



ANALYSE DE VARIANTS COMPLEXES

RNA-SEQ : ANOMALIES D'ÉPISSAGE ET FUSIONS

HCL

**HOSPICES CIVILS
DE LYON**

13/05/2024

SYLVAIN MARESCHAL

www.chu-lyon.fr

RNA-SEQ : ANOMALIES D'ÉPISSAGE ET FUSIONS

INTRODUCTION

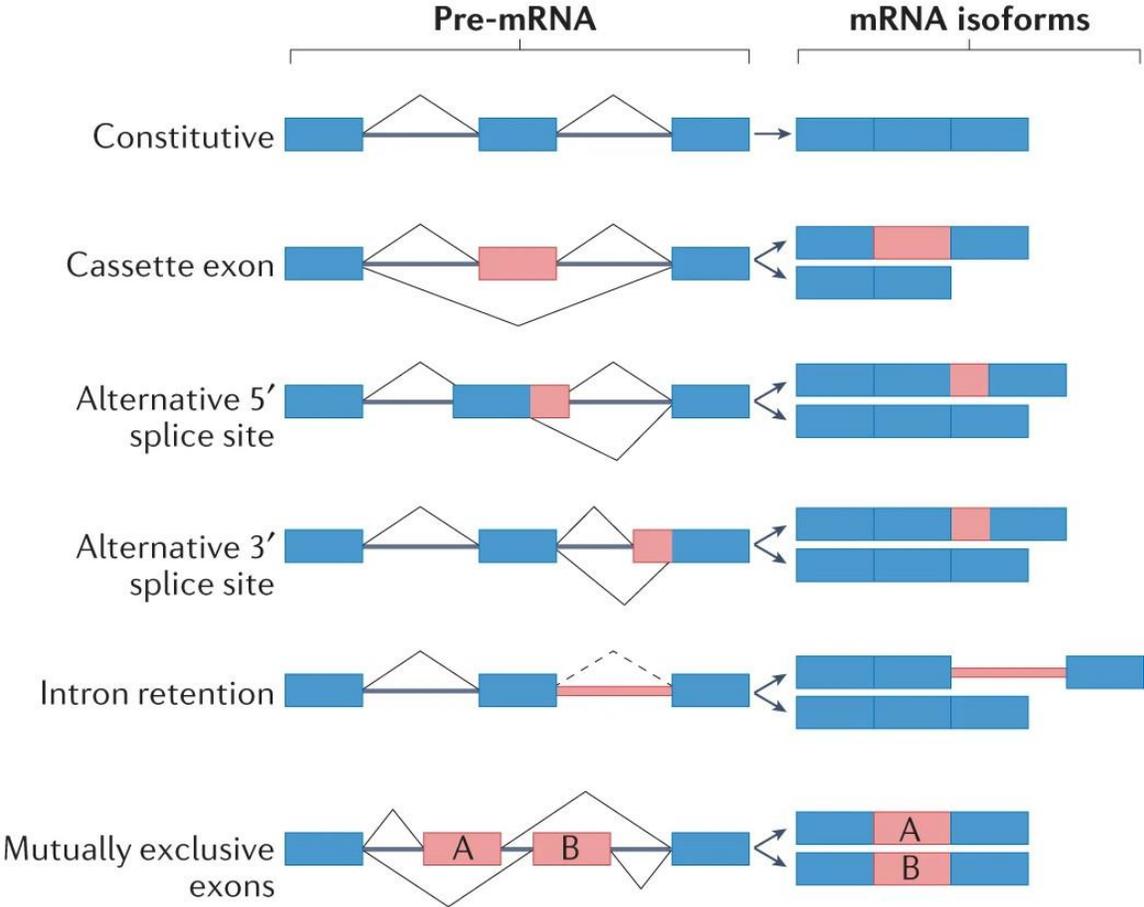
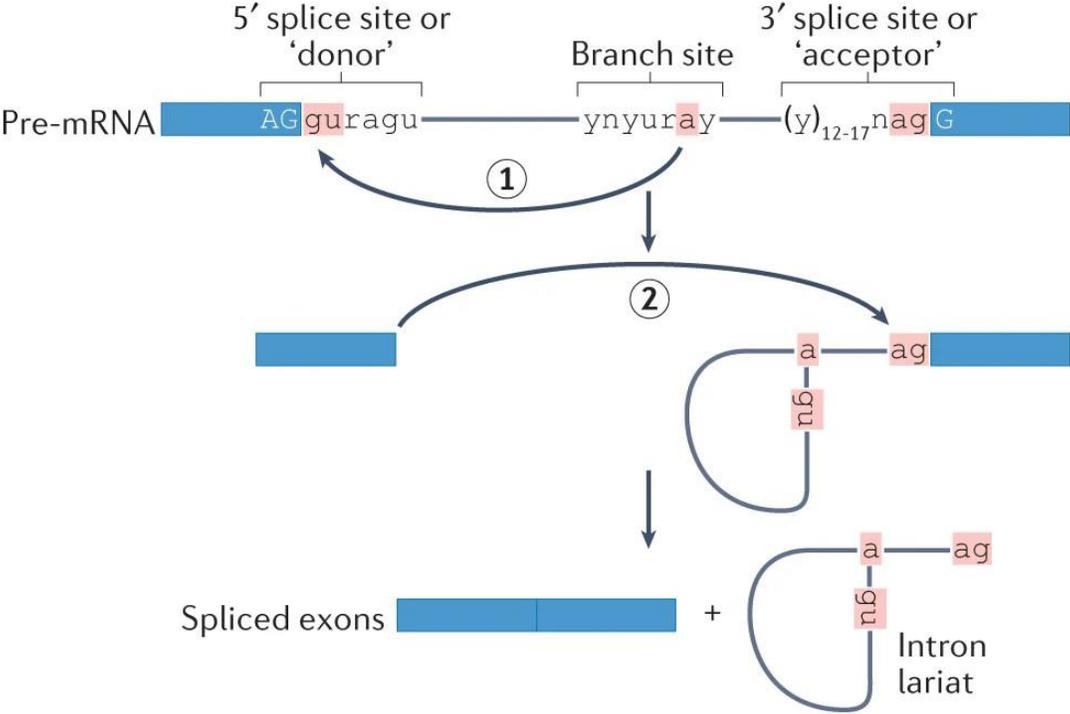
- RNA-seq aux HCL
 - **Cancer du poumon**
 - Panel MiSeq de 21 gènes (avec UMI)
 - 64 runs depuis mars 2022, ≈1200 échantillons
 - **Génétique constitutionnelle**
 - Panel NextSeq de 164 gènes (avec UMI)
 - 8 runs depuis juin 2022, ≈110 échantillons
- **Anomalies d'épissage**
 - Outils : **SpliceLauncher, SAMI**
 - Résultats sur contrôle Sereq
- **Fusions de gènes**
 - **nf-core/rnafusion** : Arriba, Star-Fusion, FusionCatcher
 - Résultats sur contrôles Sereq et Horizon

ANALYSE DE VARIANTS COMPLEXES

ANOMALIES D'ÉPISSAGE

ANOMALIES D'ÉPISSAGE

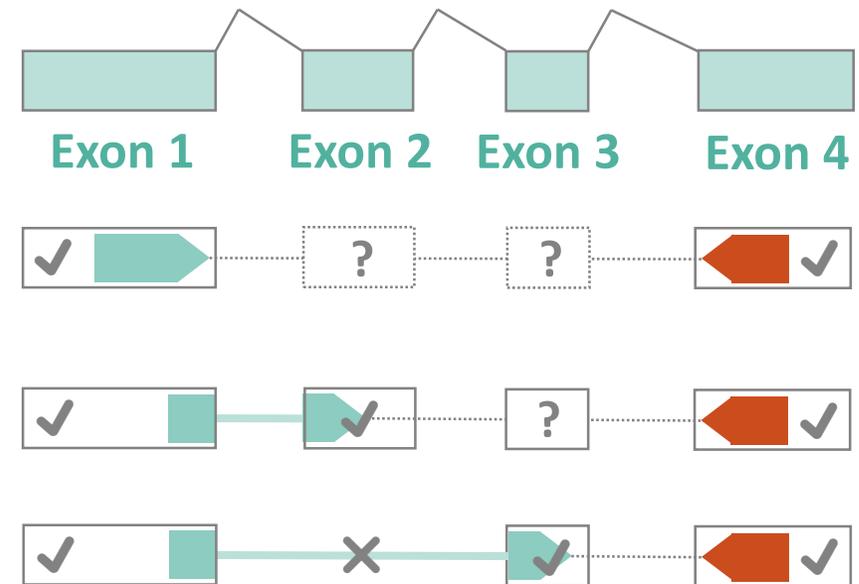
CONTEXTE : L'ÉPISSAGE ALTERNATIF



ANOMALIES D'ÉPISSAGE

CONTEXTE : CHALLENGES BIOINFORMATIQUES

- **Alignement** au génome complexe
 - Gaps de plusieurs kb
 - **STAR** reste la référence
 - Dobin et al, Bioinformatics 2013
- **Séquençage Illumina** peu informatif
 - Paired-end non-informatif
 - Split reads informatifs, mais rares en short-read
 - Approche par profondeur toujours possible
- **Forte variabilité**
 - Variabilité biologique +++, même en constit.
 - Mélange de transcrits en proportions inconnues
 - Plusieurs dizaines par gènes
 - Exprimés dans certains tissus seulement
 - Dont certains non-connus
- En diagnostic : **n=1**



ANOMALIES D'ÉPISSAGE

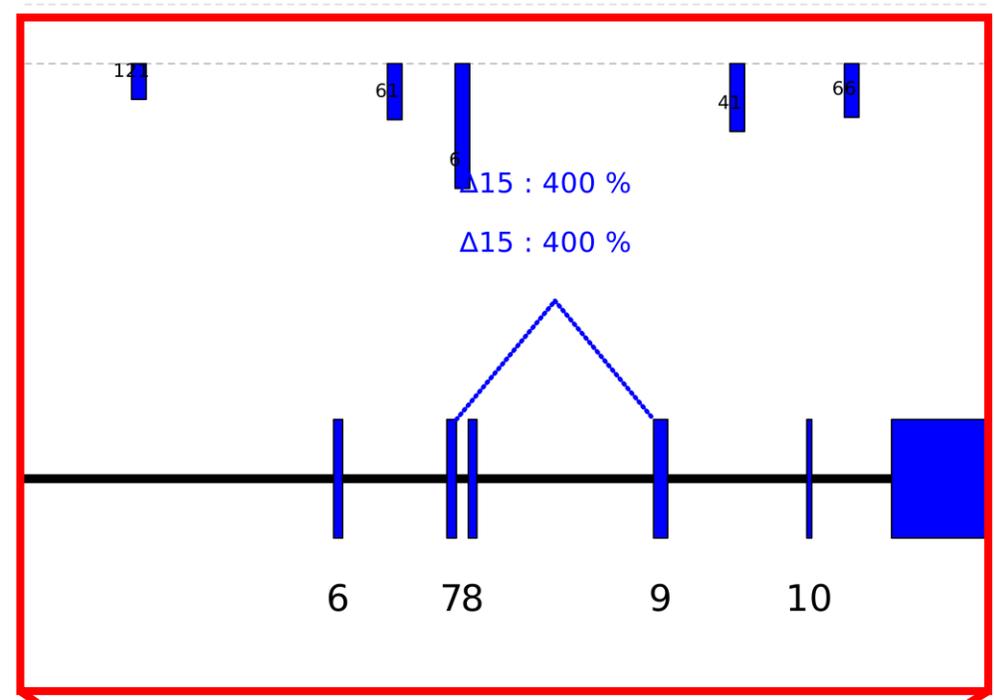
SPLICELAUNCHER

- **SpliceLauncher** : solution de Caen
 - Leman et al, Bioinformatics 2020
 - Février 2019 → Avril 2024 (pas de release)
- **Alignement STAR + post-processing**
 - Décompte personnalisé des jonctions (BAM)
 - Annotation des événements (+ nomenclature standard)
 - Ratio d'expression de l'événement (intégrant plusieurs jonctions)
 - Modélisation statistique de chaque jonction
 - Distribution Gamma ou binomiale négative si non-rejetées par pré-tests
 - Calcul d'une p-valeur (+ Bonferroni) par échantillon
- **Représentations graphiques**

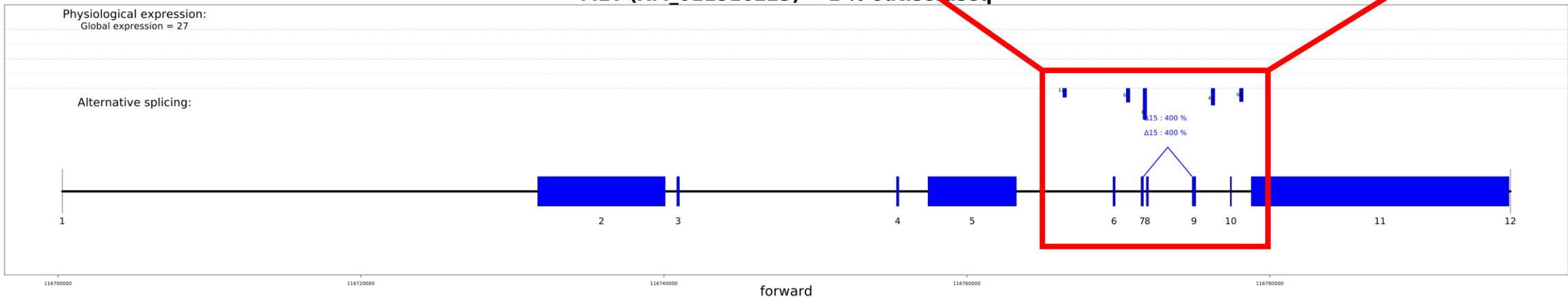
ANOMALIES D'ÉPISSAGE

SPLICELAUNCHER

	B	C	D	E	F	G	H	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR
1	chr	start	end	strand	Strand_transcript	NM	Gene	event_type	AnnotJuncs	cStart	cEnd	mean_percent	read_mean	nbSamp	DistribAjust	Significative filter	Interpretation
2	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	Physio	3_4	c.1257	c.1258	0	0,15384615	2			
3	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	Physio	5_6	c.1584	c.1585	0	0,30769231	1			
4	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	Physio	6_7	c.1758	c.1759	0	0,23076923	2			
5	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	Physio	7_8	c.1919	c.1920	0	0,69230769	2			
6	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	Physio	12_13	c.2640	c.2641	0	62,6923077	13			
7	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	Physio	13_14	c.2787	c.2788	0	55,2307692	13			
8	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	Physio	14_15	c.2944	c.2945	0	6,38461538	13			
9	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	SkipEx	$\Delta 15$	c.2944	c.3086	975,2888138	35,3076923	13			Unique junction
10	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	Physio	15_16	c.3085	c.3086	0	1,84615385	9			
11	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	NoData	Event too complex	c.3204	c.3374	0	0,07692308	1			
12	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	Physio	16_17	c.3316	c.3317	0	33,4615385	13			
13	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	NoData	Event too complex	c.3396	c.3399	0	0,15384615	2			
14	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	Physio	17_18	c.3397	c.3398	0	42,6923077	13			
15	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	NoData	Outside Transcript	c.3483	c.*169726	0	0,07692308	1			
16	NC_000007.14	1,17E+08	1,17E+08	+	forward	XM_011516223	MET	Physio	21_22	c.3992	c.3993	0	0,53846154	3			



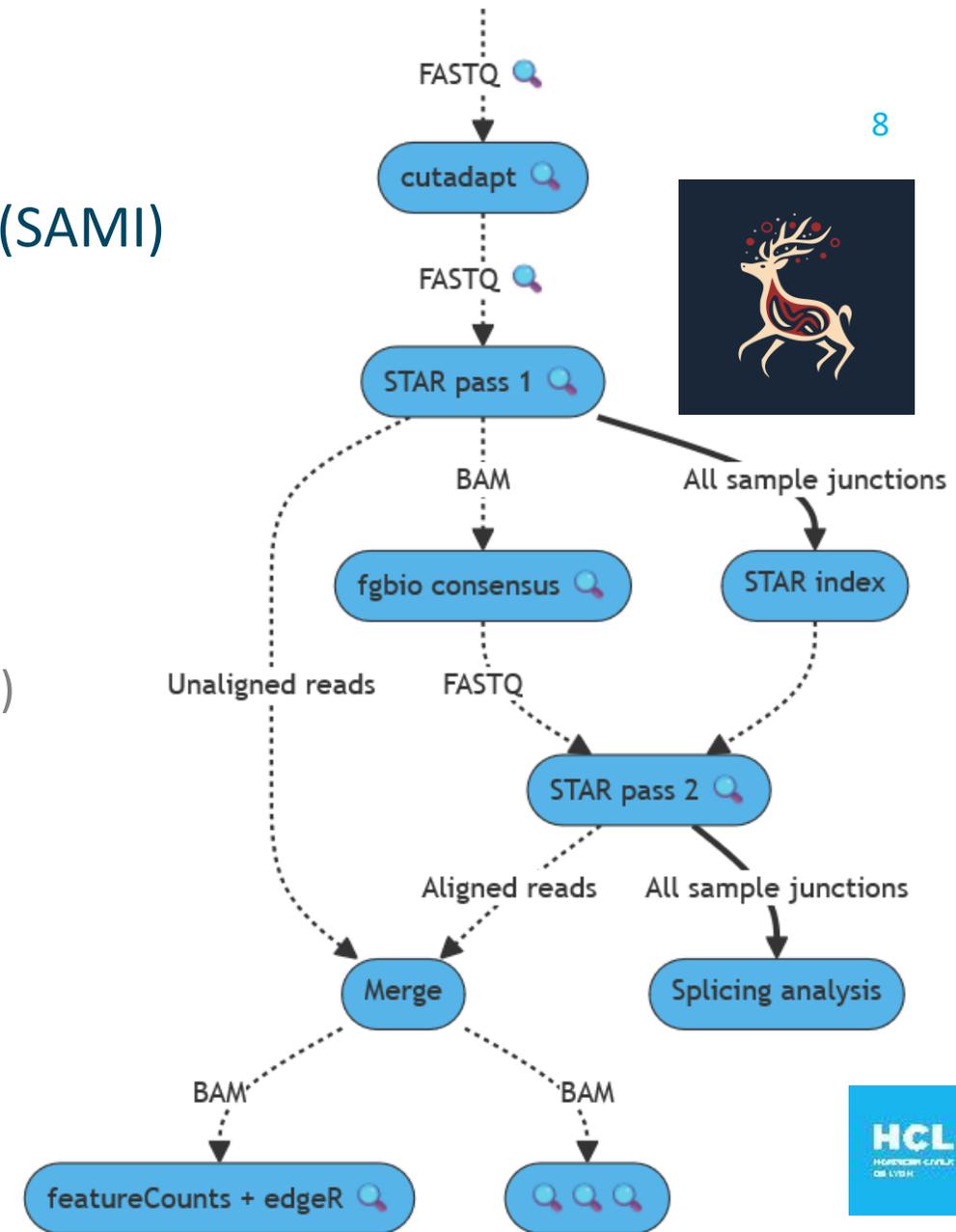
MET (XM_011516223) > 1 % out.seraseq



ANOMALIES D'ÉPISSAGE

SPLICING ANALYSIS WITH MOLECULAR INDEXES (SAMI)

- Des FASTQ aux résultats
- Extraction des UMI par **bcl-convert**
- **Alignement STAR 2-pass**
 - Collecte des jonctions pour uniformisation (pass 1)
 - Coordonnées pour déduplication (pass 1)
 - **Déduplication par consensus** avec fgbio
 - Alignement final (pass 2)
- Rapport **MultiQC**
 - FastQC, rnaSeqMetrics, soft-clipping, inserts...
- **Nextflow + Singularity** (HPC ready)



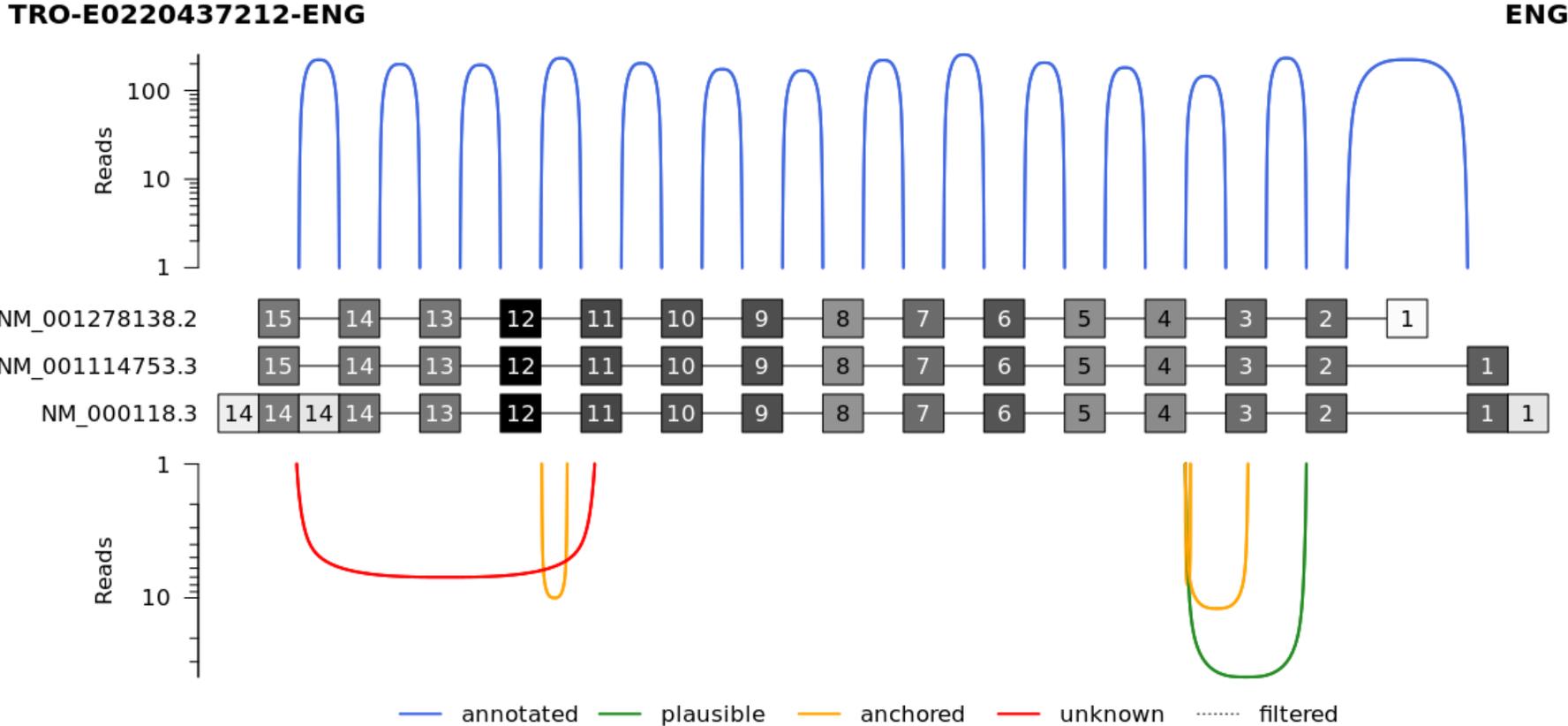
ANOMALIES D'ÉPISSAGE

SPLICING ANALYSIS WITH MOLECULAR INDEXES (SAMI)

- Jonctions collectées par STAR
- Annotation RefSeq
 - Classification annotated / plausible / anchored / unknown
- Filtration
 - Percent Spliced In (PSI)
 - Reads supportant la jonction
 - Récurrence
- Représentations graphiques par gène

ANOMALIES D'ÉPISSAGE

SPLICING ANALYSIS WITH MOLECULAR INDEXES (SAMI)



ID	junction	chrom	class	recurrence	sample	reads	fusion	left					right				
								position	genes	exons	depth	PSI	position	genes	exons	depth	PSI
DXG16	9:127826673-127843093	9	plausible	8	TRO-E0220437212-ENG_S14	40	FALSE	127826673	ENG (-)	4]	221	18,10 %	127843093	ENG (-)	[2	284	14,08 %

ANOMALIES D'ÉPISSAGE

RÉSULTATS SUR CONTRÔLE COMMERCIAL

- Micro-panel Cancer du poumon
 - 17,6 kb ciblés dans 21 gènes
 - Séquençage MiSeq 2x150 bp, 1M reads, UMI
 - Echantillons FFPE (Roche HyperPETE)
- Contrôle Seraseq (SeraCare)
 - 16 fusions
 - Saut de l'exon 14 de *MET*
 - Saut des exons 2 à 7 d'*EGFR* (EGFRvIII)
 - 7 réplicats

ANOMALIES D'ÉPISSAGE

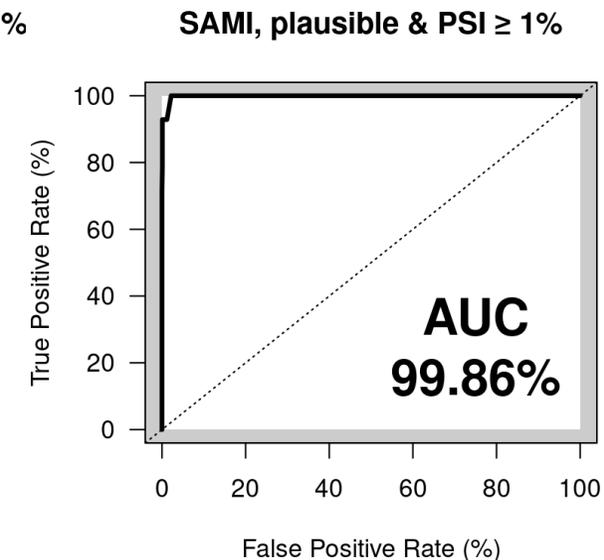
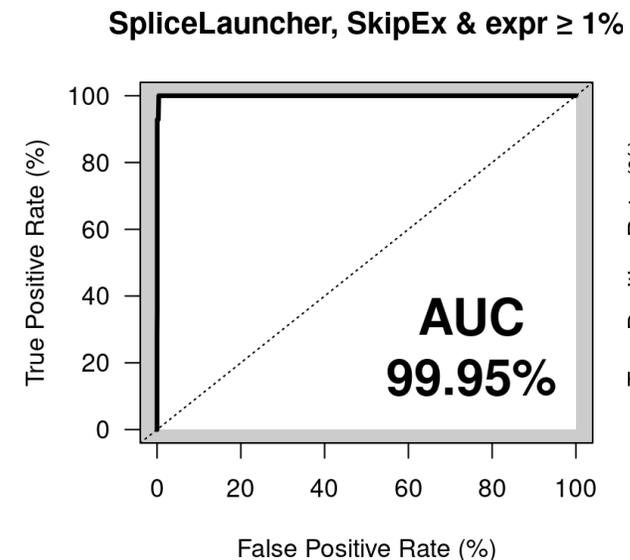
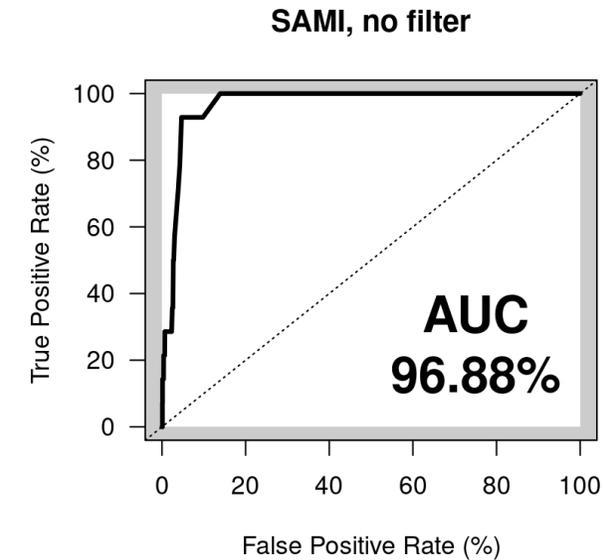
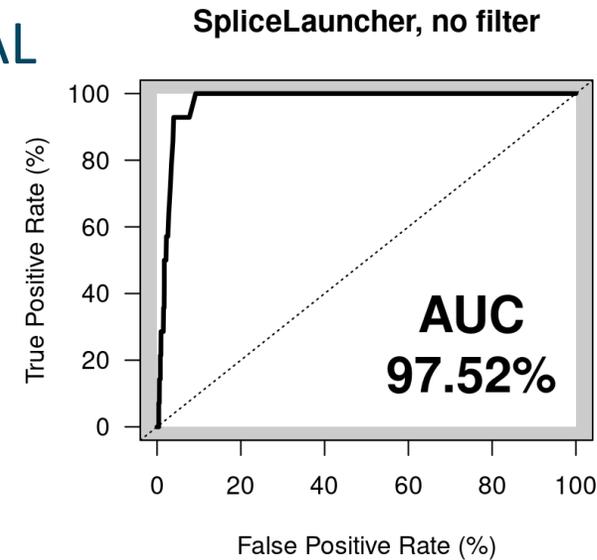
RÉSULTATS SUR CONTRÔLE COMMERCIAL

- Détection empirique
 - SAMI
 - 7/7 sauts d'exon 14 de MET
 - 7/7 sauts d'exon 2-7 d'EGFR
 - PSI \approx 100%
 - SpliceLauncher
 - 7/7 sauts d'exon 14 de MET
 - 7/7 sauts d'exon 2-7 d'EGFR
- Modèle statistique
 - SpliceLauncher
 - 0/2 modélisation réussie dans un pseudo-run de contrôles
 - 1/14 modélisation réussie dans les runs originaux (MET, $p=0$)
 - 0.2% à 1,5% des jonctions aberrantes modélisées en Gamma
 - 0,5% à 5,6% des jonctions aberrantes modélisées en binomiale négative

ANOMALIES D'ÉPISSAGE

RÉSULTATS SUR CONTRÔLE COMMERCIAL

- **Faux-positifs**
 - Seuil à 13/14 vrais positifs
 - SAMI (8+ reads)
 - 138 faux positifs sans filtre
 - **1 faux positif avec filtre**
 - « Plausible » & PSI > 1%
 - SpliceLauncher (11+ reads)
 - 117 faux positifs sans filtre
 - **1 faux positif avec filtre**
 - « SkipEx » & expr > 1%
- Spécificité ++ sur sauts d'exons
- Plus incertains sur le reste

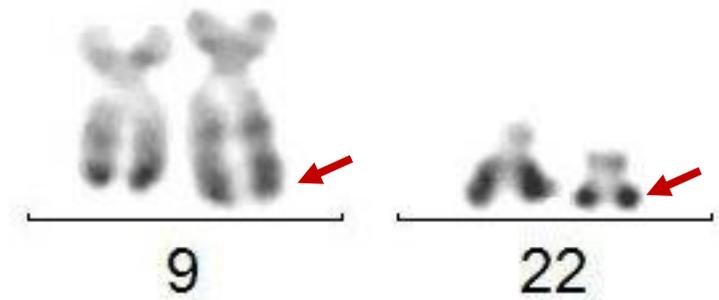


ANALYSE DE VARIANTS COMPLEXES

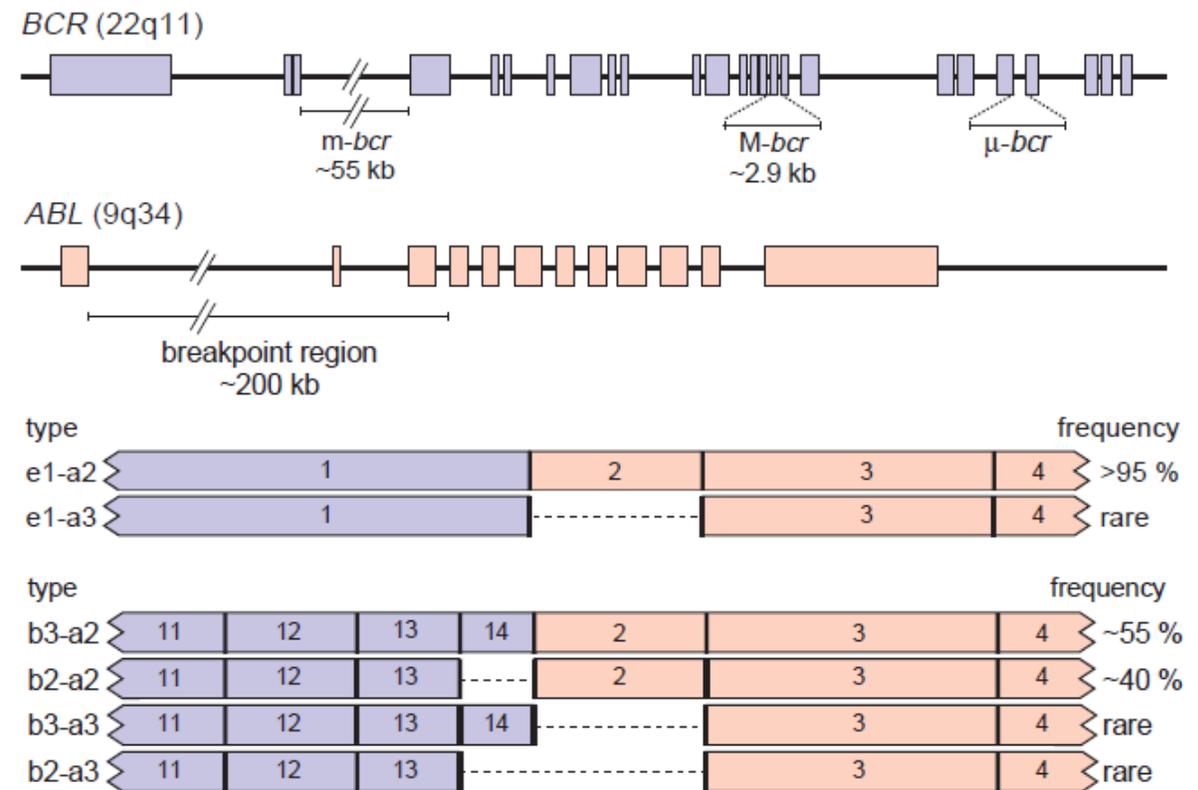
FUSIONS DE GÈNES

FUSIONS DE GÈNES

CONTEXTE : BCR-ABL1, LE CAS D'ÉCOLE



- Aspect **cytogénétique**
 - Translocation équilibrée t(9;22)
 - « Chromosome de Philadelphie » (1960)
- Aspect **moléculaire**
 - BCR : *Breakpoint Cluster Region*
 - ABL1 : *Abelson proto-oncogene 1*
 - Tyrosine kinase (mobilité, apoptose, réponse aux dégâts de l'ADN...)
 - Auto-inhibée (SH3/SH2, auto-P)
 - BCR-ABL1
 - Dimérisation & auto-activation
- Aspect **clinique**
 - >90% des Leucémies Myéloïdes Chroniques
 - Cible thérapeutique (Imatinib), suivi MRD



FUSIONS DE GÈNES

NF-CORE/RNAFUSION

● <https://nf-co.re/rnafusion>

- Nextflow DSL2
- Février 2019 → Avril 2024 (3.0.2)
- Pas d'article dédié, stabilité imparfaite mais développeurs réactifs
- Installation délicate derrière un pare-feu

● Agrégateur d'outils

- Arriba
- STAR-Fusion
- FusionCatcher

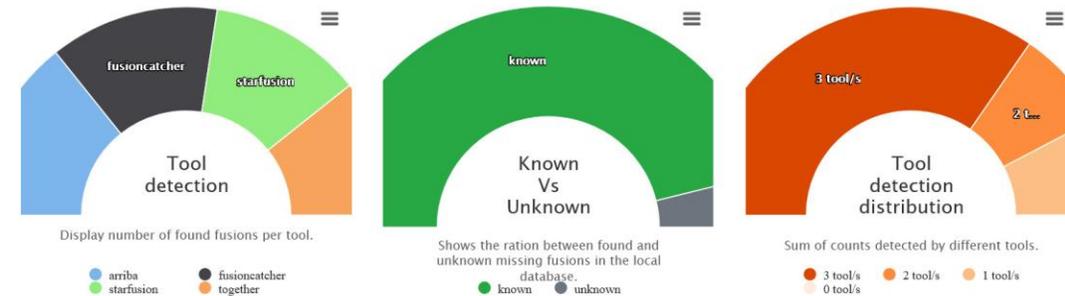
● Rapport MultiQC

● Bilan HTML par FusionReport

Summary

The pipeline ran successfully and was able to detect **13** fusions for sample **SERASEQ-RUN09**. We were able to find and describe **12** fusions from our local database. Tools used during the analysis were Fusioncatcher, Arriba, Starfusion.

Dashboard fusion summary



List of detected fusions

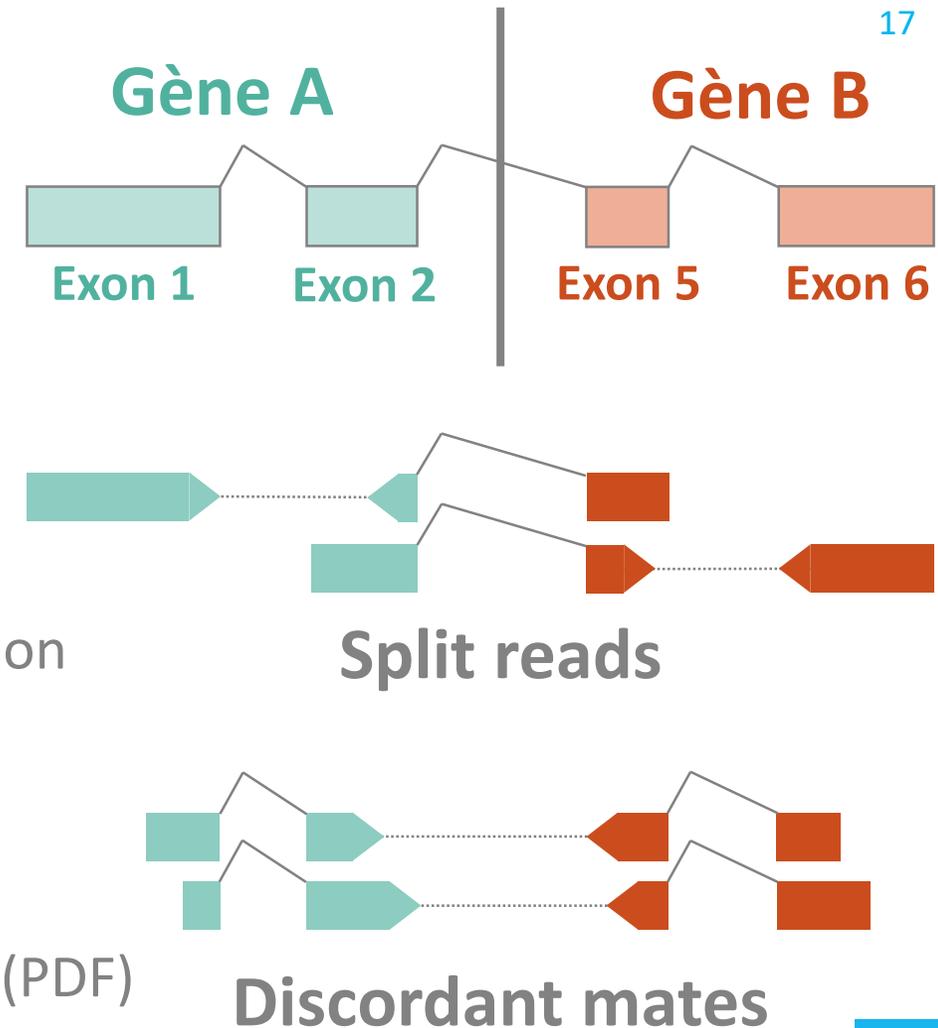
Filters fusions found by at least 1 tools. If number of chosen tools is less than 1 the filter is disabled. The whole list can be found in `./Report-/fusions.json`.

Fusion gene	Found in DB	arriba	fusioncat...	starfusion	Tools hits	Score
FGFR3--TACC3	COSMIC MitoIman	✓	✓	✓	3	0.688
TPH3--NTRK1	COSMIC MitoIman	✓	✓	✓	3	0.688
FGFR3--BATAP2L1	COSMIC MitoIman	✓	✓	✓	3	0.688
SLC45A3--BRAF	COSMIC MitoIman	✓	✓	✓	3	0.688
ETV6--NTRK3	COSMIC MitoIman	✓	✓	✓	3	0.688
TFG--NTRK1	COSMIC MitoIman	✓	✓	✓	3	0.688
EML4--ALK	COSMIC MitoIman	✓	✓	✓	3	0.688
SLC34A2--ROS1	COSMIC MitoIman	✓	✓	✓	3	0.688
CD74--ROS1	COSMIC MitoIman	✓	✓	✓	3	0.688
NCOA4--RET	COSMIC MitoIman	✓	✓	✗	2	0.625
KTF5B--RET	COSMIC MitoIman	✓	✗	✓	2	0.625
EGFR--SEPTIN14	MitoIman	✗	✓	✗	1	0.312
BRAF--EGFR	Not found	✓	✗	✗	1	0.156

FUSIONS DE GÈNES

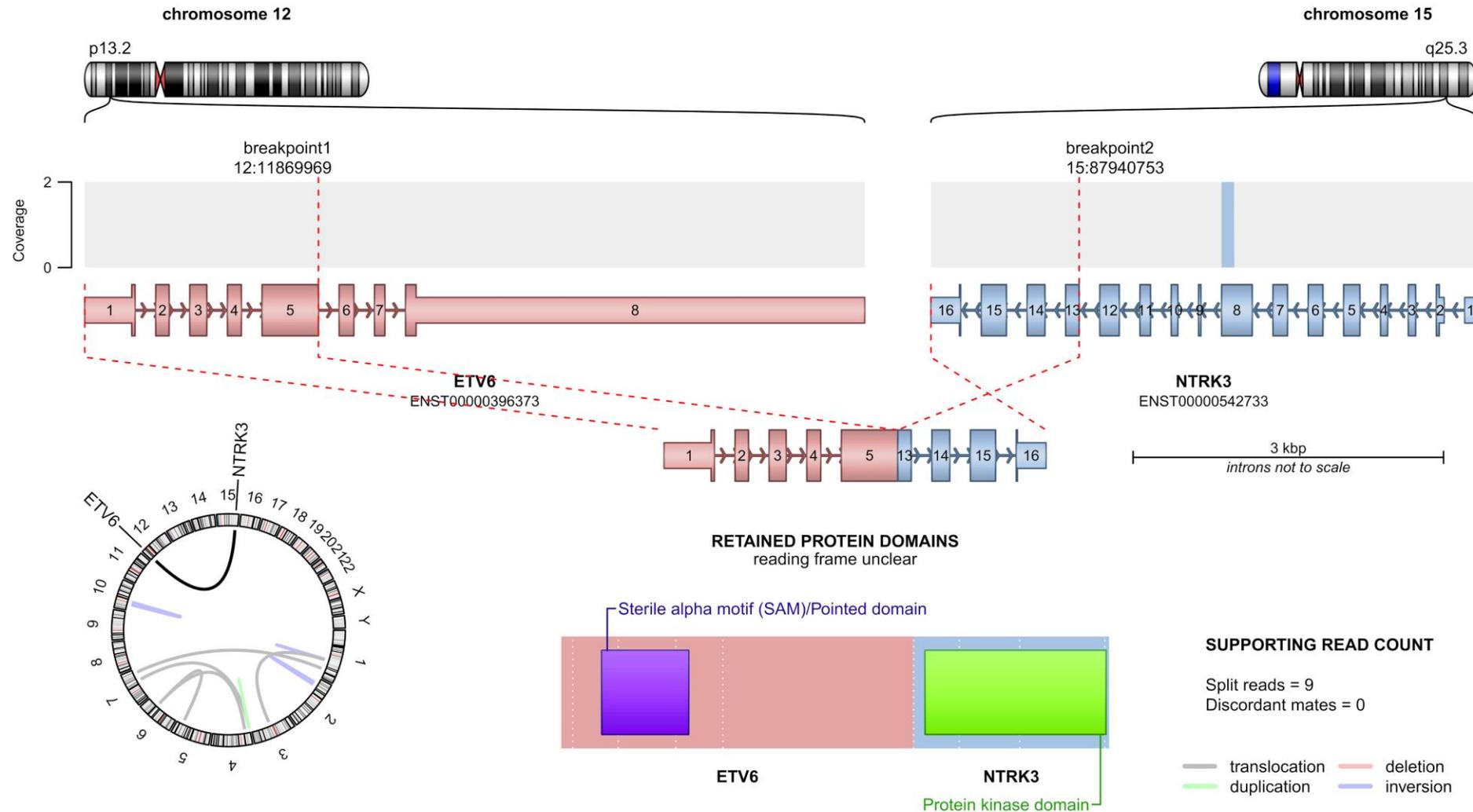
NF-CORE/RNAFUSION – ARRIBA

- **Arriba** : un bon candidat « stand alone »
 - Uhrig et al, Genome Res. 2021
 - Vainqueur du DREAM Challenge Fusion
 - F1=0,73; Creason et al, Cell syst 2021
 - Janvier 2020 → Février 2023 (version 2.4.0)
- **Alignement STAR + filtres** de Chimeric.out.junction
 - Décompte **split reads & discordant mates**
 - Support minimal (2 reads)
 - Support relatif (e-value empirique)
 - Liste noire, gènes homologues, homopolymères...
- Bonus : **représentations graphiques** des fusions (PDF)



FUSIONS DE GÈNES

NF-CORE/RNAFUSION – ARRIBA



FUSIONS DE GÈNES

NF-CORE/RNAFUSION – STAR-FUSION

- **STAR-Fusion** : prédécesseur d'Arriba
 - Haas et al, Genome Biol. 2019
 - 2e au DREAM Challenge Fusion
 - F1=0,70; Creason et al, Cell syst 2021
 - Avril 2015 → Janvier 2022 (version 1.13.0)
- **Alignement STAR + filtres** de Chimeric.out.junction
 - Pré-filtre reads dupliqués, régions homologues et fusions antisens
 - Décompte **split reads & discordant mates**
 - 2+ dont 1+ split pour des sites d'épissage annotés
 - 3+ split hors sites d'épissage annotés
 - Filtres supplémentaires basés sur l'annotation
 - Liste noire, paralogues, niveau d'expression > 0.1 FFPM...
- Outils compagnon : Fusion Inspector
 - Représentations graphiques basées sur IGV (référence chimérique)

FUSIONS DE GÈNES

NF-CORE/RNAFUSION – FUSIONCATCHER

- **FusionCatcher** : plus ancien, mais orthogonal

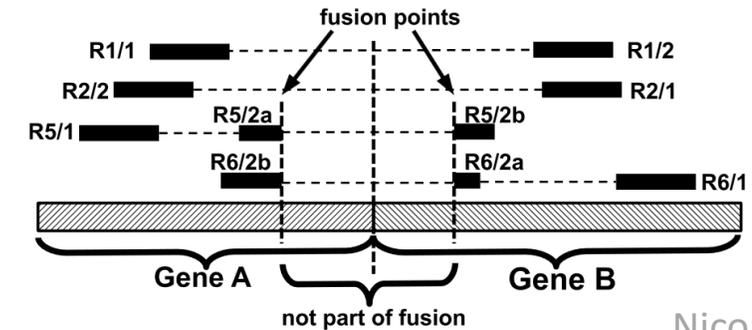
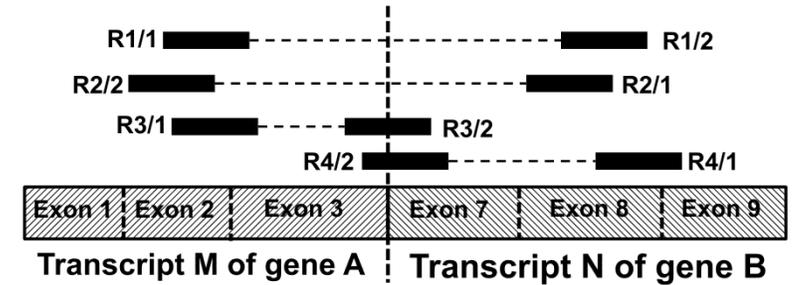
- Nicorici et al, bioRxiv 2014
- 4e au DREAM Challenge Fusion
 - F1=0,58; Creason et al, Cell syst 2021
- Juin 2013 → Janvier 2021 (version 1.33)

- **Multiples alignements**

- Première liste de candidats issue d'un alignement Bowtie sur transcriptome Ensembl
 - + filtre de contamination génomique, liste noire, paralogues...
- Second filtre par réalignement sur références chimériques
 - Bowtie sur toutes les combinaisons exon-exon possibles
 - BLAT / STAR / Bowtie2 sur l'union gène A - gène B (entiers)

- Moins gourmand en RAM (10 Go < 40 Go)

- Le temps de calcul augmente avec le nombre de candidats



Nicorici et al,
bioRxiv 2014

FUSIONS DE GÈNES

RÉSULTATS SUR CONTRÔLES COMMERCIAUX

- Micro-panel Cancer du poumon
 - 17,6 kb ciblés dans 21 gènes
 - Séquençage MiSeq 2x150 bp, 1M reads, UMI
 - Echantillons FFPE (Roche HyperPETE)
- Contrôle FFPE HD784 (Horizon)
 - 3 fusions
 - 5 réplicats
- Contrôle Seraseq (SeraCare)
 - 16 fusions (dont 1 hors panel)
 - 2 sauts d'exons
 - 7 réplicats

FUSIONS DE GÈNES

RÉSULTATS SUR CONTRÔLES COMMERCIAUX

Horizon (3 fusions x 5 réplicats)

version	caller	TP	P	sensitivity	FP	precision
1.2.0	squid	1	3	33,3 %	1	50,0 %
	ericscript	7	15	46,7 %	8	46,7 %
	pizzly	7	15	46,7 %	3	70,0 %
	fusioncatcher	12	15	80,0 %	1	92,3 %
	arriba	13	15	86,7 %	6	68,4 %
	starfusion	14	15	93,3 %	2	87,5 %
	ANY	14	15	93,3 %	20	41,2 %
3.0.1	fusioncatcher	14	15	93,3 %	11	56,0 %
	starfusion	14	15	93,3 %	3	82,4 %
	arriba	15	15	100,0 %	8	65,2 %
	ANY	15	15	100,0 %	20	42,9 %

SeraSeq (15 fusions x 7 réplicats)

version	caller	TP	P	sensitivity	FP	precision
1.2.0	squid	2	15	13,3 %	0	100,0 %
	pizzly	15	90	16,7 %	0	100,0 %
	ericscript	50	105	47,6 %	10	83,3 %
	fusioncatcher	73	105	69,5 %	0	100,0 %
	starfusion	97	105	92,4 %	0	100,0 %
	arriba	98	105	93,3 %	3	97,0 %
	ANY	101	105	96,2 %	13	88,6 %
3.0.1	starfusion	94	105	89,5 %	8	92,2 %
	fusioncatcher	100	105	95,2 %	1	99,0 %
	arriba	101	105	96,2 %	14	87,8 %
	ANY	102	105	97,1 %	17	85,7 %

- Squid, EricScript et Pizzly peu sensibles, plus disponibles en version 3
- Arriba, FusionCatcher et STAR-Fusion satisfaisants
 - Sensibilité ++ / Spécificité -- avec la version 3
 - Qualité moindre (Horizon) = plus de faux-positifs

RNA-SEQ : ANOMALIES D'ÉPISSAGE ET FUSIONS

CONCLUSION

- Anomalies d'épissage
 - [SpliceLauncher](#) disponible et performant sur sauts d'exons
 - [SAMI](#) bientôt disponible (2024)
 - Passage en DSL2 avant publication
 - Validation de méthode sur données in-silico (ART, Calib)
 - Détection de fusions et de mutations
- Fusions de gènes
 - [Arriba](#) seul plutôt satisfaisant
 - [nf-core/rnafusion](#) version 3 ajoute un peu de sensibilité
- Intérêt du long-read ?
 - Nanopore « direct RNA »
 - Challenges : capture & quantification, qualité de l'ARN issue de FFPE

MERCI

www.chu-lyon.fr



Plateforme NGS des HCL Equipe bioinformatique

Claire Bardel
Pierre-Antoine Rollat-Farnier
Thomas Simonet
Quentin Testard
Valentin Wucher

Service d'hématologie

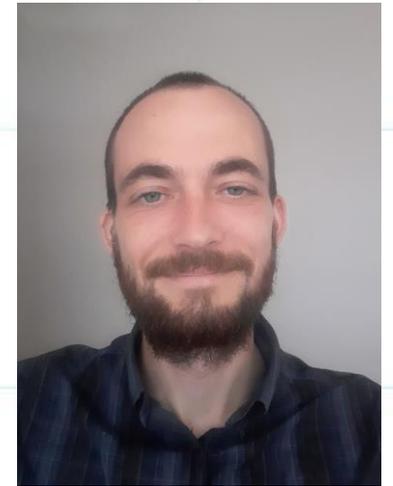
Sarah Huet
Sandrine Hayette
Kaddour Chabane

Service d'anatomopathologie

Marc Barritault
Pierre-Paul Bringuier
Camille Léonce

Service de génétique

Stéphane Pinson



HCL
HOSPICES CIVILS
DE LYON