient les séquences consensus du 5'ss ou du 3'ss (Krawczak *et al.*, 1992). Il est fort probable que cette étude sous-estime l'impact réel des mutations sur l'AS puisque ceci ne tient pas compte des mutations dans les SREs (ESEs, ESSs, ISEs et ISSs). En fait, la plupart des analyses sont limitées aux ARNms matures et ne tiennent pas compte des séquences introniques. Une analyse probabiliste chiffre maintenant à 60% la proportion de mutation ayant un impact sur l'AS (Lopez-Bigas *et al.*, 2006). Certains des gènes atteints sont des TSGs et quelques-uns sont revus dans l'ouvrage de Srebrow et Kornbliht (2006) ainsi que Brown et Pettigrew (2008). De plus, un certain nombre de gènes reliés au cancer (63 sur 487 gènes en date du 15 mars 2012, <http://www.sanger.ac.uk/genetics/CGP/cosmic>) ont certaines de leurs mutations somatiques qui affectent directement l'épissage. Fait

intéressant, aucune mutation affectant un ESS n'a été répertoriée (Pettigrew and Brown, 2008) alors que ceux affectant des ESEs sont communes. Par contre, le patron d'expression d'isoformes de certains gènes dans les tissus cancéreux n'est pas forcément la cause d'une mutation directe. L'hypothèse la plus probable demeure un débalancement des facteurs affectant l'épissage en *trans*.

2.2.3.2 Dérégulation des facteurs d'épissage

En plus de la dérégulation des isoformes observée dans différents types de cancers, une autre série d'évidences vient de l'analyse des facteurs d'épissage agissant en *trans*. Une analyse bioinformatique rétrospective sur des données issues de «serial analysis of gene expression» (SAGE) et de micropuce à ADN permet de noter la surexpression de plusieurs facteurs d'épissage dans le cancer du sein, de la prostate, du côlon et du cerveau (Kirschbaum-Slager *et al.*, 2004). Parmi l'ensemble des gènes relié au cancer et comportant des mutations somatiques, au moins 19 ont des domaines de liaison à l'ARN ou un rôle au niveau de la maturation de l'ARNm, (Skotheim and Nees, 2007). En fait, toute action permettant de moduler la concentration ou l'activité de facteurs d'épissage clé au site d'action est susceptible d'affecter l'AS de ses exons cibles. Par exemple, la surexpression d'un facteur de transcription comme Myc active la transcription de gènes codant pour des facteurs d'épissage (dont PTBP1, hnRNPA1 et 2) qui à leur tour, régulent l'AS d'exon cible dans les glioblastomes (David *et al.*, 2010). De façon générale, les domaines RS des facteurs d'épissage doivent être phosphorylés afin de préserver leur localisation nucléaire (Gui *et al.*, 1994, Yeakley *et al.*, 1999, Blaustein *et al.*, 2005). L'état de phosphorylation des protéines SR est finement régulé par la «serine/arginine-rich splicing factor kinase 1» (SRPK1), elle-même surexprimée dans les cancers du sein, de la prostate et du côlon (Hayes *et al.*, 2007).

2.2.3.3 Relation de cause à effet

Bien que le dérèglement de certains facteurs d'épissage et de plusieurs isoformes soit évident, ceci ne confère en rien un lien de cause à effet entre eux, et encore moins une

évidence comme quoi l'AS participe à la progression tumorale. En réponse indirecte à cette dernière question, plusieurs gènes reliés au cancer expriment des isoformes alternativement épissés dont la fonction est en lien avec la cancérogenèse. En fait, chacun des six « Hallmark of Cancer » de Hanahan et Weinberg (Hanahan and Weinberg, 2000) est représenté par au moins un exemple dans la littérature (Fig. 16). À plus large échelle, des prédictions bioinformatiques appuient l'impact de l'AS sur la fonction protéique (Artandi and DePinho, 2010) mais les démonstrations de leur impact sur la cancérogenèse *in vivo* tardent. Reste que la démonstration ultime est d'exploiter le défaut dans l'épissage (facteur d'épissage ou isoforme) pour éliminer les cellules cancéreuses sans affecter les cellules normales.

De façon intéressante, l'AS touche chacun des six « Hallmark of Cancer » de Hanahan et Weinberg (Hanahan and Weinberg, 2000). Auto-suffisance des signaux de croissance: La famille des gènes Ras (KRAS, Harvey rat sarcoma viral oncogene homolog (HRAS), neuroblastoma RAS viral oncogene homolog (NRAS)) est la cible de plusieurs mutations, dont une dans un intron du gène HRAS qui a pour effet d'abolir un 5'ss (Cohen *et al.*, 1989). L'isoforme court de HRAS, (p21) possède des capacités accrues de prolifération comparativement à l'isoforme de pleine longueur (p19) (Cohen and Levinson, 1988). En fait, l'exon IDX présente un codon stop dans le cadre de lecture et la quantité de cette isoforme est conséquemment régulée par la voie NMD (Barbier *et al.*, 2007). Résistance aux signaux suppresseurs de croissance: Rb1 est un régulateur clé de la progression du cycle cellulaire. En état hypophosphorylé, il se lie au facteur de transcription E2F causant la répression de son activité transcriptionnelle. Un nombre important de mutations dans les sites d'épissage de Rb1 a été répertorié et corrèle avec la sévérité du phénotype. La plupart de celle-ci introduit un codon stop prématurément et diminue conséquemment le niveau d'expression de ce gatekeeper (Lohmann and Gallie, 2004). Invasion et métastase: La protéine transmembranaire CD44 subit de l'AS dans la majorité de ses exons codant, résultant en de multiples isoformes. L'isoforme incluant l'exon 6 est requise pour l'activation du récepteur Met via la formation d'un complexe ternaire avec le ligand «hepatocyte growth factor» (HGF) (Orlan-Rousseau *et al.*, 2002), une voie de signalisation impliquée dans l'invasion et la métastase (Cecchi *et al.*, 2010). De plus, un siRNA

spécifique à l'exon 5 réduit substantiellement l'invasion des cellules HeLa dans un essai *in vitro* (Cheng and Sharp, 2006). Acquisition de l'immortalité: Les cellules humaines normales possèdent un pouvoir de réplication limité. C'est qu'à chacune des rondes de réplication de l'ADN, les extrémités des chromosomes (télomères) diminuent jusqu'à un point où les cellules cessent de proliférer et entrent en sénescence (Artandi and DePinho, 2010). Le gène encodant la sous-unité catalytique de la télomérase (telomerase reverse transcriptase, TERT) est réactivé dans plusieurs types de cancers, permettant de maintenir une taille adéquate des télomères et conséquemment l'immortalité. En condition normale, les lymphocytes produisent un nombre substantiel d'isoformes du gène TERT qui régulent négativement la faible activité basale associée à son transcrit de pleine longueur. Lors de la lymphomagenèse, il a été proposé que l'AS permet de diminuer l'expression de ces transcrits inhibiteurs via la voie NMD (Amor *et al.*, 2010). Induction d'angiogenèse: Le gène VEGF est un important régulateur de l'angiogenèse. Il est épissé alternativement de façon complexe dans sa portion 3' pour générer les isoformes VEGF₁₂₁ à VEGF₂₀₆. C'est l'utilisation d'un 3'ss distal plutôt que proximal qui lui confère ses propriétés anti-angiogéniques (Harper and Bates, 2008). Résistance à l'apoptose: Le gène BCL2L1 (Bcl-x) possède deux 5'ss en compétition sur l'exon 2 pour générer deux isoformes: Bcl-x_S (isoforme court) et Bcl-x_L (isoforme long). L'expression de Bcl-x_L en lignée cellulaire prévient la mort alors que l'expression de Bcl-x_S seule est suffisante pour induire l'apoptose (Boise *et al.*, 1993). Les propriétés antagonistiques des isoformes issus de l'AS ne sont pas une exception dans la famille Bcl; plusieurs membres ont des effets similaires dont «myeloid cell leukemia sequence 1» (MCL-1_L/MCL-1_S) et B-cell CLL/lymphoma 2 (BCL-2_α/BCL-2_β) (Akgul *et al.*, 2004).

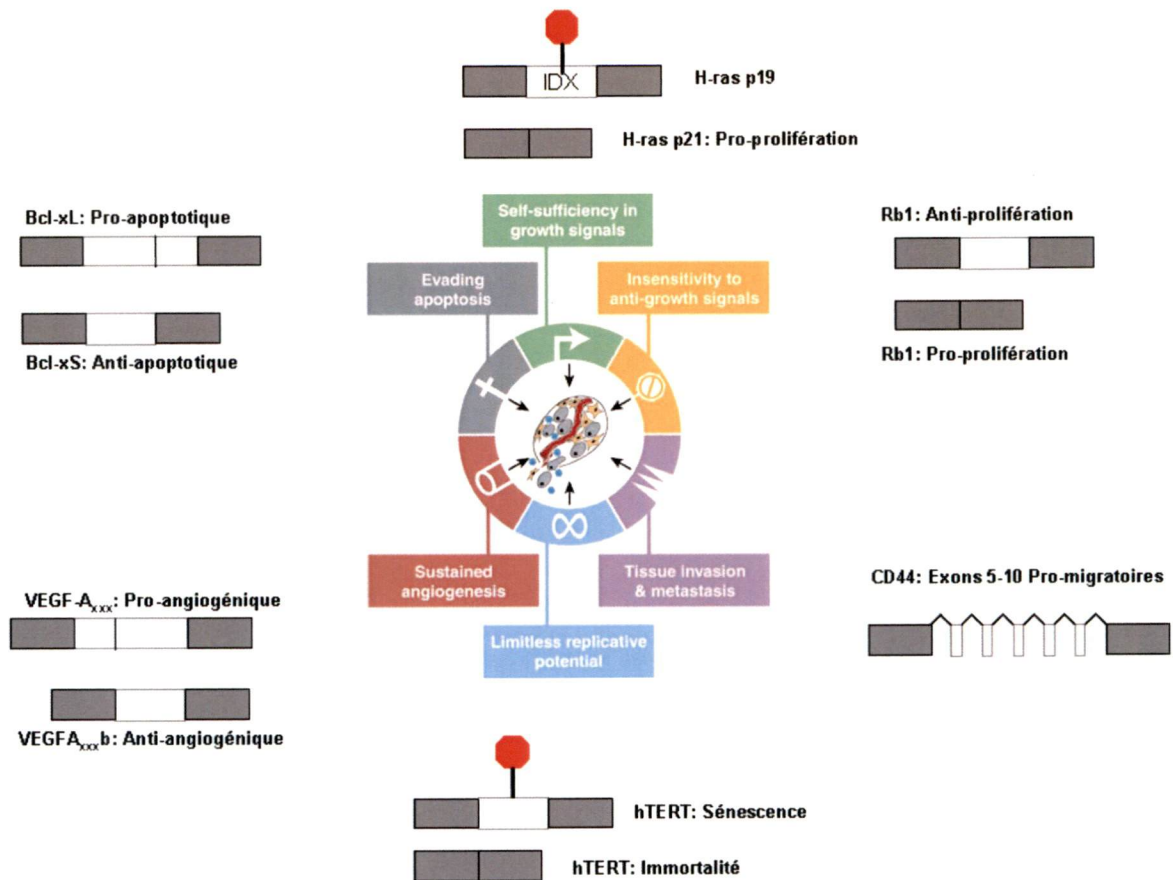


Figure 16 L'épissage alternatif est omniprésent dans les gènes liés au cancer.

Un exemple de gène dont la fonction des isoformes est en lien avec la cancérogenèse est illustré à partir du schéma des six « Hallmarks of Cancer » (Hanahan and Weinberg, 2000).

3 Quantification de l'ARNm

3.1 Méthodologies de quantification des isoformes alternativement épissés

Dans les années 1980, les isoformes issus de l'AS étaient principalement détectés par hybridation de type Northern (Cheng and Sharp, 2006). Cette technique a l'avantage de discerner les transcrits par leur poids moléculaire et permet leur quantification. Outre son faible débit, son principal défaut est relié à sa sensibilité: des quantités d'ARN total de l'ordre du microgramme sont nécessaires. Ce dernier aspect est amélioré si on applique le protocole de la protection à la ribonucléase (RNase).

La méthode classique pour la détection d'isoformes de l'AS demeure le « endpoint PCR » (Fig. 17A). C'est une méthode efficace et accessible à tous les laboratoires de biologie moléculaire. Des amorces de réaction de polymérisation en chaîne (PCR) sens et antisens sont dessinées sur les exons constitutifs flanquants l'ASE. La réaction PCR génèrera deux produits PCR, chacun d'eux correspondant à une isoforme. Par convention, on appelle isoforme long celle incluant la séquence alternative et isoforme court celle excluant la séquence alternative. Ensuite, ces deux produits sont séparés sur gel d'agarose et les bandes quantifiées au moyen de bromure d'éthidium. L'électrophorèse capillaire permet d'augmenter substantiellement le débit de cette technique (Klinck *et al.*, 2008). Le ratio des isoformes est calculé en prenant la molarité du produit long sur la somme de la molarité des produits longs et courts, ce que notre groupe de recherche a rebaptisé ψ (percent of splicing index, psi) (Klinck *et al.*, 2008). Les deux principales limites de cette technique sont reliées au pouvoir de résolution du gel et à la faible plage de valeur pour laquelle la méthode est quantitative.

Bien que les premières études de profilage par micropuce à ADN aient été rapportées au milieu des années 1990 (Schena *et al.*, 1995), les défis techniques supplémentaires qu'apporte l'étude de l'AS retardent la première publication à 2003 (Johnson *et al.*, 2003). Il n'est effectivement plus possible d'utiliser des bibliothèques d'ADN complémentaires (ADNc) de pleine longueur ou encore de courtes sondes d'oligonucléotides ciblant la

région 3' des transcrits. Une des stratégies consiste à sonder toutes les jonctions exon-exon (junction array) (Johnson *et al.*, 2003). Dans ce cas, lorsque deux jonctions consécutives donnent un signal très faible ou très fort, l'exon commun est soupçonné d'être alternatif (exclusion ou inclusion, respectivement). Une deuxième stratégie promeut l'utilisation de sondes exon-exon et de sondes exoniques (exon/junction array) (Pan *et al.*, 2004). Dans tous les cas, la haute similarité des jonctions favorise l'hybridation non-spécifique des transcrits et les méthodes d'analyse ne parviennent pas toujours à quantifier la modulation de l'AS (Cuperlovic-Culf *et al.*, 2006) (Fig. 17B).

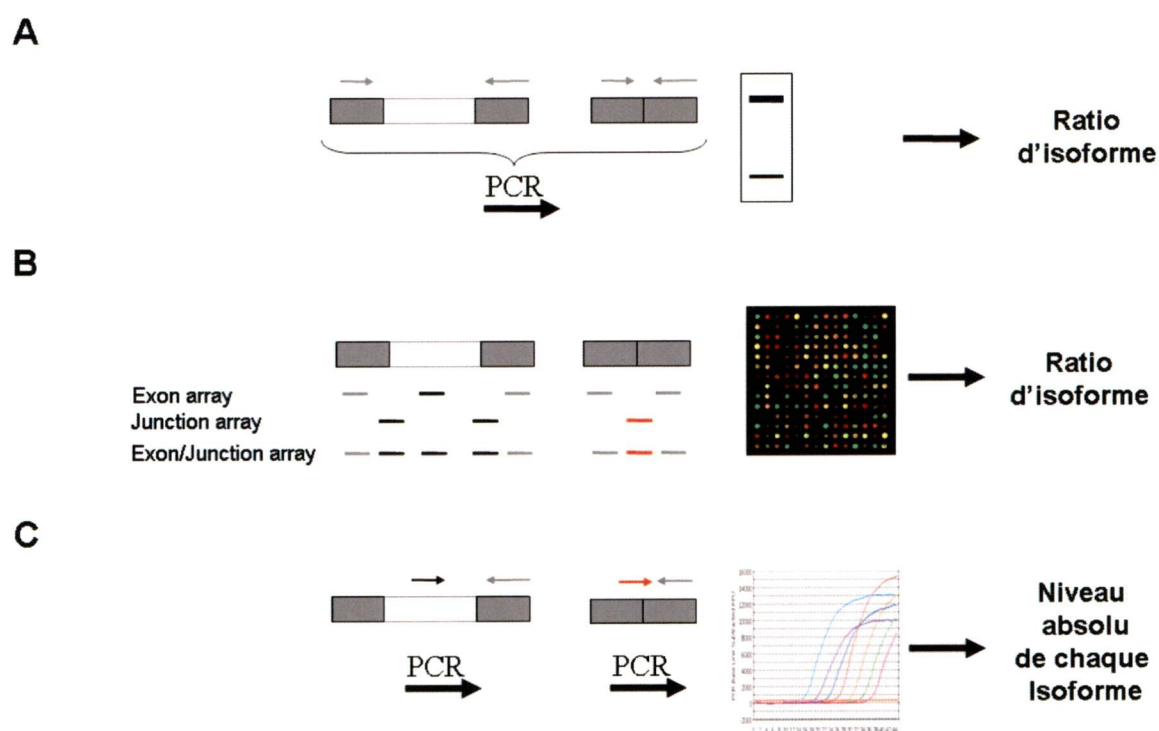


Figure 17 Stratégies de détection pour la quantification des isoformes alternativement épissés.

A) Endpoint PCR. B) Micropuce à ADN. C) PCR en temps réel.

Les trois méthodes les plus fréquemment utilisées sont le «endpoint» PCR, la PCR en temps réel et les micropuces à ADN. Une des techniques couramment utilisée pour la quantification de l'expression globale mais marginalement exploitée pour les isoformes est l'amplification par PCR en temps réel. Cette fois-ci, il s'agit de positionner les amorces de façon à générer une réaction PCR spécifique à chacune des isoformes (Fig. 17C). Pour

l'isoforme long, une des amorces est spécifique à l'ASE tandis que l'autre amorce est située sur un exon constitutif. Dans le cas de l'isoforme court, la seule région unique est la jonction exon-exon. Une des deux amorces doit obligatoirement chevaucher la jonction exon-exon, l'amorce opposée s'hybridant sur une séquence commune aux deux isoformes. Contrairement au «endpoint» PCR où la réaction est arrêtée après un certain nombre de cycles (typiquement entre 25 et 35), les produits de PCR en temps réel sont amplifiés et mesurés après chaque cycle. La quantification en temps réel est possible grâce à l'ajout d'un agent intercalant de l'ADN double-brin, typiquement le SyBr Green qui est fluorescent uniquement dans son état lié (Morrison *et al.*, 1998). Alternativement, une sonde de type Taqman™ peut être ajoutée entre les deux amorces (Morrison *et al.*, 1998, Heid *et al.*, 1996). Dans ce cas, la fluorescence provient de la dégradation de la sonde via l'activité 5' → 3' exonucléase de la Taq polymérase. Si on met en graphique l'intensité de fluorescence émise en fonction des cycles PCR, on aura une courbe présentant 4 phases : le niveau de base (typiquement dans les 15 premiers cycles), l'amplification exponentielle (3-4 cycles suivant la première phase), la phase linéaire et la phase plateau indiquant la fin de la réaction. (Fig. 17B, au centre). C'est la portion exponentielle qui est la plus reproductible et qui sert de valeur de comparaison pour la quantification. Plus précisément, c'est le point qui émerge de la ligne de base (cycle threshold, Ct). Pour des fins de normalisation, une réaction PCR amplifiant un gène de référence (housekeeping gene, HKG) ne présentant que très peu de variation entre les échantillons soumis est habituellement utilisée. L'expression relative à un échantillon non-traité est ensuite calculée selon la méthode du $\Delta\Delta Ct$ (Livak and Schmittgen, 2001). Une version améliorée de cette formule prenant en compte l'estimation de l'efficacité de chacune des paires d'amorces de PCR ainsi qu'une normalisation par moyenne géométrique de plusieurs gènes de référence est aussi de plus en plus utilisée (Hellemans *et al.*, 2007).

3.2 Utilisation d'échantillon clinique fixé

La méthode la plus simple et la plus répandue pour la conservation des tissus est de les fixer avec le formaldéhyde et de les enrouler dans la paraffine (formalin-fixed, paraffin-embedded, FFPE). L'utilisation de ces tissus pour des études d'expression génique est

longtemps demeurée controversée puisque l'ARN résultant est de très mauvaise qualité. C'est que le traitement de fixation modifie les ribonucléotides en ajoutant des groupements $\text{CH}_2\text{-OH}$ sur les bases azotées (principalement les adénines) et forme des liaisons covalentes entre l'ARN et des protéines (Masuda *et al.*, 1999). De plus l'ARN est fragmenté en segments d'au plus 100 à 200 nt. Par contre, même si les temps de «processing» (moment entre la chirurgie et l'immersion du tissu dans le formaldéhyde) et de fixation ont un impact significatif sur la qualité de l'ARN (Chung *et al.*, 2008), la normalisation par rapport à un HKG (PCR en temps réel) annule ces effets et permet une quantification adéquate (Antonov *et al.*, 2005). Pour maximiser le signal, l'étape de reverse transcription doit s'effectuer avec des amorces dégénérées ou des gènes spécifiques (pas seulement des oligonucléotides polydT) et les amplicons doivent être d'au plus 100 bp. Cette dernière contrainte restreint la technique «endpoint» PCR en FFPE à des exons d'au plus 40 nt alors que le PCR en temps réel n'est que très peu affecté par la taille des exons. En effet, la taille des deux amplicons d'une réaction «endpoint» PCR est dictée par la taille de l'ASE (Fig. 18). De plus, le dessin doit respecter la limite supérieure de 100 bp et inférieure d'au moins 60 bp (pour se différencier des amorces et dimères d'amorces), en plus de générer des amplicons résolubles par électrophorèse capillaire (> 15 bp). À noter que très peu d'exemples exploitant le SyBR Green et de l'ARN provenant de FFPE ont été rapportés (Rogerson *et al.*, 2008). La technologie de micropuce à ADN a également été utilisée avec succès, mais présentent un niveau de bruit de fond plus grand qu'avec l'utilisation de tissus frais (Waldron *et al.*, 2012).

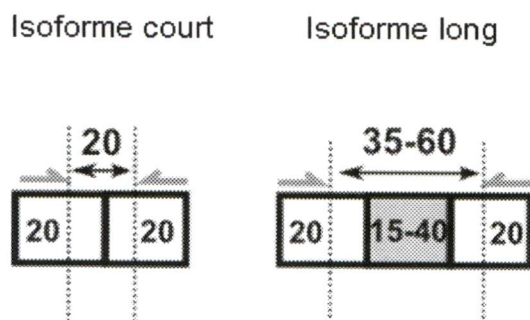


Figure 18 Contraintes de dessin d'amorces en endpoint PCR dues à l'utilisation d'échantillon clinique fixé.

Des amorces de endpoint PCR typique d'environ 20 nt amplifient un fragment PCR d'un minimum de 60 bp (isoforme court). Ceci laisse une plage de 20 nt entre les deux amorces. Comme l'amplification de l'isoforme long (amplicon d'un maximum de 100 bp) s'effectue avec les mêmes amorces PCR, cela permet l'amplification d'exon alternatif dont la taille est limitée à 40 nt ou moins. De plus, le pouvoir résolutif de l'électrophorèse capillaire exige une différence de taille d'amplicon d'au moins 15 bp.

3.3 Microdissection au laser et amplification d'ARN

Pour isoler des populations de cellules homogènes, il est possible d'exploiter des caractéristiques antigéniques, de taille ou de régularité de surface de cellules en suspension. Cette technique, le fluorescence-activating cell sorting (FACS), est difficilement transposable aux tissus solides, car les cellules sont prisonnières d'une matrice. Pour y parvenir, de fines coupes de tissus (~5 μm) sont montées sur une lame flexible et microdisséquées à l'aide d'un laser (laser capture microdissection, LCM). Cette technique a été récemment appliquée pour isoler des cellules cancéreuses de tumeur de prostate (Rogerson *et al.*, 2008), de sein (Ma *et al.*, 2003) et d'ovaire (Tone *et al.*, 2008). Comme les quantités d'ARN isolées suite au LCM sont minimales, l'ARN total doit presque obligatoirement suivre un protocole d'amplification pour en augmenter la quantité. Le protocole « Sigma whole transcriptome amplification » permet d'obtenir une amplification uniforme du transcriptome (sans biais pour la région 3' du gène) en deux étapes. D'abord des amorces dégénérées en 3' mais possédant une séquence universelle en 5' sont utilisées afin de produire une librairie de courtes séquences d'ADN. Dans un deuxième temps, cette librairie sert de matrice et est amplifiée à l'aide d'amorces de PCR dites universelles.

4 Méthodes pour reprogrammer l'épissage alternatif des cellules cancéreuses

4.1 Antisens

Les premiers effets antisens ont été remarqués à la fin des années 1970 avec de courtes séquences d'ADN s'hybridant aux séquences essentielles à la circularisation de l'ARN du virus du sarcome de Rous (Stephenson and Zamecnik, 1978). Le mécanisme proposé est l'arrêt de la traduction dû à un blocage du ribosome. À cette époque, les séquences de cibles potentielles permettant d'exploiter le mécanisme antisens sont très limitées. De plus, la synthèse d'oligonucléotides de longueur suffisante pour permettre une hybridation à 37°C demeure difficile et l'entrée des oligonucléotides en cellules est encore vivement débattue (Zamecnik, 1996). L'antisens est généralement positionné directement sur le codon d'initiation de la traduction ou dans le 5'UTR près de la structure coiffe (Crooke, 1999) (Fig. 19A). Outre l'arrêt de la traduction, il est observé que certains antisens contribuent à une dégradation de l'ARNm cible. Comme la RNase H est une enzyme ubiquitaire qui clive le brin ARN d'un duplex ADN:ARN, cela en fait un bon candidat pour expliquer les effets de dégradation-spécifique de l'ARNm (Walder and Walder, 1988) (Fig. 19B).

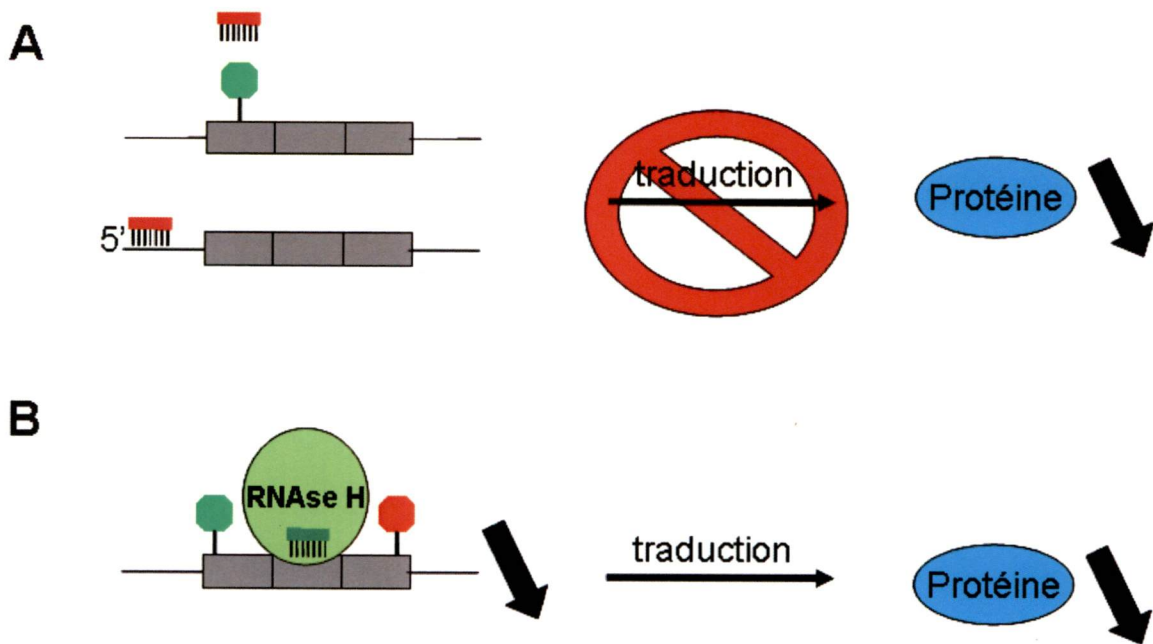


Figure 19 Mécanismes d'action des antisens.

A) Blocage de la traduction. Un antisens typique de 18 à 25 nt de longueur est positionné sur le codon d'initiation de la traduction ou près de la structure coiffe dans le 5'UTR de l'ARNm. Celui-ci interfère avec le ribosome et diminue la traduction subséquente de la protéine. À noter que l'antisens doit être modifié de façon appropriée. B) Clivage par la RNase H. Un antisens, généralement situé dans la portion codante de l'ARNm, forme un duplex ADN:ARN qui est reconnu par la RNase H. Le clivage du brin ARN et sa dégradation subséquente permet de réprimer l'expression génique et par ricochet, l'expression de la protéine correspondante.

L'ADN et particulièrement l'ARN simple brin sont sensibles aux nucléases endogènes. Fritz Eckstein fut le premier à observer l'effet de résistance aux RNases qu'offrent les nucléosides phosphorothioates (PS) (Eckstein, 1966). Comme la substitution d'un oxygène pour un soufre se prête bien au schéma de synthèse des ADNs, les oligonucléotides PS constituent la première génération d'oligonucléotides modifiés. De façon intéressante, le duplex PS-ADN:ADN est une des rares modifications qui préserve la conformation d'hélice ADN de type B, et donc la susceptibilité à la RNase H (Nakamura *et al.*, 1991). Par contre, un certain nombre d'inconvénients notamment la diminution de la température de fusion du duplex, mais plus particulièrement la liaison à des protéines de façon non-spécifique vient ternir leur réputation. La deuxième génération d'oligonucléotides fait appel à des modifications sur le 2'-OH. La modification de type 2'-O-alkyl par de gros groupements offre une forte résistance aux RNases alors que de plus petits groupes favorisent le caractère ARN (forme hélice de type A) et augmentent la force de liaison du

duplex (Manoharan, 1999). Tous ces types de modifications sont résistants à la RNase H. La modification de type 2'Ométhyl (2'OMe) demeure un choix populaire et représente un excellent compromis, balançant stabilité et affinité de liaison. Outre les modifications déjà discutées, la figure 20 présente les principales modifications explorées jusqu'à maintenant. Ces modifications concernent les groupements phosphates liant les oligonucléotides entre eux (phosphoramidate et methylphosphonate (MP)), des modifications du ribose (lock nucleic acid (LNA), morpholino, 2'Fluoro arabinose (FANA)) ou une substitution complète du squelette ribose phosphate (peptide nucleic acid). Comme la RNase H est en mesure de cliver un duplex aussi petit que 4 bp (Donis-Keller, 1979), les extrémités de l'antisens peuvent être constituées de nucléotides modifiés (excepté PS) alors que la région centrale doit être laissée sans modifications (excepté PS). C'est ce qu'on appelle des antisens de type « gapmer » (Leech *et al.*, 2000). Le positionnement sur l'ARNm et la chimie de l'oligonucléotide sont donc les deux facteurs les plus influents et dirigent le mécanisme d'action de l'antisens. L'âge d'or des antisens culmine vers la fin des années 1990 et coïncide avec l'approbation par la Federal Drug Administration (FDA) de VitraveneTM, un antisens traitant la rétinite induite par le cytomégalo virus. Depuis ce temps, plusieurs antisens ont été analysés en phase clinique, mais leur approbation par la FDA tarde (Rayburn and Zhang, 2008). Un grand nombre d'entre eux sont destinés à réduire l'ARNm de gènes impliqués dans la cancérogenèse (TP53, HRAS, BRAF, VEGF, BCL2) (Rayburn and Zhang, 2008). Les nouvelles technologies ARN comme les small interfering RNAs (siRNAs) et plus récemment encore les microRNAs (miRNAs) ont contribué à diminuer l'intérêt envers les antisens.

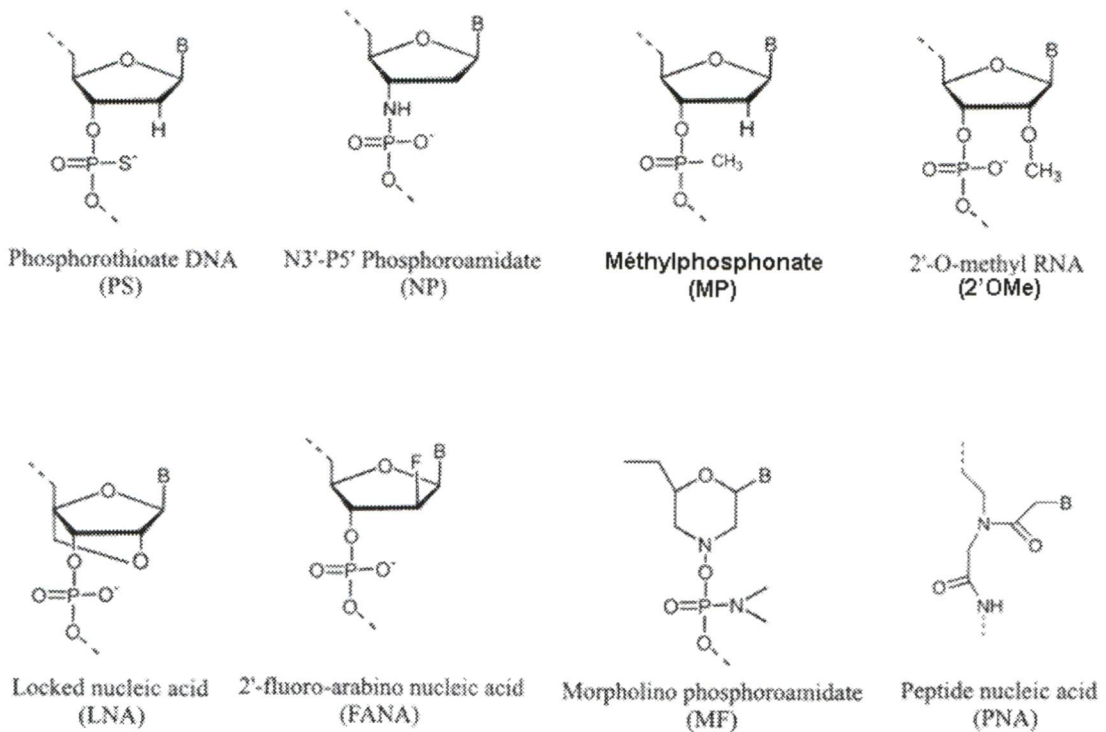


Figure 20 Structure des principaux analogues de nucléosides.

Des modifications impliquant la substitution d'un oxygène pour un soufre (phosphorothioate), une amine (phosphoroamidate) ou un groupement méthyl (méthylphosphonate) sur le groupement phosphate sont couramment employées pour augmenter la résistance aux nucléases. Certaines modifications légères du ribose telles que la protection du 2'OH par un méthyl (2'OMe) ou sa version cyclique pontant permettent de stabiliser la conformation C3'-endo, favorable à la formation d'hélice de type A. D'autres modifications telles l'utilisation du sucre arabinose (ANA, arabino nucleic acid), d'une morpholine (morpholino) ou plus drastiquement la substitution du squelette phosphoribose par un squelette peptidique (PNA, peptide nucleic acid) sont également employées. Adaptée de Kurreck (2003).

4.2 Modulateur artificiel de l'épissage alternatif

Classiquement, l'AS est étudié *in vitro* à l'aide d'extraits nucléaires. Cet extrait fournit les facteurs d'épissage essentiels et peut être déplété ou enrichi d'un ou plusieurs facteurs d'épissage afin de mettre en valeur leurs rôles respectifs. Un segment du gène à l'étude (appelé mini-gène : généralement deux 5'ss en compétition, un intron raccourci et un 3'ss unique) est inséré dans un plasmide. Pour étudier le rôle des SRE, des portions exoniques et introniques sont systématiquement omises, remplacées par d'autres SRE ou des séquences contrôles. Les différentes constructions plasmidiques sont ensuite transcrites *in vitro* et

mises en présence de l'extrait nucléaire. Ainsi, plusieurs ESE et ESS ont pu être identifiés et caractérisés dont les ASEs modèles de l'exon IIIb/c s du « fibroblast growth factor receptor 2 » (FGFR2) (Seth *et al.*, 2008), exon EDA et EDB de FN1 (Muro *et al.*, 1999) et l'alt5' de BCL2L1 (Garneau *et al.*, 2005). En intégrant une séquence promotrice au lieu d'une séquence de transcription *in vitro*, il est possible d'exprimer le transcrit en cellules. De plus, il est possible de co-transfecter à la fois le mini-gène et des plasmides codant pour des facteurs d'épissage d'intérêts et/ou des siRNAs. Par contre, le mini-gène n'est qu'une version abrégée et il est loin de représenter l'ensemble des séquences régulatrices qui se retrouvent parfois à plusieurs kb du site d'épissage.

Les antisens peuvent aussi agir directement sur le pré-ARNm endogène. L'utilisation de lipides cationiques comme agent de transfection cellulaire permet leur accumulation nucléaire (Bennett *et al.*, 1992). La β -thalassémie, une maladie génétique touchant le gène de la β -globine, entraîne un défaut de l'hémoglobine et se traduit par une anémie. De façon intéressante, la majorité des mutations dans l'intron 2 de la β -globine exposent un 3'ss cryptique et produisent un nouveau 5'ss, favorisant l'inclusion d'un nouvel exon. Des travaux ont démontré les capacités des antisens PS-2'OMe ciblant spécifiquement ces transcrits mutés à rediriger l'AS vers les sites originaux (Dominski and Kole, 1993, Lacerra *et al.*, 2000). Les antisens exploitant ce mécanisme ont été rebaptisés SSO pour «splice-switching oligonucleotides».

Les maladies neuromusculaires sont particulièrement touchées par des défauts de l'épissage (Wood *et al.*, 2010). Parmi celles-ci, l'atrophie musculaire spinale (spinal muscular atrophy, SMA) et la dystrophie musculaire de Duchenne (DMD) sont les exemples les plus étudiés. Dans le cas sévère de la DMD, ce sont des mutations non-sens causant la perte totale de la protéine dystrophine qui sont en cause alors que la dystrophie musculaire de Becker, forme plus légère de la maladie, génère des protéines tronquées retenant partiellement la fonction (Muntoni *et al.*, 2003). Ceci suggère que les patients atteints de DMD bénéficieraient d'une stratégie favorisant l'exclusion d'exon permettant de restaurer le cadre de lecture. En fait, la preuve de concept a été établie chez la souris et deux SSO sont présentement en étude clinique chez l'humain (Lu *et al.*, 2011). Dans le cas de la

SMA, le gène SMN1 est normalement responsable de la production de la protéine SMN. La sévérité de l'atrophie musculaire spinale (spinal muscular atrophy, SMA) corrèle avec le niveau protéique de SMN. Théoriquement, le gène SMN2 pourrait substituer SMN1 puisqu'il est très similaire, mais une mutation d'un C pour un T dans l'exon alternatif 7 produit un nouveau stop codon dans l'exon 8. L'ARNm résultant est ciblé par la voie NMD et la quantité de protéine SMN est nettement diminuée. La stratégie consiste donc à cibler l'exon 7 du gène SMN2 de façon à promouvoir son inclusion et ainsi restituer des taux normaux de SMN.

Ces dernières stratégies ont demandé l'élaboration de règles pour le positionnement des SSOs. Afin de promouvoir l'exclusion de l'exon, le SSO peut être situé à cheval sur la jonction intron-exon (Fig. 21A) ou exon-intron (Fig. 21B) de l'exon alternatif et interférer avec l'assemblage de U2 snRNP ou U1 snRNP, respectivement. Il peut également cibler une région complètement exonique (Fig. 21C) ou complètement intronique (Fig. 21D). Il est alors postulé que le SSO masque un site ESE ou ISE, respectivement et interfère avec des facteurs d'épissage. L'inverse est également possible, à savoir le blocage d'un ESS (Fig. 21E) ou d'un ISS (Fig. 21F). Par contre, la prédiction bioinformatique des SREs demeure difficile à ce jour. De plus, leur réelle contribution dans les effets de modulation des SSOs est parfois difficile à dissocier de paramètres intrinsèques comme la longueur du SSO et la température de fusion du duplex. Alternativement, il est possible que l'influence du SSO se fasse sentir via le remodelage de la structure secondaire de l'ARNm (Aartsma-Rus *et al.*, 2002). Bref, il est coutume de tester plusieurs SSOs afin d'en trouver un qui module efficacement l'AS (Aartsma-Rus *et al.*, 2005).

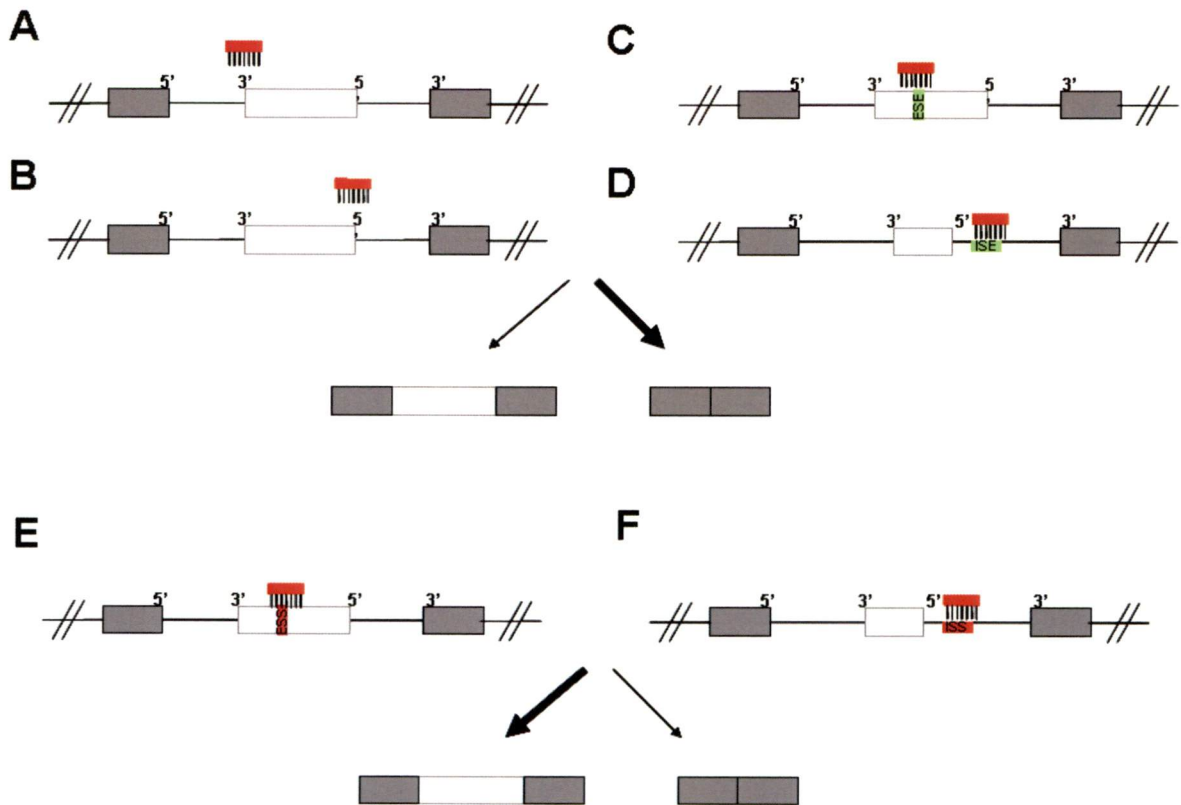


Figure 21 Mécanismes d'action des SSOs.

Afin de favoriser l'exclusion de l'exon, un SSO est positionné dans le but de masquer: A) une jonction intron-exon, B) une jonction exon-intron, C) un ESE ou D) un ISE. Afin de favoriser l'inclusion de l'exon, un SSO est positionné dans le but de masquer un E) ESS ou F) ISS.

Conceptuellement, la stratégie du «targeted oligonucleotide silencer of splicing» (TOSS) est attrayante. Il s'agit de promouvoir l'exclusion de l'exon alternatif en exhibant un ESS de façon artificielle en s'appuyant sur la spécificité antisens (Villemaire *et al.*, 2003). Le TOSS est donc un oligonucléotide bifonctionnel possédant une portion antisens complémentaire à l'exon alternatif ciblé et une portion non-hybridante reconnaissant la protéine hnRNPA1 (Fig. 22). Plus précisément, il s'agit de deux séquences de haute affinité pour le facteur d'épissage ubiquitaire hnRNPA1 (Burd and Dreyfuss, 1994). La preuve de concept a été réalisée en ciblant le pré-ARNm endogène de BCL2L1 dans plusieurs lignées cellulaires cancéreuses humaines. La généralité de ce concept (cible, type d'ASE) n'est toutefois pas connu

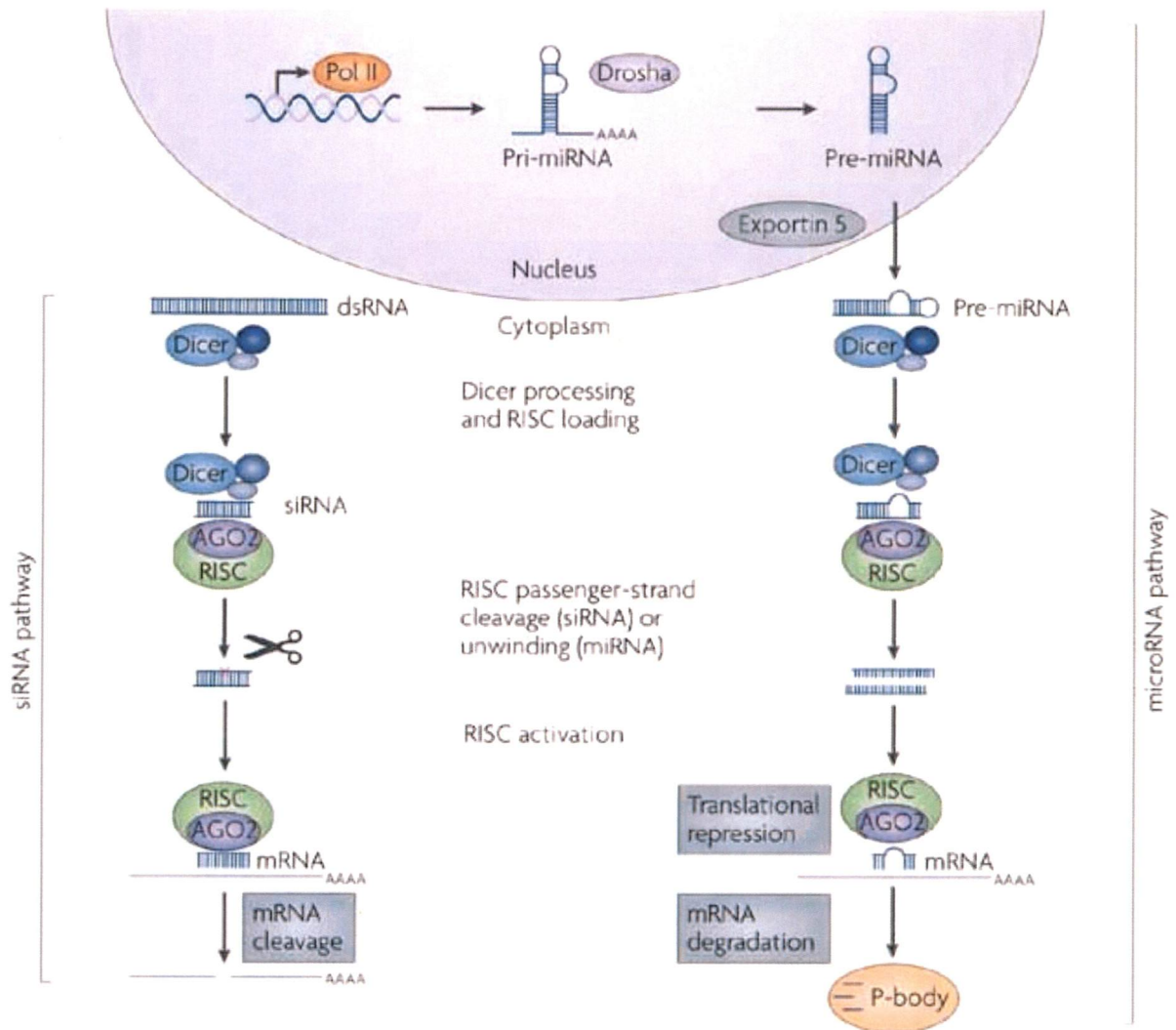


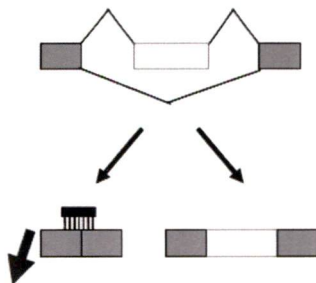
Figure 23 Mécanisme d'action des siRNAs.

Les siRNAs sont des petits ARNs double-brins d'environ 20 bp dont les extrémités flanquantes en 3' sont de 2 nt. Les siRNAs exogènes empruntent la voie des micro-ARNs (miRNAs) pour dégrader spécifiquement des ARNm au cytoplasme. Les siRNAs sont incorporés au complexe RISC. La complémentarité parfaite du duplex ARN:ARN est reconnue et AGO2 clive le brin sens de telle sorte que le brin antisens demeure dans le complexe RISC. Celui-ci s'hybride alors sur sa cible et dirige le clivage de l'ARNm cible. Le RISC relâché peut à nouveau cliver un ARNm et ainsi permettre la répression efficace du gène. Tirée de (de Fougerolles *et al.*, 2007).

Les règles de dessin de siRNA ciblant un exon alternatif sont les mêmes que ceux ciblant un exon constitutif, excepté que la séquence est restreinte. La problématique entoure plutôt le dessin de siRNA spécifique à l'isoforme court et s'apparente à la problématique associée au dessin d'amorce-jonction: la seule séquence propice au dessin concerne la jonction

exon-exon unique à l'isoforme court (Fig. 24A). Par contre, les requis conférant la spécificité de la Taq polymérase et du complexe RISC diffèrent largement, affectant conséquemment les règles de dessin. En général, la fenêtre de dessin d'un siRNA est restreinte à la séquence comprise dans le cadre de lecture en omettant les jonctions, soit de ~ 200 à 10 000 nt (Birmingham *et al.*, 2007). Des études ayant pour but de concevoir des règles de dessin de siRNAs discriminant des séquences ne différant que par 1 nucléotide ont révélé l'importance de l'intégrité d'une hélice ARN de type A pour le clivage de l'ARNm cible. Ainsi les mésappariements les plus susceptibles de la distordre, les mésappariements purine:purine, sont effectivement les plus discriminants (Huang *et al.*, 2009, Schwarz *et al.*, 2006). Outre l'identité de la paire de bases en mésappariement, la position de celle-ci joue également un rôle clé. Un mésappariement impliquant l'une ou l'autre des deux bases au site de clivage (position 10 et 11), la portion 3' du brin guide ou sa portion 5' (seed) réduit drastiquement, de façon importante ou intermédiaire l'effet de répression, respectivement (Fig. 24B). De plus, l'étude systématique de toutes les combinaisons de double mésappariement pour un siRNA donné indique que 35% des siRNAs procurent tout de même une répression de plus de 50% (Dahlgren *et al.*, 2008). En résumé, un seul mésappariement n'est pas suffisant pour empêcher l'action du RISC, à moins que celui-ci soit en position 10 ou 11.

A



B

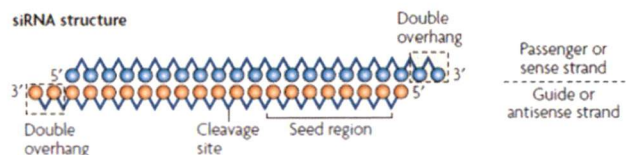


Figure 24 Utilisation de siRNAs spécifiques à l'isoforme court (siRNA court).

A) Cibler la jonction exon-exon unique à l'isoforme court avec un siRNA. B) Anatomie d'un siRNA: Les siRNAs sont des petits ARN double-brins de 19 bp. Chaque brin comporte une portion flanquante de 2 nt en 3'. La numérotation de la position se fait du 5' (position 1) vers le 3' (position 19) sur le brin antisens. La région « seed » s'étend de la position 2 à 8 et confère la spécificité du siRNA. Les nucléotides 10 et 11 aux abords du site de clivage doivent absolument être pairés pour que le clivage ait lieu.

Tout comme les antisens classiques (dépendants de la RNase H), une des applications intéressantes des siRNAs demeure la possibilité de cibler théoriquement n'importe quel ARNm et de mettre de l'avant des études fonctionnelles reliant la répression d'un gène et sa fonction. Théoriquement, il est possible de disséquer la fonction potentielle de chacune des isoformes d'un gène. Toutefois dans la majorité des études, le siRNA cible l'ensemble des transcrits, ne permettant pas de révéler les potentielles caractéristiques propres à chaque isoforme. Lorsqu'une seule des isoformes est ciblée, elle n'est que très rarement comparée avec la répression spécifique d'une autre isoforme du même gène ou encore ciblant globalement tous les transcrits du gène. Les modulateurs d'AS offrent la possibilité d'aller plus loin en explorant la contribution d'un changement de ratio d'isoformes indépendamment de l'impact sur l'expression globale.

OBJECTIFS DE CETTE THÈSE DE DOCTORAT

La cancérogenèse modifie l'expression et la maturation des ARNm, qui sont normalement finement régulés par divers mécanismes, y compris l'épissage alternatif. Bien que plusieurs exemples ponctuels de dérégulation d'isoforme de l'épissage aient été répertoriés, l'ampleur de cette dérégulation et la fonction associée à ces isoformes demeure mal comprise à ce jour. Nous émettons donc l'hypothèse selon laquelle certaines isoformes de l'épissage alternatif ont des fonctions clés dans le processus de cancérogenèse. En exploitant des échantillons humains provenant de tumeurs ovariennes et des répressions transitoires par RNAi dans des lignées cancéreuses ovariennes, nous avons cherché à comprendre l'impact de l'épissage alternatif dans le processus de cancérogenèse.

Pour déterminer le rôle des isoformes de l'épissage alternatif dans le cancer de l'ovaire, nous avons d'abord développé des méthodologies générales et à haut débit permettant de 1) quantifier les isoformes de l'épissage dans les tissus fixés et de 2) reprogrammer l'épissage de gènes endogènes *in cellulo*. Puis, nous avons exploité ces méthodologies afin de 3) déterminer la fonction anti-apoptotique d'isoformes associées au cancer de l'ovaire et de 4) identifier un programme d'épissage dans le microenvironnement des tumeurs ovariennes.

MÉTHODOLOGIES ET RÉSULTATS

Chapitre 1

High-throughput quantification of splicing isoforms

Auteurs de l'article: **Brosseau JP**, Lucier JF, Lapointe E, Durand M, Gendron D, Gervais-Bird J, Tremblay K, Perreault JP et Elela SA.

Statut de l'article: Publié, RNA2010;16(2):442-9

Avant-propos

Figure 1. JPB et JFL ont élaboré la figure A à D. JFL a produit les résultats en E

Figure 2. JPB et DG ont effectué les expériences PCR. JPB a analysé les résultats. JPB et JFL ont produit la figure

Figure 3. JFL (95%) et JGB (5%) ont conçu l'algorithme de prédiction RASE. JPB et JFL a dessiné la figure.

Table 1. JPB et MD ont effectué les expériences PCR. JPB a analysé les résultats.

Table 2. JFL a produit les résultats

Table 3. JPB et EL ont effectué les expériences PCR. JFL a produit un module d'analyse pour valider les amorces. JPB a analysé les résultats.

Matériel supplémentaire :

Interface web : JFL a produit un interface web pour le dessin d'amorce selon RASE.

Fig S1. JPB a effectué les expériences PCR. JPB et KT ont analysé les résultats. JPB et JFL ont produit la figure.

Fig S2. JFL et JPB ont produit et analysé les résultats.

Table S1. JPB et DG ont effectué les expériences PCR. JPB a analysé les résultats. JPB a produit la table.

Table S2. JPB et MD ont effectué les expériences PCR. JFL a fourni des résultats. JPB a produit la table.

Table S3. JPB a produit la table.

Table S4. JPB et JFL ont produit la table

Supervision : KT, JPP et SAE

Ecriture : JPB a écrit le brouillon. SAE, JPB, JPP et JFL ont participé à la révision de toutes les version subséquentes du manuscrit.

Mise en contexte

La cancérogenèse modifie l'expression des ARNm et nous émettons l'hypothèse selon laquelle certaines isoformes de l'épissage alternatif ont des fonctions clés dans le processus de cancérogenèse. Bien que des exemples ponctuels supportant la modification de l'expression de certaines isoformes dans les tumeurs solides ont été rapportés, il est présentement difficile de connaître l'impact de l'épissage alternatif à l'échelle du génome puisque leur quantification est laborieuse par les techniques actuelles. Pour y remédier, nous avons réussi à établir une méthodologie à haut débit basé sur le PCR quantitatif nous permettant d'amplifier spécifiquement les isoformes de l'ensemble des évènements d'épissage de la base de donnée NCBI RefSeq 36, soit 3323 évènements.

Les travaux présentés dans ce chapitre ont pour but de démontrer que 1) Le design rationnel d'amorce PCR isoforme-spécifique est possible en exploitant une condition PCR universelle; 2) la détection isoforme-spécifique est possible même en utilisant de l'ARN total provenant de tissus fixés.

Résumé

La plupart des ARNs messagers sont épissés alternativement et plusieurs de ces isoformes sont reliés à des pathologies humaines. Par contre, la contribution de ces évènements d'épissage alternatif au développement de conditions pathologiques demeure incertaine. Comme l'apport de l'épissage alternatif au diagnostic et au pronostic devient de plus en plus reconnu, il est d'autant plus pertinent de développer des méthodologies rigoureuses afin de quantifier les isoformes de l'épissage à partir d'échantillons cliniques. Dans cet ouvrage, nous présentons la procédure RASE (Real-time PCR Annotation of Splicing Events) qui permet l'identification d'un grand nombre d'évènements d'épissage dans les tissus. RASE est en mesure de dessiner des paires d'amorces PCR pour 81% de l'ensemble des évènements d'épissage alternatif de la base de données de NCBI 36. Effectivement, la plupart de ces amorces PCR permettent une amplification spécifique à partir d'ARN

provenant de lignées cellulaires ou de tissus fixés par le formaldéhyde. Avec cette méthodologie, il est maintenant possible d'identifier rapidement le patron d'épissage alternatif de tissus ou de valider les résultats issus d'autres méthodes telles que le séquençage à haut débit ou les micropuces à ADN.

Article 1: High-throughput quantification of splicing isoforms

Abstract

Most human messenger RNAs (mRNAs) are alternatively spliced and many exhibit disease-specific splicing patterns. However, the contribution of most splicing events to the development and maintenance of human diseases remains unclear. As the contribution of alternative splicing events to diagnosis and prognosis is becoming increasingly recognized, it becomes important to develop precise methods to quantify the abundance of these isoforms in clinical samples. Here we present a pipeline for Real-time PCR Annotation of Splicing Events (RASE) that allows accurate identification of a large number of splicing isoforms in human tissues. The RASE automatically designed specific primer pairs for 81 % of all alternative splicing events in the NCBI build 36 database. Experimentally, the majority of the RASE designed primers resulted in isoform-specific amplification suitable for quantification in human cell lines or in formalin-fixed, paraffin-embedded (FFPE) RNA extract. Using this pipeline it is now possible to rapidly identify splicing isoform signatures in different types of human tissues or to validate complete sets of data generated by microarray expression profiling and deep sequencing techniques.

Introduction

In eukaryotes, increased protein diversity is achieved by the alternative splicing of pre-mRNA whereby a single gene can generate several splicing isoforms coding for proteins with different functions (Fig. 1A). Nearly all multi-exon human genes are believed to be alternatively spliced and point mutations in splice sites are responsible for at least 15 % of all annotated genetic disease-causing mutations (Venables, 2006; Wang et al., 2008). Studies of alternatively spliced isoforms in vitro suggested that splicing modifies gene function and changes in splicing pattern have been linked to specific diseases (Tazi et al., 2009; Venables, 2006). However, understanding the function and association of splicing isoforms with cellular phenotypes and pathology is limited by the logistics of detecting tissue or disease-specific splicing patterns (Nagao et al., 2005; Ni et al., 2007; Pan et al., 2008). We have recently demonstrated widespread alternative splicing changes in rare frozen tumor samples with high enough RNA quality for end-point PCR characterization (Klinck et al., 2008; Venables et al., 2008). However due to degradation and sample preservation induced chemical modifications (i.e. formalin-fixed, paraffin-embedded (FFPE) storage), there is a point at which the average length of intact extracted mRNAs become too short for this type of end-point analysis; most clinically defined samples are archived in FFPE blocks (Farragher et al., 2008; Koch et al., 2006; Oberli et al., 2008). In theory, real-time PCR could circumvent this problem as amplicon length is not dependent on the size of the alternative exon.

Difficulties in producing high specificity splice-sensitive primers to quantify the expression of splicing isoforms using real-time PCR have severely limited the quality and feasibility of genome-wide data sets. The specificity of a PCR reaction depends on accurate identification of a sequence that can distinguish between different splicing isoforms. In real-time PCR reactions using the intercalating SYBR Green dye as an indicator for DNA amplification (Morrison et al., 1998) (a method favored because of its relatively low cost), the long splicing isoform is specifically amplified by using one primer in the alternative exon and an opposing primer in a constitutive exon (Fig. 1B).

On the other hand, the amplification of the short splicing isoform is achieved by a boundary-spanning primer (BSP) (Yang and Le, 1994) hybridizing to the sequence encompassing the exon-exon junction with the opposing primer in a constitutive exon (Fig. 1C). Theoretically, the BSP strategy should provide unbiased amplification of short splicing isoforms (Yang and Le, 1994; Vandenbroucke et al., 2001). In practice, achieving isoform-specific amplification using the BSP strategy is challenging (Walton et al., 2007). Mispriming due to perfect matching between either the 5' or 3' end of the BSP to the long splicing isoform often results in erroneous amplification and thus reduces the fidelity of specific splicing isoforms detection (Fig. 1D). This is particularly important if the amplicons generated by mispriming cannot be separated from the correctly primed product by gel electrophoresis (see Fig. 1C and 1D MP 5'). So far, no general rule for the design of BSPs has been established (Bracco et al., 2006; Walton et al., 2007; Williams et al., 1999) and experimentally testing the fidelity of BSPs using interference tests (Vandenbroucke et al., 2001) (amplification from mixed templates containing a fixed amount of the long isoform with serially diluted amounts of the short isoform) or the Wellman method (Wellmann et al., 2001) (amplification of fixed amount of long isoform template) is a costly and time-consuming process. Experimental validation of PCR primers is particularly unsuitable for high-throughput quantitative analysis since each primer set has to be tested and optimized individually.

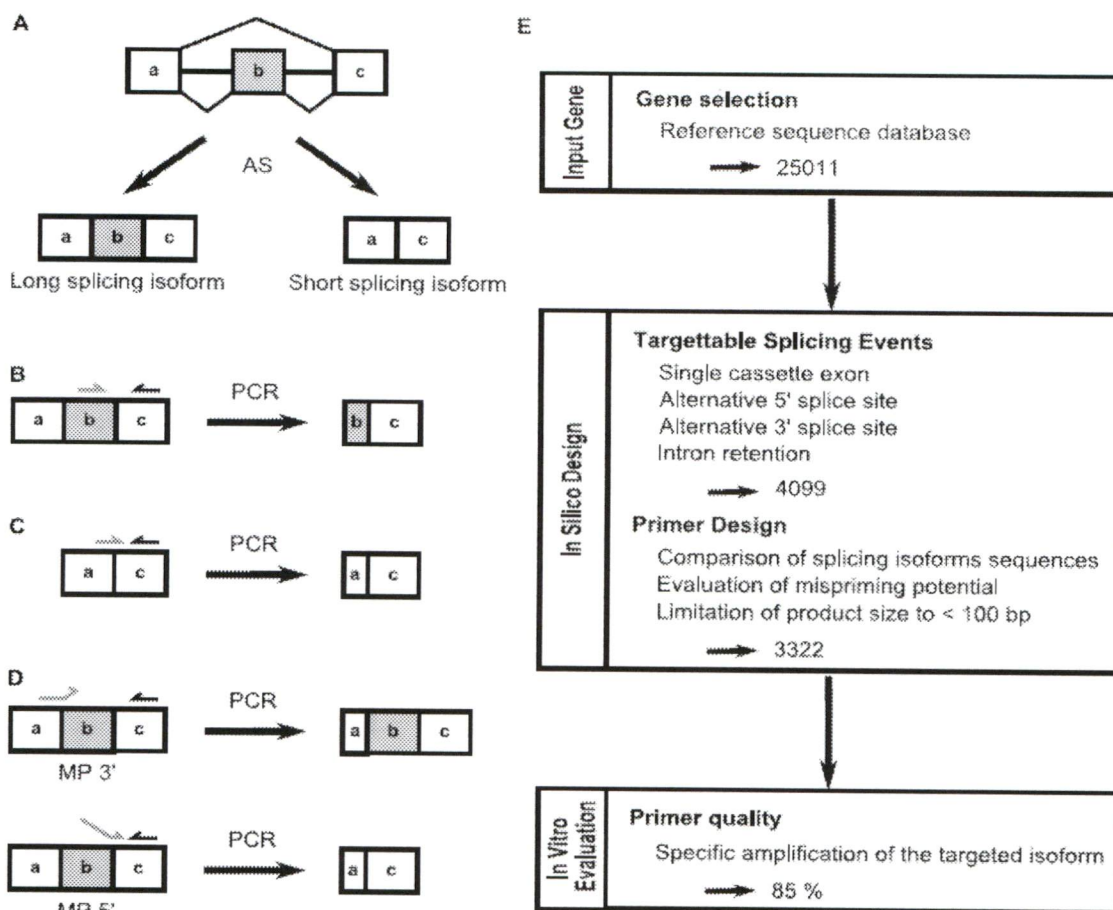


Figure 1 Problems associated with the detection of alternative splicing isoforms.

(A) Alternative splicing of a single cassette exon (gray box) generates two isoforms differing by one exon. (B) Strategy for specific amplification of the long splicing isoform. The cDNA is amplified using primers specific to the alternative (gray arrow) and the constitutive (black arrow) exons. (C) Strategy for specific amplification of the short splicing isoform. The cDNA is amplified using a boundary-spanning primer (BSP, gray arrow) and an opposing primer hybridizing to a constitutive exon (black arrow). (D) Mispriming is often caused by lack of complementarities between the alternative exon and either the 3' (MP 3') or 5' (MP 5') ends of the BSP on the long splicing isoform. (E) Summary of the RASE pipeline. In this study genes were selected based on the evidence of alternative splicing (NCBI 36.2 database) and the nature of the splicing events.

In this study we defined the primer design rules necessary for the specific amplification of splicing isoforms and integrated them into a pipeline for Real-time PCR Annotation of Splicing Events (RASE). The RASE allows rapid and accurate quantification of a large number of splicing isoforms in both human cell lines and FFPE tissues. As summarize in figure 1E, RASE successfully designed 3322 isoform-specific primer sets, representing 81 % of simple alternative splicing events of the NCBI human genome database (build 36)

(Pruitt et al., 2007) Our freely accessible web-interface provides access to RASE and allows automatic definition of target splice sites and the identification of pre-optimized primers that function under universal PCR conditions <http://designs.lgfus.ca>. The design process was validated experimentally using 51 alternative splicing events and 254 different primer sets

Results

Evaluation of essential BSP design parameters

In order to determine the key design parameters that define successful splicing isoform discrimination, we first analyzed the BSP features contributing to mispriming due to its 3' end complementary to both splicing isoforms. A training set of BSPs with systematically increasing 3' end portions (3 to 15 nucleotides) perfectly matching both isoforms were designed and tested experimentally on four different alternative splicing events (Fig. 2 and Suppl. Table 1). Briefly, the quantification of the mispriming potential of BSPs was done by looking at the capacity of a BSP to amplify a template containing the putative mispriming site (long isoform-specific splice site), compared to a second reaction with a perfectly matching primer set (long-specific primer pair, LSPP) as a reference. Using this test, some BSPs showed amplification in the presence of the long splicing isoform template. However, these unspecific amplifications due to mispriming were detected very late (high C_t) when compared to values obtained with the long-specific primer pair. We hypothesized that this low amplification level might be induced by the unnaturally high amount of template included in these artificial reactions. Therefore, we conducted tests to further evaluate BSP specificities in typical real-time PCR experiments. The results indicated that BSPs resulting in a $\Delta C_t > 16$ can accurately discriminate between the different splicing isoforms (Suppl. Fig. 1). Using this strict cut-off limit, we found that 7-8 consecutive base pairs ($T_m = 6^\circ\text{C}$) is the maximal number tolerated at the primer's 3' end to keep mispriming to a minimum level (Suppl. Fig. 2). We generalize this rule design by testing a BSP validation set targeting 36 different short splicing isoforms carrying a 3' end portion with a T_m lower than 6°C . All BSPs succeeded in giving specific amplification (Suppl. Table 2).

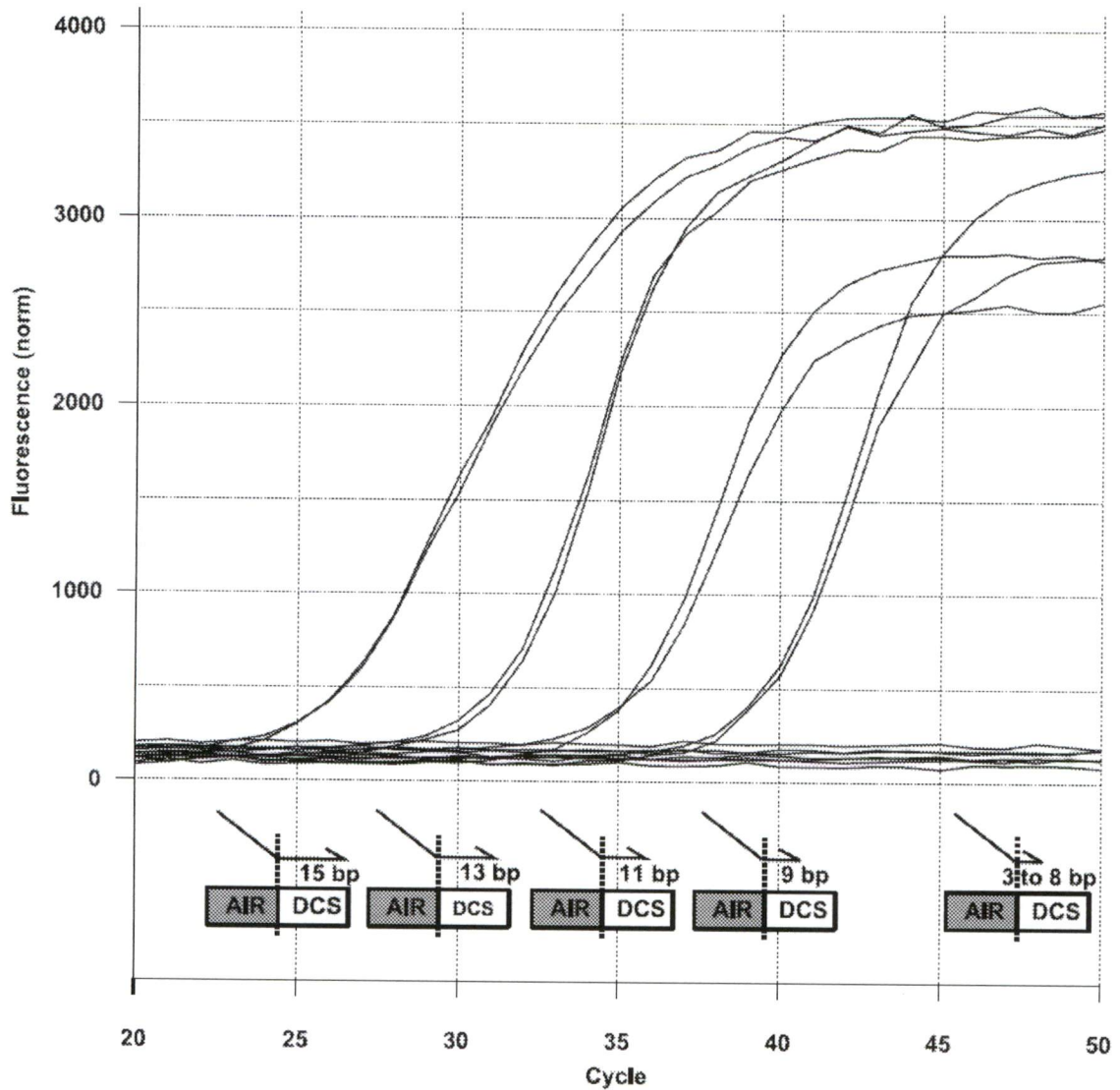


Figure 2 Systematic tiling of BSPs across an exon-exon junction.

BSPs harboring 15, 13, 11 and 9 bp complementarity to BCAS1 long splicing isoform in their 3' end show amplification curves while those with 8, 7, 5, 3 bp or no template controls show no amplification. AIR and DCS stands for alternatively included region and downstream constitutive sequence, respectively.

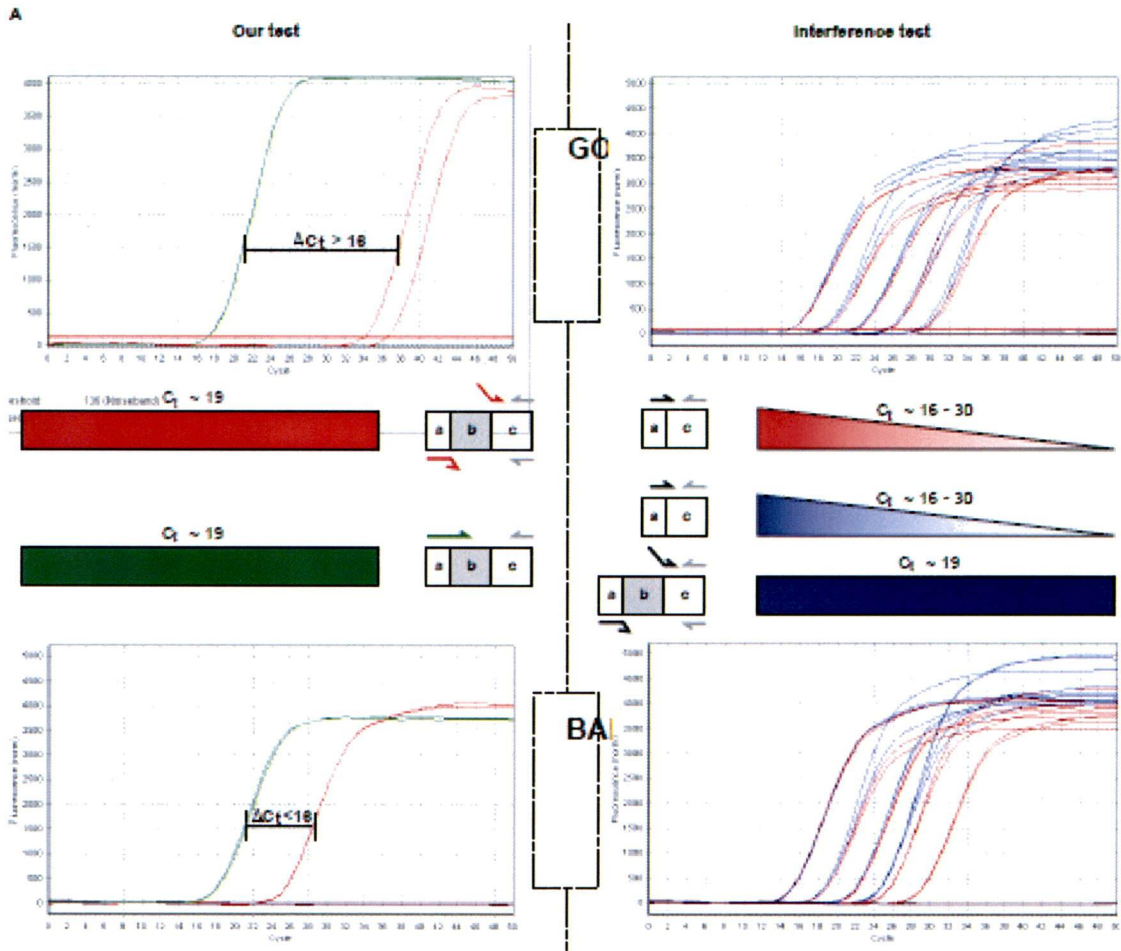


Figure S1 BSP specificity in typical qRT-PCR experiment

(A, left part) The test consists in a real-time PCR reaction containing the putative short-specific primer pair (SSPP) and an excess of the long splicing isoform amplicon (red curves). This reaction is compared to a second reaction measuring the long splicing isoform amplicon input (green curves). Ideally, C_t values from red curves should be infinite, demonstrating complete specificity. (A, right part) The interference test consists of real-time PCR reactions on two serially diluted short isoform template: One in the absence (red curves) and one in the presence (blue curves) of a fixed amount of long splicing isoform amplicon corresponding to a C_t value of ~ 19 . Ideally, amplification efficiencies should be identical, demonstrating complete specificity. (B) Summary of the BSPs specificity by our test and the interference test. BSPs were considered specific when they produced a ΔC_t of 16 or more in both reactions in our test. Note that ‘GOOD’ and ‘BAD’ examples corresponds to NRG1 and PTPRB, respectively.

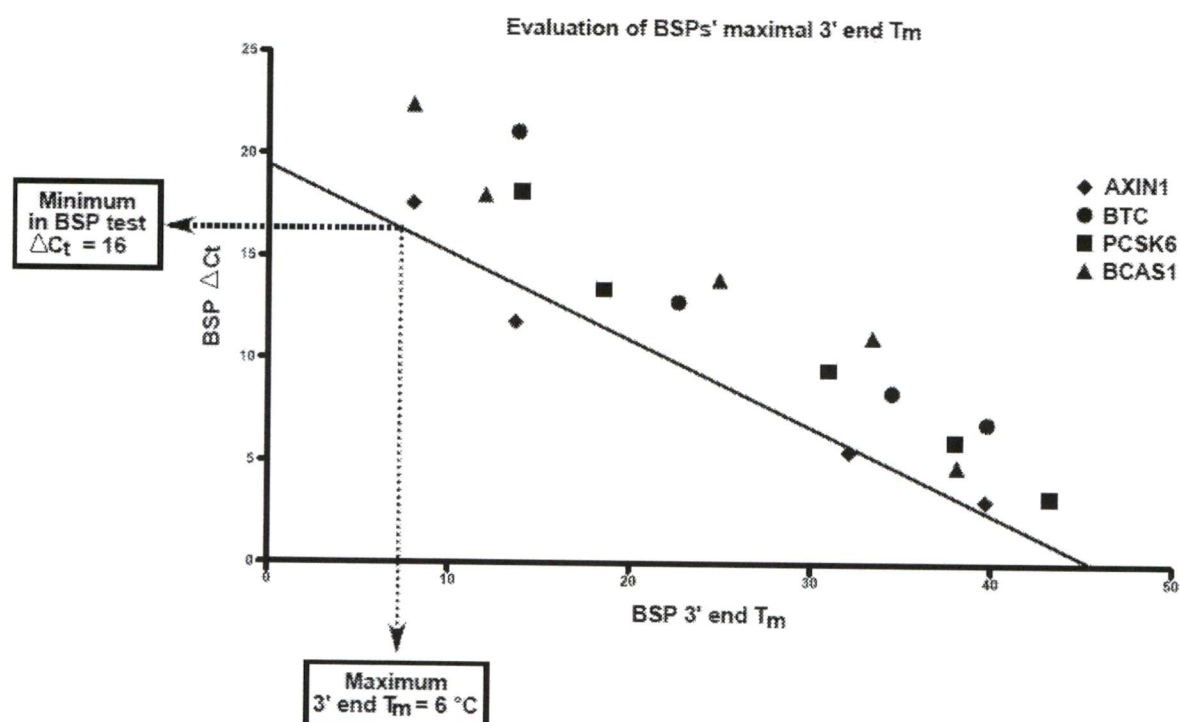


Figure S2 Evaluation of BSP's maximal 3' end T_m

BSPs for four different alternative splicing events featuring 8 to 15 bp in their 3' end matching both splicing isoform amplicon were submitted to our test as described previously for supplementary figure 1. A linear regression of all AXIN1 points predicts the maximum 3' end T_m to be 6 °C.

To determine the rule design that eliminate the mispriming due to 5' end perfect match on both splicing isoforms, we designed 59 different BSPs possessing a variety of mismatches combinations on 49 alternative splicing events (Table 1 and Suppl. Table 3). BSPs were challenged in a real-time PCR reaction using a reference cDNA, representing virtually all splicing isoforms (Universal Human Reference RNA, UHRR), and were found specific if no detectable product fall in the range expected if mispriming on the long splicing isoform would occur. We found that a C/T, A/C or G/T as 3' end terminal mismatch did not discriminate well the long and short splicing isoform. In contrast, G/G, A/A, T/T, C/C and G/A as well as any two consecutives mismatches toward the long splicing isoform were highly specific for the short splicing isoform amplification (Table 1). Together, these experiments led to the conclusion that a BSP must carry no more than 7-8 nucleotides fully

complementary and the inclusion of one terminal mismatch (G/G, A/A, T/T, C/C or G/A) whenever possible two, toward the long splicing isoform to be truly short isoform-specific.

Tableau 1 Comparison between the specificity of the boundary-spanning primers carrying different 3' end mismatches.

The specificity of 59 putative SSPPs carrying one or two mismatches with the long isoform was challenged by using a reference cDNA containing all splicing isoforms. The specificity score corresponds to the number of BSP that specifically amplify the targeted isoform for each mismatch type.

Mismatch type	Specificity score
<u>One 3' end terminal</u>	
C/T	4 on 7
A/C	5 on 6
G/T	2 on 11
A/G	7 on 7
G/G	2 on 2
C/C	3 on 3
A/A	4 on 4
T/T	7 on 7
<u>Two terminal</u>	
Any	12 on 12

Automated splice site identification and primer design

The first step to detect any splicing isoform is to identify the alternative splicing event and the unique sequence surrounding it. Therefore, we developed the RASE pipeline which starts with a computational search tool that queries publicly available sequence databases for either individual or all alternative splicing events simultaneously and identifies the upstream constitutive sequence (UCS), the alternatively included region (AIR) and the downstream constitutive sequence (DCS, Fig. 3A). As shown in figure 3B in step 1 to 4, the algorithm described in the “Methods” section calculates the similarities of the sequences found around the alternative splicing event to retrieve all BSP 3' end sequences possessing adequate length and mismatches toward the long splicing isoform. Then, the best sequence is extended in its 5' end and an opposing primer is designed to generate an amplicon that ranges in length between 70 and 100 bp (Fig. 3B, step 5 and 6). The long isoform-specific primer (LSP) was designed within the AIR (Fig. 3C). Out of 25011 genes in the NCBI build 36 database, RASE generate specific pairs of primers for 4099 simple

alternative splicing events, which corresponds to 81% of the events, regardless of the event type (i.e., single cassette exon, alternative 5' and 3' splice site and intron retention, Table 2). Specifically in the case of single cassette exons, the most frequent type of events, the design success rate reached 87%. Conversely, in the case of alternative 3' splice sites, the success rate was limited to 62%, the lowest level observed.

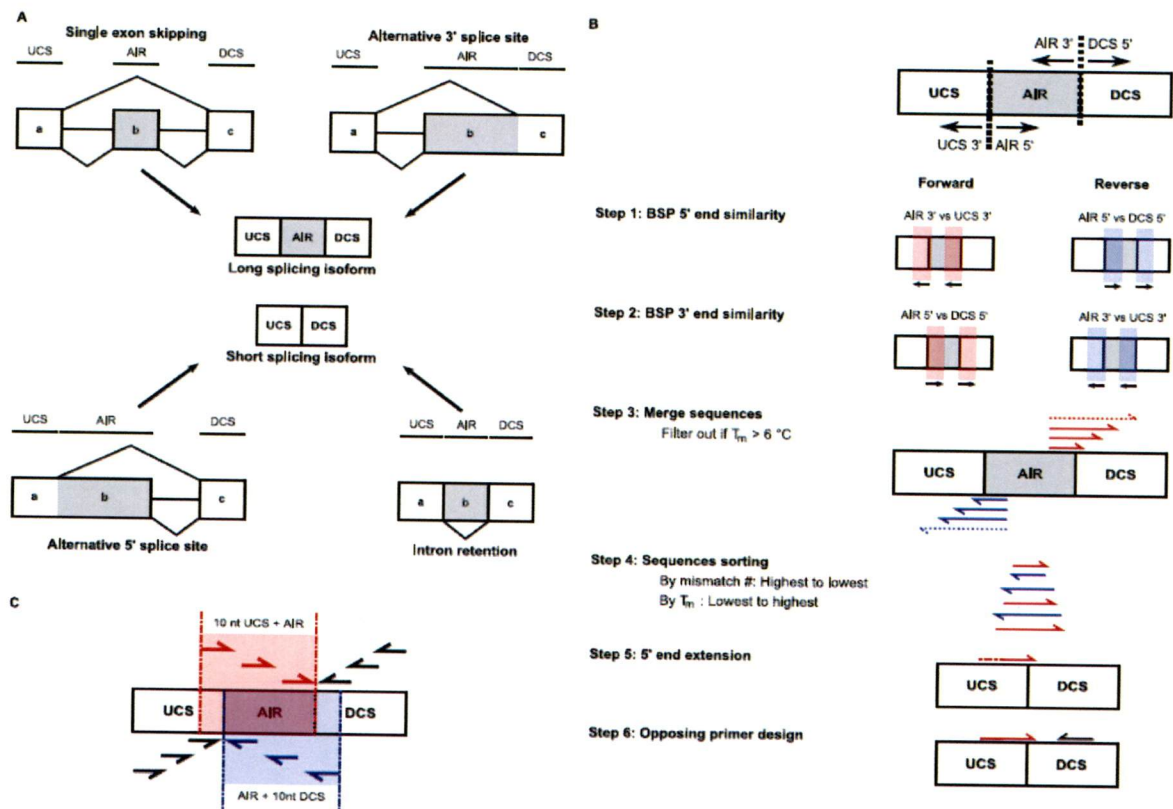


Figure 3 Strategy for automated selection of high fidelity splice sensitive PCR primers. (A) The first stage involves the identification of the upstream constitutive sequence (UCS), alternatively included region (AIR) and downstream constitutive sequence (DCS) in all simple alternative splicing events types; single cassette exon, alternative 5' and 3' splice site and intron retention. (B) The second stage generates primer sequence to quantify the short isoform of any given alternative splicing event. The schema in upper right, shows regions that are expanded on both side of junctions. Black arrows indicates the orientation of the nucleotide-by-nucleotide comparison use in step 1 and 2. In *Step 1*: the script identify the stretch of matching nucleotides between short and long splicing isoforms and the 5' end of the BSP in both orientation (BSP 5' end similarity). *Step 2*: the script analyzed the sequences for the presence of mismatches in terminal 3' end toward the long splicing isoform. (BSP 3' end similarity). *Step 3*: Sequences identified in step 1 and 2 are merged. Sequences with a $T_m > 6$ °C are removed from the valid sequence list. *Step 4*: Sequences are sorted by mismatches number and by T_m . *Step 5*: An *in silico* 5' end extension is performed on the best sequence until adequate nucleotide composition. *Step 6*: The script designs an opposing primer for the BSP yielding a SSPP (short-specific primer pair). (C) The third stage generates primer sequence to quantify the long isoform of any given alternative splicing event. The extended space design for the long-specific primer is highlight in red (forward) and blue (reverse).

Tableau 2 Summary of the primer design data generated by RASE.

Splice-sensitive primers were designed for 4099 simple alternative splicing events (e.g. single cassette exon, alternative 5' and 3' splice site, intron retention) out of 25011 genes present in NCBI build 36. Primer specific to the short splicing isoform (SSPP). Primer specific to the long splicing isoform (LSPP).

Number of alternative splicing event	Alternative splicing event type	Primer type	Design success rate (%)
2274	Single cassette exon	SSPP	90
		LSPP	97
		SSPP + LSPP	87
604	Alternative 5' splice site	SSPP	89
		LSPP	85
		SSPP + LSPP	76
642	Alternative 3' splice site	SSPP	89
		LSPP	70
		SSPP + LSPP	62
579	Intron retention	SSPP	85
		LSPP	99
		SSPP + LSPP	84
4099	All splicing event	SSPP	89
		LSPP	91
		SSPP + LSPP	81

Experimental validation of RASE primers

To evaluate the fidelity of the RASE we experimentally tested a set of primers targeting 51 cancer-specific alternative splicing events (Klinck et al., 2008; Venables et al., 2008). As much as three different primer designs for the long and short splicing isoforms of each gene were generated and examined (Table 3). The evaluation process consisted of a single test that looks at the primer capacity to amplify the targeted sequence in a cDNA from a reference RNA (UHRR). The primers were disqualified if they resulted in non-specific amplifications, abnormal amplification curves or the formation of primer-dimers when amplified in the presence of cDNA containing both splicing isoforms. Additionally, an efficiency curve was performed as described earlier (Nolan et al., 2006a). The results indicate that an average of 85 % of the cases, at least one of the three primers set designed for each splicing isoforms specifically amplified the targeted amplicon. Primers failed either due to dimerization or the amplification of unspecific sequence. Interestingly, only one primer set was completely inactive and spiking a synthetic template in UHRR cDNA shows specific amplification of this inactive primer set. All together, it demonstrates that UHRR cDNA is representative of almost all splicing isoforms investigated so far. As low as three putative primer sets specific to the short splicing isoform misprimed on the long version, a BSP specificity success rate of 94 %. All validated primer pair were tested in a cDNA derived from a pool of RNA extracted from FFPE tissues. Most of the primer pairs tested specifically amplified the targeted splicing isoform. Therefore, we conclude that amplification of splicing isoforms is possible from both cell lines and FFPE tissues using the RASE.

Tableau 3 Experimental validation of RASE designed primers.

Putative splice-sensitive primers designed by RASE were tested in a cDNA derived from Universal Human Reference RNA (UHRR). Primer set yielding the expected band size, purity and efficiency were classified as “Validated”. Alternative design was tested for primers failing to specifically amplify the targeted isoform. Cumulative results are shown as a % out of the total 51 primer pair tested. Primers were also tested for the amplification of cDNA derived from FFPE RNA extract and the results are presented as a % out of the 40 SSPP and 47 LSPP tested.

Design	UHRR			FFPE pool
	1 (%)	2 (%)	3 (%)	(%)
SSPP				
Validated	70	76	78	83
Multi-band	16	16	16	2
Primer-dimers	4	4	4	0
No amplicon	2	2	0*	15
Bad efficiency	8	2	2	0
LSPP				
Validated	75	88	92	83
Multi-band	4	2	2	2
Primer-dimers	19	10	6	0
No amplicon	0	0	0	15
Bad efficiency	2	0	0	0

Discussion

Quantitative real-time PCR, the current gold standard in mRNA quantification, has been only marginally used for the quantification of splicing isoforms. This is probably because the majority of software packages either cannot design splice-sensitive primers or design primers with low or unverified specificity (e.g. <http://www.premierbiosoft.com>, <https://www.genscript.com/ssl-bin/app/primer> or <http://www1.qiagen.com>). Another important reason is that current packages do not take into account the high degree of similarity between the exon-exon junction sequences of differentially spliced mRNAs.

In this study we circumvented these caveats by developing RASE, a pipeline that can directly identify and retrieve alternative splicing information from any public database and design primers that are specific to either the short or long splicing isoforms. To discriminate between two closely related sequences, we took advantage of *Taq* DNA polymerase requirements to prime PCR reaction: the overall duplex stability and the 3' end pairing. Numerous independent researchers addressed the minimal number of mismatches and position required to specifically amplified unique splicing isoforms, but depending on the exact experimental setting used, different conclusions were reached (Ayyadevara et al., 2000). Here, we revisited this question by performing a comprehensive coverage on multiple primer/template combinations. Requiring one, and whenever possible two, terminal mismatches to the non-cognate sequence and at most 8 nucleotides of complementarity to the cognate sequence at the 3' end were critical parameters since they allowed the largest coverage of designs in terms of the numbers of alternative splicing events while retaining the high specificity required for real-time PCR splicing isoform quantification. This automatic primer design provides ready-to-use primers for the robust detection of splicing isoforms in days, something that previously took months to determine using classical methods that required cloning of each splicing isoform, experimental testing of individual primers and manual adjustments of mismatch number and annealing temperature (Vandenbroucke et al., 2001; Walton et al., 2007; Williams et al., 1999). In addition, we have extended universal primer design capacity to the quantification of splicing isoform in fixed tissues. RNA extracted from FFPE blocks is of very poor quality, which put the reliability of quantification in FFPE tissues in question (Farragher et al.,

2008). Numerous parameters affecting raw C_t values were identified and reported and proper normalization and adequate amplicon size were shown to dramatically increase the quality of the data obtained from clinical samples (Abrahamsen et al., 2003; Antonov et al., 2005; Cronin et al., 2004). In this study, RASE designed 3322 primer sets with suitable amplicon size to quantify in fixed tissues, making it the first standardized primer design process capable of quantifying splicing isoforms in formalin-fixed, paraffin-embedded (FFPE) tissues. The difficulty in splice-sensitive primer design and validation also significantly impaired the quality of large data sets obtained from techniques such as tiling arrays (Castle et al., 2008) splicing arrays (Calarco et al., 2007), and deep sequencing (Pan et al., 2008). In all cases, few splicing isoforms were validated by a secondary approach and an even lower number have been screened in large patient cohorts to confirm their disease association. Since RASE guarantees isoform specificity and provides pre-optimized primers working under universal PCR conditions, it makes high-throughput annotation and validation of splicing isoforms feasible.

The goal of the present study was to build a pipeline for isoform-specific quantification in clinical samples using real-time PCR. To achieve this goal we built the RASE (ReaL-time PCR Annotation of Splicing Events) platform to eliminate tedious and time-consuming steps associated with primers validation and streamline the amplification process. The key features of RASE includes automated nucleotide by nucleotide genome-wide analysis of splicing isoforms and the evaluation of primers based on experimentally validated strict criteria allowing amplification under universal PCR conditions. The outcome of RASE is stored in the form of database of ready to use isoform-specific primer sequences for most alternative splicing events in the NCBI build 36 database (available upon request). In addition, custom primer design is also possible using publically available web-interface. Using these tools it is now possible to design and conduct large-scale quantitative surveys of splicing isoforms in cell lines or archival samples, paving the way towards a greater understanding of the contribution of alternative splicing to complex human diseases such as cancer.

Materials and methods

Splicing isoform-specific primer design. An application programming interface (API) and executable computer assisted scripts capable of designing isoform-specific primer pairs were developed using Perl version 5.8.8 (the Perl directory <http://www.perl.org/>) with CPAN modules (CPAN: Comprehensive Perl Archive Network <http://cpan.org/>) and the gene annotation databases were stored using a MySQL database server <http://www.mysql.com/>. The design scripts can be called using a list of alternative splicing events described in a tab-delimited file. Batch design of the whole NCBI build 36 human annotation database (Pruitt et al., 2007) takes about 10 days on an average PC. Initially, transcripts associated with any given gene were retrieved from the database and the alternative splicing events (cassette exon, intron retention, alternative 3' splice site and alternative 5' splice site) were identified using an algorithm similar to that used by ASTALAVISTA (Foissac and Sammeth, 2007).

Prior to design SSPP, the script determines if a BSP orientation must be imposed based on the desired splicing event. To do so, it verified if the 3' end of the UCS or the 5' end of the DCS undergoes alternative splicing in other transcripts. If only the 3' end of the UCS did, the script imposed the BSP to be the forward primer. Conversely, the BSP has to be the reverse primer if only the 5' end of the DCS undergoes alternative splicing. In the case where both undergo alternative splicing, the design was aborted. For the alternative splicing events where no orientation was imposed, the chosen one was based on the following rules. To determine 5' end similarity, the script compares, on a nucleotide-by-nucleotide basis, the sequences found at the 3' ends of the AIR and UCS, and 5' ends of the AIR and DCS. If the nucleotides are the same, iteration was performed on the next nucleotides of these regions until a difference was reached for each orientation. The number of identical nucleotides represents the 5' end similarity of the BSP (Fig. 3B, step 1). Next, RASE analyzed the number and position of mismatches in terminal 3' end by an algorithm similar to the one used to determine the BSP 5' end similarity. All sequences with two consecutive terminal mismatches or one strict terminal mismatch (A/A, T/T, C/C, G/G or G/A) were kept (Fig. 3B, step 2). Then, the sequences were extended in their 5' end using the BSP 5' end

similarity previously identified. The T_m of each sequence was calculated and all sequences with T_m greater than 6 °C were rejected (Fig. 3B, step 3). The remaining sequences were sorted: 1) Higher number of mismatches 2) Lower T_m (Fig. 3B, step 4). Afterwards, the scripts iterated through the sequence list and expanded the sequence using the 3' end of the UCS or the 5' end of the DCS for forward or reverse BSP, respectively, until an adequate T_m of 58 - 62 °C (optimal T_m = 60 °C), a GC content between 30-80 % (optimal 50 %) and a maximal length of 30 nucleotides (minimum 18 nucleotides) were reached (Fig. 3B, step 5). Then, an opposing primer was designed using Primer3 (Rozen and Skaletsky, 2000) with the default parameters excepted for the aforementioned T_m , GC content, primer length parameters and the maximum T_m difference between primers was reduced to 3 °C. The primer pair was scored for having no *in silico* off-targets (all possible amplification products with a size smaller than 500 bp were considered as possible *in silico* off-targets and were associated to the primer pair design when calling wublast (Gish, 1996-2003) or NCBI blast (Altschul et al., 1997) as a script parameter on an in-house transcript database based on the Aceview annotation (Thierry-Mieg and Thierry-Mieg, 2006), for giving low primer pair penalties according to Primer3 and no mispriming when checked against the Primer3 human mispriming library (Fig. 3B, step 6).

The LSPP design is described in figure 3C. To design long-specific primer (LSP), an extended space design of 10 nucleotides outside the AIR was allowed according to the LSP orientation. If the LSP was a forward, the space design consists of the AIR and the last 10 nucleotides of the UCS, allowing it to be a BSP. Conversely, if the LSP was a reverse, the space design consists of the AIR and the first 10 nucleotides of the DCS. The opposing primer was designed according to the rule described for the SSPP or aborted otherwise. All primer pairs successfully validated were analyzed for the presence of single nucleotide polymorphism (see <http://palace.lgfus.ca>, Suppl. Table 4 and UCSC Genome Browser custom track for SNP verification). A low-throughput version (only one alternative splicing event can be designed at a time) of the script was also developed and is available for the scientific community via a web interface <http://designs.lgfus.ca>. Design scripts and API are also available through <http://palace.lgfus.ca>.

RNA extraction and quality control. DNase treatment on Universal Human Reference RNA (UHRR, Stratagene) was performed using RNase-free DNase set (Qiagen, Canada). Human normal and serous epithelial ovarian cancer tissues were obtained in 10 microns FFPE shavings from the « Réseau de Recherche sur le Cancer du Fonds de la Recherche en Santé du Québec » biobank. Single-step total RNA extractions on human ovarian tissue were performed using RNeasy FFPE (Qiagen, Canada). The RNA was examined using Agilent 2100 bioanalyzer (Agilent technologies, Santa Clara, CA) and samples displaying nucleic acid density ratio of 1.8 O.D._{260/280} and RNA integrity (RIN) value higher than 9 were used. Lower RIN values were accepted for RNA extracted from FFPE tissues. The presence of PCR inhibitors was evaluated using a SYBR Green version of the SPUD assay (Nolan et al., 2006b) and genomics DNA contamination was examined using exonic primers of PLD1 gene (5'-TCATTGACCAATCGGTGGCCTT-3' and 5'-TGAAGTGACCCGCTTCACACT-3') and intron-spanning of CCNE1 gene (5'-AAGTGGCGTTTAAGTCCCCTGA-3' and 5'-ATGATACAAGGCCGAAGCAGCA-3').

Quantitative PCR analysis. RNA reverse transcription was performed using 50-250 ng of RNA with “Transcriptor”, random primer p(dN)₆, dNTPs (Roche Diagnostics, Laval, Québec, Canada) and RNase OUT (Invitrogen). Total cDNA generated from FFPE tissues were further purified on QIAquick PCR purification kit (Qiagen, Canada). Total cDNA were pooled and diluted to 0.33-3.33 ng/ μ L. All forward and reverse primers were individually resuspended at 20-100 μ M stock solution in Tris-EDTA buffer (IDT, Iowa, USA) and diluted as a primer pair to 1 μ M in RNase DNase-free water (IDT, Iowa, USA). 10 μ L Real-time PCR reactions were performed either in a 7500 ABI apparatus (Applied Biosystems, Foster City, CA) or a Realplex (Eppendorf) with 5 μ L of FastStart Universal SYBR Green Master mix (Roche Diagnostics, Laval, QC, Canada) with the following cycling: 10 min at 95°C; 50 cycles: 15 sec at 95°C, 30 sec at 60°C, 30 sec at 72 °C; and melting curve: 15 sec at 95°C, 60 sec at 60°C ; 1 °C /min temperature gradient, 15 sec at 95°C) using 3 μ L (totalizing 1-10 ng) of template cDNA or diluted amplicons in tRNA (50 ng/mL, Ambion) and 2 μ L of a 1 μ M solution of forward and reverse primer.

The primer test was performed in technical triplicate using 1 or 10 ng of cDNA generated from RNA extracted from UHRR or from a 50/50 mixture of ovarian normal/cancerous

FFPE tissues, respectively. The amplicons generated during the primer test were examined by capillary electrophoresis as previously described (Klinck et al., 2008) and stored for future use. In one case (CCL4), a chemically synthetic template of the following sequence 5' AAGCTCTGCGTGA CTGTCTCTCCTCATGCTAGTAGCTGCCTTCTGCTCT CCAGCGCTCTCAGCACCAAATTCCAAACCAAAGAAGCAAGC-3' was diluted and used as an amplicon derived from the primer test as described above. Efficiency curve were performed across a four-log serial dilution protocol. A derived value of 2,00 represents an amplification efficiency of 100 % and primer pairs with an efficiency between 90 and 105 % ($E = 1,90 - 2,05$) are considered acceptable (Nolan et al., 2006a), although we considered SSPPs with $E = 1,85 - 2,05$. See <http://palace.lgfus.ca>, Suppl. Table 4 for all validated primer sets mentioned in Table 3.

For the design of the training set used to evaluate mispriming potential (MP 5', Fig. 2), putative SSPPs were designed with a BSPs having 3 to 15 nucleotides in 3' end passes the exon-exon junction and a fixed opposing primer. Primers to generate the template were designed as LSPP design with the additional constraints of having the LSP in the same orientation as the BSP. BSP's opposing primers were designed to perform nested PCR using the long splicing isoform template. The amplification in UHRR cDNA was performed as described for the primer test. Amplicons were in the range of 100 to 300 bp, allowing evaluation of the quality of the amplification by amplicon sequencing using the Big dye chemistry on a ABI 3730xl apparatus. The mispriming potential was evaluated by performing real-time PCR reactions in technical duplicates using 3 μ L of the 1: 100 000 000 dilution of the template , 2 μ L of primer set (LSPP as the reference primer pair and SSPP as the putative primer pair to be challenged) in 10 μ L final volume as described for primer test. No template controls were run in parallel and were negative. A ΔC_t (where $\Delta C_t = C_{t\text{ SSPP}} - C_{t\text{ LSPP}}$) was calculated. See <http://palace.lgfus.ca>, Suppl. Table 1 for primer sequences.

For the evaluation of mispriming potential (MP 3', Table 1), a BSP mismatch rule set was design with BSPs fully matching the short splicing isoform and bearing one 3' end terminal mismatch or two consecutive 3' end terminal mismatches toward the long splicing isoform.

Opposing primers were designed to yield amplicon in the range of 100 to 300 bp. Primers were tested as mentioned for the primer test. See <http://palace.lgfus.ca> Suppl. Table 3 for primer sequences.

Acknowledgments

This work was supported by a grant from Génome Québec and Genome Canada. Sherif Abou Elela is a Chercheur National of the Fonds de la Recherche en Santé du Québec. Jean-Pierre Perreault holds the Canada Research Chair in Genomics and Catalytic RNA. The authors would like to thank Bruno Lamontagne, Lucien Jr Bergeron, Julian P. Venables, Benoit Chabot and Roscoe Klinck for reviewing this manuscript as well as Sonia Couture for cell culture, Maxime Lévesque and Chu Shin Koh for implementation of bioinformatic analysis module, Jules Gagnon for testing the script and web interface, Dany Namroud for server administration and Marc-André Rodrigue for sequencing. SAE and JPP are members of the Centre de Recherche Clinique Étienne-Lebel.

References

- Abrahamsen, H.N., Steiniche, T., Nexø, E., Hamilton-Dutoit, S.J. and Sorensen, B.S. 2003. Towards quantitative mRNA analysis in paraffin-embedded tissues using real-time reverse transcriptase-polymerase chain reaction: a methodological study on lymph nodes from melanoma patients. *J. Mol. Diagn.* 5: 34-41.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.
- Antonov, J., Goldstein, D.R., Oberli, A., Baltzer, A., Pirota, M., Fleischmann, A., Altermatt, H.J. and Jaggi, R. 2005. Reliable gene expression measurements from degraded RNA by quantitative real-time PCR depend on short amplicons and a proper normalization. *Lab. Invest.* 85: 1040-1050.
- Ayyadevara, S., Thaden, J.J. and Shmookler Reis, R.J. 2000. Discrimination of primer 3'-nucleotide mismatch by taq DNA polymerase during polymerase chain reaction. *Anal. Biochem.* 284: 11-18.
- Bracco, L., Throo, E., Cochet, O., Einstein, R. and Maurier, F. 2006. Methods and platforms for the quantification of splice variants' expression. *Prog. Mol. Subcell. Biol.* 44: 1-25.
- Calarco, J.A., Xing, Y., Caceres, M., Calarco, J.P., Xiao, X., Pan, Q., Lee, C., Preuss, T.M. and Blencowe, B.J. 2007. Global analysis of alternative splicing differences between humans and chimpanzees. *Genes Dev.* 21: 2963-2975.
- Castle, J.C., Zhang, C., Shah, J.K., Kulkarni, A.V., Kalsotra, A., Cooper, T.A. and Johnson, J.M. 2008. Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat. Genet.* 40: 1416-1425.
- Cronin, M., Pho, M., Dutta, D., Stephans, J.C., Shak, S., Kiefer, M.C., Esteban, J.M. and Baker, J.B. 2004. Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *Am. J. Pathol.* 164: 35-42.
- Farragher, S.M., Tanney, A., Kennedy, R.D. and Paul Harkin, D. 2008. RNA expression analysis from formalin fixed paraffin embedded tissues. *Histochem. Cell Biol.* 130: 435-445.
- Foissac, S. and Sammeth, M. 2007. ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Res.* 35: W297-9.
- Gish, W. 1996-2003. <http://blast.wustl.edu>.
- Klinck, R., Bramard, A., Inkel, L., Dufresne-Martin, G., Gervais-Bird, J., Madden, R., Paquet, E.R., Koh, C., Venables, J.P., Prinos, P. et al. 2008. Multiple alternative splicing markers for ovarian cancer. *Cancer Res.* 68: 657-663.
- Koch, I., Slotta-Huspenina, J., Hollweck, R., Anastasov, N., Hofler, H., Quintanilla-Martinez, L. and Fend, F. 2006. Real-time quantitative RT-PCR shows variable, assay-dependent sensitivity to formalin fixation: implications for direct comparison of transcript levels in paraffin-embedded tissues. *Diagn. Mol. Pathol.* 15: 149-156.
- Morrison, T.B., Weis, J.J. and Wittwer, C.T. 1998. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *BioTechniques* 24: 954-8, 960, 962.
- Nagao, K., Togawa, N., Fujii, K., Uchikawa, H., Kohno, Y., Yamada, M. and Miyashita, T. 2005. Detecting tissue-specific alternative splicing and disease-associated aberrant splicing of the PTCH gene with exon junction microarrays. *Hum. Mol. Genet.* 14: 3379-3388.

- Ni, J.Z., Grate, L., Donohue, J.P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T.A., Blume, J.E. and Ares, M., Jr. 2007. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* 21: 708-718.
- Nolan, T., Hands, R.E. and Bustin, S.A. 2006a. Quantification of mRNA using real-time RT-PCR. *Nat. Protoc.* 1: 1559-1582.
- Nolan, T., Hands, R.E., Ogunkolade, W. and Bustin, S.A. 2006b. SPUD: a quantitative PCR assay for the detection of inhibitors in nucleic acid preparations. *Anal. Biochem.* 351: 308-310.
- Oberli, A., Popovici, V., Delorenzi, M., Baltzer, A., Antonov, J., Matthey, S., Aebi, S., Altermatt, H.J. and Jaggi, R. 2008. Expression profiling with RNA from formalin-fixed, paraffin-embedded material. *BMC Med. Genomics* 1: 9.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J. and Blencowe, B.J. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40: 1413-1415.
- Pruitt, K.D., Tatusova, T. and Maglott, D.R. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 35: D61-5.
- Rozen, S. and Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132: 365-386.
- Tazi, J., Bakkour, N. and Stamm, S. 2009. Alternative splicing and disease. *Biochim. Biophys. Acta* 1792: 14-26.
- Thierry-Mieg, D. and Thierry-Mieg, J. 2006. AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol.* 7 Suppl 1: S12.1-14.
- Vandenbroucke, I.I., Vandesompele, J., Paepe, A.D. and Messiaen, L. 2001. Quantification of splice variants using real-time PCR. *Nucleic Acids Res.* 29: E68-8.
- Venables, J.P. 2006. *Alternative splicing in Cancer*. Transworld Research Network.
- Venables, J.P., Klinck, R., Bramard, A., Inkel, L., Dufresne-Martin, G., Koh, C., Gervais-Bird, J., Lapointe, E., Froehlich, U., Durand, M. et al. 2008. Identification of alternative splicing markers for breast cancer. *Cancer Res.* 68: 9525-9531.
- Walton, H.S., Gebhardt, F.M., Innes, D.J. and Dodd, P.R. 2007. Analysis of multiple exon-skipping mRNA splice variants using SYBR Green real-time RT-PCR. *J. Neurosci. Methods* 160: 294-301.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* 456: 470-476.
- Wellmann, S., Taube, T., Paal, K., Graf V Einsiedel, H., Geilen, W., Seifert, G., Eckert, C., Henze, G. and Seeger, K. 2001. Specific reverse transcription-PCR quantification of vascular endothelial growth factor (VEGF) splice variants by LightCycler technology. *Clin. Chem.* 47: 654-660.
- Williams, D.M., Koduri, S., Li, Z., Hankins, W.D. and Poola, I. 1999. Primer design strategies for the targeted amplification of alternatively spliced molecules. *Anal. Biochem.* 271: 194-197.
- Yang, D. and Le, J. 1994. Targeted amplification of alternatively spliced transcripts of major histocompatibility complex class I heavy chain. *J. Immunol. Methods* 176: 265-270.

Chapitre 2

Développement d'outils moléculaires isoforme-spécifiques

Auteurs du chapitre : **Jean-Philippe Brosseau**, Elvy Lapointe, Jean-François Lucier, Sonia Couture, Philippe Thibault, Daniel Gendron et Eric Paquet.

Avant-propos

Figures 1, 2, 3, 6, 8, 9, 15, 16 et tableau 1. JPB a produit les résultats, interprété et réalisé tous les figures et tableaux.

Figures 4, 5 et tableau 2. EL a produit les résultats. JPB a interprété et produit les figures et le tableau.

Figure 7. DG a produit les résultats. DG et JPB ont produit la figure.

Figures 11, 13 et 14. JPB (PCR quantitatif) et EL (endpoint PCR) ont produit les résultats. JPB a interprété les résultats et produit la figure.

Les transfections ont été réalisées par SC.

Les analyses PCR ont été analysées par PP et JPB

Les dessins de TOSS et antisens ont été produits par JFL (85%), JPB (5%), EP (10%).

Les dessins de TOES et les stratégies afin d'inclure l'exon ont été dessinés par JPB et JFL.

Les dessins de siRNAs ont été dessinés par JPB et JFL, basés sur un algorithme conçu par JFL.

Mise en contexte

Le développement des technologies antisens (dépendantes de la RNase H) et plus tard du RNAi, a démontré qu'il est possible de cibler virtuellement n'importe quelle séquence d'un ARNm mature afin de réprimer son expression. Par contre, la plupart des gènes sont épissés alternativement et possèdent plusieurs isoformes très semblables dont la fonction est mal comprise. Dans quelques cas, il a été possible d'adapter la stratégie antisens ou RNAi pour cibler spécifiquement une des isoformes et en révéler sa fonction. Par contre, la plupart des exemples cible un exon et il n'existe pas de règle de dessin pour les siRNAs à cheval sur la jonction, la seule région à cibler dans le cas de l'isoforme court.

La technologie antisens peut également être adapté plus profondément afin d'influencer la décision d'épissage du pré-ARNm. Le TOSS (targeted oligonucleotide silencer of splicing) en est un exemple et a été publié récemment avant le début de notre étude par Villemaine et collaborateurs (2003). Le TOSS est un oligonucléotide bifonctionnel possédant une portion antisens complémentaire à l'exon alternatif ciblé et une portion non hybridante reconnaissant le facteur d'épissage hnRNPA1. La preuve de concept a été réalisé en ciblant le pré-ARNm endogène de BCL2L1 dans plusieurs lignées cellulaires cancéreuses humaines. Toutefois, il n'est pas clair à quel point ce concept peut être généralisé à d'autres gènes, type d'évènement d'épissage et si oui, quel en sont les règles de dessin.

Les travaux présentés dans ce chapitre ont pour but de démontrer que 1) Le TOSS reprogramme réellement l'épissage de cible endogène, 2) Le design rationnel des TOSS est possible; 3) Le design rationnel de siRNA jonction est possible.

Methodologies

Transfection de TOSS ou SSO en 6 puits

D'abord, chacun des puits d'une plaque 6 puits (1 puits par oligonucléotide à transférer) estensemencé avec une suspension de 250 000 cellules dans du milieu complet (1 mL) et laissé adhérer à 37°C, 5% CO₂ pour 4 h. 30 min avant la fin du 4 h, les oligonucléotides (2,4 µL; 200 µM) sont dilués dans l'Opti-MEM I (97,6 µL) dans un microtube de 1,5 mL. Puis, la lipofectamine 2000 (3 µL/puits) est diluée dans l'Opti-MEM I (97 µL/puits) et laissée reposer 5 min. Cette solution de lipofectamine 2000 (100 µL/puits) est ajoutée dans chaque microtube contenant un oligonucléotide dilué et dans un tube contenant seulement de l'Opti-MEM I (contrôle négatif), puis il faut la laisser reposer 20 min. La totalité du contenu de chacun des complexes oligonucléotide-lipofectamine 2000 (200 µL) est ajoutée goutte-à-goutte dans leur puits respectif et incubée à 37°C, 5% CO₂ pour 24 h.

Électrophorèse capillaire (Caliper 90)

Dessin d'amorces de endpoint PCR

Fait tel que décrit dans Klinck et collaborateurs (2008).

Essai endpoint PCR One step (Qiagen)

Fait selon les recommandations du manufacturier.

Bloc de TOSS *in vitro*.

D'abord, chaque TOSS est dilué en série par un facteur 10 (20 nM – 20 µM). Ensuite, les différentes concentrations de TOSS (5 µL) sont diluées individuellement avec une solution d'ARN total de HeLa S3 (5 µL, 100 ng) et du tampon Qiagen 5X One step (2,5 µL). Cette solution est chauffée (95°C, 3 min) et subséquemment refroidie (t.p., 15 min). Puis, une solution de déoxynucléoside triphosphate (dNTPs, 1 µL), le tampon 5X (2,5 µL), l'eau

(4,66 μ L), le mélange d'enzyme One step Qiagen (1 μ L) et les amorces PCR (1,67; 9 μ M) sont ajoutées. Les tubes sont ensuite transférés sur un thermocycleur PCR et la PCR est exécutée selon le protocole suivant : 50°C, 30 min; 95°C, 15 min; 35 cycles (94°C, 30 sec; 55°C, 30 sec; 72°C, 1 min); 72°C, 10 min. Après la réaction, les produits PCR sont séparés par électrophorèse capillaire.

Quantification du TOSS résiduel au cours d'une extraction au Trizol®

L'expérience suivante est faite en parallèle pour les 8 puits excepté l'ajout de TOSS à des moments bien définis. D'abord, huit puits de plaque six puits sontensemencés avec une suspension de 250 000 cellules dans du milieu complet (1 mL) et laissés adhérer à 37°C, 5% CO₂ pour 24 h. Le lendemain, le surnageant est enlevé et les cellules sont délicatement lavées au D-PBS (1 mL). Du Trizol® (1 mL/puits) est ajouté et le TOSS A1_Bclx4 (480 nmol) est ajouté dans le puits 1. 5 min plus tard, le lysat cellulaire est passé 10 fois avec la pipette avant d'être transféré dans des microtubes de 1,5 mL. À ce moment, A1_Bclx4 (480 nmol) est ajouté dans le microtube correspondant au puits 2. Ensuite, du CHCl₃ (200 μ L) est ajouté ainsi que A1_Bclx4 (480 nmol) dans le microtube correspondant au puits 3. Les microtubes sont ensuite fermés, agités manuellement (15 sec) et laissés reposer (2 min, t.p.), avant d'être centrifugés (13 000 rpm, 15 min, 4°C). La phase aqueuse (supérieure, 0,5 mL) est ensuite transférée dans des microtubes contenant du glycogène (40 μ g) et de l'iPrOH (500 μ L). À ce moment, A1_Bclx4 (480 nmol) est ajouté dans le microtube correspondant au puits 4. Les microtubes sont fermés, agités manuellement (15 sec) et laissés reposer (10 min, t.p.) avant d'être centrifugés (10 000 rpm, 10 min, 4°C). Le surnageant est enlevé délicatement et de l'EtOH 75% (1 mL) y est ajouté. À ce moment, A1_Bclx4 (480 nmol) est ajouté dans le microtube correspondant au puits 5. Les microtubes sont refermés, vortexés (2 sec) et centrifugés (10 000 rpm, 5 min, 4°C). Le surnageant est enlevé délicatement et du A1_Bclx4 (480 nmol) est ajouté dans le microtube correspondant au puits 6. Les culots d'ARN totaux sont séchés à l'air libre (10 min) et subséquemment resuspendus dans H₂O (30 μ L). À ce moment, A1_Bclx4 (480 nmol) est ajouté dans le microtube correspondant au puits 7. Tous les microtubes sont laissés reposer sur glace (2 min), vortexés (10 sec) et centrifugés (13 000 rpm, 15 sec, t.p.). Finalement, les

extraits d'ARN sont chauffés (10 min, 60°C) et A1_Bclx4 (480 nmol) est ajouté dans le microtube correspondant au puits 8.

Marquage de sonde au P³²

20 ng de sonde (anti-TOSS : CACCCTAAGTCCCTATCATA ou anti-U6 : TCTTCTCTGTATCGTTCC) est dilué avec de l'H₂O (3 µL), du tampon kinase (1 µL), du P³² gamma-ATP et de la T4 polynucléotide kinase (1 µL) dans un volume final de 10 µL et placé dans un bain (37°C, 45 min). La réaction est arrêtée en ajoutant du tampon tris-EDTA (TE) (10 µL) et en plaçant le microtube dans un bloc chauffant (65°C, 5 min). Le mélange réactionnel est dilué avec de l'H₂O (20 µL) et passé sur colonne S-200.

Northern-Blot

Un gel 12% d'acrylamide avec urée 8,3 M dans du tampon TBE (tris-borate-EDTA) 1X est moulé et laissé reposer. Puis, les échantillons sont chauffés (60°C, 2 min) et déposés sur gel. Ceux-ci doivent d'abord pénétrer le gel pendant 20 min à 12 watts, puis 45 min à 19 watts. Une fois la migration terminée, les échantillons sont transférés sur membrane Hybrid XL (4°C, 180 min, 400 mA). Une fois le transfert terminé, la membrane est soumise aux rayons ultra-violet de chaque côté (2 x 2 min). La membrane est ensuite lavée avec 50 mL de solution SSPE (NaCl-NaH₂PO₄-EDTA) 6X pH =7.7 contenant du Denhardt's 2X, et du SDS 0,1% (40°C, 60 min). La membrane est ensuite hybridée avec les sondes anti-TOSS (7,5 µL) et anti-U6 (7,5 µL) fraîchement radiomarquées au P³² (40°C, 16 h) dans une solution SSPE 6X (50 mL) contenant du Denhardt's 2X et du SDS 0,1%. Le lendemain, la membrane est lavée deux fois avec une solution de SSPE 6X (50 mL) contenant du Denhardt's 2X et du SDS 0,1% (40°C, 30 min). Finalement, la membrane est enveloppée avec un saran et exposée sur film 24 h dans une cassette.

Essai de retard sur gel

Les oligonucléotides sont d'abord purifiés sur gel d'acrylamide. Puis, une fraction de ceux-ci (1 μmol) sont marqués au P^{32} tel qu'il a été décrit précédemment. Ces oligonucléotides marqués sont ensuite incubés avec SF2-GST (165 nM final) (généreusement donné par Benoit Chabot) 10 min dans un tampon HEPES. Finalement, les complexes sont séparés sur gel d'acrylamide 4% non-dénaturant.

Essai de PCR en temps réel

Intégrité de l'ARN (Agilent RNA chip)

Quantification de l'ARN (Nanodrop)

Test de présence d'ADN génomique

Tel qu'il a été décrit dans le chapitre 1.

Extraction d'ARN total au Trizol® (Invitrogen)

Celle-ci a été faite selon les recommandations du manufacturier excepté l'ajout d'acrylamide linéaire (5 μg) lors de la précipitation à l'isopropanol.

Extraction d'ARN sur colonne (Absolutely RNA 96 Microprep kit de Stratagene)

Tel qu'il a été décrit dans le chapitre 3.

Résultats : Développement d'outils moléculaires isoforme-spécifiques

2.1 Développement d'outils moléculaire favorisant l'inclusion d'exon alternatif

2.1.1 Méthodologie pour quantifier l'effet des TOSS sur les isoformes de l'épissage

2.1.1.1 Le TOSS peut produire un changement de ratio artéfactuel *in vitro*.

Le dessin de TOSS original (A1_Bclx4) est conçu pour rediriger l'épissage vers le 5'ss en amont dans l'exon 2 du gène BCL2L1 (Bcl-x) et a été publié récemment avant le début de notre étude par Villemaire et collaborateurs (2003). Bien que la démonstration de modulation soit évidente, les résultats ne sont pas représentés par le ratio conventionnel de l'isoforme long sur le total, mais bien directement par un ratio de l'isoforme long sur l'isoforme court. Cette dernière méthode permet difficilement de juger de l'amplitude de modulation d'un TOSS. Nous avons donc re-analysé l'effet du TOSS contre BCL2L1 (A1_Bclx4) par endpoint PCR selon la méthode d'analyse conventionnelle (ψ). Brièvement, des cellules PC-3 ont été transfectées transitoirement en présence de lipofectamine 2000 (LF) par A1_Bclx4. Après 24 h, l'ARN total a été extrait au Trizol®, dosé par Nanodrop et la qualité évaluée par Agilent Bioanalyzer 2100. Puis, un essai de endpoint PCR en une seule étape spécifique au alt5' du gène BCL2L1 a été effectué. Une fois les produits PCR séparés par électrophorèse capillaire, le ratio d'isoforme est exprimé sous forme de ψ tel que décrit dans l'introduction. Lorsque comparé au véhicule de transfection seul (LF), on note que A1_Bclx4 réduit la valeur de ψ (Fig. 1). Pour apprécier la spécificité de cet effet, un oligonucléotide de même chimie, longueur et séquence de recrutement hnRNP A1, mais dont la portion antisens n'est pas complémentaire à l'ARNm de BCL2L1 a été dessiné. En fait, cette séquence antisens est l'une des moins susceptibles de s'apparier à des séquences codantes lorsque criblée *in silico* sur une banque d'ADNc codant. Lorsque les cellules sont transfectées en parallèle par ce TOSS contrôle (ctrl TOSS), son effet est similaire à celui observé par l'effet de la LF seule (Fig. 1), démontrant ainsi le recours à une séquence antisens spécifique pour obtenir l'effet de modulation.

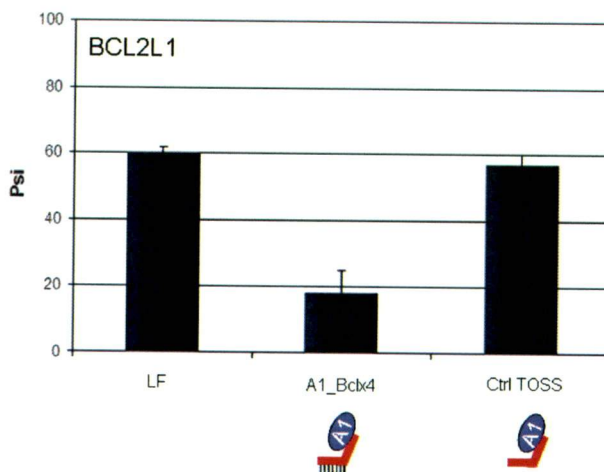


Figure 1 Effet de A1_Bclx4 sur la modulation de l'épissage de BCL2L1 *in cellulo*.

Des cellules PC-3 ont été transfectées pendant 24 h par 480 nmol de A1_Bclx4 ou de TOSS contrôle (ctrl TOSS). L'ARN total a été extrait au Trizol®. Les histogrammes présentent les valeurs moyennes de ψ obtenues par endpoint PCR pour trois réplicats biologiques. L'amplitude de la modulation est obtenue en comparant les valeurs de A1_Bclx4 et LF.

Bien que cet effet de reprogrammation de l'AS semble réellement découler de l'hybridation du TOSS sur sa cible pré-ARNm, il est possible que l'effet soit partiellement artificiel dû à son hybridation sur l'isoforme long (ARNm). Effectivement, il a déjà été rapporté dans la littérature que les oligonucléotides comportant des modifications de type 2'OMe peuvent efficacement bloquer l'élongation d'enzymes de transcription inverse virale (Boiziau *et al.*, 1996). Les TOSS étant constitués d'oligonucléotides modifiés de type 2'OMe, il est probable que ceux-ci puissent également bloquer la réaction de transcription inverse utilisée dans l'essai endpoint PCR. Comme le TOSS ne peut s'hybrider que sur l'isoforme long, cela conduirait à une sous-évaluation et une surévaluation de l'abondance de l'isoforme long et court, respectivement (Fig. 2A). Cette hypothèse fut testée avec A1_Bclx4 et deux autres TOSS modèles (A1_APAF12 et A1_CAPN3_A). Chacun des TOSS est chauffé en présence d'ARN total de HeLa S3, puis lentement refroidi afin d'hybrider le TOSS (4 nM – 4 μ M final) à sa cible. La réaction PCR est effectuée et les valeurs de ψ sont calculées et comparées à un TOSS contrôle ne possédant pas la séquence de reconnaissance spécifique à l'ASE à l'étude. Dans tous les cas, on note un effet artificiel dose-dépendant pouvant atteindre près de 30% à 4 μ M et se faire sentir à des concentrations aussi basses que 4 nM (Fig. 2B). Ceci démontre qu'une faible quantité de TOSS présente dans l'extrait d'ARN total induit une modulation artificielle de l'AS *in vitro*.

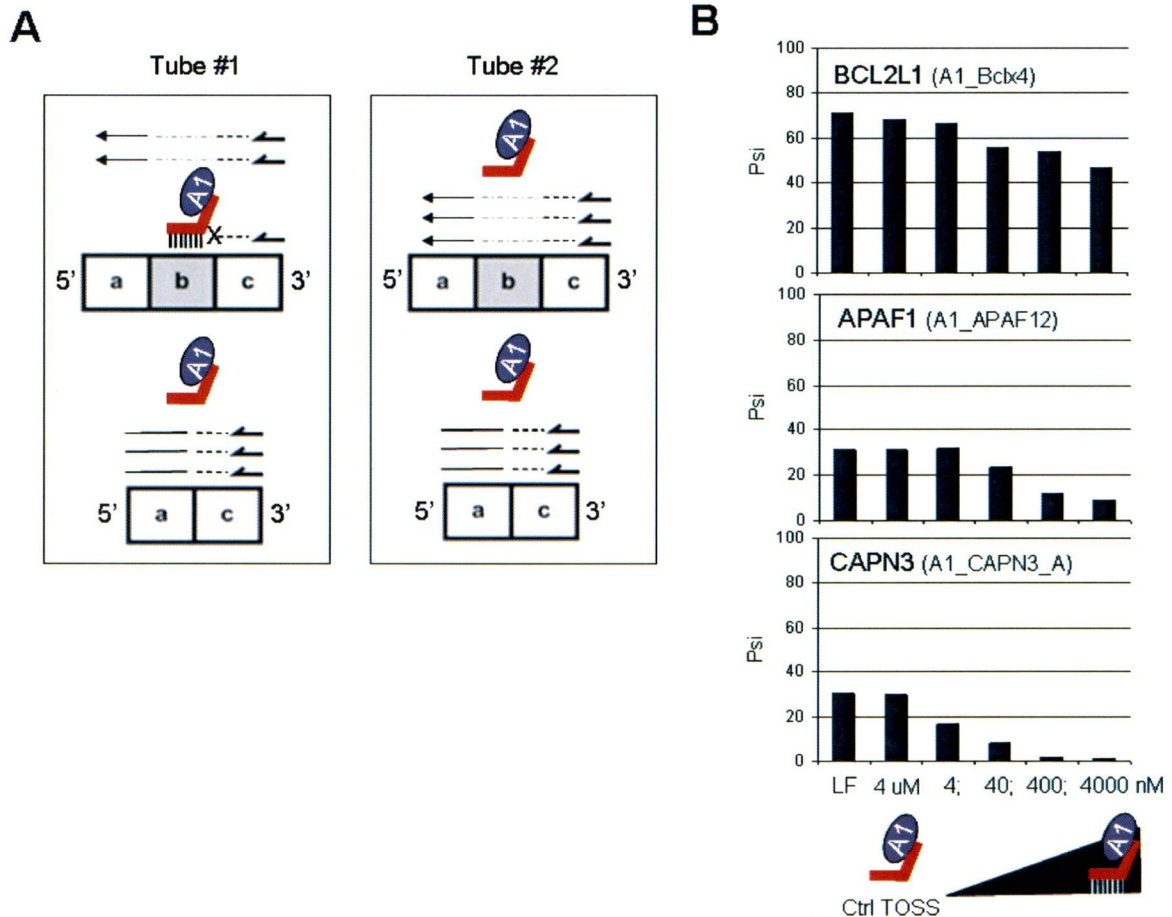


Figure 2 Blocage de la transcription inverse par les oligonucléotides 2'OMe.

A) Schéma explicatif de la conséquence du blocage de la transcription inverse par le TOSS. Dans le tube #1, un oligonucléotide 2'OMe (TOSS) hybridé sur l'exon alternatif de l'isoforme long (exon b) bloque partiellement la synthèse d'ADNc. Cet oligonucléotide ne possède pas de séquence complémentaire à l'isoforme court. Dans le tube #2, un oligonucléotide contrôle ne s'hybridant sur aucune séquence d'ARNm ne bloque pas la synthèse d'ADNc. Le résultat net postulé de la sous-évaluation de l'abondance de l'isoforme long conduirait à une diminution du ψ , soit le même résultat qu'une réelle modulation par le TOSS *in cellulo*. B) Démonstration expérimentale du bloc du TOSS *in vitro*. Les histogrammes présentent la valeur de ψ obtenue par endpoint PCR avec des extraits d'ARN total de HeLa S3 traités avec des concentrations croissantes de TOSS choisis (A1_Bclx4, A1_APAF12 et A1_CAPN3_A). On note un effet artéfactuel dose-dépendant pouvant atteindre près de 30% à 4 uM et se faire sentir à des concentrations aussi basses que 4 nM. Ceci démontre qu'une faible quantité de TOSS présente dans l'extrait d'ARN total induit une modulation artéfactuelle de l'AS *in vitro*.

2.1.1.1 Quantification du TOSS résiduel post-transfection.

Bien que le changement de ratio induit artificiellement par le TOSS *in vitro* soit évident, plusieurs variables (dont la proportion de TOSS effectivement entrée dans les cellules et le pourcentage de TOSS dégradé post-transfection), rendent l'estimation de la concentration résiduelle de TOSS dans les échantillons d'ARN total difficile à évaluer. Dans le but de quantifier le TOSS résiduel 24 h post-transfection, un essai par buvardage de type Northern a été réalisé suivant une transfection par un TOSS. Des cellules HeLa S3 ont été transfectées par A1_Bclx4 (480 nmol). Une fois l'ARN total extrait au Trizol®, celui-ci est mis en présence de deux sondes d'oligonucléotides d'ADN radiomarqués au P³² en 5': la première possédant une complémentarité parfaite avec la portion « tail » du TOSS, la deuxième reconnaissant U6 snRNA. De façon surprenante, aucun TOSS n'a pu être détecté dans l'extrait d'ARN transfecté alors qu'une faible quantité de A1_Bclx4 exogène (5 pmol) et de U6 snRNA endogène sont détectables (résultats non montrés). Pour déterminer à quel moment de la procédure d'extraction d'ARN le TOSS devient indétectable, la transfection des cellules fut omise et huit extractions d'ARN furent menées en parallèle. A1_Bclx4 (480 nmol) a été ajouté (*in vitro*) une seule fois par extraction, chacune d'elle à différent point stratégique du protocole d'extraction. Chacun des huit extraits d'ARN total a été soumis aux conditions de détection décrites précédemment. Clairement, tous les extraits dont le TOSS a été ajouté après la séparation des phases aqueuse et organique (phénolique) révèlent la présence de TOSS (Fig. 3). Ces nouveaux résultats suggèrent que le TOSS serait retenu dans la phase phénolique plutôt que dans la phase aqueuse lors de l'extraction d'ARN. En résumé, cet essai démontre que l'extraction d'ARN à l'aide du Trizol® enlève le TOSS, permettant ainsi d'obtenir des extraits d'ARN sans artefact PCR.

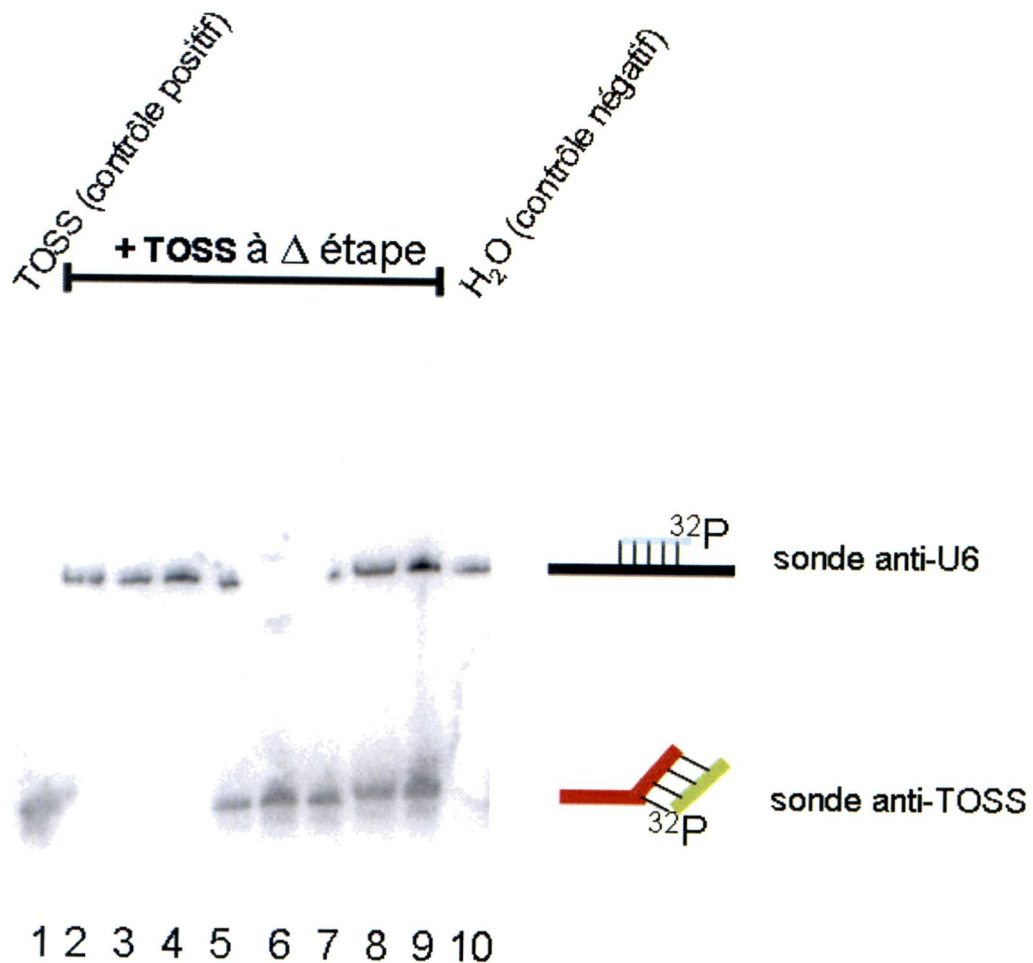


Figure 3 Le traitement au Trizol® enlève le TOSS d'un extrait d'ARN total.

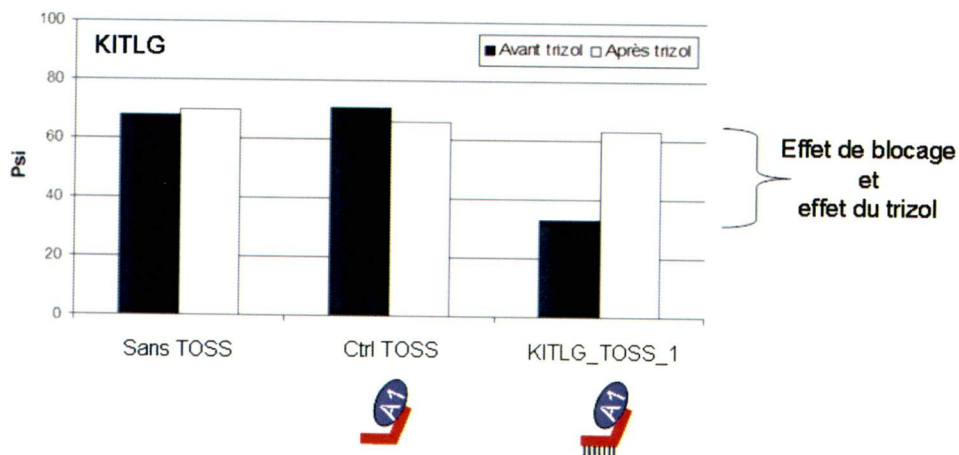
Des cellules HeLa S3 ont été ensemencées en plaque 6 puits en omettant la transfection. Huit extractions d'ARN furent menées en parallèle. A1_Bclx4 (480 nmol) a été ajouté (*in vitro*) une seule fois par extraction, chacune d'elle à différent point stratégique du protocole d'extraction au Trizol®. Piste 1: Diluer dans l'eau (contrôle positif). Piste 2: Ajout du TOSS suivant l'addition du Trizol® dans la plaque 6 puits. Piste 3: Ajout du TOSS suivant le transfert du Trizol® dans le microtube. Piste 4: Ajout du TOSS suivant l'addition de chloroforme. Piste 5: Ajout du TOSS suivant l'ajout de glycogène et d'isopropanol au surnageant. Piste 6: Ajout du TOSS suivant l'ajout de l'éthanol 75%. Piste 7: Ajout du TOSS sur le culot d'ARN avant le séchage. Piste 8: Ajout du TOSS lors de la resuspension de l'ARN total dans l'eau. Piste 9: Ajout du TOSS suite au chauffage de l'ARN total à 60°C. Piste 10: L'eau servant à la dilution du TOSS (contrôle négatif). Suivant la séparation sur gel de polyacrylamide dénaturant et le transfert sur membrane de nitrocellulose, celle-ci est mise en présence de sondes radiomarquées: l'une reconnaissant le petit ARN U6 et l'autre s'hybridant sur la portion « tail » de A1_Bclx4. On note que la détection du petit ARN non codant U6 est uniforme dans tous les puits où de l'ARN total est présent (pistes 2 à 10) alors que le TOSS est détecté uniquement lorsqu'il a été ajouté suivant la séparation des phases organique et aqueuse (pistes 5 à 9).

2.1.1.3 Le TOSS module vraiment l'épissage *in cellulo*.

Comme le Trizol® enlève le TOSS présent dans un extrait d'ARN total, une expérience montrant cette fois que la composante artéfactuelle due à la présence de TOSS peut être éliminée grâce à l'action du Trizol® a été entreprise. Brièvement, une fois les extraits d'ARN traités *in vitro* avec KITLG_TOSS_1 et le bloc confirmé par endpoint PCR tel qu'il a été décrit à la section 3.1.1, les extraits d'ARN ont été resoumis au protocole d'extraction au Trizol®. Les résultats endpoint PCR suivants cette expérience confirment clairement le pouvoir du Trizol® à enlever le TOSS, et donc, la composante artéfactuelle qui lui est associée (Fig. 4A). La même expérience a été reprise avec AXIN1_cass_t, BTC_cass_t, et SYK_TOSS_1 avec des conclusions similaires (Fig. 4B).

Bien que l'extraction d'ARN total au Trizol® est une méthode simple et peu coûteuse, celle-ci est difficilement envisageable à haut débit. Des procédures exploitant des colonnes de silice en format 96 puits comme le « Absolutely RNA 96 Microprep kit » de Stratagène sont commercialement disponibles. Dans le but de connaître la capacité de cette technique d'extraction d'ARN à se débarrasser des oligonucléotides 2'OMe tel le TOSS, des extraits d'ARN ayant été préalablement traités *in vitro* avec du TOSS (KITLG_TOSS_1, A1_Bclx4) ont été soumis au protocole d'extraction d'ARN tel que recommandé par Stratagène. Suivant l'essai endpoint PCR spécifique à chacun des ASEs, les résultats indiquent un effet de blocage et l'incapacité de l'extraction par colonne à contrer cet effet, et ce, dans les deux cas (résultats non montrés). Nous avons donc entrepris des variations du protocole original de Stratagène ayant comme objectif de réduire la complémentarité entre le TOSS et l'ARNm, mais en vain (lavage avec une solution à faible pH, chauffage suivi d'un refroidissement rapide; résultats non montrés). Pour contourner ce problème, des extraits d'ARN contenant du TOSS ont systématiquement été soumis à des températures croissantes de transcription inverse avec ou sans ajout de tampon Qiagen défavorisant la formation de structure secondaire. Les résultats indiquent un effet favorable du tampon Qiagen dans un cas sur six, alors que l'élévation de la température de transcription inverse semble diminuer la quantité de produit PCR et fait augmenter la variabilité (résultats non montrés).

A



B

	Sans TOSS		Ctrl TOSS		TOSS		Effet de blocage	Effet du trizol
	Avant	Après	Avant	Après	Avant	Après		
AXIN1_cass_t	13	21	13	19	2	17	11	15
BTC_cass_t	30	44	31	38	10	39	20	29
SYK_TOSS_1	62	62	63	61	29	58	33	29

Figure 4 Le Trizol® permet d'enlever la composante artificielle.

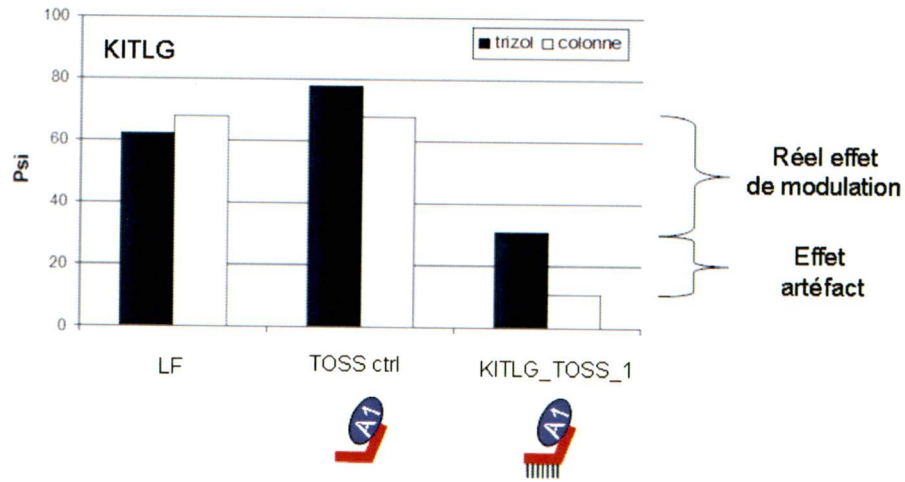
KITLG_TOSS_1 ou ctrl TOSS (480 nmol) a été ajouté (*in vitro*) à des extraits d'ARN de NIH-OVCAR-3. A) Les histogrammes en noir (Avant Trizol®) présentent les valeurs de ψ obtenues par endpoint PCR et confirment le blocage du TOSS. Les histogrammes en blanc (après Trizol®) présentent les valeurs de ψ obtenues par endpoint PCR suivant une soumission de chacun des extraits au protocole d'extraction au Trizol® et confirment l'effet du Trizol® à enlever la composante artificielle. B) La même expérience a été reprise avec AXIN1_cass_t, BTC_cass_t et SYK_TOSS_1 avec des conclusions similaires. L'effet de blocage est calculé en comparant les valeurs de ψ «Avant» du TOSS et sans TOSS. L'effet du Trizol® est obtenu en comparant les valeurs de ψ «Avant» et «Après» du TOSS. On remarque que les valeurs effet de blocage et effet du Trizol® sont semblables pour un TOSS donné.

Puisque l'extraction au Trizol® n'est pas envisageable à haut débit et qu'aucune alternative n'est totalement efficace afin d'éliminer la composante artéfactuelle, une expérience permettant d'estimer la composante artéfactuelle *in cellulo* a été mise au point. Des cellules NIH-OVCAR-3 ont été transfectées par KITLG_TOSS_1. Après 24 h, les cellules ont été décollées avec de la trypsine et séparées en deux parties égales. La première a été extraite selon la procédure au Trizol®, la deuxième selon la stratégie en colonne de Stratagène. Ensuite, la détection des isoformes a été effectuée telle que décrite précédemment. De cette façon, les données résultantes de la méthode au Trizol® seront considérées comme authentiques (réel effet de modulation) alors que celles provenant de la méthode par colonne seront analysées comme étant la somme du changement de ratio réel plus la composante artéfactuelle. Cette expérience a été répétée avec AXIN1_cass_t, BTC_cass_t et SYK_TOSS_1. Ainsi, la composante artéfactuelle atteint un $\Delta \psi$ ($\psi_{LF} - \psi_{TOSS}$) de près de 10 dans trois cas sur quatre et se chiffre à 20 dans le cas de KITLG_TOSS_1 (Fig. 5A). Comme l'amplitude de l'effet réel de modulation varie de 15 à 60, la composante artéfactuelle joue un rôle mineur dans le cas de fort modulateur (KITLG_TOSS_1) et possède un impact majeur dans le cas de faible modulateur (AXIN1_cass_t, Fig. 5B). Ainsi, il est acceptable d'utiliser la stratégie en colonne lors d'expérience à haut débit et de purifier les extraits au Trizol® advenant le cas de faible modulateur ($\Delta \psi < 25$).

Tel qu'illustré à la figure 2A, l'effet de blocage du TOSS ne s'applique qu'à l'isoforme long. Conséquemment, l'expression de l'isoforme court ne devrait pas varier en fonction du traitement au TOSS. Nous avons cherché à évaluer la variation de l'isoforme court en fonction du traitement au TOSS. Cette fois, il n'est plus possible d'utiliser les valeurs de ψ extraites des réactions endpoint PCR, car cette technique ne permet pas d'évaluer l'expression des isoformes un à la fois. La PCR en temps réel permet de mesurer l'isoforme court et l'isoforme long séparément. De plus, la quantification de l'expression globale (tous les isoformes ARNm d'un gène) devient possible par PCR en temps réel. Pour ce faire, l'ARN total extrait au Trizol® subit une transcription inverse en ADN complémentaire (ADNc) à l'aide d'amorces hexamériques dont les nucléotides sont dégénérés. L'ADNc est subséquemment dilué et introduit dans trois réactions PCR dont les amorces sont spécifiques à l'isoforme court, l'isoforme long ou ciblent une portion commune aux deux

ARNm (global). Pour des fins de normalisation, des réactions PCR sur deux gènes ne présentant que très peu de variation entre les échantillons à l'étude sont utilisées, soit un total de cinq réactions PCR par échantillon. La méthode d'analyse de qBase (Hellemans *et al.*, 2007) est appliquée pour générer les valeurs d'expression relative pour chacun des isoformes.

A



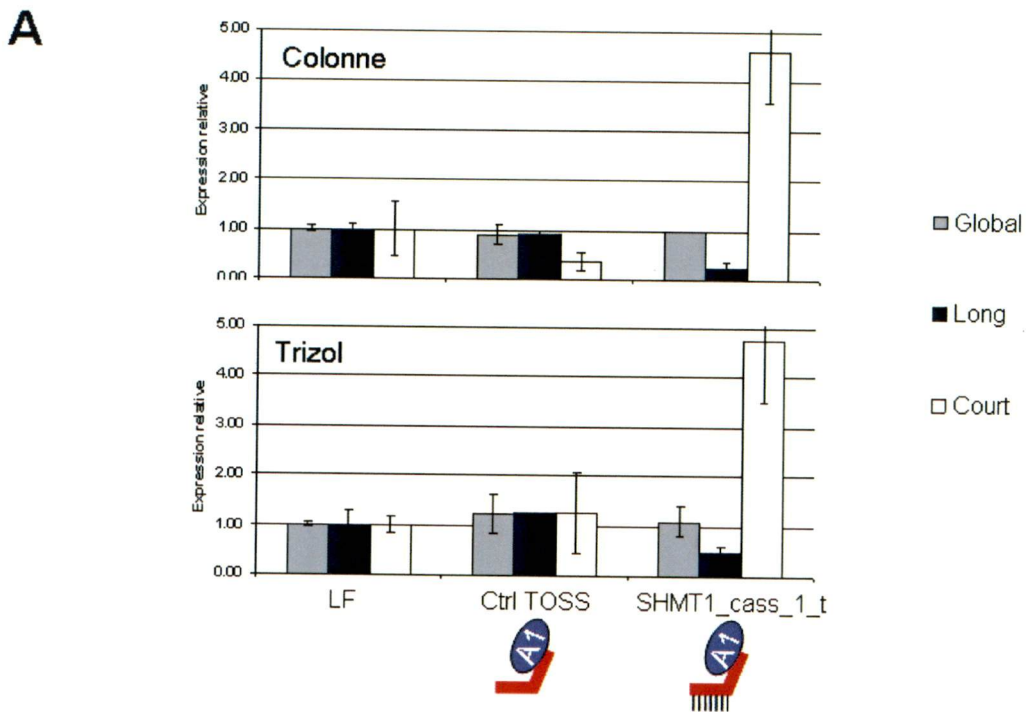
B

	LF		Ctrl TOSS		TOSS		Réel effet de modulation	Effet artéfact
	trizol	colonne	trizol	colonne	trizol	colonne		
AXIN1_cass_t	19	20	21	21	15	6	4	9
BTC_cass_t	41	41	42	44	14	7	27	7
SYK_TOSS_1	69	69	68	69	17	8	52	9

Figure 5 Estimation de la réelle proportion de la composante artéfactuelle *in cellulo*.

Des cellules NIH-OVCAR-3 ont été transfectées par KITLG_TOSS_1 et ctrl TOSS. A) Les histogrammes présentent les valeurs de ψ obtenues par endpoint PCR lorsque l'ARN est extrait au Trizol® (noir) ou selon la stratégie en colonne de Stratagène (blanc). B) La même expérience a été reprise avec AXIN1_cass_t, BTC_cass_t et SYK_TOSS_1 avec des conclusions similaires. L'effet de réelle modulation est calculé en comparant les valeurs de ψ du TOSS et de LF. La composante artéfactuelle est estimée en comparant les valeurs obtenues pour le TOSS pour les deux méthodes d'extraction. On note que l'effet artéfact représente une faible proportion de l'effet total dans trois cas sur quatre.

Pour démontrer les réels effets de modulation du TOSS par PCR en temps réel, des cellules SKOV3ip1 ont été transfectées par SHMT1_cass_1_t. Après 24 h, l'ARN total a été extrait selon la stratégie en colonne de Stratagène. Puis, 10 μ L d'ARN ont été subséquemment purifiés selon la méthode d'extraction au Trizol®. Finalement, la détection des isoformes a été effectuée par PCR en temps réel. Contrairement à l'isoforme long, on note que les niveaux d'isoformes courts ne sont pas affectés par SHMT1_cass_1_t, démontrant leur réelle modulation (Fig. 6A). La même expérience a été reprise avec FN1_TOSS_2, MCL1_TOSS_3, SYK_TOSS_1 et UTRN_cass_t avec des conclusions similaires (Fig. 6B).



B

	Colonne			Trizol		
	Global	Long	Court	Global	Long	Court
FN1_TOSS_2	1.00	0.15	2.06	1.25	0.69	1.96
MCL1_TOSS_3	0.99	0.53	18.0	0.94	0.52	17.9
SYK_TOSS_1	1.42	0.31	2.24	1.58	0.20	2.32
UTRN_cass_t	1.20	0.24	19.1	1.10	0.30	16.4

Figure 6 Confirmation du réel effet modulateur du TOSS *in cellulo* par PCR en temps réel.

Des cellules SKOV3ip1 ont été transfectées par SHMT1_cass_1_t et ctrl TOSS. A) Les histogrammes présentent les valeurs moyennes d'expression relative globale (gris), de l'isoforme long (noir) et de l'isoforme court (blanc) par rapport à l'exon 117 de SHMT1 par PCR en temps réel pour trois réplicats biologiques. Dans le haut, l'ARN a été extrait selon la stratégie en colonne de Stratagène. Dans le bas, les mêmes échantillons d'ARN ont été purifiés selon la méthode d'extraction au Trizol®. Contrairement aux niveaux d'isoformes longs, les niveaux d'isoformes courts ne sont pas affectés par SHMT1_cass_1_t. B) La même expérience a été reprise avec FN1_TOSS_2, MCL1_TOSS_3, SYK_TOSS_1 et UTRN_cass_t avec des conclusions similaires.

Si le TOSS module réellement la proportion d'isoformes ARNm, ceci devrait se refléter dans la quantité relative des isoformes protéiques. Pour le confirmer, nous avons choisi deux gènes dont les isoformes ciblés par un TOSS sont techniquement séparables par électrophorèse sur gel d'acrylamide, soit CASP8 et CASP10. L'immuno-buvardage indique une baisse de l'isoforme long et une hausse concomitante de l'isoforme court, démontrant le réel effet de modulation des TOSS.

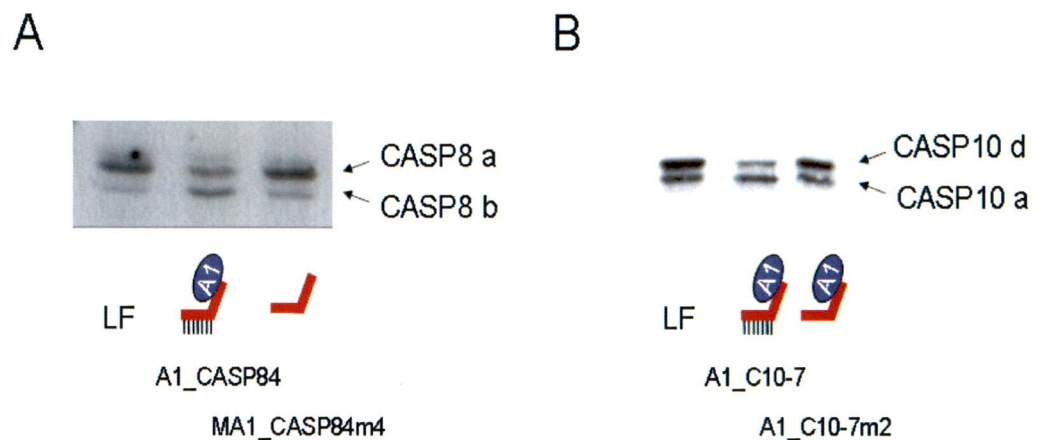


Figure 7 Effets des TOSS sur les isoformes de CASP8 et CASP10 au niveau protéique.

Des cellules HeLaS3 ont été transfectées par A) A1_CASP84 et A1_CASP84m4 (4 mésappariements dans la portion antisens) ainsi que B) A1_C10-7 et A1_C10-7m2 (2 mésappariements dans la portion antisens). Après 72h, les protéines ont été extraites et l'impact des TOSS sur le ratio d'isoformes a été vérifié qualitativement par western-blot.

2.1.1.4 Modulation de type A et de type B.

Parmi les TOSS testés jusqu'à présent, tous produisent une réelle modulation de l'AS, soit une diminution de l'isoforme long, une augmentation de l'isoforme court et aucun changement d'expression globale. Or, nous avons testé un TOSS additionnel ciblant l'exon cassette 273 de FN1. De façon surprenante, ce TOSS (FN1_TOSS_1) se comporte comme un siRNA spécifique à l'isoforme long. On observe une diminution de l'isoforme long sans augmentation compensatoire de l'isoforme court et une diminution marquée de l'expression globale (Fig. 8). Nous avons nommé ce comportement TOSS de type B, par opposition à l'effet modulateur attendu que nous appellerons maintenant type A. À noter que pour ce même évènement, FN1_TOSS_2 procure une modulation de type A.

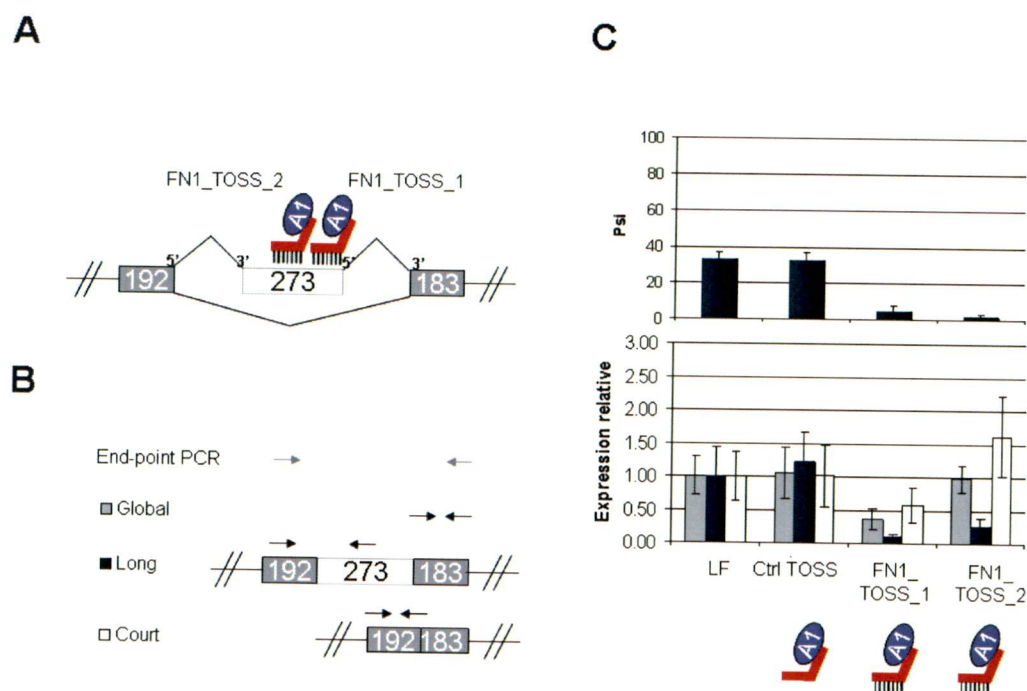


Figure 8 Modulation de type A et de type B.

A) Schéma représentant la position de FN1_TOSS_1 et FN1_TOSS_2 sur l'exon 273 de FN1. B) Schéma représentant la position des amorces PCR utilisées pour la détection des isoformes de FN1. C) Des cellules SKOV3ip1 ont été transfectées par FN1_TOSS_1, FN1_TOSS_2, ctrl TOSS, et l'ARN total a été extrait en colonne. C) Dans le haut, les histogrammes présentent les valeurs moyennes de ψ obtenues par endpoint PCR pour trois réplicats biologiques. Dans le bas, les histogrammes présentent les valeurs moyennes d'expression relative globale (gris), de l'isoforme long (noir) et de l'isoforme court (blanc) de FN1 par rapport à l'exon 273 par PCR en temps réel pour trois réplicats biologiques. FN1_TOSS_2 procure une réelle modulation (type A) alors que FN1_TOSS_1 se comporte comme un siRNA spécifique à l'isoforme long (type B).

2.1.1.5 Marche à suivre permettant la validation des TOSS à haut débit.

Les ASEs ciblés par les TOSS décrits dans ce chapitre ont été choisis en fonction de leur association avec le cancer de l'ovaire (Klinck *et al.*, 2008), du sein (Venables *et al.*, 2008) ou reliés à des cascades de signalisation apoptotique. Une grande proportion est décrite dans le chapitre 3 et l'ensemble des résultats sont comptabilisés dans l'annexe 1. Pour être en mesure de transfecter un grand nombre de TOSS en parallèle, nous avons entrepris la transfection de cellules par les TOSS en format 96 puits. Les cellules cancéreuses sont transfectées par les TOSS à 480 nmol (400 uM final) pour une période de 24 h en triplicats biologiques, c'est-à-dire selon trois journées différentes. Après 24 h, les cellules sont congelées à -80°C dans un tampon de lyse au fur et à mesure. L'isolation de l'ARN des trois réplicats est effectuée simultanément avec le Absolutely RNA 96 Microprep kit (colonne) avec de légères modifications au protocole original et inclut un traitement à la Dnase. L'intégrité de l'ARN est confirmée par Agilent Bioanalyzer en effectuant l'analyse du RNA integrity number (RIN) pour 10% des échantillons pris au hasard. Les valeurs sont typiquement de 9.0 et plus. Des échantillons possédant un RIN inférieur à 8.0 indiquent une détérioration de la qualité de l'ARN et les échantillons de cette plaque doivent être questionnés. La présence d'ADN génomique contaminant est effectuée en soumettant les échantillons d'ARN au test de présence d'ADN génomique. Les échantillons non-conformes indiquent qu'un traitement supplémentaire à la désoxyribonucléase (DNase) doit être fait ou les cellules doivent être transfectées à nouveau. La quantification est effectuée par spectrométrie (Nanodrop) et les valeurs se situent généralement entre 5 et 20 ng/uL.

Les échantillons peuvent être directement soumis au protocole de endpoint PCR. Pour ce faire, des amorces PCR sens et antisens doivent être dessinées sur les exons constitutifs flanquant la région alternative (Klinck *et al.*, 2008). Ensuite, l'amorce est utilisée pour la synthèse de l'ADNc selon une procédure de transcription inverse et PCR en une seule étape avec un volume fixe de 5,5 μL . Les produits PCR sont séparés par électrophorèse capillaire (Klinck *et al.*, 2008). Les valeurs de ψ sont rapportées pour chacun des échantillons et le $\Delta\psi$ ($\psi_{\text{LF}} - \psi_{\text{TOSS}}$) est utilisé pour connaître l'amplitude de modulation du TOSS.

Pour la quantification des isoformes par PCR en temps réel, une étape de synthèse d'ADNc avec des amorces hexamériques dégénérées est utilisée. Bien qu'il soit préférablement de normaliser les quantités d'ARN selon la concentration pour la synthèse d'ADNc, elle est effectuée sur un volume fixe (5,5 μ L) d'ARN. La normalisation via des gènes de référence (SDHA et PSMC4) dont le changement d'expression est minimal à travers les échantillons assure une normalisation adéquate. Les amorces isoforme-spécifiques sont dessinées, diluées et testées selon la procédure décrite au chapitre 1 (RASE pipeline). Pour les amorces globales, un algorithme permettant de déterminer la région la plus commune à tous les isoformes d'un gène est appliqué tel qu'il a été décrit au chapitre 3. Ces amorces sont testées selon les mêmes standards que les amorces isoforme-spécifiques. La figure 9 résume les étapes clés de la marche à suivre afin de valider les effets de modulation des TOSS à haut débit.

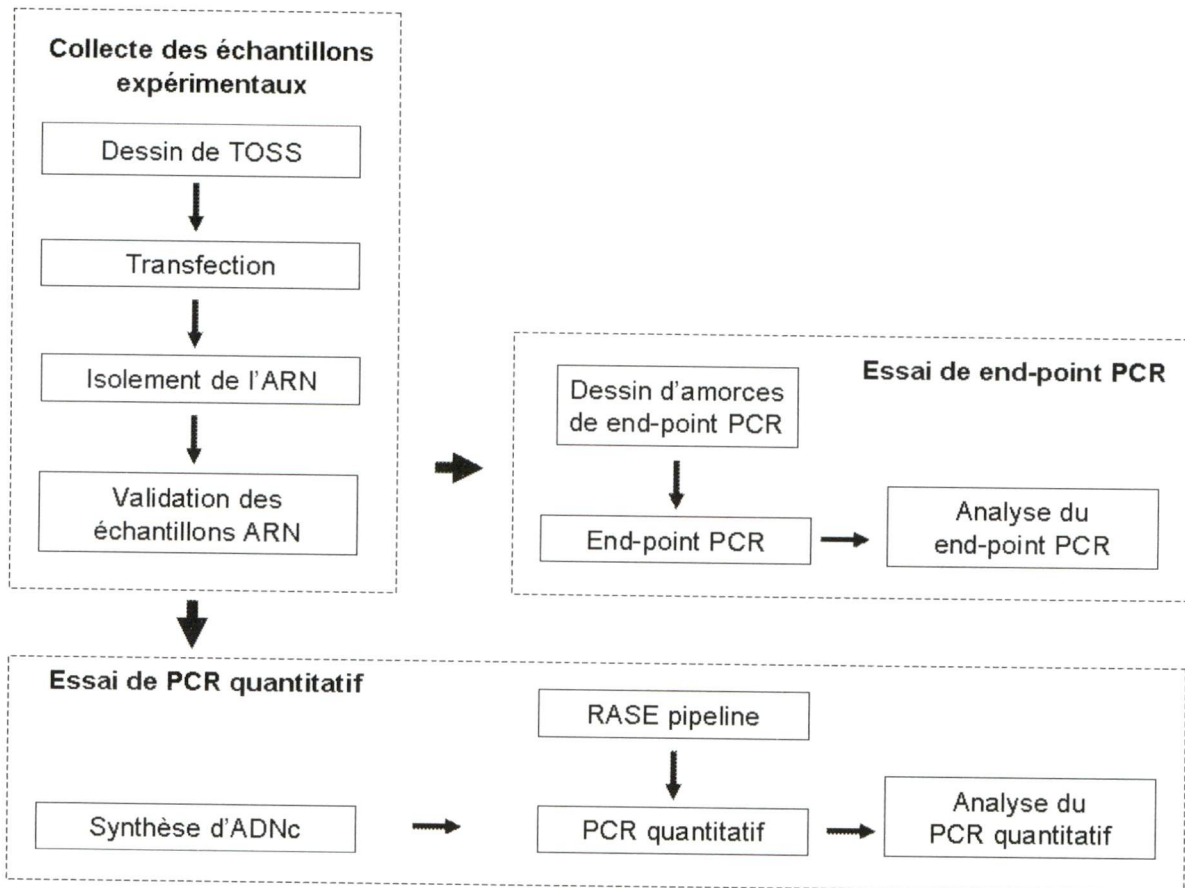


Figure 9 Marche à suivre permettant la validation des TOSS à haut débit.

Les TOSS sont dessinés tel que décrit au chapitre 3. Des cellules sont transfectées en format 96 puits pour une période de 24 h en triplicats biologiques par des TOSS (480 nmol). L'isolement de l'ARN des trois réplicats est effectué simultanément avec le Absolutely RNA 96 Microprep kit (colonne). L'intégrité et la quantité d'ARN sont vérifiées par Agilent Bioanalyzer et Nanodrop, respectivement. L'absence d'ADN génomique contaminant est confirmée avec le test de présence d'ADN génomique. Des amorces endpoint PCR sens et antisens sont dessinées tel que décrit par Klinck et collaborateurs (2008). Celles-ci sont utilisées pour la procédure de transcription inverse et PCR en une seule étape et les produits PCR sont subséquentement séparés par électrophorèse capillaire. Les valeurs de ψ sont rapportées pour chacun des échantillons et le $\Delta \psi$ ($\psi_{LF} - \psi_{TOSS}$) est utilisé pour connaître l'amplitude de modulation du TOSS. Pour la quantification des isoformes par PCR en temps réel, des amorces isoforme-spécifiques et globales sont dessinées, diluées et testées tel qu'indiqué au chapitre 1 (RASE pipeline). Ensuite, des amorces hexamériques dégénérées sont utilisées pour l'étape de synthèse d'ADNc. La quantification relative des isoformes d'intérêt se fait en normalisant via des gènes de référence dont le changement d'expression est minimal à travers les échantillons.

2.2.2 Dessin rationnel de TOSS.

2.2.2.1 Paramètres propres aux TOSS.

2.2.2.1.1 Complémentarité intra et intermoléculaire.

Les oligonucléotides ont tendance à se replier sur eux-mêmes en formant des paires de bases Watson-Crick afin de minimiser leurs énergies. Le TOSS ne faisant pas exception à cette règle, ce processus est susceptible d'interférer avec l'hybridation de la portion antisens sur la cible du TOSS et/ou la portion « tail » pour la protéine hnRNP A1. Afin de déterminer la complémentarité d'un TOSS avec lui-même, on utilise la différence d'énergie libre de Gibbs : plus la valeur est négative, plus l'oligonucléotide a tendance à se replier sur lui-même. Pour trouver la valeur seuil de $\Delta G_{\text{intramoléculaire}}$ au-dessus de laquelle les TOSS ont un repliement minimum, un script a été confectionné à l'aide de Unafold (Markham and Zuker, 2005). Au total, 10000 séquences de TOSS prises au hasard ont été évaluées afin de déterminer un $\Delta G_{\text{intramoléculaire}}$ moyen et la déviation standard. Nous avons établi la limite acceptable au $\Delta G_{\text{intramoléculaire}}$ moyen plus deux déviations standards, soit -9,4 kcal/mol.

La portion « tail », ayant pour objectif de recruter le facteur d'épissage hnRNP A1, possède deux séquences UAGGG de haute affinité (Burd and Dreyfuss, 1994). Or, les motifs GGG sont propices à la formation de structures secondaires de type G-quartet (Davis, 2004). Puisque la formation d'une telle structure demande l'association de quatre séquences GGG, celle-ci pourrait avoir lieu avec une association intermoléculaire entre deux molécules de TOSS. Lorsque la séquence antisens possède également deux séquences GGG, la formation d'un G-quartet intramoléculaire est alors favorisée. Pour stabiliser ce type de structure, des cations comme le Na^+ , mais préférablement le K^+ sont requis. Comme le milieu extracellulaire et intracellulaire possède des quantités généralement suffisantes de Na^+ et de K^+ , respectivement, il est possible que la concentration du TOSS disponible au site d'action soit modifiée par la capacité d'un TOSS à former des structures secondaires tel le G-quartet. Pour tester cette hypothèse, cinq TOSS comportant des aptitudes à former des structures secondaires et des capacités de modulation d'AS de différentes amplitudes ont été choisis. Chacun des TOSS a d'abord été purifié sur gel d'acrylamide, chauffé (100°C, 10 min) et rapidement refroidi dans l'azote liquide. Puis, un aliquot (150 μL , 15 μM) a été

dilué dans des tampons (200 μ L) mimant le milieu extracellulaire ($\text{Na}^+ = 110$ mM, $\text{K}^+ = 5$ mM) et intracellulaire ($\text{Na}^+ = 25$ mM, $\text{K}^+ = 133$ mM) selon les recommandations de Dasdia et collaborateurs (1979). Des spectres de dichroïsme circulaire ont été enregistrés à 37°C au temps = 0, 1, 4 et 24 h pour chacune des conditions expérimentales. Peu importe le temps d'acquisition, les spectres sont demeurés parfaitement superposables pour un TOSS donné, indiquant l'inertie du TOSS à former des structures secondaires, et ce, tant pour le tampon élevé en Na^+ que K^+ , (résultats non montrés). En somme, nous n'avons pas d'évidence que les TOSS forment des structures secondaires de type G-quartet. Toutefois, les séquences comportant des répétitions de GGGG sont à éviter dans le dessin de TOSS puisqu'elles sont difficiles à synthétiser.

2.2.2.1.2 Chimie de la portion « tail » du TOSS.

Tous les TOSS étudiés jusqu'ici sont composés entièrement de nucléotides modifiés de type 2'OMe. Il est important que la portion antisens soit composée de nucléotides modifiés afin de préserver son effet indépendant de la RNase H. Par contre, il a été rapporté que hnRNP A1 se lie autant aux séquences d'ADN que d'ARN (Ding *et al.*, 1999). Dans l'optique d'améliorer l'effet de modulation du TOSS, nous avons systématiquement varié la chimie de la portion « tail » (2'OMe, ADN, ARN) et mesuré l'impact sur la modulation de l'ASE modèle de KITLG. Des cellules SKOV3ip1 ont été transfectées tel que décrit précédemment et l'ARN total extrait au Trizol®. Tel qu'illustré à la figure 10, la version de la portion « tail » originale en 2'OMe demeure la plus puissante pour moduler l'AS.

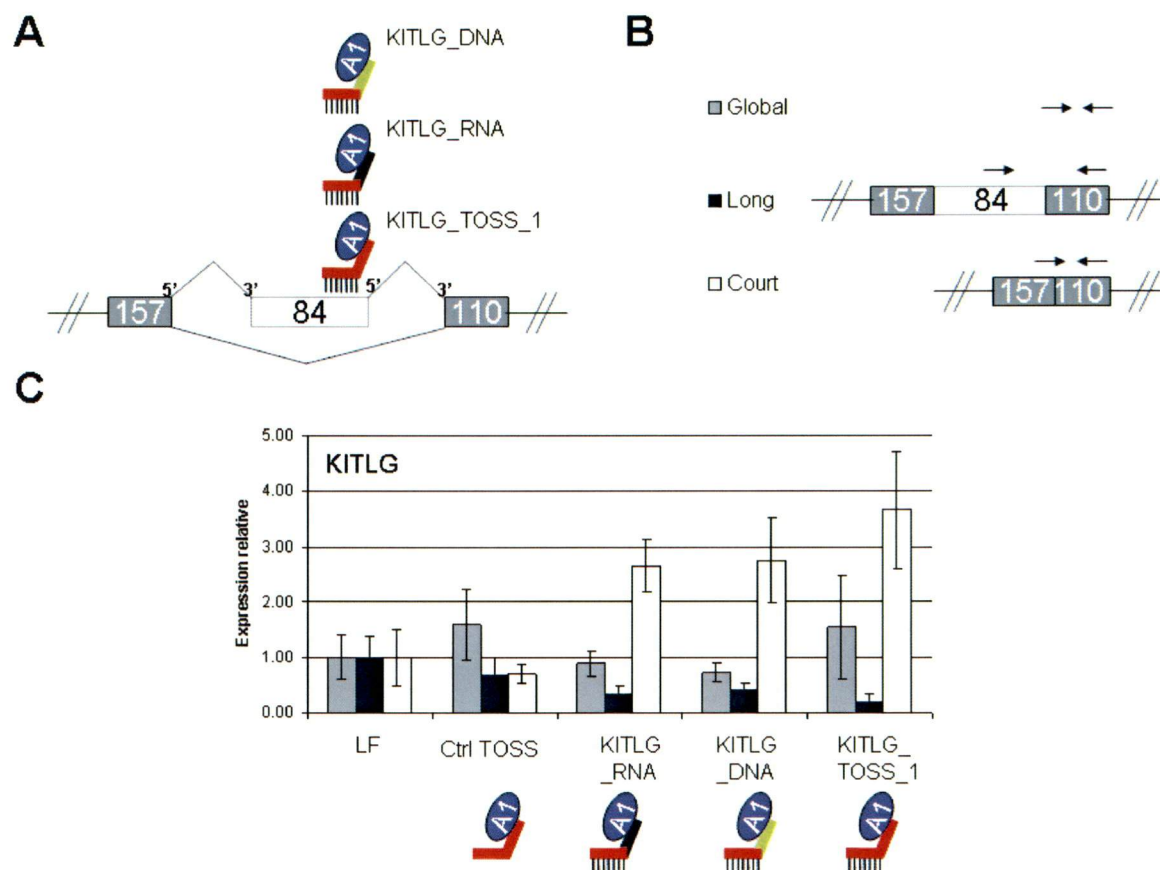


Figure 10 Effet de la modification de la chimie de la portion « tail ».

A) Schéma représentant la position de KITLG_TOSS_1 sur l'exon 84 de KITLG. Deux autres TOSS de séquence identique, mais dont la chimie est différente: KITLG_RNA (ARN) et KITLG_DNA (ADN) ont été dessinés. B) Schéma représentant la position des amorces PCR utilisées pour la détection des isoformes de KITLG. C) Des cellules SKOV3ip1 ont été transfectées par KITLG_TOSS_1, KITLG_RNA, KITLG_DNA, ctrl TOSS et l'ARN total a été extrait au Trizol®. Les histogrammes présentent les valeurs moyennes d'expression relative globale (gris), de l'isoforme long (noir) et de l'isoforme court (blanc) par rapport à l'exon 84 de KITLG par PCR en temps réel pour trois réplicats biologiques. La version de portion « tail » originale en 2'OMe demeure la plus puissante pour moduler l'épissage.

2.2.2.1.3 Effet de la séquence de la portion « tail » du TOSS.

De façon générale, il existe deux grandes catégories de SSOs définies selon le positionnement. Le plus souvent, l'antisens est situé directement sur la jonction exon-intron ou encore complètement exonique. Dans ce dernier cas, il est postulé que celui-ci entre en compétition avec un facteur d'épissage pour un site de liaison ayant un impact sur le ratio d'AS. Comme le TOSS est lui aussi exonique, bloquant potentiellement un facteur d'épissage quelconque, il est possible que l'effet de modulation du TOSS provienne

majoritairement de la portion antisens. Pour démontrer clairement la contribution de la portion « tail » du TOSS, une série de 10 TOSS ainsi que leur portion antisens correspondante ont été testés en cellules. Parmi les neuf TOSS qui fonctionnent, huit ont un effet de modulation équivalent ou supérieur à l'antisens (Tableau 1). Cela indique le rôle prépondérant que joue la portion « tail » du TOSS.

Tableau 1 Effet de la portion « tail » du TOSS.

Des cellules SKOV3ip1 ont été transfectées par 10 TOSS et leur portion antisens (ASO) et l'ARN total a été extrait en colonne. Les valeurs moyennes d'expression relative globale, de l'isoforme long et de l'isoforme court sont obtenues par PCR en temps réel représentant un triplicats biologique. Aug. = l'isoforme est détecté, mais non quantifiable. NON = Sans effet. A = modulation de type A. B = modulation de type B.

Oligonucléotide	PCR en temps réel			Verdict
	Global	Court	Long	
C11ORF17_ASO_1	0.90	1.05	0.16	NON
C11ORF17_TOSS_1	0.85	1.46	0.09	A
CHEK2_ASO_1	0.87	0.81	0.40	NON
CHEK2_TOSS_1	0.78	1.46	0.34	A
FGFR1_ASO_1	1.19	1.66	0.05	A
FGFR1_TOSS_1	0.95	1.57	0.06	A
FGFR2_ASO_1	0.81	0.90	0.83	NON
FGFR1_TOSS_1	1.04	0.82	0.60	NON
CCNE1_ASO_1	1.05	Aug.	0.40	A
CCNE1_TOSS_1	1.32	Aug.	0.72	A
KITLG_ASO_1	1.10	2.25	0.20	A
KITLG_TOSS_1	0.92	4.00	0.12	A
FN1_ASO_2	0.76	1.01	0.36	A
FN1_TOSS_2	1.00	2.06	0.15	A
MCL1_ASO_3	1.15	2.08	0.87	NON
MCL1_TOSS_3	0.99	18.04	0.53	A
SYK_ASO_1	0.34	0.82	0.14	B
SYK_TOSS_1	0.63	1.20	0.03	A
LIG4_ASO_1	0.88	1.82	0.06	A
LIG4_TOSS_1	0.74	4.24	0.14	A

Pour démontrer la contribution de hnRNP A1 dans l'effet de modulation du TOSS, Villemaire et collaborateurs (2003) avaient comparé A1_Bclx4 à un TOSS dont les deux séquences de liaison pour hnRNP A1 sont mutées (MA1_Bclx4). Ce dernier ne présente que très peu d'affinité pour une version écourtée de hnRNP A1 (UP1) comparativement à A1_Bclx4 *in vitro* et ne possède aucun effet notable sur la modulation de l'alt5' de BCL2L1 lorsque transfecté dans les cellules HCT-116 (Villemaire J., mémoire de maîtrise, Université de Sherbrooke). Nous avons non seulement cherché à confirmer ce résultat, mais avons également comparé l'effet d'un contrôle de la portion « tail » muté sur deux autres ASEs (exon 45 de CASP8 et exon 84 de KITLG). Dans tous les cas, l'effet d'un TOSS ayant la version de la portion « tail » mutée est en mesure de moduler l'AS, mais dans une plus faible mesure que la version originale (voir annexe 1). Dans un cas, une portion « tail » comportant un 5'ss a également été testée et donne des résultats similaires à la version mutée précédemment décrite (voir annexe 1). Cela permet d'affirmer que la version originale est celle permettant la plus grande modulation.

2.2.2.2 Paramètres propres à la cible

2.2.2.2.1 Effet de la position du TOSS vis-à-vis le site 5'.

Conceptuellement, il est postulé que le recrutement de hnRNP A1 via la portion « tail » du TOSS augmente l'effet stérique du TOSS et bloque l'accès de U1 snRNP à son 5'ss (Villemaire *et al.*, 2003). Conséquemment, l'AS a lieu préférentiellement sur le site le plus disponible. En ce sens, plus un TOSS est situé loin du site alt5', moins il devrait être en mesure de moduler l'AS. Nous avons vérifié cette hypothèse en variant la position du TOSS par rapport au site 5', et ce, pour deux événements d'exon cassette (Tableau 2). De façon générale, il n'y a pas de corrélation entre les valeurs de $\Delta\psi$ ($\psi_{LF} - \psi_{TOSS}$) et la distance en nucléotides par rapport au site 5'. Nous concluons en nuancant que l'action de modulation du TOSS ne passe pas nécessairement par un blocage stérique du site 5', mais plutôt par la présence de hnRNP A1 dans un contexte exonique.

Tableau 2 Effet de la position du TOSS par rapport au site 5' d'épissage.

Des cellules SKOV3ip1 ont été transfectées par six TOSS dont la distance varie vis-à-vis le site 5' et l'ARN total a été extrait en colonne. Les valeurs moyennes de $\Delta\psi$ ($\psi_{LF} - \psi_{TOSS}$) ont été calculées à partir des valeurs obtenues par endpoint PCR pour trois réplicats biologiques.

TOSS	Distance (nt)	$\Delta\psi$
MCL1_cass_1_t	-38	66
MCL1_cass_2_t	-31	61
MCL1_cass_t	-2	25
SHMT1_cass_1_t	-58	54
SHMT1_cass_2_t	-49	41
SHMT1_cass_t	0	52

Pour appuyer cette hypothèse, un TOSS (A1_KITLG_int8396_p15) dont la portion antisens s'hybride dans l'intron en aval de l'exon alternatif 84 de KITLG (+15 à +35) a été dessiné. Pour favoriser le blocage stérique, la portion « tail » du TOSS est positionnée en 3' (Gendron *et al.*, 2006). Contrairement au contrôle positif KITLG_TOSS_1, A1_KITLG_int8396_p15 n'a généré aucune modulation (voir annexe 1). L'ensemble de ces résultats converge vers un modèle permettant l'action du TOSS à distance et suggère sa présence exonique comme nécessaire pour la modulation.

2.2.2.2.2 Type d'évènement d'épissage.

Originellement, le TOSS a été employé pour moduler l'AS de l'alt5' de BCL2L1. Nous avons également ciblé l'alt5' de BMP4 à l'aide de deux TOSS. BMP4_345_2_t produit une forte modulation de l'isoforme court alors que BMP4_345_1_t génère une augmentation modeste de l'isoforme court. Dans les deux cas, l'isoforme long est diminué et l'expression globale n'est pas affectée (voir annexe 1). La reprogrammation d'un alt5' n'est donc pas exclusive à BCL2L1. Dans les paragraphes précédents, nous avons généralisé son emploi avec succès aux exons cassettes simples (BTC, C11orf17, CCNE1, CHEK2, FN1, KITLG, LIG4, MCL1, SHMT1, SYK) qui constituent en fait le type d'évènement alternatif le plus fréquent chez l'humain.

2.2.2.2.2.1 Exons cassettes multiples.

Le TOSS module efficacement la plupart des exons cassettes simples. Qu'en est-il lorsque deux exons alternatifs sont consécutifs? Est-ce que le TOSS reprogramme l'AS vers

l'isoforme le plus court possible (les exons alternatifs sont alors perçus comme un seul grand bloc d'exon alternatif) ou l'action du TOSS est restreinte à l'exon alternatif ciblé? Pour répondre à ces questions, trois évènements de cassette double (APP, FGFR2 et FANCA) ainsi que trois évènements de cassette triple (PTPN13, NRG1 et BCAS1) ont été sélectionnés. Pour chaque évènement, un TOSS a été dessiné sur le premier (FGFR2, FANCA) ou sur le dernier (APP, PTPN13, NRG1, BCAS1) exon alternatif. En plus des amorces de endpoint et de PCR en temps réel amplifiant spécifiquement l'exon alternatif ciblé par le TOSS, des amorces endpoint PCR circonscrivant l'ensemble des exons alternatifs consécutifs ont été conçues. Lorsque c'était possible, des amorces de PCR en temps réel ont été élaborées de façon à quantifier des isoformes potentiellement modulées par le TOSS. Dans le cas de FGFR2_TOSS_1 et NRG1_cass_L_t aucune modulation n'a été détectée. Seul PTPN13_cass_t a généré une modulation de type A alors que les actions de APP_cass_t, BCAS1_cass_t et FANCA_cass_L_t ont été classifiées comme étant de type B (voir annexe 1). Dans l'ensemble, la reprogrammation du bloc d'exon alternatif entier n'a jamais été observée. L'action du TOSS est donc spécifique à l'exon ciblé.

2.2.2.2.2 *Alt3'*.

Selon notre modèle actuel, le TOSS situé sur un exon alternatif agit en reprogrammant l'AS vers le site 5' en amont, favorisant ainsi une version d'isoforme plus courte. Qu'en est-il des alt3'? Des études ont démontré que hnRNP A1 peut créer de l'interférence avec U2 snRNP (Tange *et al.*, 2001, Martins de Araujo *et al.*, 2009). Il est alors possible que l'action du TOSS puisse influencer la décision d'épissage au 3'ss. Un TOSS (DRF1_alt3_L_t) a été dessiné sur l'exon 117 de DRF1 entre les deux sites alternatifs 3' (Fig. 11A). Avec des amorces circonscrivant ces deux 3'ss alternatifs, on note une baisse marquée de la valeur de ψ (Fig. 11C, en haut). Par contre, lorsqu'une paire d'amorces circonscrivant l'exon alternatif 117 en entier est utilisée, on note également une diminution de la valeur de ψ (Fig. 11C, au milieu). Cela est surprenant puisqu'aucune isoforme sans l'exon 117 en entier n'a été rapportée jusqu'ici (Thierry-Mieg and Thierry-Mieg, 2006). Des réactions de PCR en temps réel spécifique à l'isoforme long ou commune à l'isoforme long et moyen (global) indiquent une diminution de l'expression (Fig. 11C, en bas). Bien que l'isoforme court contenant l'exon 52 n'a pu être quantifiée adéquatement dû à sa faible expression, l'action

du TOSS ne fait pas augmenter pour autant son expression (résultat non montré). Tout cela suggère que DRF1_alt3_L_t agit comme modulateur de type B en réprimant l'exon alternatif 117 en entier, mais ne reprogramme pas vers le site alt3' en aval tel que souhaité.

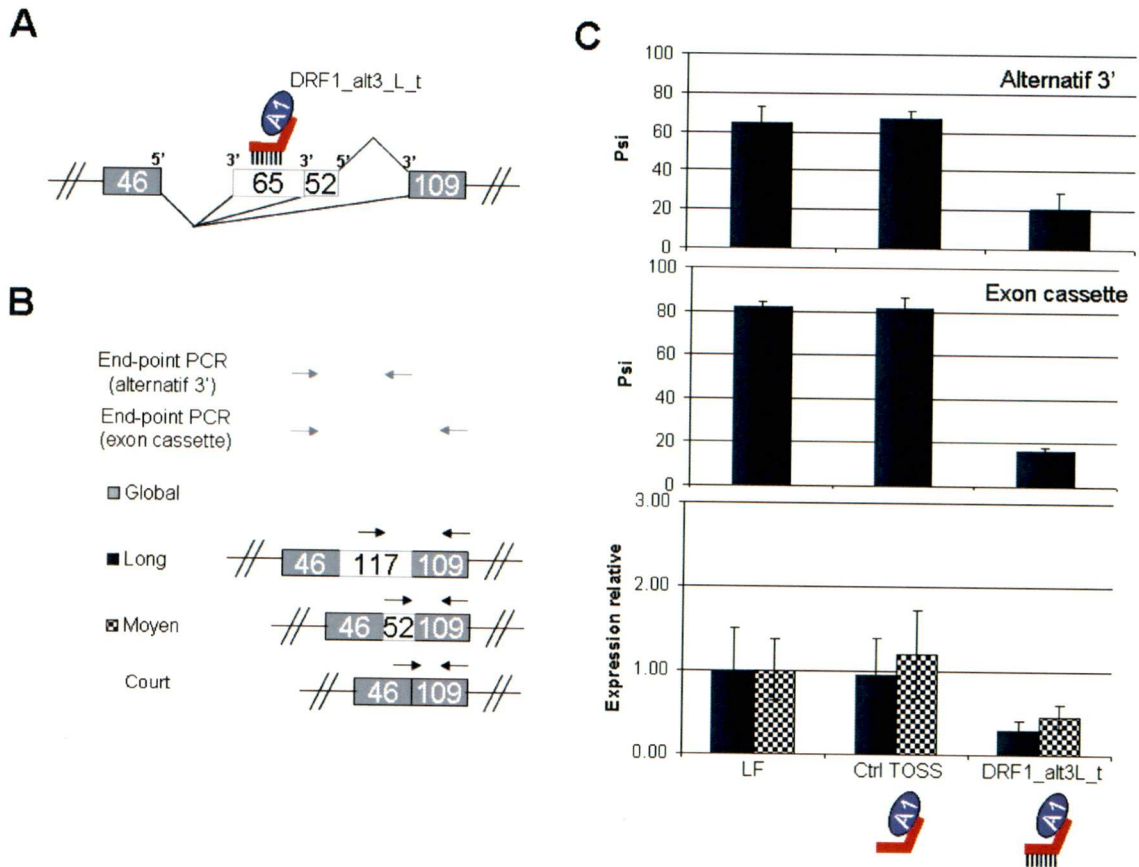


Figure 11 Modulation de l'alt3' de DRF1.

A) Schéma représentant la position de DRF1_alt3_L_t sur l'exon 117 de DRF1. B) Schéma représentant la position des amorces PCR utilisées pour la détection des isoformes de DRF1. C) Des cellules SKOV3ip1 ont été transfectées par DRF1_alt3_L_t, ctrl TOSS et l'ARN total a été extrait en colonne. Dans le haut (alternatif 3') et le milieu (exon cassette), les histogrammes présentent les valeurs moyennes de ψ par endpoint PCR pour trois réplicats biologiques. Dans le bas, les histogrammes présentent les valeurs moyennes d'expression de l'isoforme long (noir) et de l'isoforme moyen (carrelé) de DRF1 par rapport à l'exon 117 par PCR en temps réel pour trois réplicats biologiques. DRF1_alt3_L_t se comporte comme un siRNA spécifique à l'isoforme long (modulation de type B).

2.2.2.2.3 Cas complexe de type exon cassette-alt5'.

Comme le TOSS module l'AS tant sur des alt5' que sur les exons cassettes, nous nous sommes demandé quel serait le résultat de reprogrammation d'un TOSS situé sur un pré-ARNm comportant à la fois un alt5' et un exon cassette en compétition. Le gène NUP98 est

un exemple où le transcrit subit de l'AS pour produire trois isoformes. L'exon 222 peut être présent (long) ou non (court), mais il existe également un site alt5' à l'intérieur de l'exon 222, produisant l'exon 131 (moyen, Fig. 12A). Deux TOSS ayant des séquences antisens complètement différentes ont été dessinés pour cibler la région située entre les deux sites 5'ss alternatifs sur l'exon 222 (Fig. 12A). Lorsque des amorces de endpoint PCR circonscrivent entièrement l'exon 222, on remarque que l'effet de modulation est plus prononcé dans le cas de NUP98_cass_1_t comparativement à NUP98_cass_t (Fig. 12C, en haut). De plus, l'effet de ce dernier est nul sur la modulation du site alt5' alors que NUP98_cass_1_t parvient à le moduler (Fig. 12C, au milieu). Les effets de NUP98_cass_1_t sont confirmés par l'augmentation du transcrit court et une diminution du transcrit long, tel que quantifié par PCR en temps réel (Fig. 12C, en bas). De plus, les quantités basales de transcrit moyen sont indétectables, mais un traitement avec NUP98_cass_1_t permet de le détecter, corroborant la reprogrammation vers le site alternatif 5' produisant l'exon 131. Comme l'expression globale n'est pas affectée, NUP98_cass_1_t est un modulateur de type A. Les mêmes tendances sont observées pour NUP98_cass_t, mais dans une moindre mesure, corroborant les résultats endpoint PCR.

Dans le but de déterminer le rôle de la portion « tail » (donc indirectement de hnRNP A1) vis-à-vis l'antisens, des contrôles sans portion « tail » (portion antisens seulement) ont été dessinés. Ceux-ci sont en fait identiques aux deux TOSS précédemment décrits excepté qu'ils ne possèdent pas de séquence de recrutement hnRNP A1. Les expériences endpoint PCR confirment non seulement la capacité unique de NUP98_cass_1_t à moduler le site 5' alternatif, mais que cet effet dépend de la portion « tail » du TOSS (Fig. 12D). En conclusion, NUP98_cass_t promeut l'exclusion de l'exon 222 alors que NUP98_cass_1_t permet de moduler à la fois l'exon 222 et le site alt5'.

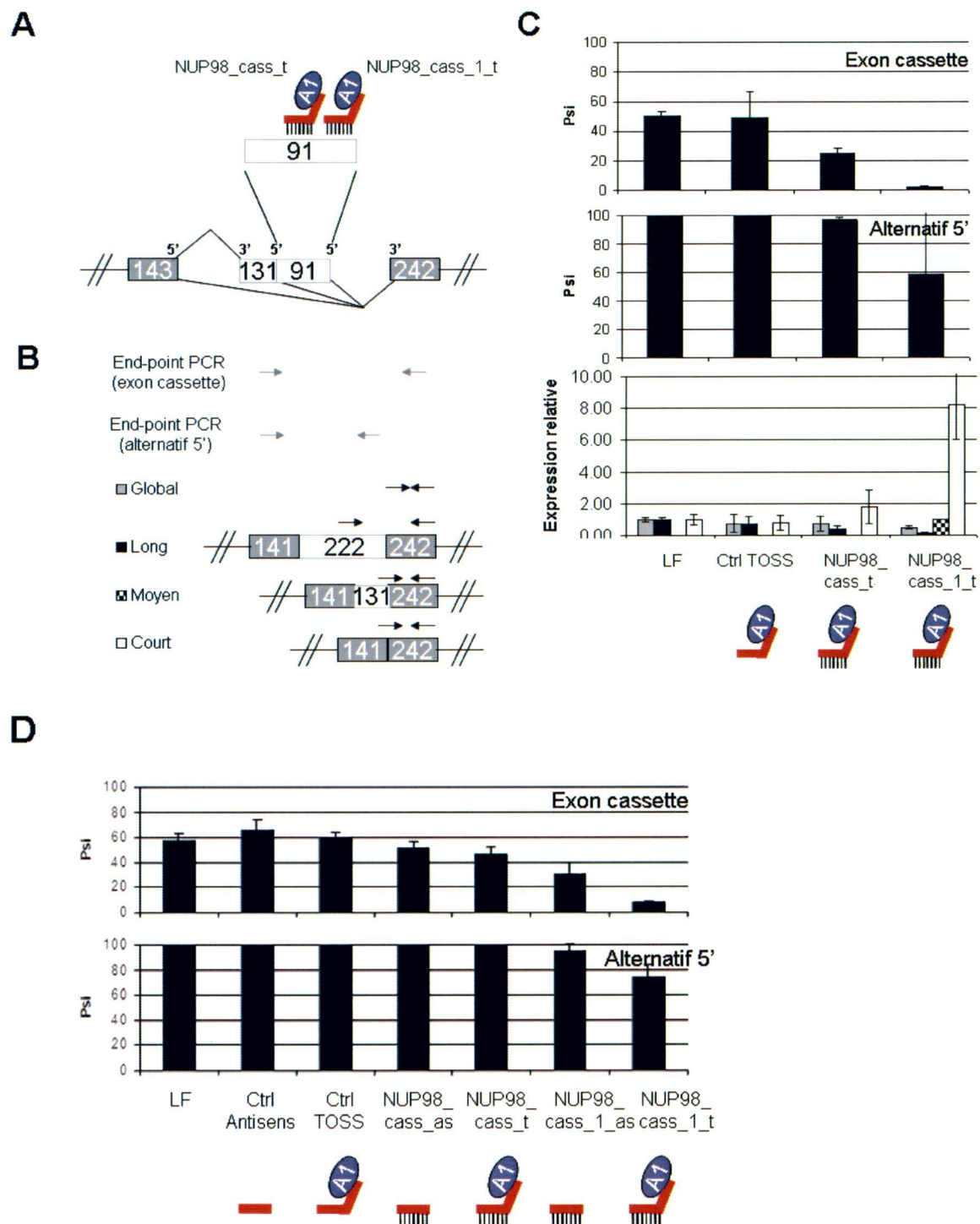


Figure 12 Modulation de l'évènement d'épissage alternatif complexe (exon cassette et alt5') de NUP98.

A) Schéma représentant la position de NUP98_cass_t et NUP98_cass_1_t sur l'exon 222 de NUP98. B) Schéma représentant les amorces PCR utilisées pour la détection des isoformes de NUP98. C) Des cellules SKOV3ip1 ont été transfectées par NUP98_cass_t, NUP98_cass_1_t, ctrl TOSS et l'ARN total a été extrait en colonne. Dans le haut (exon cassette) et le milieu (alternatif 5'), les histogrammes présentent les valeurs moyennes de ψ par endpoint PCR pour trois réplicats biologiques. Dans le bas, les histogrammes présentent les valeurs moyennes d'expression relative globale (gris), de l'isoforme long (noir), de l'isoforme moyen (carrelé) et de l'isoforme court (blanc) de NUP98 par rapport à l'exon 222 par PCR en temps réel pour trois réplicats biologiques. D) Des cellules SKOV3ip1 ont été transfectées par NUP98_cass_as, NUP98_cass_1_as, des versions sans portion « tail » des TOSS étudiés en C. Dans le haut (exon cassette) et le bas (alternatif 5'), les histogrammes présentent les valeurs moyennes de ψ par endpoint PCR pour trois réplicats biologiques.

2.2.2.2.4 Cas complexe de type exon cassette-alt3'.

Qu'arrive-t-il maintenant lorsqu'un site alt3' est en compétition avec l'exclusion d'un exon cassette? En nous basant sur les résultats acquis pour les gènes DRF1 et NUP98, nous spéculons qu'un TOSS situé sur l'exon cassette en aval du site alt3' produirait uniquement l'exclusion de l'exon alors qu'un TOSS situé entre les deux sites alternatifs 3' produirait une répression spécifique de l'isoforme le plus long. Pour le démontrer, nous avons d'abord choisi de dessiner un TOSS sur l'évènement complexe de TOPBP1. Comme la portion alternative entre les deux sites 3' ne fait que 15 nucléotides, le TOSS TOPBP1_alt3_L_t a été exceptionnellement dessiné à la jonction intron-exon de l'intron 715 et de l'exon 167 (Fig. 13A). La valeur de ψ calculée suivant la réaction endpoint PCR circonscrivant entièrement l'exon 167 indique un effet positif de TOPBP1_alt3_L_t (Fig. 13C, en haut). Par contre, les résultats de la réaction spécifique à l'évènement alt3' indiquent que TOPBP1_alt3_L_t n'a aucun effet sur le ratio isoforme long/isoforme moyen (Fig. 13C, au milieu). De plus, les résultats d'expression des isoformes long et moyen sont corroborés par PCR en temps réel (Fig. 13C, en bas). Cela est surprenant puisque TOPBP1_alt3_L_t n'est pas supposé affecter les niveaux de transcrit moyen. Un alignement de séquence entre la portion « tail » et la séquence exonique suivant le site alt3' suggère que le TOSS s'hybride également à l'isoforme moyen. Nous en concluons qu'il s'agit d'une répression des isoformes long et moyen de TOPBP1: une modulation de type B.

Nous avons également procédé à l'évaluation d'un autre cas d'AS complexe de type cassette-alt3': l'exon 551 du gène *LIG3*. Cette fois, nous avons dessiné deux TOSS; l'un en amont (*LIG3_cass_2_t*), l'autre en aval (*LIG3_cass_4_t*) du site alt3' de l'exon 551. Tel qu'il a été attendu, l'essai endpoint PCR spécifique à l'évènement alt3' démontre clairement que seul *LIG3_cass_2_t* possède un effet sur le ratio du long (exon 551) vis-à-vis le moyen (exon 331) alors que les deux TOSS indiquent des valeurs de ψ diminuées avec la réaction PCR spécifique à l'exon 551 entier (voir annexe 1).

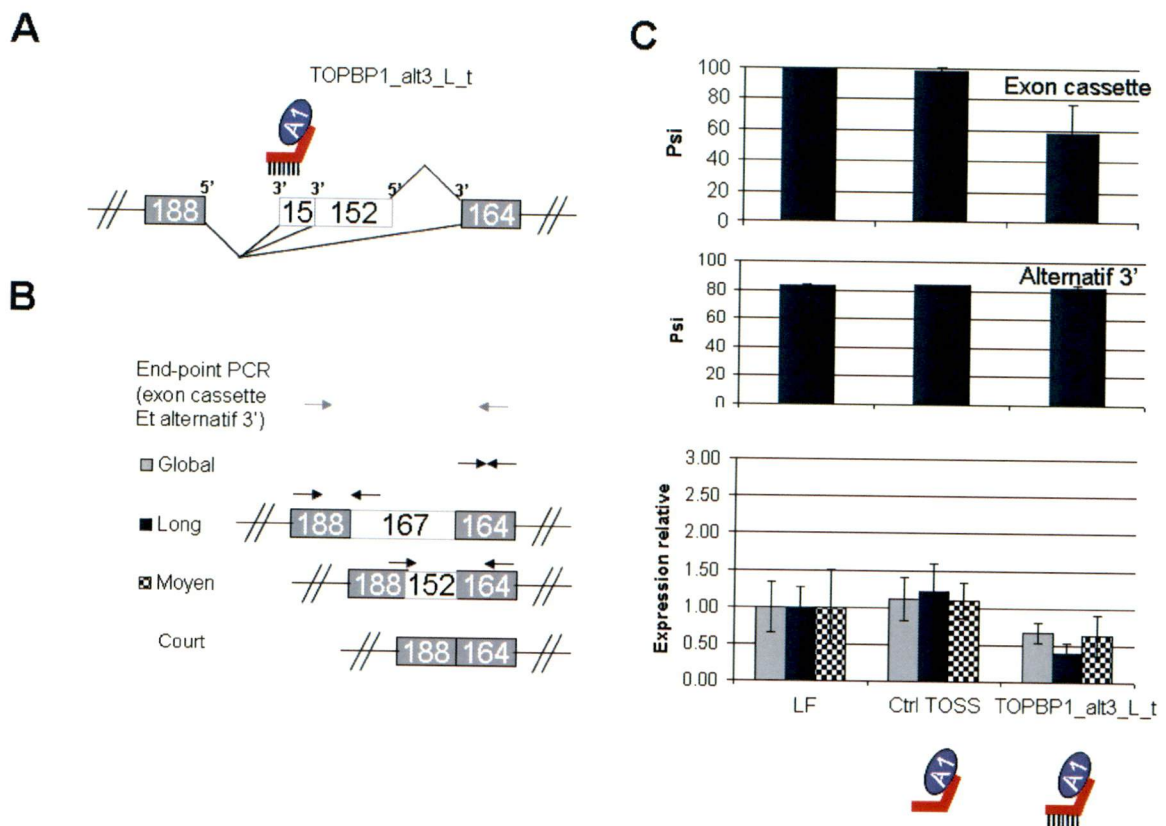


Figure 13 Modulation de l'évènement d'épissage alternatif complexe (cassette et alt3') de TOPBP1.

A) Schéma représentant la position de TOPBP1_alt3_L_t sur l'exon 167 de TOPBP1. B) Schéma représentant les amorces PCR utilisées pour la détection des isoformes de TOPBP1. C) Des cellules SKOV3ip1 ont été transfectées par TOPBP1_alt3_L_t, ctrl TOSS et l'ARN total a été extrait en colonne. Dans le haut (exon cassette) et le milieu (alternatif 3'), les histogrammes présentent les valeurs moyennes de ψ par endpoint PCR pour trois réplicats biologiques. Dans le bas, les histogrammes présentent les valeurs moyennes d'expression relative globale (gris), de l'isoforme long (noir) et de l'isoforme moyen (carrelé) de TOPBP1 par rapport à l'exon 167 par PCR en temps réel pour trois réplicats biologiques. Le TOSS TOPBP1_alt3_L_t procure une modulation de type B.

2.2.2.2.5 Cas complexe de type exon cassette-alt3'-rétention d'intron.

Outre l'exon cassette, les alt3' et alt5', il existe un autre type d'évènement simple: la rétention d'intron. Quelques exemples d'AS *in vitro* impliquent hnRNP A1 pour sa capacité à promouvoir l'épissage d'intron (Tange *et al.*, 2001, Wang *et al.*, 2006). De ce fait, nous émettons l'hypothèse qu'un TOSS situé dans l'intron pourrait augmenter l'efficacité d'épissage du 5'ss en amont et du 3'ss en aval. Pour illustrer la rétention d'intron, nous avons choisi l'évènement complexe (alt3', exon cassette et rétention d'intron) de FN1. Deux TOSS ont été dessinés; l'un exonique, en amont du 3'ss alternatif (FN1_alt3_L_t) et l'autre intronique, en amont du site 3'ss de l'exon 107 (FN1_intron_L_t, Fig. 14A). Dans un premier temps, on note que seul FN1_alt3_L_t module le site alt3' (Fig. 14C, en haut). Dans un deuxième temps, les deux TOSS font chuter le ratio isoforme long/isoforme moyen sans intron (Fig. 14C, au milieu). Dans le cas de FN1_alt3_L_t, ceci est dû à une répression des deux isoformes contenant la portion alternative 3' (isoforme long et isoforme moyen sans intron) tel qu'il a été montré par PCR en temps réel (Fig. 14C, en bas). Dans le cas de FN1_intron_L_t, l'expression globale n'est pas affectée et les niveaux d'ARNm de l'isoforme long sont diminués, concluant avec l'inhibition de la rétention d'intron (Fig. 14B, en bas). Par contre, nous avons été en mesure de détecter l'isoforme court seulement qu'en présence de FN1_intron_L_t. De plus, l'exclusion du bloc en entier (exon 467 au complet) n'a pas été observée pour aucun des TOSS (résultats non-montrés). Cela signifie que le TOSS FN1_intron_L_t agit non pas sur la rétention d'intron, mais sur l'exon cassette en aval, permettant l'exclusion de l'exon.

Par la suite, nous avons tenté de cibler deux autres gènes ayant des événements de rétention d'intron, soit POLI (POLI_intron_L_t) et TSSC4 (TSSC4_intron_L_t). Dans les deux cas, l'essai endpoint PCR avec des amorces circonscrivant l'intron alternatif permet d'observer une diminution de la valeur de ψ lorsque des cellules SKOV3ip1 sont traitées avec ces TOSS ciblant l'intron. Cela suggère que le TOSS favorise l'épissage de l'intron. Par contre, il n'a jamais été possible de quantifier l'augmentation de l'isoforme épissé (court) par PCR en temps réel, et ce, dans les deux cas. En fait, seul l'isoforme long de TSSC4 a pu être quantifié, mais celui-ci n'est pas affecté par le traitement de TSSC4_intron_L_t. Quant à la variation de l'expression globale, elle demeure inchangée (voir annexe 1). Vu les résultats

contradictoires, il n'est donc pas possible de conclure que les TOSS permettent une modulation d'évènement de rétention d'intron.

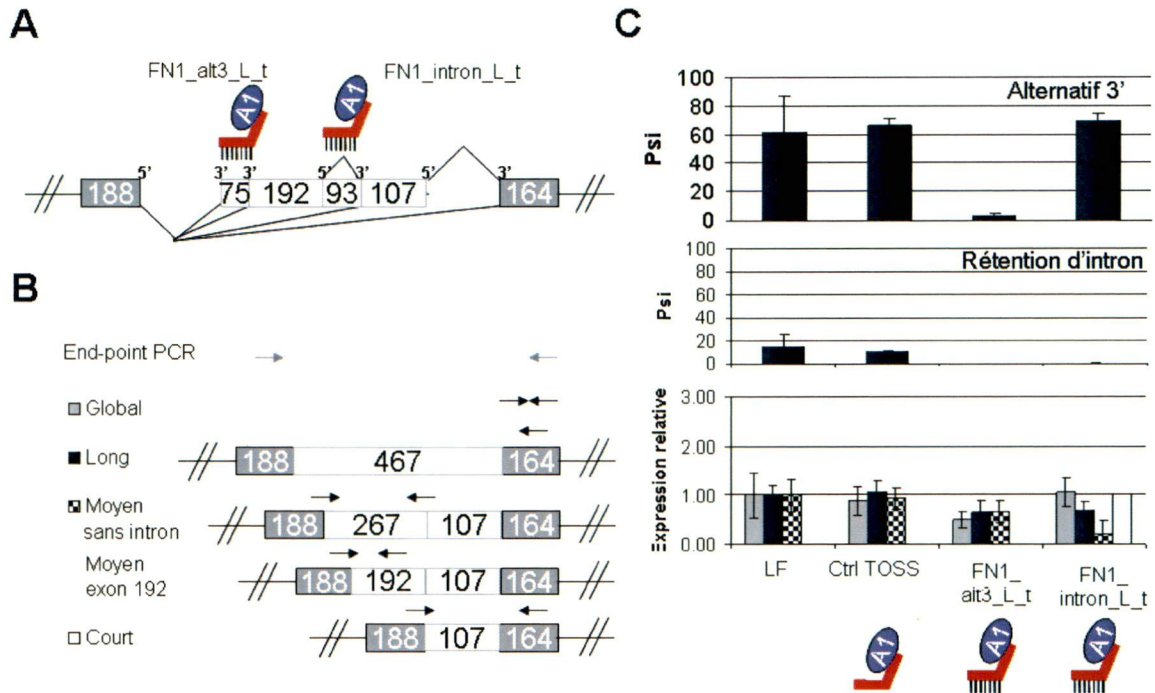


Figure 14 Modulation de l'évènement d'épissage alternatif complexe (exon cassette, alt3' et rétention d'intron) de FN1.

A) Schéma représentant la position de FN1_alt3_L_t, FN1_intron_L_t, ctrl TOSS sur l'exon 467 de FN1. B) Schéma représentant les amorces PCR utilisées pour la détection des isoformes de FN1. C) Des cellules SKOV3ip1 ont été transfectées par FN1_alt3_L_t et FN1_intron_L_t et l'ARN total a été extrait en colonne. Dans le haut (alternatif 3') et le milieu (rétention d'intron), les histogrammes présentent les valeurs moyennes de ψ par endpoint PCR pour trois réplicats biologiques. Dans le bas, les histogrammes présentent les valeurs moyennes d'expression relative globale (gris), de l'isoforme long (noir), de l'isoforme moyen sans intron (carrelé) et de l'isoforme court (blanc) de FN1 par rapport à l'exon 467 par PCR en temps réel pour trois réplicats biologiques.

2.3 Développement de stratégie générale réprimant l'expression de l'isoforme court

2.3.1 TOES dépendant de SF2

Le TOSS reprogramme les exons cassettes et les alt5' avec un haut taux de succès en favorisant l'isoforme court. Nous nous sommes demandé s'il était possible de concevoir un outil aussi général favorisant cette fois l'inclusion d'exon (l'isoforme long). Cet outil pluriel est d'autant plus attrayant puisqu'il est susceptible de procurer un gain de fonction. Des exemples sporadiques sur le développement de stratégies favorisant l'inclusion d'exon ont été rapportés dans la littérature, la plupart ciblant le gène survival of motor neuron 2 (SMN2) (Skordis *et al.*, 2003, Cartegni and Krainer, 2003, Hua *et al.*, 2007, Madocsai *et al.*, 2005). Nous nous sommes inspirés du dessin de TOES (targeted oligonucleotide enhancer of splicing) de Skordis et collaborateurs (2003), (même chimie, portion « tail », longueur d'antisens et position sur l'exon) pour cibler deux ASEs dont l'exclusion de l'exon est associée avec le cancer de l'ovaire (Klinck *et al.*, 2008). Dans les deux cas, l'essai endpoint PCR indique une augmentation du ψ lorsque des cellules cancéreuses sont transfectées par nos TOES (Fig. 15A). Comme les modifications PS procurent des effets hors cibles notoires (Levin, 1999), nous avons contrôlé cette expérience avec des TOES de même séquence, mais modifié complètement de 2'OMe (même chimie que le TOSS, en omettant toutes modifications PS). Les résultats sont marquants: pour les deux gènes testés, l'effet d'inclusion est aboli suite au traitement des TOES sans PS (Fig. 15A).

Il est possible que la perte de modulation de l'AS soit directement reliée à la perte d'affinité de SF2 pour la portion « tail » en 2'OMe. Pour répondre à cette question, nous avons fait synthétiser des oligonucléotides de différentes chimies (ADN, ARN et 2'OMe) avec ou sans phosphorothioate (PS) correspondant à la séquence originale de la portion « tail » du TOES. Une quantité de SF2-GST correspondante au EC₅₀ (165 nM final) a été incubée avec chacun des oligonucléotides 10 min dans un tampon HEPES avant d'être séparée sur gel d'acrylamide 4% non-dénaturant. Dans cet essai de retard sur gel, nous avons observé que la portion « tail » a plus d'affinité pour l'ARN et le 2'OMe, mais pas pour l'ADN (ARN \cong 2'OMe \gg ADN). L'affinité de liaison avec SF2 est légèrement majorée par les modifications PS (Fig. 5B). Ceci suggère que les versions 2'OMe et ARN-PS des TOES de

la figure 5A se lie de manière similaire à SF2, ce qui n'explique donc pas la différence d'impact au niveau des ASEs ciblés. En effet, des cellules SKOV3ip1 transfectées par une version de SF2_KITLG_2'OMe dont la portion « tail » est uniquement constituée d'ARN (SF2_KITLG_RNA) ne procurent pas les effets d'inclusion de son homologue SF2_KITLG_PS (résultats non montrés). Ce qui tend à montrer que les effets de modulation de l'AS sont dépendants des modifications PS. En contrepartie, les oligonucléotides en ARN simple brin sont plus propices à la dégradation que ceux modifiés 2'OMe et/ou PS. Il est donc possible que l'absence d'effets de SF2_KITLG_RNA soit principalement due à sa dégradation précoce. Or, nous avons déjà démontré à la figure 9C qu'un TOSS dont la portion « tail » est en ARN peut procurer des effets sur l'AS qui se situent à mi-chemin entre l'effet de l'antisens résiduel et celui de la version en 2'OMe (Tableau 1). Ceci suggère qu'une portion « tail » en ARN est en mesure de résister suffisamment à la dégradation pour induire un effet sur l'AS. Conséquemment, l'ensemble des résultats suggère que les modifications PS sont responsables des effets sur l'inclusion d'exon.

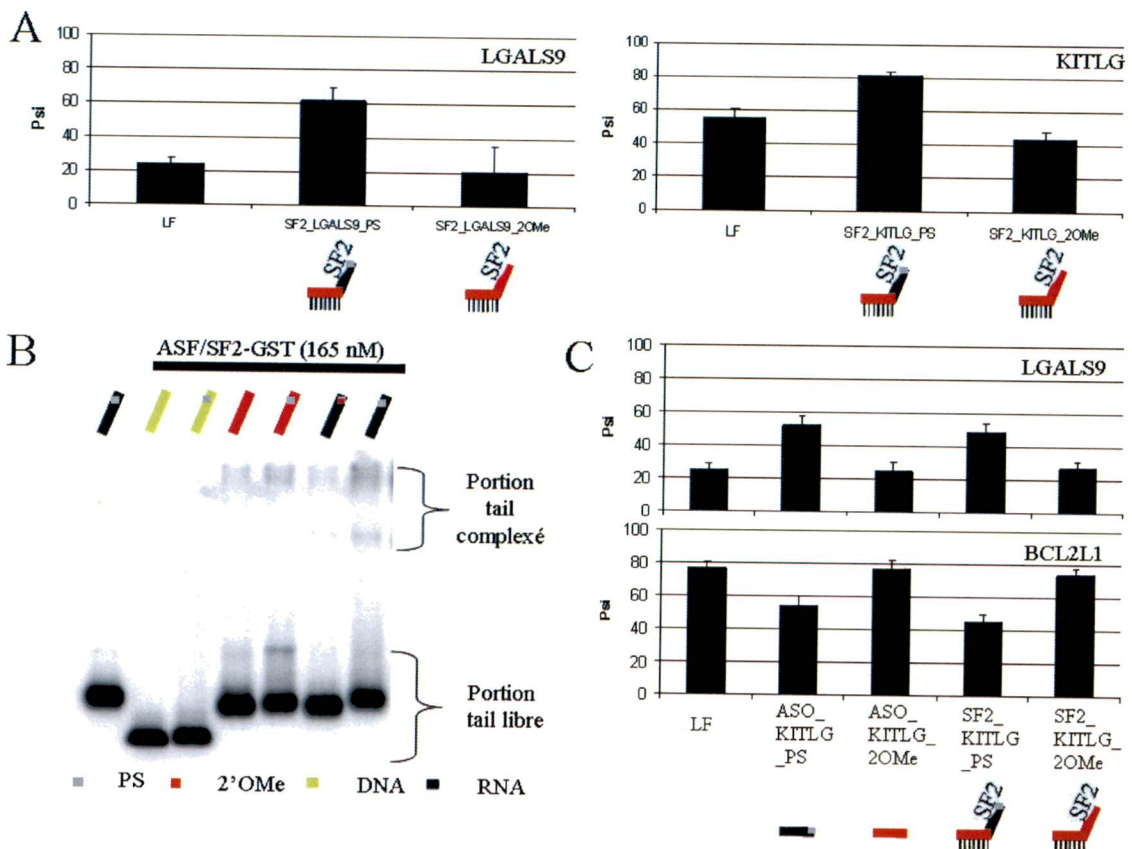


Figure 15 Impact de la chimie du TOES sur la modulation de l'AS.

A) Effet de la modification PS sur la modulation de l'AS du gène cible. À droite: Des cellules SKOV3ip1 (droite) et NIH-OVCAR-3 (gauche) ont été transfectées pendant 24 h par 480 nmol de SF2_LGALS9_PS, SF2_LGALS9_2OMe (droite) et SF2_KITLG_PS, SF2_KITLG_2OMe (gauche). L'ARN total a été extrait au Trizol®. Les histogrammes présentent les valeurs moyennes de ψ obtenues par endpoint PCR pour trois réplicats biologiques. L'augmentation de ψ du gène cible est dépendante de la modification PS. B) Effet de la chimie de la portion « tail » sur la liaison à SF2. Des oligonucléotides d'ADN (vert), d'ARN (noir) et 2'OMe (rouge) avec ou sans PS (carré gris) correspondant à la séquence de la portion « tail » ont été testés pour leur affinité pour SF2-GST. Piste 1: portion « tail » RNA sans SF2-GST. Piste 2 à 7: Différentes chimies de portion « tail » avec SF2-GST. SF2 se lie préférentiellement à une séquence de chimie ARN. C) Effet de PS sur l'AS de gène hors cible. Des cellules SKOV3ip1 ont été transfectées pendant 24 h par 480 nmol de ASO_KITLG_PS, ASO_KITLG_2OMe, SF2_KITLG_PS ou SF2_KITLG_2OMe. L'ARN total a été extrait au Trizol®. Les histogrammes présentent les valeurs moyennes de ψ obtenues par endpoint PCR et les barres d'erreur représentent la déviation standard obtenue pour trois réplicats biologiques.

Pour démontrer définitivement l'effet toxique des PS sur l'AS, des cellules SKOV3ip1 ont été transfectées par des oligonucléotides de différentes longueurs et séquences possédant ou non des modifications PS ciblant KITLG. Un essai endpoint PCR a été effectué pour vérifier la modulation de l'AS de gènes hors cibles (LGALS9, BCL2L1). Les résultats sont nets: la modulation de l'AS est dépendante de la modification PS indépendamment des séquences utilisées (Fig. 15C). De plus, un examen des cellules au microscope permet d'identifier à l'œil les puits traités aux PS des puits non traités, car une grande proportion des cellules se décolle et flottent dans le milieu sous l'action des PS. Il est donc fort probable que le dérèglement de l'AS reflète en fait l'état de crise des cellules et ne soit pas du tout une conséquence directe de l'action du modulateur. Ceci corrèle avec des travaux sur l'antisens Genasense qui ont montré que des oligonucléotides PS induisent l'apoptose de cellules cancéreuses avant même leur effet sur l'ARNm (Lai *et al.*, 2006). Ceci appuie notre argumentation à l'égard des modifications PS. Bien que celles-ci confèrent des propriétés de modulation de l'ASE cible, elles permettent la modulation de plusieurs autres ASEs, diminuant du même coup la spécificité d'action du TOES. À noter que le ctrl TOSS est entièrement composé de chimie 2'OMe et ne donne que rarement de tels effets hors cibles (comparer les Figs. 1,5,6,8,10,11,12,13 et 14). En résumé, la modulation de l'AS est

une conséquence indirecte due aux oligonucléotides PS et non une cause directe de l'action du modulateur.

Les effets de la portion « tail » des TOES testés semblent dépendants du contexte de l'exon pour lequel cela a été originalement dessiné. Pour réussir à moduler l'AS spécifiquement et indépendamment de PS, nous avons choisi d'exploiter une autre séquence de recrutement pour SF2 (5'-GAUGGUGAAGAAGACACUGC-3'), cette fois-ci préalablement validée dans un contexte hétérologue. Cette séquence initialement retrouvée dans l'exon EDA du gène FN1 favorise l'inclusion lorsqu'elle est insérée dans l'exon EDB (Muro *et al.*, 1999). Ainsi, nous postulons que le TOES élaboré à partir de cette séquence permettra de favoriser l'inclusion de l'exon EDB. Des cellules SKOV3ip1 ont été transfectées par deux TOES possédant cette nouvelle portion « tail », mais aucun n'a été en mesure de stimuler l'exon EDB (résultats non montrés). Ceci indique clairement que l'inclusion d'exon via le recrutement de SF2 est difficile à généraliser selon notre modèle de TOES.

2.3.2 TOES dépendant de RBFOX2

Comme le TOES dépendant de SF2 semble difficile à généraliser, nous nous sommes tournés vers une stratégie de TOES où la portion « tail » recrute un facteur d'épissage dont la séquence de liaison est un ISE. Ainsi, il est postulé que l'effet basal d'exclusion de la portion antisens sera moins grand si nous avons recours à des TOES dont l'antisens se lie à des séquences introniques. RBFOX2 est un facteur d'épissage qui manifeste ses effets favorisant l'inclusion d'exon via la liaison à l'hexamère UGCAUG souvent présent en aval d'exon cassette. Notre expérience avec RBFOX2 nous a permis d'identifier des ASEs qui lui sont sensibles et indifférents par des stratégies de répression par siRNA en cellules (Venables *et al.*, 2009). Dans un premier temps, nous avons ciblé quatre ASEs (ECT2, INSR, MPRIP et NDEL1) des plus sensibles à RBFOX2 avec des SSO 2'OMe. Ceux-ci sont destinés à bloquer l'action de RBFOX2 en ciblant une séquence UGCAUG intronique située dans les 100 nt suivants le 5'ss. Il est postulé que l'effet sera similaire à celui d'une répression de l'expression de RBFOX2. Ainsi, il serait ultimement possible de restituer la dépendance à RBFOX2 en ajoutant une portion « tail » munie d'une séquence UGCAUG (Fig. 16A). De façon surprenante, aucun des antisens ne parvient à moduler l'épissage des

cibles testées (résultats non montrés). Ces résultats nous laissent avec au moins trois hypothèses: 1) L'action de RBFOX2 est principalement médiée par sa liaison à UGCAUG, mais n'est pas gênée par la présence de l'antisens et se lie au double brin; 2) L'action de RBFOX2 est médiée de façon redondante via plusieurs séquences de liaisons UGCAUG; et 3) l'action de RBFOX2 est principalement indépendante de sa liaison à UGCAUG. Nous ne sommes pas en mesure de rejeter l'une ou l'autre de ces hypothèses présentement, bien que la deuxième soit préférée.

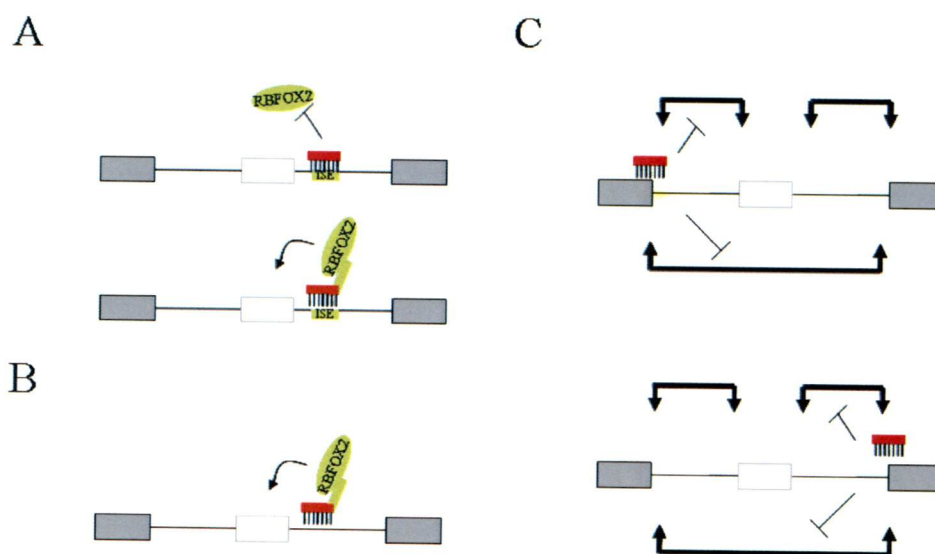


Figure 16 Autres stratégies favorisant l'inclusion de l'exon alternatif.

A) ASE dépendant de RBFOX2. Dans le haut: Masquage du site de liaison de RBFOX2 (ISE) par un SSO 2'OMe permettant de diminuer l'inclusion de l'exon alternatif. Dans le bas: Récupération de la dépendance à RBFOX2 en introduisant une séquence liant RBFOX2 via le même antisens 2'OMe. B) ASE indépendant de RBFOX2. Introduction d'une séquence liant RBFOX2 via une portion antisens 2'OMe positionné en aval du 5'ss de l'exon alternatif pour augmenter son inclusion. C) Cibler le 5'ss constitutif en amont (haut) ou le 3'ss constitutif en aval (bas) avec un SSO 2'OMe.

Dans un deuxième temps, nous avons ciblé quatre ASEs (AXIN1, BTC, FN1-EDA et STIM1) indifférents à RBFOX2 avec des portions antisens 2'OMe et une portion « tail » munie d'une (5'-CCCACCGCCUGCAUGCACUGCUCAC-3') ou deux (5'-UUGCAUGCUGUUGCAUGCAGGU-3') séquences UGCAUG afin de les rendre dépendants à RBFOX2 (Fig. 16B). Dans tous les cas, nous n'avons pas observé de stimulation de l'exon alternatif, pas plus que l'exclusion d'exon souvent retrouvée lorsque l'exon est directement ciblé (résultats non montrés). Nous en concluons que les stratégies

de type TOES favorisant l'inclusion d'exon basée sur SF2 ou RBFOX2 ne sont pas en mesure d'influencer suffisamment la décision d'AS pour s'avérer efficace dans la plupart des cas.

2.3.4 Antisens ciblant les jonctions constitutives

Vu les résultats infructueux de la stratégie TOES, nous avons pensé à une stratégie indépendante du recrutement de facteur d'épissage. En bloquant le 5'ss constitutif en amont d'un exon cassette, ceci interfèrera avec l'épissage du 3'ss de l'exon alternatif et le 3'ss de l'exon constitutif en aval, mais pas avec l'épissage du 5'ss de l'exon alternatif et le 3'ss de l'exon constitutif en aval (Fig. 16C). Suivant la même logique, il est théoriquement possible de favoriser l'inclusion de l'exon alternatif en ciblant le 3'ss constitutif en aval d'un exon cassette. Bien qu'il soit possible que le bloc d'une jonction constitutive se répercute sur l'épissage constitutif et donc, sur l'expression globale, cette stratégie pourrait s'avérer particulièrement avantageuse dans le cas d'ASE dont la séquence est courte et donc difficile à dessiner. Nous avons mis en pratique cette stratégie en ciblant trois ASEs : deux dont le 5'ss en amont a été ciblé (C11orf17 et SLIT2) et un dont le 3'ss en aval a été ciblé (SYNE2). Dans tous les cas, nous n'avons pas été en mesure d'observer une augmentation notable de l'inclusion de l'exon alternatif (résultats non montrés). En conclusion, les stratégies tentées pour généraliser la reprogrammation de l'inclusion de l'exon ne se sont pas avérées un succès.

2.3.5 siRNAs spécifiques à l'isoforme court

Dans le but de permettre une modulation de l'AS en faveur de l'isoforme long, nous nous sommes tournés vers une technologie beaucoup plus mature : les siRNAs. En ciblant spécifiquement l'isoforme court, la valeur de ψ ou $Q\psi$ sera à la hausse. Par contre, dans le cas particulier des siRNAs ciblant spécifiquement l'isoforme court, aucune étude systématique permettant de révéler des règles de dessin générales n'a été réalisée. Comme les siRNAs sont des duplex d'une longueur de 19 bp, nous avons délimité la fenêtre de dessin des siRNAs jonctions à 30 nt, soit 15 nt provenant de chacun des exons de la jonction à cibler. Chacune des jonctions est flanquée d'un G constitutif, ce qui ne laisse qu'un minimum de 3 mésappariements théoriques vis-à-vis l'isoforme long. La séquence

de 30 nt est soumise aux mêmes critères de dessin qu'un siRNA ciblant un exon, soit le respect de l'asymétrie et l'absence de séquences hautement homologues (décrits dans le chapitre 3). Règle générale, nous avons choisi deux siRNAs jonctions par ASEs. Pour ceux-ci, les séquences guides potentielles ont été alignées sur les jonctions exon-exon propres à l'isoforme long et les mésappariements ont été soigneusement annotés (voir annexe 2). Dans tous les cas, les siRNAs jonctions possèdent un minimum de 3 mésappariements vis-à-vis chacune des jonctions exon-exon propres à l'isoforme long.

Afin de tester leur spécificité, des cellules SKOV3ip1 ont été transfectées par des siRNAs jonctions (10 nM final) ciblant TACC2, KIAA1217, SYNE2, APP, CD47 et C11orf17. Nous avons exploité la détection isoforme-spécifique par PCR en temps réel pour quantifier chacun des isoformes séparément, seul moyen pour véritablement valider la spécificité du siRNA jonction. En effet, la mesure du ratio d'isoforme (valeur de ψ) donnée par endpoint PCR ne donne aucune information sur la quantité de l'un ou l'autre des isoformes, pas plus que la valeur de la quantité globale. Tous les siRNAs dessinés s'avèrent spécifiques à l'isoforme court sauf pour SYNE2 (annexe 2). Ce résultat incohérent nous a poussés à analyser la similarité de la jonction exon 277-exon 140 de SYNE2 par rapport à la jonction exon 277- intron 1777. Comme le pré-ARNm contient cette jonction exon-intron, l'effet d'un siRNA ciblant l'intron serait similaire à l'effet d'un siRNA global, ce qui expliquerait les effets de SYNE2_hit1_1_s sur l'isoforme long. En effet, une analyse de la séquence intronique révèle une différence d'un seul nucléotide en position 1 (par rapport au siRNA). Ceci suggère que la séquence intronique, autant que la séquence exonique doit être tenue en compte lors du dessin de siRNAs jonctions.

2.4 Conclusions

Dans un premier temps, nous avons observé un artéfact PCR dû à la présence de TOSS dans l'extrait d'ARN. En effet, une faible quantité de TOSS ajouté sur un extrait d'ARN total suffit pour induire une modulation artéfactuelle du ratio d'isoforme *in vitro*. De façon intéressante, nous avons remarqué qu'une extraction au Trizol® permet d'enlever le TOSS d'un extrait d'ARN, et ainsi sa composante PCR artéfactuelle associée; alors qu'une extraction en colonne ne le permet pas. La capacité à retenir ou non le TOSS dans l'extrait d'ARN a été exploitée afin d'évaluer le véritable impact du TOSS résiduel 24 h post-transfection. L'ensemble des résultats indique que la composante artéfactuelle joue un rôle mineur dans le cas de fort modulateur et possède un impact majeur dans le cas de faible modulateur. De plus, l'augmentation de l'isoforme court et l'absence d'effet au niveau de l'expression globale par PCR en temps réel démontre clairement les effets réels de modulation des TOSS dans la plupart des cas. Nous avons donc utilisé la stratégie en colonne pour le haut débit et les échantillons ont été purifiés au Trizol® en cas de doute.

Dans un deuxième temps, nous avons réalisé une série d'expériences afin de mettre au point des règles de dessin de TOSS. Au total, 94 TOSS, TOSS contrôles et antisens ont été testés. Les meilleures amplitudes de modulation sont obtenues avec un TOSS composé entièrement de nucléotides modifiés en 2'OMe, une portion « tail » de séquence 5'-UAU GAU AGG GAC UUA GGG UG-3' et une portion antisens de 20 nucléotides complémentaires à l'exon alternatif. Les effets sont indépendants de la position de l'antisens sur l'exon alternatif. Bien que la plupart des TOSS qui fonctionnent produisent effectivement une réelle modulation que nous avons nommée type A, un certain nombre de TOSS procurent un changement du ratio d'isoforme, mais également un changement d'expression globale. Un TOSS doit donc être testé au moins par PCR en temps réel pour son effet sur l'expression globale et par endpoint PCR ou PCR en temps réel pour son effet sur le ratio d'isoforme. Le TOSS n'est pas limité aux sites alternatifs 5' et permet de moduler les exons cassettes simples et les sites alternatifs 3' (type B seulement). Il permet également de cibler spécifiquement un exon ou une région alternative d'un ASE complexe ou de cassette multiple.

Dans un troisième temps, les résultats ont été compilés selon le type de modulation (type A, type B) et le type d'ASE pour le premier TOSS dessiné par ASE afin d'évaluer le taux de succès de dessin (Tableau 3). En faisant abstraction du faible nombre d'entrées pour les cas de site alt3' et 5', on remarque que le taux de succès global est similaire peu importe le type d'évènement. Seuls les cas de rétention d'intron sont infructueux. Sauf pour les cas de site alt3', il ne semble pas y avoir de corrélation entre le type d'évènement et le type de modulation. En effet, outre le cas de site alt3' du gène DRF1, deux autres cas d'évènements complexes mettant en cause des sites alternatifs 3' (TOPBP1 et LIG3) procurent des modulations de type B, haussant le nombre à 3 en 3. Cela permet de conclure que les TOSS ont une moyenne de 84% de succès lorsqu'ils ciblent des sites alternatifs 3', 5' ou des exons cassettes, et ce, même s'il s'agit d'un évènement complexe ou d'un exon cassette multiple. De plus, la plupart des modulations sont de type A, faisant des TOSS des outils moléculaires particuliers et distincts des siRNAs.

Finalement, plusieurs stratégies ont été élaborées pour tenter de favoriser l'inclusion d'exon cassette. Par analogie à la stratégie TOSS qui recrute un facteur d'épissage se liant à des ESS, nous avons testé la stratégie TOES qui consiste à recruter un facteur d'épissage se liant à des ESE ou ISE. Ces approches n'ont pas été en mesure de stimuler l'AS de façon notable, pas plus que le ciblage des jonctions constitutives adjacentes. Le fait le plus marquant est la modulation non spécifique de l'AS de façon PS dépendante. À l'inverse, nous avons démontré la spécificité de plusieurs siRNAs ciblant la jonction exon-exon de l'isoforme court et nous permet d'entrevoir la possibilité d'un criblage avec des siRNAs longs (ciblant l'isoforme long) et siRNAs courts (ciblant l'isoforme court) afin de révéler la fonction propre à chaque isoforme.

Tableau 3 Taux de succès des TOSS.

Les résultats ont été compilés selon le type de modulation (type A, type B) et le type d'ASE pour le premier TOSS dessiné pour chacun des ASEs étudiés. Dans certain cas, les données d'expression globale sont incomplètes et empêchent de définitivement conclure entre type A et type B (A ou B). Les TOSS ont un taux de succès moyen de 84% (A ou B) et de 43 à 67% si l'on considère seulement la modulation de type A.

	Modulation					Taux de succès (%)	
	A	B	A ou B	NON	Total	A	B
Alt3'	0	1	0	0	1	0	100
Alt5'	1	0	1	0	2	50-100	0-50
Exon cassette	17	3	6	4	30	57-77	10-30
Complexe	2	2	2	1	7	29-50	25-50
Multiple	3	2	2	1	8	38-63	25-50
Rétention d'intron	0	0	0	2	2	0	0
Grand total	23	8	11	8	50	43-67	14-39

Chapitre 3

Alternative splicing of SYK regulates mitosis and cell survival

Auteurs de l'article: Prinos P, Garneau D, Lucier JF, Gendron D, Couture S, Boivin M, **Brosseau JP**, Lapointe E, Thibault P, Durand M, Tremblay K, Gervais-Bird J, Nwilati H, Klinck R, Chabot B, Perreault JP, Wellinger RJ, Elela SA.

Statut de l'article : Publié, *Nature Structural & Molecular Biology* 2011;18(6):673-9

Avant-propos

J'ai développé une méthodologie de validation des siRNAs et TOSS à haut débit qui a été utilisée pour produire la figure 2. De plus, je suis intellectuellement impliqué dans la conception de la figure 1. J'ai participé activement à la rédaction du texte de la section « Matériels et méthodes supplémentaires ». De plus, le développement des outils moléculaires présenté au chapitre 2 y est explicitement référé. Pour plus de détails, consulter le « Author contribution » à la fin de l'article 2.

Mise en contexte

La cancérogenèse modifie l'expression des ARNm. Nous émettons l'hypothèse selon laquelle certaines isoformes de l'épissage alternatif ont des fonctions clés dans le processus de cancérogenèse. Même si des analyses bioinformatiques prédisent que la plupart des évènements d'épissage affectant la portion codante des gènes produisent des isoformes ARNm susceptible d'être traduites en protéine, il n'est pas clair quel en est l'impact sur leur fonction.

En théorie, la relation fonctionnelle entre deux isoformes peut rejoindre l'un des quatre modèles du tableau 1. D'abord, si un des deux isoformes produit une protéine excluant un domaine fonctionnel donné ou est ciblé vers la voie de dégradation NMD, alors seul le premier isoforme est doté d'une fonction cellulaire donnée. La quantité de transcrit de celui-ci dictera donc l'effet phénotypique et l'action d'un TOSS ou d'un siRNA long devrait être effectif. Par contre, l'effet d'un TOSS ou d'un siRNA long se manifestera également lorsque les deux isoformes proposent des fonctions opposées ou ayant une relation de type dominant-négatif. Même si c'est le rapport entre les deux isoformes (ratio) qui prime, ces deux relations fonctionnelles distinctes ne pourront être mises en évidence que par un siRNA global. En revanche, l'action d'un TOSS sera sans effet dans une situation où les deux isoformes ont des fonctions identiques dans un contexte donné (compensatoire).

Tableau 1 Relation fonctionnelle entre les isoformes de l'AS et influence théorique des outils moléculaires sur le phénotype.

Quatre modèles exprimant la relation fonctionnelle entre un gène théorique et ses deux isoformes (A et B) sont présentés. L'analyse globale de l'impact d'un TOSS, siRNA long (A), siRNA court (B), et siRNA global (A et B) sur un phénotype choisi permet de déduire la relation fonctionnelle des isoformes pour un contexte donné. L'isoforme long et l'isoforme court ont été attribués arbitrairement aux isoformes A et B, respectivement. \checkmark = effet phénotypique. \checkmark' = effet phénotypique inverse. --- = sans effet phénotypique.

	Monofonctionnel : Seul l'isoforme possède une fonction	Compensatoire : Les deux isoformes ont des fonctions similaires	Antagoniste type I : L'isoforme B agit en dominant-négatif	Antagoniste type II : Les deux isoformes ont des fonctions opposées
Isoforme A	Fonction A	Fonction A	Fonction A	Fonction A
Isoforme B		Fonction B	Fonction anti-A	Fonction Z
Dépendance du phénotype	[A]	[A] + [B]	[A] / [B]	[A] / [B]
TOSS	\checkmark	---	\checkmark	\checkmark
siRNA A	\checkmark	\checkmark	\checkmark	\checkmark
siRNA B	---	\checkmark	\checkmark	--- (ou \checkmark')
siRNA global	\checkmark	\checkmark	---	---

Le gène BCL2L1 est un modèle d'étude attrayant puisque l'isoforme court (Bcl-x_S) code pour une protéine pro-apoptotique alors que l'isoforme long (Bcl-x_L) a des propriétés anti-apoptotiques. Puisque leurs fonctions sont opposées, le ratio des deux isoformes devrait théoriquement dicter le phénotype. Or en pratique, il n'est pas clair si une simple altération du ratio des isoformes sans changement d'expression globale peut induire un phénotype en cellules. En effet, peu d'études ont directement comparé les effets de répression (et conséquemment, d'un changement de ratio d'isoforme) et d'une simple modulation d'isoforme endogène sur le phénotype de cellules humaines. En fait, les résultats les plus complets proviennent d'études sur le gène modèle BCL2L1 (Mercatante *et al.*, 2001, Mercatante *et al.*, 2002). Même dans ce cas bien étudié, il n'est pas clair si un certain ratio de Bcl-x_L/Bcl-x_S dicte l'induction d'apoptose ou si un niveau minimal de Bcl-x_S doit être respecté.

Pour répondre à cette question, un modulateur d'AS (splice-switching oligonucleotide, SSO) de 18 nt de longueur en 2'OMe-PS (5' Bcl) a été utilisé pour moduler le 5'ss de Bcl-

x_L vers Bcl- x_S . Pour réprimer globalement l'expression de BCL2L1, un antisens ciblant la jonction constitutive intron2-exon3 (3' Bcl) fut utilisé (Mercatante *et al.*, 2001). En temps normal, un tel antisens favoriserait Bcl- x_L tout en diminuant globalement l'expression de BCL2L1, mais comme Bcl- x_S est indétectable par immuno-buvardage dans les PC-3, 3' Bcl agit à toute fin pratique comme répresseur spécifique de Bcl- x_L . Dans un essai quantifiant l'ADN sous-diploïde par FACS, 5' Bcl est environ deux fois plus puissant que 3' Bcl à induire l'apoptose. Les auteurs en concluent que l'augmentation de Bcl- x_S est nécessaire pour générer de l'apoptose et que la diminution de Bcl- x_L n'est pas suffisante. En appui à cette conclusion, le niveau de cellules apoptotiques semble corrélér positivement avec la force promotrice des vecteurs de surexpression de Bcl- x_S dans les cellules MCF-7 (Sumantran *et al.*, 1995, Clarke *et al.*, 1995). Dans une deuxième publication, l'équipe de Kole corrèle l'efficacité à induire l'apoptose et les niveaux protéiques de Bcl- x_L et Bcl- x_S (Mercatante *et al.*, 2002). Plus une lignée cellulaire exprime fortement le gène Bcl-x, plus le traitement de 5' Bcl sera efficace, ce qui est consistant avec une augmentation des niveaux absolus de Bcl- x_S .

Ceci fait contraste avec les résultats obtenus précédemment par les chercheurs d'Isis Pharmaceuticals et le groupe de Zangemeister-Wittke (voir tableau 3). Alors que les premiers n'observent pas d'induction d'apoptose dans les cellules A549 avec un SSO (Taylor *et al.*, 1999), les derniers sont en mesure d'augmenter le nombre de cellules apoptotiques en ciblant spécifiquement Bcl- x_L dans la même lignée cellulaire à l'aide d'un antisens dépendant de la RNaseH, le gapmer 4259 (Leech *et al.*, 2000). Plus encore, le traitement au gapmer 4259 décroît considérablement la viabilité cellulaire des cellules MCF-7 traitées, tel que jugé par un essai d'exclusion au bleu de trypan. À noter toutefois que les effets semblent être également lignée-dépendants puisqu'ils ne sont pas observés dans les cellules SW2 malgré une répression similaire de l'ARNm. À la lumière de ces résultats, c'est le ratio Bcl- x_L / Bcl- x_S qui dicte le phénotype et non le niveau absolu de Bcl- x_S . Un peu plus tard, l'équipe de Massimo Caputi s'est lancée dans la course en employant une technologie d'oligonucléotide bifonctionnel conceptuellement apparenté au TOES nommé ESSENCE (exon-specific splicing enhancement by small chimeric effectors) (Wilusz *et al.*, 2005). Il s'agit d'un antisens en PNA (peptide nucleic acid) couplé à un peptide mimant un domaine RS de facteur d'épissage SR. Ce chimère est positionné en

amont du 5'ss de Bcl-x_S et favorise la génération de Bcl-x_S au détriment de Bcl-x_L. Contrairement à l'équipe de Kole, ceux-ci sont en mesure d'observer une induction d'apoptose dans les cellules HeLa. Finalement, Dodier et Piché ont ciblé Bcl-x_L à l'aide de siRNA isoforme-spécifique. Bien qu'ils observent une augmentation d'apoptose des cellules CaOV3, les essais ont été infructueux dans le cas de deux autres lignées ovariennes (SKOV3ip1 et NIH-OVCAR3) soulignant les effets type cellulaires dépendants (Dodier and Piche, 2006).

Tableau 2 Comparaison des effets de modulateurs d'AS et de répression isoforme-spécifique du gène BCL2L1 sur l'induction d'apoptose dans des cellules cancéreuses humaines.

Lignées cellulaires	Induction d'apoptose suivant une modulation de Bcl-x _L vers Bcl-x _S	Induction d'apoptose suivant une répression spécifique de Bcl-x _L
A549	Non (Taylor <i>et al.</i> , 1999)	Oui (Leech <i>et al.</i> , 2000)
MCF-7	Non (Mercatante <i>et al.</i> , 2001)	Oui (Simoes-Wust <i>et al.</i> , 2000)
MDA-MB-231	Oui (Mercatante <i>et al.</i> , 2002)	Oui (Simoes-Wust <i>et al.</i> , 2000)
HeLa	Non (Mercatante <i>et al.</i> , 2002) Oui (Wilusz <i>et al.</i> , 2005)	N.D.
CaOV3	N.D.	Oui (Dodier and Piché, 2006)
PC-3	Oui (Mercatante <i>et al.</i> , 2001)	N.D.

N.D = Non-disponible

Pour éclaircir ce mystère, nous avons testé le SSO 5' Bcl en parallèle avec le TOSS A1_Bclx4 dans le meilleur modèle cellulaire décrit par Kole R : les cellules PC-3 (Mercatante *et al.*, 2001). Bien que la modulation vers Bcl-x_S soit évidente pour tous par RT-PCR, la viabilité cellulaire décroît seulement lorsque les cellules sont traitées avec 5' Bcl. Comme nous avons déjà expérimenté des effets non-spécifiques dûs aux modifications PS, nous avons contrôlé cette expérience en comparant les effets de 5' Bcl à une version de 5' Bcl sans PS et à un autre oligonucléotide PS ciblant un autre gène (β -globin). Résultat: la viabilité cellulaire est dépendante de la présence de PS et indépendante de la modulation des isoformes. Des essais par immuno-buvardage afin d'évaluer la présence de PARP clivé (un indicateur d'apoptose) se sont avérés négatifs, et ce même si les cellules PC-3 sont sensibilisées avec une dose de cisplatine correspondant à sa valeur de IC₁₀ (concentration requise pour obtenir 10% de l'effet maximal). Comme l'induction d'apoptose est infructueuse avec un modulateur de Bcl-x_L/Bcl-x_S efficace, nous avons abandonné le gène

modèle BCL2L1 pour nous concentrer sur l'étude d'un ASE qui donne un fort phénotype d'apoptose indépendamment de la présence de PS.

Les travaux présentés dans ce chapitre ont pour but de démontrer que 1) pour un gène donné, un phénotype pro-cancer peut être obtenu par seule altération du ratio d'isoforme de gène clé (sans altération de l'expression globale); 2) pour un gène donné, la dépendance au ratio d'isoforme varie en fonction du contexte (essai phénotypique).

Résumé

Chez l'humain, la plupart des gènes produisent plusieurs isoformes ARNm dues à l'épissage alternatif. Par contre, leur pertinence biologique demeure incertaine. Dans cette étude, nous avons évalué l'impact de l'épissage alternatif sur des cellules cancéreuses. Le patron d'épissage de 41 évènements d'épissage alternatif reliés au cancer a été systématiquement modulé et leurs effets sur la croissance cellulaire, la viabilité et l'apoptose ont été notés. Nous avons identifié 3 évènements d'épissage associés à la survie des cellules cancéreuses. De façon plus détaillée, la reprogrammation de l'épissage du gène encodant la kinase SYK promeut la progression du cycle cellulaire et stimule la croissance des cellules en suspension. De plus, un traitement exogène des cellules cancéreuses au facteur de croissance épidermique (EGF) module le ratio d'épissage de SYK en faveur de l'isoforme pro-survie tel qu'observé *in vivo*. Ces données suggèrent que l'épissage alternatif de certains gènes est spécifiquement modulé durant la cancérogenèse pour permettre l'expression d'isoformes favorisant la survie des cellules cancéreuses.

Article 2: Alternative splicing of SYK regulates mitosis and cell survival

Abstract

The majority of human genes produce multiple mRNA isoforms through alternative splicing. However, the biological relevance of most splice variants remains unclear. In this study, we evaluated the functional impact of alternative splicing in cancer cells. The splicing pattern of 41 cancer-associated splicing events was modulated and its impact on cell growth, viability and apoptosis scored. We identified 3 splicing-dependent defects in genes associated with cell survival. Specifically, changing the splicing pattern of the spleen tyrosine kinase gene (SYK) promoted cell cycle progression and stimulated anchorage-independent growth. Strikingly, exposure of cancer cells to epithelial growth factor modulated the SYK splicing pattern to promote the pro-survival isoform associated with cancer tissues in vivo. The data suggest that splicing of selected genes is specifically modified during tumor development to allow the expression of isoforms that promote cancer cell survival.

Introduction

Currently, the majority of therapeutic approaches define a target as one protein, one gene or one function. Yet, whole transcriptome sequencing has confirmed that only a minority of genes produce a single protein^{1,2}. Indeed, the vast majority of genes use alternative splicing (AS) to produce several protein isoforms. In eukaryotes, AS plays a central role both in protein diversity and post-transcriptional gene regulation. While AS can introduce or remove regulatory elements to affect translation, localization or degradation of an mRNA³, most often it will produce multiple and functionally diverse protein isoforms⁴. Perturbations in alternative splicing are common and extensive in cancer⁵ and can affect cancer development and maintenance⁶. For example, a switch in the alternative splicing of CD44 has been associated with the acquisition of metastatic potential⁷. Furthermore, a splicing switch in Pyruvate Kinase, a glycolytic enzyme, is essential for cancer metabolism and tumor growth⁸. However, the general impact of alternative splicing in cancer biology and its impact on drug targeting remain largely unexplored⁹. Indeed, it is not clear in many cases whether changing the proportion of a given set of splice variants contributes to cancer biology independently of the level of gene expression.

RNA interference (RNAi) has become a preferred tool for discovering and evaluating the impact of gene expression on the viability of cancer cells¹⁰. However, RNAi-based screens commonly target genes without considering the variety of alternative mRNA isoforms generated by AS. This limits our understanding of individual isoform functions and strongly reduces the numbers of potential drug targets. Our recent studies in ovarian¹¹ and breast cancer¹² identified 74 new cancer-associated splicing events. Here we describe a systematic isoform-specific functional screen of 41 breast and ovarian cancer-associated alternatively spliced variants with the goal of evaluating the impact of alternative splicing on cancer biology. Our targeting approach utilizes a dual system for isoform-specific silencing and modulation. This comprehensive approach harnesses the power of siRNA for isoform-specific silencing¹³ and complements it with alternative splicing modulation using bifunctional antisense oligonucleotides¹⁴. This approach uncouples the impact of splicing modulation on cell function from phenotypic defects caused by changes in gene expression levels. The results identify alternative splicing in genes associated with cell survival as

major targets for breast and ovarian cancer. Changing the splicing pattern of the tyrosine kinase (SYK) altered cell survival and mitotic progression, while global knockdown of the same gene had no effect, suggesting that at least for some genes, global knockdown of gene expression cannot mimic the effects of alternative splicing on cell survival and apoptosis.

Results

Functional analysis of cancer-associated splice variants

A large number of alternative splicing events (ASEs) are modified in tumor cells⁵. However, little is known about the contribution of alternative splicing to cancer biology. To gain insights into this aspect, we have developed a platform for high-content functional analysis of alternative splicing events (FASE) (Supplemental Figure 1) and have used it to examine the function of 29 ovarian¹¹ and 21 breast cancer variants¹², including 9 common to both cancers (supplemental Table 1). For a graphical representation of the protein domains affected in these variants see http://palace.lgfus.ca/link/40_ases. Cancer-associated ASEs were classified based on their splicing pattern, and isoform-specific inhibitors were designed for all events with simple alternative 5' events or cassette exons (Figure 1 and Supplemental Figure 2). Two exon-specific short interfering RNAs (siRNAs) and 1 targeted oligonucleotide silencer of splicing (TOSS)¹⁴ were designed for each splicing event and their impact on the phenotype of the SKOV3ip1 ovarian cancer cell line was scored using multiplexed fluorescent dyes for cell count, viability and apoptosis assays (Supplemental Figures 1 and 3). Assays were performed in biological and technical triplicates and quality metrics (Z-factors) for each assay plate were calculated (Supplemental Figures 1 and 3). Phenotypic defects reproduced by at least two independent isoform specific inhibitors (ISIs) against the same splicing event with a robust Z-score of greater than 3 were considered functionally relevant and retained for further analysis. As expected, mock transfection or transfection of unrelated siRNA or TOSS did not affect cellular phenotypes, while transfection of siRNA against the positive control, RBM8A, consistently induced apoptosis and reduced cell counts and viability (Supplemental Figure 3). While the majority of the targeted splicing events did not exhibit any phenotypic effect (Z-score <3), four splicing events in the AGR3, FN1, MCL-1 and SYK genes produced robust apoptosis induction (Suppl. Table 2). ISI dependent changes in the expression level of AGR3 could not be validated by PCR due to its low expression level and thus it was not pursued further. The remaining 3 genes affecting apoptosis have previously been implicated in cell survival pathways^{15,16}. Inhibiting the expression of the long isoforms of these 3 genes by siRNAs resulted in a 3 to 8 fold increase in the ratio of apoptotic marker

positive cells (Supplemental Table 2 and Figure 2, left panels). Similar, 4 to 12 fold, increases in the ratio of PI-positive, non-viable cells were observed by siRNA-mediated silencing of the long isoforms of these 3 genes (Supplemental table 2 and Figure 2, left panels). Consistently, reprogramming FN1-EDB and SYK splicing using TOSS resulted in increased apoptosis and decreased viability (Figure 2a and supplemental Table 2). In contrast, modulation of MCL-1 alternative splicing with TOSS had modest effect on apoptosis (Figure 2 b). Silencing of CCNE1 and LIG4 by splice isoform specific siRNA produced marked decreases in cell viability (Z-scores>3, suppl. Table 2) but did not result in substantial apoptosis (suppl. Table 2 and data not shown). Overall, the screen results indicate that 10 % of the ASEs tested contribute to the viability of cancer cells in vitro.

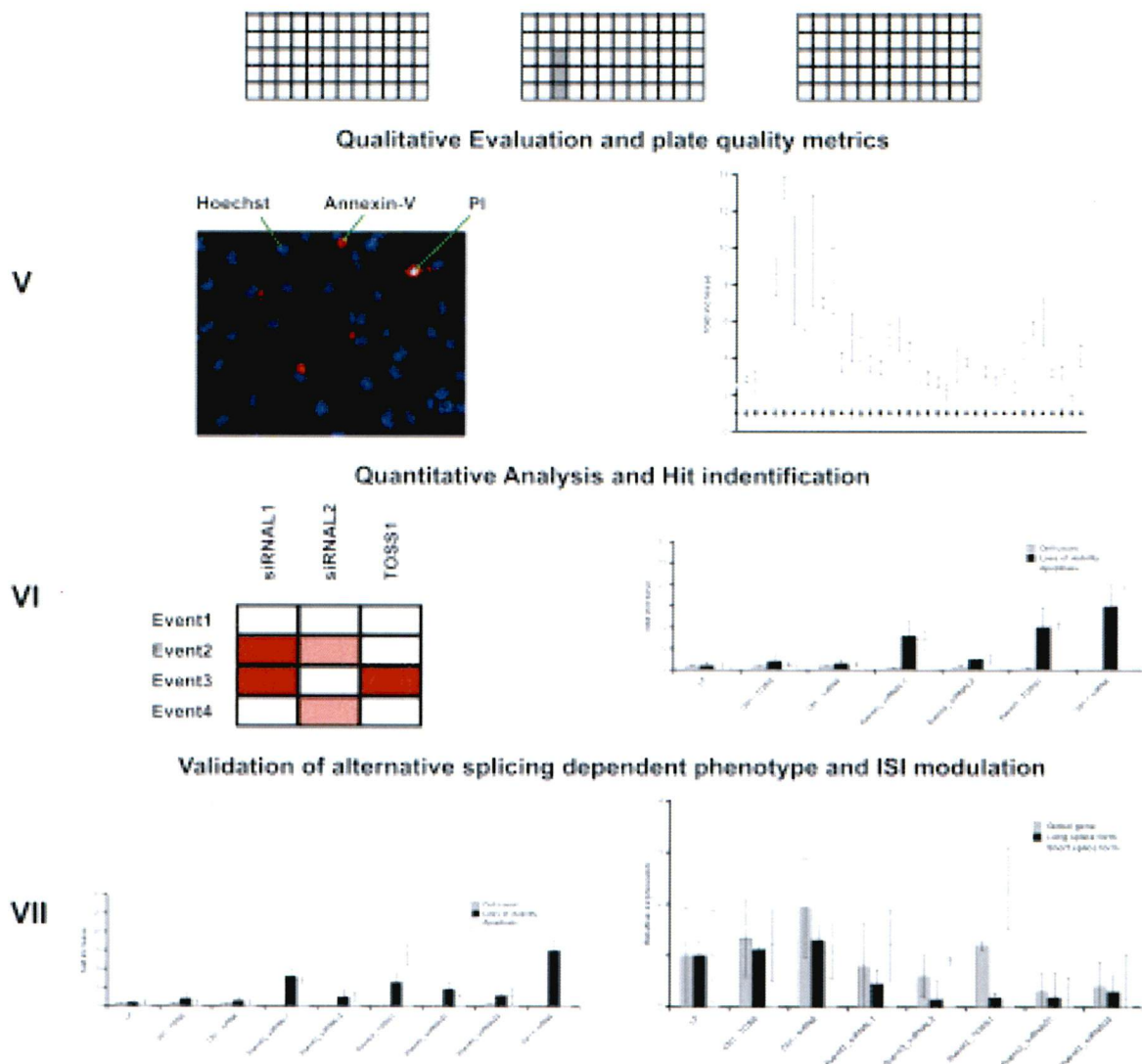


Figure S1 FASE (Platform for High-Content Functional Analysis of Alternative Splicing Events) screen pipeline.

One of the unique aspects of our platform is that it uses as its primary input a list of experimentally annotated, cancer-associated alternative splicing events. The first step of our workflow is identification of the alternative region that will be used for our bioinformatics design of isoform-specific inhibitors (ISIs). An automated design algorithm for efficient siRNA and TOSS design was developed in-house based on published and empirical criteria. Our experimental design pipeline consists of randomized transfection plate set-up to avoid systematic errors (row, column effects). ISIs are then transfected in triplicate per plate (technical replicates) three independent times. Each plate also contains several positive and negative controls for assay quality evaluation. Plate scanning and image acquisition are performed following addition of fluorescent dyes 96 hours post-transfection. 9 images are scanned per well and a montage is generated representing each well. The data file generated for each well is loaded into a database and analyzed bioinformatically with various modules. The analysis output is graphically displayed in bar graphs per assay. Quality metrics allow plate and assay validation by means of Z-factors. Data triage and normalization by subtracting the negative control are then performed. This allows for hit identification by Z-score calculation. Successful modulation of alternative splicing by Q-PCR and western blots allow for hit validation.

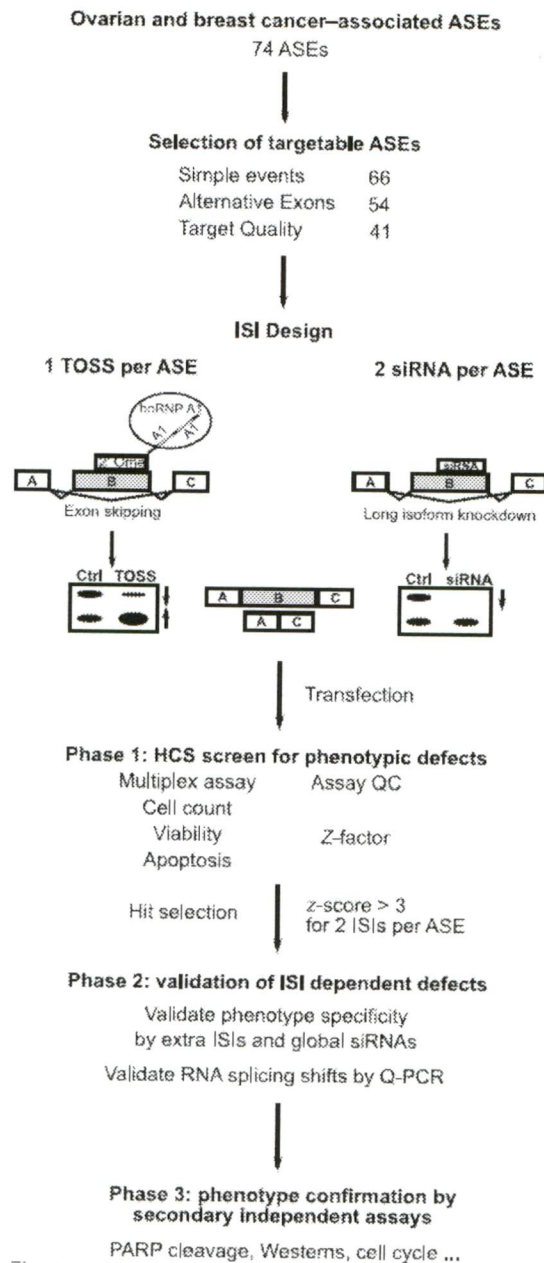


Figure 1 Strategy for functional annotation of cancer-associated alternative splicing events (ASEs).

Ovarian and breast cancer-associated alternative cassette exons were identified and their expression was modulated using isoform specific inhibitors (ISIs). Each simple alternative cassette exon or alternative 5' splice site was inhibited using 2 independent exon-specific siRNAs and its splicing was modulated using a Targeted Oligonucleotide Silencer of Splicing (TOSS). The impact of modulating the expression of splice isoforms on cell proliferation, viability and induction of apoptosis was monitored using the high content FASE platform (**F**unctional **A**nalysis of **A**lternative **S**plicing **E**vents). Phenotypes caused by selective modulation of splice isoform ratios and not the overall inhibition of gene expression were considered as indicators of alternative splicing isoform-specific functions.

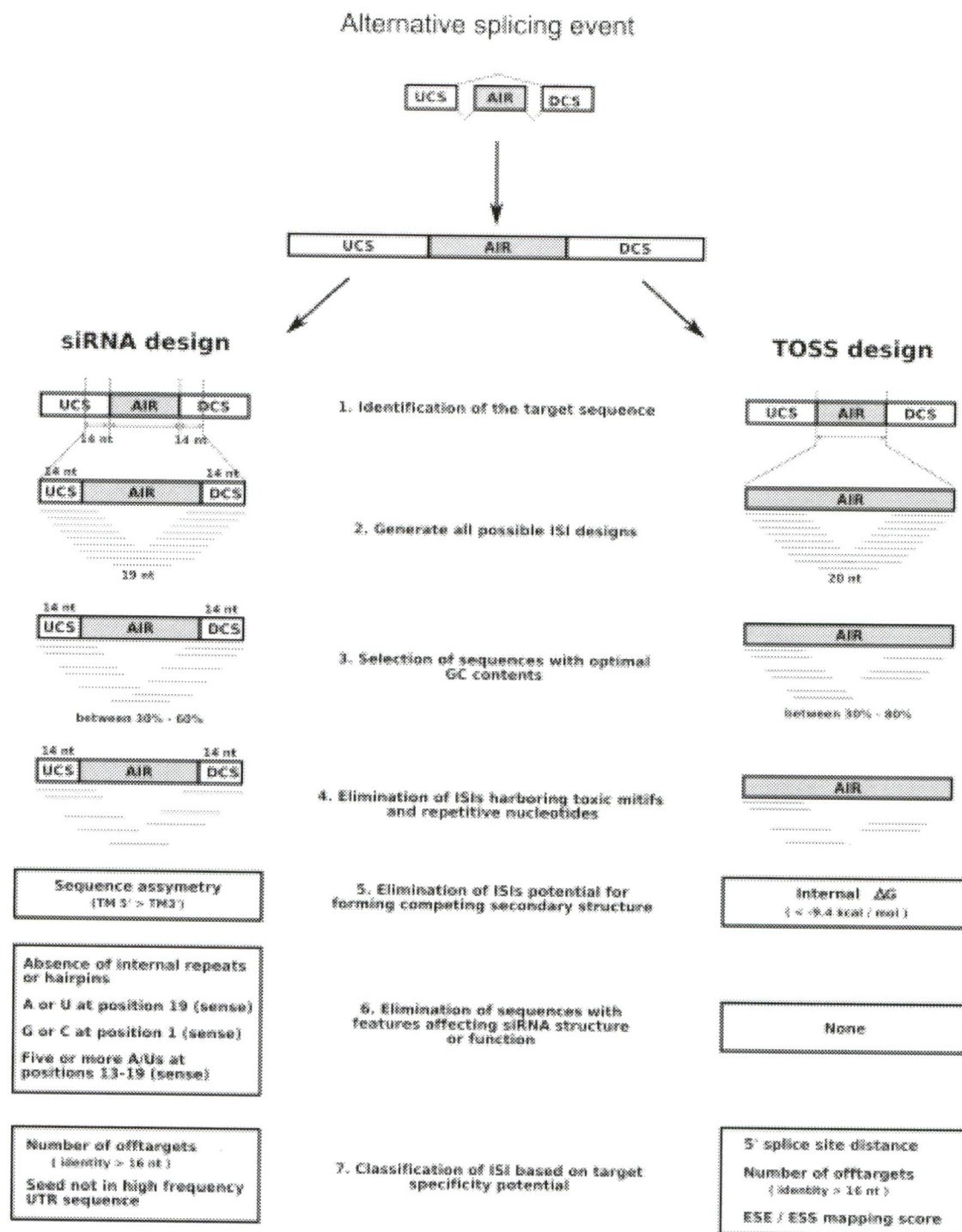


Figure S2 ISI design rules.

Long-isoform specific siRNA (left panel) and TOSS (right panel) selection criteria and design rules. An algorithm for efficient siRNA and TOSS design was developed in-house based on published and empirical criteria. The design algorithm is automated in a user-friendly web-based interface. The minimum exon length for ISI design is 20 nucleotides. AIR: Alternatively included region. UCS: upstream constitutive sequences; DCS: downstream constitutive sequences.

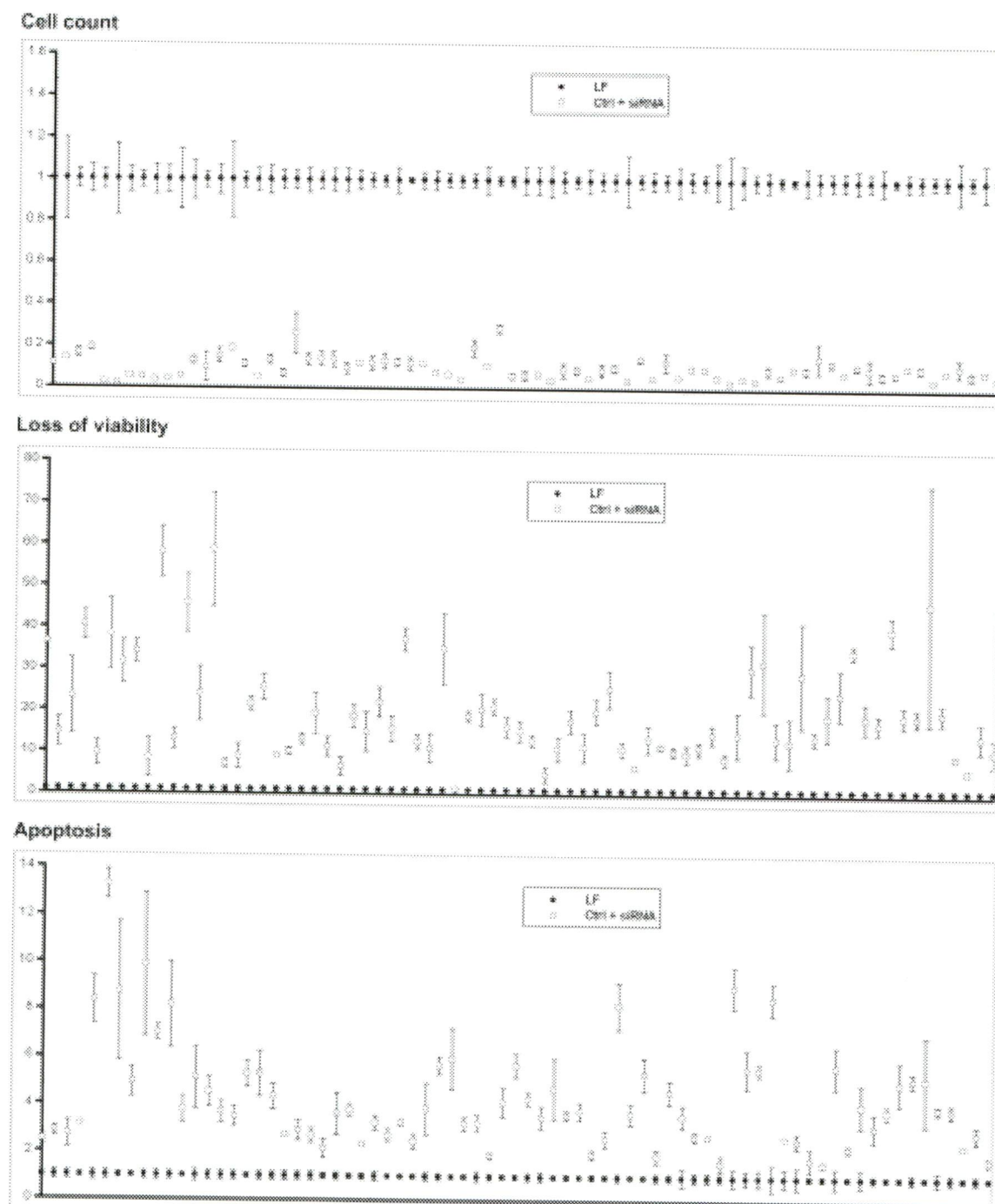


Figure S3 FASE screen quality control.

Graphical representations of positive (RBM8 siRNA) and negative (lipofectamine) control segregation for cell count, viability and apoptosis assays. Z-factors were derived from these three different assay outputs of the multiplex screen.

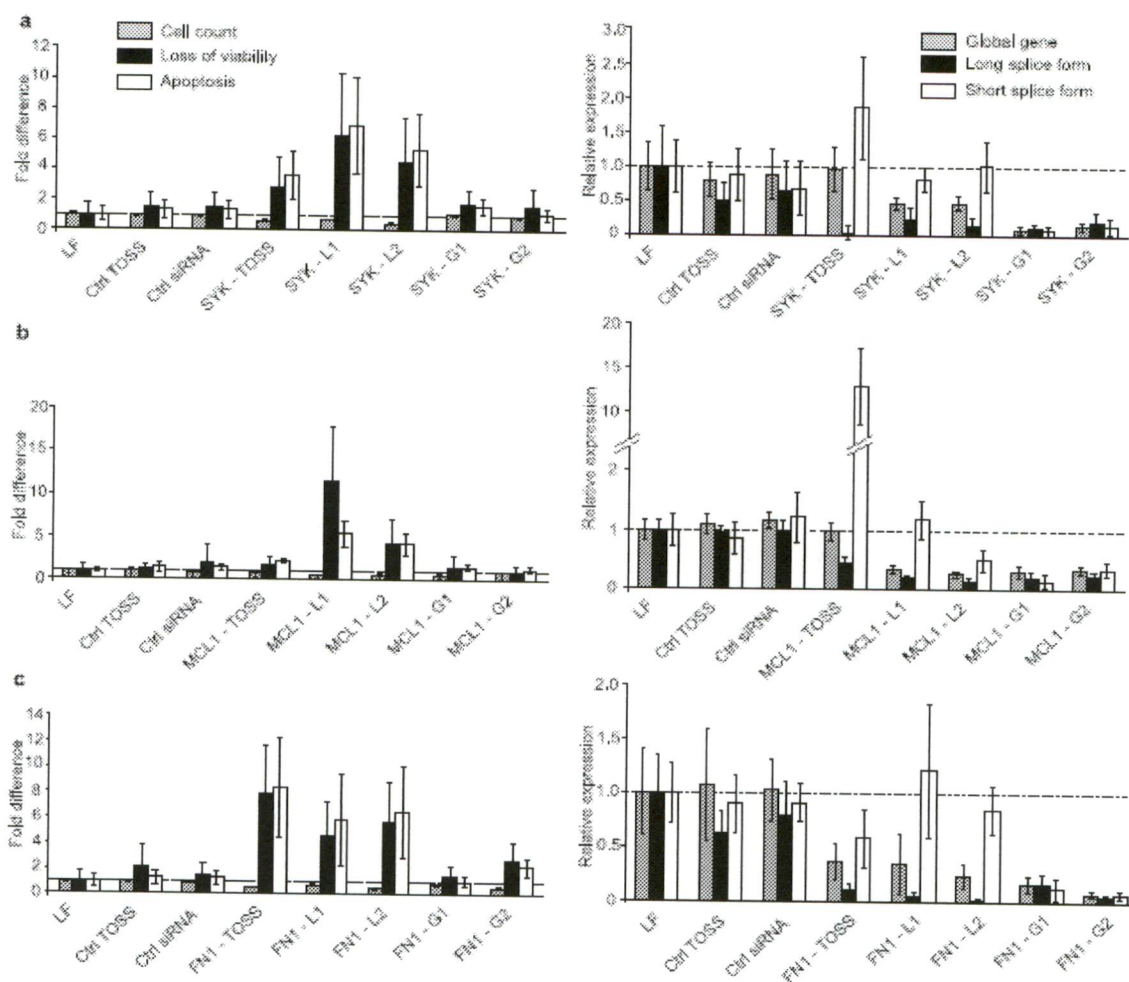


Figure 2 Targeting of SYK, MCL-1 or FN1-EDB alternatively spliced isoforms induces apoptosis.

The impact of TOSS and siRNAs targeting the pre-mRNAs coding for SYK (a), MCL-1 (b) and FN1 (c) proteins on cellular phenotype (left panels) and RNA expression monitored by Q-PCR (right panels) were evaluated and compared to mock transfected (LF) SKOV3ip1 ovarian cancer cells. Experiments were performed in three biological and three technical replicates and an average for each cellular phenotype was calculated for each ISI. The impact of TOSS and alternative exon-specific siRNAs on cell count (grey), loss of viability (black) and apoptosis (white) are shown as bar graphs on the left panels. The Q-PCR relative expression values for the overall gene expression, long splice isoform and short splice isoform are indicated in the right panels by grey, black and white bars respectively. Error bars represent the s.d. of the replicates. siRNA and TOSS with unrelated sequences were used as negative controls. The 3 genes shown were analyzed with a full set of ISIs and siRNAs that inhibit global gene expression (G1, G2).

Inhibition of tumor-associated variants induces apoptosis

The phenotypic effects observed could be due to changes in alternative splice forms or a general reduction in gene expression. To differentiate between these two effects, we compared the impact of siRNA and TOSS on global and isoform-specific expression. Global siRNAs equally inhibited the expression of constitutive, long and short splice isoforms of SYK, MCL1 and FN1 (Figure 2, right panels). Isoform-specific siRNAs inhibited the accumulation of the long, but not the short splice isoforms, while proportionally reducing the overall gene expression. In contrast, reprogramming splicing with TOSS simultaneously reduced the expression of the long and increased the expression of the short splice isoform to various extents for SYK and MCL-1 transcripts without altering overall mRNA levels (Figure 2, right panels). In contrast, TOSS against FN1-EDB decreased the expression of the EDB+ FN1 isoform accompanied by a concomitant decrease in the overall FN1 transcripts levels. Strikingly, while decreasing the overall expression of these genes failed to trigger apoptosis, inhibiting the expression of the long splice isoforms of all three genes induced apoptosis by 4 to 8 fold (Figure 2, left panels), confirming the results of the high-content screen. Consistently, reprogramming the splicing of SYK with TOSS induced apoptosis (Figure 2a, left panels, white bars). TOSS targeting FN1-EDB produced apoptosis by decreasing the long EDB+ isoform and global FN1 expression levels thus behaving in a manner similar to an isoform-specific siRNA (Figure 2c). A second TOSS that modulated FN1-EDB splicing ratio without decreasing overall FN1 expression did not produce apoptosis (data not shown). TOSS targeting MCL-1 did not induce apoptosis despite a significant increase in RNA corresponding to the short isoform (Figure 2b).

The inability of TOSS targeting MCL-1 to induce apoptosis suggests that either the change in the splicing ratio of MCL-1 is not sufficient to induce apoptosis or that TOSS-induced changes in MCL-1 mRNA expression do not lead to corresponding changes in protein isoforms. To differentiate between these two possibilities we performed western blots against MCL-1 in cells transfected with MCL-1 ISIs or control ISIs. MCL-1 isoform-specific siRNAs diminished the expression of the long MCL-1 protein (Suppl. Figure 4a). However, MCL-1 TOSS did not change the levels of the long MCL-1 protein isoform nor

increase the short protein isoform (Suppl. Figure 4a). Therefore, TOSS against MCL-1 fails to induce apoptosis because it does not change the ratio of the corresponding protein isoforms, suggesting the existence of mechanisms that compensate for changes in the RNA expression of this gene in the cell line tested. Indeed, MCL-1 transcripts are known to be extremely unstable and the MCL-1 protein has a rapid turnover of about 30 min¹⁷. Furthermore, exon skipping of MCL-1 produces a frameshift resulting in a premature stop codon thus making this message susceptible to nonsense mediated decay and highly unstable¹⁸. In support of this, we failed to detect any endogenous MCL-1 protein bands corresponding to the short isoform even following TOSS treatment (Suppl. Figure 4a). We also could not verify FN1-EDB protein isoform levels due to their large size (>220kDa), which prevented the SDS gel separation of the isoforms (data not shown), and the absence of suitable isoform specific antibodies. Consequently, we focused our studies on SYK and did not pursue the analysis of FN1 and MCL-1 any further.

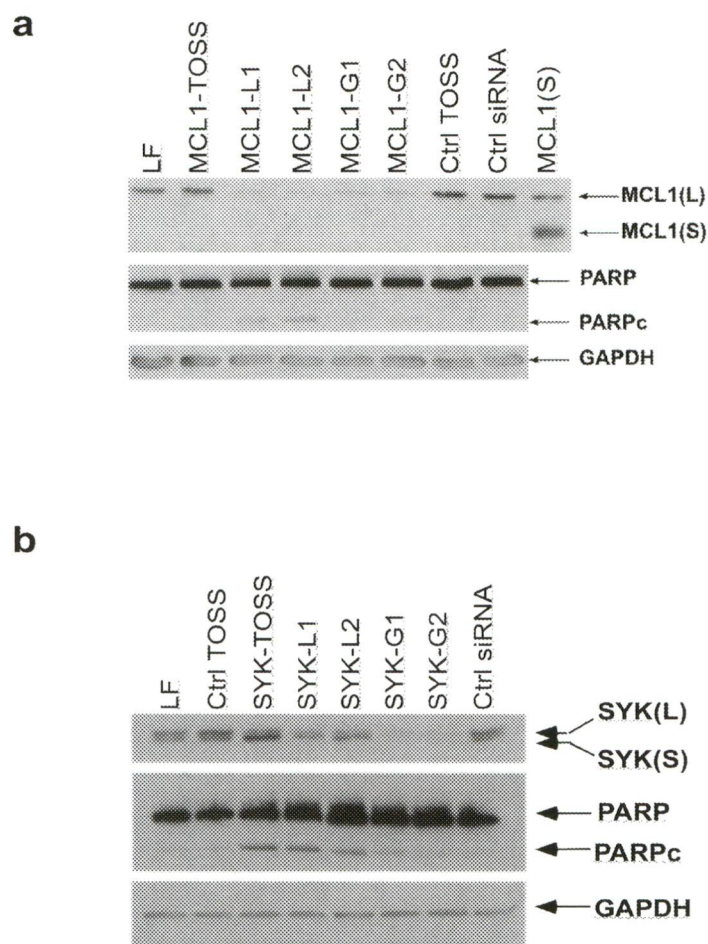


Figure S4 Western validation for MCL-1 and SYK.

The impact of TOSS and siRNAs targeting (a) MCL-1 and (b) SYK pre mRNAs on their respective protein isoforms (top panels), and PARP protein cleavage (middle panels) were evaluated by western blot and compared to mock transfected (LF) cells. Total proteins were extracted 72 hours post-transfection and an equal amount of proteins loaded on an 8% SDS-PAGE. The antibodies used are as described in supplemental materials & methods below. In the last lane of the b) panel, extracts transfected with a plasmid encoding for MCL-1 (S) protein were loaded as a reference for the size of the MCL-1 short isoform.

To verify whether the SYK ISI-induced RNA splicing shifts had an impact at the protein level, western analyses using antibodies against SYK proteins were performed following treatment with different ISIs or control oligonucleotides (Figure 3 and Suppl. Figure 4b). Global siRNAs against constitutively spliced SYK exons decreased all protein isoforms while siRNA against the long SYK isoform decreased the expression of the long but not the short protein isoforms. In contrast, treatment of the cells with TOSS reduced the expression of the long and increased the expression of the short protein isoform of SYK (Figure 3a and Suppl. Figure 4b). Successful modulation of SYK alternative splicing was associated with enhanced cleavage of the apoptosis marker PARP, confirming the specific impact of alternative splicing on apoptosis (Suppl. Figure 4b). The effects of SYK splicing modulation were also investigated in a variety of normal and cancer cell lines expressing different levels and splicing isoform ratios of SYK RNA in order to evaluate the specificity of the observed effects on apoptosis. As expected, transfection of SYK specific ISIs in cells that do not express SYK (e.g. MDA-MB231 and Hs578t breast cancer cell lines) or normal cells (i.e. MCF-10A mammary epithelial cells or BJ-Tielf fibroblasts) did not promote apoptosis (data not shown and Suppl. Figure 5c). In contrast, cells exhibiting the same SYK splice isoform ratio observed in cancer tissues (e.g. HCT116 colorectal carcinoma and TOV-112D endometrioid carcinoma cells) entered into apoptosis upon the modulation of SYK alternative splicing (supplemental figure 5a and 5b). We conclude that the alternative splicing of cancer-associated genes can specifically influence cancer cell viability and apoptosis, in a manner that is largely independent of the overall level of gene expression.

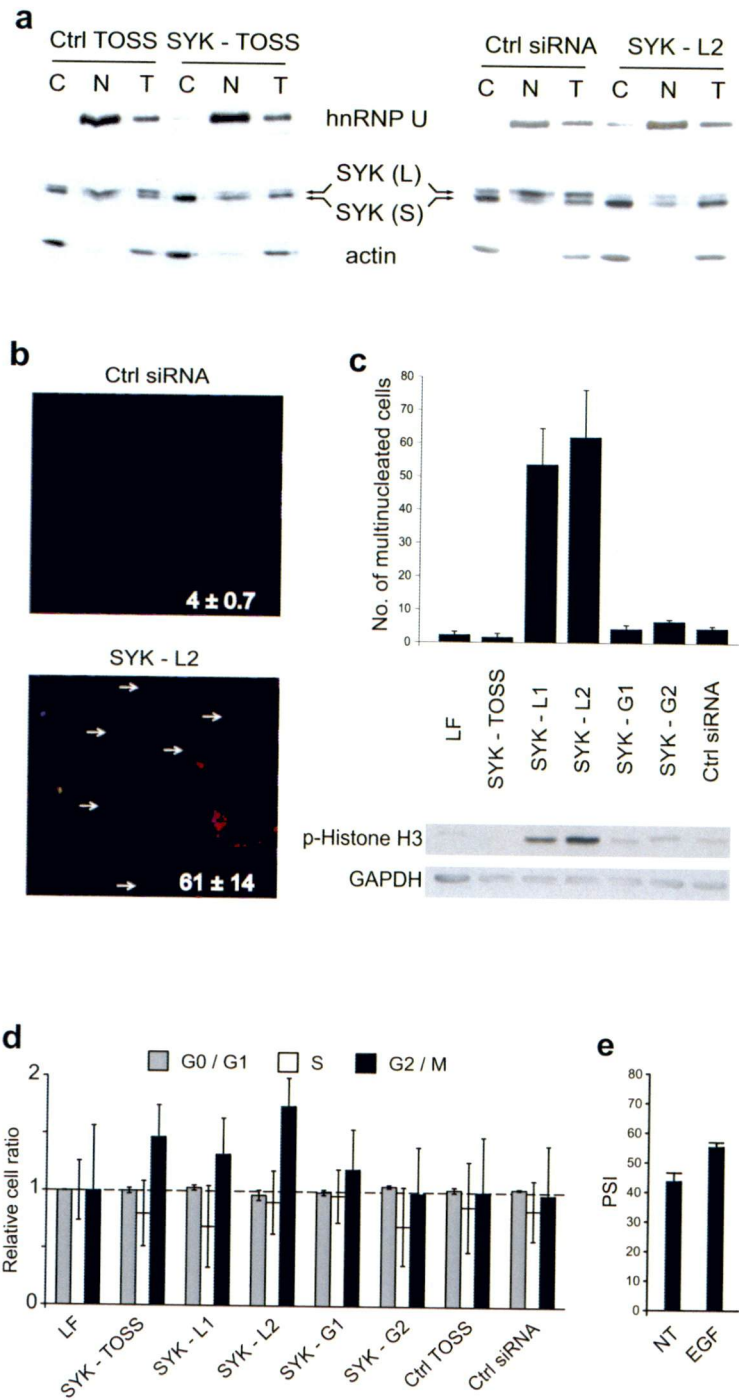


Fig. 3
 Prinos et al., 2010

Figure 3 SYK alternative splicing regulates cell cycle progression and mitosis.

(a) Reducing the expression of the long SYK splice isoform induces the accumulation of the SYK protein in the cytoplasm. Cells were treated by the different ISIs, the nuclei (N) and cytoplasm (C) were separated using differential extraction and centrifugation, proteins extracted and visualized by western blot as indicated in Figure 3 a. Total cellular proteins (T) were also loaded as reference. The nuclear splicing factor hnRNP U and the cytoplasmic protein actin were used as fractionation markers. SYK L and SYK S indicate the position of the long and short splicing isoforms of SYK, respectively. (b) Inhibiting the expression of SYK(L) isoform induces changes in nuclear morphology. The cells were treated with control siRNA or SYK(L) siRNA as indicated and nuclei were stained with Hoechst 96 hours post-transfection. The non-viable cells were stained with propidium iodide. Arrows indicate abnormal or supernumerary nuclei. The numbers in the bottom of each image indicate the average numbers of abnormal nuclei +/- s.d. observed per treatment. (c) SYK alternative splicing regulates mitosis and cytokinesis. Top panel: images from SYK or control ISI transfected cells were visually inspected and cells with abnormal mitosis/supernumerary nuclei were counted and plotted as bar graphs. The numbers are the average of 3 independent transfections. Bottom panel: Western blot of the same samples with a phosphohistone H3 antibody, a known mitosis marker⁴⁵. (d) SYK alternative splicing regulates progression through the G2/M phase of the cell cycle. The cell cycle distribution of SKOV3ip1 cells 48hrs post-transfection with different ISIs was analyzed using flow cytometry. The accumulation of transfected cells in G0-G1 phases (gray), S phase (white) and G2-M phase (black) relative to that of untreated cells is shown as bar graph. The data are an average of three independent transfection experiments and error bars represent the s.d. (e) EGF promotes SYK (L) exon inclusion. Bar graph depicting the changes in SYK splicing upon EGF treatment. The PSI values were calculated as described previously¹¹. Error bars represent the s.d. from 3 independent treatments.

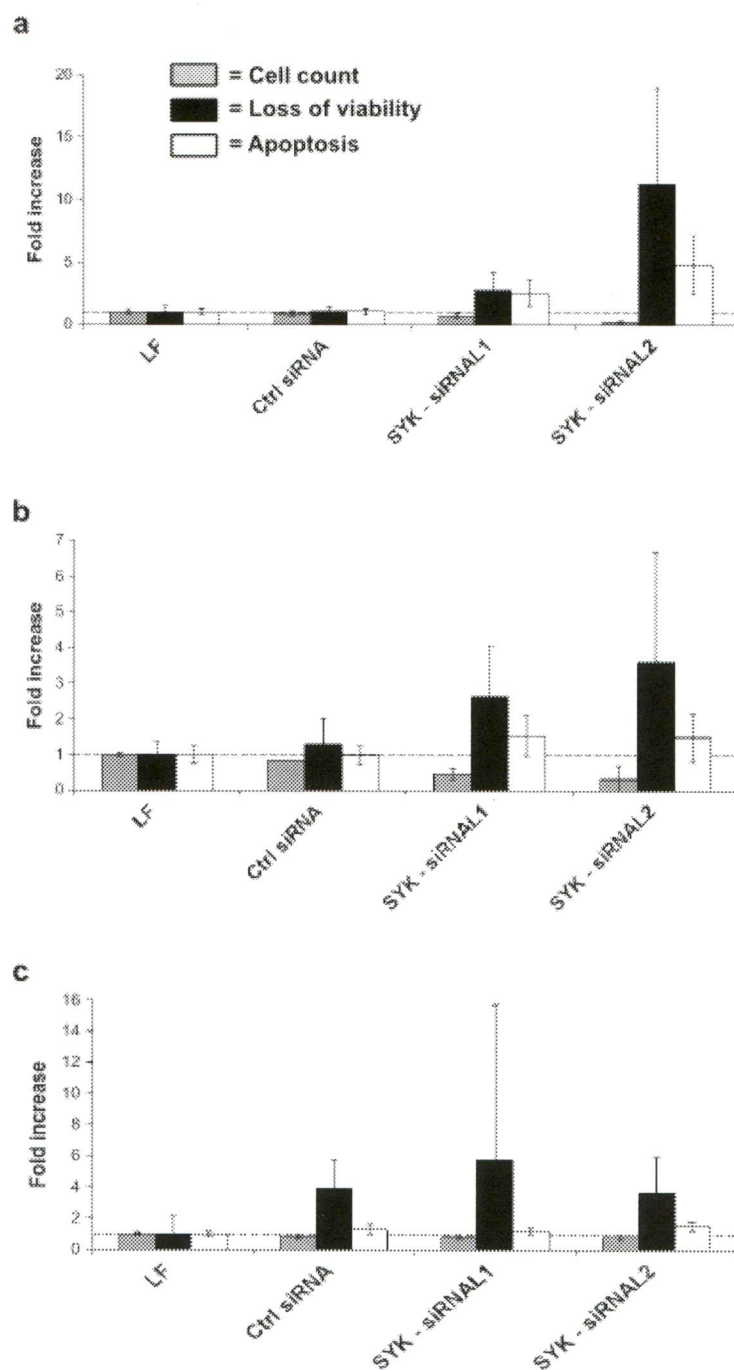


Figure S5 SYK phenotype validation in other cell lines.

The impact of long-isoform specific siRNAs targeting SYK on cell proliferation, viability and apoptosis in a) HCT116 colorectal carcinoma; b) TOV112D endometrioid carcinoma and c) BJ-Tief immortalized fibroblast cell lines. Cell numbers, viability and apoptosis were monitored using fluorescence markers and the relative changes were calculated and used to generate the bar graphs. The error bars represent the variation observed in three technical and three biological replicates.

Alternative splicing of SYK controls cell cycle and mitosis

Previous reports suggested that SYK splice isoforms are localized differently in the cell^{19,20}. Therefore, we investigated the possibility that changes in SYK alternative splicing induce apoptosis by changing the cellular localization of the protein. Cytoplasmic and nuclear proteins were fractionated following treatment with different SYK ISIs and SYK protein isoforms were detected by western blot. In untreated cells, the long SYK isoform, SYK(L), is predominantly nuclear and the short isoform, SYK(S), is predominantly cytoplasmic (Figure 3a). Treatment of the cells with ISIs reduced the expression of the long isoform of SYK in the nucleus and the majority of SYK proteins became cytoplasmic. This suggests that altering the cellular localization of the protein may induce apoptosis. To determine whether the impact of SYK alternative splicing is limited to the regulation of apoptosis/cell survival or extends to other known SYK cellular activities like cell proliferation, we monitored the impact of the ISIs on cell cycle progression. Upon close visual examination of cell images from SYK ISI-treated SKOV3ip1 cells, we noted an increased amount of cells displaying abnormal nuclear morphology with larger, aberrant nuclei and of cells bearing two or more nuclei (Figures 3b, 3c) indicating a role for SYK(L) in mitotic exit and cytokinesis. Changing the alternative splicing of SYK led to an accumulation of cells in the G2/M phase of the cell cycle (Figure 3d) consistent with defective mitosis. The mitotic arrest was confirmed by the increase in phospho-histone H3 (Figure 3c). Notably, the mitotic arrest was isoform-specific as it was not observed by general SYK knockdown with global siRNAs (Figure 3d). Intriguingly, SYK TOSS produced the opposite effect than that of SYK siRNAs i.e. decrease in abnormal mitosis and phospho-histone H3 levels (Figure 3c) suggesting that increasing the expression of the short SYK isoform can compensate for the long isoform dependent effect on mitosis. Given the importance of mitogenic signaling in oncogenesis and cell proliferation, we investigated the effect of oncogenic growth factors on SYK splicing. The pro-survival alternative splicing pattern of SYK was observed *in vitro* upon exposure to epidermal growth factor (EGF, Figure 3e) known to promote cell proliferation and differentiation, underscoring the role of alternative splicing in regulating the activity of this gene. Therefore, we conclude that cancer cell mitosis is regulated by both the amount and ratio of SYK splicing isoforms.

SYK(L) isoform induces JUN-dependent cell survival

To understand how SYK alternative splicing promotes malignant growth, we tested the impact of the different SYK ISIs on the expression of a panel of apoptosis, cell cycle and cancer-relevant genes by Q-PCR. Reducing the expression of the long isoform of SYK resulted in prominent increases in the expression of JUN²¹ as measured by Q-PCR in SKOV3ip1 cells (Supplemental Table 3). Notably, this result was observed when targeting the expression of the long SYK isoform by either siRNA or TOSS, but not when the overall expression of SYK was downregulated by siRNA against constitutive exons (Supplemental Table 3). This suggests that changing the splicing pattern of SYK and not the global expression impacts the MAPK/JNK signalling pathway that acts as a conduit for apoptosis²². To directly test this hypothesis, we evaluated the effect of a JNK inhibitor SP600125²³ on SYK-induced apoptosis. As shown in Figure 4a, the inhibition of JNK with SP600125 partly rescued the apoptosis induced by modulation of SYK(L) isoform expression in SKOV3ip1 cells (Figure 4a), indicating that the SYK-induced phenotype is mediated at least in part by JUN activation. We conclude that the production of SYK(L) by alternative splicing promotes JNK-dependent cell survival.

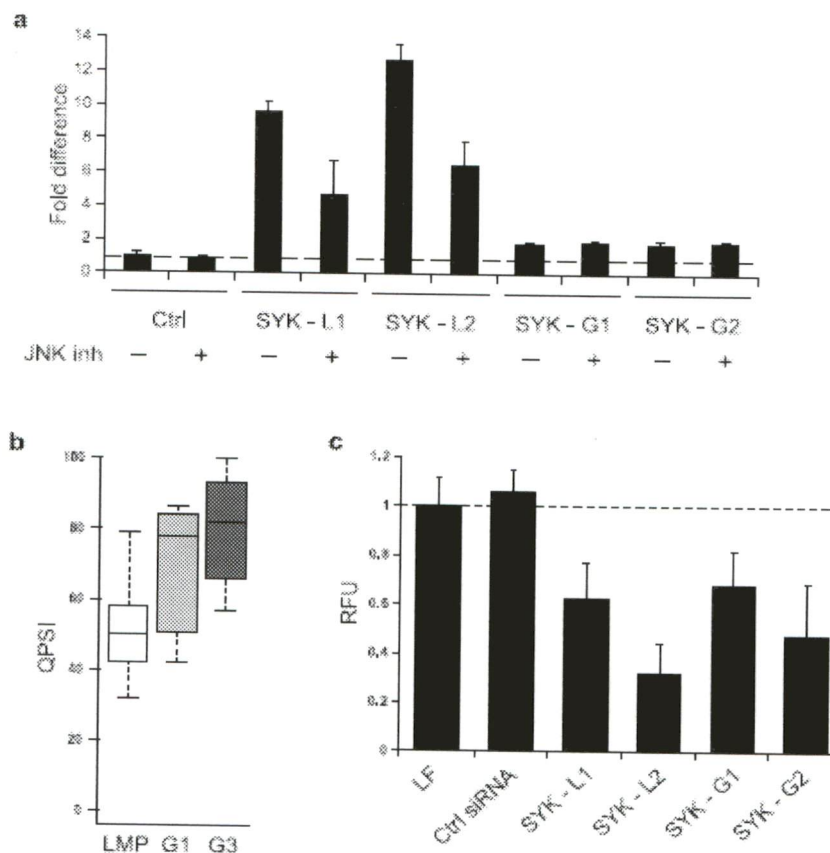


Figure 4 SYK(L) promotes malignancy and anchorage-independent growth.

(a) SYK alternative splicing regulates JUN-dependent cell survival. SKOV3ip1 cells were transfected with siRNAs targeting SYK (L). 24 hours later cells were treated with the JNK inhibitor SP600125 or DMSO control. The cells were allowed to grow for 72 hours and subsequently apoptotic cells were stained with annexin-V as in figure 2a. Results represent the average of three individual experiments. (b) Expression of SYK long splice isoform is a marker for tumor malignancy. The ratio of SYK splice isoforms was monitored by Q-PCR in RNA extracted from 6 borderline (LMP), grade 1 (G1) or grade 3 (G3) ovarian tumors. The percent splicing index (Q-PSI) was calculated as described¹² and the relative values are presented as box plots. (c) Altering the alternative splicing pattern of SYK inhibits anchorage-independent growth. SYK ISI-transfected SKOV3ip1 cells, or cells transfected with control ISIs, were seeded in 96-well plates containing soft agar and incubated for 7 days in order to form colonies. Cells were subsequently stained with Hoechst and total fluorescence was measured and plotted against ISI controls. The time frame of the assay precluded the use of the TOSS. Error bars represent the s.d. of three independent experiments.

SYK(L) promotes tumor malignancy and anchorage-independent growth

Our results suggest that increased expression of SYK long isoform in tumors may not only increase cell survival, but also cellular proliferation and tumor aggressiveness. To test this hypothesis we monitored SYK splicing in a total of 18 ovarian cancer tissues from tumors of low malignant potential (LMP), grade 1 tumors (G1) and the aggressive grade 3 tumors (G3). Consistent with the notion that SYK regulates tumor aggressiveness, the expression of the SYK long isoform was higher in G1 tumors compared to LMP tissues and higher still in G3 tumors (P-value between LMP and G3 = 0.001, Figure 4b). To further dissect the mechanism by which SYK splicing promotes aggressive tumor growth, we monitored the impact of silencing the SYK(L) isoform on anchorage-independent growth in vitro. As shown in Figure 4c, the ability of SKOV3ip1 cells to form colonies in soft agar was markedly inhibited with siRNAs against the long isoform or against both long and short isoforms. This indicates that, unlike the impact of SYK on apoptosis and mitosis that requires a minimum expression of the short isoform, the effect on tumorigenesis is only influenced by the expression level of the long form independent of the splicing ratio. We conclude that the expression level of the SYK long splicing isoform regulates tumor aggressiveness by promoting anchorage-independent growth and cell proliferation.

Alternative splicing regulation is a common feature of the MAPK/JNK pathway

In order to determine whether or not alternative splicing dependent regulation of gene function is limited to SYK or extends to other members of the MAPK/JNK pathway, we tested the functional impact of all cancer-associated ASEs⁵ in this pathway. A list of all cancer associated splicing events in genes with ontology terms related to phosphoinositide-3 kinase signal transduction or MAPK/JNK (http://palace.lgfus.ca/link/pheno_screen_pi3k-akt) was examined for function in vitro as described in Figure 1. From this, we identified 13 cancer-associated isoforms (see http://palace.lgfus.ca/link/mapk-jnk_annotation for an annotation table) and validated previously undefined isoform specific function for MAP3K7 (TAK1) and MINK1 kinases. In both cases, silencing the long splicing isoform of these two genes reduced cell viability by inducing apoptosis in SKOV3ip1 cells (Supplemental Figure 6). We conclude that alternative splicing plays a critical role in modulating the production of isoforms of players regulated by the MAPK/JNK pathway.

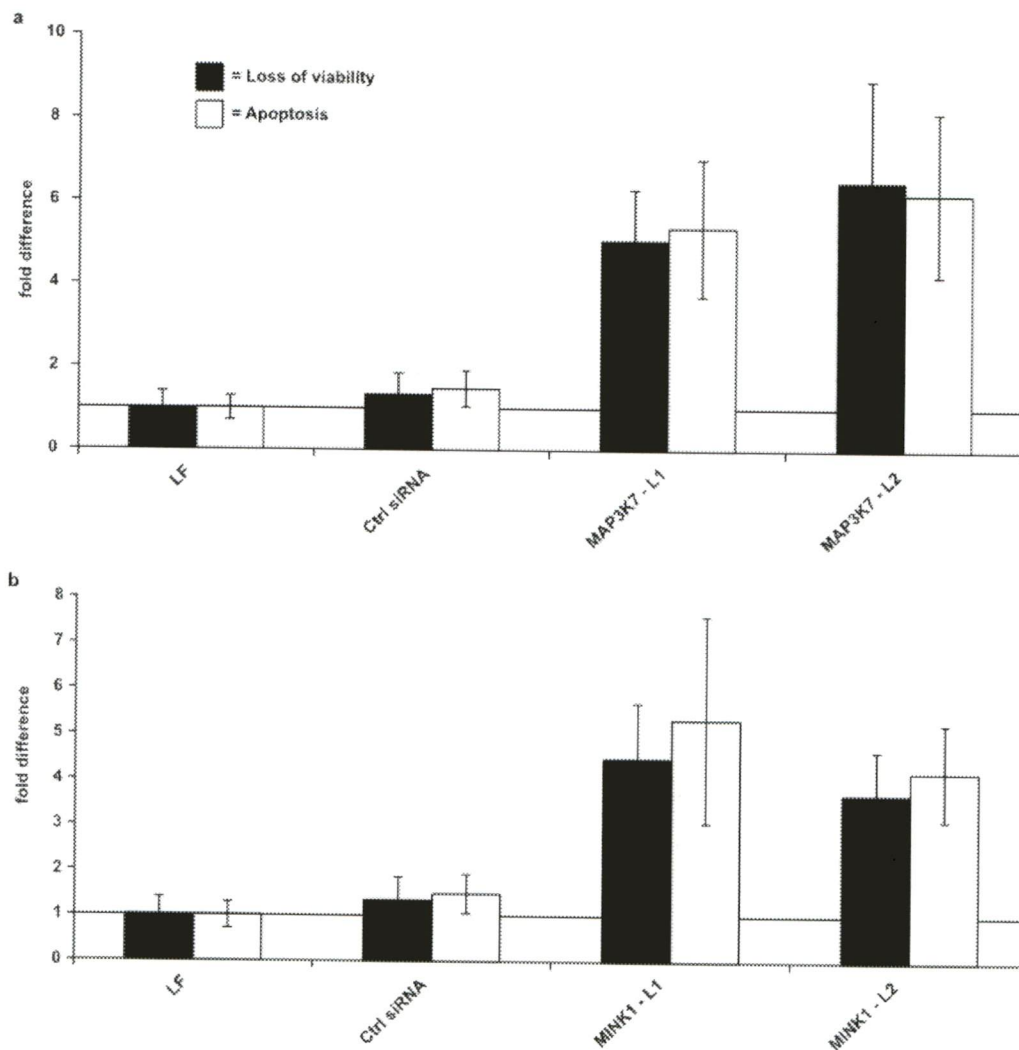


Figure S6 MAPK isoforms essential for cell viability.

The impact of ISIs against the pre-mRNAs coding for MAP3K7 and MINK1 isoforms on the viability (black) and apoptosis (white) phenotypes of SKOV3ip1 cells is shown relative to lipofectamine and ISI controls. The error bars represent the variation observed in three technical and three biological replicates.

Discussion

In this study we have presented a procedure for evaluating the functional impact of alternative splicing on phenotypic parameters that are relevant to cancer. Moreover, we have uncovered cancer-specific alternative splicing events that promote cell proliferation and survival. Approximately, 10% of all tested breast and ovarian cancer-associated splicing events significantly affected cell viability *in vitro* confirming that splicing pattern changes in cancer are not simply an inconsequential anomaly of diseased cells, but instead may have been selected for functional advantages.

The strongest impact we observed on apoptosis was after inhibiting the expression level of the long isoform of MCL-1, a member of the BCL2 mitochondrial apoptotic regulator family. MCL-1 is primarily localized to the outer mitochondrial membrane where it promotes cell survival by interacting with other Bcl-2 family members²⁴. Active MCL-1 promotes cell viability by inhibiting the activation of caspases leading to apoptosis. MCL-1 has a short half-life due to its regulation at multiple levels, including alternative splicing, allowing the rapid induction by survival signals or elimination of the protein by apoptotic signals²⁵. Alternative splicing leads to MCL-1 exon 2 skipping and deletion of the C-terminal transmembrane domain resulting in protein mislocalization to the cytosol²⁴. Our results suggest that apoptosis is regulated by the expression level of the long isoform of MCL-1 (Figure 2 and Suppl. Figure 4a). This observation is somewhat different from earlier experiments showing that over-expression of the MCL-1 short isoform, which lacks the pro-survival transmembrane and BH2 domains, causes apoptosis^{18,26}. In our hands apoptosis was induced upon decreasing the expression level of the long isoform even when the expression of short isoform was not detectable by western blot (Supplemental Figure 4a), suggesting that apoptosis is not strictly dependent on increasing the amount of the short isoform. Apoptosis was not induced to the same levels by siRNA against constitutive exons compared with those against those targeting the long isoform. However, close inspection of the knockdown level suggests that global inhibition in these experiments did not reduce the amount of the long isoform to the same level observed with the long isoform specific siRNA (supplemental figure 4a and data not shown). Once again the results are consistent with a long isoform specific effect on apoptosis. Indeed, inhibition of the MCL-1 long

isoform by itself in skin cancer cell lines was also shown to cause apoptosis²⁷. Therefore, in the case of MCL-1, apoptosis appears to be regulated at least in part by changing the expression level of the long isoform.

In addition to regulating the splicing of apoptotic genes, tumorigenesis also impairs apoptosis by altering the splicing of kinases such as SYK²⁸. We have found that ovarian tumors preferentially express the long isoform of SYK. Intriguingly, SYK is required for leukemic B cell proliferation and survival by regulating MCL-1 expression^{15,16}. This isoform was previously reported to affect the subcellular distribution of the SYK protein kinase via a non-classical nuclear localization signal (NLS) located in the alternative exon¹⁹. Indeed, our data show that the SYK(L) isoform is nuclear and changes in SYK splicing affect its nuclear localization, consistent with these observations. In addition, we find that modulation of SYK splicing causes cell cycle perturbations and defective mitosis. The involvement of SYK in cell proliferation and mitotic progression has been documented in breast cancer cells^{29,30}. Furthermore, SYK has been shown to localize to the centrosomes where it associates with γ -tubulin³¹. SYK is also found in microtubules³² and it phosphorylates α -tubulin³³, thus providing a possible explanation for why changing the splicing of this gene impairs mitosis. On the other hand, the mechanism by which SYK splicing exerts its effect on apoptosis appears to be mediated by the activity of c-JUN. The expression of this protein that regulates cell cycle and apoptosis²² was found to be modified upon changing SYK splicing (suppl. table 3). In addition, the SYK isoform-dependent apoptotic phenotype was rescued in the presence of JNK inhibitors underlining the importance of c-JUN expression for the induction of SYK-dependent apoptosis. Together the results are consistent with a model where cancer cells promote cell survival by inducing the expression of SYK pro-survival isoform that in turn suppresses the expression of c-JUN and its target genes thus inhibiting apoptosis (figure 5).

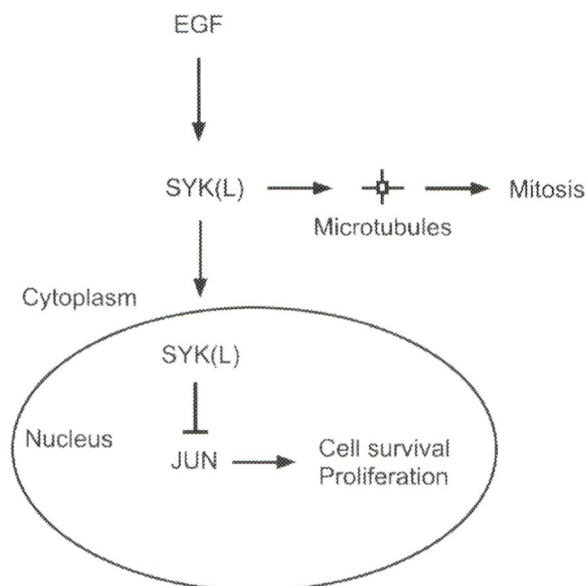


Figure 5 A model depicting the different functions of the SYK(L) pro-survival isoform.

Extracellular growth factors such as EGF activate signal transduction pathways. The EGF signal promotes a splicing shift in the SYK long isoform, which affects SYK protein isoform subcellular localization and activities. SYK(L) can have at least two distinct functions depending on its cellular location: it can enter the nucleus where it normally represses JUN expression and promotes cell survival; cytoplasmic SYK(L) can associate with the centrosome and microtubules where it is essential for mitotic spindle assembly and cytokinesis.

SYK acts as a tumor suppressor in breast cancer and loss of its expression is required for malignant growth and tumor cell invasion²⁹. Paradoxically, SYK can also act as an oncogene in other types of tumors and SYK over-expression has been associated with T-cell lymphomas³⁴ and chronic leukemias³⁵, as well as gastric cancer³⁶ and head and neck carcinomas³⁷ where its expression correlates with tumor progression and is of prognostic value. These contradictory expression profiles of SYK in different tumor types are difficult to explain based on global gene expression. Confusion about the role of SYK in different cancers may be due in part to the methods of evaluating the expression and the functional impact of SYK because no distinction was made in these studies between the global and splice isoform-specific increase in SYK expression. It is possible that differential expression of SYK(L) and SYK(S) may contribute to these disparate observations in

different cancers. In support of this, Wang and coworkers¹⁹ observed elevated SYK(S) expression in human breast carcinomas, but not in matched normal mammary tissues, suggesting a role for SYK(S) in mammary tumor progression. Consistent with a role for SYK alternative splicing in tumorigenesis, Goodman et al. have observed aberrant SYK splicing in childhood acute B-cell leukemias³⁸. Alternatively, the differences in SYK behavior in different cancer types may be due to differences in the contribution of SYK to tumor biology. In this study, we have found that while apoptosis is affected only by changes in the splicing pattern of SYK (figure 2a), mitosis and cell proliferation are affected by both changes in SYK splicing and expression levels (figure 3), while anchorage-independent growth (figure 4c) is mostly driven by overall SYK expression and not splicing. Accordingly, tumors depending on SYK for viability will be mostly dependent on SYK splicing and its impact on c-JUN, which is cell type specific, while those tumors using SYK to modulate cell adhesion, will be mostly affected by changes in the overall expression level of SYK. In our hands, we found that ovarian cancer and colorectal cancer cell lines are most sensitive to changes in SYK splicing while invasive breast cancer cell lines generally do not express SYK²⁹ and therefore are not sensitive to SYK modulation in vitro (supplemental Fig. 5 and data not shown). Indeed, SYK splicing pattern was linked to ovarian and not breast cancer in vivo^{11,12}. In all cases, our observations underline the need to re-examine previously established links between SYK gene expression and tumor behavior with a clear distinction between isoform and global changes in gene expression.

The functional contribution of alternative splicing to cancer cell biology reported here is not static but responds to extracellular signals. Indeed, we observed increased exon inclusion in SYK and FN1-EDB alternative exons with short-term EGF treatment (figure 3e). This finding is consistent with previous reports indicating that FN1 alternative splicing is regulated by growth factors through the PI3K/AKT and Ras signal transduction pathways³⁹. The effect of EGF on alternative splicing of SYK and FN1 is likely to be mediated by splicing modulators such as proteins of the SR or hnRNP families⁴⁰. Indeed, knocking down hnRNP-K in SKOV3ip1 cells changes the splicing pattern of these genes and promotes apoptosis (supplementary figure 7). Strikingly, EGF has been reported to regulate hnRNP-K expression and activity⁴¹, thus suggesting that the EGF effects on SYK and FN1

alternative splicing could be mediated by hnRNP-K. Intriguingly, exon junction complex components have recently been shown to regulate MAPK splicing and EGF signaling in *Drosophila*^{42,43}. It is thus tempting to speculate on the existence of an additional layer of complexity involving alternative splicing in the EGF / Ras signaling axis with functional consequences in tumor progression (supplemental figure 8). Indeed, we found that apoptosis is induced upon altering the splicing of the MAP3K7 and MINK genes, both part of the MAPK cell survival pathway, which were previously linked to ovarian and breast cancers⁵ and Ras-induced growth arrest in ovarian epithelial cells⁴⁴, respectively. Based on these data, we propose that splicing acts as a modulator of cancer cell survival / cell death by regulating the EGF/MAPK/JNK pathway. It is also important to note that the impact of alternative splicing on the cell survival pathway is likely to be stronger than what is shown in this study since the functional impact of many splicing events associated with cancer in vivo might be missed due to the limited conditions and cell type that may be sampled in vitro.

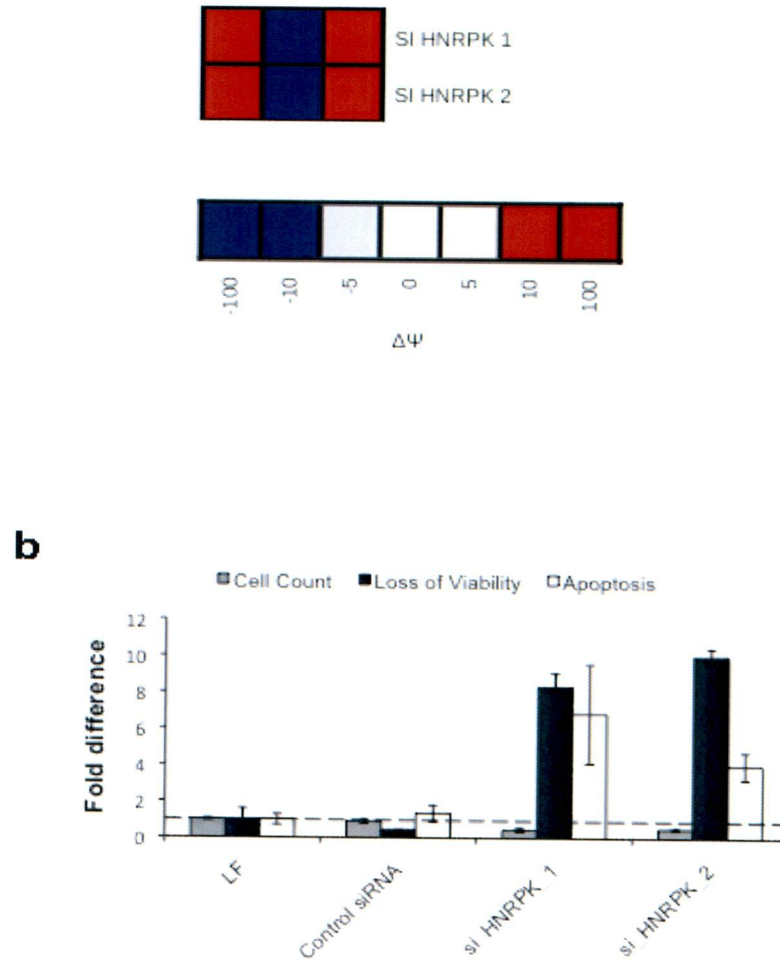


Figure S7 hnRNP-K controls SYK, MCL-1 and FN1 alternative splicing ratios (a) and is essential for SKOV3ip1 cell survival (b).

HnRNP-K silencing was performed as described previously⁴⁸. (a) The effect of hnRNP-K siRNA-mediated knockdown on the alternative splicing profiles of SYK, FN-EDB and MCL-1 was assessed by RT-PCR as described before⁴⁸. (b) The effect of hnRNP-K silencing in SKOV3ip1 cells was assessed by the multiplex viability, apoptosis, proliferation assay as described in Suppl. Figure 1.

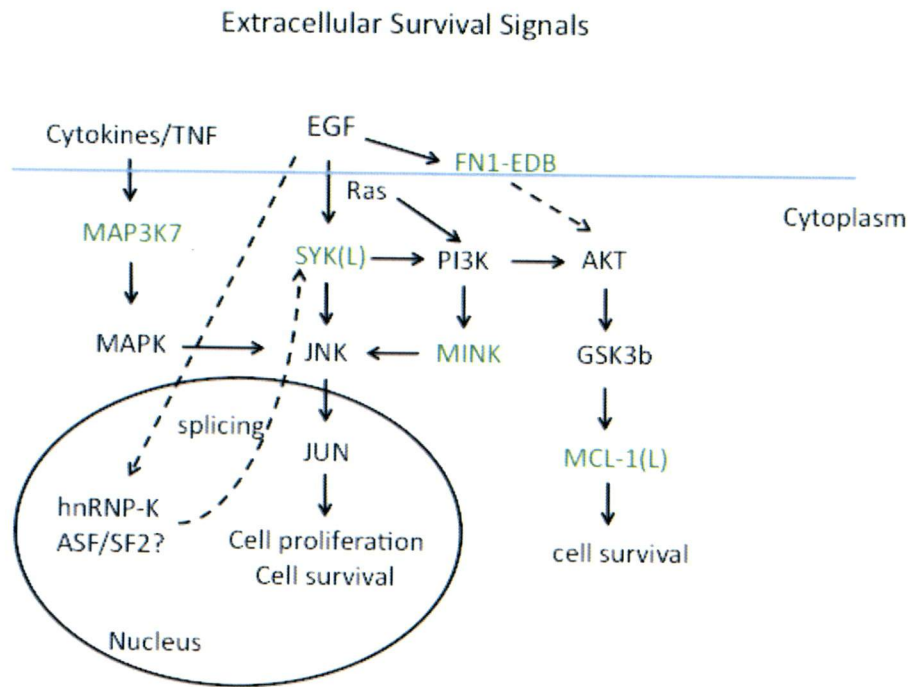


Figure S8 Alternative splicing regulation of EGF/MAPK signaling.

Model depicting how EGF signaling may affect alternative splicing and subsequently diversify the signaling repertoire in the EGF/ MAPK/ JNK pathways by regulating the expression of the functional isoforms identified in this study (depicted in green). The functional isoforms (FN1, SYK, MCL-1, MAP3K7 and MINK) required for cell viability are highlighted in green.

Acknowledgements

We are indebted to Claudine Rancourt (Departement de Microbiologie, Université de Sherbrooke) for providing cell lines and help in the initial set-up phase of the project, Bruno Lamontagne and Lucien Jr. Bergeron for assay set-up and help with troubleshooting. We thank the Réseau de Recherche sur le Cancer FRSQ tissue bank for ovarian tumor tissues and Leonid Volkov for help with flow cytometry. This work was funded by a grant from Genome Canada and Génome Québec. B. Chabot is the Canada Research Chair in Functional Genomics. Jean-Pierre Perreault is the Canada Research Chair on Genomics and Catalytic RNA. S. Abou Elela is a Chercheur National of the Fonds de la Recherche en Santé du Québec (FRSQ).

Author contributions

D.G, M.B., D.G., S.C., J.P.B. and E.L. carried out experiments and analyzed data and J-F.L., P.T. and J.G-B. analyzed data and prepared figures. J-F.L. developed the ISI design program and the FASE statistics and bioinformatics analysis tools. J.P.B., K.T., E.L. and P.T. developed the Q-PCR procedures used to evaluate ISI impact on splicing. H.N. did the histopathological review of tissue specimens. P.P., K.T., J.V., R.K., J. P.P., B.C., R.J.W. and S.A.E. designed experiments, discussed data and participated in the writing of the paper. P.P. and K.T. supervised experiments and analyzed data. P.P. and S.A.E. wrote the manuscript.

Methods

Cell culture and transfection. All cell lines were cultured in antibiotic- and antimycotic-free media (Wisent). The ovarian adenocarcinoma SKOV3ip1 cell line was grown in DMEM/F12 (50/50) medium supplemented with 10 % fetal bovine serum and 2 mM L-Glutamine. TOV112D endometrioid carcinoma cells and HCT116 colorectal adenocarcinoma were propagated in OSE medium or McCoy's modified 5a medium (Wisent) respectively supplemented with 10% FBS and 2mM L-Glutamine. Cell propagation and passaging were as recommended by ATCC (American Type Culture Collection). Recombinant human EGF (Pepro-Tech) was used at 50ng ml⁻¹ for the EGF induction experiment. For image analyses and RNA extractions, SKOV3ip1 cells were seeded in 96-well tissue culture-treated imaging plates (BD Biosciences) at a density of 5,000 cells per well. For immunoblotting and FACS analyses, cells were seeded at 350,000 cells per well in 6-well plates (BD Biosciences). Transfections were performed in suspension using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol and 10 nM of 21-mer siRNA duplexes or 400 nM of TOSS oligonucleotides with 2'OMe chemical modifications (IDT). Each oligonucleotide was transfected in randomly positioned wells in triplicate (technical replicates) on at least three different transfection dates (biological replicates). The transfection efficiencies were confirmed by transfecting a fluorescent AlexaFluor-488 conjugated negative control siRNA (AllStars siRNA, Qiagen) in every plate. For the rescue experiments, JNK inhibitor II (SP600125, EMD Biosciences) was added 24 hours post-transfection at 10 μ M final concentration.

SiRNA and TOSS oligo design. To design long isoform-specific siRNA and TOSS, an algorithm identifying the alternatively included region was employed⁴⁶. All possible stretches of 19 (siRNA) or 20 (TOSS) nucleotides were extracted and scored (Supplementary Figure 2). The top scoring siRNAs and TOSS were used for the FASE screen. Refer to the supplementary material for all sequences. To perform global gene knockdown using siRNA, a global gene junction-ranking algorithm was developed (JF Lucier, unpublished).

RNA extraction and quantitative PCR. Total RNA extractions were performed in 96-well plates with the Absolutely RNA Microprep Kit (Stratagene) as recommended by the manufacturer except that DNase treatments were done at 37°C. RNA quality and presence of contaminating genomic DNA was verified as described⁴⁶. RNA integrity was assessed for 11 random samples per 96-well plate with an Agilent 2100 Bioanalyzer. Reverse transcription was performed at 55°C on a fixed volume (5.5 µl) of RNA sample with Transcriptor (Roche Diagnostics), random hexamers, dNTPs and 10 units of RNaseOUT (**Invitrogen**) in a total volume of 10 µl. All cDNAs were diluted by adding 65 µl of RNase DNase-free water. Quantitative PCR reactions were performed on 1 ng cDNA as described elsewhere⁴⁸. Relative expression levels were calculated using the qBASE framework⁴⁷ using PSMC4 and SDHA as reference genes. Primer design and validation were evaluated as described elsewhere⁴⁶. Additionally, an algorithm was developed to design primer pairs targeting constitutive exon junctions of a given gene and were tested as described⁴⁶. In every run, a no template control and a no reverse transcription control (no RT) on a mock-transfected sample was performed for each primer pair. These controls were consistently negative.

Apoptosis/viability/cell enumeration triplex assay. A 96-well multiplex assay was performed to simultaneously monitor viability, apoptosis and cell count in each well. Cells were stained 96 h post-transfection with the following three fluorescent dyes: 1 µg ml⁻¹ of propidium iodide (Invitrogen), 1:350 of annexin V-Alexa Fluor 647 (Invitrogen) and 800 ng ml⁻¹ of Hoechst 33342 (Invitrogen) in an annexin binding buffer (pH 7.4) composed of 10 mM HEPES, 140 mM NaCl and 2.5 mM CaCl₂. High-content endpoint fluorescence images were acquired on a BD Pathway 855 BioImager (BD Biosciences) using a 20X objective with a laser-based auto focus and excitation and emission filters appropriate for each fluorescent dye. Experimental error was reduced with the acquisition of nine images (800 µm gap between each frame) per well per dye. Image pre-processing and segmentation were performed with Attovision 1.6 software (BD Biosciences). Thresholds were applied manually to identify propidium iodide and annexin V positive cells.

Soft Agar Colony Formation Assay. SKOV3ip1 cells were seeded at 5,000 cells per well and transfected in 96-well tissue culture-treated plates. 24 hrs after transfection, the cells were trypsinized and seeded in a 1.2% agar base layer (MJS BioLynx) containing DMEM/F12 supplemented with 20 % FBS and 4 mM L-Glutamine. Once the cell layer was solidified, 100 μ l of medium was added to each well and the plates were incubated for 7 days at 37°C under 5 % CO₂. On the seventh day, medium was removed, and the cells were stained for 1 hour with Hoechst 33342 (5 μ g ml⁻¹). Fluorescence was measured on a FLX800 microplate fluorescence reader (BioTek Instruments).

Immunoblotting. Cells were collected from 6-well plates at 72 h post-transfection, pooled and lysed as described previously⁵⁰. Refer to Supplemental Methods for the conditions and the antibodies used.

Statistical analysis and quality control metrics. A pilot screen was employed to establish the dynamic range and robustness of the assay with z-factor scores routinely in excess of 0.5 (Supplementary Fig. 3). The z-factor⁴⁹ was determined by calculating the average and standard deviation of the randomly distributed triplicate wells of the positive control RBM8A versus the lipofectamine control wells for each plate.

Data processing and hit identification. The primary screening assay produced a multiplex readout of cell count, viability and apoptosis by a combination of 3 vital dyes (Fig 1). Triplicate technical replicates of triplicate biological transfections were performed. A complete randomization of the plate was employed each time to eliminate systematic instrument measurement errors (Supplemental Figure 1). All valid plates were normalized by comparing to lipofectamine treatment (Supplemental Figure 9). A stringent statistical cut-off of 3 median absolute deviations^{50,51} above the median measurements per assay (corresponding to a robust z-score>3) was employed to minimize false positives and ensure true positive phenotypes⁵⁰. This method is robust to outliers and can identify weak hits and is thus deemed ideal for RNAi screens⁵⁰. All the steps for data processing, normalization and hit calling are outlined in Supplemental Figure 9.

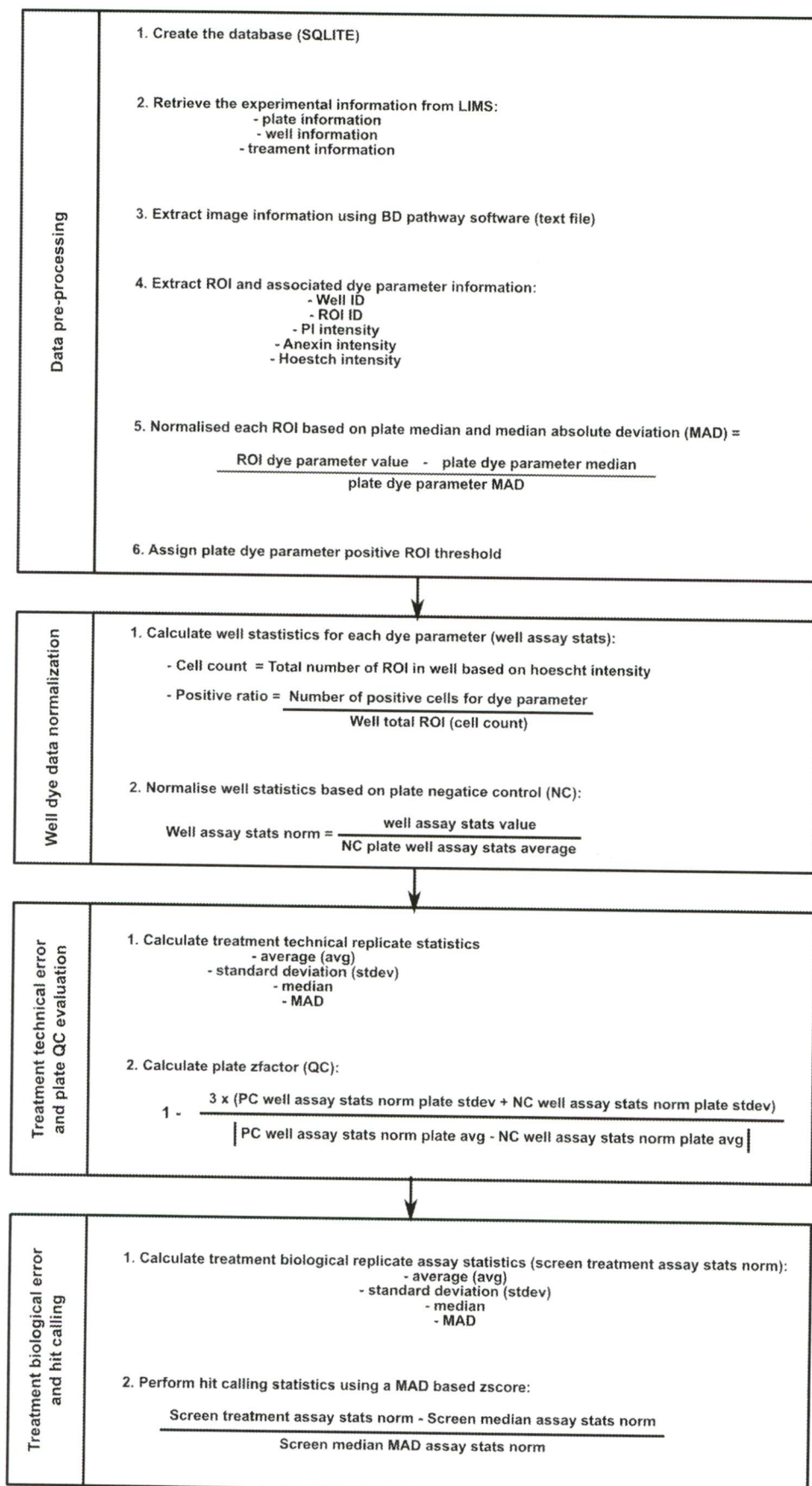


Figure S9 Data analysis flowchart displaying all the data processing, normalization and bioinformatic and statistical analysis steps for the FASE high-content screen.

After import of raw data in a database, the data is processed to extract meaningful predefined parameters for each fluorescence measurement. Quantitative phenotypic data is then normalized to a standard treatment (LF) and ranked according to Z-scores. ROI: region of interest; MAD: median absolute deviation⁵¹; NC: negative control; PC: positive control; QC quality control.

References

1. Pan, Q., Shai, O., Lee, L.J., Frey, B.J. & Blencowe, B.J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**, 1413-5 (2008).
2. Wang, E.T. et al. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-6 (2008).
3. Matlin, A.J., Clark, F. & Smith, C.W. Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* **6**, 386-98 (2005).
4. Castle, J.C. et al. Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat Genet* **40**, 1416-25 (2008).
5. Venables, J.P. et al. Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol* **16**, 670-676 (2009).
6. Venables, J.P. Aberrant and alternative splicing in cancer. *Cancer Res* **64**, 7647-54 (2004).
7. Cooper, D.L. & Dougherty, G.J. To metastasize or not? Selection of CD44 splice sites. *Nat Med* **1**, 635-7 (1995).
8. Christofk, H.R. et al. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452**, 230-3 (2008).
9. Levanon, E.Y. & Sorek, R. The importance of alternative splicing in the drug discovery process. *TARGETS* **2**, 109-114 (2003).
10. Iorns, E., Lord, C.J., Turner, N. & Ashworth, A. Utilizing RNA interference to enhance cancer drug discovery. *Nat Rev Drug Discov* **6**, 556-68 (2007).
11. Klinck, R. et al. Multiple alternative splicing markers for ovarian cancer. *Cancer Res* **68**, 657-63 (2008).
12. Venables, J.P. et al. Identification of alternative splicing markers for breast cancer. *Cancer Res* **68**, 9525-31 (2008).
13. Celotto, A.M. & Graveley, B.R. Exon-specific RNAi: a tool for dissecting the functional relevance of alternative splicing. *RNA* **8**, 718-24 (2002).
14. Villemaire, J., Dion, I., Elela, S.A. & Chabot, B. Reprogramming alternative pre-messenger RNA splicing through the use of protein-binding antisense oligonucleotides. *J Biol Chem* **278**, 50031-9 (2003).
15. Baudot, A.D. et al. The tyrosine kinase Syk regulates the survival of chronic lymphocytic leukemia B cells through PKCdelta and proteasome-dependent regulation of Mcl-1 expression. *Oncogene* **28**, 3261-73 (2009).
16. Gobessi, S. et al. Inhibition of constitutive and BCR-induced Syk activation downregulates Mcl-1 and induces apoptosis in chronic lymphocytic leukemia B cells. *Leukemia* **23**, 686-97 (2009).
17. Adams, K.W. & Cooper, G.M. Rapid turnover of mcl-1 couples translation to cell survival and apoptosis. *J Biol Chem* **282**, 6192-200 (2007).
18. Bingle, C.D. et al. Exon skipping in Mcl-1 results in a bcl-2 homology domain 3 only gene product that promotes cell death. *J Biol Chem* **275**, 22136-46 (2000).
19. Wang, L. et al. Alternative splicing disrupts a nuclear localization signal in spleen tyrosine kinase that is required for invasion suppression in breast cancer. *Cancer Res* **63**, 4724-30 (2003).

20. Zhou, F., Hu, J., Ma, H., Harrison, M.L. & Geahlen, R.L. Nucleocytoplasmic trafficking of the Syk protein tyrosine kinase. *Mol Cell Biol* **26**, 3478-91 (2006).
21. Shaulian, E. AP-1--The Jun proteins: Oncogenes or tumor suppressors in disguise? *Cell Signal* **22**, 894-9.
22. Wisdom, R., Johnson, R.S. & Moore, C. c-Jun regulates cell cycle progression and apoptosis by distinct mechanisms. *EMBO J* **18**, 188-97 (1999).
23. Bennett, B.L. et al. SP600125, an anthrapyrazolone inhibitor of Jun N-terminal kinase. *Proc Natl Acad Sci U S A* **98**, 13681-6 (2001).
24. Akgul, C. Mcl-1 is a potential therapeutic target in multiple types of cancer. *Cell Mol Life Sci* **66**, 1326-36 (2009).
25. Michels, J., Johnson, P.W. & Packham, G. Mcl-1. *Int J Biochem Cell Biol* **37**, 267-71 (2005).
26. Bae, J., Leo, C.P., Hsu, S.Y. & Hsueh, A.J. MCL-1S, a splicing variant of the antiapoptotic BCL-2 family member MCL-1, encodes a proapoptotic protein possessing only the BH3 domain. *J Biol Chem* **275**, 25255-61 (2000).
27. Shieh, J.J., Liu, K.T., Huang, S.W., Chen, Y.J. & Hsieh, T.Y. Modification of alternative splicing of Mcl-1 pre-mRNA using antisense morpholino oligonucleotides induces apoptosis in basal cell carcinoma cells. *J Invest Dermatol* **129**, 2497-506 (2009).
28. Mocsai, A., Ruland, J. & Tybulewicz, V.L. The SYK tyrosine kinase: a crucial player in diverse biological functions. *Nat Rev Immunol* **10**, 387-402.
29. Coopman, P.J. et al. The Syk tyrosine kinase suppresses malignant growth of human breast cancer cells. *Nature* **406**, 742-7 (2000).
30. Moroni, M. et al. Progressive loss of Syk and abnormal proliferation in breast cancer cells. *Cancer Res* **64**, 7346-54 (2004).
31. Zyss, D. et al. The Syk tyrosine kinase localizes to the centrosomes and negatively affects mitotic progression. *Cancer Res* **65**, 10872-80 (2005).
32. Faruki, S., Geahlen, R.L. & Asai, D.J. Syk-dependent phosphorylation of microtubules in activated B-lymphocytes. *J Cell Sci* **113** (Pt 14), 2557-65 (2000).
33. Peters, J.D., Furlong, M.T., Asai, D.J., Harrison, M.L. & Geahlen, R.L. Syk, activated by cross-linking the B-cell antigen receptor, localizes to the cytosol where it interacts with and phosphorylates alpha-tubulin on tyrosine. *J Biol Chem* **271**, 4755-62 (1996).
34. Feldman, A.L. et al. Overexpression of Syk tyrosine kinase in peripheral T-cell lymphomas. *Leukemia* **22**, 1139-43 (2008).
35. Buchner, M. et al. Spleen tyrosine kinase is overexpressed and represents a potential therapeutic target in chronic lymphocytic leukemia. *Cancer Res* **69**, 5424-32 (2009).
36. Nakashima, H. et al. Clinical significance of nuclear expression of spleen tyrosine kinase (Syk) in gastric cancer. *Cancer Lett* **236**, 89-94 (2006).
37. Luangdilok, S. et al. Syk tyrosine kinase is linked to cell motility and progression in squamous cell carcinomas of the head and neck. *Cancer Res* **67**, 7907-16 (2007).
38. Goodman, P.A., Wood, C.M., Vassilev, A., Mao, C. & Uckun, F.M. Spleen tyrosine kinase (Syk) deficiency in childhood pro-B cell acute lymphoblastic leukemia. *Oncogene* **20**, 3969-78 (2001).
39. Blaustein, M. et al. Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nat Struct Mol Biol* **12**, 1037-44 (2005).

40. Shin, C. & Manley, J.L. Cell signalling and the control of pre-mRNA splicing. *Nat Rev Mol Cell Biol* **5**, 727-38 (2004).
41. Mandal, M. et al. Growth factors regulate heterogeneous nuclear ribonucleoprotein K expression and function. *J Biol Chem* **276**, 9699-704 (2001).
42. Ashton-Beaucage, D. et al. The exon junction complex controls the splicing of MAPK and other long intron-containing transcripts in *Drosophila*. *Cell* **143**, 251-62.
43. Roignant, J.Y. & Treisman, J.E. Exon junction complex subunits are required to splice *Drosophila* MAP kinase, a large heterochromatic gene. *Cell* **143**, 238-50.
44. Nicke, B. et al. Involvement of MINK, a Ste20 family kinase, in Ras oncogene-induced growth arrest in human ovarian surface epithelial cells. *Mol Cell* **20**, 673-85 (2005).
45. Hendzel, M.J. et al. Mitosis-specific phosphorylation of histone H3 initiates primarily within pericentromeric heterochromatin during G2 and spreads in an ordered fashion coincident with mitotic chromosome condensation. *Chromosoma* **106**, 348-60 (1997).
46. Brosseau, J.P. et al. High-throughput quantification of splicing isoforms. *RNA* **16**, 442-9 (2010).
47. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* **8**, R19 (2007).
48. Venables, J.P. et al. Multiple and specific mRNA processing targets for the major human hnRNP proteins. *Mol Cell Biol* **28**, 6033-43 (2008).
49. Zhang, J.H., Chung, T.D. & Oldenburg, K.R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J Biomol Screen* **4**, 67-73 (1999).
50. Birmingham, A. et al. Statistical methods for analysis of high-throughput RNA interference screens. *Nat Methods* **6**, 569-75 (2009).
51. Chung, N. et al. Median absolute deviation to improve hit selection for genome-scale RNAi screens. *J Biomol Screen* **13**, 149-58 (2008).

Chapitre 4

Programmed modulation of alternative splicing in the tumor microenvironment

Auteurs de l'article: **Brosseau JP**, Lucier JF, Nwilati H, Thibault P, Garneau D, Durand M, Couture S, Lapointe E, Prinos P, Klinck R, Perreault JP, Chabot B and Elela SA.

Statut de l'article : Soumis et en révision depuis le 19 juillet 2012, *Nature Structural & Molecular Biology*

Avant-propos

JPB a produit toutes les figures. JPB a écrit le brouillon. SAE et JPB ont participé à la révision de toutes les versions subséquentes du manuscrit. Voir la section « author contribution » pour plus de détails.

Mise en contexte

La cancérogenèse modifie l'expression et la maturation des ARNm, qui sont normalement finement régulés par divers mécanismes, y compris l'épissage alternatif. Bien que plusieurs exemples ponctuels de dérégulation d'isoforme de l'épissage aient été répertoriés, il n'est pas clair à quel point les isoformes de l'épissage alternatif contribuent à la cancérogenèse. Nous avons exploité la méthodologie de quantification d'isoformes développée au chapitre 1 afin de mettre en lumière un réseau d'épissage alternatif dérégulé dans le microenvironnement des tumeurs ovariennes.

Les travaux présentés dans ce chapitre ont pour but de démontrer que 1) l'expression de certaines isoformes sont modifiés dans les tumeurs indépendamment de leur composition cellulaire hétérogène; 2) l'expression globale de facteurs d'épissage associés aux tumeurs corrèle avec le changement d'expression d'isoformes associés aux tumeurs; 3) certains de ces changements ont lieu dans le microenvironnement de la tumeur.

Résumé

Il est bien établi que l'épissage alternatif est perturbé dans plusieurs types de cancers. Par contre, il n'est pas clair si l'épissage alternatif est une conséquence secondaire ou est directement impliqué dans le développement de la tumeur. Dans cet ouvrage, nous avons identifié des isoformes de l'épissage alternatif spécifiques au cancer qui ne sont pas affectées par la composition cellulaire des tissus. De façon surprenante, les résultats les plus intéressants proviennent des cellules stromales adjacentes à la tumeur (microenvironnement). En effet, nos résultats indiquent que la répression de RBFOX2 et QKI provoque l'exclusion d'un sous-ensemble d'exon à caractère mésenchymal. En somme, nos résultats indiquent que le transcriptome des cellules du microenvironnement est modifié pour permettre aux tumeurs épithéliales de se développer.

*Article 3: Programmed modulation of alternative splicing in the tumor microenvironment**Abstract*

Pre-mRNA alternative splicing is modified in cancer, but the origin and specificity of these changes remain unclear. Here, we probed ovarian tumors to identify cancer-specific splice isoforms that are not affected by variations in tumor tissue composition. Surprisingly, the most informative events occurred in the tumor microenvironment where the suppression of a subset of mesenchymal exons was caused by the downregulation of splicing factors RBFOX2 and QKI. Together, our results indicate that the architecture of transcripts is systematically and specifically redesigned in the tumor microenvironment.

Introduction

Ovarian cancer is the fourth most common cause of cancer-related death in women, largely because it is typically diagnosed at advanced stages¹. Like most solid tumors, the majority of ovarian cancers are epithelial in origin. By definition, ovarian cancer leads to an enrichment of the ovarian epithelial content. Ovaries that are normally composed of a thin layer of epithelial cells and a large mass of stromal cells that include fibroblasts, smooth muscle cells and endothelial cells are dramatically transformed into a large epithelial mass in advanced cancer stages². In addition, the composition of the cells surrounding the tumor (the tumor microenvironment) changes as the tumor grows and gets established³. The tumor microenvironment is composed of extracellular matrix (ECM), genetically stable cancer-associated fibroblasts (CAFs), immune cells and soluble factors required for cancer progression and metastasis^{4,5}. Interplay between tumor and adjoining stromal tissues has been observed to be an important aspect of the tumorigenic process^{6,7}. Indeed, it was shown that epithelial ovarian cancer cells could directly induce a CAF phenotype (i.e. changes of normal fibroblasts into CAFs) via secretion of transforming growth factor beta^{8,9}. At the same time, CAFs may secrete growth factors such as hepatocyte growth factor to promote cancer cell proliferation and invasion¹⁰.

Gene expression, and more recently alternative splicing, a highly regulated and cell type-specific process, have been found to be globally altered in cancer cells¹¹⁻¹³. Profiling the global expression resulted in several sets of biomarkers capable of detecting cancer subtypes and was particularly successful in differentiating between different breast cancer subtypes^{14,15}. However, most expression profiling techniques focus on changes in the levels of gene expression and simply ignore changes in the transcript architecture due to alternative splicing. Profiling the expression of splicing isoforms identified potential markers for ovarian and breast cancers, and some of these were shown to be required for cell survival *in vitro*^{11,16,17}. Hundreds of splicing isoforms associated with the cell cytoskeleton and other tissue-specific genes were associated with ovarian cancer when RNA extracted from whole tumor tissues was compared to that extracted from normal ovaries¹². This made the distinction between cancer and tissue-specific markers difficult^{12,18}. The search for biomarkers is further complicated by the high epithelial plasticity of

cancer cells^{19,20}; it is believed that some cancer cells undergo an epithelial-to-mesenchymal transition (EMT) to increase their migratory and invasive potential²¹⁻²³. The splicing of several EMT-associated genes was shown to be modified in cancer cell lines^{24,25}, but the role of alternative splicing in defining the epithelial or mesenchymal status of cancer cells in patient tissues remains unclear. Comparison between different cell lines of different tissue origin or during EMT identified several EMT and mesenchymal-to-epithelial (MET) specific alternatively spliced isoforms as well as associated splicing factors like ESRP1/2^{24,25}. Some of these splice variants and their regulators may be linked to cancer as suggested by *in vitro* studies²⁵. However, their general association with tumor tissues and their specific roles in cancer development remains to be confirmed.

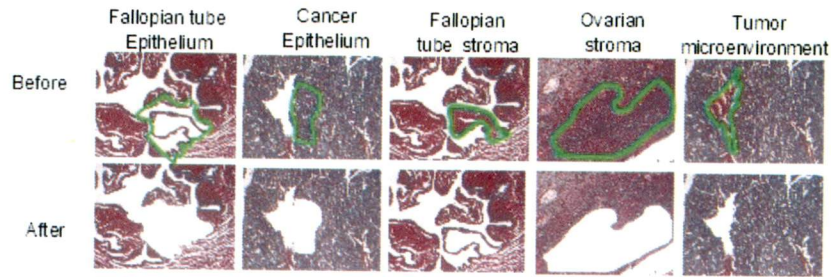
Recently, enriched cell populations generated by laser capture microdissection (LCM) have been useful in the identification of tumor-specific expression markers^{26,27}. In this study we have used this approach to identify a highly specific group of cancer-associated ASEs by isolating tumors and their microenvironment and comparing them with their reciprocal normal tissues. Strikingly, most of the ASEs associated with the tumor microenvironment resulted in the exclusion of mesenchymal-enriched exons, and many of these events were regulated by the cancer-associated splicing factors QKI and RBFOX2. Our results indicate that cancer cells induce programmed changes in a restricted set of functionally related mRNAs.

Results

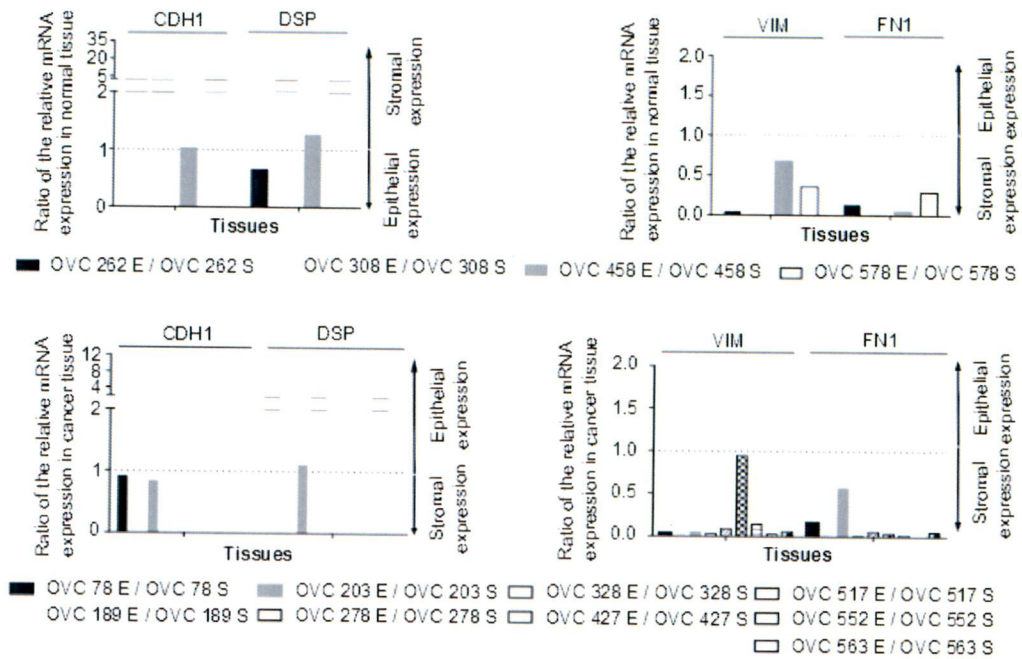
Identification of cancer-specific modulation of alternative splicing

Identifying the molecular characteristics of cancer by comparing tumor and normal tissues is often hampered by tissue heterogeneity. Therefore, we sought to obtain relatively pure cell populations from representative tumor and normal tissues by laser capture microdissection (LCM). Normal ovaries and high-grade serous ovarian tumors were dissected and visual inspection indicated that homogeneous populations of stromal and epithelial cells were obtained from each tissue (Fig. 1a). Given the low number of epithelial cells in normal ovaries, fallopian tube, which is considered the origin of serous ovarian cancer^{28,29}, was used as a source for normal epithelium. The dissected portions of the tumor microenvironment and normal stroma were mostly composed of fibroblasts (Fig. 1a), and only samples that had a similar number of fibroblasts, immune cells and endothelial cells were used for RNA extraction (see Material and Methods for details). The identity of the isolated cells was also confirmed molecularly by monitoring the global expression levels of established epithelial and stromal markers³⁰ by quantitative RT-PCR (Fig. 1b). As expected, epithelial-specific genes including the surface protein E-cadherin CDH1³¹ and the desmosome component desmoplakin DSP^{32,33} were overexpressed in the epithelial cell populations (Fig. 1b). Conversely, overexpression of the intermediate filament protein VIM³⁴ and the cell adhesion factor FN1³⁴ (stromal markers) were observed specifically in the stromal cell populations (Fig. 1b). To compare the overall epithelial and mesenchymal status of the different samples we monitored the global expression of a panel of established tissue markers in normal and cancer tissues. In general, the expression of 53 epithelial and stromal markers confirmed the similar overall epithelial and stromal nature of both cancer and normal tissues (Fig. 1c). Consistently, immunohistochemical analysis of ovarian cancer tissues confirms the predominant expression of stromal markers (FN1 and VIM) in the ovarian tumor microenvironment and the expression of the epithelial marker (CDH1) in ovarian cancer cells (<http://www.proteinatlas.org/>). Together these tests confirm the homogeneity and comparable nature of the normal and cancer cell populations isolated by LCM.

a



b



c

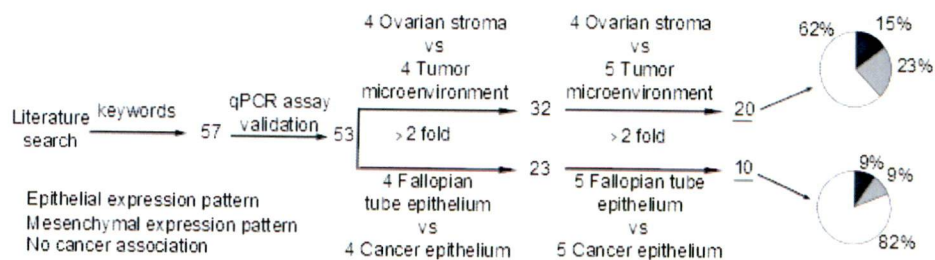


Figure 1 Brousseau 2012

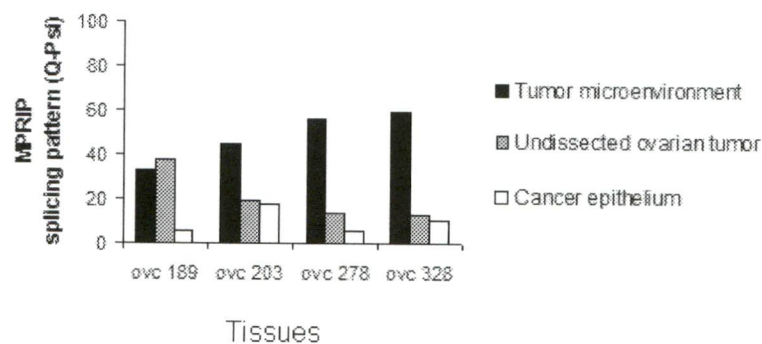
Figure 1 Laser capture microdissection (LCM) of normal and cancer ovarian tissues.

(a) Shown are hematoxylin and eosin stain samples from Fallopian tubes, ovaries, and high-grade serous ovarian tumors visualized by light microscopy before and after dissection at 10x magnification. The limits of dissected tissues are indicated in green and the type of tissues examined is indicated on top. Dissected tissues with similar cell population in both normal and cancer samples were chosen for RNA extraction. Typically, dissected Fallopian tube epithelium was composed of 90% epithelial and 10% fibroblast cells; dissected cancer epithelium was composed of more than 95% epithelial and less than 5% fibroblast and other non epithelial cells (e.g. endothelial and inflammatory cells); dissected Fallopian tube stroma, dissected ovarian stroma and dissected tumor microenvironment was composed of 60-80 % fibroblast, 10-20% endothelial and 10-20% of inflammatory cells (see Materials and Methods for details) (b) The quality of the dissected tissues was assessed by examining the expression of established epithelial (e.g. CDH1 and DSP; left panels) and stromal (e.g. VIM and FN1, right panels) markers by quantitative RT-PCR. The global expression of the different markers was determined in the epithelium and stroma of four-paired normal Fallopian tube (upper panels) and nine paired high-grade serous ovarian cancer samples (bottom panels). Relative values were normalized to housekeeping genes as previously described¹⁷ and the expression ratio (epithelium / stroma) was plotted as bar graph. The dashed line indicates the threshold separating stromal and epithelial expression levels. Detailed description (Supplementary Table 13) and individual expression values (Supplementary Table 14) of the dissected tissues are provided as supplementary information. (c) Expression pattern of classical epithelial and mesenchymal markers in dissected ovarian cancer tissues. Quantitative RT-PCR assays were developed for the classical gene expression markers associated with epithelial and stromal tissues in NCBI PubMed database⁶⁴. The expression of 53 stromal mesenchymal and epithelial markers was monitored in four dissected Fallopian tube and cancer epithelium as well as four ovarian and tumor microenvironment. Relative values were normalized to housekeeping genes as described in a. Markers that distinguish between normal and cancer tissues by at least two fold in expression level were re-examined using an independent set of epithelial and stromal tissues (see Materials and Methods for the details). The resulting number of markers capable of discriminating between cancer and normal tissues is underlined. The epithelial / mesenchymal expression pattern of the newly identified cancer-associated markers is illustrated in the form of pie charts.

To investigate cancer-specific alternative splicing, we examined the splicing profiles of 870 simple ASEs (cassette exons, alternative 5' or 3' splice sites and intron retention events) found in the RefSeq database³⁵ in dissected normal and cancer tissues by quantitative RT-PCR (Fig. 2a) (the raw data are accessible through <http://palace.lgfus.ca/pcreactiongroup/list/226>). Comparison of the splicing patterns obtained from the Fallopian tube and cancer epithelium identified eight cancer-specific ASEs, referred to hereafter as the cancer epithelial signature (CES) (Supplementary Table 1, column 1). On the other hand, comparison of the tumor microenvironment to normal

stroma splicing patterns revealed 10 cancer-specific ASEs, referred to hereafter as the cancer stromal signature (CSS) (Supplementary Table 2, column 1). The splicing profiles obtained from dissected tissues displayed distinct patterns that cannot be reproduced by simple subtraction from the profile obtained from the original non-dissected samples (Supplementary Fig. 1). Accordingly, the systematic and consistent changes in splicing cannot be explained by random variation in the quality of dissection. The CSS and CES splicing signatures were not linked to changes in the expression levels of the cognate genes, indicating that the splicing differences are not simply the result of a change in mRNA levels (Supplementary Table 1, column 2 and Supplementary Table 2, column 2). We conclude that the changes in alternative splicing not only distinguish between tumor and normal tissues, but also permit the discrimination between the tumor microenvironment from normal stroma.

a



b

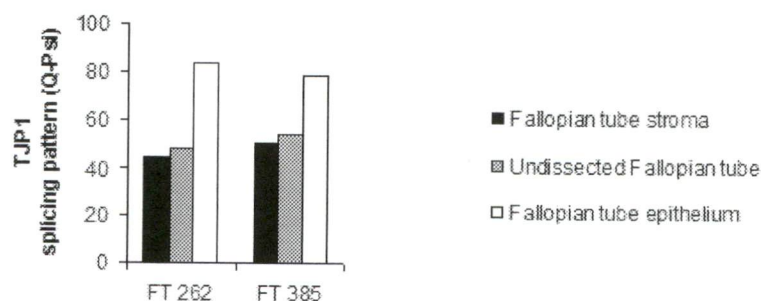


Figure S1 Comparison between the splicing patterns obtained in undissected and dissected ovarian cancer and Fallopian tube mirror image samples.

Shown are histograms representing the quantitative splicing pattern ($Q\psi$)¹ ($Q\psi$ of 100 mean that only the long isoform is present and $Q\psi$ of 0 mean that only the short isoform is present) of (a) a tumor microenvironment associated splicing event (MPRIP) in LCM tumor microenvironment (black), undissected ovarian tumor (gray) and LCM cancer epithelium (white) from four ovarian serous high-grade tissues and (b) a cancer epithelium associated splicing event (TJP1) in LCM Fallopian tube stroma (black), undissected Fallopian tube (gray) and LCM Fallopian tube epithelium (white) from two normal Fallopian tubes as detected by quantitative RT-PCR. Detailed description of the tissue compared are provided in Supplementary Table 13.

The clinical usefulness of the newly identified cancer-specific signatures was evaluated by monitoring their capacity to distinguish between undissected normal and cancer tissues. Five out of the 10 CSS ASEs (*ABII*, *CAST*, *CD46*, *MPRIP* and *PBRM1*), were able to distinguish normal from cancer tissues with $P < 5.0e-07$ (t-test), while three others (*CD68*, *GIT2*, *MSRB3*) distinguished between the normal and cancer tissues with $P < 0.01$ (t-test) (Supplementary Table 2, column 4). In the case of CES, three out of the eight ASEs distinguished between cancer and normal tissues with $P < 0.01$ (t-test) (Supplementary Table 1, column 4). These results indicate that the CSS is more apt than the CES at detecting cancer in RNA extracted from undissected tumors, perhaps due to the low epithelial content of tissues extracted from normal ovary. Indeed, the CES misidentified all Fallopian tube tissues as cancer samples (Fig. 2b, left panel), while the CSS correctly classified four out of six Fallopian tube samples as normal tissues in spite of their epithelial contents (Fig. 2b, right panel). Overall, the different analyses of the CSS and CES behavior in undissected tissues identified five CSS ASEs (*CAST*, *CD46*, *MPRIP*, *MSRB3* and *PBRM1*) that can detect cancer regardless of the tissue purity and the cellular composition of the control (normal ovaries or Fallopian tube) tissues (Supplementary Table 2, column 4). This later group of five CSS ASEs, which could be useful in the clinic, was named the ovarian cancer signature (OCS). Three ASEs (*CAST*, *CD46* and *PBRM1*) from this OCS were also able to discriminate between undissected tissues originating from 18 normal breast and 20 ductal breast cancer with $P < 5.0e-07$ (t-test) further confirming the value of this signature (Supplementary Table 3). ASEs that are not included in this group nevertheless remain interesting for the purpose of understanding cancer biology and the

mechanism regulating alternative splicing in cancer using dissected tissues. We conclude that while both the cancer epithelium and the tumor microenvironment may exhibit cancer-specific alternative splicing patterns, the ASEs that emerged by comparing the tumor microenvironment and normal stromal tissues are more powerful at identifying cancer and less dependent on tissue origin and cell- type composition.

Figure 2 Splicing of a subset of mesenchymal-enriched exons is repressed in the ovarian tumor microenvironment.

(a) Strategy for the identification of cancer-specific alternative splicing patterns. Expression of the 3313 simple ASEs (i.e. cassette exons, alternative 5' and 3' splice sites, and intron retention) annotated in the RefSeq database 36.3³⁵ was examined using endpoint PCR amplification of 5 RNA samples extracted from normal Fallopian tube tissue and ovarian cancer tissues (see Materials and Methods for the details). Validated quantitative RT-PCR assays were developed for a total of 870 expressed ASEs and the resulting PCR values were used to calculate the mean quantitative splicing shift¹⁷ ($\Delta Q\psi = Q\psi_{\text{TUMOR}} - Q\psi_{\text{NORMAL}}$) in the different normal and cancer tissues described in Fig. 1c and Supplementary Table 13 (see Materials and Methods for the details). ASEs were ranked based on a t-test of the mean quantitative splicing shift and ASEs displaying statistically significant differences between normal and cancer tissues ($P < 0.05$, t-test) and a $\Delta Q\psi > 15$ were re-examined in an independent set of dissected tissues (see Materials and Methods for the details). The final number of validated ASEs capable of discriminating between normal and cancer epithelium (CES) or stroma (CSS) is indicated. (b) Stromal markers detected in undissected cancer tissues. The quantitative percent splicing pattern ($Q\psi$)¹⁷ of the CES and CSS ASEs was determined by quantitative RT-PCR in an independent set of undissected ovarian tissues containing 9 high-grade serous ovarian cancer, 6 Fallopian tube and 14 normal ovarian tissues (Supplementary Table 15). The results are displayed in the form of two heatmaps representing the splicing patterns of the CES (left panel) and the CSS (right panel) in the different tissues. Gene names and the gene clusters are shown on the y-axis. The type of tissues is indicated at bottom and the tissue clusters are shown on top. Cancer and normal tissues are indicated by black and white boxes, respectively. Epithelial contents are indicated by gray gradients at bottom, where white indicated no epithelial contents and dark gray indicated 100% epithelial cells. The color code representing the different splicing patterns is indicated at bottom. Dark orange indicates complete exon exclusion and bright yellow signifies complete exon inclusion. (c) Epithelial splicing isoforms are enriched in the tumor microenvironment. The cancer epithelial signature (CES) and the cancer stromal signature (CSS), which include the ASEs capable of detecting cancer in both dissected tissues and whole tumors were divided into groups based on their expression patterns in normal mesenchymal and epithelial tissues (see Material and Methods for details). The constitutive exons are indicated as white boxes and introns as lines. Epithelial- and mesenchymal-enriched exons are indicated in black and gray boxes, respectively. Gene names are indicated at bottom.

Cancer represses the inclusion of a subset of mesenchymal-enriched exons

Many of the ASEs in the cancer epithelium signature (CES) were found in genes linked to different aspects of tumorigenesis such as apoptosis (e.g. *CD47*³⁶) and other related cellular functions. In contrast, the ASEs in the tumor microenvironment (CSS) found in genes linked to EMT (e.g. *PBRM1*^{37,38} and *CD46*³⁹) or associated with EMT-related

processes such as cell migration and invasion (e.g. *ABII*⁴⁰, *CAST*⁴¹ and *MPRIIP*⁴²) (Supplementary Table 1, column 5 and Supplementary Table 2, column 5). This suggests that ASEs making up the CES may influence basic cellular functions required for tumor survival, while those occurring in CSS modulate their mesenchymal characteristics. Indeed, three out of the five CSS ASEs with known EMT-related functions (*ABII*⁴⁰, *PBRMI*^{37,38} and *MPRIIP*⁴²) exhibited switch-like splicing shifts (splicing pattern differing by more than 50 percentage points), when normal epithelial and stromal tissues were compared, while only one ASE (*TJPI*⁴³) from the CES was spliced in a tissue-specific manner (Supplementary Table 1, column 2 and Supplementary Table 2, column 2). Therefore, at least in the dissected cell population, the tumor microenvironment harbored more changes in the splicing of EMT-related genes than the tumor itself (Supplementary Table 1, column 5 and Supplementary Table 2, column 5). These changes in the tumor microenvironment promoted the expression of epithelial splicing isoforms through the exclusion of five mesenchymal-enriched exons (exons that are predominantly included in mesenchymal cells compared to epithelial cells) and the inclusion of two epithelial-enriched exons (exons that are predominantly included in epithelial cells compared to mesenchymal cells) (Fig. 2c, Supplementary Table 2, column 2). This indicates that the epithelial-like splicing pattern in the tumor microenvironment is mostly due to the repression of mesenchymal-enriched exons (Supplementary Fig. 2). This directional splicing shift is specifically associated with the CSS since ASEs linked to cancer epithelium exhibited either the mesenchymal splicing pattern (through epithelial-enriched exon inclusion) or simply had non-preferential splicing pattern (Fig. 2c, Supplementary Table 1 column 2). We conclude that alternative splicing in the tumor microenvironment is modified to favor the suppression of a subset of mesenchymal-enriched exons.

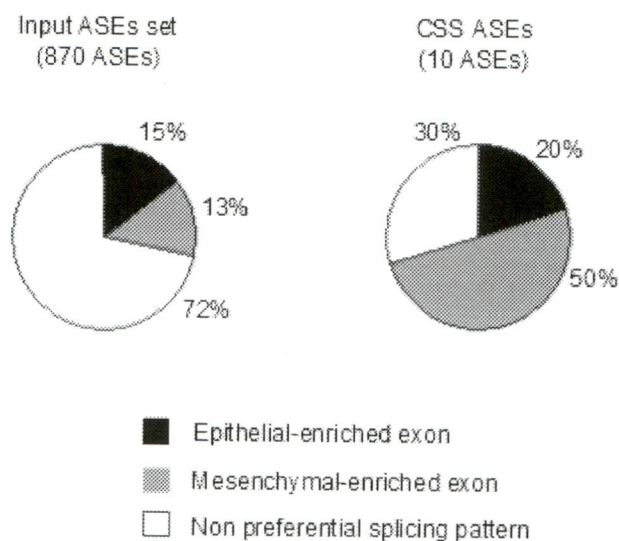


Figure S2 Mesenchymal-enriched exons are preferentially regulated in the tumor microenvironment.

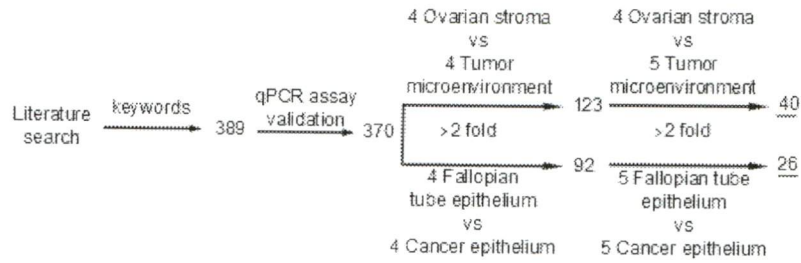
The number of epithelial- and mesenchymal-enriched exons in the discovery screen input ASEs set (IAS) and the cancer stromal signature (CSS) ASEs (see Fig. 2) are shown in the form of pie chart. Only 13% of the exons in the IAS were enriched in mesenchymal cells, while 50% of the CSS ASEs were mesenchymal-enriched exons ($P = 0.004$, Fischer's exact test).

Identification of cancer-specific splicing factors

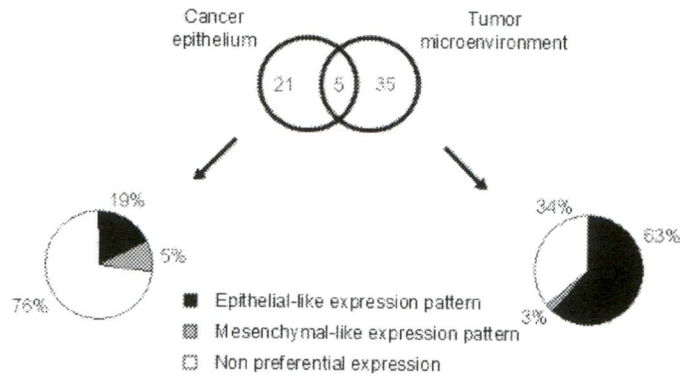
To understand how the ASEs of the CES and CSS are regulated, we examined the expression of 370 putative splicing factors in epithelial and stromal cells isolated from normal and cancer tissues using quantitative RT-PCR (raw data are accessible through <http://palace.lgfus.ca/pcrreactiongroup/list/219>). <http://palace.lgfus.ca/pcrreactiongroup/list/227>). Comparison of the RNA levels in normal and cancer tissues identified 26 and 40 splicing factors whose expression was changed in the cancer epithelium and the tumor microenvironment, respectively (Fig. 3a, Supplementary Table 4, columns 1-3). The overlap between the tumor epithelium and its microenvironment was small (8%) (Fig. 3b), as would be expected from tissues with distinct profiles of alternative splicing (Fig. 2c). The relatively small numbers of splicing factors associated with the tumor (7%) and its microenvironment (11%) indicates that the deregulated expression of splicing factors is not global (Fig. 3a). The epithelial and or mesenchymal relationship of the cancer-associated

splicing factors was established by comparing their expression in normal dissected epithelium and stroma and determining the direction of their expression levels in cancer (raw data are accessible through <http://palace.lgfus.ca/pcrreactiongroup/list/227>). As was the case with CSS ASEs (Fig. 2c), the cancer-associated splicing factors also displayed epithelial-like expression pattern in the tumor microenvironment, while no directional expression pattern was observed in the cancer epithelium (Fig. 3b, Supplementary Table 4 column 4). Therefore, ovarian cancer appears to induce epithelial-like expression of a small subset of splicing factors and ASEs in adjacent stromal cells. This focused reprogramming of alternative splicing appears specific to the tumor microenvironment as it occurs to a much lower extent in the tumor itself. Together these observations suggest that cancer alters the expression of splicing factors in the microenvironment to reprogram the splicing of a specific group of mesenchymal-enriched exons.

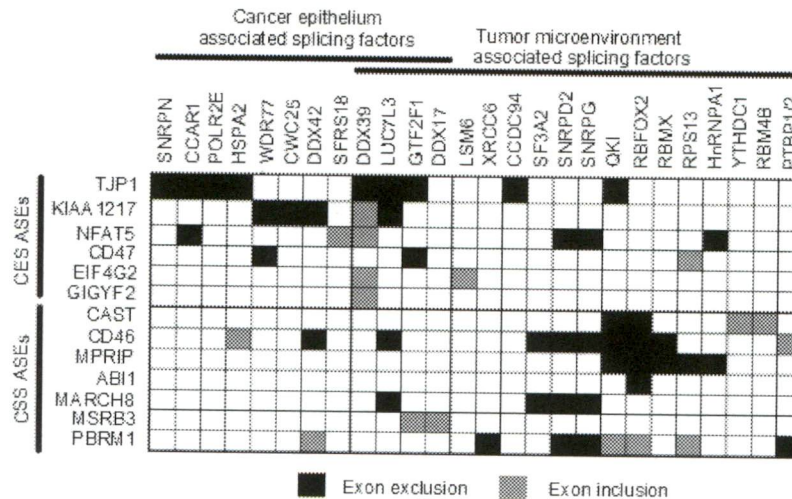
a



b



c



d

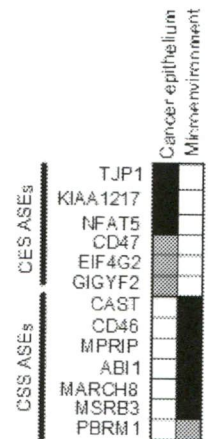


Figure 3 Brosseau 2012

Figure 3 The expression of cancer-associated splicing isoforms is regulated by a small group of splicing factors.

(a) Strategy for the identification of cancer-associated splicing factors. The expression of all splicing factors identified in both the RefSeq database and the NCBI PubMed database⁶⁴ were evaluated by quantitative RT-PCR in the different normal and cancer tissues described in Fig. 1c. Splicing factors capable of discriminating between cancer and normal tissues by at least two fold are underlined. (b) The tumor microenvironment is associated with epithelial-enriched splicing factors. The Venn diagram illustrates the tissue distribution of the cancer-associated splicing factors obtained in a. The pie chart indicates the epithelial / mesenchymal expression of the cancer-associated splicing factors as identified by quantitative RT-PCR analysis of RNA extracted from Fallopian tube epithelium and ovarian stroma (see Supplementary Table 4 column 4 and Materials and Methods section for values and details, respectively). (c) Depletion of the tumor microenvironment associated splicing factors alters the splicing pattern of the cancer stromal signature (CSS). The cancer-associated splicing factors were depleted using siRNA in the ovarian cancer cell line SKOV3ip1. The impact on the CES and CSS ASEs identified in Fig. 2, was evaluated using quantitative RT-PCR. Exons inclusion (gray boxes) or exclusion (black boxes) generating a quantitative splicing shift¹⁷ ($\Delta Q\psi = Q\psi_{\text{KNOCKDOWN}} - Q\psi_{\text{CONTROL}}$) of at least 10 was considered significant and presented in the form of a table. For simplicity, only the ASEs regulated by at least one splicing factor and the splicing factors regulating at least one ASE are shown. The expression of all splicing factors except DDX39 was downregulated in cancer tissues and therefore the illustrated *in vitro* knockdown of these factors are expected to induce a splicing pattern similar to that detected in cancer tissues. (d) Illustration of the CES and CSS ASEs as exon exclusion and inclusion. To enable direct comparison between the splicing shift ($\Delta Q\psi$)¹⁷ produced by the splicing factors knockdown (illustrated in c) and that detected in tissues (Fig. 2), the cancer-associated genes were listed in the same order used in c and their expression in cancer tissues is indicated as exon exclusion (black boxes) and inclusion (gray boxes) (Supplementary Table 1 column 1 and Supplementary Table 2 column 1).

All mesenchymal-enriched exons associated with CSS, are repressed in the tumor microenvironment (Fig. 2c) and most of the cancer-associated splicing factors presented in figure 3c, with the exception of DDX39, are downregulated in cancer. Therefore, silencing of splicing factors in cancer cell lines is expected to induce the exclusion of all cancer-associated exons that are directly regulated by these splicing factors. Accordingly, we depleted 47 out of 61 cancer-associated splicing factors by at least two fold using two independent siRNAs in the model ovarian cancer cell line SKOV3ip1 (Supplementary Table 5 and raw data are accessible through <http://palace.lgfus.ca/pcrreactiongroup/list/23219>) and tested the impact on the splicing pattern of both CES and CSS ASEs. As shown in Fig. 3c, splicing factors linked to cancer epithelium affected mostly CES events while those repressed in the tumor microenvironment led to the exclusion of mesenchymal-enriched exons (Fig. 3c). In 71% (22 out of 31) of the cases, the depletion of splicing factors in SKOV3ip1 altered alternative splicing in the same direction as that observed in cancer tissues (Fig. 3d). These

results demonstrate that the splicing pattern detected in the tumor microenvironment results at least in part by changes in the expression of splicing factors.

Examining the effects of knocking down cancer-associated splicing factors on CES and CSS ASEs revealed two groups of splicing factors. The first group includes core spliceosomal factors such as SF3A2⁴⁴, SNRPD2⁴⁵ and SNRPG⁴⁶, while the second includes alternative splicing regulators RBFOX2⁴⁷ and QKI⁴⁸. The knockdown of the first group, which are underexpressed in the tumor microenvironment modified the splicing of 2 out of 8 CSS ASEs expressed in the SKOV3ip1 cell line (compare Fig. 3c and Fig. 3d), but they also induced apoptosis (data not shown), as would be expected from the knockdown of essential splicing factors. Induction of apoptosis and general perturbation of splicing patterns makes the distinction between direct and indirect effects very difficult and as such this group of splicing factors were not pursued further. In contrast, reproducing the tumor microenvironment associated downregulation of the RBFOX2 or QKI in SKOV3ip1 cells replicated the cancer-associated splicing profile of 4 out of 8 CSS ASEs without dramatically affecting cell viability (compare Figs. 3c and Fig. 3d and data not shown). This indicates that a part of the splicing profiles observed in the tumor microenvironment is reproduced by the inhibition of RBFOX2 and QKI expression.

The tumor microenvironment exhibits limited changes in the mesenchymal-to-epithelial (MET) associated splicing program

The expression profile of established stromal markers demonstrated clearly the stromal nature of the tumor microenvironment (Fig. 1c) indicating that the majority of cells in the tumor microenvironment maintain their mesenchymal phenotype. In contrast, the expression of most CSS ASEs (Fig. 2c) and cancer-associated splicing factors (Fig. 3b) displayed epithelial-like splicing patterns and expression, respectively. These observations suggest either that the tumor induces the MET splicing program without affecting the global expression levels of known markers (Fig. 1c) or that it modifies a subclass of genes associated with MET without generating the canonical epithelial phenotype. All MET and EMT splicing factors were not preferentially expressed in the cancer epithelium or the tumor microenvironment when compared to normal tissues except for ESRP1/2, which was overexpressed in the tumor microenvironment (Supplementary Table 6). This indicates that, overall, the expression of the EMT and MET splicing regulators is not perturbed in cancer tissues. In addition, only 4 out of 17 ASEs previously associated with EMT²⁵ (*PBRM1*, *ENAH*, *BAIAP2* and *CLSTN1*) were linked to the tumor microenvironment (Fig. 4, Supplementary Table 2 and Supplementary Table 7). However, for simplicity, they were

not incorporated in the CSS if they were not identified in the original tumor microenvironment association screen (Fig. 2a). This further confirms that cancer displays limited changes in the splicing of a subset of mesenchymal-enriched exons and no global modification of the MET splicing program.

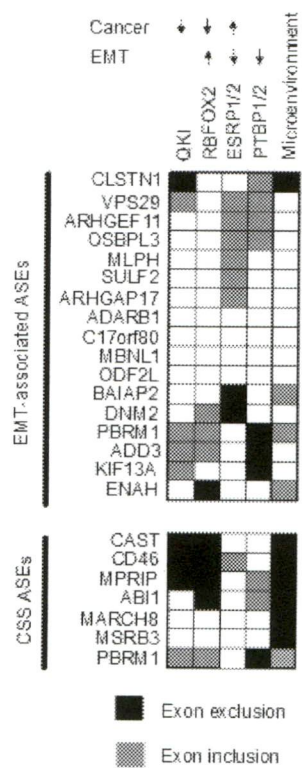


Figure 4 Brosseau 2012

Figure 4 The expression of splicing factors associated with mesenchymal- to- epithelial transition (MET) are not systematically modified in cancer tissues.

The association between the EMT²⁵ (e.g. ESRP1/2 and PTBP1/2) or the tumor microenvironment (e.g. QKI and RBFOX2) associated splicing factors and the splicing of ASEs linked to EMT (upper panel) or tumor microenvironment (lower panel) was evaluated by knocking down these splicing factors and calculating the mean quantitative splicing shift¹⁷ ($\Delta Q\psi = Q\psi_{\text{KNOCKDOWN}} - Q\psi_{\text{CONTROL}}$) of each ASE as described in Fig. 3c. The knockdowns were performed in the ovarian cancer cell line SKOV3ip1 with the exception of ESRP1/2, which was knocked down in the epithelial cell line ZR-75 because it is not expressed in SKOV3ip1. The impact on splicing patterns was evaluated by quantitative RT-PCR as described in Fig. 3c. Exons inclusion (gray boxes) or exclusion (black boxes) generating a quantitative splicing shift of at least 10 was considered significant (see Material and Methods for details) and presented in the form of a table. The expression of the different splicing factors in the tumor microenvironment (cancer) and *in vitro* twist-induced EMT in epithelial cells (EMT)²⁵ is shown on top, where the upward and downward arrows indicate an increase and decrease in expression, respectively.

To determine the overlap between the tumor microenvironment and the EMT / MET splicing programs, we compared the effect of knocking down the MET splicing factors ESRP1/2 and the EMT splicing factor PTBP1 (Supplementary Table 5) to that generated by the knockdown of the cancer-associated factors RBFOX2 and QKI. Comparison of the splicing profiles of the different knockdowns indicated that 71% (12 out 17) of the EMT associated ASEs tested are regulated by either PTBP1 or ESRP1/2, but that only two were also regulated by the tumor microenvironment-associated splicing factors RBFOX2 and QKI (Fig. 4, upper panel). Consistently, a low number of CSS ASEs were affected by the EMT or MET-specific splicing factors in a way that mimics the expression pattern in the tumor microenvironment (Fig. 4, lower panel). We conclude that the splicing pattern of CSS ASEs is not directly linked to the expression of EMT or MET associated splicing factors.

The tumor microenvironment-associated splicing factors QKI and RBFOX2 regulate common splicing targets

To determine the extent of overlap between RBFOX2 and QKI splicing targets and uncover potentially new cancer-associated splicing events, we compared the effects of depleting the expression of these two proteins on a larger set of ASEs. We monitored the splicing

profiles of 48 pre-established RBFOX2-regulated ASEs¹² and 57 newly identified QKI-regulated ASEs (see Material and Methods for details) after QKI or RBFOX2 knockdown in SKOV3ip1 (Supplementary Table 8, columns 3-5). As indicated in Fig. 5a, 66% (37 out of 56) of ASEs modified by RBFOX2 knockdown were also affected by the depletion of QKI, and half (37 out of 74) of the QKI regulated ASEs were perturbed by RBFOX2 knockdown. The effect of RBFOX2 and QKI knockdown on the common group of ASEs (Fig. 5a) displayed a Pearson correlation of 0.65 ($P = 1.5e-05$), confirming their status as shared splicing targets (Fig. 5b). This indicates that there is a broad overlap between QKI and RBFOX2 splicing targets. To determine the functional significance of the common RBFOX2 and QKI splicing targets, we classified them according to their epithelial / mesenchymal preferential splicing pattern and calculated the number of exon inclusions and exclusions in both mesenchymal and epithelial groups (Fig. 5c, Supplementary Table 8, column 7 and Supplementary Table 9). The results indicate that RBFOX2 promotes inclusion of QKI-independent exons, while QKI favors the exclusion of exons that are not controlled by RBFOX2, irrespective of their relative epithelial or mesenchymal-enrichment. In contrast, most exons regulated by both RBFOX2 and QKI were mesenchymal-enriched. In the tumor microenvironment the expression of QKI and RBFOX2 is downregulated (Supplementary Table 4) and their expression level is correlated with the changes in alternative splicing. Therefore, while RBFOX2 and QKI may individually affect different splicing targets, collectively they converge to favor the inclusion of a subset of mesenchymal-enriched exons, which explains why the downregulation of these splicing factors leads to exon exclusion in the tumor microenvironment.

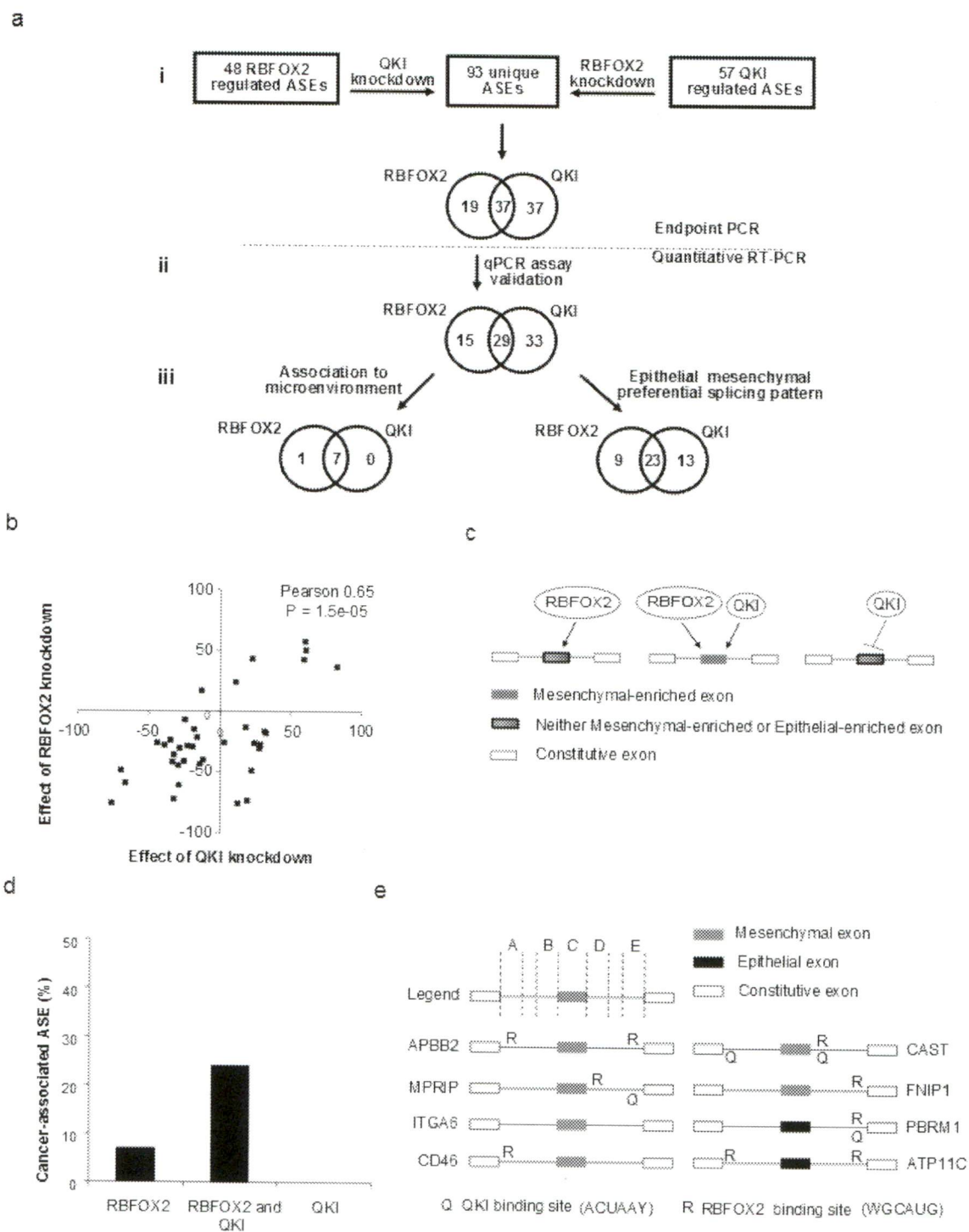


Figure 5 Brosseau 2012

Figure 5 RBFOX2 and QKI regulate the expression of common epithelial splicing isoforms in the tumor microenvironment.

(a) Identification of common RBFOX2 and QKI splicing targets. RBFOX2 and QKI were individually knocked down using two independent siRNAs in SKOV3ip1. The effect of the different knockdowns was evaluated by endpoint PCR using a set of 48 pre-established RBFOX2 targets and a set of 57 newly identified QKI targets (see Material and Methods for details). The 93 ASEs responding to the knockdown of at least one splicing factor are presented in the form of Venn diagram (i). Quantitative RT-PCR primers were designed and validated for 77 out of 93 ASEs (ii) and their expression in the tumor microenvironment or the epithelial and stromal normal tissues was tested as described in Fig. 2 (iii). (b) Comparison of the impact of RBFOX2 and QKI knockdown on the splicing of 37 common ASEs. The impact of RBFOX2 and QKI knockdown on the common set of 37 splicing targets discovered in a, was plotted as a quantitative splicing shift¹⁷ ($\Delta Q\psi = Q\psi_{\text{KNOCKDOWN}} - Q\psi_{\text{CONTROL}}$) to generate a scatter graph. The Pearson correlation between the effect of RBFOX2 and QKI knockdown on splicing pattern and its p value is indicated on the top right of the graph. (c) Schematic representation of the effects of RBFOX2 and QKI on alternative splicing. In normal tissues, RBFOX2 favors exon inclusion, while QKI favors exon exclusion regardless of the epithelial / mesenchymal splicing patterns of the targeted alternative splicing events. In contrast, the knockdown of RBFOX2 and QKI mostly lead to the inclusion of mesenchymal-enriched exons in cases where the ASEs are affected by both splicing factors (i.e. common ASEs). The exons and introns are presented in the form of boxes and lines, respectively. Arrows indicate exon inclusion and bar-headed lines indicate exon exclusion. Gray boxes indicate mesenchymal exons and the boxes with gray and black gradient indicate an exon with non-preferential epithelial or mesenchymal splicing pattern. (d) Common RBFOX2 and QKI targets are more likely to be associated with the tumor microenvironment than those affected by only one splicing factor. The bar graphs represent the percentage of the RBFOX2, QKI or common RBFOX2 and QKI splicing targets (identified in a) associated with the tumor microenvironment as described in Fig. 2a. (e) Schematic representation of the protein binding sites near RBFOX2- and QKI-responsive exons. The position of RBFOX2 and QKI binding sites (WGCAUG and ACUAAY) in 8 common RBFOX2 and QKI splicing targets associated with the tumor microenvironment (identified in Fig. 3c and Fig. 5d) are indicated as “R” and “Q”, respectively. The binding sites were found in five regions: 250 nts from the 5’ splice site of the upstream intron (A); 250 nts from the 3’ splice site of the upstream intron (B); within the exon (C); 250 nts from the 5’ splice site of the downstream intron (D); or 250 nts from the 3’ splice site of the downstream intron (E).

If RBFOX2 and QKI regulate splicing in the tumor microenvironment, we expect at least some of the alternative splicing events affected by the knockdown of these proteins *in vitro* to be associated with cancer *in vivo*. Accordingly, we monitored the splicing profiles of RBFOX2 and QKI-dependent ASEs in the tumor microenvironment. Out of 29 ASEs tested, seven ASEs including three that overlap with Fig. 2a (Supplementary Table 2, columns 1-2) were associated to the tumor microenvironment (Supplementary Table 10,

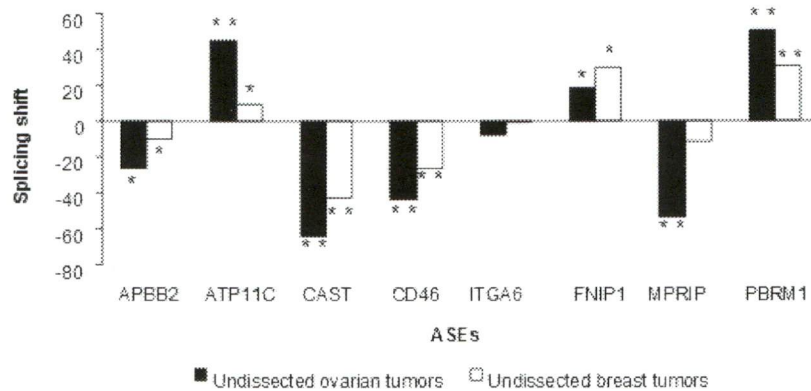
columns 1-2). A total of five were mesenchymal-enriched exons excluded in the tumor microenvironment (Supplementary Table 2, column 2 and Supplementary Table 10, column 2). Indeed, these four additional ASEs (APBB2, ATP11C, ITGA6, FNIP1) could be considered potential markers for ovarian cancer. However, for simplicity they were not incorporated in the CSS, which refer only to ASEs identified in the original association screen using dissected tissues (Fig. 2a). Overall, 24% (7 out of 29) of the RBFOX2 and QKI-dependent ASEs were associated with the tumor microenvironment, whereas less than 7% of the ASEs affected by only one of the two splicing factors were linked to cancer (Fig. 5d). This indicates that while a substantial fraction of ASEs under the control of RBFOX2 and QKI are associated with the tumor microenvironment a large number of RBFOX2 and QKI *in vitro* splicing targets are not. Together these results suggest that while RBFOX2 and QKI are important regulators of cancer-associated genes, other factors influence their capacity to modulate the splicing of their target genes in the tumor microenvironment. Examining the sequence surrounding the RBFOX2 and QKI-dependent ASEs revealed consensus binding motifs^{49,50} for at least one of the two splicing factors, in all cases except ITGA6 (Fig. 5e). This suggests that RBFOX2 and QKI directly contribute to the splicing decisions of this specific subset of mesenchymal-enriched exons.

Repression of mesenchymal-enriched exons is not an exclusive feature of ovarian cancer

It was previously suggested that RBFOX2 might be inactivated in breast cancer through the skipping of a 40 nucleotides C-terminal cassette exon (exon 11)^{12,51}. Therefore, at least some of RBFOX2 splicing targets, including those associated with the ovarian tumor microenvironment, are likely to be similarly modified in breast cancer. To examine this possibility we compared the splicing pattern of the set of RBFOX2 and QKI-dependent ASEs associated with the tumor microenvironment in undissected ovarian and breast tumors using quantitative RT-PCR. As expected, the splicing patterns in both ovarian and breast cancers were tightly correlated, indicating that RBFOX2 and QKI splicing targets are also repressed in high-grade breast cancer (Fig. 6a, Supplementary Table 3), and suggesting that breast and ovarian cancer use similar mechanisms for modifying the splicing patterns of tumor associated ASEs. Consistently, an independent study using a large number of

breast cancer tissues concluded that QKI expression is reduced in breast tumors ($P < 3.0e-03$, t-test)⁴⁸. In our hands, however, the most striking difference between tumor and normal breast tissues was not the change in global expression but rather in the pattern of QKI splicing (Fig. 6b, Supplementary Table 11 and Supplementary Table 12). In the tumor, QKI-6 isoform was preferentially expressed over the QKI-7 isoform. This change was previously shown to affect QKI activity⁵². Therefore, both changes in the expression and splicing of QKI are altered depending on the type of cancer or cancer-subtype examined. Together the data suggest that RBFOX2 and QKI play an important role in regulating the splicing of a subset of mesenchymal-enriched exons in both ovarian and breast cancer.

a



b

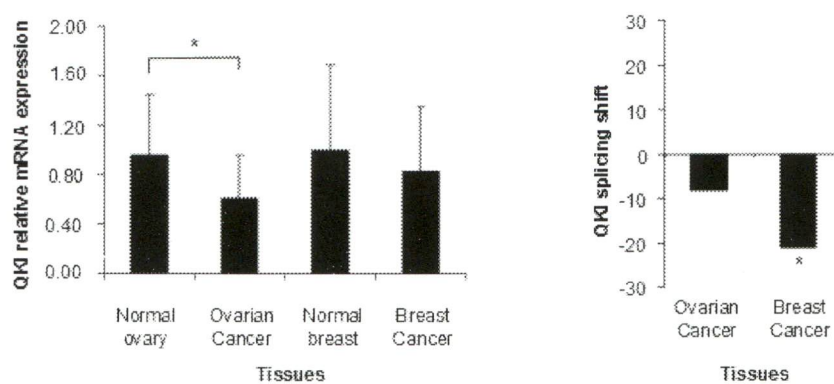


Figure 6 Brosseau 2012

Figure 6 The common RBFOX2 and QKI splicing targets are deregulated in both ovarian and breast cancer.

(a) The behaviour of 8 common RBFOX2 and QKI splicing targets associated to the tumor microenvironment was monitored using quantitative RT-PCR in 14 normal and 13 serous high-grade ovarian cancer tissues and their quantitative splicing pattern ($Q\psi$)¹⁷ compared to that detected in 18 normal and 20 ductal breast cancer tissues (Supplementary Table 3). Shown are histograms representing the mean quantitative splicing shift¹⁷ ($\Delta Q\psi = Q\psi_{TUMOR} - Q\psi_{NORMAL}$). (b) QKI expression is altered by different mechanisms in ovarian and breast cancer. The expression (left panel) and splicing pattern (right panel) of QKI was monitored by quantitative RT-PCR in breast and ovarian samples as described in a. The global expression pattern was calculated relative to housekeeping genes as previously described¹⁷ and the relative value presented in the form of a bar graph. The splicing pattern of QKI isoforms 6 (short) and 7 (long) was calculated as mean quantitative splicing shift¹⁷ ($\Delta Q\psi = Q\psi_{TUMOR} - Q\psi_{NORMAL}$) as previously described in a and plotted in the form of bar graphs. When significant, the p-value (t-test) of difference in expression or splicing shift is displayed on top of histogram by asterisks (* means $P < 0.05$; ** means $P < 5.0e-07$).

Discussion

In this study we compared the splicing program of a homogenous population of normal and cancer cells with the goal of identifying cancer-specific splicing isoforms. The results indicate that differences between normal and cancer tissues are much smaller than previously thought. Analyzing the expression of the Ref. Seq alternative splicing events in RNA extracted from whole ovarian tumors identified 336 cancer-associated ASEs^{11,12}, while comparison of normal and cancer epithelial or stromal tissues identified a total of 25 cancer-specific ASEs (Fig. 2c, Supplementary Table 7 column 1 and Supplementary Table 10 column 1). Most of the 336 whole tumor markers affected cell-type specific genes including those linked to cell architecture (*MPRIP*⁴²), cell plasticity (*MAP3K7*) and cell movement (*CEACAM1*)¹². In contrast, those identified using dissected tissues were mostly linked to tumorigenesis or membrane (signaling) associated functions (Fig. 2, Supplementary Table 1 column 5 and Supplementary Table 2 column 5). Therefore, while whole tumor screens identified cell-type specific genes, comparison of samples with similar cell composition produced gene signatures that are more directly associated with tumor biology. Consistently, the overlap between the two types of screens was limited to 13 ASEs commonly identified as cancer-specific markers by the two approaches (Supplementary

Fig. 3). Strikingly, all the common ASEs except one (*OSBPL9*) were identified as tumor microenvironment markers, suggesting that most of ASEs identified in whole tumor screens are epithelial markers that could easily be eliminated as false positives when samples with similar epithelial contents are compared. In any case, it is clear that comparison of dissected tissue of similar composition is much more likely to reveal true cancer-specific markers.

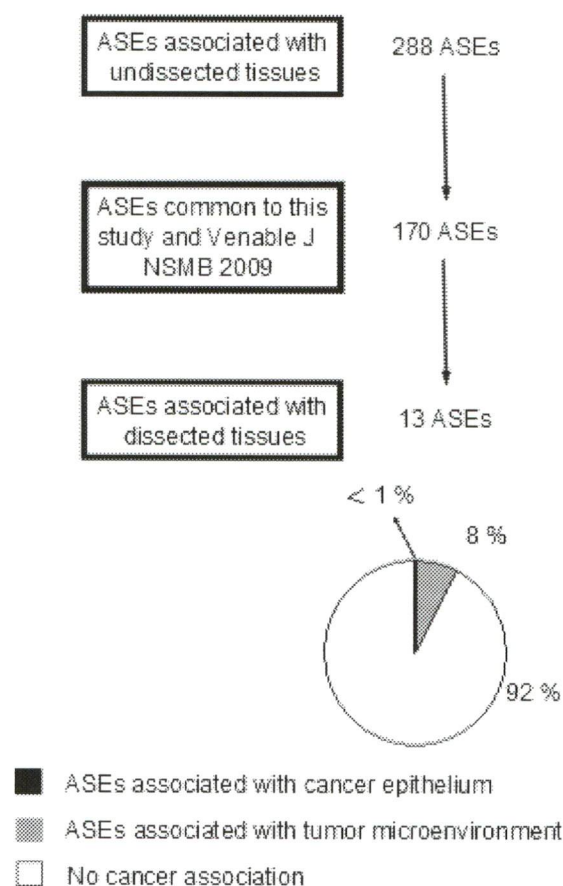


Figure S3 Monitoring the splicing pattern of ASEs associated with whole tumors in dissected tissues.

Out of 288 ASEs that were previously associated with whole ovarian tumors ², 170 were included in the input set used in the dissected tissue screen described in Fig. 2. The pie chart illustrates the number of whole tumor associated events that were preferentially expressed in dissected cancer tissues. The majority of ASEs identified in whole tumor did not distinguish between dissected normal and cancer tissues.

Contrary to *in vitro* assumptions^{24,25}, we found no evidence of strong mesenchymal splicing patterns in high-grade ovarian cancer epithelium. The expression of well-established epithelial expression markers like CDH1 and DSP (Fig. 1c) and the histological inspection of ovarian tumors (<http://www.proteinatlas.org/>) confirmed the epithelial nature of cancer cells. In addition, no mesenchymal-enriched exons were detected in the tumor cancer epithelium (Fig. 2c) and none of the splicing factors linked to EMT were differentially expressed in tumor tissues (Supplementary Table 6, column 2). These data suggest that if indeed EMT was active in the samples tested, the fraction of attached epithelial cells acquiring mesenchymal character is relatively small. This conclusion is also supported by earlier studies that failed to identify a strong EMT signature in dissected ovarian cancer tissues²⁷. Therefore, while cancer may induce changes in the splicing program of certain genes in the direction of EMT *in vitro*, these changes may be restricted to a very small population of cancer cells in tumor tissues.

The changes in splicing in the tumor microenvironment were mostly repression of mesenchymal-enriched exons (Fig. 2c). However, the majority of the established mesenchymal markers were expressed normally suggesting that the observed changes in splicing are not due to MET (Fig. 1c). In addition, the expression of most EMT and MET associated ASEs (Fig. 4) and splicing factors (Supplementary Table 6 column 3) were not altered in the tumor microenvironment, suggesting that the expression of EMT and MET-dependent exons is not globally deregulated in cancer. Consistently, the splicing factors associated with the tumor microenvironment (e.g. QKI) were not necessarily linked to EMT²⁵. Therefore, it appears that cancer suppresses the mesenchymal characters of certain exons for reasons other than the induction of MET. Indeed, overexpression of QKI blocks proliferation *in vitro* and acts as a potential tumor suppressor in colon cancer⁵³ and astrocytic tumors⁵⁴.

Reducing the expression of splicing factors in any given cell line *in vitro* lead to changes in large number of ASEs, few of which can be associated with a given condition *in vivo*. In our hands, repression of QKI or RBFOX2 *in vitro* changed the splicing pattern of a total of 39 common splicing targets in the SKOV3ip1 ovarian adenocarcinoma cell line (Fig. 3c,

Fig. 5a and Supplementary Table 8, columns 3-5). However, only a small subset of these were associated with the tumor microenvironment *in vivo* (Fig. 2a and Fig. 5d). This may reflect differences between cell lines grown *in vitro* and patient tissues. The splicing events affected by QKI and RBFOX2 knockdowns *in vitro* might be affected by other cancer-associated splicing factors not present in the cell lines tested. Indeed, the outcome of any splicing target is determined by many splicing factors that may vary depending on the cell line and conditions tested. For examples, RBFOX2 and RBFOX1 splicing activities were found to be affected by hnRNP H1 and TFG in brain⁴⁷. On the other hands, RBFOX2 and the neuron-specific splicing factor NOVA were found to regulate common splicing targets in neuronal tissues⁵⁵. It was also shown that the splicing factor QKI controls the expression of other splicing factors like hnRNPA1, suggesting that one splicing factor may lead to a cascade of changes in splicing that goes beyond its direct splicing targets⁵⁶. In the case of ovarian cancer, the effect of QKI on splicing appears not to be mediated by hnRNPA1 as the expression level of hnRNPA1 is associated with the tumor microenvironment (Supplementary Table 4, column 3) but its depletion affects the splicing of only one (MPRIP) common RBFOX2 and QKI splicing target in SKOV3ip1 cell line (Fig. 3c). Therefore, while the data presented in this work points to a group of splicing targets that are specifically regulated by QKI and RBFOX2 in the tumor microenvironment, it is likely that the generation of the cancer splicing program involves a large number of factors that together define the splicing outcome in the tumor microenvironment.

The cancer stromal signature (CSS) discovered in ovarian cancer was also detected in breast cancer, indicating similarity in the splicing patterns of these types of cancer. The CSS was tested in a variety of breast cancer types including high-grade estrogen receptor positive and negative ductal breast cancer samples (Fig. 6). Overall, the splicing profile of these ASEs was similar to that detected in the microenvironment of ovarian tumors regardless of the estrogen receptor status or cancer subtype, underlining the generality of the newly discovered signature. Earlier study of luminal (estrogen receptor positive) and basal (estrogen receptor negative) breast cancer cell lines identified cancer subtype specific splicing signatures^{25,57}. In contrast, we did not find any statistically significant differences in the splicing patterns of tumor microenvironment and RBFOX2 and QKI dependent

ASEs in estrogen receptor positive and negative breast cancer subtypes (Supplementary Table 3). This contradiction might be due to differences between the minimalistic, *in vitro* growth conditions of monolayer cell cultures and the intricate, three-dimensional complex growth of cells in patient tissues *in vivo* or the number and nature of splicing events examined. The expression pattern of splicing factors was also perturbed in both ovarian and breast cancer but in different ways: in ovarian cancer the expression of RBFOX2 and QKI was downregulated, while their splicing patterns were altered in breast cancer (Fig. 5)¹². Other studies of QKI expression in breast cancer suggested that QKI might also be downregulated in breast cancer⁵⁸, suggesting that the expression of QKI varies between breast cancer tumors. Therefore, while there is similarity between the breast and ovarian cancer splicing patterns, there is difference in the way these patterns are produced in each type of cancer.

Oncogenesis and tumor progression rely on a reciprocal secretory communication program with the tumor microenvironment⁵⁹. This communication results in a tumor microenvironment replete with inflammatory mediators, growth factors, matrix remodeling enzymes and angiogenic factors⁶⁰. Growth factors produced by many kinds of cancer cells including breast and ovarian carcinomas attract fibroblasts and stimulate their proliferation⁶¹. Differentiation of fibroblasts in the vicinity of cancer cells to myofibroblasts (i.e. activated fibroblasts) is important for tumor growth and has been observed in several *in vitro* studies using ovarian cells⁶². This transformation is required for tumor development since fibroblasts derived from the omentum, the richly vascularized fatty subperitoneal layer surrounding the ovaries, increased ovarian cancer cell adhesion and invasive behavior, whereas omentum-derived mesothelial cells reduced growth of ovarian cancer cells⁶³. In this study, we uncovered a new layer of tumor-dependent modification of stromal cells defined by changes in alternative splicing. These changes occur preferentially in the tumor microenvironment and include differences in the splicing of genes related to diverse functions linked to cancer development and were associated with changes in the expression of splicing factors. Interestingly, the cancer-specific splicing markers that are least sensitive to tissue composition are found in the tumor microenvironment and not the tumor itself, suggesting that cancer can be detected more by changes in the

microenvironment that may even take place early in tumor development. However, it is currently unclear whether these changes in stromal cells precede tumor development or occur after the tumor is established. *In vitro* studies using cell and animal models will be required to clarify this issue. In any case, the results presented here pave the way for in depth studies of a new level of tumor dependent modification of gene expression in stromal cells. All together, it suggests that cancer cells modulate their environment in a specific and programmed fashion and that the monitoring of splicing provides sensitive and specific markers that may prove to have clinical values.

Acknowledgments

We thank the “Réseau de Recherche sur le Cancer FRSQ” tissue bank for ovarian and breast normal and tumor tissues, David Huntsman and the British Columbia tumor bank for the Fallopian tube samples, the Alberta tumor bank for the low malignant potential sample and Vincent Normandeau-Babin for clinical data management. We also thank David Huntsman for discussions and advice in the early stages of this work. We thank Fernand-Pierre Gendron, the “Service de phénotypage” and members of the Department of pathology from the Université de Sherbrooke for help with LCM and tissue preparation. This work was funded by a grant from the Canadian Cancer Society Research Institute and the Canadian Institute of Health Research (CIHR). Jean-Pierre Perreault is the Canada Research Chair on Genomics and Catalytic RNA. Benoit Chabot is the Canada Research Chair in Functional Genomics. Sherif Abou-Elela is a Chercheur National of the Fonds de la Recherche en Santé du Québec (FRSQ).

Authors contribution

Designed experiments: JPB, SAE, JFL, HN

Analyzed data: JPB, JFL, PT, SAE

Paper writing: JPB, SAE, BC

Supervised experiments: RK, PP, JPP, BC, SAE, RK

Carried out experiments: JPB, MD, DG, SC, EL, HN

Competing interests

We declare no competing interests.

References

1. Coleman, M.P. et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* **377**, 127-38 (2011).
2. Auersperg, N., Wong, A.S., Choi, K.C., Kang, S.K. & Leung, P.C. Ovarian surface epithelium: biology, endocrinology, and pathology. *Endocr Rev* **22**, 255-88 (2001).
3. Saad, A.F., Hu, W. & Sood, A.K. Microenvironment and pathogenesis of epithelial ovarian cancer. *Horm Cancer* **1**, 277-90 (2010).
4. Mandai, M. et al. Ovarian clear cell carcinoma as a stress-responsive cancer: influence of the microenvironment on the carcinogenesis and cancer phenotype. *Cancer Lett* **310**, 129-33 (2011).
5. Agarwal, A. et al. Identification of a metalloprotease-chemokine signaling system in the ovarian cancer microenvironment: implications for antiangiogenic therapy. *Cancer Res* **70**, 5880-90 (2010).
6. Barbolina, M.V., Burkhalter, R.J. & Stack, M.S. Diverse mechanisms for activation of Wnt signalling in the ovarian tumour microenvironment. *Biochem J* **437**, 1-12 (2011).
7. Yang, N. et al. Transcriptional regulation of PIK3CA oncogene by NF-kappaB in ovarian cancer microenvironment. *PLoS One* **3**, e1758 (2008).
8. Iwatsuki, M. et al. Epithelial-mesenchymal transition in cancer development and its clinical significance. *Cancer Sci* **101**, 293-9 (2010).
9. Casey, T.M. et al. Cancer associated fibroblasts stimulated by transforming growth factor beta1 (TGF-beta 1) increase invasion rate of tumor cells: a population study. *Breast Cancer Res Treat* **110**, 39-49 (2008).
10. Jing, Y., Han, Z., Zhang, S., Liu, Y. & Wei, L. Epithelial-Mesenchymal Transition in tumor microenvironment. *Cell Biosci* **1**, 29 (2011).
11. Klinck, R. et al. Multiple alternative splicing markers for ovarian cancer. *Cancer Res* **68**, 657-63 (2008).
12. Venables, J.P. et al. Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol* **16**, 670-6 (2009).
13. Sotiriou, C. et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* **98**, 262-72 (2006).
14. Munirah, M.A. et al. Identification of different subtypes of breast cancer using tissue microarray. *Romanian journal of morphology and embryology = Revue roumaine de morphologie et embryologie* **52**, 669-77 (2011).
15. Prat, A., Ellis, M.J. & Perou, C.M. Practical implications of gene-expression-based assays for breast oncologists. *Nature reviews. Clinical oncology* **9**, 48-57 (2012).
16. Venables, J.P. et al. Identification of alternative splicing markers for breast cancer. *Cancer Res* **68**, 9525-31 (2008).
17. Prinos, P. et al. Alternative splicing of SYK regulates mitosis and cell survival. *Nat Struct Mol Biol* **18**, 673-9 (2011).

18. Biermann, K. et al. Genome-wide expression profiling reveals new insights into pathogenesis and progression of testicular germ cell tumors. *Cancer Genomics Proteomics* **4**, 359-67 (2007).
19. Zhu, Y., Nilsson, M. & Sundfeldt, K. Phenotypic plasticity of the ovarian surface epithelium: TGF-beta 1 induction of epithelial to mesenchymal transition (EMT) in vitro. *Endocrinology* **151**, 5497-505 (2010).
20. Strauss, R. et al. Analysis of epithelial and mesenchymal markers in ovarian cancer reveals phenotypic heterogeneity and plasticity. *PLoS One* **6**, e16186 (2011).
21. Gallo, D., Ferlini, C. & Scambia, G. The epithelial-mesenchymal transition and the estrogen-signaling in ovarian cancer. *Curr Drug Targets* **11**, 474-81 (2010).
22. Thiery, J.P., Acloque, H., Huang, R.Y. & Nieto, M.A. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871-90 (2009).
23. Venables, J.P. et al. RBFOX2 is an important regulator of mesenchymal-specific splicing in both normal and cancer tissues. *Molecular and cellular biology* (2012).
24. Warzecha, C.C. et al. An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *EMBO J* **29**, 3286-300 (2010).
25. Shapiro, I.M. et al. An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genet* **7**, e1002218 (2011).
26. Rogerson, L. et al. Application of transcript profiling in formalin-fixed paraffin-embedded diagnostic prostate cancer needle biopsies. *BJU Int* **102**, 364-70 (2008).
27. Tone, A.A. et al. Gene expression profiles of luteal phase fallopian tube epithelium from BRCA mutation carriers resemble high-grade serous carcinoma. *Clin Cancer Res* **14**, 4067-78 (2008).
28. Mehra, K. et al. STICS, SCOUTs and p53 signatures; a new language for pelvic serous carcinogenesis. *Front Biosci (Elite Ed)* **3**, 625-34 (2011).
29. Kurman, R.J. & Shih Ie, M. The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. *Am J Surg Pathol* **34**, 433-43 (2010).
30. Nofech-Mozes, S. et al. Immunophenotyping of serous carcinoma of the female genital tract. *Mod Pathol* **21**, 1147-55 (2008).
31. Zhu, Z.G. et al. Germline mutational analysis of CDH1 and pathologic features in familial cancer syndrome with diffuse gastric cancer/breast cancer proband in a Chinese family. *Eur J Surg Oncol* **30**, 531-5 (2004).
32. Mitsui, S. et al. Multiple promoters regulate tissue-specific alternative splicing of the human kallikrein gene, KLK11/hippostasin. *FEBS J* **273**, 3678-86 (2006).
33. Ichinose, H. et al. Tissue-specific alternative splicing of the first exon generates two types of mRNAs in human aromatic L-amino acid decarboxylase. *Biochemistry* **31**, 11546-50 (1992).
34. Collino, F. et al. Epithelial-mesenchymal transition of ovarian tumor cells induces an angiogenic monocyte cell population. *Exp Cell Res* **315**, 2982-94 (2009).
35. Pruitt, K.D., Tatusova, T. & Maglott, D.R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35**, D61-5 (2007).
36. Mateo, V. et al. Mechanisms of CD47-induced caspase-independent cell death in normal and leukemic cells: link between phosphatidylserine exposure and cytoskeleton organization. *Blood* **100**, 2882-90 (2002).

37. Shain, A.H. et al. Convergent structural alterations define SWItch/Sucrose NonFermentable (SWI/SNF) chromatin remodeler as a central tumor suppressive complex in pancreatic cancer. *Proc Natl Acad Sci U S A* **109**, E252-9 (2012).
38. Huang, X., Gao, X., Diaz-Trelles, R., Ruiz-Lozano, P. & Wang, Z. Coronary development is regulated by ATP-dependent SWI/SNF chromatin remodeling component BAF180. *Dev Biol* **319**, 258-66 (2008).
39. Strauss, R. et al. Epithelial phenotype confers resistance of ovarian cancer cells to oncolytic adenoviruses. *Cancer Res* **69**, 5115-25 (2009).
40. Wang, C. et al. Abelson interactor protein-1 positively regulates breast cancer cell proliferation, migration, and invasion. *Mol Cancer Res* **5**, 1031-9 (2007).
41. Leloup, L., Mazeres, G., Daury, L., Cottin, P. & Brustis, J.J. Involvement of calpains in growth factor-mediated migration. *Int J Biochem Cell Biol* **38**, 2049-63 (2006).
42. Ono, R. et al. M-RIP, a novel target of JNK signaling and a requirement for human cancer cell invasion. *Int J Mol Med* **22**, 199-203 (2008).
43. Kiener, T.K., Sleptsova-Friedrich, I. & Hunziker, W. Identification, tissue distribution and developmental expression of tjp1/zo-1, tjp2/zo-2 and tjp3/zo-3 in the zebrafish, *Danio rerio*. *Gene Expr Patterns* **7**, 767-76 (2007).
44. Hong, W. et al. Association of U2 snRNP with the spliceosomal complex E. *Nucleic Acids Res* **25**, 354-61 (1997).
45. Schumperli, D. & Pillai, R.S. The special Sm core structure of the U7 snRNP: far-reaching significance of a small nuclear ribonucleoprotein. *Cell Mol Life Sci* **61**, 2560-70 (2004).
46. Ma, Y., Dostie, J., Dreyfuss, G. & Van Duyne, G.D. The Gemin6-Gemin7 heterodimer from the survival of motor neurons complex has an Sm protein-like structure. *Structure* **13**, 883-92 (2005).
47. Sun, S., Zhang, Z., Fregoso, O. & Krainer, A.R. Mechanisms of activation and repression by the alternative splicing factors RBFOX1/2. *RNA* **18**, 274-83 (2012).
48. Novikov, L. et al. QKI-mediated alternative splicing of the histone variant MacroH2A1 regulates cancer cell proliferation. *Mol Cell Biol* **31**, 4244-55 (2011).
49. Galarneau, A. & Richard, S. Target RNA motif and target mRNAs of the Quaking STAR protein. *Nat Struct Mol Biol* **12**, 691-8 (2005).
50. Auweter, S.D. et al. Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J* **25**, 163-73 (2006).
51. Underwood, J.G., Boutz, P.L., Dougherty, J.D., Stoilov, P. & Black, D.L. Homologues of the *Caenorhabditis elegans* Fox-1 protein are neuronal splicing regulators in mammals. *Mol Cell Biol* **25**, 10005-16 (2005).
52. Pilotte, J., Larocque, D. & Richard, S. Nuclear translocation controlled by alternatively spliced isoforms inactivates the QUAKING apoptotic inducer. *Genes Dev* **15**, 845-58 (2001).
53. Yang, G. et al. RNA-binding protein quaking, a critical regulator of colon epithelial differentiation and a suppressor of colon cancer. *Gastroenterology* **138**, 231-40 e1-5 (2010).
54. Ichimura, K. et al. Small regions of overlapping deletions on 6q26 in human astrocytic tumours identified using chromosome 6 tile path array-CGH. *Oncogene* **25**, 1261-71 (2006).

55. Zhang, C. et al. Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science* **329**, 439-43 (2010).
56. Zhang, G. et al. Ion sensing in the RCK1 domain of BK channels. *Proc Natl Acad Sci U S A* **107**, 18700-5 (2010).
57. Lapuk, A. et al. Exon-level microarray analyses identify alternative splicing programs in breast cancer. *Molecular cancer research : MCR* **8**, 961-74 (2010).
58. Novikov, L. et al. QKI-mediated alternative splicing of the histone variant MacroH2A1 regulates cancer cell proliferation. *Molecular and cellular biology* **31**, 4244-55 (2011).
59. Bhowmick, N.A., Neilson, E.G. & Moses, H.L. Stromal fibroblasts in cancer initiation and progression. *Nature* **432**, 332-7 (2004).
60. Tlsty, T.D. & Coussens, L.M. Tumor stroma and regulation of cancer development. *Annual review of pathology* **1**, 119-50 (2006).
61. Orimo, A. & Weinberg, R.A. Heterogeneity of stromal fibroblasts in tumors. *Cancer biology & therapy* **6**, 618-9 (2007).
62. Schauer, I.G., Sood, A.K., Mok, S. & Liu, J. Cancer-associated fibroblasts and their putative role in potentiating the initiation and development of epithelial ovarian cancer. *Neoplasia* **13**, 393-405 (2011).
63. Kenny, H.A., Krausz, T., Yamada, S.D. & Lengyel, E. Use of a novel 3D culture model to elucidate the role of mesothelial cells, fibroblasts and extra-cellular matrices on adhesion and invasion of ovarian cancer cells to the omentum. *International journal of cancer. Journal international du cancer* **121**, 1463-72 (2007).
64. Lu, Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford)* **2011**, baq036 (2011).
65. Brosseau, J.P. et al. High-throughput quantification of splicing isoforms. *RNA* **16**, 442-9 (2010).
66. Chen, J., Feng, Y. & Zhou, X. [Experimental study on resistance to cisplatin and taxol in human ovarian cancer SKOV3ip1 multicellular aggregates]. *Zhonghua Fu Chan Ke Za Zhi* **36**, 489-92 (2001).
67. Spellman, R., Llorian, M. & Smith, C.W. Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Mol Cell* **27**, 420-34 (2007).

Online Methods

Tissue selection

Clinically homogeneous (high grade, high stage and chemo-naïve) tumors were selected for both dissected and undissected analyses. Tissue sourcing was performed as previously described^{11,16} and separated into 6 sample sets; 1) the active ASEs set containing one Fallopian tube, one low malignant potential, one serous ovarian cancer grade 1, one serous ovarian cancer grade 3 and one ovarian clear cell carcinoma, 2) the LCM discovery set containing 4 normal ovarian stroma, 4 Fallopian tube epithelium, 4 cancer epithelium and 4 tumor microenvironment samples, 3) the LCM validation set containing 4 normal ovarian stroma, 5 Fallopian tube epithelium, 5 cancer epithelium and 5 tumor microenvironment samples, 4) the fresh frozen (FF) and formalin-fixed, paraffin-embedded (FFPE) matched sample-sets containing mirror images of 4 undissected normal ovaries and 4 undissected ovarian serous carcinomas, 5) the undissected serous ovarian samples-set is composed of 14 undissected normal ovaries, 6 Fallopian tube (FT) and 13 undissected ovarian serous carcinomas and 6) the undissected ductal breast samples-set composed of 18 normal breast tissues and 20 ductal breast cancer samples. The samples labelled OVN 250 S and OVN 308 S did not pass the RNA quality control tests and therefore were rejected from further analyses. The percentage of epithelial and stromal cells in each samples were determined using hematoxylin and eosin stained mirror image samples. The description of individual samples is listed in Supplementary Table 13, Supplementary Table 15 and Supplementary Table 16. Dissected regions were selected based on the epithelium (hematoxylin) and stromal cells (eosin) specific staining patterns. Cancer epithelial cells were selected based on their morphology (cylindrical or cubic cells with round to ovoid nuclei) and the absence of necrosis. Cells in the tumor microenvironment were selected based on their morphology (spindle cells with spindle nuclei) from the region adjacent to tumors visualized at 10x magnification.

Cell culture, transfection, RNA extraction and quantitative RT-PCR

Cell culture, transfection, RNA extraction and quantitative RT-PCR were performed as previously described^{17,65} except for FFPE and LCM samples. The breast cancer cell line ZR-75 was grown in RPMI 1640 medium supplemented with 10% (v/v) FBS and 2mM L-Glutamine. LCM samples isolated using several successive 8 µm cut from FFPE bloc of each LCM sample were deposited on membrane slide (Molecular Machines & Industries, Canada). In general, 1-3 cuts were used for ovarian stroma and cancer epithelium, while 3-5 and 8-15 cuts were used for tumor microenvironment and Fallopian tube, respectively. LCM slides were stained with hematoxylin and eosin (Fischer, Canada) and were dissected using MMICellCut (Molecular Machines & Industries, Canada). Samples with the same cell type and tissue source were pooled. RNA extraction was performed using RNeasy FFPE kit (Qiagen, Canada) with the following modifications: dissected material were incubated in Aqua DePar 1X (Davis Diagnostics, Canada) for 15 minutes at 95°C then lysed by incubation in tris-guanidinium thiocyanate 4M (Sigma-Aldrich, Canada) and 10 µL of proteinase K (Qiagen, Canada) for 15 minutes at 55°C followed by 15 minutes incubation at 80°C. Finally, demodification was achieved by the addition of NH₄Cl (Sigma-Aldrich, Canada) to a final concentration of 2.5M and incubation for 20 min at 95°C. DNase treatment was performed using RNase-free DNase set (Qiagen, Canada). The resulting RNA (250 ng / reaction) was subjected to a single round of 25 PCR cycles of amplification using Sigma whole transcriptome amplification kit (Sigma-Aldrich, Canada). The resulting DNA libraries were purified using Qiaquick PCR purification kit (Qiagen, Canada) and diluted to a final volume of 12.5 mL. Global¹⁷ and isoform-specific⁶⁵ quantitative RT-PCR primers were designed and validated as previously described⁶⁵. The reliability of the gene expression assay using RNA samples from archived material was evaluated by comparing technical duplicates of amplification (Supplementary Fig. 4b), global gene expression (Supplementary Fig. 4a) and splicing pattern (Supplementary Fig. 5) obtained from the fresh frozen (FF) and fresh frozen formalin-fixed, paraffin- embedded (FFPE) matched samples-set and different methodology (e.g. quantitative RT-PCR vs endpoint PCR, see Supplementary Fig. 4c and Supplementary Fig. 5). The quality of the tissue dissection was verified visually (Fig. 1a) and through the expression of known cell-type markers (Fig. 1b). The expression levels of the different markers did not vary greatly between tissues of the same type confirming the absence of random sample-to-sample

variations in cell type (Fig. 1b and Supplementary Table 14). Furthermore, the changes in gene expression observed in tumor samples were gene-specific and not marker type specific (e.g. epithelial or stromal, see Supplementary Fig. 1c and Supplementary Table 14) confirming that any changes observed in cancer tissues are not due to variation in the level of cell type contaminations. All raw PCR data are available through <http://palace.lgfus.ca/pcrreactiongroup/list/219>.

Primer design and alternative splicing screens

The previously described LISA platform (LISA)¹¹ was updated with RefSeq database build 36.3³⁵ and used for automatic identification of ASEs and the design of isoform-specific PCR primers as described before¹². ASEs with strong evidence of the expression of both splicing isoforms (active ASEs) were identified by endpoint PCR using four individual ovarian tissues (one clear cell, one low malignant potential, one serous grade 1 and one serous grade 3) and 1 Fallopian tube tissue as previously described¹². Isoform-specific quantitative RT-PCR primers were designed for all active ASEs and validated using a universal human reference RNA as described earlier⁶⁵. Primer pairs showing no PCR amplifications were subsequently tested using RNA extracted from FFPE pool that include two normal ovaries and two serous ovarian cancer tissues, totalizing 1740 validated primer pairs (covering 870 ASEs). The screen for cancer-specific ASE markers was performed in two steps: 1) Discovery screen, which consists of surveying 870 ASEs in the LCM discovery samples-set and 2) Validation screen, which assess the expression of all ASEs obtained from the discovery screen in an independent LCM samples-set. Splicing pattern ($Q\psi$) values were calculated as described elsewhere¹⁷. In the discovery screen, ASEs were selected for further studies if they were expressed with a splicing shift ($\Delta Q\psi = Q\psi_{\text{TUMOR MICROENVIRONMENT}} - Q\psi_{\text{OVARIAN STROMA}}$ or $Q\psi_{\text{CANCER EPITHELIUM}} - Q\psi_{\text{FALLOPIAN TUBE EPITHELIUM}}$) of at least 15 in at least 3 out of 4 tissues. A t-test was used to rank ASEs and an arbitrary cut-off was applied at $P < 0.05$. In the validation screen, ASEs were considered cancer-associated if they were expressed in at least 3 out of 4 tissues (ovarian stroma) or 3 out of 5 tissues (tumor microenvironment, cancer epithelium and Fallopian tube epithelium) with a splicing shift ($\Delta Q\psi = Q\psi_{\text{TUMOR MICROENVIRONMENT}} - Q\psi_{\text{OVARIAN STROMA}}$ or $Q\psi_{\text{CANCER EPITHELIUM}} - Q\psi_{\text{FALLOPIAN TUBE EPITHELIUM}}$) of at least 15. The validity of the

markers identified in this screen and their capacity to detect cancer were further evaluated using a third independent set of undissected ovarian serous tumors. ASEs association with cancer was considered specific if they displayed a splicing shift ($\Delta Q\psi = Q\psi_{\text{OVARIAN CANCER}} - Q\psi_{\text{NORMAL OVARY}}$ OR $Q\psi_{\text{OVARIAN CANCER}} - Q\psi_{\text{NORMAL FALLOPIAN TUBE}}$) of at least 15 and a $P < 0.01$ (t-test). Data clustering were performed using the R package (www.r-project.org).

Splicing factors screen

Search of the NCBI literature database PubMed⁶⁴ retrieved 389 spliceosome associated genes. Validated quantitative RT-PCR assays were successfully developed for 370 out of the 389 splicing factors tested. The splicing factors discovery screen consisted of surveying the global expression of all 370 splicing factors in the LCM discovery samples-set (Supplementary Table 13) by quantitative RT-PCR. Genes expression levels relative to housekeeping genes¹⁷ was determined and those that vary between the tumor microenvironment and ovarian stroma, between cancer epithelium and Fallopian tube epithelium by at least two folds in at most one normal and at least three cancer samples or at most one cancer and at least three normal tissue were selected and validated using LCM validation sample set (Supplementary Table 13). Genes with expression that vary by at least two folds in tumor microenvironment / ovarian stroma or cancer epithelium / Fallopian tube epithelium in at most two normal and at least five cancer or at most two cancer and at least five normal tissue were considered associated. A similar strategy was used to screen the epithelial and established mesenchymal markers described in Fig. 1c.

Evaluation of the epithelial / mesenchymal expression pattern of splicing factors and variants

Cancer-associated exons were assigned as mesenchymal-enriched or epithelial-enriched by examining the mean exon inclusion level (ψ values) found in four⁶⁶ ovarian stroma and four Fallopian tube epithelium from the LCM discovery sample set (Supplementary Table 15 column 2 and Supplementary Table 2 column 2). Exons were considered to be tissue enriched (either mesenchymal or epithelial) when they exhibit a quantitative splicing shift ($\Delta Q\psi = Q\psi_{\text{OVARIAN STROMA}} - Q\psi_{\text{FALLOPIAN TUBE EPITHELIUM}}$) of at least 15. The preferential

epithelial / mesenchymal splicing pattern in cancer tissues was determined by comparing the quantitative splicing shift ($\Delta Q\psi = Q\psi_{\text{TUMOR MICROENVIRONMENT}} - Q\psi_{\text{OVARIAN STROMA}}$ or $Q\psi_{\text{CANCER EPITHELIUM}} - Q\psi_{\text{FALLOPIAN TUBE EPITHELIUM}}$) and the epithelial / mesenchymal-enriched exon assignment (Supplementary Table 1 column 2, Supplementary Table 2 column 2, Supplementary Table 7 column 2 and Supplementary Table 10 column 2). The same method was used to evaluate the preferential epithelial or mesenchymal expression of splicing factors (Supplementary Table 4 column 4) and those displaying at least two fold difference in a specific tissue where considered tissue-specific.

Identification of splicing factors dependent ASEs in cell lines

The expression levels of the 61 cancer-associated splicing factors were determined using quantitative RT-PCR in SKOV3ip1 cell line⁶⁶. A raw Ct of 30 or less for a 1 ng cDNA input was considered as “expressed”. Forty-seven splicing factors were expressed and successfully knocked down with two non-overlapping siRNAs in SKOV3ip1 (see Supplementary Table 17 for sequences). PTBP1 and PTBP2 (PTB1/2) were targeted simultaneously to provide overall PTB knockdown⁶⁷. ESRP1 and ESRP2 (ESRP1/2) were not expressed in SKOV3ip1 and thus were simultaneously knocked down in the epithelial cell ZR-75²⁴. RNA was harvested 72 hours after transfection and the quality of knockdown was evaluated by quantitative RT-PCR (Supplementary Table 5). RNA extracted from cells where splicing factors were knocked down by at least two folds or more was used to monitor the impact on 18 cancer-associated ASEs by quantitative RT-PCR. ASEs displaying quantitative splicing shift ($\Delta Q\psi$) of at least 10 between the knockdowns and mock transfection were considered significant. All cancer-associated ASEs regulated by at least one splicing factor and splicing factors regulating at least one ASE are shown in Fig. 3c. (the raw data are accessible through <http://palace.lgfus.ca/pcrreactiongroup/list/219229>).

Identification of RBFOX2 and QKI dependent ASEs

To determine QKI-responsive ASEs, QKI was knockdown using two independent siRNAs in SKOV3ip1 cell line. Effective knockdown of at least two folds was confirmed by

quantitative RT-PCR and the splicing pattern (ψ values) of 382 alternative cassette exons (ACE) were monitored by high-throughput endpoint PCR. A splicing shift ($\Delta\psi$) of at least 10 between the mock transfection and each siRNA was considered significant. The effect of QKI on RBFOX2 splicing targets was evaluated using an established set of 48 RBFOX2 splicing targets¹² as described above (Supplementary Table 8). The tendency of RBFOX2 and QKI co-regulated exons to be associated to tumor microenvironment was evaluated as described in Fig. 2a.

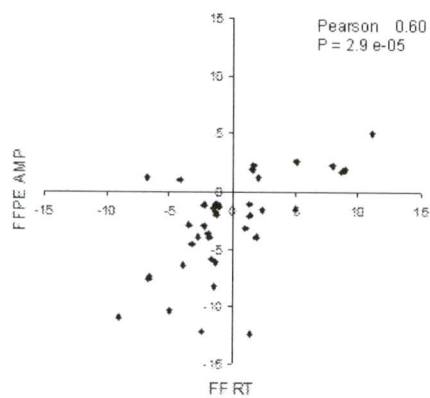
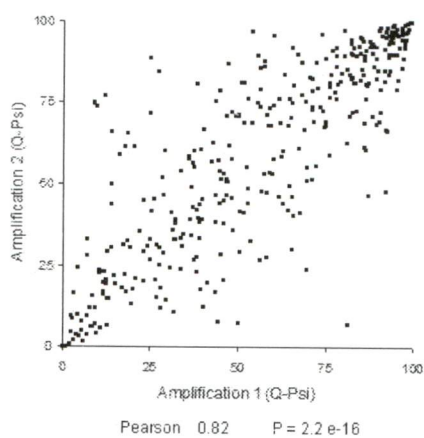
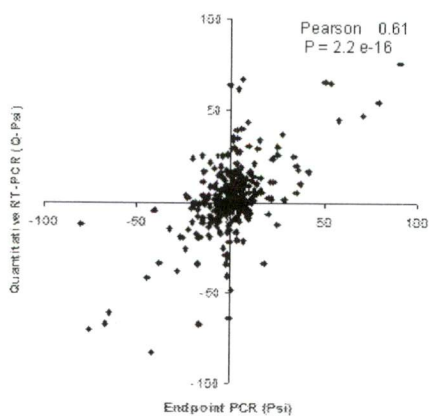
a**b****c**

Figure S4 Evaluation of the reliability of the formalin-fixed, paraffin-embedded (FFPE) based RNA expression assays by quantitative RT-PCR.

(a) Quantitative RT-PCR was conducted using reverse transcribed (RT) RNA extracted from fresh frozen (FF) undissected normal and serous ovarian cancer tissues or using amplified RNA (AMP) extracted from matching sets of formalin-fixed, paraffin embedded (FFPE) tissues (Supplementary Table 13). The scatter graph represents the mean fold difference in the expression levels of 42 genes detected in four normal ovaries and four high-grade serous ovarian cancer and normalized to housekeeping genes as previously described ¹. Most of the genes showing two or more fold difference in gene expression between normal and cancer FFPE tissues (21 out of 26) displayed the same expression pattern in the matching FF tissues. The Pearson correlation between the gene expression levels detected in RNA extracted from FF and FFPE and its accompanying p value are indicated on top right. (b) The reliability of RNA amplifications was evaluated by comparing the quantitative splicing patterns ($Q\psi$)¹ of 870 ASEs obtained from quantitative RT-PCR performed using a RNA sample independently amplified from the same tissue (OVC 278). Shown is a scatter plot of the quantitative splicing pattern ($Q\psi$) of 359 ASEs detectable in both samples. The Pearson correlation and its accompanying p value are indicated at bottom. (c) Comparison of splicing pattern obtained by endpoint PCR (ψ) and quantitative RT-PCR ($Q\psi$). The quantitative splicing shift ($\Delta Q\psi$) of 370 splicing events monitored by quantitative RT-PCR in dissected material was compared to the splicing shift ($\Delta\psi$) of an overlapping set of splicing event previously recorded by endpoint PCR in whole ovarian tumors using FF RT samples ². The Pearson correlation and its p value are indicated at the top right.

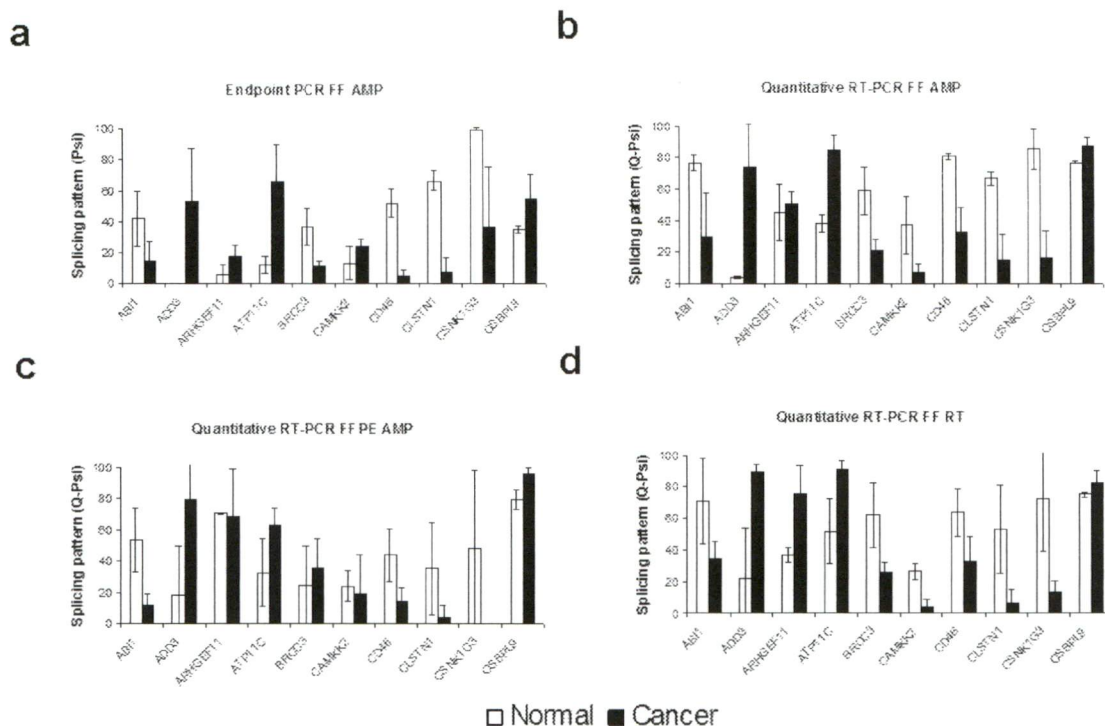


Figure S5 Comparison between quantitative and end-point PCR using amplified and non-amplified RNA extracted from fresh and fixed tissues.

(a) The mean splicing pattern (ψ) was monitored by endpoint PCR in four normal ovaries (white column) and four ovarian serous high-grade cancer tissues (black column) from amplified (AMP) and fresh frozen (FF) RNA in a panel of 10 splicing events. The mean quantitative splicing pattern ($Q\psi$)² were monitored in (b) FF and AMP RNA, (c) formalin-fixed, paraffin-embedded (FFPE) and AMP RNA and (d) FF and RT RNA using the same panel of 10 splicing events described in a and a FF and FFPE matched set (see Supplementary Table 13). Error bars indicate the mean standard deviation.

Supplementary Table legends

Tableau S1 Description of the cancer epithelial signature (CES) alternative splicing events (ASEs).

Official gene symbol are listed in column 1. The mean quantitative splicing pattern ($Q\psi$)¹ of the 8 ASEs associated with cancer epithelium (CES) was determined in Fallopian tube epithelium, cancer epithelium, ovarian stroma, undissected Fallopian tube and undissected ovarian tumor from the LCM discovery sample set (column 2), the LCM dissected validation set (column 3) or the undissected sample set (column 4). The mean quantitative splicing shift ($\Delta Q\psi$) of the difference between cancer epithelium and Fallopian tube epithelium, ovarian stroma and Fallopian tube epithelium, undissected ovarian tumor and undissected Fallopian tube are indicated. The global expression levels of the host genes harboring the alternative splicing events were determined using quantitative RT-PCR primer pairs targeting common exonic sequence and normalized to housekeeping genes as previously described¹. The t-test score of the difference between Fallopian tube and ovarian tumor are indicated. “No” indicates the absence of preferential epithelial / mesenchymal splicing pattern, “NE” means not expressed and “ND” means not determined. Cancer-related gene function and function related to the epithelial or mesenchymal cells are indicated in column 5. Laser capture microdissection (LCM) indicates dissected tissues.

Tableau S2 Description of the cancer stromal signature (CSS) alternative splicing events (ASEs).

Official gene symbol are listed in column 1. The mean quantitative splicing pattern ($Q\psi$)¹ of the 10 cancer stromal signature (CSS) ASEs was determined in ovarian stroma, tumor microenvironment, Fallopian tube epithelium, undissected normal ovary and undissected ovarian tumor from the LCM discovery sample set (column 2), LCM validation sample set (column 3) or the undissected sample set (column 4). The global expression levels of the host genes harboring the alternative splicing events were determined using quantitative RT-PCR primer pairs targeting common exonic sequence and normalized to housekeeping genes as previously described¹. The mean quantitative splicing shift ($\Delta Q\psi$) of the difference between tumor microenvironment and ovarian stroma, ovarian stroma and Fallopian tube epithelium, undissected ovarian tumor and undissected normal ovary are listed. The t-test score of the difference between undissected normal ovary and undissected ovarian tumor are listed. “No” indicates the absence of preferential epithelial / mesenchymal splicing pattern. Alternative splicing events that belong to the ovarian cancer signature (OCS) are indicated. Cancer-related gene function and function related to the epithelial or mesenchymal cells are indicated in column 5. Laser capture microdissection (LCM) indicates dissected tissues.

Tableau S3 Description of the common RBFOX2 and QKI splicing targets associated with the tumor microenvironment.

The mean quantitative splicing patterns ($Q\psi$)¹ of all tumor microenvironment associated ASEs that are affected by both RBFOX2 and QKI in cell lines (listed in Fig. 6a) were monitored in undissected tissues extracted from normal breast, breast tumor, normal ovary, Fallopian tube and ovarian tumors. The mean quantitative splicing pattern ($\Delta Q\psi$) is indicated for individual tissue and tissue type. The splicing shift and t-test score of the

difference between normal breast and breast tumor, estrogen receptor positive and negative breast tumor, Fallopian tube and ovarian tumor, normal ovary and ovarian tumor are listed. “NE” indicates not expressed.

Tableau S4 Expression patterns of cancer-associated splicing factors.

The global expression of each splicing factors gene associated to cancer (Fig. 3) was determined using quantitative RT-PCR in the different discovery and validation samples-set and normalized relative to housekeeping genes as previously described ¹. The fold difference between cancer epithelium and Fallopian tube epithelium (column 2), tumor microenvironment and ovarian stroma (column 3) or Fallopian tube epithelium and ovarian stroma (column 4) are shown. Association with cancer is indicated by yes or no. Splicing factors that are only expressed in cancer epithelium or tumor microenvironment and not in the reciprocal normal tissues are indicated by “CS” because numerical values could not be determined. Similarly, the factors that are only expressed in normal tissues are indicated by “NS”. “ND” indicates not determined and “NE” indicates not expressed. “No” indicates the absence of preferential epithelial / mesenchymal splicing pattern.

Tableau S5 Validation of the splicing factors knockdown.

The knockdown of the different splicing factors was validated using quantitative RT-PCR and the extent of the knockdown determined as fold difference relative to mock transfection (knockdown). The name of each factor (Gene name), names of the different siRNA (siRNAs), the cell type in which splicing factors are preferentially expressed (Associated cell type) and the cell line where the knockdown was performed (Cell line) are indicated.

Tableau S6 The expression of splicing factors associated with mesenchymal-to-epithelial transition (MET) are not systematically modified in cancer tissues.

The global expression of the EMT-associated splicing factors ³ was determined using quantitative RT-PCR in the different discovery and validation samples-set and normalized relative to housekeeping genes as previously described ¹. The fold difference between cancer epithelium and Fallopian tube epithelium (column 2) and tumor microenvironment and ovarian stroma (column 3) are shown. Association with cancer is indicated by yes or no. EMT / MET direction is indicated. Splicing factors that are only expressed in tumor microenvironment and not in the reciprocal normal tissues are indicated by “CS” because numerical values could not be determined. “ND” indicates not determined. The knockdowns were performed in the ovarian cancer cell line SKOV3ip1 with the exception of that of ESRP1/2, which was knocked down in the epithelial cell line ZR-75 because it is not expressed in SKOV3ip1.

Tableau S7 List of EMT associated alternative splicing events that are preferentially expressed in tumor tissues.

Official gene symbol are listed in column 1. The mean quantitative splicing pattern ($Q\psi$) ¹ of three splicing events associated with EMT ³ and tumor microenvironment was determined by quantitative RT-PCR in ovarian stroma, tumor microenvironment and Fallopian tube epithelium from the discovery LCM sample set (column 2) and validation sample set (column 3). The mean quantitative splicing shift ($\Delta Q\psi$) between tumor microenvironment and ovarian stroma, ovarian stroma and Fallopian tube epithelium are also indicated. Laser capture microdissection (LCM) indicates dissected tissues.

Tableau S8 Identification of RBFOX2 and QKI common splicing targets.

The splicing pattern (ψ) of splicing events that were previously associated with either RBFOX2 (RBFOX2 ASEs) or QKI (QKI ASEs) were determined after the knockdown of either RBFOX2 or QKI in the SKOV3ip1 cell line (see Fig. 5a). The name of each gene (Official gene symbol), the endpoint PCR primers used (ED PCR primers), source of each RNA examined (i.e. mock transfection (LF); RNA from cells transfected with unrelated siRNA (siCTL); RNA from cells transfected with QKI specific siRNA1 and siRNA2; RNA extracted from cells transfected with RBFOX2 specific siRNA1 and siRNA2), the RBFOX2 and QKI specificity of each splicing event (RBFOX2 ASEs; RBFOX2 and QKI ASEs; QKI ASEs), the primers used for quantitative PCR (qPCR primers) and the preferential expression of splicing events in epithelial and mesenchymal tissues (Epithelial mesenchymal preferential expression) are indicated on top. “ND” indicates not determined, “NA” indicates not available and “NE” indicates not expressed.

Tableau S9 RBFOX2 and QKI preferentially regulate the expression of mesenchymal-enriched exons.

Exons affected by RBFOX2 or QKI knockdown were classified based on their expression levels in epithelial (Epithelial-enriched exons) and stromal (Mesenchymal-enriched exons) tissues. The number of events leading to exon exclusion or inclusion, after the knockdowns of either RBFOX2 or QKI and the knockdown of both RBFOX2 and QKI in SKOV3ip1 cell line are indicated. Mixed behaviour indicates exons that are regulated by both QKI and RBFOX2 knockdowns in opposite direction (i.e. one factor knockdown induce inclusion while the other induce exclusion).

Tableau S10 Splicing pattern of RBFOX2 and QKI splicing targets in cancer tissues.

Official gene symbol are listed in column 1. The mean quantitative splicing pattern ($Q\psi$)¹ of all RBFOX2 and QKI regulated ASEs (Fig. 5a) was determined in ovarian stroma, tumor microenvironment, Fallopian tube epithelium, undissected normal ovary and undissected ovarian tumor from the LCM discovery sample set (column 2), LCM validation sample set (column 3) or the undissected sample set (column 4). The mean quantitative splicing shift ($\Delta Q\psi$) and t-test of the difference between undissected ovarian tumor and undissected normal ovary is listed. Laser capture microdissection (LCM) indicates dissected tissues.

Tableau S11 QKI-6 and QKI-7 splicing pattern in undissected tissues.

The quantitative splicing pattern ($Q\psi$) of QKI-6 and QKI-7 isoform was monitored in undissected tissues extracted from normal breast, breast tumor, normal ovary and ovarian tumor. The tissue type is indicated. The mean quantitative splicing shift ($\Delta Q\psi$)¹ and t-test of the difference between normal breast and breast tumor, normal ovary and ovarian tumor are listed. “NE” indicates not expressed.

Tableau S12 QKI relative global expression in undissected tissues.

The relative global expression of QKI was monitored in undissected tissues extracted from normal breast, breast tumor, normal ovary and ovarian tumor and was calculated relative to housekeeping genes as previously described¹. The tissue type is indicated. The relative expression is indicated for individual tissue and the mean for each tissue type. The mean

quantitative splicing shift ($\Delta Q\psi$) and t-test score of the difference between normal breast and breast tumor, normal ovary and ovarian tumor are listed.

Tableau S13 Clinical data of tissues used for laser capture microdissection (LCM) samples.

Patient identification, patient diagnostic, the cell type isolated after dissection, its new identification and the presence of matched fresh frozen tissues (mirror images existed in both paraffin and fresh frozen forms) are indicated.

Tableau S14 Global expression levels of established epithelial and stromal markers in normal and cancer tissues.

Quantitative RT-PCR was performed using primers against all known epithelial and stromal expression markers (Fig. 1c). The expression value normalized to housekeeping genes is indicated for each tissues examined. The description of each tissue is listed in Supplementary Table 13.

Tableau S15 Clinical data associated with the undissected serous ovarian tumors used in this study.

Patient identification, simplified identification, patient diagnostic, the percentage of cancer cells in the tumor samples and the percentage of epithelial cells in the tumor are indicated. "ND" indicates not determined. * indicated the percentage of epithelial cells generally found in distal Fallopian tube samples (<http://www.proteinatlas.org/>).

Tableau S16 Clinical data associated with the undissected ductal breast tumors used in this study.

Patient identification, patient diagnostic and estrogen receptor (ER) status are indicated. "ND" indicates not determined.

Tableau S17 List of siRNAs used for the knockdown of splicing factors.

Each siRNA name and its antisense sequence are indicated.

References

1. Prinos, P. et al. Alternative splicing of SYK regulates mitosis and cell survival. *Nat Struct Mol Biol* **18**, 673-9 (2011).
2. Venables, J.P. et al. Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol* **16**, 670-6 (2009).
3. Shapiro, I.M. et al. An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. *PLoS Genet* **7**, e1002218 (2011).

DISCUSSION

L'épissage alternatif (AS) est un mécanisme qui permet à un seul pré-ARNm de produire deux ou plusieurs isoformes ARNm. Les séquences résultantes sont hautement similaires et il est difficile de les détecter. Pour y remédier, nous avons élaboré une méthodologie basée sur la PCR en temps réel. Nous avons démontré qu'il est possible d'exploiter cette technique pour cribler un grand nombre d'évènements d'épissage alternatif (ASEs) dans des conditions PCR universelles, à partir d'ARN totaux de tissus fixés (FFPE), tout en demeurant quantitatif.

Des analyses bioinformatiques prédisent des changements de patron d'AS dans certains types de cancers et sont confirmés par des validations expérimentales sur un petit nombre de tissus dans certains cas. Par contre, l'AS est un processus qui est propre à chaque type cellulaire et les tumeurs analysées ont un contenu cellulaire hétérogène. Il n'est donc pas clair si ces changements de patron d'AS sont véritablement associés aux tissus cancéreux. Nos préparations homogènes de principaux types cellulaires ont permis de révéler la présence d'isoformes associées au cancer indépendamment de l'hétérogénéité des tumeurs. Étonnamment, les changements les plus intéressants ne se retrouvent pas dans les cellules cancéreuses proprement dites, mais dans les cellules « normales » en bordure de la tumeur. Notre analyse des facteurs de régulation de ces isoformes a permis d'affirmer que ces changements ne sont pas dus à un dérèglement anarchique des cellules cancéreuses, mais découlent plutôt d'un programme destiné à avantager les cellules cancéreuses.

Des analyses bioinformatiques prédisent que la plupart des ASEs affectant différentes portions codantes des gènes produisent des isoformes ARNm susceptibles d'être traduits en protéine. Dans quelques cas, des démonstrations expérimentales faites à partir de vecteur de surexpression ou de répression spécifique (siRNA, antisens) permettent de discriminer et mettre en valeur les fonctions des isoformes de l'AS. Bien que certains concepts d'outils moléculaires pouvant moduler le patron d'AS aient été publiés, il n'existe pas de règle générale de dessin. Nous avons non seulement conçu des règles de dessin pour le TOSS (targeted oligonucleotide silencer of splicing) mais également élargi son usage aux exons

cassettes, les ASEs les plus courants chez l'humain. Appuyé par notre méthodologie basée sur la PCR en temps réel, nous avons démontré que la plupart des TOSS reprogramment l'AS sans affecter l'expression globale. Puis, nous avons comparé directement l'effet de siRNA (répression de l'isoforme long et modulation du patron d'AS) et de TOSS (modulation du patron d'AS sans répression globale) sur 41 ASEs de gènes associés au cancer. L'analyse phénotypique de cellules cancéreuses indique que dans quatre cas, les propriétés pro-cancer (résistance à l'apoptose) ont pu être reliées à la répression d'un isoforme en particulier, évidence qui encore une fois, appuie notre hypothèse selon laquelle la production de certains isoformes sont en mesure de promouvoir le développement de tumeurs. De plus, dans le cas de l'exon 69 de la kinase SYK, la résistance à l'apoptose et la croissance en agar mou sont strictement dépendantes du ratio des isoformes et de l'expression globale, respectivement. Ces derniers résultats permettent d'expliquer certaines discordances retrouvées dans la littérature et accentuent le besoin d'intégrer l'analyse des isoformes d'AS dans les études d'expression génique.

1 Quantification de séquences hautement similaires

1.1 Quantification de l'expression globale et isoforme-spécifique sur une même plateforme

De façon générale, les données d'expression par PCR en temps réel sont rapportées sous forme de quantification relative pour chacune des isoformes. De cette façon, il est souvent difficile de comparer directement ces résultats avec ceux provenant d'autres plateformes (endpoint PCR, micropuces à ADN ou le séquençage de deuxième génération). Contrairement au endpoint PCR, il est possible de quantifier séparément chacun des isoformes par PCR en temps réel. Pour obtenir une valeur de ratio d'isoforme, le nombre de copie absolue d'une isoforme donnée doit être déterminé par rapport à une courbe standard afin de relier la quantité de fluorescence libérée au nombre de copies du transcrit (Bustin, 2000). Comme la précision de cette méthode dépend en grande partie du dosage des analytes utilisés pour produire la courbe standard, un plasmide unique incluant une copie de chacune des isoformes est idéal (Vandenbroucke *et al.*, 2001). Ainsi, il sera non seulement possible d'obtenir la proportion d'isoforme long et court (ratio d'isoforme) mais également l'expression globale du gène.

Afin d'éviter cette procédure laborieuse, nous avons imaginé une quantification par ratio des isoformes par PCR en temps réel sans faire intervenir une courbe standard. Par analogie au endpoint PCR, nous l'avons nommé $Q\psi$. De plus, puisque la formule du $Q\psi$ est indépendante des gènes de référence (HKGs), cela facilite le processus de quantification, particulièrement lorsque des tissus sont utilisés. Ceci permet donc d'obtenir une valeur de ratio des isoformes absolue simplifiée.

À l'heure actuelle, la détermination d'un changement d'expression globale entre deux échantillons ou plus par PCR en temps réel est effectuée grâce à une troisième paire d'amorces commune à tous les transcrits du gène et suivant le schéma de quantification relative avec normalisation par « housekeeping » (HKG). Or, il serait possible d'estimer un changement d'expression globale à partir des données obtenues avec les paires d'amorces isoformes-spécifiques. En effet, si les deux isoformes présentent un changement d'expression dans le même sens (les deux isoformes augmentent ou diminuent), c'est qu'il y a un changement d'expression globale. Pour confirmer ce changement, la troisième paire d'amorces ciblant la totalité des transcrits du gène doit être utilisée pour valider cette nouvelle approche. Il serait possible d'exploiter cette stratégie à la ré-analyse de données brutes issues des deux criblages effectués à la figure 2 du chapitre 4. Pour donner un exemple, la portion du transcriptome qui est affectée uniquement par la transcription et non par l'AS dans le stroma adjacent et dans les cellules épithéliales du cancer séreux avancé de l'ovaire pourrait être révélée.

En théorie, les micropuces à ADN de type « exon array », « exon/junction array » ou le séquençage de deuxième génération peuvent tous être exploités tant pour l'expression génique que pour déterminer les isoformes d'AS différenciellement exprimés (Costa *et al.*, 2010). En pratique, les algorithmes de normalisation complexes et le manque de pouvoir statistique préviennent l'utilisation adéquate des deux types de données (expression globale et ratio d'isoforme). De plus, et ce, malgré les récentes avancées (Waldron *et al.*, 2012) l'utilisation de FFPE est encore laborieuse pour les micropuces à ADN et le séquençage de deuxième génération. Bien que nous n'ayons criblé qu'un peu moins de 900 ASEs, il est

théoriquement possible de cribler la totalité des 3322 ASEs de la base de données RefSeq prédit par RASE, soit un nombre comparable aux études de micropuces à ADN (Pan *et al.*, 2004, Fagnani *et al.*, 2007).

1.2 Quantification transcrit-spécifique

Il n'est pas rare que plusieurs ASEs aient lieu sur un même pré-ARNm. Dans un certain nombre de cas d'ASEs complexes (par exemple un exon ayant un 5'ss et un 3'ss alternatif), RASE est en mesure de simplifier le tout en décortiquant les ASEs sous-jacents. Ainsi, la région IIICS du gène FN1 composée d'un alt3' et d'une rétention d'intron est perçue par RASE comme quatre transcrits distincts (chapitre 2, Fig. 14). Le positionnement d'amorce et la proximité des deux ASEs permettent la quantification de tous les transcrits séparément. De façon intéressante, l'isoforme moyen sans intron de FN1 IIICS nécessite deux BSPs (chapitre 2, Fig. 14). La situation se complexifie lorsqu'un pré-ARNm possède deux ASEs ou plus qui sont distancés par plus de 300 nt. Le gène FN1 possède également deux autres régions d'AS, dont EDB, située à ~ 3 kb en amont de IIICS (chapitre 2, Fig. 8). Comme l'amplification d'un amplicon de 2000 bp est impraticable par PCR en temps réel, il est impossible de déterminer la co-régulation des deux ASEs. En d'autres termes, il est possible de quantifier l'isoforme long et l'isoforme court de EDB, mais il n'est pas possible de préciser si le transcrit court par rapport à EDB comprend ou non l'intron 93 de IIICS. On parle alors du concept de famille d'isoformes longs et famille d'isoformes courts.

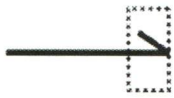



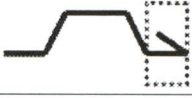



Seule la technique de buvardage de type Northern permet d'apprécier chacun des transcrits d'un gène, à la condition que chacune des régions alternatives soit de longueur significativement plus élevée que la limite de résolution du gel, ce qui est somme toute rare. Le séquençage de deuxième génération ne permettra de « connecter » deux ASEs que si les « reads » sont de taille suffisante pour englober ces ASEs. En ce sens, l'ARN fragmenté provenant de FFPE ne pourra être utilisé et de l'ARN de haute qualité sera nécessaire pour ce type d'expérience. La quantification transcrit-spécifique à haut débit est le prochain défi des techniques de quantification des isoformes de l'AS.

1.3 Application des règles de dessin d'amorce-jonction pour d'autres cas de séquences très semblables

Outre les isoformes de l'AS, il existe des variations dans la séquence de notre ADN d'un seul nucléotide nommé « single nucleotide polymorphism » (SNP). La détection de ces SNPs par PCR est laborieuse. La PCR en temps réel a été utilisée maintes fois pour la détection de SNP, mais presque toujours conjointement avec une sonde fluorescente de type Taqman™ (Hampe *et al.*, 2001). Il n'est pas toujours évident d'obtenir une hybridation différentielle suffisamment grande pour régler le problème de faux positif. Nous avons montré au chapitre 1 que certaines de nos amorces-jonction (BSP) sont capables de discerner des séquences distinctes d'un seul nucléotide lorsque le mésappariement sur l'isoforme long est strict. Nous émettons donc l'hypothèse qu'il est possible d'exploiter nos règles de dessin de BSP afin d'amplifier spécifiquement l'allèle porteur de SNP sans amplifier l'allèle sauvage (Tableau 1). Nous avons fait l'étude de trois cas d'amorce-jonction comportant des mésappariements en position Nx et Nz, mais dont le Ny est pairé, soit une situation similaire à la stratégie proposée par Zhou et collaborateurs (2004) (voir Tableau 1). En accord avec ces derniers, nous n'avons pas été en mesure d'obtenir un produit PCR même après 50 cycles. De façon intéressante, une des trois amorces possède des mésappariements permissifs en position Nx et Nz. Ceci signifie que l'introduction volontaire d'un mésappariement en position Nx (permissif ou strict) est en mesure d'éliminer le problème de faux positif puisque nous n'observons aucun signal. Malheureusement, nous n'avons pas vérifié l'amplification d'une amorce dont le Nx est mésapparié, mais dont le Ny et Nz sont pairés. Cette réaction est nécessairement moins efficace vis-à-vis un appariement parfait, mais il est fort probable qu'un mésappariement permissif (A/C, G/T et C/T) soit bien toléré. L'avantage de cette stratégie réside du fait que peu importe le nucléotide en Nx, il sera toujours possible de concevoir un mésappariement permissif.

Tableau 1 Stratégies de dessin d'amorce allèle-spécifique.

Pour amplifier spécifiquement l'allèle SNP sans amplifier l'allèle sauvage, divers positionnements de mésappariement ont été proposés. La position du SNP est encadrée en pointillé.

Allèle SNP	Allèle sauvage	Références
Nx Ny Nz	Nx Ny Nz	
		Metcalfe P. et al., 1995
		Zhou G.H. et al., 2004
		Yaku H. et al., 2008
		Aono T. et al., 2000

Plusieurs stratégies endpoint PCR ont été mises de l'avant (Tableau 1) et le design original consiste à païrer le nucléotide en 3' terminal de l'amorce avec le SNP (Metcalfe *et al.*, 1995). Conséquemment, cette amorce aura un mésappariement en 3' terminal sur l'allèle sauvage. S'il s'agit d'une substitution de type T pour A, G pour C, A pour C, C pour G, G pour T ou G pour A (en rouge dans la figure 1A), ceci a pour effet de produire un mésappariement strict en 3' terminal. Dans ces cas, nos résultats suggèrent que l'amplification de l'allèle porteur du SNP sera spécifique (Fig. 1A). Par contre, la fréquence de ce type de substitution est faible, et la plupart des SNPs provoqueront des mésappariements permissifs (A/C, G/T et C/T) lorsque cette stratégie est employée (Fig. 1B). Dans ces cas, il est possible de substituer volontairement un ou des nucléotides de la portion 3' terminal afin de diminuer l'amplification non-désirée de l'allèle sauvage. Les données de notre étude systématique des paramètres essentiels à la spécificité des amorces-jonctions suggèrent donc qu'il serait possible de réaliser des règles générales de dessin d'amorce SNP afin de discriminer tous les types de substitution SNP vis-à-vis leur séquence de type sauvage, ce qui demeure encore laborieux à ce jour. Cette stratégie

pourrait se révéler particulièrement avantageuse lorsque des échantillons FFPE sont utilisés. En effet, les modifications de bases qu'entraîne la fixation se prêtent mal aux techniques de séquençage (Qiu *et al.*, 2008).

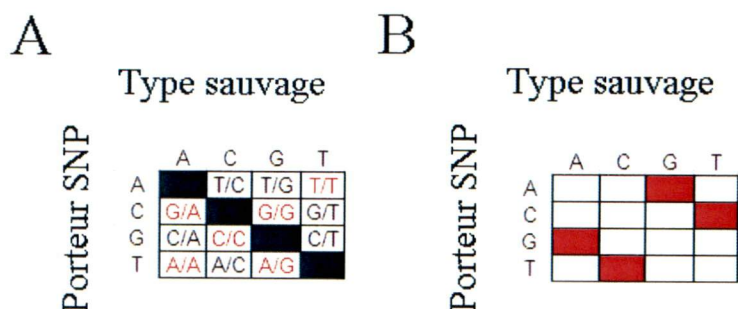


Figure 1 Fréquence et mésappariement de l'amorce allèle-spécifique sur l'allèle de type sauvage selon le type de SNP.

Tableaux indiquant toutes les substitutions nucléotidiques possibles entre le type sauvage et le porteur SNP. A) L'intersection représente le mésappariement amorce SNP/allèle de type sauvage et les mésappariements stricts sont indiqués en rouge. Par exemple, pour la substitution d'un C pour un A, l'amorce SNP possède un T en 3' terminal pour s'apparier avec le A de l'allèle SNP et se mésapparier avec le C de l'allèle de type sauvage (T/C). B) Les intersections en rouge sont les substitutions les plus fréquentes.

Dans cet ordre d'idées, les gènes dupliqués des protéines ribosomales de *Saccharomyces cerevisiae* sont excessivement semblables. Afin de quantifier l'expression d'un gène à la fois, Parenteau et collaborateurs (2011) se sont inspirés des règles de dessin d'amorce-jonction issues du chapitre 1. Ainsi, la quantification de l'expression de chacun des gènes dupliqués de protéines ribosomales a permis de révéler la contribution des séquences introniques dans leur inter-régulation.

2 Nouvelles stratégies pour la modulation de l'AS de cibles endogènes

2.1 Modulation d'ASEs simples par les TOSS

Les TOSS sont des outils moléculaires conçus pour moduler l'AS en favorisant l'isoforme court sans changer l'expression globale. Au chapitre 2, nous avons exploité notre stratégie de quantification isoforme-spécifique afin de confirmer la réelle modulation de 42 ASEs simples, faisant du TOSS un outil simple à utiliser en cellules. Le design rationnel de TOSS permet de garantir une modulation de l'épissage dans 87% des cas d'exon cassette simple. Au contraire, il est recommandé de tester plusieurs SSOs (splice switch oligonucleotides) afin de s'assurer que l'un d'entre eux fonctionne adéquatement (Aartsma-Rus *et al.*, 2005; Hua *et al.*, 2007). Contrairement au TOSS, l'effet de ceux-ci est beaucoup plus dépendant de la séquence ciblée sur l'exon (Aartsma-Rus *et al.*, 2005; Hua *et al.*, 2007).

De façon surprenante, la modulation de certain TOSS est également accompagnée d'une diminution globale de l'expression, un effet similaire à ce que l'on observe avec un siRNA ciblant spécifiquement l'isoforme long (siRNA long). Il est possible que ce comportement, rebaptisé modulation de type B, tienne origine de l'interférence du TOSS avec les facteurs d'épissage constitutif. Ainsi, dans une fraction des cas, le TOSS de type B hybridé à sa cible culminerait vers un complexe d'épissage non-productif (qui ne formerait pas d'ARNm mature). En ce sens un antisens ciblant une jonction constitutive peut altérer l'expression globale d'un gène (Mercatante *et al.*, 2001). Par contre, les effets sur l'expression globale ne sont généralement pas systématiquement vérifiés. Il n'est donc pas possible de déterminer si la modulation de type B est plus susceptible via un TOSS ou un SSO. Alternativement, le même résultat pourrait découler d'une dégradation isoforme-spécifique, quoique moins probable. L'ASE en soi ne semble pas être un facteur puisque nous avons observé de la modulation de type A et de type B pour des TOSS situés sur un même ASE (chapitre 2, Fig. 7). De plus, cette modulation de type B ne semble pas être dépendante de hnRNPA1 puisque nous en avons observé pour un antisens (chapitre 2, Tableau 5). Il serait intéressant d'approfondir le mécanisme d'action du comportement de type B de façon à mieux concevoir les TOSS de type A. En comparant les TOSS de type A

et B et leur cibles, il est probable qu'un dénominateur commun puisse être identifié (séquence antisens, séquence cible à proximité du TOSS).

2.2 Modulation d'ASEs complexes par les TOSS

Originellement, le TOSS a été employé pour la modulation d'un alt5' (Villemaire *et al.*, 2003). Nous avons élargi le spectre des ASEs pouvant être ciblé par le TOSS en incluant les exons cassettes simples et les alt3'. Nous avons testé le comportement des TOSS dans des cas d'ASE complexes, comme par exemple une compétition de plusieurs 5'ss et/ou 3'ss. Un TOSS situé entre deux 5'ss ou deux 3'ss sur un exon lui-même alternatif favorise l'exclusion de l'exon entier. Cette observation rejoint le modèle de Wang et collaborateurs (2006), qui propose que des ESS, dont la séquence est apparentée à hnRNPA1 (classe I), bloque le 3'ss en amont et le 5'ss en aval (Fig. 2A).

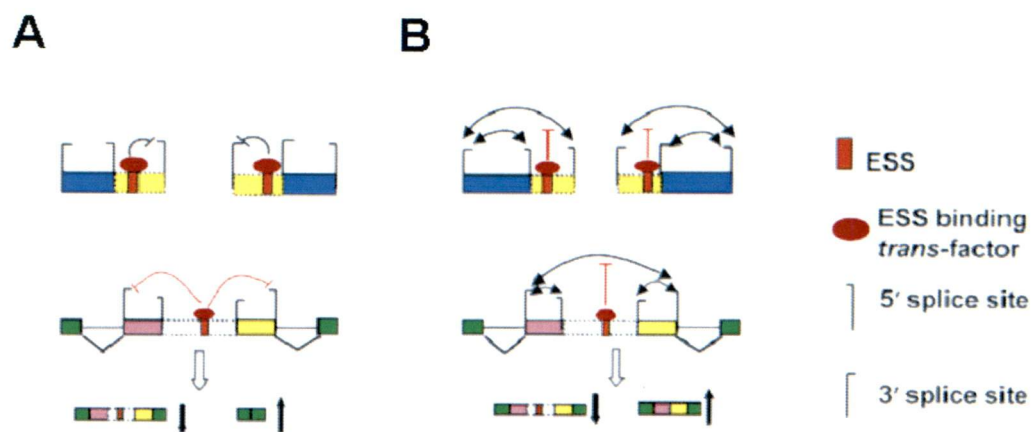


Figure 2 Mécanismes d'action des ESS selon le modèle de Wang et collaborateurs (2006).

Les ESS dont le motif est apparenté à A) hnRNPA1 (classe I) agissent en bloquant le 5'ss en aval et le 3'ss en amont et B) hnRNPF/H ou au 5'ss consensus (classe II) agissent en bloquant la communication inter 3'ss et 5'ss tant exonique qu'intronique. Images adaptées de Wang (2006).

Ce modèle prédit également que des séquences apparentées à hnRNPF/H ou similaires au 5'ss consensus agissent plutôt en bloquant la définition de l'exon (Fig. 2B). Si ce modèle

dit vrai, alors un TOSS munit d'une telle séquence serait en mesure de favoriser le site d'épissage distal même dans le cas où l'exon est lui-même alternatif. Ainsi, il serait possible de favoriser préférentiellement le 5'ss alternatif ou le 3'ss alternatif (Fig. 3A), une possibilité de design rationnel qui n'a pas été démontré par les outils moléculaires actuels. De plus, un TOSS situé sur l'exon alternatif mais en dehors des deux sites d'AS favoriserait l'exclusion de l'exon entier, tel que nous l'avons rapporté pour LIG3_cass_4_t au chapitre 2 (Fig. 3B). Dans les cas d'exon cassette multiple, le TOSS demeure spécifique à l'exon ciblé. Il serait intéressant de vérifier l'effet combiné de TOSS ciblant chacun des exons alternatifs contigus. Ainsi, il serait possible de promouvoir l'exclusion en bloc. Entropiquement parlant, la saturation de tous les exons d'un même pré-ARNm est défavorable et cette stratégie risque d'être également dispendieuse. Alternativement, le modèle de Wang et collaborateurs (2006) prédit qu'une séquence de classe I située dans un intron flanqué de deux exons alternatifs promouvoit l'exclusion des deux exons. Ainsi, un seul TOSS serait nécessaire dans le cas d'un double exon cassette (Fig. 3C). Les résultats du TOSS FN1_intron_L_t (chapitre 2, Fig. 14) sont consistents avec le modèle de Wang et collaborateurs (2006). En effet, ce TOSS de classe I favorise l'exclusion de l'exon alternatif 267 alors que l'exon 107 en aval n'est pas exclu puisqu'il possède un 5'ss constitutif. En revanche si une séquence de classe II était utilisée sur un intron difficile à épisser, la définition de l'intron serait favorisée et permettrait ultimement d'inhiber la rétention d'intron (Fig. 3D).

Comme il est connu que la longueur des introns est proportionnelle à leur épissage dans des extraits nucléaires, les longs introns étant difficilement épissés, nous avons imaginé une deuxième série de nouveaux dessins de TOSS autour de ce concept. Pour stimuler leur épissage, des TOSS pourraient cibler les extrémités d'un intron de grande longueur. Il est postulé que hnRNPA1 se dimérise une fois liée à sa séquence, permettant ainsi de rapprocher les exons à se joindre et conséquemment, l'inclusion de l'exon alternatif. Bien que le traitement simultané par deux TOSS (chacun ciblant une extrémité) est efficace *in vitro*, ils ne semblent pas agir selon un mécanisme coopératif, probablement à cause de motif de liaison UAGGG déjà présent dans la séquence intronique étudiée (Martinez-Contreras *et al.*, 2006). Il serait néanmoins intéressant de vérifier l'efficacité de cette

stratégie sur des ASEs endogènes flanqués d'intron long (Fig. 3E). En extrapolant ce concept, il est possible d'imaginer que des motifs TOSS positionnés de part et d'autre d'un exon alternatif puissent recruter et promouvoir la dimérisation de hnRNPA1, favorisant par le fait même l'exclusion d'exon, voire un groupe d'exon multiple (Blanchette and Chabot, 1999) (Fig. 3E). Ainsi, le ciblage simultané de TOSS sur des séquences introniques permettrait de favoriser tant l'exclusion ou l'inclusion d'exon, selon les besoins. Il serait intéressant de tester ces nouveaux dessins de TOSS aux cas étudiés en détail au chapitre 2. Le TOSS pourrait donc s'avérer être un outil complémentaire général qui permettrait non seulement de moduler l'AS et d'en évaluer la fonction des isoformes, mais également d'étudier le rôle des SREs dans un contexte endogène. En effet, au lieu de modifier la séquence du gène pour y implanter un ESS, il est possible grâce au TOSS d'exhiber spécifiquement un ESS ou ISS sur le pré-ARNm endogène.

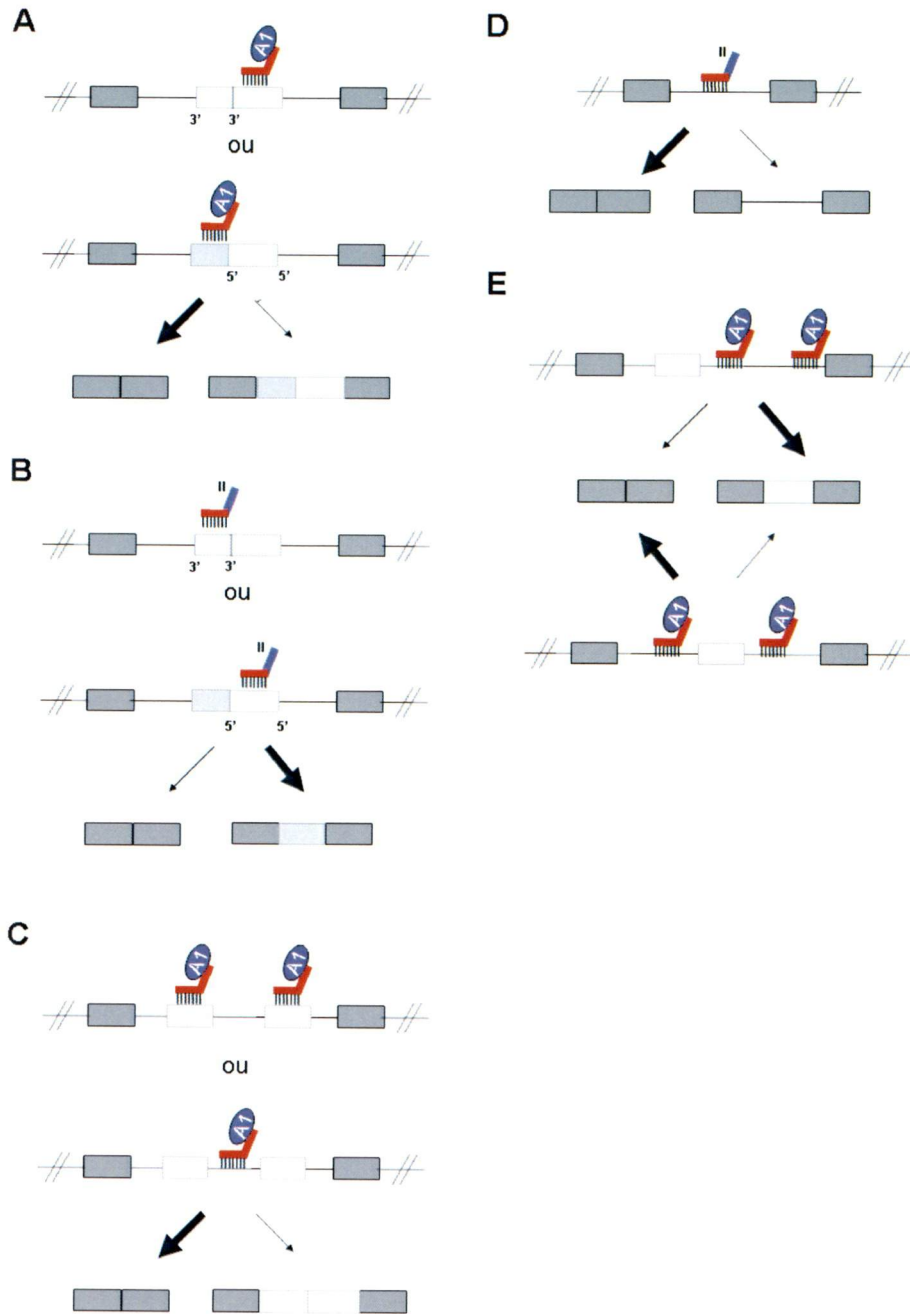


Figure 3 Nouveaux dessins de TOSS.

A) Cas complexe de type exon cassette-alt3.' B) Cas complexe de type exon cassette-alt5'. C) Cas d'exon cassette multiple. D) Cas de rétention d'intron. E) Cas d'exon cassette. II = Séquence de classe II selon Wang et collaborateurs (2006).

2.3 Nouvelles stratégies pour stimuler l'inclusion d'exon

Les stratégies illustrées à la figure 3B et 3E sont dessinées dans le but de promouvoir l'isoforme long. Ces outils sont d'autant plus attrayants puisqu'ils sont susceptibles de procurer un gain de fonction. Bien que des siRNAs courts (ciblant l'isoforme court) furent testés avec succès, nous n'avons pas réussi à produire un outil moléculaire capable de stimuler l'inclusion d'exon sans effet sur l'expression globale de cible endogène. Très récemment, une étude ayant pour but de diminuer la toxicité associée aux modifications PS et d'augmenter l'effet d'inclusion des TOES a été publiée (Owen *et al.*, 2011). Malgré de nombreuses variations de séquences et de chimie de la portion « tail », aucun TOES (targeted oligonucleotide enhancer of splicing) n'est plus puissant pour induire l'inclusion de l'exon 7 de SMN2 que celui de la publication originale (Skordis *et al.*, 2003). Deux conclusions en accord avec nos données et pertinentes pour de futurs travaux ressortent de cette étude: 1) Il est possible de reprogrammer l'AS avec une portion « tail » en ARN ou 2'OMe; 2) L'effet d'inclusion est très sensible à la position et l'orientation de la portion « tail ». En effet, nous avons observé que dans 7 cas sur 10, l'utilisation d'un SSO reprogramme l'AS vers la forme courte, soit l'inverse de l'effet attendu avec un TOES (voir chapitre 2, Tableau 1). Bien que les TOSS soient a priori dessinés afin de cibler des ESEs putatifs, des antisens ciblant des ESS putatifs n'ont pas été en mesure de stimuler l'inclusion d'exon (Annexe 1). Bref, cela signifie qu'il est plus facile d'induire l'exclusion d'un exon cassette. En d'autres termes, les ESEs sont beaucoup plus dépendants de leurs contextes (séquences avoisinantes et facteur d'épissage en *trans*) que les ESS.

3 Impact fonctionnel des isoformes de l'épissage alternatif dans la cancérogenèse

3.1 Impact fonctionnel des isoformes de l'épissage alternatif de gènes associés au cancer

Des analyses bioinformatiques prédisent que la plupart des ASEs affectant différentes portions codantes des gènes produisent des isoformes ARNm susceptibles d'être traduites en protéine. Par contre, un nombre limité de gènes reliés au cancer ont été ciblés par des modulateurs d'AS (Tableau 2) ou par des siRNAs isoforme-spécifiques (Tableau 3), et l'impact sur le phénotype vérifié. Dans de très rares cas, les effets des modulateurs ont été comparés directement (Jiang *et al.*, 2008) ou peuvent être comparés indirectement (Ghigna *et al.*, 2005, Ghigna *et al.*, 2010) aux effets de répresseurs, mais il est toujours difficile de réconcilier l'ensemble des résultats et de conclure sur la nécessité de diminuer l'expression globale pour induire un phénotype en cellule.

Tableau 2 Exemples de gènes associés au cancer ayant été ciblés avec des modulateurs d'AS (SSO, TOSS ou TOES) dans des cellules cancéreuses humaines.

Le type d'évènement ciblé, le type de chimie utilisé pour le modulateur, son positionnement et l'effet sur l'épissage et phénotypique résultant est indiqué. La définition de chacune des chimies de modulateur est mentionnée à la figure 20 de l'introduction excepté 2'-MOE (2' Ométhoxyéthyl). N.D. = non disponible.

Gène	Type de ASE	Chimie du modulateur	Positionnement	Effet sur l'épissage	Effet phénotypique associé	Référence
MDM2	Cassette	PNA	Chevauche la jonction intron-exon du 3'ss,	Induit la rétention de l'intron en amont et l'exclusion de l'exon ciblé.	Sensibilise les cellules JAR à la camptothécine	(Shiraishi <i>et al.</i> , 2010)
ERBB2 (HER-2)	Cassette	PNA	Chevauche la jonction intron-exon du 3'ss	Induit l'exclusion de l'exon ciblé	N.D.	(Pankratova <i>et al.</i> , 2010)
	Cassette	2'-MOE	Chevauche la jonction exon-intron du 5'ss	Induit l'exclusion de l'exon ciblé	N.D.	(Wan <i>et al.</i> , 2009)
MCL-1	Cassette	Morpholino	Deux oligonucléotides ciblent l'exon désiré : Un chevauchant la jonction intron-exon du 3'ss, l'autre chevauchant la jonction exon-intron du 5'ss	Induit l'exclusion de l'exon ciblé	Induit l'apoptose dans les cellules BCC et AGS.	(Shieh <i>et al.</i> , 2009)
BCL2L1 (Bcl-x)	Alt5'	2'-MOE	Complètement exonique (séquence exacte non dévoilée)	Redirige vers le 5'ss en amont	Sensibilise les cellules A549 au cisplatine et aux UVB	(Taylor <i>et al.</i> , 1999)

		2'OMe (TOSS)	Complètement exonique en aval du 5'ss	Redirige vers le 5'ss en amont	N.D.	(Villemaire <i>et al.</i> , 2003)
		PS-2'OMe	Chevauche la jonction exon-intron du 5'ss	Redirige vers le 5'ss en amont	Induit l'apoptose dans les cellules PC-3	(Mercatan <i>et al.</i> , 2001)
		Chimère PNA-peptide Ser-Arg (ESSENCE)	Complètement exonique en amont du 5'ss	Redirige vers le 5'ss en amont	Induit l'apoptose dans les cellules HeLa	(Wilusz <i>et al.</i> , 2005)
		PS-2'OMe	Chevauche la jonction exon-intron du 5'ss	Redirige vers le 5'ss en amont.	Induit l'arrêt du cycle cellulaire en G1	(Zhaozui <i>et al.</i> , 2006)
VEGF	Cassette	Morpholino	Chevauche la jonction intron-exon du 3'ss	Induit l'exclusion de l'exon ciblé.	Induit la vascularisation dans les cellules PC-3	(Catena <i>et al.</i> , 2007)
TAF6	Alt5'	PS-2'OMe	Complètement exonique, en aval du 5'ss	Redirige vers le 5'ss en amont.	Induit l'apoptose dans les cellules HeLa	(Wilhelm <i>et al.</i> , 2008)
BRCA1	Cassette	Chimère PNA-peptide Ser-Arg (ESSENCE)	Complètement exonique	Stimule l'inclusion de l'exon 18 portant la mutation E1694X	N.D.	(Cartegni and Krainer, 2003)
FGFR1	cassette	Morpholino	Masque un ISS	Stimule l'inclusion de l'exon ciblé	N.D.	(Bruno <i>et al.</i> , 2004)
MSTR1 (Ron)	Cassette	2'OMe (TOES)	Complètement exonique	Stimule l'inclusion de l'exon ciblé	N.D.	(Ghigna <i>et al.</i> , 2010)

PKC	Alt5'	2'MOE	Chevauche le 5'ss en aval	Redirige vers le 5'ss en amont	N.D.	(Jiang <i>et al.</i> , 2008)
-----	-------	-------	---------------------------	--------------------------------	------	------------------------------

Tableau 3 Exemples de gènes associés au cancer et ciblés avec des siRNAs isoforme-spécifiques dans des cellules cancéreuses humaines.

Le type d'évènement ciblé, l'isoforme et l'effet phénotypique associé est indiqué.

Gène	Type de ASE	Isoforme ciblé	Effet phénotypique associé	Référence
MSTR1 (Ron)	Cassette	Court (exon 11)	Réduit la migration des cellules 293 surexprimant ASF/SF2	(Ghigna <i>et al.</i> , 2005)
H-Ras	Cassette	Long (p19)	Induit la prolifération des cellules HeLa	(Camats <i>et al.</i> , 2009)
BIRC7	Alt3'	Court (isoforme β)	Réduit la prolifération des cellules HeLa	(Crnkovic-Mertens <i>et al.</i> , 2006)
CCND1	Rétention d'intron	Long (b isoforme)	Réduit l'invasion des cellules T24	(Kim <i>et al.</i> , 2009)
eIF4H	Cassette	Long (isoforme 1)	Réduit la prolifération des cellules LOVO	(Wu <i>et al.</i> , 2011)
BCL2L1 (Bcl-x)	Alt5'	Long (Bcl-xL)	Réduit la prolifération des cellules CaOV3	(Dodier and Piche, 2006)
DNMT3B	Cassette	Court (DNMT3B4)	Réduit la méthylation du promoteur RASSF1A	(Wang <i>et al.</i> , 2007)
BRD8	Cassette	Long (p120 β)	Réduit l'activation du récepteur de la thyroïde dans les cellules PC-3	(Hosoya <i>et al.</i> , 2008)
KLF6	Alt5' et 3'ss cryptique	Court (isoforme SV1)	Induit l'apoptose des cellules A549	(DiFeo <i>et al.</i> , 2008)
PKC	Alt5'	Long (PKC δ VIII)	Induit l'apoptose des cellules NT2	(Jiang <i>et al.</i> , 2008)
CXCR3	Alt3'	Long (isoforme B)	Augmente la prolifération des cellules MDA-MB-435	(Datta <i>et al.</i> , 2006)
CD99	Cassette	Court	Réduit l'invasion des cellules MDA-MB-435	(Byun <i>et al.</i> , 2006)
MDM2	Cassette multiple	Long (DMX211)	Réduit la prolifération des cellules ARO	(Giglio <i>et al.</i> , 2005)
NEK2	Rétention d'intron	Long (NEK2B)	Induit un délai lors de la mitose des cellules HeLa	(Fletcher <i>et al.</i> , 2005)
MYLK	Cassette	Long (Isoforme 1)	Augmente la résistance trans-épithélial (réduit la perméabilité)	(Clayburgh <i>et al.</i> , 2004)

Afin de déterminer l'impact fonctionnel de l'AS dans la cancérogenèse, nous avons sélectionné des gènes dont les isoformes sont associés au cancer du sein et/ou de l'ovaire (Klinck *et al.*, 2008, Venables *et al.*, 2008). Puis, nous avons confronté directement l'impact de siRNAs (répression de l'isoforme long) et de TOSS (répression de l'isoforme long et augmentation de l'isoforme court) sur l'induction d'apoptose de cellules cancéreuses ovariennes (SKOV3ip1) de 41 ASEs endogènes. Nous avons d'abord émis l'hypothèse qu'il serait plus facile d'identifier des isoformes avec des relations fonctionnelles de type antagoniste, c'est pourquoi nous avons systématiquement criblé les ASEs en trois phases lors du chapitre 3. D'abord une sélection positive avec des siRNAs longs et des TOSS. Puis, une sélection négative en évaluant l'effet de siRNAs globaux. Enfin, nous avons changé le contexte (l'essai phénotypique) dans le cas de SYK pour en déterminer les conséquences sur la relation fonctionnelle des deux isoformes. Ainsi, nous avons identifié quatre ASEs dont le phénotype résultant est relié au ratio d'isoforme plutôt qu'à l'expression globale du gène, démontrant l'impact de l'AS sur la fonction de gènes associés au cancer. De plus, pour un même ASE, nous avons observé que la dépendance du phénotype quant à l'expression globale et le ratio d'isoforme peut varier drastiquement en fonction du contexte phénotypique (apoptose, prolifération et croissance en suspension). Bien que certains gènes ciblés soient effectivement impliqués dans le processus d'apoptose (MCL1, PTPN13), la plupart des gènes criblés ne le sont pas *a priori*. Ce criblage systématique permet donc d'annoter de nouvelles fonctions isoforme-spécifique à des gènes associés au processus de cancérogenèse.

Dans le cas du gène MCL1, le TOSS devrait donner le même effet que les siRNAs longs, mais aucune modulation des isoformes protéiques n'est détectée par immuno-buvardage, malgré la modulation évidente des isoformes ARNm et contrairement à ce que nous avons observé pour SYK, CASP8 et CASP10 (voir Fig. S4 de l'article 2 et Fig. 7 du chapitre 2). En appui avec la relation antagoniste des isoformes long (MCL-1_L) et court (MCL-1_S) mise en lumière par notre approche, plusieurs membres de la famille Bcl ont des effets similaires dont : Bcl-x_S/Bcl-x_L et BCL-2_α/BCL-2_β (Akgul *et al.*, 2004). De plus, un essai de levure à double hybride avec MCL-1_S indique qu'il n'interagit qu'avec MCL-1_L et non avec plusieurs partenaires de la famille Bcl, indiquant qu'il s'agit probablement d'un

antagonisme de type dominant-négatif (Bae *et al.*, 2000). Dans ce cas, une répression effective de MCL-1_S devrait générer un phénotype d'apoptose similaire à une répression de MCL-1_L. Malheureusement, les séquences particulières de la jonction unique des isoformes courts ne nous ont pas permis d'évaluer le phénotype dû à leur répression. Notre stratégie permet néanmoins de prédire la relation fonctionnelle entre deux isoformes de l'AS.

Le gène FN1 produit la fibronectine 1, une glycoprotéine de la matrice extracellulaire importante pour l'adhésion cellulaire. Dans le cas particulier de ces isoformes incluant l'exon EDB, bien peu de rôles lui ont été clairement attribués, malgré plus de 20 années de recherche (White *et al.*, 2008). Au cours de notre travail, nous avons évalué des siRNAs longs ciblant l'exon EDB, mais aussi ciblant l'exon EDA et la jonction exon 267 – intron 93 de la région alternative IIICS. Seuls ceux spécifiques à l'exon EDB permettent d'induire l'apoptose des cellules SKOV3ip1, permettant d'identifier un rôle fonctionnel pour l'isoforme EDB. À l'inverse, ce sont les siRNAs longs ciblant la région alternative IIICS qui atténuent la migration des cellules SKOV3ip1. De façon importante, ces phénotypes ne peuvent pas être reproduits par une simple répression globale (résultats non-montrés).

Dans le cas de l'exon 69 de SYK, celui-ci code pour un domaine de liaison entre deux domaines SH2 (Src homology 2). Il a été proposé que celui-ci encode un signal de localisation nucléaire qui permet la redistribution intracellulaire de la kinase SYK (Wang *et al.*, 2003a). Effectivement, seul le TOSS ou les siRNAs longs parviennent à relocaliser la protéine SYK vers le cytoplasme et affecte la survie cellulaire, qui est en partie médiée par la voie de signalisation de JUN. Par contre, lorsqu'il est question de croissance indépendante de l'adhésion, tant les siRNAs longs que les siRNAs globaux sont en mesure de perturber la croissance des cellules dans l'agar mou. Dans ce dernier contexte, la théorie prédit un comportement similaire pour un siRNA court si les deux isoformes ont des propriétés similaires (modèle compensatoire), mais la séquence particulière de la jonction spécifique à l'isoforme court ne se prête pas au dessin de siRNA.

Afin de démontrer clairement l'impact de l'AS sur la fonction, nous avons procédé à l'évaluation de l'induction d'apoptose par un 2^o criblage systématique de siRNAs globaux dans la même lignée cellulaire et sur la même série de gènes décrite au chapitre 3. Cette

fois-ci, c'est la répression globale des gènes BMP4 et SYNE2 qui sensibilise les cellules à l'apoptose et non les siRNA longs (résultats non-montrés). Bien que nous n'ayons pas évalué les effets de siRNAs courts, ceux-ci doivent potentiellement révéler le phénotype décrit par les siRNAs globaux. Ceci expliquerait pourquoi le TOSS n'induit pas le phénotype soutenu par les siRNAs globaux, puisque l'on s'attend à la même absence d'effet qu'avec les siRNAs longs. Il s'agirait donc de cas monofonctionnels dont l'isoforme court détient la fonction. Par contre, ceci est contre intuitif pour BMP4 puisque l'isoforme court constitue moins de 10% de l'ensemble des transcrits. De plus, l'ASE est situé en 5'UTR, donnant lieu à des protéines identiques. Il est possible que ce gène possède d'autres isoformes qui ne sont pas annotées à l'heure actuelle, ce qui expliquerait les ambiguïtés présentées. Ces deux ASEs potentiels exigent donc une validation et une attention supplémentaire avant de conclure. Reste que l'évaluation systématique des gènes possédant un phénotype dû à la répression globale (siRNAs globaux) ou isoforme-spécifique (siRNA long) ne présente pas de chevauchement.

Comme un changement de ratio des isoformes de SYK, FN1 et MCL1 a initialement été observé entre des ovaires normaux et d'autres cancéreux (Klinck *et al.*, 2008), il est tentant d'impliquer ces isoformes dans la résistance à l'apoptose des cellules cancéreuses de l'ovaire *in vivo*. Par contre, seul le ciblage de MCL-1_L permet la reprogrammation vers un ratio « normal ». En effet, on note une diminution de la valeur de ψ pour l'exon 69 de SYK et EDB de FN1 dans les tumeurs ovariennes. Conséquemment, c'est donc une augmentation du ψ (siRNA court ou TOES) qui devrait sensibiliser les cellules à l'apoptose et non l'inverse. Par contre, l'ovaire normal comme source de normalisation n'est pas un choix judicieux pour les études du cancer de l'ovaire séreux de haut grade pour deux raisons. Premièrement, il est constitué presque exclusivement de cellules stromales alors que les tumeurs sont d'origine épithéliale. Comme l'AS est un processus qui est propre à chaque type cellulaire, il est alors normal de constater des différences d'expression d'isoformes entre ces deux types cellulaires (sans nécessairement être associé au cancer). Deuxièmement, les cellules épithéliales en surface de l'ovaire ne seraient pas les cellules d'origine de la plupart des cas de cancer séreux de haut grade; ce serait plutôt les cellules épithéliales sécrétrices de la trompe de Fallope (Colgan *et al.*, 2001; Medeiros *et al.*, 2006;

Callahan *et al.*, 2007; Kim *et al.*, 2012). Ce sont en fait ces motifs qui nous ont poussés à utiliser la technologie LCM afin de comparer directement les cellules épithéliales entre elles (chapitre 4). Des trois ASEs discutés, seul l'exon 69 de SYK a été criblé dans les tissus LCM du chapitre 4 et n'est ni associé aux cellules cancéreuses épithéliales ni au microenvironnement de la tumeur. De plus, l'expression du facteur d'épissage hnRNPK, sensé réguler l'ensemble de ces trois ASEs, n'est pas affectée lorsque l'on normalise l'expression avec les cellules normales épithéliales de la trompe de Fallope (résultat non-montré). L'extrapolation des résultats *in vitro* présentés avec SYK demeure donc limitée.

3.2 Cibles potentielles pour le TOSS

Le TOSS est un excellent outil pour mettre en valeur l'impact fonctionnel de deux isoformes à l'égard de leur ratio. Sur les 41 ASEs criblés, nous avons été en mesure d'en identifier un seul ASE (SYK) pour lequel le ratio des isoformes (TOSS) est critique pour les phénotypes d'apoptose et de viabilité. En revanche, la répression spécifique d'un seul isoforme (par siRNA long) est nécessaire et suffisante pour induire l'apoptose (MCL1 et FN1) ou diminuer la viabilité (CCNE1 and LIG4) de plusieurs cas. Le fait que peu de ASEs produisent un phénotype par le TOSS peut être dû au choix du type d'essai phénotypique ou à la classe de gène ciblé. Comme le criblage par siRNAs longs génère 4 fois plus de ASEs associés à un phénotype que le criblage par TOSS, ceci suggère que le phénotype d'apoptose est moins dépendant du ratio d'isoforme est beaucoup plus dépendant de l'expression d'un isoforme en particulier (le long), du moins pour ce sous-ensemble de gènes. Comme les résultats présentés au chapitre 4 suggèrent que les ASEs criblés sont spécifiques au type cellulaire plutôt que spécifiques aux cellules cancéreuses comme tel, il serait intéressant d'exécuter le même type de criblage, mais pour un ensemble de ASEs vraiment associé au cancer (CES et CSS ASEs du chapitre 4). Est-ce que l'association au cancer augmente les chances de produire un phénotype *in vitro*? Si oui, est-ce plus probable par l'action d'un TOSS ou d'un siRNA long? Si la majorité des phénotypes sont produits par le TOSS, cela supporterait notre hypothèse voulant qu'il existe un programme d'épissage dont l'impact fonctionnel est indépendant de la transcription. Dans la négative,

il est tout à fait plausible que le contexte (essai phénotypique) ne soit pas approprié pour valoriser le TOSS à ca juste mesure.

3.3 Impact fonctionnel du programme RBFOX2 et QKI dans la cancérogenèse

Des analyses bioinformatiques prédisent des changements de patron d'AS dans certains types de cancers et sont confirmés par des validations expérimentales sur un petit nombre de tissus dans certain cas. Il n'est pas clair si ces changements de patron d'AS proviennent d'un effet secondaire des cellules cancéreuses (conséquence) ou plutôt le résultat d'un programme destiné à promouvoir le développement de la tumeur (causal). Afin de supporter notre hypothèse générale voulant que les isoformes de l'AS soient importantes pour le développement des tumeurs, nous avons identifié un certains nombres de ASEs dérégulés dans le cancer de l'ovaire séreux de haut grade. Puis, nous avons déterminé les facteurs d'épissages qui sont altérés et finalement, nous avons associé le changement de ratio d'isoforme observé et l'expression globale des facteurs d'épissage dérégulé par siRNA *in vitro*. Cette démarche nous a conduite à l'identification d'une paire de facteurs d'épissage alternatif (RBFOX2 et QKI) impliquées dans la régulation d'exon mésenchymaux dans le microenvironnement des tumeurs ovariennes.

Cette stratégie pourrait être appliquée afin de déterminer un programme d'AS responsable de l'induction d'apoptose des cellules cancéreuses ovariennes. En partant des facteurs d'épissage associés aux cellules épithéliales cancéreuses de l'ovaire du chapitre 4, la surexpression individuelle de ceux-ci dans une lignée ovarienne cancéreuse permettrait d'identifier les joueurs responsables de la résistance à l'apoptose des cellules cancéreuses. Alternativement, le criblage pourrait être fait en évaluant la résistance accrue dans une lignée épithéliale normale lors de la répression de facteurs d'épissage par siRNA. L'étape suivante serait de vérifier l'impact sur l'AS de gènes apoptotiques par PCR à haut débit ou sur le transcriptome entier par séquençage à haut débit. Ainsi, le suivi de l'impact sur l'AS en parallèle avec des marqueurs de l'apoptose permettrait d'identifier un programme d'AS qui promouvoit la résistance à l'apoptose des cellules cancéreuses.

Tel que relaté à la section 2.2.2.2 de l'introduction, RBFOX2 est un facteur d'épissage qui est exprimé de façon tissu-spécifique et qui influence le ratio d'isoformes de plusieurs ASEs. Son expression est importante afin de coordonner le développement normal du système nerveux (Gehman *et al.*, 2012), du système cardiaque ainsi que des muscles squelettiques (Gallagher *et al.*, 2011). En combinaison avec d'autres facteurs, il est soupçonné d'être impliqué dans des programmes d'AS promouvant l'identité des cellules musculaires squelettiques (Llorian and Smith, 2011) et neuronale (Zhang *et al.*, 2010).

Une mutation autosomale récessive affectant l'expression de QKI originalement découverte par Sidman et collaborateurs (1964) produit des tremblements caractéristiques chez la souris dus à une myéline défectueuse (qk^v). La délétion du gène QKI est létale chez la souris, ne permettant pas aux embryons de se développer plus de 10 jours *in utero* (Li *et al.*, 2003). Le gène QKI subit de l'AS dans sa portion 3' pour générer 3 isoformes ARNm de 5, 6 et 7 kDa d'où leur nom: QKI-5, QKI-6 et QKI-7. QKI-5 est le seul isoforme à posséder un signal de localisation nucléaire (Wu *et al.*, 1999). La protéine $qki-7$ est exclusivement cytoplasmique alors que QKI-6 peut se retrouver tant au cytoplasme qu'au noyau, un résultat qui pourrait s'expliquer par une hétérodimérisation avec la protéine nucléaire QKI-5 (Larocque *et al.*, 2002). Ces trois isoformes ont un domaine de liaison à l'ARN de type KH et font partie d'une famille de protéines qui regroupe entre autres hnRNPK et Sam68 (Chenard and Richard, 2008). Ils ont un rôle à jouer dans la maturation des ARNm. Plus précisément, les isoformes cytoplasmiques QKI-6 et QKI-7 sont impliquées dans la stabilisation ou la traduction des ARNm de gènes impliqués dans la différenciation cellulaire via la liaison à un motif ACUAAY en 3'UTR (Larocque *et al.*, 2005). En revanche, l'isoforme nucléaire QKI-5 est présumé responsable de la régulation directe de l'AS de gènes impliqués dans le maintien de la gaine de myéline comme la glycoprotéine MAG (myelin-associated glycoprotein) (Wu *et al.*, 2002). À noter que ce sont les isoformes QKI-6 et QKI-7 qui sont réprimées dans les oligodendrocytes du modèle murin qk^v . En ce sens, il n'est pas exclu que QKI-6 ou QKI-7 pourrait aussi agir indirectement sur les ARNm de certains facteurs d'épissage comme hnRNPA1 (Zhao *et al.*, 2010, Zearfoss *et al.*, 2011). D'autre part, la localisation nucléaire partielle de QKI-6 et ses effets potentiellement directs sur l'AS pourraient expliquer le changement de ratio des isoformes favorisant

l'isoforme QKI-6 au détriment de l'isoforme QKI-7 dans les échantillons de sein de la figure 6 du chapitre 4.

Bien que RBFOX2 et QKI semblent avoir des rôles distincts, le chapitre 4 met en lumière un ensemble d'ASEs régulé de façon redondante par ces deux facteurs d'épissage. En comparant le profil d'AS et l'expression de facteurs d'épissage de tissus épithéliaux et mésenchymaux normaux et cancéreux, nous avons découvert que les cellules du microenvironnement de tumeurs de l'ovaire possèdent un profil qui ressemble beaucoup aux cellules épithéliales normales. Tel que discuté dans le chapitre 4, nos résultats indiquent qu'il ne s'agit vraisemblablement pas d'une transition mésenchymale-à-épithéliale (MET) (Fig. 4, voie 1). En fait, il n'est pas clair en quoi l'acquisition de propriétés épithéliales favoriserait directement ou indirectement la progression de la tumeur épithéliale. Il est possible que la cancérogenèse induise la différenciation épithéliale des cellules avoisinant la tumeur afin de lui ressembler, mais que la cascade de signalisation est interrompue prématurément, d'où le maintien du phénotype stromal. En ce sens, il serait intéressant d'évaluer la présence du programme RQ dans un système où des cellules précurseuses se différencient en cellules épithéliales.

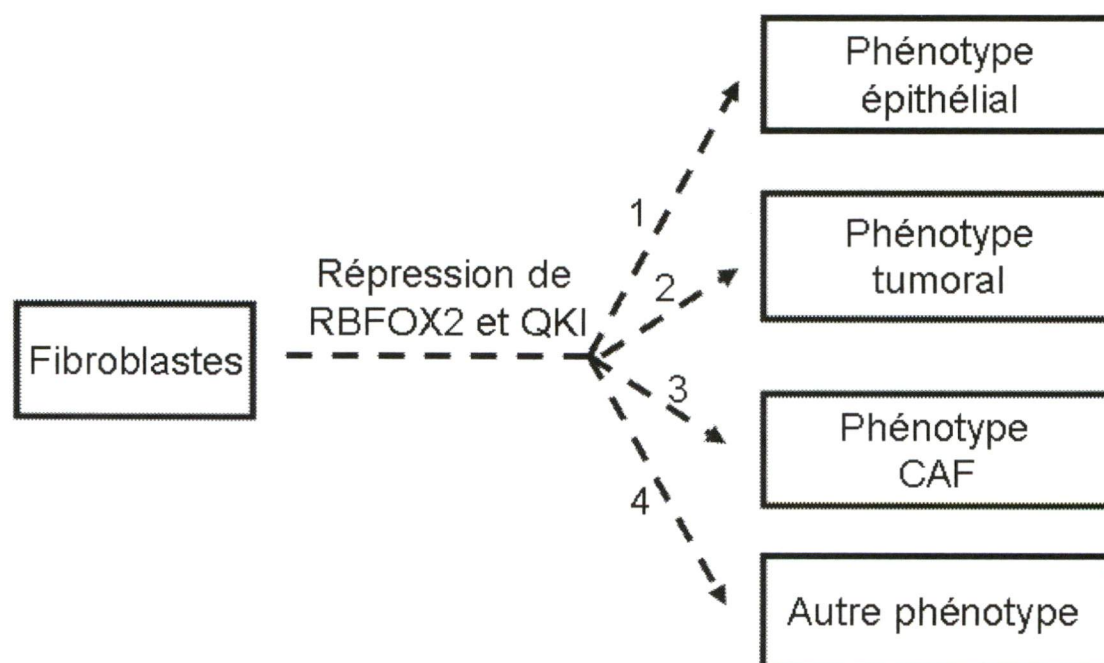


Figure 4 Impact du programme RBFOX2 et QKI sur les fibroblastes normaux.

Schématisation hypothétique de l'impact de la désactivation du programme RBFOX2 et QKI sur le phénotype des fibroblastes. La répression de RBFOX2 et QKI pourrait promouvoir un phénotype 1) épithélial; 2) tumoral; 3) de type CAF ou 4) autre dans le but de promouvoir le développement de la tumeur.

Il est possible que la répression de RBFOX2 et QKI induise une transformation des fibroblastes au point de leur faire acquérir des propriétés tumorigéniques (Fig. 4, voie 2). De façon intéressante, une surexpression du facteur d'épissage SF2 (Karni *et al.*, 2007) ou SRp20 (Jia *et al.*, 2010) dans des fibroblastes immortalisés de rat est suffisante pour permettre la formation de sarcome lorsqu'injecté, dans des souris immuno-déficientes, ce qui les placent au rang de proto-oncogène. Curieusement, un modèle murin du cancer de l'ovaire originant de la trompe de Fallope indique que le carcinome prend place dans le stroma de la trompe, suggérant que les cellules d'origine seraient mésenchymales (Kim *et al.*, 2012). Par contre, QKI est suspecté comme gène suppresseur de tumeur dans les cancers du système gastro-intestinal (Yang *et al.*, 2010, Bian *et al.*, 2012). Par définition, la suppression du gène favorise la formation de tumeur, mais requiert la présence d'oncogène. En effet, la suppression de TP53 ou PTEN dans les fibroblastes du microenvironnement accélère la formation de tumeur épithéliale adjacente, mais ne parvient pas à transformer les fibroblastes en adénocarcinome ou en sarcome (Kiaris *et al.*, 2005, Trimboli *et al.*, 2009). Malgré tout, il serait intéressant de déterminer l'impact du programme RQ sur la tumorigénicité de fibroblastes normaux par des essais de croissance en agar mou et par injection sous-cutanée dans un modèle murin immuno-déficient suite à une répression de RBFOX2 et QKI.

Si la répression de RBFOX2 et QKI ne permet pas aux fibroblastes d'acquérir des propriétés tumorigéniques, alors elle pourrait permettre leur différenciation en cancer-associated fibroblasts (CAFs) (Fig. 4, voie 3). Ceux-ci sont reconnus pour sécréter des facteurs qui influencent positivement la prolifération et l'invasion, et contribuent à la résistance à l'apoptose de la tumeur (Shimoda *et al.*, 2010). Tel que décrit à la section 1.2.7, les CAFs se distinguent par l'expression de marqueurs et de phénotype spécifique. Il serait intéressant de vérifier la prolifération, l'invasion et la résistance à l'apoptose de cellules cancéreuses épithéliales suite à la désactivation du programme RQ de fibroblastes

en co-culture. D'une façon ou d'une autre, le programme RQ doit contribuer à la progression de la tumeur d'une quelconque façon pour justifier son existence (Fig. 4, voie 4).

4 Potentiel clinique des signatures moléculaires obtenues par microdissection

En théorie, les signatures moléculaires issues de biopsies peuvent également permettre de révéler le pronostic d'une patiente. Par exemple, l'analyse non-supervisée de 285 tumeurs ovariennes a permis une sous-classification moléculaire de quatre types séreux de haut grade dont un qui corrèle avec la sévérité du phénotype clinique. En effet, l'expression de gènes habituellement exprimés dans les cellules stromales caractérise une sous-population de tumeurs séreuses de haut grade et est associée à une durée de survie plus courte que les autres sous-types moléculaires (Tothill *et al.*, 2008). Fait intéressant, le faible contenu épithélial des tumeurs de ce sous-groupe corrèle avec la forte expression de ces gènes stromaux dans le microenvironnement de la tumeur, révélant que cette signature ne provient pas des cellules épithéliales cancéreuses mais bien des cellules stromales. Par contre, l'association clinique n'a été corroborée que partiellement puisqu'une étude ultérieure de « The Cancer Genome Atlas » n'a pas été en mesure d'associer un de ces sous-groupes au degré de survie (Bell D *et al.*, 2011, Bentink *et al.*, 2012). Ces études, ainsi que d'autres dans le cancer du sein (Finak *et al.*, 2008, Farmer *et al.*, 2009) suggèrent que les signatures stromales peuvent être utilisées à des fins pronostiques.

En date du dépôt de cette thèse, aucune étude de profilage des isoformes épissées alternativement dans le cancer de l'ovaire n'a été réalisée par un autre groupe de recherche que le nôtre. Il est donc impossible d'analyser rétrospectivement une étude antérieure de grande envergure telle qu'il est usuel de le faire pour l'expression globale. Le pronostic de la patiente et la durée de sa survie suivant son diagnostic sont intimement liés à la sensibilité de sa tumeur à la chimiothérapie de première ligne et à l'efficacité de la cytoréduction. La plupart des tumeurs ovariennes de haut grade sont initialement sensibles, mais finissent par progresser au cours des deux années suivantes. Le temps pris pour récidiver est une bonne mesure de la survie globale de la patiente et en fait un bon

marqueur pronostique. Malheureusement, les tissus utilisés au cours du chapitre 4, pour lesquels le suivi complet de la patiente y est associé, sont très limités et ne permettent pas de procéder à des analyses de corrélation avec les données cliniques.

CONCLUSIONS

L'épissage alternatif (AS) est un mécanisme par lequel un pré-ARNm produit plusieurs isoformes ARNm. Bien que l'AS influence plus de 90% des gènes, la fonction associée à chacune des isoformes est difficile à évaluer. Ceci est en partie dû au défi que représente la quantification des isoformes. Dans cette thèse, nous avons élaboré une méthodologie à haut débit permettant une quantification isoforme-spécifique ou globale (toutes les isoformes) dans des conditions PCR universelle. Nous avons illustré le potentiel de cette approche en l'appliquant à deux problématiques importantes de la biologie actuelle. Dans un premier temps, il nous a été possible de systématiquement disséquer la contribution de chacun des isoformes de gènes impliqués dans le cancer de l'ovaire sur le phénotype de cellules cancéreuses. Pour ce faire, nous avons d'abord développé une méthodologie générale permettant de caractériser des outils moléculaires (siRNA, TOSS) à haut débit. Le TOSS (targeted oligonucleotide silencer of splicing) est un antisens bifonctionnel qui favorise l'exclusion d'un exon alternatif. Nous avons développé des règles de dessin et généralisé son utilisation en démontrant par deux techniques (endpoint PCR et PCR en temps réel) que le TOSS reprogramme l'AS en favorisant l'isoforme court; et qu'il n'affecte pas l'expression globale du gène dans la plupart des cas. En s'appuyant sur cette méthodologie, nous avons clairement démontré qu'il est possible d'induire l'apoptose des cellules cancéreuses SKOV3ip1 en ciblant spécifiquement une seule des isoformes du gène SYK et que ce phénotype est indépendant de l'altération de son expression globale. Dans un deuxième temps, il a été possible de caractériser la dérégulation de l'AS dans les cellules cancéreuses de l'ovaire séreux de haut grade et leur microenvironnement. Pour ce faire, nous avons utilisé un appareil de microdissection au laser pour isoler les cellules épithéliales des cellules stromales de trompe de Fallope et d'ovaire. De cette façon, nous avons été en mesure de quantifier l'expression globale de 370 gènes associés au spliceosome ainsi que l'expression de l'ensemble des isoformes répertoriés dans NCBI RefSeq 36.3 (870 ASEs). En combinant les résultats de criblage des tissus et d'approche par répression transitoire (siRNA), nous avons identifié un programme d'AS régulé par RBFOX2 et QKI. En comparant les signatures de profil d'AS et de facteurs d'épissage dans

des tissus épithéliaux et mésenchymaux normaux, nous avons découvert que le microenvironnement des tumeurs ovariennes possède un profil moléculaire épithélial. Bien que RBFOX2 et QKI aient leurs propres cibles ASEs et fonctions, leur cibles communes sont presque exclusivement des exons mésenchymaux inclus dans les tissus normaux. Comme RBFOX2 et QKI sont réprimés dans le microenvironnement de tumeur ovarienne, ceci explique pourquoi ce sont majoritairement des exons mésenchymaux qui soient exclus. Ce programme est également désactivé dans d'autres adénocarcinomes (sein), ce qui laisse croire que ce programme est général et important pour le développement des tumeurs épithéliales.

REMERCIEMENTS

Je suis reconnaissant envers mes superviseurs Jean-Pierre Perreault et Sherif Abou Elela, ainsi que mon jury de thèse composé d'Anne-Marie Mes-Masson, Nathalie Rivard et Rodney Ouellette. Merci pour votre précieux temps passé à lire cette thèse afin d'en faire non seulement un meilleur ouvrage, mais également faire de moi un auteur scientifique plus accompli.

Sur le plan professionnel, trois personnes m'ont apporté une aide inestimable dans trois circonstances différentes; sans quoi le travail présenté dans cette thèse serait franchement différent. La première personne m'a transmis sa fougue et ses connaissances techniques nécessaires à la réalisation de l'article 1. Son passage dans notre laboratoire a transformé nos façons de faire et nous a doté d'une solide structure de travail dont on se sert encore et dont on se servira longtemps, et qui sous-tendent tous les chapitres de cette thèse. Merci Karine Tremblay.

La deuxième personne est aussi fougueuse que la première. Notre collaboration de tous les jours s'est concrétisée par 1) le développement des outils moléculaires isoformes-spécifiques, 2) la réalisation d'un premier article dans RNA, 3) une amitié. Ensemble, nous avons mis à notre disposition des moyens à la hauteur de nos ambitions pour bâtir nos rêves. Merci Jean-Francois Lucier.

La troisième personne a apportée la crédibilité nécessaire à la réalisation du chapitre 4. L'embauche d'un pathologiste senior demeure un des bons coups de Sherif à ce jour. Cela nous a permis de faire le ménage dans notre banque de tissus et conséquemment, nous a dicté les études réalisables, mais surtout celles qui ne sont pas possibles. Merci Hanad Nwilati.

J'aimerais remercier les artisans du LGFUS qui sont dans l'ombre, notamment Roscoe Klinck, celui qui a bâti la plateforme à partir de rien. Son équipe actuelle (Philippe Thibault, Elvy Lapointe et Mathieu Durand) a non seulement vivement participé à la

réalisation de certaines portions de la thèse, mais plus important encore, je peux les considérer comme des amis. Plusieurs membres passés et présents du laboratoire ont su forger la personnalité du LGFUS et faire avancer les projets à leur façon. Je pense ici à ceux que j'ai connu (Anne et Lyna, le vénérable du sommet, Takis, Dan Garneau, Geneviève, Marianne, Junior, Bruno). Bonne chance à Daniel Gendron sur les épaules de qui repose maintenant la pression de faire rayonner ce projet.

Sainte-Julie a le droit à un paragraphe à part. Bien qu'elle ne travaille pas sur le « vrai projet » du LGFUS, je lui dois tout mon respect pour son dévouement envers la science, et sa touche magique qui lui permet d'obtenir ce qu'elle veut de Sherif tout en lui donnant l'impression que c'était l'idée de Sherif. Elle a aussi eu la générosité, tout comme Rachel, Guylaine, Laurence et Brigitte, de se taper la correction de mes fautes d'orthographe, nombreuses et répétitives.

Ma famille m'a été d'un grand soutien tout au long de mes 13 ans passés à Sherbrooke, particulièrement dans la dernière année là où ma sœur (Marie-Bri) et son conjoint (Francois-Xavier) m'ont hébergé pour l'été, avant de passer la relève à mes parents (Jacques et Brigitte).

Merci pour tout Amélie, Mariana, Guylaine.

LISTE DES PUBLICATIONS

- Aartsma-Rus, A, Bremmer-Bout, M, Janson, AA, den Dunnen, JT, van Ommen, GJ & van Deutekom, JC (2002). Targeted exon skipping as a potential gene correction therapy for Duchenne muscular dystrophy. *Neuromuscul Disord* **12 Suppl 1**: S71-7.
- Aartsma-Rus, A, De Winter, CL, Janson, AA, Kaman, WE, Van Ommen, GJ, Den Dunnen, JT, *et al.* (2005). Functional analysis of 114 exon-internal AONs for targeted DMD exon skipping: indication for steric hindrance of SR protein binding sites. *Oligonucleotides* **15**: 284-97.
- Abrahams, PJ, Houweling, A, Cornelissen-Steijger, PD, Jaspers, NG, Darroudi, F, Meijers, CM, *et al.* (1998). Impaired DNA repair capacity in skin fibroblasts from various hereditary cancer-prone syndromes. *Mutat Res* **407**: 189-201.
- Ahmed, AA, Etemadmoghadam, D, Temple, J, Lynch, AG, Riad, M, Sharma, R, *et al.* (2010). Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. *J Pathol* **221**: 49-56.
- Akgul, C, Moulding, DA & Edwards, SW (2004). Alternative splicing of Bcl-2-related genes: functional consequences and potential therapeutic applications. *Cell Mol Life Sci* **61**: 2189-99.
- Amor, S, Remy, S, Dambrine, G, Le Vern, Y, Rasschaert, D & Laurent, S (2010). Alternative splicing and nonsense-mediated decay regulate telomerase reverse transcriptase (TERT) expression during virus-induced lymphomagenesis in vivo. *BMC Cancer* **10**: 571.
- Antonov, J, Goldstein, DR, Oberli, A, Baltzer, A, Pirotta, M, Fleischmann, A, *et al.* (2005). Reliable gene expression measurements from degraded RNA by quantitative real-time PCR depend on short amplicons and a proper normalization. *Lab Invest* **85**: 1040-50.
- Aono T, Takarada Y, Segawa M & Yoshiga S (2000). *Patent Japan 101544370; PCT/WO/JP/00/08657.*
- Armitage, P, Doll, R (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer* **8**: 1-12.
- Artandi, SE, DePinho, RA (2010). Telomeres and telomerase in cancer. *Carcinogenesis* **31**: 9-18.
- Auersperg, N, Wong, AS, Choi, KC, Kang, SK & Leung, PC (2001). Ovarian surface epithelium: biology, endocrinology, and pathology. *Endocr Rev* **22**: 255-88.
- Auweter, SD, Fasan, R, Reymond, L, Underwood, JG, Black, DL, Pitsch, S, *et al.* (2006). Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J* **25**: 163-73.
- Bae, J, Leo, CP, Hsu, SY & Hsueh, AJ (2000). MCL-1S, a splicing variant of the antiapoptotic BCL-2 family member MCL-1, encodes a proapoptotic protein possessing only the BH3 domain. *J Biol Chem* **275**: 25255-61.
- Barbier, J, Dutertre, M, Bittencourt, D, Sanchez, G, Gratadou, L, de la Grange, P, *et al.* (2007). Regulation of H-ras splice variant expression by cross talk between the p53 and nonsense-mediated mRNA decay pathways. *Mol Cell Biol* **27**: 7315-33.
- Bast, RC, Jr, Hennessy, B & Mills, GB (2009). The biology of ovarian cancer: new opportunities for translation. *Nat Rev Cancer* **9**: 415-28.
- Bell D, Berchuck A, Birrer M, Chien J, Cramer DW & Dao F, (2011). Integrated genomic analyses of ovarian carcinoma. *Nature* **474**: 609-15.

- Bennett, CF, Chiang, MY, Chan, H, Shoemaker, JE & Mirabelli, CK (1992). Cationic lipids enhance cellular uptake and activity of phosphorothioate antisense oligonucleotides. *Mol Pharmacol* **41**: 1023-33.
- Bentink, S, Haibe-Kains, B, Risch, T, Fan, JB, Hirsch, MS, Holton, K, *et al.* (2012). Angiogenic mRNA and microRNA gene expression signature predicts a novel subtype of serous ovarian cancer. *PLoS One* **7**: e30269.
- Bian, Y, Wang, L, Lu, H, Yang, G, Zhang, Z, Fu, H, *et al.* (2012). Downregulation of tumor suppressor QKI in gastric cancer and its implication in cancer prognosis. *Biochem Biophys Res Commun* **422**: 187-93.
- Birmingham, A, Anderson, E, Sullivan, K, Reynolds, A, Boese, Q, Leake, D, *et al.* (2007). A protocol for designing siRNAs with high functionality and specificity. *Nat Protoc* **2**: 2068-78.
- Blanchette, M, Chabot, B (1999). Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J* **18**: 1939-52.
- Blaustein, M, Pelisch, F, Tanos, T, Munoz, MJ, Wengier, D, Quadrana, L, *et al.* (2005). Concerted regulation of nuclear and cytoplasmic activities of SR proteins by AKT. *Nat Struct Mol Biol* **12**: 1037-44.
- Boise, LH, Gonzalez-Garcia, M, Postema, CE, Ding, L, Lindsten, T, Turka, LA, *et al.* (1993). Bcl-X, a Bcl-2-Related Gene that Functions as a Dominant Regulator of Apoptotic Cell Death. *Cell* **74**: 597-608.
- Boiziau C, Tarrago-Litvak L, Sinha ND, Moreau S, Litvak S, Toulmé (1996). Antisense oligonucleotides inhibit in vitro cDNA synthesis by HIV-1 reverse transcriptase. *Antisense Nucleic Acid Drug Development* **6**:103-9.
- Brinkman, BM (2004). Splice variants as cancer biomarkers. *Clin Biochem* **37**: 584-94.
- Bruno, IG, Jin, W & Cote, GJ (2004). Correction of aberrant FGFR1 alternative RNA splicing through targeting of intronic regulatory elements. *Hum Mol Genet* **13**: 2409-20.
- Burd, CG, Dreyfuss, G (1994). RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J* **13**: 1197-204.
- Bustin, SA (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* **25**: 169-93.
- Byun, HJ, Hong, IK, Kim, E, Jin, YJ, Jeoung, DI, Hahn, JH, *et al.* (2006). A splice variant of CD99 increases motility and MMP-9 expression of human breast cancer cells through the AKT-, ERK-, and JNK-dependent AP-1 activation signaling pathways. *J Biol Chem* **281**: 34833-47.
- Caceres, JF, Stamm, S, Helfman, DM & Krainer, AR (1994). Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science* **265**: 1706-9.
- Calabrese, P, Tavaré, S & Shibata, D (2004). Pretumor progression: clonal evolution of human stem cell populations. *Am J Pathol* **164**: 1337-46.
- Callahan, MJ, Crum, CP, Medeiros, F, Kindelberger, DW, Elvin, JA, Garber, JE, *et al.* (2007). Primary fallopian tube malignancies in BRCA-positive women undergoing surgery for ovarian cancer risk reduction. *J Clin Oncol* **25**: 3985-90.
- Camats, M, Kokolo, M, Heesom, KJ, Ladomery, M & Bach-Elias, M (2009). P19 H-ras induces G1/S phase delay maintaining cells in a reversible quiescence state. *PLoS One* **4**: e8513.
- Cartegni, L, Krainer, AR (2003). Correction of disease-associated exon skipping by synthetic exon-specific activators. *Nat Struct Biol* **10**: 120-5.

- Castle, JC, Zhang, C, Shah, JK, Kulkarni, AV, Kalsotra, A, Cooper, TA, *et al.* (2008). Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat Genet* **40**: 1416-25.
- Catena, R, Muniz-Medina, V, Moralejo, B, Javierre, B, Best, CJ, Emmert-Buck, MR, *et al.* (2007). Increased expression of VEGF121/VEGF165-189 ratio results in a significant enhancement of human prostate tumor angiogenesis. *Int J Cancer* **120**: 2096-109.
- Cecchi, F, Rabe, DC & Bottaro, DP (2010). Targeting the HGF/Met signalling pathway in cancer. *Eur J Cancer* **46**: 1260-70.
- Charlet-B, N, Logan, P, Singh, G & Cooper, TA (2002). Dynamic antagonism between ETR-3 and PTB regulates cell type-specific alternative splicing. *Mol Cell* **9**: 649-58.
- Chen, M, Manley, JL (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10**: 741-54.
- Chenard, CA, Richard, S (2008). New implications for the QUAKING RNA binding protein in human disease. *J Neurosci Res* **86**: 233-42.
- Cheng, C, Sharp, PA (2006). Regulation of CD44 alternative splicing by SRm160 and its potential role in tumor cell invasion. *Mol Cell Biol* **26**: 362-70.
- Chi, DS, Schwartz, PE (2008). Cytoreduction vs. neoadjuvant chemotherapy for ovarian cancer. *Gynecol Oncol* **111**: 391-9.
- Chung, JY, Braunschweig, T, Williams, R, Guerrero, N, Hoffmann, KM, Kwon, M, *et al.* (2008). Factors in tissue handling and processing that impact RNA obtained from formalin-fixed, paraffin-embedded tissue. *J Histochem Cytochem* **56**: 1033-42.
- Cirri, P, Chiarugi P (2011). Cancer-associated fibroblasts: the dark side of the coin. *Am J Can Res* **1**: 482-97.
- Clarke, MF, Apel, IJ, Benedict, MA, Eipers, PG, Sumantran, V, Gonzalez-Garcia, M, *et al.* (1995). A recombinant bcl-x s adenovirus selectively induces apoptosis in cancer cells but not in normal bone marrow cells. *Proc Natl Acad Sci U S A* **92**: 11024-8.
- Clayburgh, DR, Rosen, S, Witkowski, ED, Wang, F, Blair, S, Dudek, S, *et al.* (2004). A differentiation-dependent splice variant of myosin light chain kinase, MLCK1, regulates epithelial tight junction permeability. *J Biol Chem* **279**: 55506-13.
- Cohen, JB, Broz, SD & Levinson, AD (1989). Expression of the H-ras proto-oncogene is controlled by alternative splicing. *Cell* **58**: 461-72.
- Cohen, JB, Levinson, AD (1988). A point mutation in the last intron responsible for increased expression and transforming activity of the c-Ha-ras oncogene. *Nature* **334**: 119-24.
- Colgan TJ, Murphy J, Cole DE, Narod S & Rosen B. (2001). Occult carcinoma in prophylactic oophorectomy specimens: prevalence and association with BRCA germline mutation status. *Am J Surg Pathol* **25**:1283-1289.
- Costa, V, Angelini, C, De Feis, I & Ciccodicola, A (2010). Uncovering the complexity of transcriptomes with RNA-Seq. *J Biomed Biotechnol* **2010**: 853916.
- Crnkovic-Mertens, I, Semzow, J, Hoppe-Seyler, F & Butz, K (2006). Isoform-specific silencing of the Livin gene by RNA interference defines Livin beta as key mediator of apoptosis inhibition in HeLa cells. *J Mol Med* **84**: 232-40.
- Crooke, ST (1999). Molecular mechanisms of action of antisense drugs. *Biochim Biophys Acta* **1489**: 31-44.
- Cuperlovic-Culf, M, Belacel, N, Culf, AS & Ouellette, RJ (2006). Data analysis of alternative splicing microarrays. *Drug Discov Today* **11**: 983-90.

- Dahlgren, C, Zhang, HY, Du, Q, Grahn, M, Norstedt, G, Wahlestedt, C, *et al.* (2008). Analysis of siRNA specificity on targets with double-nucleotide mismatches. *Nucleic Acids Res* **36**: e53.
- Dasdia, T, Di Marco, A, Goffredi, M, Minghetti, A & Necco, A (1979). Ion level and calcium fluxes in HeLa cells after adriamycin treatment. *Pharmacol Res Commun* **11**: 19-29.
- Datta, D, Flaxenburg, JA, Laxmanan, S, Geehan, C, Grimm, M, Waaga-Gasser, AM, *et al.* (2006). Ras-induced modulation of CXCL10 and its receptor splice variant CXCR3-B in MDA-MB-435 and MCF-7 cells: relevance for the development of human breast cancer. *Cancer Res* **66**: 9509-18.
- David, CJ, Chen, M, Assanah, M, Canoll, P & Manley, JL (2010). HnRNP proteins controlled by c-Myc deregulate pyruvate kinase mRNA splicing in cancer. *Nature* **463**: 364-8.
- Davis, JT (2004). G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew Chem Int Ed Engl* **43**: 668-98.
- de Fougerolles, A, Vornlocher, HP, Maraganore, J & Lieberman, J (2007). Interfering with disease: a progress report on siRNA-based therapeutics. *Nat Rev Drug Discov* **6**: 443-53.
- DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**: P3.
- DiFeo, A, Feld, L, Rodriguez, E, Wang, C, Beer, DG, Martignetti, JA, *et al.* (2008). A functional role for KLF6-SV1 in lung adenocarcinoma prognosis and chemotherapy response. *Cancer Res* **68**: 965-70.
- Ding, J, Hayashi, MK, Zhang, Y, Manche, L, Krainer, AR & Xu, RM (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev* **13**: 1102-15.
- Dodier, P, Piche, A (2006). Bcl-X(L) is functionally non-equivalent for the regulation of growth and survival in human ovarian cancer cells. *Gynecol Oncol* **100**: 254-63.
- Doig, T, Monaghan, H (2006). Sampling the omentum in ovarian neoplasia: when one block is enough. *Int J Gynecol Cancer* **16**: 36-40.
- Dominski, Z, Kole, R (1993). Restoration of correct splicing in thalassemic pre-mRNA by antisense oligonucleotides. *Proc Natl Acad Sci U S A* **90**: 8673-7.
- Donis-Keller, H (1979). Site specific enzymatic cleavage of RNA. *Nucleic Acids Res* **7**: 179-92.
- Druker, BJ, Sawyers, CL, Kantarjian, H, Resta, DJ, Reese, SF, Ford, JM, *et al.* (2001). Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *N Engl J Med* **344**: 1038-42.
- Dvorak, HF (1986). Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. *N Engl J Med* **315**: 1650-9.
- Eckstein, F (1966). Nucleoside phosphorothioates. *J Am Chem Soc* **88**: 4292-3.
- Eperon, IC, Makarova, OV, Mayeda, A, Munroe, SH, Caceres, JF, Hayward, DG, *et al.* (2000). Selection of alternative 5' splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol Cell Biol* **20**: 8303-18.
- Eyden B (2009). The myofibroblast, electron microscopy and cancer research. *Int J Cancer* **125**: 1743-5.

- Faa, G, Gerosa, C, Fanni, D, Monga, G, Zaffanello, M, Van Eyken, P, *et al.* (2012). Morphogenesis and molecular mechanisms involved in human kidney development. *J Cell Physiol.* **227**:1257-68.
- Fagnani, M, Barash, Y, Ip, JY, Misquitta, C, Pan, Q, Saltzman, AL, *et al.* (2007). Functional coordination of alternative splicing in the mammalian central nervous system. *Genome Biol* **8**: R108.
- Fairbrother, WG, Yeh, RF, Sharp, PA & Burge, CB (2002). Predictive identification of exonic splicing enhancers in human genes. *Science* **297**: 1007-13.
- Farmer, P, Bonnefoi, H, Anderle, P, Cameron, D, Wirapati, P, Becette, V, *et al.* (2009). A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nat Med* **15**: 68-74.
- Fathalla, MF (1971). Incessant ovulation--a factor in ovarian neoplasia? *Lancet* **2**: 163.
- Fearon, ER, Vogelstein, B (1990). A genetic model for colorectal tumorigenesis. *Cell* **61**: 759-67.
- Finak, G, Bertos, N, Pepin, F, Sadekova, S, Souleimanova, M, Zhao, H, *et al.* (2008). Stromal gene expression predicts clinical outcome in breast cancer. *Nat Med* **14**: 518-27.
- Fletcher, L, Cerniglia, GJ, Yen, TJ & Muschel, RJ (2005). Live cell imaging reveals distinct roles in cell cycle regulation for Nek2A and Nek2B. *Biochim Biophys Acta* **1744**: 89-92.
- Futreal, PA, Coin, L, Marshall, M, Down, T, Hubbard, T, Wooster, R, *et al.* (2004). A census of human cancer genes. *Nat Rev Cancer* **4**: 177-83.
- Gallagher, TL, Arribere, JA, Geurts, PA, Exner, CR, McDonald, KL, Dill, KK, *et al.* (2011). Rbfox-regulated alternative splicing is critical for zebrafish cardiac and skeletal muscle functions. *Dev Biol* **359**: 251-61.
- Gardina, PJ, Clark, TA, Shimada, B, Staples, MK, Yang, Q, Veitch, J, *et al.* (2006). Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* **7**: 325.
- Garneau, D, Revil, T, Fisette, JF & Chabot, B (2005). Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J Biol Chem* **280**: 22641-50.
- Garneau, NL, Wilusz, J & Wilusz, CJ (2007). The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* **8**: 113-26.
- Gatenby, RA, Gillies, RJ (2008). A microenvironmental model of carcinogenesis. *Nat Rev Cancer* **8**: 56-61.
- Gehman, LT, Meera, P, Stoilov, P, Shiue, L, O'Brien, JE, Meisler, MH, *et al.* (2012). The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. *Genes Dev* **26**: 445-60.
- Gendron, D, Carriero, S, Garneau, D, Villemaire, J, Klinck, R, Elela, SA, *et al.* (2006). Modulation of 5' splice site selection using tailed oligonucleotides carrying splicing signals. *BMC Biotechnol* **6**: 5.
- Ghigna, C, De Toledo, M, Bonomi, S, Valacca, C, Gallo, S, Apicella, M, *et al.* (2010). Pro-metastatic splicing of Ron proto-oncogene mRNA can be reversed: therapeutic potential of bifunctional oligonucleotides and indole derivatives. *RNA Biol* **7**: 495-503.
- Ghigna, C, Giordano, S, Shen, H, Benvenuto, F, Castiglioni, F, Comoglio, PM, *et al.* (2005). Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. *Mol Cell* **20**: 881-90.

- Giglio, S, Mancini, F, Gentiletti, F, Sparaco, G, Felicioni, L, Barassi, F, *et al.* (2005). Identification of an aberrantly spliced form of HDMX in human tumors: a new mechanism for HDM2 stabilization. *Cancer Res* **65**: 9687-94.
- Graveley, BR (2000). Sorting out the complexity of SR protein functions. *RNA* **6**: 1197-211.
- Gromak, N, Matlin, AJ, Cooper, TA & Smith, CW (2003). Antagonistic regulation of alpha-actinin alternative splicing by CELF proteins and polypyrimidine tract binding protein. *RNA* **9**: 443-56.
- Gui, JF, Lane, WS & Fu, XD (1994). A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature* **369**: 678-82.
- Haggie, JA, Sellwood, RA, Howell, A, Birch, JM & Schor, SL (1987). Fibroblasts from relatives of patients with hereditary breast cancer show fetal-like behaviour in vitro. *Lancet* **1**: 1455-7.
- Halazonetis, TD, Gorgoulis, VG & Bartek, J (2008). An oncogene-induced DNA damage model for cancer development. *Science* **319**: 1352-5.
- Hampe, J, Wollstein, A, Lu, T, Frevel, HJ, Will, M, Manaster, C, *et al.* (2001). An integrated system for high throughput TaqMan based SNP genotyping. *Bioinformatics* **17**: 654-5.
- Hanahan, D, Weinberg, RA (2011). Hallmarks of cancer: the next generation. *Cell* **144**: 646-74.
- Hanahan, D, Weinberg, RA (2000). The hallmarks of cancer. *Cell* **100**: 57-70.
- Harper, SJ, Bates, DO (2008). VEGF-A splicing: the key to anti-angiogenic therapeutics? *Nat Rev Cancer* **8**: 880-7.
- Hayes, GM, Carrigan, PE & Miller, LJ (2007). Serine-arginine protein kinase 1 overexpression is associated with tumorigenic imbalance in mitogen-activated protein kinase pathways in breast, colonic, and pancreatic carcinomas. *Cancer Res* **67**: 2072-80.
- Heid, CA, Stevens, J, Livak, KJ & Williams, PM (1996). Real time quantitative PCR. *Genome Res* **6**: 986-94.
- Hellemans, J, Mortier, G, De Paepe, A, Speleman, F & Vandesompele, J (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* **8**: R19.
- Hethcote, HW, Knudson, AG, Jr (1978). Model for the incidence of embryonal cancers: application to retinoblastoma. *Proc Natl Acad Sci U S A* **75**: 2453-7.
- Hosoya, T, Monden, T, Fukabori, Y, Hashimoto, K, Satoh, T, Kasai, K, *et al.* (2008). A novel splice variant of the nuclear coactivator p120 functions strongly for androgen receptor: characteristic expression in prostate disease. *Endocr J* **55**: 657-65.
- Hu, M, Yao, J, Cai, L, Bachman, KE, van den Brule, F, Velculescu, V, *et al.* (2005). Distinct epigenetic changes in the stromal cells of breast cancers. *Nat Genet* **37**: 899-905.
- Hua, Y, Vickers, TA, Baker, BF, Bennett, CF & Krainer, AR (2007). Enhancement of SMN2 exon 7 inclusion by antisense oligonucleotides targeting the exon. *PLoS Biol* **5**: e73.
- Huang, H, Qiao, R, Zhao, D, Zhang, T, Li, Y, Yi, F, *et al.* (2009). Profiling of mismatch discrimination in RNAi enabled rational design of allele-specific siRNAs. *Nucleic Acids Res* **37**: 7560-9.
- Hughes, TA (2006). Regulation of gene expression by alternative untranslated regions. *Trends Genet* **22**: 119-22.

- Infante, JR, Matsubayashi, H, Sato, N, Tonascia, J, Klein, AP, Riall, TA, *et al.* (2007). Peritumoral fibroblast SPARC expression and patient outcome with resectable pancreatic adenocarcinoma. *J Clin Oncol* **25**: 319-25.
- Jia, R, Li, C, McCoy, JP, Deng, CX & Zheng, ZM (2010). SRp20 is a proto-oncogene critical for cell proliferation and tumor induction and maintenance. *Int J Biol Sci* **6**: 806-26.
- Jiang, K, Apostolatos, AH, Ghansah, T, Watson, JE, Vickers, T, Cooper, DR, *et al.* (2008). Identification of a novel antiapoptotic human protein kinase C delta isoform, PKCdeltaVIII in NT2 cells. *Biochemistry* **47**: 787-97.
- Johnson, JM, Castle, J, Garrett-Engele, P, Kan, Z, Loerch, PM, Armour, CD, *et al.* (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**: 2141-4.
- Kalluri, R, Zeisberg, M (2006). Fibroblasts in cancer. *Nat Rev Cancer* **6**: 392-401.
- Kalsotra, A, Xiao, X, Ward, AJ, Castle, JC, Johnson, JM, Burge, CB, *et al.* (2008). A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proc Natl Acad Sci U S A* **105**: 20333-8.
- Karni, R, de Stanchina, E, Lowe, SW, Sinha, R, Mu, D & Krainer, AR (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol* **14**: 185-93.
- Kiaris, H, Chatzistamou, I, Trimis, G, Frangou-Plemmenou, M, Pafiti-Kondi, A & Kalofoutis, A (2005). Evidence for nonautonomous effect of p53 tumor suppressor in carcinogenesis. *Cancer Res* **65**: 1627-30.
- Kim, CJ, Nishi, K, Isono, T, Okuyama, Y, Tambe, Y, Okada, Y, *et al.* (2009). Cyclin D1b variant promotes cell invasiveness independent of binding to CDK4 in human bladder cancer cells. *Mol Carcinog* **48**: 953-64.
- Kim, J, Coffey, DM, Creighton, CJ, Yu, Z, Hawkins, SM & Matzuk, MM (2012). High-grade serous ovarian cancer arises from fallopian tube in a mouse model. *Proc Natl Acad Sci U S A* **109**: 3921-6.
- Kirschbaum-Slager, N, Lopes, GM, Galante, PA, Riggins, GJ & de Souza, SJ (2004). Splicing factors are differentially expressed in tumors. *Genet Mol Res* **3**: 512-20.
- Klinck, R, Bramard, A, Inkel, L, Dufresne-Martin, G, Gervais-Bird, J, Madden, R, *et al.* (2008). Multiple alternative splicing markers for ovarian cancer. *Cancer Res* **68**: 657-63.
- Knudson, AG, Jr (1971). Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**: 820-3.
- Kopelovich, L, Conlon, S & Pollack, R (1977). Defective organization of actin in cultured skin fibroblasts from patients with inherited adenocarcinoma. *Proc Natl Acad Sci U S A* **74**: 3019-22.
- Krawczak, M, Reiss, J & Cooper, DN (1992). The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet* **90**: 41-54.
- Krummel, DA, Nagai, K & Oubridge, C (2010). Structure of spliceosomal ribonucleoproteins. *Fl1000 Biol Rep* **2**: 39.
- Kuehl, WM, Bergsagel, PL (2002). Multiple myeloma: evolving genetic events and host interactions. *Nat Rev Cancer* **2**: 175-87.
- Kurman, RJ, Shih, I (2010). The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. *Am J Surg Pathol* **34**: 433-43.
- Kurman, RJ, Shih, I (2008). Pathogenesis of ovarian cancer: lessons from morphology and molecular biology and their clinical implications. *Int J Gynecol Pathol* **27**: 151-60.

- Kurose, K, Gilley, K, Matsumoto, S, Watson, PH, Zhou, XP & Eng, C (2002). Frequent somatic mutations in PTEN and TP53 are mutually exclusive in the stroma of breast carcinomas. *Nat Genet* **32**: 355-7.
- Kuroyanagi, H (2009). Fox-1 family of RNA-binding proteins. *Cell Mol Life Sci* **66**: 3895-907.
- Kurreck, J (2003). Antisense technologies. Improvement through novel chemical modifications. *Eur J Biochem* **270**: 1628-44.
- Lacerra, G, Sierakowska, H, Carestia, C, Fucharoen, S, Summerton, J, Weller, D, *et al.* (2000). Restoration of hemoglobin A synthesis in erythroid cells from peripheral blood of thalassemic patients. *Proc Natl Acad Sci U S A* **97**: 9591-6.
- Lai, JC, Tan, W, Benimetskaya, L, Miller, P, Colombini, M & Stein, CA (2006). A pharmacologic target of G3139 in melanoma cells may be the mitochondrial VDAC. *Proc Natl Acad Sci U S A* **103**: 7494-9.
- Landen, CN, Jr, Birrer, MJ & Sood, AK (2008). Early events in the pathogenesis of epithelial ovarian cancer. *J Clin Oncol* **26**: 995-1005.
- Lander, ES, Linton, LM, Birren, B, Nusbaum, C, Zody, MC, Baldwin, J, *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.
- Larocque, D, Galarneau, A, Liu, HN, Scott, M, Almazan, G & Richard, S (2005). Protection of p27(Kip1) mRNA by quaking RNA binding proteins promotes oligodendrocyte differentiation. *Nat Neurosci* **8**: 27-33.
- Larocque, D, Pilote, J, Chen, T, Cloutier, F, Massie, B, Pedraza, L, *et al.* (2002). Nuclear retention of MBP mRNAs in the quaking viable mice. *Neuron* **36**: 815-29.
- Leech, SH, Olie, RA, Gautschi, O, Simoes-Wust, AP, Tschopp, S, Haner, R, *et al.* (2000). Induction of apoptosis in lung-cancer cells following bcl-xL anti-sense treatment. *Int J Cancer* **86**: 570-6.
- Levanon, K, Ng, V, Piao, HY, Zhang, Y, Chang, MC, Roh, MH, *et al.* (2010). Primary ex vivo cultures of human fallopian tube epithelium as a model for serous ovarian carcinogenesis. *Oncogene* **29**: 1103-13.
- Levin, AA (1999). A review of the issues in the pharmacokinetics and toxicology of phosphorothioate antisense oligonucleotides. *Biochim Biophys Acta* **1489**: 69-84.
- Lewis, BP, Green, RE & Brenner, SE (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc Natl Acad Sci U S A* **100**: 189-92.
- Li, R, Liang, J, Ni, S, Zhou, T, Qing, X, Li, H, *et al.* (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7**: 51-63.
- Li, Z, Takakura, N, Oike, Y, Imanaka, T, Araki, K, Suda, T, *et al.* (2003). Defective smooth muscle development in qkI-deficient mice. *Dev Growth Differ* **45**: 449-62.
- Livak, KJ, Schmittgen, TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**: 402-8.
- Llorian, M, Smith, CW (2011). Decoding muscle alternative splicing. *Curr Opin Genet Dev* **21**: 380-7.
- Loeb, LA (1991). Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res* **51**: 3075-9.
- Lohmann, DR, Gallie, BL (2004). Retinoblastoma: revisiting the model prototype of inherited cancer. *Am J Med Genet C Semin Med Genet* **129C**: 23-8.

- Lopez-Bigas, N, Blencowe, BJ & Ouzounis, CA (2006). Highly consistent patterns for inherited human diseases at the molecular level. *Bioinformatics* **22**: 269-77.
- Lu, QL, Yokota, T, Takeda, S, Garcia, L, Muntoni, F & Partridge, T (2011). The status of exon skipping as a therapeutic approach to duchenne muscular dystrophy. *Mol Ther* **19**: 9-15.
- Lutz, CS (2008). Alternative polyadenylation: a twist on mRNA 3' end formation. *ACS Chem Biol* **3**: 609-17.
- Ma, XJ, Salunga, R, Tuggle, JT, Gaudet, J, Enright, E, McQuary, P, *et al.* (2003). Gene expression profiles of human breast cancer progression. *Proc Natl Acad Sci U S A* **100**: 5974-9.
- Madocsai, C, Lim, SR, Geib, T, Lam, BJ & Hertel, KJ (2005). Correction of SMN2 Pre-mRNA splicing by antisense U7 small nuclear RNAs. *Mol Ther* **12**: 1013-22.
- Manoharan, M (1999). 2'-Carbohydrate Modifications in Antisense Oligonucleotide Therapy: Importance of Conformation, Configuration and Conjugation. *Biochim Biophys Acta* **1489**: 117-30.
- Markham, NR, Zuker, M (2005). DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res* **33**: W577-81.
- Martin, D, Gutkind, JS (2008). Human tumor-associated viruses and new insights into the molecular mechanisms of cancer. *Oncogene* **27 Suppl 2**: S31-42.
- Martinez-Contreras, R, Fisette, JF, Nasim, FU, Madden, R, Cordeau, M & Chabot, B (2006). Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biol* **4**: e21.
- Martins de Araujo, M, Bonnal, S, Hastings, ML, Krainer, AR & Valcarcel, J (2009). Differential 3' splice site recognition of SMN1 and SMN2 transcripts by U2AF and U2 snRNP. *RNA* **15**: 515-23.
- Masuda, N, Ohnishi, T, Kawamoto, S, Monden, M & Okubo, K (1999). Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples. *Nucleic Acids Res* **27**: 4436-43.
- Mayeda, A, Krainer, AR (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68**: 365-75.
- Medeiros, F, Muto, MG, Lee, Y, Elvin, JA, Callahan, MJ, Feltmate, C, *et al.* (2006). The tubal fimbria is a preferred site for early adenocarcinoma in women with familial ovarian cancer syndrome. *Am J Surg Pathol* **30**: 230-6.
- Mercatante, DR, Bortner, CD, Cidlowski, JA & Kole, R (2001). Modification of alternative splicing of Bcl-x pre-mRNA in prostate and breast cancer cells. analysis of apoptosis and cell death. *J Biol Chem* **276**: 16411-7.
- Mercatante, DR, Mohler, JL & Kole, R (2002). Cellular response to an antisense-mediated shift of Bcl-x pre-mRNA splicing and antineoplastic agents. *J Biol Chem* **277**: 49374-82.
- Metcalfe, P, Bleeker, PM, von dem Borne, AE & Ouwehand, WH (1995). Human platelet antigen-2 and -3 genotyping by PCR-SSP. *Transfus Med* **5**: 285-8.
- Morrison, TB, Weis, JJ & Wittwer, CT (1998). Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *BioTechniques* **24**: 954,8, 960, 962.
- Mueller, MM, Fusenig, NE (2004). Friends or foes - bipolar effects of the tumour stroma in cancer. *Nat Rev Cancer* **4**: 839-49.
- Muntoni, F, Torelli, S & Ferlini, A (2003). Dystrophin and mutations: one gene, several proteins, multiple phenotypes. *Lancet Neurol* **2**: 731-40.

- Muro, AF, Caputi, M, Pariyarath, R, Pagani, F, Buratti, E & Baralle, FE (1999). Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. *Mol Cell Biol* **19**: 2657-71.
- Nachman, MW, Crowell, SL (2000). Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**: 297-304.
- Nakamura, H, Oda, Y, Iwai, S, Inoue, H, Ohtsuka, E, Kanaya, S, *et al.* (1991). How does RNase H recognize a DNA.RNA hybrid? *Proc Natl Acad Sci U S A* **88**: 11535-9.
- Nakamura, M, Tokura, Y (2011). Epithelial-mesenchymal transition in the skin. *J Dermatol Sci* **61**: 7-13.
- Nordling, CO (1953). A new theory on cancer-inducing mechanism. *Br J Cancer* **7**: 68-72.
- Nowell, PC (1976). The clonal evolution of tumor cell populations. *Science* **194**: 23-8.
- Okamura, K, Lai, EC (2008). Endogenous small interfering RNAs in animals. *Nat Rev Mol Cell Biol* **9**: 673-8.
- Olumi, AF, Grossfeld, GD, Hayward, SW, Carroll, PR, Tlsty, TD & Cunha, GR (1999). Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. *Cancer Res* **59**: 5002-11.
- Orian-Rousseau, V, Chen, L, Sleeman, JP, Herrlich, P & Ponta, H (2002). CD44 is required for two consecutive steps in HGF/c-Met signaling. *Genes Dev* **16**: 3074-86.
- Orimo, A, Gupta, PB, Sgroi, DC, Arenzana-Seisdedos, F, Delaunay, T, Naeem, R, *et al.* (2005). Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* **121**: 335-48.
- Owen, N, Zhou, H, Malygin, AA, Sangha, J, Smith, LD, Muntoni, F, *et al.* (2011). Design principles for bifunctional targeted oligonucleotide enhancers of splicing. *Nucleic Acids Res* **39**: 7194-208.
- Pan, Q, Shai, O, Misquitta, C, Zhang, W, Saltzman, AL, Mohammad, N, *et al.* (2004). Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol Cell* **16**: 929-41.
- Pankratova, S, Nielsen, BN, Shiraishi, T & Nielsen, PE (2010). PNA-mediated modulation and redirection of Her-2 pre-mRNA splicing: specific skipping of erbB-2 exon 19 coding for the ATP catalytic domain. *Int J Oncol* **36**: 29-38.
- Parenteau, J, Durand, M, Morin, G, Gagnon, J, Lucier, JF, Wellinger, RJ, *et al.* (2011). Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. *Cell* **147**: 320-31.
- Parkin, DM (2006). The global health burden of infection-associated cancers in the year 2002. *Int J Cancer* **118**: 3030-44.
- Patel, AA, Steitz, JA (2003). Splicing double: insights from the second spliceosome. *Nat Rev Mol Cell Biol* **4**: 960-70.
- Patel, SB, Bellini, M (2008). The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res* **36**: 6482-93.
- Pettigrew, CA, Brown, MA (2008). Pre-mRNA splicing aberrations and cancer. *Front Biosci* **13**: 1090-105.
- Polyak, K, Weinberg, RA (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer* **9**: 265-73.
- Prat, J, Ribe, A & Gallardo, A (2005). Hereditary ovarian cancer. *Hum Pathol* **36**: 861-70.
- Qiu, W, Hu, M, Sridhar, A, Opeskin, K, Fox, S, Shipitsin, M, *et al.* (2008). No evidence of clonal somatic genetic alterations in cancer-associated fibroblasts from human breast and ovarian carcinomas. *Nat Genet* **40**: 650-5.

- Quaggin, SE, Kapus, A (2011). Scar wars: mapping the fate of epithelial-mesenchymal-myofibroblast transition. *Kidney Int* **80**: 41-50.
- Rayburn, ER, Zhang, R (2008). Antisense, RNAi, and gene silencing strategies for therapy: mission possible or impossible? *Drug Discov Today* **13**: 513-21.
- Religio, A, Ben-Dov, C, Baum, M, Ruggiu, M, Gemund, C, Benes, V, *et al.* (2005). Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells. *J Biol Chem* **280**: 4779-84.
- Risch, HA (1998). Hormonal etiology of epithelial ovarian cancer, with a hypothesis concerning the role of androgens and progesterone. *J Natl Cancer Inst* **90**: 1774-86.
- Rogerson, L, Darby, S, Jabbar, T, Mathers, ME, Leung, HY, Robson, CN, *et al.* (2008). Application of transcript profiling in formalin-fixed paraffin-embedded diagnostic prostate cancer needle biopsies. *BJU Int* **102**: 364-70.
- Rowley, JD (1973). Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**: 290-3.
- Samavarchi-Tehrani, P, Golipour, A, David, L, Sung, HK, Beyer, TA, Datti, A, *et al.* (2010). Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**: 64-77.
- Santarius, T, Shipley, J, Brewer, D, Stratton, MR & Cooper, CS (2010). A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* **10**: 59-64.
- Schena, M, Shalon, D, Davis, RW & Brown, PO (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467-70.
- Scheffzek, K, Ahmadian, MR, Kabsch, W, Wiesmuller, L, Lautwein, A, Schmitz, F, *et al.* (1997). The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* **277**: 333-8.
- Schildkraut, JM, Schwingl, PJ, Bastos, E, Evanoff, A & Hughes, C (1996). Epithelial ovarian cancer risk among women with polycystic ovary syndrome. *Obstet Gynecol* **88**: 554-9.
- Schwarz, DS, Ding, H, Kennington, L, Moore, JT, Schelter, J, Burchard, J, *et al.* (2006). Designing siRNA that distinguish between genes that differ by a single nucleotide. *PLoS Genet* **2**: e140.
- Seth, P, Miller, HB, Lasda, EL, Pearson, JL & Garcia-Blanco, MA (2008). Identification of an intronic splicing enhancer essential for the inclusion of FGFR2 exon IIIc. *J Biol Chem* **283**: 10058-67.
- Shieh, JJ, Liu, KT, Huang, SW, Chen, YJ & Hsieh, TY (2009). Modification of alternative splicing of Mcl-1 pre-mRNA using antisense morpholino oligonucleotides induces apoptosis in basal cell carcinoma cells. *J Invest Dermatol* **129**: 2497-506.
- Shimoda, M, Mellody, KT & Orimo, A (2010). Carcinoma-associated fibroblasts are a rate-limiting determinant for tumour progression. *Semin Cell Dev Biol* **21**: 19-25.
- Shiraishi, T, Eysturskarth, J & Nielsen, PE (2010). Modulation of mdm2 pre-mRNA splicing by 9-aminoacridine-PNA (peptide nucleic acid) conjugates targeting intron-exon junctions. *BMC Cancer* **10**: 342.
- Shook, D, Keller, R (2003). Mechanisms, mechanics and function of epithelial-mesenchymal transitions in early development. *Mech Dev* **120**: 1351-83.
- Sidman, RL, Dickie, MM & Appel, SH (1964). Mutant Mice (Quaking and Jimpy) with Deficient Myelination in the Central Nervous System. *Science* **144**: 309-11.

- Simoës-Wust, AP, Olie, RA, Gautschi, O, Leech, SH, Haner, R, Hall, J, *et al.* (2000). Bcl-xl antisense treatment induces apoptosis in breast carcinoma cells. *Int J Cancer* **87**: 582-90.
- Skordis, LA, Dunckley, MG, Yue, B, Eperon, IC & Muntoni, F (2003). Bifunctional antisense oligonucleotides provide a trans-acting splicing enhancer that stimulates SMN2 gene expression in patient fibroblasts. *Proc Natl Acad Sci U S A* **100**: 4114-9.
- Skotheim, RI, Nees, M (2007). Alternative splicing in cancer: noise, functional, or systematic? *Int J Biochem Cell Biol* **39**: 1432-49.
- Srebrow, A, Kornblihtt, AR (2006). The connection between splicing and cancer. *J Cell Sci* **119**: 2635-41.
- Stam, K, Heisterkamp, N, Grosveld, G, de Klein, A, Verma, RS, Coleman, M, *et al.* (1985). Evidence of a new chimeric bcr/c-abl mRNA in patients with chronic myelocytic leukemia and the Philadelphia chromosome. *N Engl J Med* **313**: 1429-33.
- Stephenson, ML, Zamecnik, PC (1978). Inhibition of Rous sarcoma viral RNA translation by a specific oligodeoxyribonucleotide. *Proc Natl Acad Sci U S A* **75**: 285-8.
- Sumantran, VN, Ealovega, MW, Nunez, G, Clarke, MF & Wicha, MS (1995). Overexpression of Bcl-XS sensitizes MCF-7 cells to chemotherapy-induced apoptosis. *Cancer Res* **55**: 2507-10.
- Tabin, CJ, Bradley, SM, Bargmann, CI, Weinberg, RA, Papageorge, AG, Scolnick, EM, *et al.* (1982). Mechanism of activation of a human oncogene. *Nature* **300**: 143-9.
- Takahashi, Y, Ishii, G, Taira, T, Fujii, S, Yanagi, S, Hishida, T, *et al.* (2011). Fibrous stroma is associated with poorer prognosis in lung squamous cell carcinoma patients. *J Thorac Oncol* **6**: 1460-7.
- Tange, TO, Damgaard, CK, Guth, S, Valcarcel, J & Kjems, J (2001). The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J* **20**: 5748-58.
- Taylor, JK, Zhang, QQ, Wyatt, JR & Dean, NM (1999). Induction of endogenous Bcl-xS through the control of Bcl-x pre-mRNA splicing by antisense oligonucleotides. *Nat Biotechnol* **17**: 1097-100.
- Thierry-Mieg, D, Thierry-Mieg, J (2006). AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol* **7 Suppl 1**: S12.1-14.
- Thiery, JP, Acloque, H, Huang, RY & Nieto, MA (2009). Epithelial-mesenchymal transitions in development and disease. *Cell* **139**: 871-90.
- Tian, M, Schiemann, WP (2009). The TGF-beta paradox in human cancer: an update. *Future Oncol* **5**: 259-71.
- Tone, AA, Begley, H, Sharma, M, Murphy, J, Rosen, B, Brown, TJ, *et al.* (2008). Gene expression profiles of luteal phase fallopian tube epithelium from BRCA mutation carriers resemble high-grade serous carcinoma. *Clin Cancer Res* **14**: 4067-78.
- Tohill, RW, Tinker, AV, George, J, Brown, R, Fox, SB, Lade, S, *et al.* (2008). Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res* **14**: 5198-208.
- Trimboli, AJ, Cantemir-Stone, CZ, Li, F, Wallace, JA, Merchant, A, Creasap, N, *et al.* (2009). Pten in stromal fibroblasts suppresses mammary epithelial tumours. *Nature* **461**: 1084-91.
- Trope, C, Davidson, B, Paulsen, T, Abeler, VM & Kaern, J (2009). Diagnosis and treatment of borderline ovarian neoplasms "the state of the art". *Eur J Gynaecol Oncol* **30**: 471-82.

- Tuhkanen, H, Anttila, M, Kosma, VM, Heinonen, S, Juhola, M, Helisalimi, S, *et al.* (2006). Frequent gene dosage alterations in stromal cells of epithelial ovarian carcinomas. *Int J Cancer* **119**: 1345-53.
- Vandenbroucke, II, Vandesompele, J, Paepe, AD & Messiaen, L (2001). Quantification of splice variants using real-time PCR. *Nucleic Acids Res* **29**: E68-8.
- Vaughan, S, Coward, JI, Bast, RC, Jr, Berchuck, A, Berek, JS, Brenton, JD, *et al.* (2011). Rethinking ovarian cancer: recommendations for improving outcomes. *Nat Rev Cancer* **11**: 719-25.
- Venables, JP, Klinck, R, Bramard, A, Inkel, L, Dufresne-Martin, G, Koh, C, *et al.* (2008). Identification of alternative splicing markers for breast cancer. *Cancer Res* **68**: 9525-31.
- Venables, JP, Klinck, R, Koh, C, Gervais-Bird, J, Bramard, A, Inkel, L, *et al.* (2009). Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol* **16**: 670-6.
- Villemaire J, mémoire de maîtrise, Université de Sherbrooke
- Villemaire, J, Dion, I, Elela, SA & Chabot, B (2003). Reprogramming alternative pre-messenger RNA splicing through the use of protein-binding antisense oligonucleotides. *J Biol Chem* **278**: 50031-9.
- Voelker, RB, Berglund, JA (2007). A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing. *Genome Res* **17**: 1023-33.
- Walder, RY, Walder, JA (1988). Role of RNase H in hybrid-arrested translation by antisense oligonucleotides. *Proc Natl Acad Sci U S A* **85**: 5011-5.
- Waldron, L, Simpson, P, Parmigiani, G & Huttenhower, C (2012). Report on emerging technologies for translational bioinformatics: a symposium on gene expression profiling for archival tissues. *BMC Cancer* **12**: 124.
- Wan, J, Sazani, P & Kole, R (2009). Modification of HER2 pre-mRNA alternative splicing and its effects on breast cancer cells. *Int J Cancer* **124**: 772-7.
- Wang, ET, Sandberg, R, Luo, S, Khrebtkova, I, Zhang, L, Mayr, C, *et al.* (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470-6.
- Wang, J, Bhutani, M, Pathak, AK, Lang, W, Ren, H, Jelinek, J, *et al.* (2007). Delta DNMT3B variants regulate DNA methylation in a promoter-specific manner. *Cancer Res* **67**: 10647-52.
- Wang, L, Duke, L, Zhang, PS, Arlinghaus, RB, Symmans, WF, Sahin, A, *et al.* (2003a). Alternative splicing disrupts a nuclear localization signal in spleen tyrosine kinase that is required for invasion suppression in breast cancer. *Cancer Res* **63**: 4724-30.
- Wang, Z, Lo, HS, Yang, H, Gere, S, Hu, Y, Buetow, KH, *et al.* (2003b). Computational analysis and experimental validation of tumor-associated alternative RNA splicing in human cancer. *Cancer Res* **63**: 655-7.
- Wang, Z, Xiao, X, Van Nostrand, E & Burge, CB (2006). General and specific functions of exonic splicing silencers in splicing control. *Mol Cell* **23**: 61-70.
- Warzecha, CC, Jiang, P, Amirikian, K, Dittmar, KA, Lu, H, Shen, S, *et al.* (2010). An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. *EMBO J* **29**: 3286-300.
- Wells, A, Chao, YL, Grahovac, J, Wu, Q & Lauffenburger, DA (2011). Epithelial and mesenchymal phenotypic switchings modulate cell motility in metastasis. *Front Biosci* **16**: 815-37.
- White, ES, Baralle, FE & Muro, AF (2008). New insights into form and function of fibronectin splice variants. *J Pathol* **216**: 1-14.

- Wilhelm, E, Pelay, FX, Benecke, A & Bell, B (2008). TAF6delta controls apoptosis and gene expression in the absence of p53. *PLoS One* **3**: e2721.
- Wilusz, JE, Devanney, SC & Caputi, M (2005). Chimeric peptide nucleic acid compounds modulate splicing of the bcl-x gene in vitro and in vivo. *Nucleic Acids Res* **33**: 6547-54.
- Wood, MJ, Gait, MJ & Yin, H (2010). RNA-targeted splice-correction therapy for neuromuscular disease. *Brain* **133**: 957-72.
- Wu, D, Matsushita, K, Matsubara, H, Nomura, F & Tomonaga, T (2011). An alternative splicing isoform of eukaryotic initiation factor 4H promotes tumorigenesis in vivo and is a potential therapeutic target for human cancer. *Int J Cancer* **128**: 1018-30.
- Wu, J, Zhou, L, Tonissen, K, Tee, R & Artzt, K (1999). The quaking I-5 protein (QKI-5) has a novel nuclear localization signal and shuttles between the nucleus and the cytoplasm. *J Biol Chem* **274**: 29202-10.
- Wu, JI, Reed, RB, Grabowski, PJ & Artzt, K (2002). Function of quaking in myelination: regulation of alternative splicing. *Proc Natl Acad Sci U S A* **99**: 4233-8.
- Xing, Y (2007). Genomic analysis of RNA alternative splicing in cancers. *Front Biosci* **12**: 4034-41.
- Yaku, H, Yukimasa, T, Nakano, S, Sugimoto, N & Oka, H (2008). Design of allele-specific primers and detection of the human ABO genotyping to avoid the pseudopositive problem. *Electrophoresis* **29**: 4130-40.
- Yang, G, Fu, H, Zhang, J, Lu, X, Yu, F, Jin, L, *et al.* (2010). RNA-binding protein quaking, a critical regulator of colon epithelial differentiation and a suppressor of colon cancer. *Gastroenterology* **138**: 231,40.e1-5.
- Yeakley, JM, Tronchere, H, Olesen, J, Dyck, JA, Wang, HY & Fu, XD (1999). Phosphorylation regulates in vivo interaction and molecular targeting of serine/arginine-rich pre-mRNA splicing factors. *J Cell Biol* **145**: 447-55.
- Yeo, GW, Van Nostrand, EL & Liang, TY (2007). Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. *PLoS Genet* **3**: e85.
- Zamecnik, PC (1996). History of antisense oligonucleotides. *Methods Mol Med* **1**: 1-11.
- Zearfoss, NR, Clingman, CC, Farley, BM, McCoig, LM & Ryder, SP (2011). Quaking regulates Hnrnpa1 expression through its 3' UTR in oligodendrocyte precursor cells. *PLoS Genet* **7**: e1001269.
- Zhang, C, Frias, MA, Mele, A, Ruggiu, M, Eom, T, Marney, CB, *et al.* (2010). Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science* **329**: 439-43.
- Zhang, C, Li, HR, Fan, JB, Wang-Rodriguez, J, Downs, T, Fu, XD, *et al.* (2006). Profiling alternatively spliced mRNA isoforms for prostate cancer classification. *BMC Bioinformatics* **7**: 202.
- Zhang, C, Zhang, Z, Castle, J, Sun, S, Johnson, J, Krainer, AR, *et al.* (2008). Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev* **22**: 2550-63.
- Zhang, W, Liu, H, Han, K & Grabowski, PJ (2002). Region-specific alternative splicing in the nervous system: implications for regulation by the RNA-binding protein NAPOR. *RNA* **8**: 671-85.
- Zhang, XH, Chasin, LA (2004). Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev* **18**: 1241-50.

- Zhao, L, Mandler, MD, Yi, H & Feng, Y (2010). Quaking I controls a unique cytoplasmic pathway that regulates alternative splicing of myelin-associated glycoprotein. *Proc Natl Acad Sci U S A* **107**: 19061-6.
- Zhaozui, Z, Xing, S, Ping, C & Fuqing, Z (2006). *Journal of Huazhong University of Science and Technology* **26**: 213-6.
- Zhou, GH, Shirakura, H, Kamahori, M, Okano, K, Nagai, K & Kambara, H (2004). A gel-free SNP genotyping method: bioluminometric assay coupled with modified primer extension reactions (BAMPER) directly from double-stranded PCR products. *Hum Mutat* **24**: 155-63.
- Zhu, J, Mayeda, A & Krainer, AR (2001). Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell* **8**: 1351-61.
- Zorn, KK, Jazaeri, AA, Awtrey, CS, Gardner, GJ, Mok, SC, Boyd, J, *et al.* (2003). Choice of normal ovarian control influences determination of differentially expressed genes in ovarian cancer expression profiling studies. *Clin Cancer Res* **9**: 4811-8.

ANNEXES

Annexe 1 Base de données de TOSS

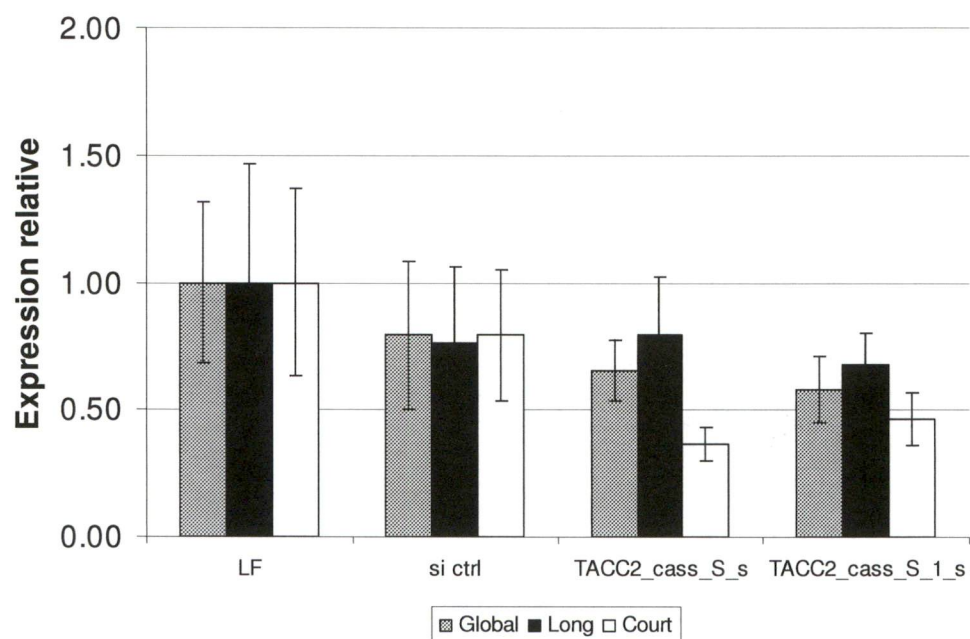
Pour chacun des ASEs ciblés par les TOSS, le type d'ASE, le nom de l'oligonucléotide et les amorces PCR sont indiqués. Aug. Détection possible seulement lors d'un traitement avec le TOSS ou SSO; A = modulation de type A; B = modulation de type B; NON = sans effet; A ou B = Le TOSS ou SSO fonctionne mais les résultats ne permettent pas de trancher en faveur d'une modulation de type A ou B.

Disponible à partir de <http://palace.lgfus.ca/pcrreactiongroup/list/265>.

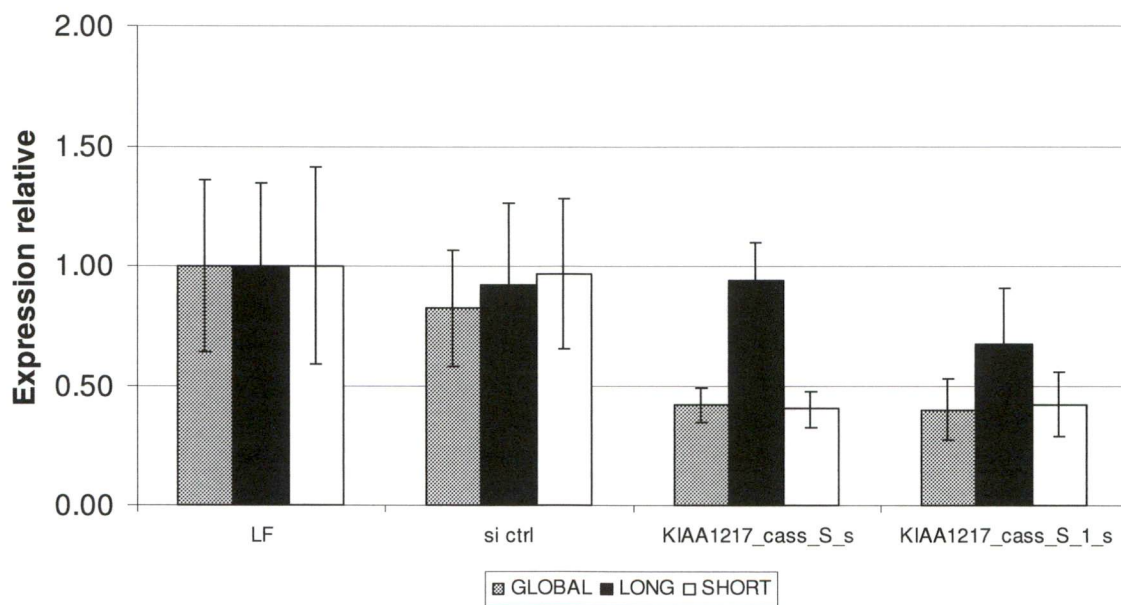
Annexe 2 Étude de spécificité des siRNAs spécifiques à l'isoforme court

Dans la colonne « Séquences », les nucléotides qui diffèrent par rapport à la jonction ciblée sont soulignés. Dans la colonne « Position des mésappariements », la position du mésappariement est notée en fonction de la numérotation du brin antisens du siRNA de 1 à 19.

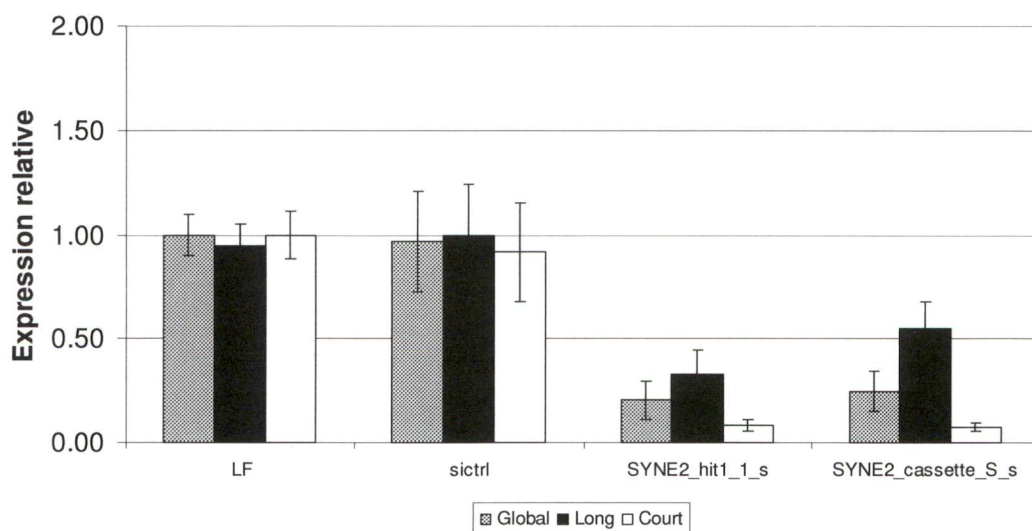
TACC2	Séquences		Position des mésappariements	
siRNA 1	GCCCAGAG GAGTTCCGATT		TACC2_cass_S_s	
siRNA 2	CAGAG GAGTTCCGATTCTG		TACC2_cass_S_1_s	
Exon126/ exon137	TCTGATAGCCCAGAG	GAGTTCCGATTCTGA		
Exon126/ intron	TCTGATAGCCCAGAG	<u>GTACGGTGGGGGCC</u>	1 à 3 et 5 à 10	1 à 6 et 7 à 13
Exon126/ exon135	TCTGATAGCCCAGAG	<u>CATCTCCCAGCTGC</u>	1 à 4, 8, 9 et 11	4 à 7, 11, 12 et 14
Intron/ exon137	<u>TCCCTCTCTCATCAG</u>	GAGTTCCGATTCTGA	14 à 16, 18 et 19	17 à 19
Exon135/ exon 137	<u>CAAGGACCTCAGCAG</u>	GAGTTCCGATTCTGA	14 à 16, 18 et 19	17 à 19



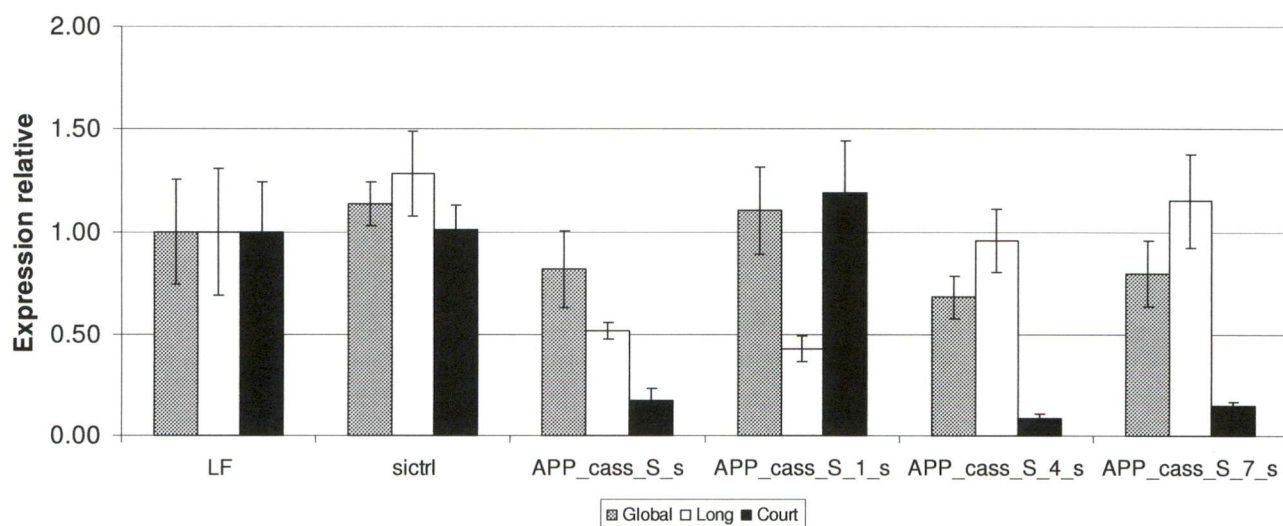
KIAA1217	Séquences		Position des mésappariements	
siRNA 1	GAGAGACCAG CGAGAAAAT		KIAA1217_cass_S_s	
siRNA 2	GACCAG CGAGAAAATGATG		KIAA1217_cass_S_1_s	
Exon833/ exon50	AGACAGAGAGACCAG	CGAGAAAATGATGAA		
Exon833/ intron	AGACAGAGAGACCAG	<u>GTAAGGTGCAGTGAG</u>	1 à 6, 8 et 9	2 à 10, 12 et 13
Exon833/ exon105	AGACAGAGAGACCAG	<u>AGAGAGGATGCAAGC</u>	3, 4 et 9	1 à 3, 7, 8 et 13
Intron/ exon50	<u>TTTTCTTTTATTCAG</u>	CGAGAAAATGATGAA	13 à 19	17 à 19
Exon105/ exon50	<u>CAAAGATGCGTCTAG</u>	CGAGAAAATGATGAA	12, 14, 16, 18 et 19	16 et 18



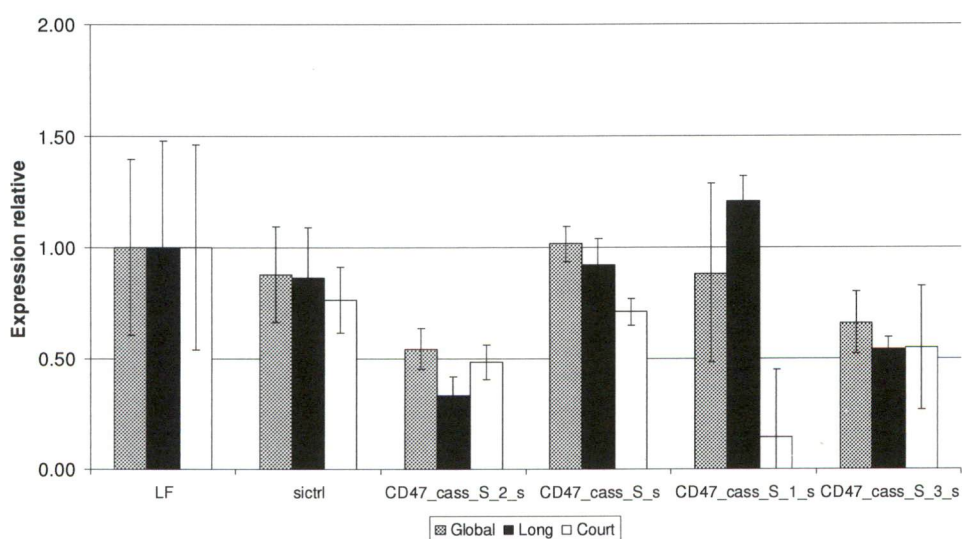
SYNE2	Séquences		Position des mésappariements		
siRNA 1	CAGCGCACTGTCAG GTAAA		SYNE2_hit1 _1_s		
siRNA 3	AGCGCACTGTCAG GTAAAT			SYNE2_cassett e_S_s	
siRNA 2	GCACTGTCAG GTAAATCCA				SYNE2_hit1 _2_s
Exon277/ exon140	ACAGCGCACTGTCAG	GTAAATCCATTTCCGG			
Exon277/ intron	ACAGCGCACTGTCAG	GTAACAGCTGGGTTTC	1	1 et 2	2 à 4
Exon277/ exon69	ACAGCGCACTGTCAG	ATGTAGAAATCCCTG	2,3 et 5	1, 3, 4 et 6	2 à 4, 6, 7 et 9
Intron/ exon140	TGTCCTTCGTTTCAG	GTAAATCCATTTCCGG	10, 12 à 19	11, 13 à 19	14, 16 à 19
Exon69/ exon140	TGAAAGCGTCTTCTG	GTAAATCCATTTCCGG	7, 10 à 13, 16, 17 et 19	8, 11 à 14, 17 et 18	11, 14 à 17 et 18



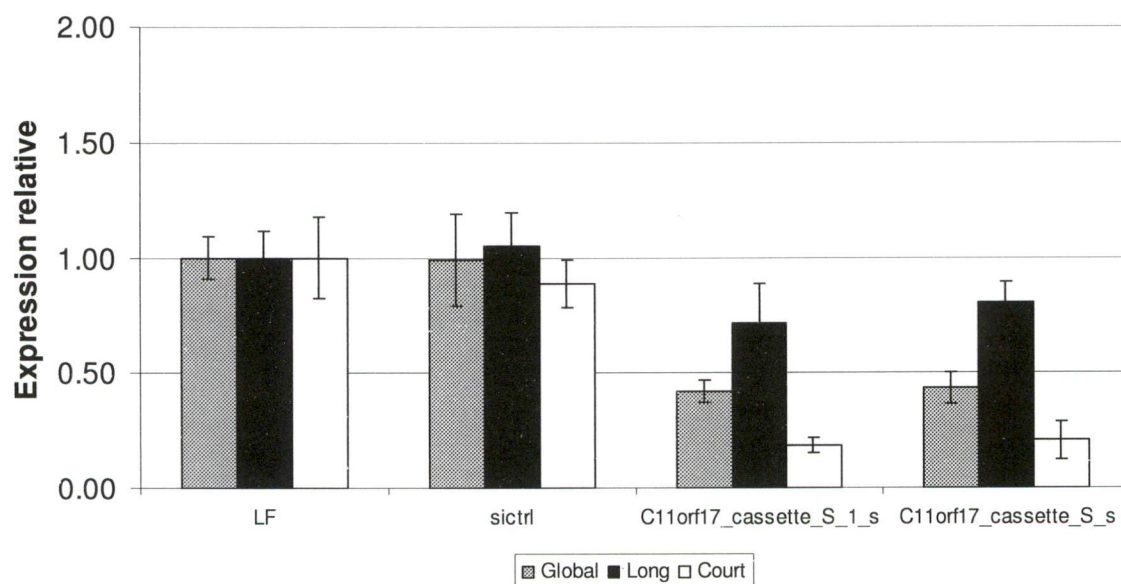
APP	Séquences		Position des mésappariements			
siRNA 4	TGTGGCAGCGCCA TTCCTA		APP_cass _S_7_s			
siRNA 2	GGCAGCGCCA TTCCTACAA			APP_cass _S_1_s		
siRNA 1	CAGCGCCA TTCCTACAACA				APP_cass_S _s	
siRNA 3	CGCCA TTCCTACAACAGCA					APP_cass_S _4_s
Exon168/ exon134	TGTGTGGCAGCGCCA	TTCCTACAACAGCAG				
Exon168/ intron	TGTGTGGCAGCGCCA	<u>GTAAGTGGACCCTTC</u>	1 à 4 et 6	2 à 7 et 9	1, 4 à 9 et 11	1 à 4, 7 à 12 et 14
Exon168/ exon57	TGTGTGGCAGCGCCA	<u>TGTCCTCAAAGTTTAC</u>	1, 2, 4 et 5	3 à 5, 7 et 8	1, 2, 5 à 7, 9 et 10	2 à 5, 8 à 10, 12 et 13
Intron/ exon134	<u>TCTTCTACTTTATAG</u>	TTCCTACAACAGCAG	7 à 13, 15 à 18	10 à 16, 18 et 19	12 à 18	15 à 19
Exon57/ exon134	<u>GAGATCCTGTAAAC</u>	TTCCTACAACAGCAG	7 à 16, 17 et 19	10 à 19	12 à 19	15 à 19
Exon203/ exon134	<u>AAGAGGTGGTTCGAG</u>	TTCCTACAACAGCAG	7 à 15, 17 à 19	10 à 18	12 à 19	15 à 19



CD47	Séquences		Position des mésappariements			
siRNA 3	GGAACCCCTTAATG AATAA		CD47_cas s_S_2_s			
siRNA 1	AACCCCTTAATG AATAACT			CD47_cass_ S_s		
siRNA 2	CCCCTTAATG AATAACTGA				CD47_cas s_S_1_s	
siRNA 4	CCCTTAATG AATAACTGAA					CD47_cass_ S_3_s
Exon25/ exon213	AGGAACCCCTTAATG	AATAACTGAAGTGAA				
Exon25/ intron	AGGAACCCCTTAATG	<u>GTATGTGGTTATTC</u>	1 à 5	1 à 7	1, 3 à 9	1, 2, 4 à 10
Exon25/ exon33	AGGAACCCCTTAATG	<u>CATCAAAGAATCAA</u>	1, 2 et 5	1 à 4 et 7	1 à 6 et 9	2 à 7 et 10
Intron/ exon213	<u>TTTTCAATTTCCAG</u>	AATAACTGAAGTGAA	7 à 9 et 12 à 19	9 à 11 et 13 à 19	11 à 13 et 15 à 19	12 à 14 et 16 à 19
Exon33/ exon213	<u>GAATGATGAATGATG</u>	AATAACTGAAGTGAA	9 et 11 à 19	11 et 13 à 19	13 et 15 à 19	14 et 16 à 19
Exon32/ exon213	<u>ATACAACCTCCTAGG</u>	AATAACTGAAGTGAA	7, 10 à 12 et 17 à 19	9, 12 à 14 et 19	11, 14 à 16	12, 15 à 17
Exon93/ exon213	<u>ATATGAAATTTGTGG</u>	AATAACTGAAGTGAA	7 à 9 et 12 à 19	9 à 11 et 14 à 19	11 à 13 et 16 à 19	12 à 14 et 17 à 19



C11orf17	Séquences		Position des mésappariements	
siRNA 2	CTCCCGGGAGAG AAATATT		C11orf17_cassette_S_1_s	
siRNA 1	CCCGGGAGAG AAATATTAT			C11orf17_cassette_S_s
Exon228/ exon105	GTTCTCCCGGGAGAG	AAATATTATTCATCT		
Exon228/ intron	GTTCTCCCGGGAGAG	<u>GTGAGGGTCGCTGTG</u>	1 à 7	1 à 9
Exon228/ exon81	GTTCTCCCGGGAGAG	<u>AGAGAAGAGAGACCC</u>	1, 2, 4 et 6	2, 3, 5 et 7
Intron/ exon105	<u>ACATTACGTCTCTAG</u>	AAATATTATTCATCT	10 à 15, 17 et 19	12 à 17 et 19
Exon81/ exon105	<u>TCAAGTCAGTGTGGG</u>	AAATATTATTCATCT	9, 11, 13, 15, 17 à 19	11, 13, 15, 17 et 19



Annexe 3 Tableau S1-17 du chapitre 4

Disponible à partir de <http://palace.lgfus.ca/pcrreactiongroup/list/219>

Username : brosseau_review

Password : SX1IT4_LR