
Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total

Table des matières

1	OBJET	3
2	CHAMP D'APPLICATION POUR LES PLATEFORMES DU PLAN FMG 2025	3
3	PRESENTATION DES 2 KITS D'EXTRACTION ARN FFPE SELECTIONNES.....	3
3.1	Critères de sélection.....	3
3.2	Comparaison des 2 kits sélectionnés et de leurs équivalents pour tissus congelés	4
4	PLAN D'EXPERIENCE	4
4.1	Echantillons	4
4.2	Expériences.....	5
4.3	Analyses bioinformatiques	6
4.4	Vérification par qPCR de la présence des fusions identifiées en NGS.....	7
5	RESULTATS DES EXPERIENCES	8
5.1	Extraction d'ARN total	8
5.2	Préparation des librairies <i>TruSeq® Stranded Total RNA Library Prep Gold</i>	9
5.3	Contrôle qualité post-séquençage	9
5.3.1	Taille des inserts	10
5.3.2	Alignement et comptage des séquences dupliquées.....	10
5.3.3	Répartition des séquences alignées	11
5.3.4	Cartographie génomique des séquences	12
5.4	Analyse des fusions de gènes	13
5.4.1	SeraSeq : 16 fusions de référence	13
5.4.2	Echantillons de patients	16
5.5	Pertes d'exons	20
6	VALIDATION PAR LES PLATEFORMES	21
7	CONCLUSION GENERALE : AVANTAGES ET INCONVENIENTS.....	22

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 2/23		

8	HISTORIQUE DES MODIFICATIONS	23
9	VALIDATION DU DOCUMENT	23

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 3/23		

1 OBJET

Le CRefIX a réalisé une évaluation de 2 kits d'extraction d'ARN issus d'échantillons FFPE (*Formalin-Fixed Paraffin Embedded*) afin d'identifier et de sélectionner le kit qui :

- Offrira une quantité et une qualité d'ARN après extraction suffisante pour construire les librairies ARN total ;
- Permettra une détection de fusion de transcrits se rapprochant le plus d'une analyse faite à partir d'échantillons FF (*Fresh Frozen*) ;
- Offrira une flexibilité et une facilité d'utilisation pour le déploiement possible sur les plateformes du plan FMG 2025;
- Aura un coût abordable.

2 CHAMP D'APPLICATION POUR LES PLATEFORMES DU PLAN FMG 2025

Cette recommandation n'est valable que pour une analyse finale de séquençage d'ARN total avec le kit de librairie *TruSeq® Stranded Total RNA Library Prep Gold* incluant le kit de ribodéplétion *Illumina Ribo-Zero plus rRNA Depletion Kit*, combiné aux indexes *IDT for Illumina – TruSeq RNA UD Indexes v2*.

3 PRESENTATION DES 2 KITS D'EXTRACTION ARN FFPE SELECTIONNES

3.1 Critères de sélection

Le kit commercial sélectionné doit :

- Être automatisable ;
- Être compatible avec des expériences NGS à partir d'échantillons FFPE ;
- Contenir le moins de produits CMR possible ;
- Conserver les mutations et fusions somatiques de l'ARN

Les 2 kits sélectionnés par le CRefIX pour cette évaluation sont les kits :

- RNeasy® FFPE Kit, Qiagen.
- Maxwell® RSC RNA FFPE Kit, Promega.

La sélection de ces 2 kits résulte du fait qu'ils remplissent tous les 2 les conditions précitées. De plus, le kit Qiagen semble être un standard utilisé dans plusieurs publications^{1,2}, et le kit Promega est le kit le plus utilisé sur les plateformes de diagnostic.

Afin d'évaluer ces kits, une comparaison avec leurs homologues utilisés pour l'extraction d'ARN total de tissus congelés a été effectuée, à savoir :

- RNeasy® Mini Kit, Qiagen.
- Maxwell® RSC simplyRNA Tissue Kit, Promega.

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 4/23		

3.2 Comparaison des 2 kits sélectionnés et de leurs équivalents pour tissus congelés

<u>FFPE</u>	RNeasy® FFPE Kit (Qiagen)	Maxwell® RSC RNA FFPE Kit (Promega)
Automatisable?	Oui (Qiacube)	Obligatoire (Maxwell)
Nombre scroll	Jusqu'à 4 scrolls de : épaisseur : 20µm max surface : 250 mm2 max	2mm3 total épaisseur : 5-10µm surface : 20-200 mm2
Déparaffinage	Deparaffinization solution (Qiagen)	Huile minérale
Type de purification	Colonne	Billes magnétiques
Digestion PK	56°C 1H	56°C 15min à 1H
DNase	Oui	Oui
Decrosslinking	80°C - 15min	80°C - 1h

<u>FF</u>	RNeasy® Mini Kit (Qiagen)	Maxwell® RSC simplyRNA Tissue Kit (Promega)
Automatisable?	Oui (Qiacube)	Obligatoire (Maxwell)
Quantité de matériel de départ (congelé)	30 mg max	De 10 à 20mg
Broyage et homogénéisation	Mécanique *	Mécanique *
Type de purification	Colonne	Billes magnétiques
DNase	Oui	Oui

* Type rotor-stator ou Mortier/Pilon

Tableau récapitulatif des caractéristiques techniques des kits comparés dans l'étude

4 PLAN D'EXPERIENCE

4.1 Echantillons

La comparaison des kits a été effectuée à partir de 2 biopsies solides de patients, ayant entre 50 et 60 ans, présentant une tumeur du poumon et d'un échantillon commercial standard. Ces échantillons de patients proviennent de la Biobanque CRB-Tumorotheque de Nice BB-0033-00025. Le séquençage WGS PCR free de ces échantillons de patients a révélé la présence de réarrangements

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 5/23		

génomiques de type fusions, argument dans la sélection de ces échantillons. L'échantillon *Seraseq® FFPE Tumor Fusion RNA v4 Reference Material* (Seracare, <https://www.seracare.com/Seraseq-FFPE-Fusion-RNA-RM-v4-0710-0496>) a été utilisé comme contrôle positif. Ce dernier possède 18 régions de transcrits de fusion (dont 2 transcrits alternatifs) représentatives de différents types de tumeurs solides, dont la fréquence d'apparition a été déterminée par dPCR et validée en NGS par le fabricant.

	PATIENT 05 *	PATIENT 10 *	Seraseq® FFPE Tumor Fusion RNA v4 Reference Material
Origine de l'échantillon	Biopsie solide de tumeur pulmonaire	Biopsie solide de tumeur pulmonaire	Lignée WT GM24385 + Transcrits synthétiques (18 régions de fusion de transcrits)
Cellularité tumorale mesurée en anatomopathologie	80 %	70 %	NA
Cellularité tumorale estimée en NGS	20 – 30 %	30 %	NA
Taille section	20µm	20µm	10µm
Nombre de sections utilisées	1	1	1
Type fixateur	Formol	Formol	Formol
% fixateur	10%	10%	Non transmis par le fabricant

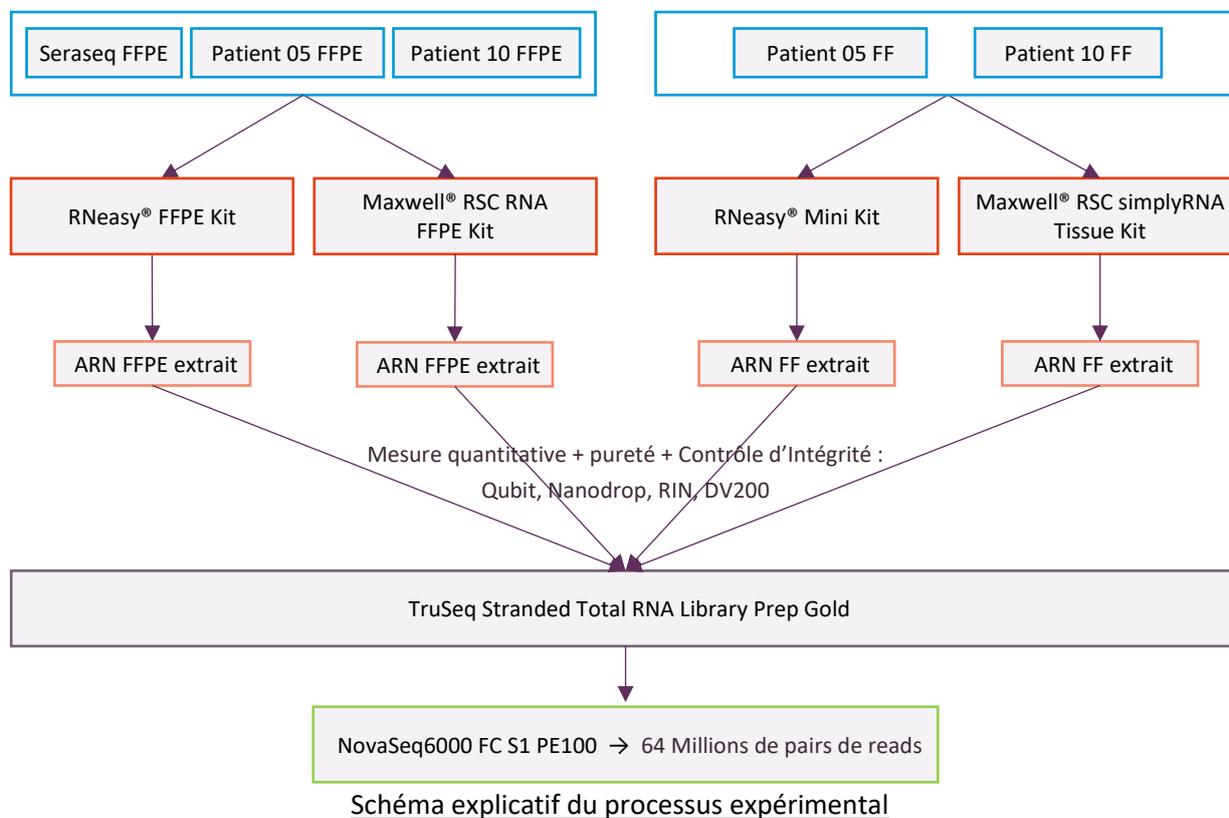
*Ces échantillons sont aussi disponibles en tissu congelé (FF)

Tableau récapitulatif des caractéristiques des échantillons utilisé dans l'étude

4.2 Expériences

Les kits d'extractions Promega nécessitent obligatoirement l'utilisation du Maxwell RCS Instrument. Ces extractions ont donc été automatisées grâce à cet instrument. Concernant les extractions Qiagen, elles ont été réalisées manuellement en suivant les recommandations du fabricant, mais il est également possible d'automatiser ces expériences avec le QIACUBE (non testé par le CRefIX). Le kit d'Illumina, *TruSeq® Stranded Total RNA Library Prep Gold* incluant le kit de ribodéplétion Illumina *Ribo-Zero plus rRNA Depletion Kit* de construction de bibliothèques RNA-Seq total est défini comme gold standard, en effet il est couramment utilisé sur les plateformes de recherche et de diagnostic en utilisant 100ng d'ARN total de départ. Les échantillons sont séquencés sur Novaseq 6000 avec une profondeur de 64 millions de clusters, comme visée par les plateformes du plan.

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 6/23		



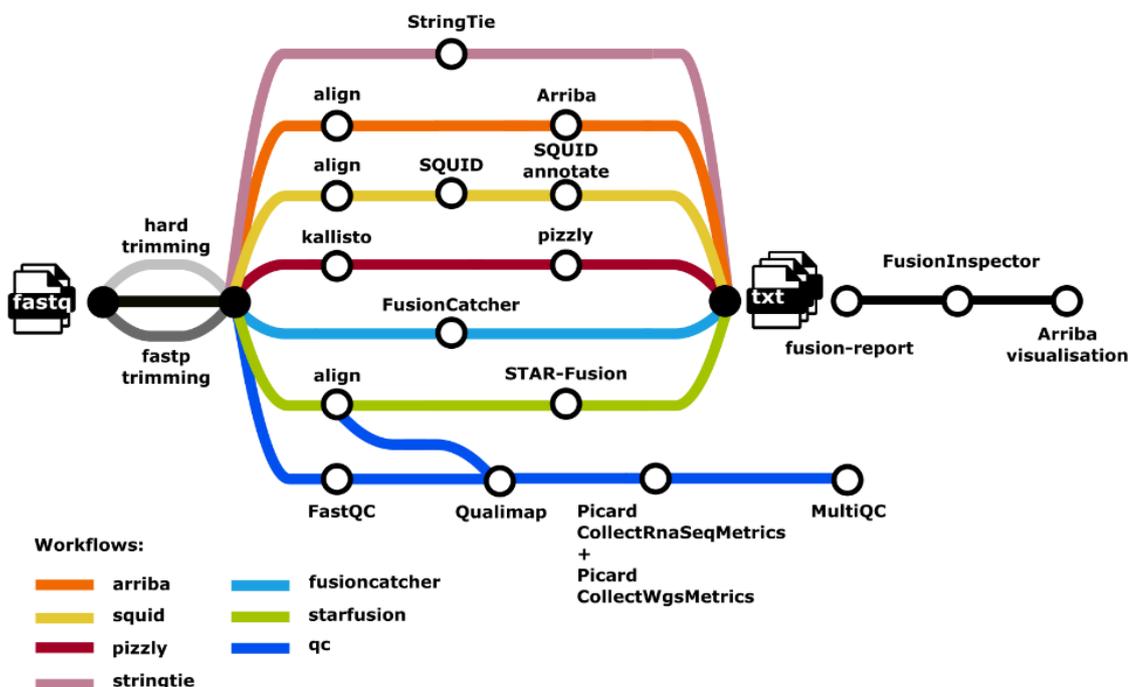
Toutes les extractions ont été réalisées en duplicats et des bibliothèques ont été préparées pour chacune d'entre elles.

4.3 Analyses bioinformatiques

- Dans un souci d'uniformité, tous les fichiers de séquences fastq sont downsampled à une profondeur de 64 millions de paires à l'aide de l'outil seqkit 2.4.0.
- Le contrôle qualité et l'analyse des fusions ont été effectués au moyen du pipeline nextflow (v23.04) *nf-core rnafusion v2.3.4*. Dans ce pipeline, les outils FastQC 0.11.9, Qualimap 2.2.2d, Picard 2.27.4 et MultiQC 1.14 ont permis d'effectuer le contrôle qualité tandis que Arriba 2.3.0, FusionCatcher 1.33 et STAR-Fusion 1.10.1 ont réalisé l'analyse de fusions. Le workflow mise en place est basé, entre autre, sur celui publié par *alsania, K., Shen, Tw., Chen, X. et al. Structural variant analysis of a cancer reference cell line sample using multiple sequencing technologies. Genome Biol 23, 255 (2022)*
- Le génome de référence utilisé est le hg38

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 7/23		

- Les statistiques et les figures sont générées à l'aide du logiciel R (v4.1.1).



Détail du pipeline nf-core/rnafusion

Suivant le pipeline nf-core/rnafusion présenté ci-dessus, le contrôle qualité est effectué indépendamment de l'analyse de fusion. Dans celle-ci seuls les trois outils sélectionnés (arriba, FusionCatcher et STARFusion) sont exécutés, les résultats sont concaténés par fusion-report. Les fusions retrouvées par au moins deux outils sur trois font l'objet de rapports détaillés tandis que les données brutes servent à la réalisation de statistiques et graphiques sous R.

4.4 Vérification par qPCR de la présence des fusions identifiées en NGS

Les fusions de transcrits identifiées à l'issue de ces analyses ont fait l'objet d'une vérification par qPCR, à partir des extractions Qiagen. Une reverse transcription a été faite à partir des extractions d'ARN sur tissus FF et FFPE, à l'aide du kit Transcriptor First Strand cDNA Synthesis Kit (Roche, 04896866001) + Random Primers. Des primers de qPCR entourant les points de fusions ont ensuite été synthétisés (2 couples par fusion) et utilisés pour vérifier l'amplification ou non des régions identifiées. Ces qPCR ont été effectuées à l'aide du kit KAPA Library Quantification Kit for Illumina (Roche, 07960140001), en triplicats.

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 8/23		

5 RESULTATS DES EXPERIENCES

5.1 Extraction d'ARN total

	Kits FFPE		Kits FF	
	<i>RNeasy® FFPE Kit (Qiagen)</i>	<i>Maxwell® RSC RNA FFPE Kit (Promega)</i>	<i>RNeasy® Mini Kit (Qiagen)</i>	<i>Maxwell® RSC simplyRNA Tissue Kit (Promega)</i>
Quantité d'ARN total extrait (Qubit)	1 scroll : 9605ng (6310-14152) Importante	1 scroll : 4203ng (743-7918) Suffisante	8 à 15mg : 12470ng (6325-19352) Importante	6 à 8mg : 4548ng (1431-10702) Suffisante
Absorbance 260/280 (Nanodrop)	2,0 (1,9-2,0) Très Bonne	2,15 (2,1-2,2) Très Bonne	2,1 (2,1-2,1) Très Bonne	2,1 (2,1-2,1) Très Bonne
Valeur du RIN (Bioanalyzer)	2,95 (2,4-3,9) Mauvaise	3,3 (1-4,3) Mauvaise avec variabilité	5,6 (4-7,7) Mauvaise	4,8 (3,2-6,5) Mauvaise avec variabilité
Valeur du DV200 (Bioanalyzer)	74% (56-80) Bien pour FFPE	70% (50-86) Bien pour FFPE	88% (85-94) Très bien	88% (80-94) Très bien
Reproductibilité entre duplicats	Très Bonne	Satisfaisante	Très Bonne	Satisfaisante
Facilité d'utilisation	Très simple	Simple	Très simple	Simple
Temps total estimé	~ 2H30	~ 3H30	~ 2H	~ 2H

Résultats basés sur les échantillons de patients uniquement

Tableau synthétique des performances d'extraction pour les deux kits étudiés

Les différents critères listés dans le tableau ci-dessus montrent que les résultats obtenus à partir des kits Qiagen et Promega sont relativement similaires. A noter tout de même de meilleurs rendements d'extractions avec Qiagen, et un outlier présentant un RIN de 1 avec Promega.

Par ailleurs, bien que les extractions Qiagen aient été réalisées à la main, au contraire de celles réalisées avec Promega, le kit semble plus facile à utiliser. On soulignera que l'automatisation du Maxwell ne couvre pas tout le protocole, dont la nécessité de réaliser plusieurs transferts des échantillons pour réaliser les extractions Promega. Ceci ajoute des étapes chronophages, ainsi qu'un risque non négligeable d'erreurs.

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 9/23		

5.2 Préparation des bibliothèques *TruSeq® Stranded Total RNA Library Prep Gold*

	<u>Kits FFPE</u>		<u>Kits FF</u>	
	<i>RNeasy® FFPE Kit (Qiagen)</i>	<i>Maxwell® RSC RNA FFPE Kit (Promega)</i>	<i>RNeasy® Mini Kit (Qiagen)</i>	<i>Maxwell® RSC simplyRNA Tissue Kit (Promega)</i>
Quantité de bibliothèques obtenues (qPCR) *	94,8nM (60,3-118,1)	92nM (74,5-106,9)	142,7nM (121,4-181,6)	131,2nM (76,2-201,6)
Tailles et qualités des bibliothèques **	322pb (317-336)	311pb (297-320)	358pb (342-384)	359pb (330-391)
Reproductibilité entre duplicats	Très bonne	Très bonne	Très bonne	Très bonne
Similarité FF Vs FFPE	Très bonne	Très bonne	Très bonne	Très bonne

Résultats basés sur les échantillons de patients uniquement

* : concentration minimale= 2nM

** : taille attendue dans les recommandations Illumina (~260 pb)

Récapitulatif des performances de la construction des bibliothèques de séquençage pour les différents kits

Les quantités de bibliothèques obtenues sont similaires et très importantes indépendamment du kit d'extraction utilisé. La taille des bibliothèques est petite et induit un *overlap* de séquençage. La conservation de ces petits fragments est indispensable pour préserver au mieux la diversité du transcriptome des échantillons FFPE. On précisera aussi que toutes les bibliothèques (FFPE comme FF) présentent des dimères d'adaptateurs, dus aux kits de préparation de bibliothèques, que l'on retrouve après séquençage dans des proportions inférieures à 5%. Les résultats entre duplicats d'échantillons sont similaires. La comparaison des kits FFPE avec leurs homologues FF montre que les bibliothèques réalisées à partir d'échantillons FFPE sont légèrement plus petites et moins concentrées que celles réalisées à partir d'échantillon FF, que ce soit en utilisant les kits d'extractions Qiagen comme ceux de Promega.

5.3 Contrôle qualité post-séquençage

Le contrôle qualité et l'analyse des fusions ont été effectués sur un sous-échantillonnage des fichiers fastq à 64 millions de paires de *reads*. Les échantillons ont ensuite été traités en deux groupes (échantillons contrôles SeraSeq et échantillons de patients) par le processus « qc » de nf-core/rnafusion, composé de FastQC, Qualimap et Picard. Ces logiciels permettent d'évaluer la qualité générale des séquences, fournissant de nombreuses informations et statistiques sur le % de GC, le %

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 10/23		

d'adaptateurs, la taille d'insert de la librairie. Les résultats des échantillons SeraSeq et patients ont ici été rassemblés, les résultats sont comparables entre eux.

Les rapports de contrôle qualité (QC) des échantillons ont été regroupé dans un unique fichier grâce à l'outil MultiQC.

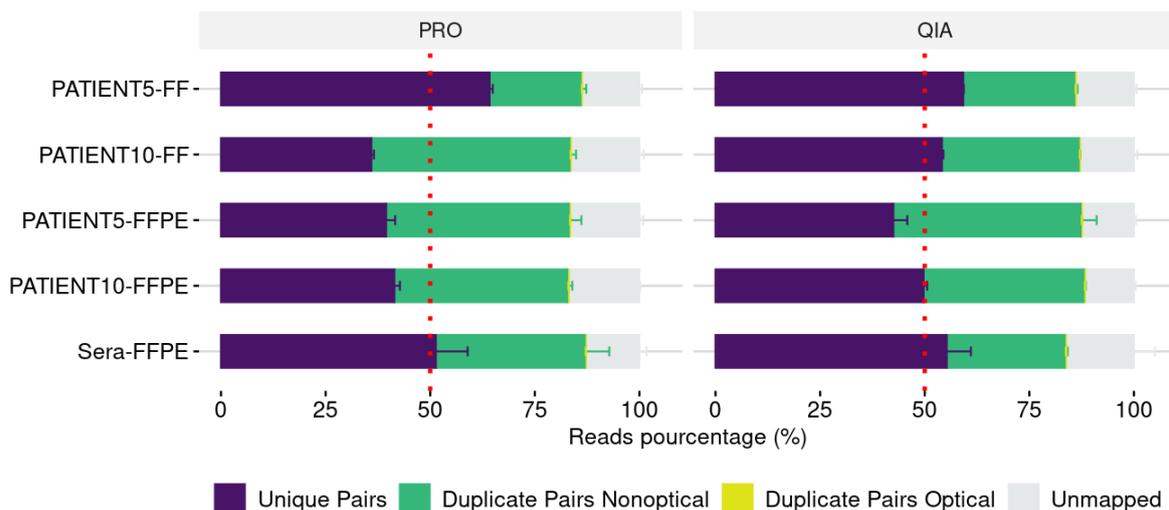
5.3.1 Taille des inserts

Tissu	Promega (PRO)	Qiagen (QIA)
FF	151 (149-153)	150 (149-153)
FFPE	141 (134-170)	142 (139-155)

Taille des inserts médiane FF et FFPE par kit d'extraction

La taille des inserts (taille des molécules d'ARN séquencées) des librairies TruSeq séquencées depuis chacun des kits d'extraction est homogène, bien que l'on observe une perte d'environ 10 paires de base entre les échantillons FF et FFPE.

5.3.2 Alignement et comptage des séquences dupliquées



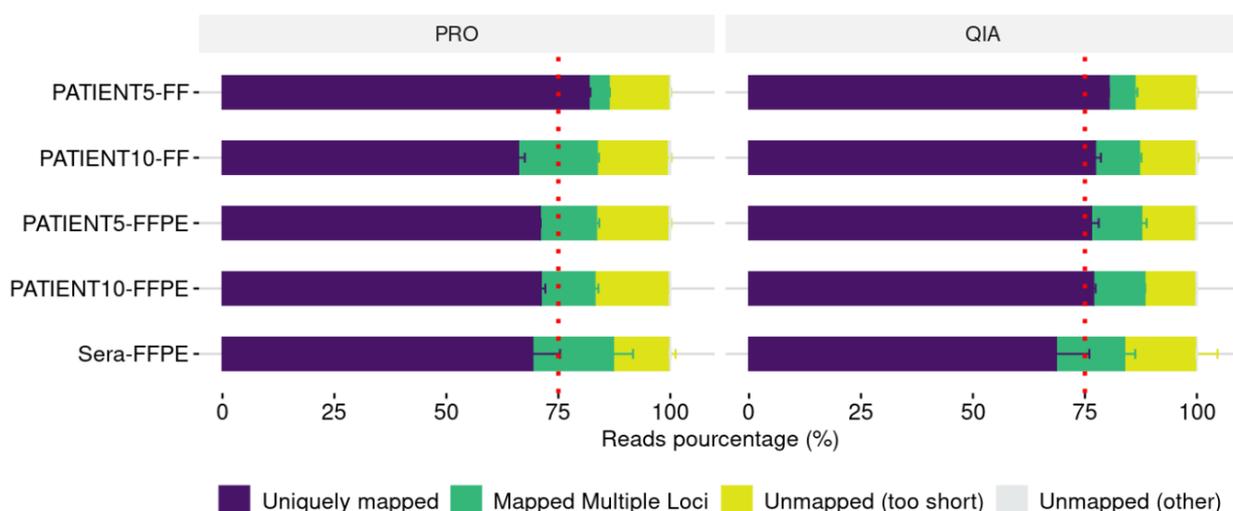
Tissu	Kit Extraction	Unique Pairs	Duplicates Pairs Nonoptical	Duplicates Pairs Optical	Unmapped Pairs
FF	Promega	50.06 %	34.53 %	0.34 %	15.06 %
FF	Qiagen	56.65 %	29.60 %	0.34 %	13.40 %
FFPE	Promega	44.08 %	40.16 %	0.35 %	15.41 %
FFPE	Qiagen	49.16 %	37.04 %	0.36 %	13.44 %

Taux de duplicats des échantillons

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 11/23		

Le graphique ci-dessus permet de visualiser le nombre de duplicats de PCR introduits durant l'étape d'amplification de la librairie. Le taux de duplicats est comparable entre les deux kits d'extraction, bien qu'il soit, comme attendu, légèrement supérieur pour les échantillons FFPE (+ 6-9%). Le pourcentage de séquences mappées (*Unique Pairs*), également représenté sur le graphique, est un bon indicateur pour mesurer la performance globale du séquençage. Ici, le kit d'extraction Qiagen fournit respectivement + 6% de séquences uniques pour le tissu FF et + 5% pour le tissu FFPE, directement lié à un pourcentage de duplicats plus faible. Ce résultat implique un plus grand nombre de séquences utiles pour l'analyse du transcriptome avec le kit Qiagen.

5.3.3 Répartition des séquences alignées



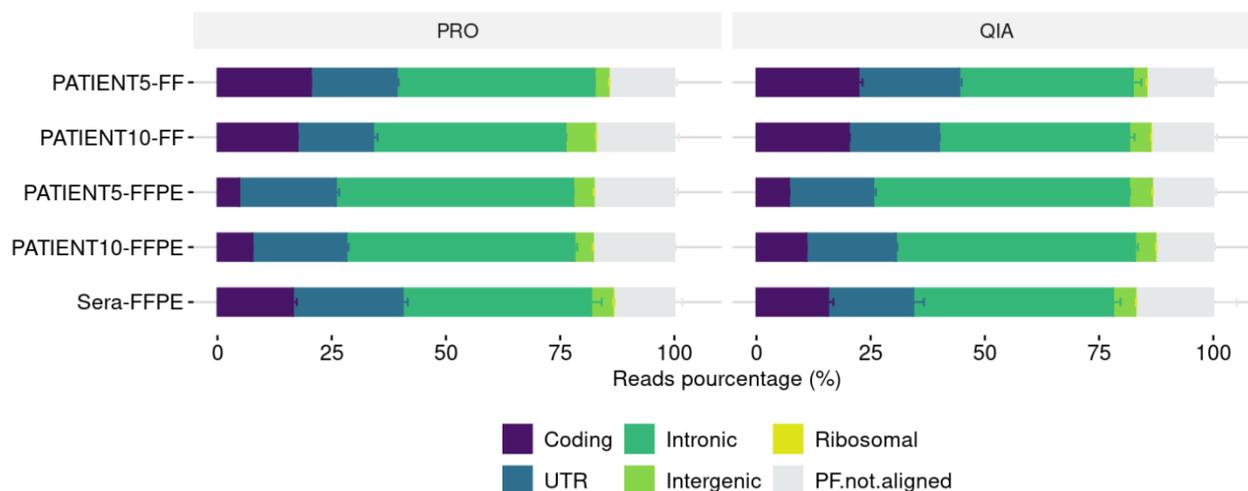
Tissu	Kit Extraction	Uniquely mapped	Mapped multiple loci	Unmapped (too short)	Unmapped (other)
FF	PRO	73.90 %	11.04 %	14.47 %	0.60%
FF	QIA	78.79 %	7.81 %	12.92 %	0.48%
FFPE	PRO	70.40 %	14.19 %	14.87 %	0.54%
FFPE	QIA	73.93 %	12.64 %	12.91 %	0.53%

Statistiques d'alignement des séquences

Le taux d'alignement des reads *Uniquely mapped* est légèrement plus élevé avec le kit Qiagen par rapport au kit Promega (+ 3-4%), ce qui confirme un nombre plus grand de séquences utiles pour la suite de l'analyse avec le kit Qiagen.

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 12/23		

5.3.4 Cartographie génomique des séquences



Tissu	Kit Extraction	Coding	UTR	Intronic	Intergenic	PF not aligned	Ribosomal	Genes
FF	PRO	18.96 %	17.67 %	42.71 %	4.71 %	15.76 %	0.18 %	21 204
FF	QIA	21.28 %	20.90 %	39.81 %	3.73 %	14.10 %	0.19 %	21 650
FFPE	PRO	9.65 %	21.91 %	47.69 %	4.38 %	16.26 %	0.10 %	19 135
FFPE	QIA	11.25 %	18.90 %	50.66 %	4.74 %	14.33 %	0.13 %	19 718

Régions d'alignement des bases séquencées

La cartographie génomique des séquences sur le génome de référence est une visualisation permettant de vérifier les proportions de séquences codantes, non codantes ou intergénomiques dans nos échantillons. La version v41 de GENCODE a été utilisée pour l'annotation des transcrits. Cette répartition permet également de vérifier le pourcentage de contamination des ARN ribosomiaux restant dans les échantillons. Ici, ce dernier est extrêmement faible, ce qui indique que la déplétion des rRNA a fonctionné correctement lors de la préparation des bibliothèques.

Globalement, la répartition génomique des séquences est en adéquation avec la distribution théorique d'un séquençage issu d'ARN total observée pour les échantillons FF, c'est-à-dire environ 60% de régions introniques et environ 20% de régions codantes. La proportion de séquences codantes dans les échantillons FFPE est plus faible que celle observée dans leur homologues FF, en cohérence avec leur état de dégradation.

Cependant, une différence dans la proportion de séquence codantes entre les deux kits d'extraction testés est notable. En effet, le kit Promega génère une proportion de séquences codantes plus faible (FFPE : 9.7%) que le kit Qiagen (FFPE : 11.3%), et ce pour les deux types de bibliothèques testés : FF & FFPE.

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 13/23		

Cette observation est directement liée au fait que le nombre de séquences « non alignées » est plus important pour le kit Promega (16.3 % contre 14.3% pour Qiagen).

L'extraction des échantillons FFPE avec le kit Qiagen fournit un nombre de séquences alignées utiles plus important, et par conséquent une cartographie génomique sensiblement plus efficace dans les régions codantes.

5.4 Analyse des fusions de gènes

L'une des applications du séquençage d'ARN est d'identifier des transcrits de fusion, qui sont des altérations moléculaires consécutives, entre autre, à des cassures et réarrangements chromosomiques. L'analyse des transcrits de fusion est effectuée sur trois échantillons : l'échantillon de référence Seracare FFPE, qui présente 16 fusions caractérisées et sur deux échantillons de patients.

Les fusions considérées comme validées sont celles qui ont été retrouvées par au moins deux des trois callers de nf-core/rnafusion sélectionnés : Arriba, FusionCatcher et STARfusion. Les transcrits de fusion détectés par un seul outil sont donc considérés comme faux positifs et éliminés de l'analyse. De même, les fusions réelles détectées par un seul outil ont été considérées comme faux négatifs au même titre que celles n'étant détectées par aucun (affichées dans le graphique des faux négatifs pour faciliter le comptage).

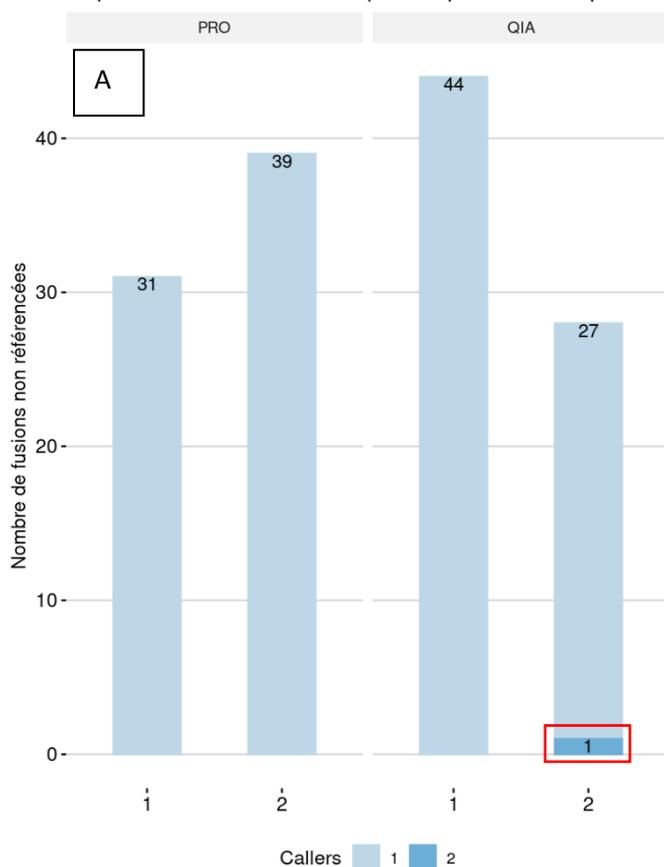
5.4.1 SeraSeq : 16 fusions de référence

Pour évaluer l'impact des kits d'extraction sur la performance de détection lors de l'analyse des gènes de fusion, les indicateurs de Précision, de Sensibilité et de F1-Score ont été calculés. Ce dernier correspond à la moyenne harmonique de la Précision et de la Sensibilité.

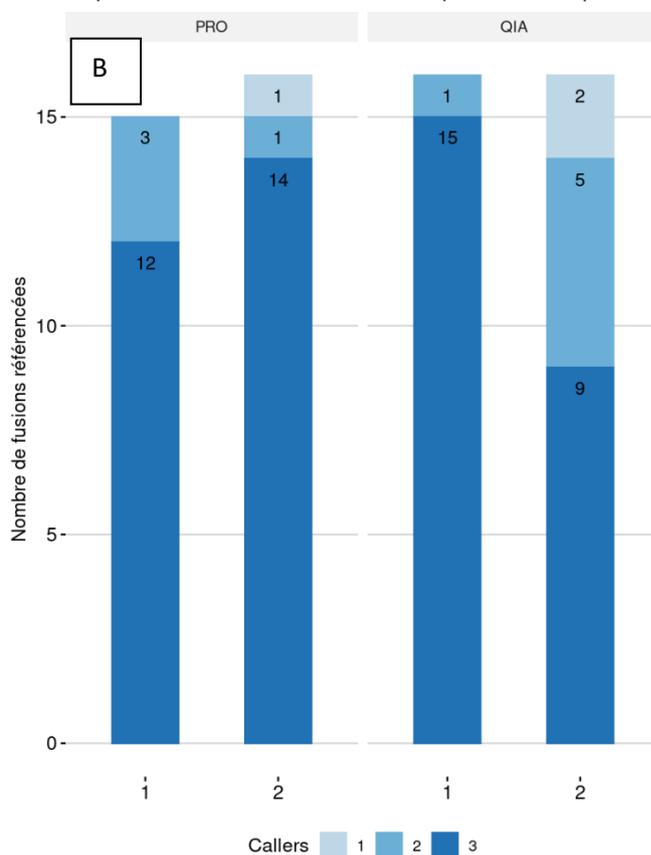
Le nombre de faux positifs (à gauche) et de vrais positifs (à droite) sont représentées par les graphes ci-dessous. En abscisse, les résultats pour chaque réplicat (1 et 2) de chaque kit d'extraction ont été ajoutés.

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 14/23		

Répartition des fusions faux positifs pour SeraSeq



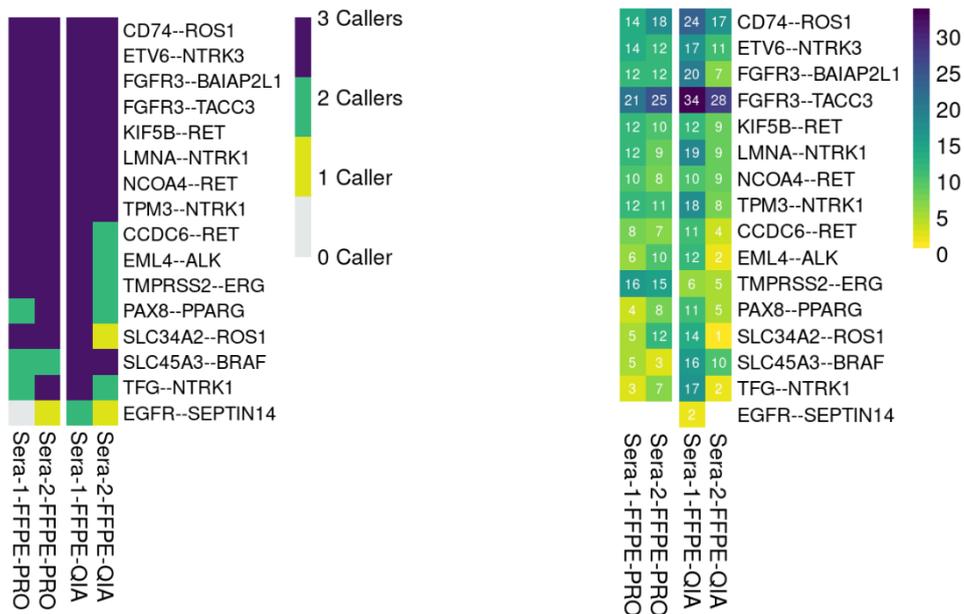
Répartition des fusions référencées pour SeraSeq



Concernant les faux positifs, le nombre de fusions appelées par 1 seul caller est comparable pour les 4 conditions expérimentales, allant de 27 à 44 (figure A). Une seule fusion a cependant été retrouvée par un deuxième outil et est donc considérée comme un réel « faux positif » dans le calcul de la précision pour l'extraction Qiagen du réplikat n°2 (encadré rouge dans la figure A).

Les 16 fusions référencées n'ont été retrouvées que dans un seul des quatre échantillons (réplikat 1 – Qiagen). Les faux négatifs, fusions non trouvées par plus de 2 callers, de deux échantillons (réplikat 2 – Promega ; réplikat 2 – Qiagen) ont été trouvés par un seul outil et pour le dernier échantillon contenant un faux négatif (réplikat 1 – Promega), la fusion n'a été retrouvée par aucun des trois outils et est donc impossible à « récupérer ».

La récurrence des 16 fusions référencées en fonction du nombre de caller est représentée par les heatmap ci-après, le nombre de support split read du logiciel Arriba y est également ajouté pour chaque kit d'extraction afin d'estimer le nombre de reads supportant la fusion.



Représentation en heatmap des 16 fusions référencées

Le nombre moyen de support split read (read1 + read2) est supérieur à 10 pour les fusions référencées, avec pour chaque réplicat :

Echantillon	Moyenne Support Read Split (Arriba)
Sera-1-FFPE-PRO	10,2
Sera-2-FFPE-PRO	10,4
Sera-1-FFPE-QIA	15,2
Sera-2-FFPE-QIA	7,9

Nombre de support split read par réplicat

Le nombre de support reads est plus homogène pour le kit Promega, et plus hétérogène pour le kit Qiagen entre les deux réplicats.

Pour résumer, les performances globales sur les fusions de références sont :

Kit extraction	Sensibilité VP / (VP + FN)	Précision VP / (VP + FP)	F1 score
FFPE Promega (Rép 1)	93,8% (15/16)	100%	96,8%
FFPE Promega (Rép 2)	93,8% (15/16)	100%	96,8%
FFPE Qiagen (Rép 1)	100% (16/16)	100%	100%
FFPE Qiagen (Rép 2)	87,5% (14/16)	93,3% (1 fusion)	90,3%

Performances d'analyse de fusions (trouvées par 2 callers et plus)

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 16/23		

Malgré les variations liées entre réplicats, les deux kits semblent avoir des performances comparables en termes de fusions de référence.

Les différences de résultats obtenus entre les 4 réplicats, que soit pour le nombre de faux positifs (identifiés par un seul caller) ou le nombre de vrais positifs (validés par au moins 2 callers), s'expliquent par le nombre de reads utiles de chaque conditions (PRO-1 : 41M, PRO-2 : 48M, QIA-1 : 48M, QIA-2 : 39M) : plus le nombre de reads utiles est élevé, plus le nombre de vrais positifs et de faux positifs/artefacts sont également élevés.

Concernant les fusions non référencées (figure A), identifiées par 1 seul caller, le nombre de Support Reads est très largement inférieur à 3. Ce qui appuie l'hypothèse de fusions « faussement positives », sauf pour les 3 fusions ci-dessous qui sont surreprésentées, en particulier pour la fusion IGHV3-48--AHCTF1 du réplicat 2 de l'extraction Promega.

Fusions	PRO		QIA	
	Sera-FFPE-1	Sera-FFPE-2	Sera-FFPE-1	Sera-FFPE-2
<i>NBEAL2--NBEAL2</i>	23	13	14	24
<i>EGFR--AC069287.1</i>	0	6	4	5
<i>IGHV3-48--AHCTF1</i>	0	75	0	0

Tableau des fusions non référencées, identifiées par 1 seul caller et surexprimées

5.4.2 Echantillons de patients

5.4.2.1 Patient 5

Le patient 5 présente majoritairement des fusions détectées avec un seul caller, du même ordre de grandeur de ce qui a pu être observé sur la référence SeraSeq (figure A).

Une seule fusion, RPS6KB1--VMP1, récurrente, est identifiée avec les 3 callers, mais uniquement dans les échantillons FF. Néanmoins, seul un réplicat FF du kit d'extraction Promega ne la détecte pas. Cette fusion est quant à elle totalement absente des librairies FFPE.

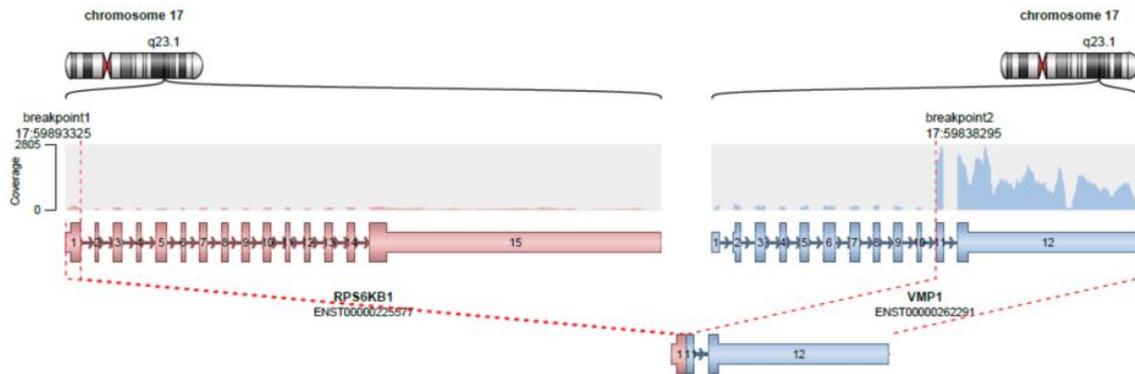
Cette fusion a été validée par qPCR, confirmant sa présence dans les librairies FF et FFPE.

	Noms Fusions	Moyenne des Ct des réplicats				Différences de Ct avec H ₂ O	Détection de la fusion par qPCR
		FF	Ctrl négatif	Ctrl positifs			
NICE 5 FF	RPS6KB1-- VMP1	28,27	32,07	21,18	17,98	3,79	Oui
		28,31	35,00			6,69	
NICE 5 FFPE	RPS6KB1-- VMP1	30,33	32,07	27,64	25,77	1,74	Oui
		35,00	35,00			0,00	

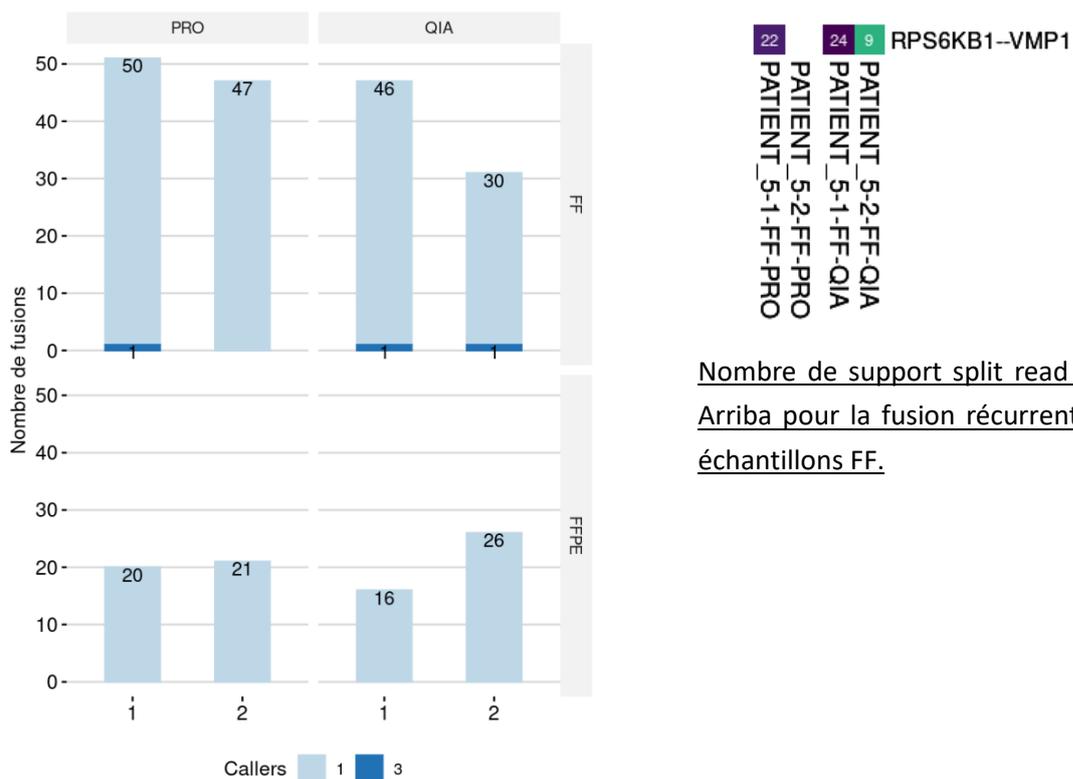
Résultats des valeurs de Ct mesurées par qPCR pour la fusion du patient 5

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 17/23		

Cette fusion de duplication intra-chromosomique sur le chr17 est connue dans différentes base de données : Mitelman, FusionGDB et FusionGDB2.



Représentation de la fusion RPS6KB1--VMP1 à l'aide du logiciel Arriba



Nombre de support split read du logiciel Arriba pour la fusion récurrente dans les échantillons FF.

Comptage de fusions détectées en fonction du nombre de caller selon les librairies FF ou FFPE et les kits d'extraction

Le nombre de séquences utiles plus faible pour les librairies FFPE, avec un pourcentage de séquences codantes seulement de 5% contre 22% pour les librairies FF, il est fort probable que la détection de la fusion RPS6KB1--VMP1 ne soit pas assez sensible lors de l'analyse bio-informatique pour ces librairies

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 18/23		

FFPE. Il est à noter que les puretés des échantillons sont faibles (20% FF et 30% FFPE), ce qui peut complexifier la détection.

5.4.2.2 Patient 10

Pour ce patient, la présence de 6 fusions détectées en NGS, par au moins 3 callers dans un ou plusieurs échantillons (voir graphe page 17), a été testée par qPCR (voir tableau ci-après). Dans l'échantillon FF, 5 fusions ont été validées par la qPCR, l'une fut en échec lors de l'amplification. Dans l'échantillon FFPE, 3 sont positives, l'une en échec et 2 sont négatives.

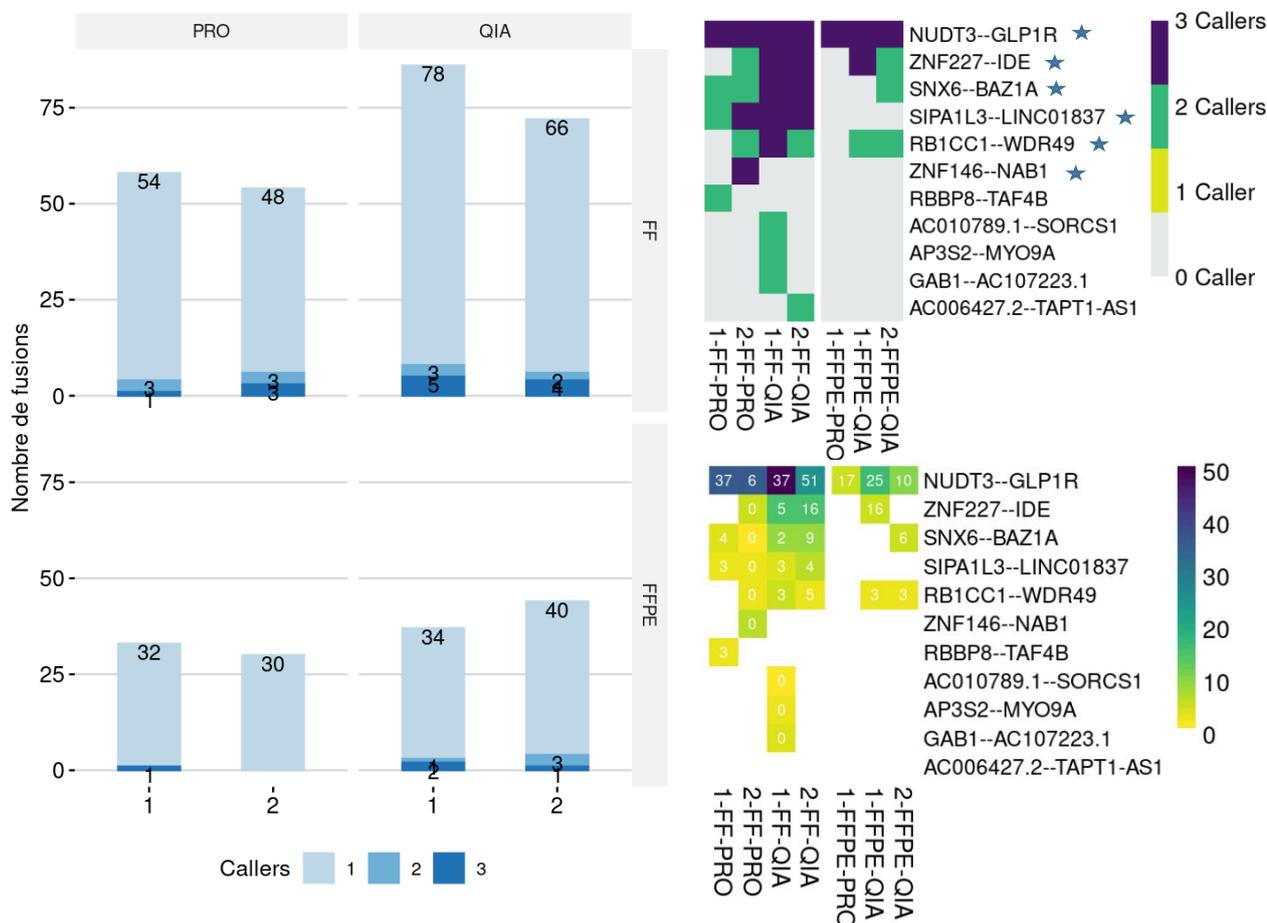
Il est à noter que certaines fusions n'ont été vérifiées que par un seul couple de primer en raison de la difficulté de synthétiser des primers optimaux entourant le point de fusion.

	Noms Fusions	Moyenne des Ct des réplicats				Différences de Ct avec H2O	Détection de la fusion par qPCR	
		FF	Ctrl négatif		Ctrl positifs			
			Ctrl H2O	Ctrl GAPDH	Ctrl ACTB			
NICE 10 FF	NUDT3-- GLP1R	26,75	35,00	21,37	20,68	8,25	Oui	
		26,88	35,00			8,12		
	RB1CC1-- WDR49	29,47	35,00			5,53	Oui	
	SIPA1L3-- LINC01837	35,00	35,00			0,00	Echec amplification	
	SNX6-- BAZ1A	30,49	35,00			4,51	Oui	
		30,78	35,00			4,22		
	ZNF146-- NAB1	30,92	35,00			4,08	Oui	
	ZNF227--IDE	28,59	34,68			6,09	Oui	
	27,56	33,92	6,35					
NICE 10 FFPE	NUDT3-- GLP1R	31,89	35,00	25,07	24,18	3,11	Oui	
		31,88	35,00			3,12		
	RB1CC1-- WDR49	31,60	35,00			3,40	Oui	
	SIPA1L3-- LINC01837	35,00	35,00			0,00	Echec amplification	
	SNX6-- BAZ1A	35,00	35,00			0,00	Non	
		35,00	35,00			0,00		
	ZNF146-- NAB1	35,00	35,00			0,00	Non	
	ZNF227--IDE	31,11	34,68			3,57	Oui	
	29,30	33,92	4,62					

Tableau répertoriant les valeurs de Ct mesurées par qPCR pour les fusions détectées par NGS

Lors de l'analyse bio-informatique, une fusion (NUDT3—GLP1R) apparaît dans tous les échantillons (FF & FFPE) hormis le réplicat FFPE n°2 du kit d'extraction Promega. Ce dernier ne présente aucune fusion validée par au moins deux callers (et n'a pas été représenté sur la heatmap page 19).

Ceci semble indiquer une meilleure performance et régularité lors de l'extraction avec le kit Qiagen pour cet échantillon.



Comptage de fusions détectées en fonction du nombre de caller selon les bibliothèques FF ou FFPE et les kits d'extraction

Représentation en heatmap des fusions du patient 10, selon le nombre de callers et le comptage des splits reads du logiciel Arriba

L'évaluation de la performance de la sensibilité de détection des fusions est basée sur les 5 validées en qPCR (★)

	FFPE Promega (Rép 1)	FFPE Promega (Rép 2)	FFPE Qiagen (Rép 1)	FFPE Qiagen (Rép 2)
Sensibilité	20%	0%	60%	80%
VP / (VP + FN)	(1/5)	(0/5)	(3/5)	(4/5)

Performance de la sensibilité de détection des 5 fusions validées par qPCR

Le pourcentage de région codante pour les bibliothèques FFPE est plus important pour le kit Qiagen, ~11% contre 7% pour le kit Promega. Ceci peut expliquer la plus grande sensibilité de détection, et donc la meilleure performance, pour les bibliothèques extraites avec le kit Qiagen.

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 20/23		

5.4.2.3 Synthèse de la performance de la détection de fusions

		FF		FFPE	
		Nombre de fusions détectées par au moins 2 callers	Nombre de fusions détectées par qPCR	Nombre de fusions détectées par au moins 2 callers	Nombre de fusions détectées par qPCR
Patient 5	Promega	1/2	1	0/2	1
	Qiagen	2/2		0/2	
Patient 10	Promega	7/10	5	1/6	3
	Qiagen	8/10		6/6	

Les bibliothèques FFPE extraites avec le kit Qiagen sont celles qui présentent les meilleures performances pour la détection des fusions.

5.5 Pertes d'exons

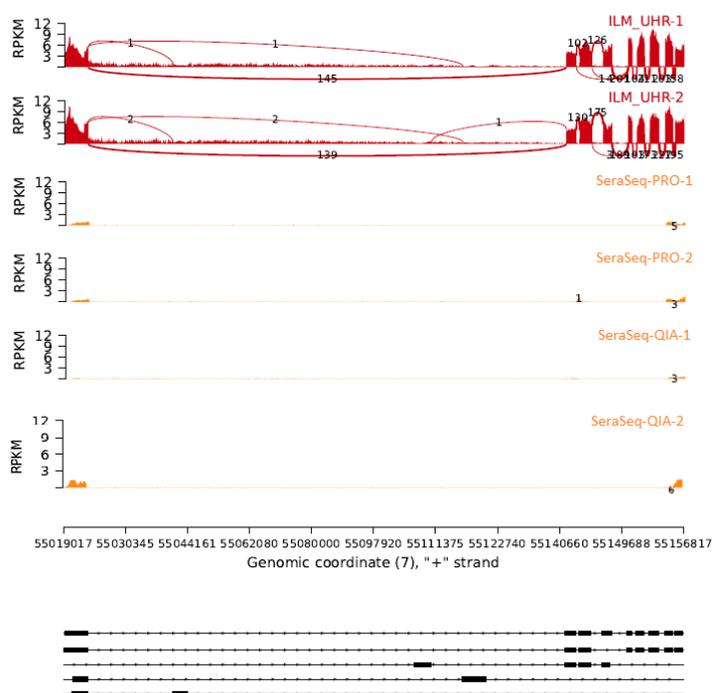
Le RNA-Seq peut révéler des variations de séquences dans les transcrits dues à des pertes d'exons et fournir ainsi les informations sur l'expression génique spécifique à un isoforme. Cette caractérisation est une notion importante dans l'analyse du RNA-Seq.

L'échantillon standard de référence SeraSeq FFPE utilisé est constitué de 2 transcrits alternatifs que sont EGFR VIII (délétion des exons 2 à 7) et MET *exon skipping* 14, en plus des 16 fusions étudiées précédemment.

Ces transcrits alternatifs ont été visualisés au moyen du logiciel *rmats2sashimiplot* (v2.0.4)³, à partir des fichiers bam, issus de l'alignement par le logiciel STAR. A des fins de contrôles positifs, l'ARN de référence humain universel, UHR, est utilisé lors de l'analyse. Cet échantillon de référence est composé d'ARN total provenant de 10 lignées cellulaires humaines, ses séquences ont été récupérées via le site BaseSpace d'Illumina (basespace.illumina.com).

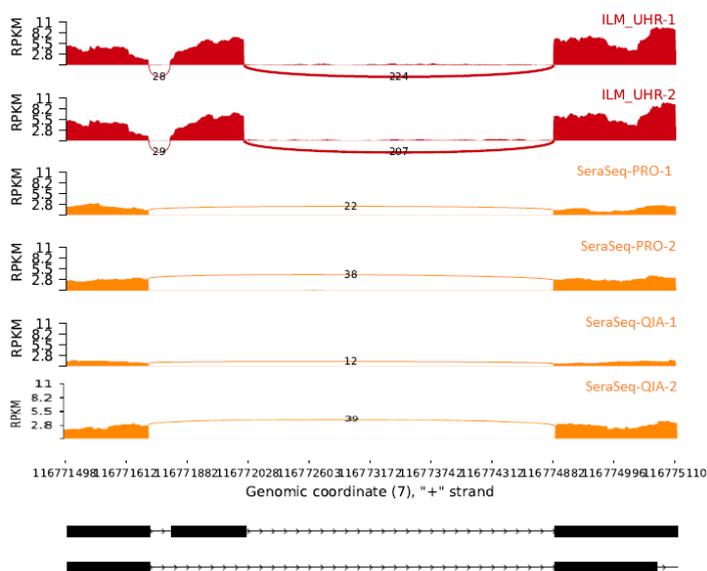
N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 21/23		

7:55019017:55156819:+



EGFR variant III

7:116771497:116775112:+



MET exon skipping 14

Visualisation des deux transcrits alternatifs selon le logiciel rmats2sashimplot

L'*exon skipping* de MET est détecté pour les deux extractions, en revanche la délétion plus importante d'EGFR s'avère plus problématique. L'épissage alternatif d'EGFR n'est identifié dans aucun des cas, quel que soit le kit d'extraction. Le faible nombre de séquences observé (3 – 6 reads) de part et d'autre de la délétion pourrait expliquer le manque de puissance et donc une sensibilité moins performante pour ce gène.

Une analyse complémentaire augmentant le nombre de séquences initiales utiles dans les échantillons FFPE de 118 à 126 millions de paires de reads au lieu de 64 millions a été réalisée. Cependant, aucune différence dans les résultats n'a pu être observée, notamment dans le cas d'EGFR variant III.

6 VALIDATION PAR LES PLATEFORMES

Le CRefIX n'étant pas un organisme d'accréditation, les tests de répétabilité, de robustesse, de limite de détection (non applicable) n'ont pas été faits.

Les plateformes désirant utiliser ces kits d'extractions devront refaire une validation de méthode.

Le CRefIX n'a pas réalisé de test en utilisant une plateforme robotique.

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 22/23		

7 CONCLUSION GENERALE : AVANTAGES ET INCONVENIENTS

		RNeasy® FFPE Kit (QIAGEN)	Maxwell® RSC RNA FFPE Kit (Promega)
Extraction d'ARN total	Quantité extraite	Elevée	Suffisante
	Pureté ARN (A260/280)	Très bonne	Très bonne
	RIN	Mauvais	Mauvais avec variabilité
	DV200	Corrects	Corrects
	Reproductibilité	Correcte	Médiocre
	Facilité d'utilisation	Bonne	Complexe et chronophage
Création de librairies	Quantité (qPCR)	Elevée	Elevée
	Qualité (BioA)	Bonne	Bonne
	Similarité FF Vs FFPE	Très proche	Très proche
Analyse bioinformatique	Alignements uniques	74%	70%
	F1 score détection de fusions (SeraCare)	90%-100%	97%-97%
	Variants d'épissage	MET détecté, pas EGFR VIII	MET détecté, pas EGFR VIII

Tableau synthétique des performances des kits sur l'extraction, l'efficacité de création de librairies et sur la puissance de l'analyse bio-informatique.

L'évaluation des 2 kits montre une similarité des résultats, les rendant tout 2 utilisables pour ce type d'analyse. En revanche, le kit Qiagen présente une performance globale sensiblement supérieure au kit Promega, notamment au niveau de la quantité et reproductibilité des extractions, ainsi que des alignements.

Pour cette raison, le CRefIX recommande l'utilisation du RNeasy FFPE Kit de QIAGEN pour l'extraction d'ARN total à partir d'échantillons FFPE, en vue de réaliser une analyse de fusions de transcrits.

N°: 2024xxxx_RT10 Version : 01 Date : 04/2024	Recommandation Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	Comparaison de 2 kits d'extraction ARN FFPE pour un séquençage d'ARN total	
Page 23/23		

DOCUMENTS

- RNeasy® FFPE Kit, Qiagen.
 - For purification of total RNA from formalin-fixed, paraffin-embedded tissue sections (July 2021)
- Maxwell® RSC RNA FFPE Kit, Promega.
 - Instructions for Use of Product AS1440 (Revised 7/22 TM436)
- RNeasy® Mini Kit, Qiagen.
 - For purification of total RNA from animal cells, animal tissues, bacteria, and yeast, and for RNA cleanup (October 2019)
- Maxwell® RSC simplyRNA Tissue Kit, Promega.
 - Instructions for Use of Products AS1390 and AS1340 (Revised 4/22 TM416)
- TruSeq® Stranded Total RNA Library Prep Gold
 - TruSeq Stranded Total RNA Reference Guide Ref Document # 1000000040499 v00 (October 2017)
- IDT for Illumina – TruSeq RNA UD Indexes v2
 - Index Adapters Pooling Guide Ref Document # 1000000041074 v12
- Seraseq® FFPE Tumor Fusion RNA v4 Reference Material
 - <https://www.seracare.com/Seraseq-FFPE-Fusion-RNA-RM-v4-0710-0496>

Références :

1. Jacobsen, S. B., Tfelt-Hansen, J., Smerup, M. H., Andersen, J. D. & Morling, N. Comparison of whole transcriptome sequencing of fresh, frozen, and formalin-fixed, paraffin-embedded cardiac tissue. *PloS One* 18, e0283159 (2023).
2. Newton, Y. et al. Large scale, robust, and accurate whole transcriptome profiling from clinical formalin-fixed paraffin-embedded samples. *Sci. Rep.* 10, 17597 (2020).
3. Shen, S., et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S A.* 111 (51), E5593-E5601 (2014).

8 HISTORIQUE DES MODIFICATIONS

Version	Date	Nature de la modification
V01		Relecture
V00	08/08/2023	Création

9 VALIDATION DU DOCUMENT

RÉDACTEURS	RELECTEURS
Kévin Gorrichon, Jasmin Cévest, Thomas Eychenne, Mélanie LETEXIER	Jean François Deleuze, Alain Viari, Violette Turon
PLAN EXPERIMENTAL	Fonction : Direction Unité
Kévin Gorrichon	Date :
ANALYSE	Visa :
Jasmin Cévest, Mélanie Letexier	