

Evaluation de l'impact des tissus fixés au *Formalin-Fixed Paraffin Embedded* (FFPE) par comparaison aux tissus congelés (FF) en WGS PCR-free (Truseq DNA PCR-free, Illumina)

Table des matières

1	OBJET	2
2	CHAMP D'APPLICATION POUR LES PLATEFORMES DU PLAN FMG 2025	2
3	PRESENTATION DES ECHANTILLONS TUMORAUX.....	2
4	PLAN D'ACTION	3
4.1	Plan d'expérience	3
4.2	Matériels & Méthodes bioinformatiques	3
5	RESULTATS.....	4
5.1	Extraction ADN : quantités et qualités obtenues	4
5.2	Création des librairies WGS PCR-free : quantités et qualités librairies obtenues	4
5.3	Validation Technique de séquençage	4
5.4	Validation Biologique.....	6
5.4.1	Calcul du pourcentage de cellules tumorales	6
5.4.2	Analyse des variants	6
5.4.3	Distribution des variants SNVs et InDels dans les échantillons FFPE et FF	7
5.4.4	Répartition des VAF des variants SNVs et InDels dans les échantillons FFPE et FF	8
5.4.5	Signature mutationnelle	11
5.4.6	Hotspots: catalogue Cancer Genome Interpreter (2700 hotspots)	14
5.4.7	Copy Number Alterations	15
5.4.8	Variants structurels.....	16
6	CONCLUSION GENERALE	19
7	DOCUMENTS	20
8	HISTORIQUE DES MODIFICATIONS.....	20
9	VALIDATION DU DOCUMENT	21

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	FFPE en WGS Truseq DNA PCR-free	
Page 2/21		

1 OBJET

Il a été demandé au CRefIX de comparer les variants somatiques (leur nombre, position, etc) obtenus en WGS PCR-free avec le kit d'Illumina, Truseq DNA PCR-free (TRU) entre des échantillons cancer FFPE et leur équivalent en tissus frais/congelés (FF). Le but : identifier les artéfacts introduits par le formaldéhyde dans les échantillons FFPE par rapport à des échantillons FF.

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

Cette recommandation n'est valable que pour une analyse finale de WGS PCR-free.


3 PRESENTATION DES ECHANTILLONS TUMORAUX

Pour cette étude, 4 patients ayant un cancer du poumon ont été sélectionnés. Pour chaque patient, 3 types d'échantillons ont été utilisés:

- du tissu congelé tumoral (échantillon FF),
- des rubans FFPE tumoraux, miroir avec le tissu congelé,
- du PBMC (cellules mononuclées du sang périphérique).

Ces échantillons proviennent de la Biobanque CRB-Tumorothèque de Nice BB-0033-00025 et ont entre 50 ans et 60 ans. Ces données ont été collectées dans le respect de la norme RGPD.

	PATIENT_05	PATIENT_08	PATIENT_10	PATIENT_11
Type de cancer	Adénocarcinome papillaire	Adénocarcinome micro-papillaire	Adénocarcinome solide	Adénocarcinome entérique
Fumeur ?	oui	oui	oui	oui
Temps ischemie	25 min	20 min	35 min	30 min
% nécrose	10%	Focale	30%	30%
% Cellules tumorales en histologie	FF : 80% FFPE : 80%	FF : 70% FFPE : 70%	FF : 80% FFPE : 70%	FF : 70% FFPE : 70%
Age FFPE	2,5 ans	3 ans	2,5 ans	2 ans

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	FFPE en WGS Truseq DNA PCR-free	
Page 3/21		

4 PLAN D'ACTION

4.1 Plan d'expérience

Les expériences ont été réalisées manuellement.

Les ADN des échantillons FF et PBMC ont été extraits avec le kit QIAamp DNA mini de Qiagen (Ref : 51304).

Les ADN des échantillons FFPE ont été extraits avec le kit QIAamp DNA FFPE de Qiagen 5Ref : 56404) Afin de se rapprocher au mieux de ce qui est fait au niveau des plateformes du plan FMG2025, le kit de création de bibliothèques utilisé est le kit TruSeq DNA PCR-free d'Illumina (=TRU) avec fragmentation mécanique et nécessitant 1µg d'ADN de départ. Le séquenceur utilisé est le Novaseq 6000. Le dépôt pour les échantillons FFPE s'est toujours fait en mode XP.

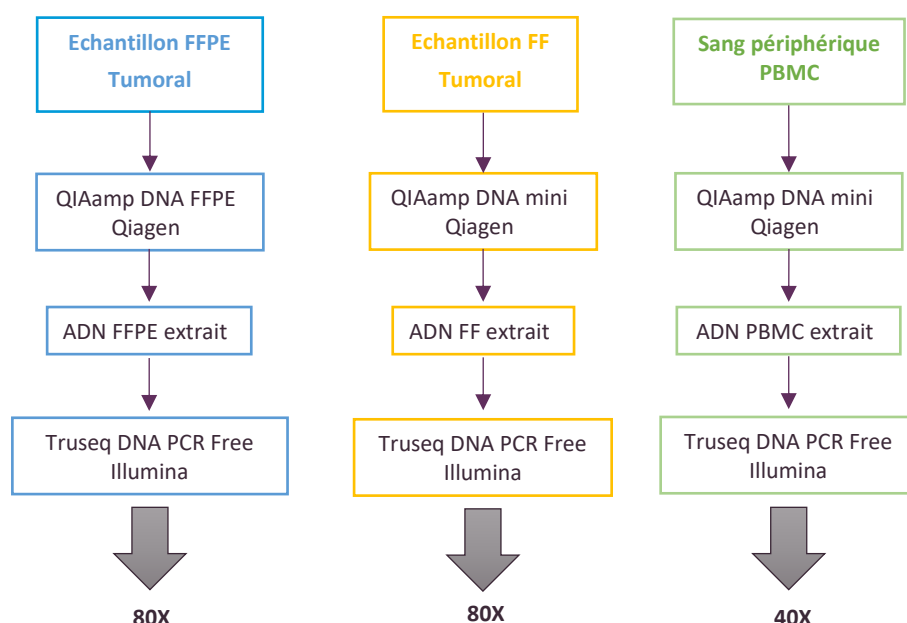



Figure 1 : Déroulement du plan d'action depuis l'extraction d'ADN jusqu'au séquençage

4.2 Matériels & Méthodes bioinformatiques

- Les fichiers bam proviennent de l'alignement des séquences (fastq) sur la référence GRCh38 à l'aide du logiciel bwa-mem2 (v2.1.1) [1].
- Dans un souci d'uniformité, tous les fichiers d'alignements sont downsampled à une profondeur de 80X à l'aide de l'outil samtools (v1.13) [2]. Les profondeurs obtenues sont validées en amont par le logiciel Mosdepth (v0.3.2) [3].
- La détection des variants somatiques est effectuée par Mutect2 (GATK v4.1.7.0) [4].
- Les statistiques et les figures sont générées à l'aide du logiciel R (v4.1.1).

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	FFPE en WGS Truseq DNA PCR-free	
Page 4/21		

5 RESULTATS

5.1 Extraction ADN : quantités et qualités obtenues

Indicateurs de performance	Seuils recommandés	FF Médiane (Min-Max)	FFPE Médiane (Min-Max)	PBMC Médiane (Min-Max)
Quantité ζ extraites	-	23,9 mg (23-34)	2 scrolls 20 μ m (2-3)	5,75E+06 ζ (2,2E+06 - 7,85E+06)
Quantité ADN obtenue	>1 μ g	14,6 μ g (8,9-18,3)	8,8 μ g (2,4-15,1)	14,05 μ g (6,8-19,6)
A260/280	1,8-2,2	1,85 (1,84-1,86)	1,84 (1,83-1,87)	1,86 (1,84-1,87)
A260/230	1,8-2,2	2,14 (2,06-2,39)	1,94 (1,83-2,24)	2,12 (1,75-2,36)
Valeur du DIN*	FF & PBMC >7 FFPE: pas de valeur	6,85 (6,5-7,5)	5,8 (4,4-6,1)	8,55 (8,3-8,7)

Les quantités et les qualités des ADN obtenues pour les différents échantillons sont correctes. Seule la valeur du DIN des ADN FF est légèrement plus basse que la valeur seuil recommandée.

* Voir Recommandation Technique du CRefIX – RT06 – « Valeur seuil du DIN »


5.2 Création des librairies WGS PCR-free : quantités et qualités librairies obtenues

Indicateurs de performance	Seuils recommandés	FF Médiane (Min-Max)	FFPE Médiane (Min-Max)	PBMC Médiane (Min-Max)
Concentration librairies (qPCR)	FF > 10nM FFPE > 1,33nM	19,6 nM (13,18-30,24)	1,58 nM (1,52-2,72)	23,7 nM (18,92-28,4)
Qualité librairies (Profil BioA)	1: Bon; 2:Moyen; 3: Mauvais	1 (1-1)	1 (1-2)	1 (1-1)

Les valeurs seuils recommandées sont données ici à titre indicatif. Pour les échantillons FFPE, la valeur donnée pour la concentration des librairies est la concentration minimale qu'il faut avoir pour pouvoir déposer sur une flowcell Novaseq 6000 en mode XP. **Avec le kit TruSeq PCR free, les quantités de librairies obtenues sont suffisantes pour déposer les échantillons à la profondeur cible. La qualité des librairies est bonne.**

5.3 Validation Technique de séquençage

Cette phase de validation technique, en amont de la validation biologique, est une étape importante car elle permet d'apprécier la qualité intrinsèque des librairies et d'anticiper les erreurs d'interprétation biologique éventuelles. Il a été choisi d'uniformiser les fichiers fastq des librairies à un sous-échantillonnage à 350 Millions de paires de reads. La profondeur attendue est d'environ 30 X.

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	FFPE en WGS Truseq DNA PCR-free	
Page 5/21		

Des indicateurs de performance propres à la méthode testée ont été définis et permettent une comparaison ciblée entre les différents échantillons. Aussi, des seuils recommandés pour chaque indicateur, selon notre expérience, sont ajoutés comme référence attendue.

Indicateurs de performance	Seuils recommandés (FFPE)	FF	FFPE	PBMC
		Méd (Min-Max)	Méd (Min-Max)	Méd (Min-Max)
Couverture 30X	>30%	45% (21-58)	43% (39-54)	38% (12-63)
Taille insert (fragment)	360 +/- 50 pb	382 (341-422)	354 (329-382)	381 (343-411)
IQR	< 12	10 (9-11)	13 (10-14)	8 (8-9)
% Adaptateurs	< 5%	0.16 (0.08-0.22)	0.96 (0.57-1.14)	0.15 (0.11-0.23)
% Alignement	>95%	97.84 (97.42-98.16)	95.44 (92.56-96.64)	97.89 (97.13-98.13)
% Chimère	< 5%	1.74 (1.51-2.01)	4.56 (3.15-7.15)	1.77 (1.61-2.07)
% Duplicat Optique		11.36 (6.02-19.58)	8.3 (7.8-9.65)	12.85 (5.08-21.88)
% Duplicat Autre	< 10%	1.17 (0.49-5.99)	2.16 (1.37-2.91)	1.34 (0.51-6.6)

Seule l'uniformité de couverture (IQR) pour les librairies FFPE dévie du seuil recommandé. Tous les autres indicateurs de performances sont en accord avec les seuils fixés. Le pourcentage d'alignement « properly paired » est plus faible et par conséquent entraine un pourcentage de chimères plus élevé pour les librairies FFPE, mais reste cependant dans des seuils acceptables.

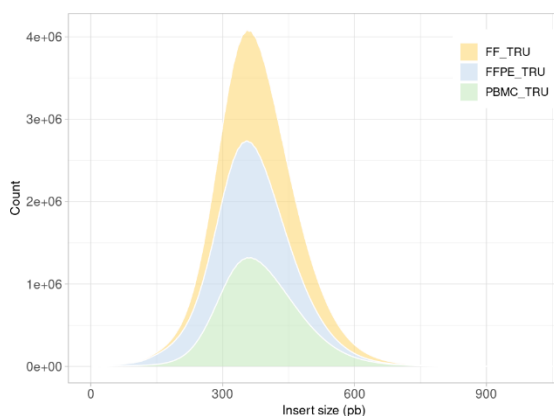


Figure 2 : Répartition de la taille des inserts (fragments)

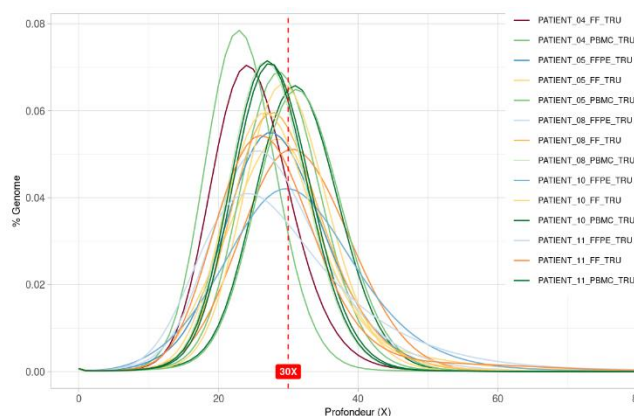



Figure 3 : Répartition de la profondeur des reads sur le génome

Globalement, les librairies FFPE ont une taille d'insert plus petites que les autres librairies mais leur profil de répartition reste identique. La distribution des profondeurs pour chacun des échantillons est plus ou moins centrée sur la profondeur de 30X.

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	FFPE en WGS Truseq DNA PCR-free	
Page 6/21		

5.4 Validation Biologique

5.4.1 Calcul du pourcentage de cellules tumorales


La vérification de la teneur en cellules tumorales dans nos échantillons est réalisée par calcul bioinformatique, à partir des fichiers bam. Ce contrôle est réalisé à l'aide de l'outil FACETS (bioconductor/3.13) [5], package R qui analyse l'hétérogénéité clonale. La pureté, la ploïdie et les fractions cellulaires de la tumeur sont ainsi déduites.

Echantillon		ℓ tumorales histologie	ℓ tumorales WGS	FFPE/FF ℓ tumorales WGS	Recall VP / (VP + FN)
PATIENT_05	FF	80 %	20 %	1,5	78,46 %
	FFPE	80 %	30 %		
PATIENT_08	FF	70 %	30 %	1,5	80,43 %
	FFPE	70 %	45 %		
PATIENT_10	FF	80 %	32 %	1	95,12 %
	FFPE	70 %	33 %		
PATIENT_11	FF	70 %	54 %	1	95,56 %
	FFPE	70 %	52 %		

Le pourcentage de cellules tumorales semble avoir été surestimé par le(s) pathologiste(s) ou bien sous-estimé par le logiciel FACETS. Pour la suite de l'analyse, les résultats moléculaires seront pris en compte. Le PATIENT_05 avec 20% de cellularité en FF, est en limite de détection pour l'appel de variants. Aussi, plus de 10% de différence est observé sur le calcul de la pureté entre les extractions FF & FFPE pour les patients PATIENT_05 et PATIENT_08. **La baisse du % de Recall (nombre de variants Vrais Positifs (VP) retrouvés dans la librairie FFPE parmi ceux détectés au sein de la librairie FF, qui sert de référence (VP + Faux Négatifs (FN)) serait expliquée par le fait que ces échantillons ne sont pas des blocs miroirs comme attendu et fausseraient l'interprétation biologique finale. Ainsi, les patients PATIENT_05 et PATIENT_08 seront écartés de la validation biologique.**

5.4.2 Analyse des variants

Une analyse globale et générale des résultats de Mutect2 est présentée dans le tableau ci-dessous. Le logiciel rtg vcftstats (v.3.12.1) [6] a permis de sortir différentes statistiques, dont le nombre de variants « PASS », « Failed », le nombre d'insertions et de délétions, ainsi que le ratio Ti/Tv. Seules les positions incluses dans les régions de haute confiance sont conservées.

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	FFPE en WGS Truseq DNA PCR-free	
Page 7/21		

Echantillon		« Failed Filters »	« Passed Filters »	FFPE / FF	SNVs	MNPs	SNVs Ti/Tv	INS / DEL	InDels / SNVs+MNPs
PATIENT_10	FF	146 445	40 803	1,4	38 789	717	0.86 (17 938/20 851)	0.59 (482/815)	0.03 (1 297/39 506)
	FFPE	226 640	58 790		47 094	2 711	1.03 (23 901/23 193)	0.09 (778/8 207)	0.18 (8 985/49 805)
PATIENT_11	FF	137 190	65 299	2,6	60 256	2 038	0.32 (14 764/45 492)	0.49 (983/2 022)	0.05 (3 005/62 294)
	FFPE	518 251	169 248		123 436	27 213	0.88 (57 654/65 782)	0.22 (3 301/15 298)	0.12 (18 599/150 649)

Les résultats de l'analyse montre qu'il y a :

- **1,4 à 2,6 fois** plus de variants pour les échantillons FFPE que dans les échantillons FF
- **8 fois plus** de délétions pour les échantillons FFPE, impliquant un ratio Insertion/Deletion (INS/DEL) plus faible que pour les échantillons FF
- **2 à 3,5 fois plus** d'insertions pour les échantillons FFPE que pour les FF
- Ratio Ti/Tv **1,2 à 2,8 fois plus** élevé pour les FFPE que pour les FF

5.4.3 Distribution des variants SNVs et InDels dans les échantillons FFPE et FF

Les variants (SNVs et InDels) présents dans les régions de haute confiance ont été sélectionnés puis analysés à l'aide de scripts R internes

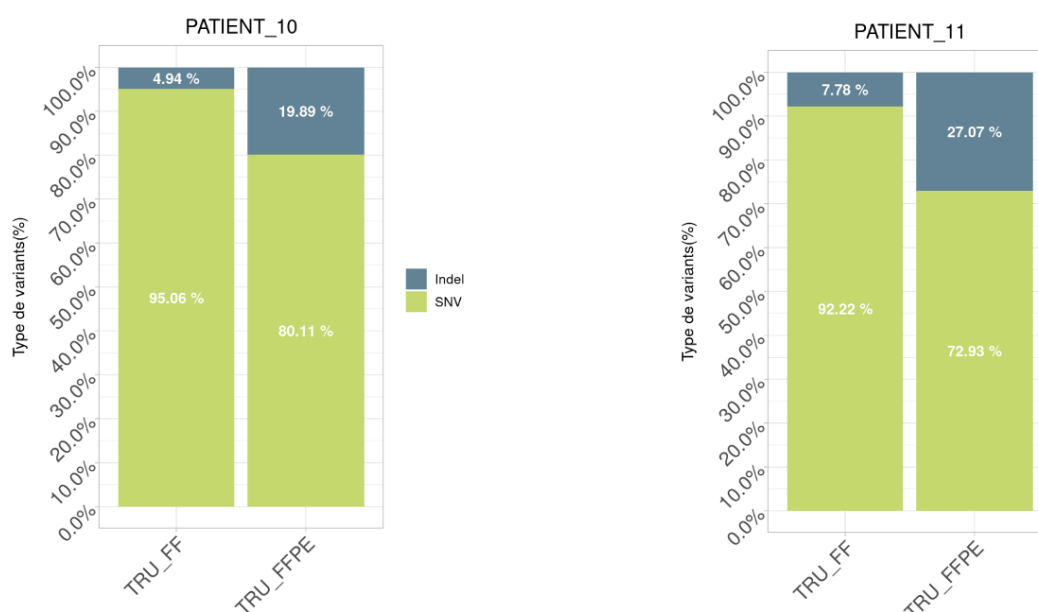



Figure 4 : Distribution des variants SNVs et InDels dans les régions de haute confiance et downsamplés à 80 X, pour les échantillons FFPE et FF

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet Page 8/21	FFPE en WGS Truseq DNA PCR-free	

Les échantillons FFPE possèdent 15-20% de plus d'InDels que les échantillons FF.

5.4.4 Répartition des VAF des variants SNVs et InDels dans les échantillons FFPE et FF

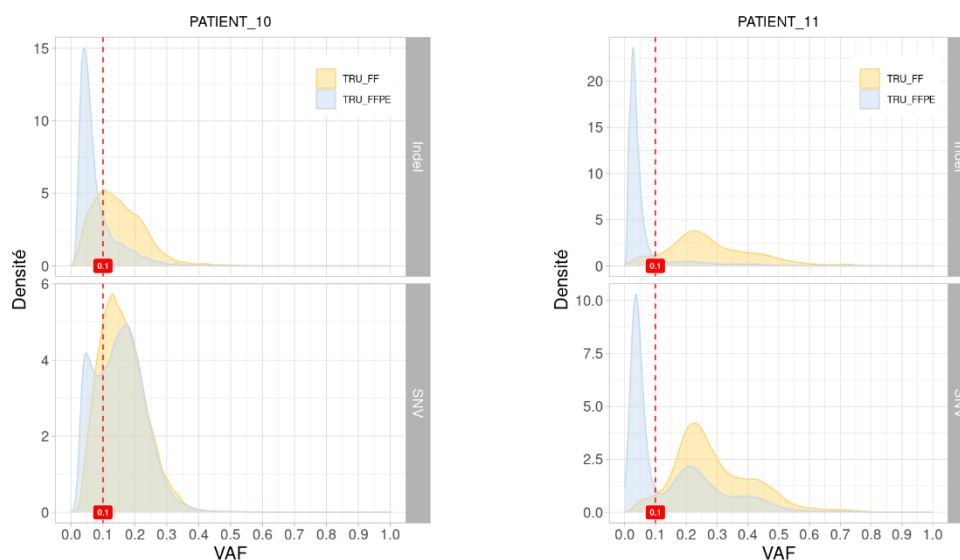


Figure 5 : Comparaison de la répartition des variants SNVs et InDels FF vs FFPE selon leur VAF pour les Patients PATIENT_10 et PATIENT_11

Les SNVs présents dans l'échantillon de type FFPE montrent un profil avec deux pics, un premier pic avec des faibles VAF (<0.1) et un second pic avec des VAF plus élevées. Ce deuxième pic a le même profil que son équivalent en FF. Pour les InDels, l'exploitation de l'analyse de la VAF est plus complexe.


Pour la suite de l'analyse biologique, nous visualiserons deux seuils de VAF, l'un > 0 et l'autre > 0.1 . Les variants de faibles VAF (<0.1) semblent être spécifiques à l'échantillon FFPE ; les éliminer permettrait d'augmenter la précision, sans diminuer la sensibilité.

5.4.4.1 Comparaison des SNVs détectés dans les échantillons FF et FFPE respectivement

Pour mesurer la performance des SNVs trouvés dans les échantillons FFPE vs ceux trouvés dans les échantillons FF (=référence), les indicateurs de performance Précision et Recall ont été calculés avec les deux seuils de VAF.

La Précision se concentre uniquement sur les SNVs communs (vrais positifs, VP) et donne une indication sur les faux positifs (= FP, c'est-à-dire les SNVs trouvés uniquement dans les échantillons FFPE).

Le Recall se concentre uniquement sur les SNVs communs (VP) et donne une indication sur la part de faux négatifs (=FN, cad les SNVs trouvés uniquement dans les échantillons FF).

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	FFPE en WGS Truseq DNA PCR-free	
Page 9/21		

Le F1-score est une moyenne harmonique de la Precision et du Recall.

Patient	VAF	SNVs commun	SNVs FF spé	SNVs FFPE spé	Precision VP / (VP + FP)	Recall VP / (VP + FN)	F1-score $2*VP / (2*VP+FN+FP)$
PATIENT_10	> 0	34 973	4 137	12 121	74,26 %	89,42 %	81,14 %
	> 0.1	26 494	4 905	7 254	78,51 %	84,38 %	81,34 %
PATIENT_11	> 0	56 270	4 088	67 167	45,59 %	93,23 %	61,23 %
	> 0.1	54 438	3 219	2 852	95,02 %	94,42%	94,72 %

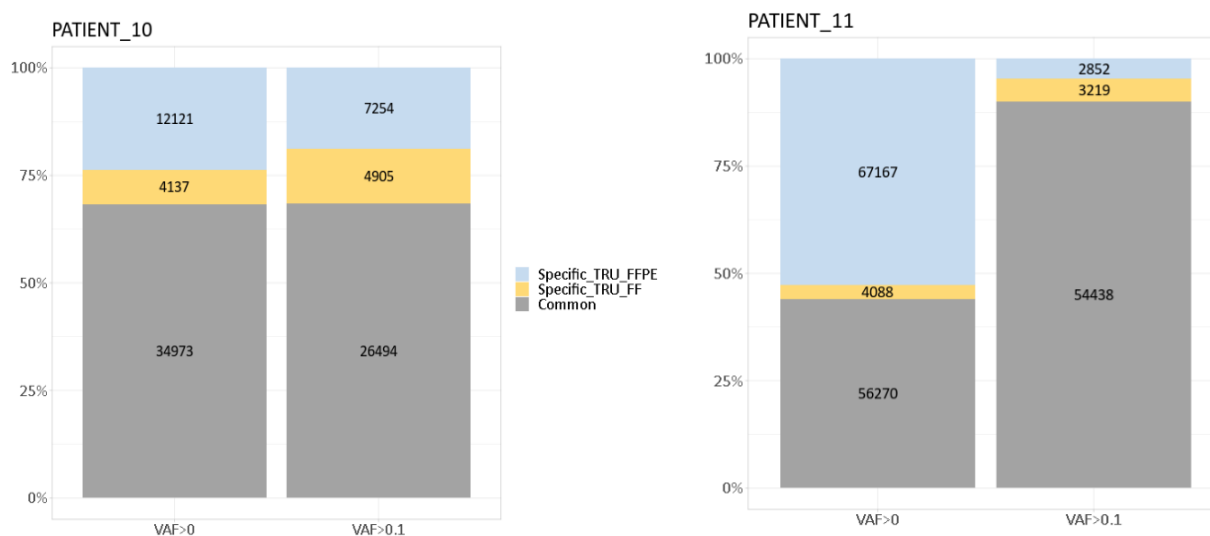


Figure 6 : Comparaison de la répartition des SNVs commun vs spécifiques avec VAF>0 et VAF >0.1 entre FF et FFPE

Pour le PATIENT_10, le seuil de VAF à 0.1 ne permet pas d'améliorer significativement la précision et provoque une baisse du recall. En effet, cette méthode élimine de nombreux vrais positifs de faible VAF en raison de la faible cellularité (30%), causant également l'apparition de faux négatifs. La méthode du seuil doit donc être utilisée avec prudence, de façon adaptée à chaque échantillon. Pour le PATIENT_11, le changement de seuil de VAF augmente de façon significative (1,5 fois plus) la valeur de F1-score. En effet, l'échantillon FFPE de ce patient possède énormément de SNVs spécifiques de faible VAF (voir le violin plot ci-dessous). Dans ce cas, un seuil à de VAF >0.1 a permis d'obtenir une précision de plus de 95% sans trop affecter les vrais positifs (recall) et donc un F1-score de plus de 90%.

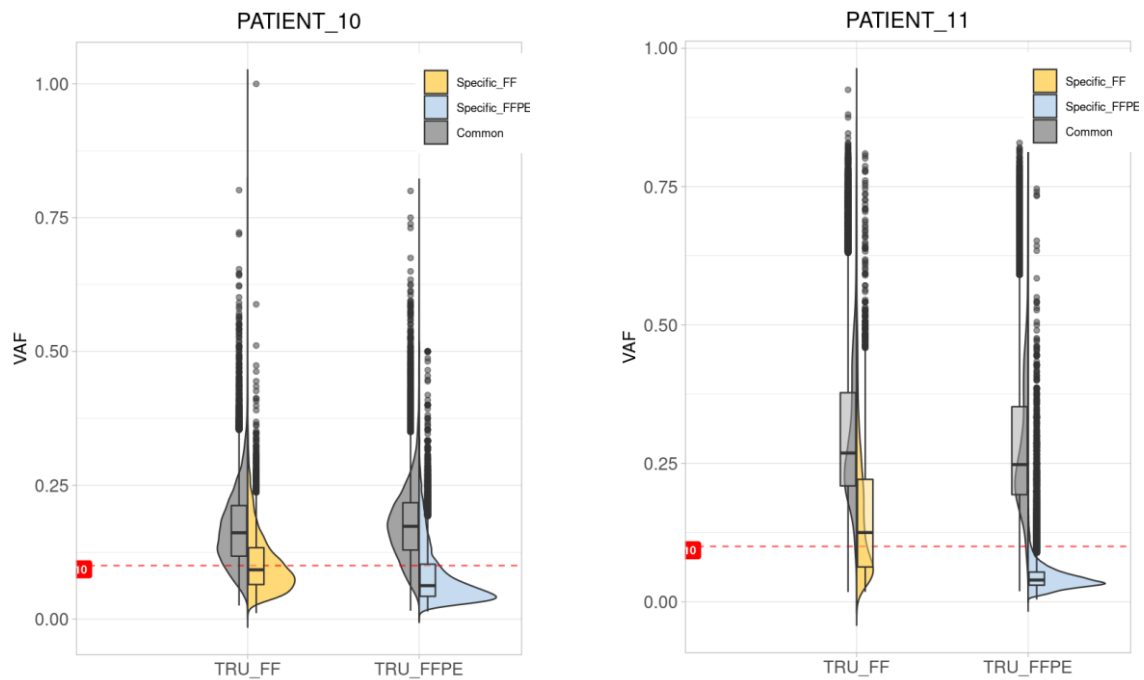


Figure 7 : Violin plot des SNVs commun vs spécifiques en fonction de l'échantillon FF/FFPE

5.4.4.2 Comparaison de la répartition des SNVs par région génomique: Whole Genome, Broadex, Exome, Cancer Genes

Dans cette partie, nous allons regarder la répartition des SNVs communs et spécifiques selon 4 granularités : le Whole Genome (complet), le Whole Genome filtré sur les régions de hautes confiance dérivé du BROAD (nommé Broadex), l'Exome (exons de RefSeq) et les exons Cancer Genes (Cancer gene Census, COSMIC v97).

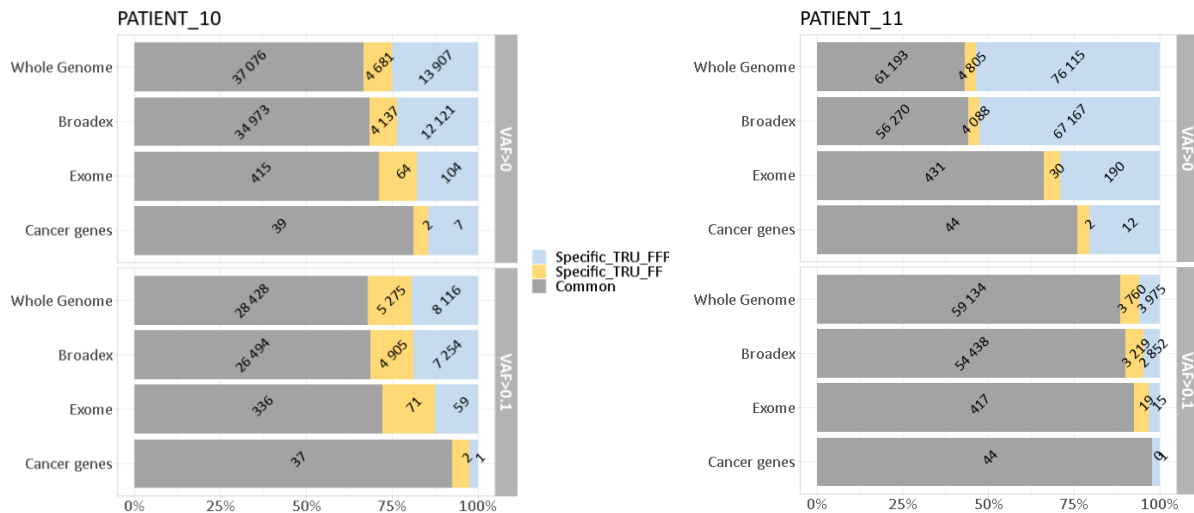



Figure 8 : Comparaison de la répartition des SNVs commun vs spécifiques avec VAF > 0 et VAF > 0.1 entre FF et FFPE selon les différentes régions génomiques

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	FFPE en WGS Truseq DNA PCR-free	
Page 11/21		

Patient	Région	VAF	Precision	Recall	F1-score
			VP / (VP + FP)	VP / (VP + FN)	2*VP / (2*VP+FN+FP)
PATIENT_10	Exome	> 0	79,96 %	86,64 %	83,17 %
		> 0.1	85,06 %	82,56 %	83,79 %
	Cancer genes	> 0	84,78 %	95,12 %	89,66 %
		> 0.1	97,37 %	94,87 %	96,10 %
PATIENT_11	Exome	> 0	69,40 %	93,49 %	79,67 %
		> 0.1	96,53 %	95,64 %	96,08 %
	Cancer genes	> 0	78,57 %	95,65 %	86,27 %
		> 0.1	97,78 %	100 %	98,88 %

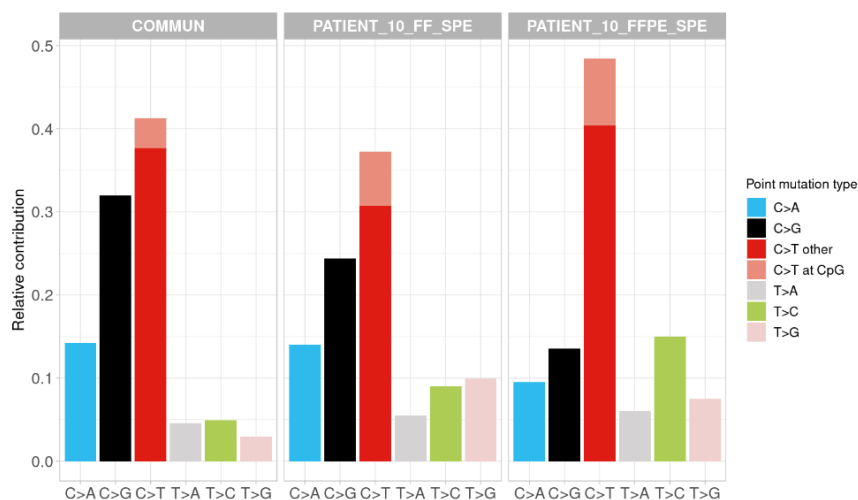
Pour les 2 patients :

- Sans filtre sur la VAF, il y a plus de variants spécifiques FFPE dans les régions *Whole genome* et *Broadex* (entre 5 et 20% de plus) que dans les régions *Exome* et *Cancer genes*.
- Le filtre de VAF >0.1 a permis de réduire drastiquement les variants spécifiques FFPE (les faux positifs) d'un facteur 2 voire 23 dans certaines régions.
- Pour les régions *Exome* et *Cancer gene*, le F1-score est toujours plus élevé avec le filtre de VAF>0.1.

5.4.5 Signature mutationnelle

5.4.5.1 Spectrum mutation/96 mutational profile

La caractérisation des signatures mutationnelles est réalisée en utilisant uniquement les substitutions de base et les informations incluses dans la séquence contextuelle de chaque mutation. Il y a donc 6 classes de substitutions et 96 mutations possibles dans la classification regardée. Le package R MutationalPatterns (V3.8.0 ; <http://bioconductor.org/packages/MutationalPatterns>) [7] a été utilisé à cet effet.



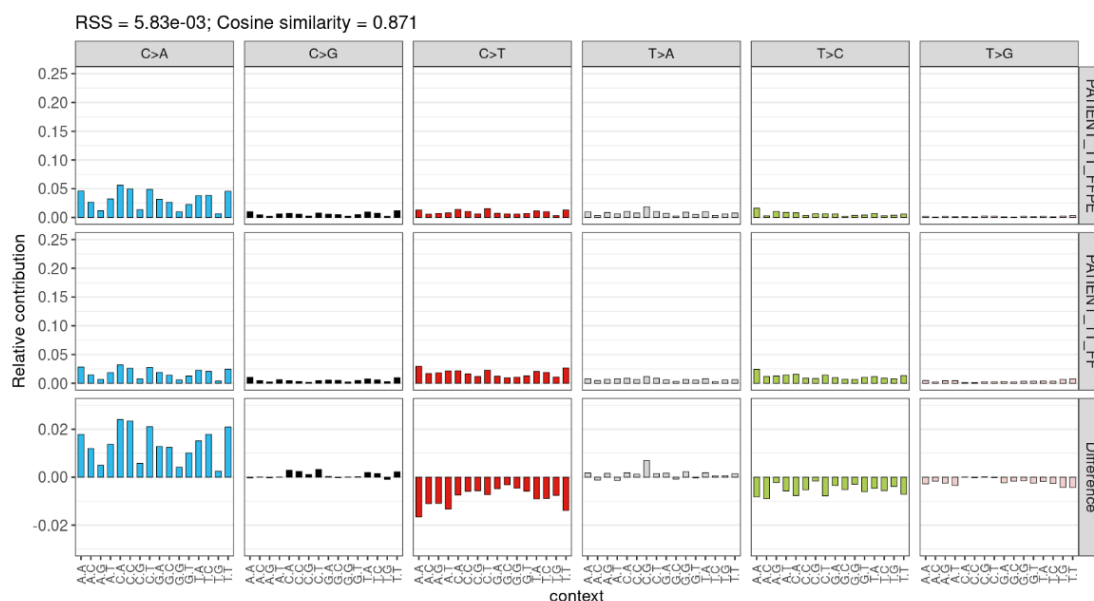


Figure 10 : Calcul de la différence de la contribution relative entre deux profils de signatures mutationnelles sur la base des 96 mutations. La différence est la résultante du profil FFPE sur FF. Un coefficient de similarité de cosme est obtenu, compris entre 0 et 1, 1 étant le degré de similarité le plus élevé.


Le coefficient de similarité de cosme est égal à 0.995 pour le PATIENT_10, signifiant que les profils mutationnels entre les deux extractions sont fortement similaires. Celui observé pour le PATIENT_11 est légèrement plus faible, son coefficient de cosme étant de 0.871. Cela fait écho aux observations précédentes faites sur les substitutions de bases.

5.4.5.2 Cosmic SBS signatures

Le profil des 96 mutations est comparé au catalogue des signatures mutationnelles connues défini par COSMIC (<https://cancer.sanger.ac.uk/signatures/sbs/>). Les Single Base Substitutions ou SBS présentant une contribution relative d'au moins 10% avec le profil mutationnel du patient sont conservées.

PATIENT_10_FF			PATIENT_10_FFPE		
% relative contribution	SBS	Ethiologie	% relative contribution	SBS	Ethiologie
35,21%	SBS13	Activity of APOBEC family of cytidine deaminases	29,03%	SBS13	Activity of APOBEC family of cytidine deaminases
32,18%	SBS2	Activity of APOBEC family of cytidine deaminases	27,50%	SBS2	Activity of APOBEC family of cytidine deaminases

Tableau 1: Noms et étiologie des signatures SBS COSMIC retrouvées avec une contribution relative >= à 10% pour chaque extraction FF & FFPE pour le PATIENT_10.

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	FFPE en WGS Truseq DNA PCR-free	
Page 14/21		

PATIENT_11_FF			PATIENT_11_FFPE		
% relative contribution	SBS	Ethiologie	% relative contribution	SBS	Ethiologie
50,92%	SBS4	Tobacco smoking	24,54%	SBS4	Tobacco smoking
12,62%	SBS40	Unknown	12,73%	SBS37	Unknown
0%	SBS37	Unknown	8.69 %	SBS40	Unknown

Tableau 2: Noms et étiologie des signatures SBS COSMIC retrouvées avec une contribution relative > = à 10% pour chaque extraction FF & FFPE pour le PATIENT_11.

Les SBS COSMIC retrouvées chez le PATIENT_10 sont identiques entre les échantillons FFPE et FF. Les SBS les plus majoritaires sont SBS13 et SBS2. Concernant le PATIENT_11, les SBS ne sont pas similaires. La SBS37 observée pour l'échantillon FFPE est absente du FF. La SBS40 vue pour l'échantillon FF est retrouvée avec une contribution relative de 8.7%, en dessous du seuil fixé à 10%.


5.4.6 Hotspots: catalogue Cancer Genome Interpreter (2700 hotspots)

La liste de hotspots du catalogue de CancerGenomeInterpreter (CGI ; <https://www.cancergenomeinterpreter.org/mutations>) est utilisée pour faciliter l'identification des altérations tumorales de nos deux patients qui pourrait être cliniquement exploitables. CGI s'appuie sur des méthodes computationnelles ainsi que sur des connaissances collectées dans le domaine public (DoCM, ClinVar, OncoKB). Cette liste de hotspots est composée de 2 749 mutations. L'idée ici est de s'assurer que le FFPE n'altère pas la découverte des hotspots s'il en existe dans le tissu FF.

Les résultats et positions de l'analyse de Mutect2 sont comparés à celles de la liste de hotspots CGI, aucun critère sur le champ FILTER, ni de filtre sur la VAF n est appliqué.

Echantillon	Gene	HGVSp	HGVSc	Profondeur	AD (Ref, Alt)	VAF	FILTER Mutect	
PATIENT_10	FF	TP53	p.R213Q	c.638G>A	91 x	63,28	0.308	PASS
	FFPE	TP53	p.R213Q	c.638G>A	39 x	32,7	0.179	PASS
PATIENT_11	FF	KRAS	p.G12V	c.35G>T	110 x	70,40	0.364	PASS
	FFPE	KRAS	p.G12V	c.35G>T	101 x	61,40	0.396	PASS

Un hotspot de la liste CGI est détecté pour chaque patient. Ces derniers sont bien retrouvés également dans les tissus FFPE et en étant taggés PASS par l'analyse Mutect2.

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	
Sujet Page 15/21	FFPE en WGS Truseq DNA PCR-free	

5.4.7 Copy Number Alterations

Les CNAs (Copy Number Alteration) sont obtenus en analysant les paires appariées PBMC-FF et PBMC-FFPE par le logiciel FACETS (bioconductor/3.13) [5], l'échantillon PBMC servant de normal apparié. Nous avons choisi de se focaliser uniquement sur le graphique général de FACETS qui est le 3^{ème} plot correspondant aux résultats du nombre de copies (total, mineur) et la fraction cellulaire suivant les chromosomes.

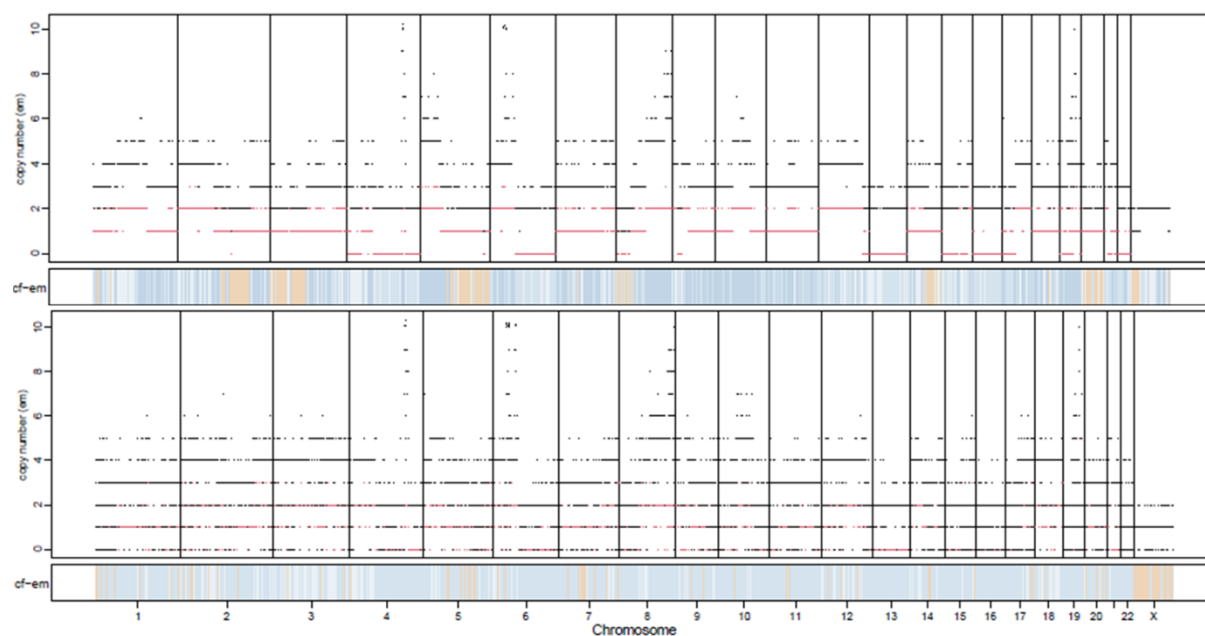


Figure 11: Comparaison des CNAs entre FF (graphe haut) et FFPE (graphe bas) pour le PATIENT_10

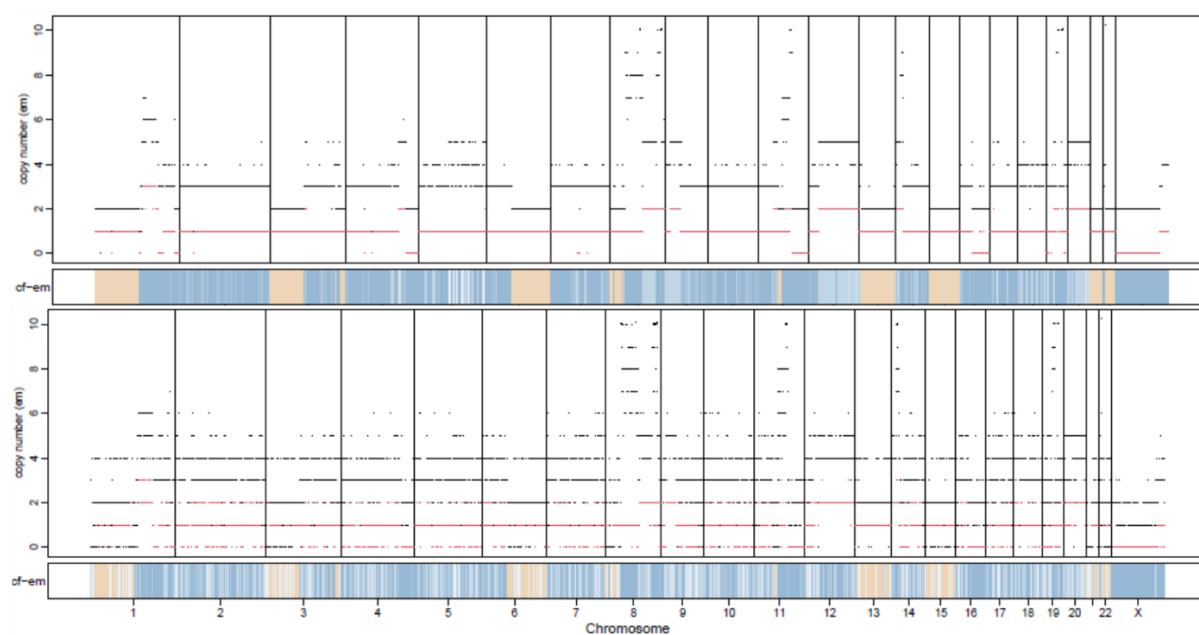

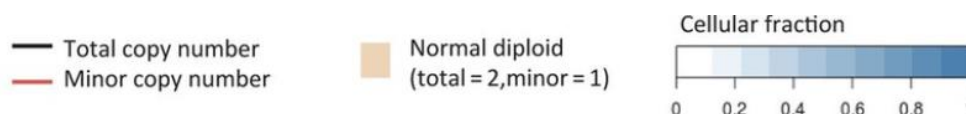


Figure 12: Comparaison des CNAs entre FF (graphe haut) et FFPE (graphe bas) pour le PATIENT_11

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	FFPE en WGS Truseq DNA PCR-free	
Page 16/21		



Bien que le FFPE garde globalement le même profil CNA que le FF, on observe qu'il occasionne un signal perturbé dans la détermination du nombre de copies, la cause principale étant une plus grande dispersion dans les valeurs de \log_2R . Egalement, l'apparition de nombreuses régions délétées (Total copy number = 0), absentes du FF sont observées. Cette variabilité semble avoir une distribution aléatoire et uniforme sur le génome.

Le tableau ci-après montre ce phénomène, où le FFPE présente 90 fois à 900 fois plus de régions perdues par rapport au FF (colonne Loss). Ces régions comptabilisent 139 Mb pour le FFPE, contre seulement 0,049 Mb pour le FF. De même, 26 fois à 90 fois plus de régions de LOH sont observées pour les échantillons FFPE.

Echantillon	# CNAs	FFPE/F F ratio	Amplif (>=6 copies)	Gain	cnLOH	LOH	FFPE/FF ratio	Loss	FFPE/FF ratio	Uniparental gain
PATIENT_10_FF	1 565		109	1 169	126	8		1		152
PATIENT_10_FFPE	4 903	3,1	152	2 382	401	752	94	981	981	235
PATIENT_11_FF	1 252		122	1047	65	6		2		10
PATIENT_11_FFPE	2 741	2,2	201	1 763	252	153	26	184	92	188

Tableau 3: Répartition des évènements de CNAs entre les échantillons FF et FFPE, obtenus à l'aide du logiciel FACETS

5.4.8 Variants structurels

Le logiciel d'Illumina Manta (v1.0.3) [8] est utilisé pour la détection de variants structurels ou SV de tous types (délétions, duplications, insertions, translocations et remaniements complexes) en utilisant les approches « split-reads » (ou reads chimériques) et « paired-reads ». L'approche de cette méthode est l'assemblage local d'un sous ensemble de lectures d'intérêt, qui permet de valider la présence d'un variant de structure et d'affiner sa position sur le génome.

La fusion et la comparaison des évènements au sein de chaque patient est réalisé à l'aide du logiciel SURVIVOR (v1.0.3) [9] à partir de fichiers vcf (régions de haute confiance et downsampling à 80X). Une distance maximale de 1kb est paramétrée permettant la fusion des évènements. Les résultats sont ensuite annotés avec le logiciel AnnotSV (v3.1) [10], un chevauchement de 70% est autorisé permettant la caractérisation génomique pertinentes des résultats : annotation des gènes, des éléments régulateurs, annotation de base de données (1000 Genomes, GnomAD, ...), pathogénicité ...

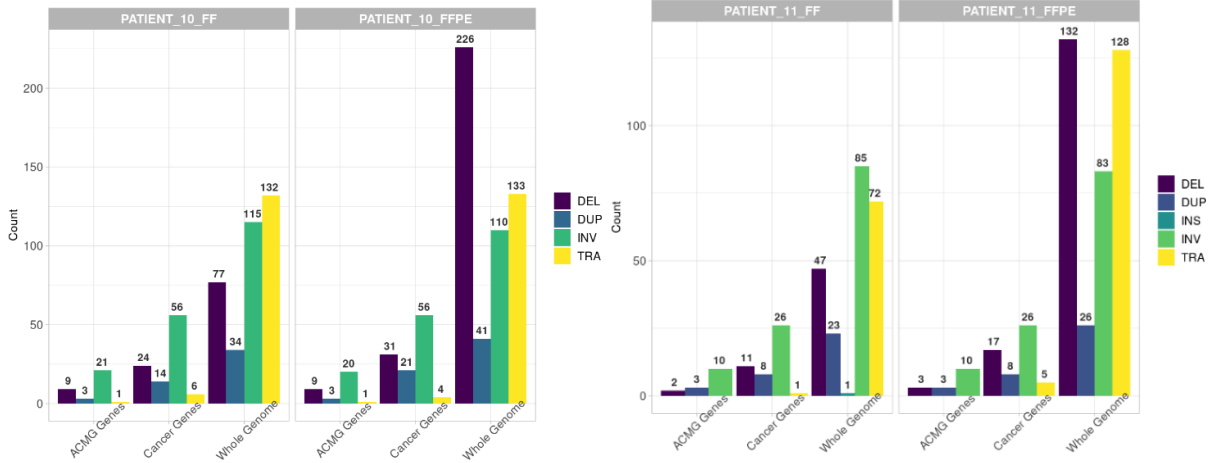



Figure 13 : Répartition des événements de SV entre les deux types d'extractions. Seuls les SV considérés comme PASS par le logiciel Manta sont représentés ici.

Région	Type SV	FF	FFPE	FFPE / FF	FF Spé	FFPE Spé	Commun	Recall VP / (VP + FN)
Whole Genome	DEL	77	226	2.9	10,67 %	69,57 %	19,76 %	64,94 %
	DUP	34	41	1.2	8,89 %	24,44 %	66,67 %	88,24 %
	INV	115	110	1	25 %	22,16 %	52,84 %	67,88 %
	TRA	132	133	1	30,73 %	31,25 %	38,02 %	55,3 %
	SUM	358	510	1.4				
Cancer Genes	DEL	24	31	1.3	16,22 %	35,14 %	48,65 %	75 %
	DUP	14	21	1.5	4,55 %	36,36 %	59,09 %	92,86 %
	INV	56	56	1	22,22 %	22,22 %	55,56 %	71,43 %
	TRA	6	4	0.7	55,56 %	33,33 %	11,11 %	16,67 %
	SUM	100	112	1.1				
ACMG Genes	DEL	9	9	1	10 %	10 %	80 %	88,89 %
	DUP	3	3	1	0 %	0 %	100 %	100 %
	INV	21	20	1	13,04 %	8,7 %	78,26 %	85,71 %
	TRA	1	1	1	0 %	0 %	100 %	100 %
	SUM	34	33	1				


Tableau 4 : Répartition des événements de SV selon 3 catégories de régions génomiques pour le PATIENT_10. Seuls les SV considérés comme PASS par le logiciel Manta sont représentés ici.

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	FFPE en WGS Truseq DNA PCR-free	
Page 18/21		

Région	Type SV	FF	FFPE	FFPE/FF	FF Spé	FFPE Spé	Commun	Recall VP / (VP + FN)
Whole Genome	DEL	47	132	2.8	7,04 %	66,9 %	26,06 %	78,72 %
	DUP	23	26	1.1	13,33 %	23,33 %	63,33 %	82,61 %
	INS	1	0	0	100 %	0 %	0 %	0 %
	INV	85	83	1	13,08 %	10,28 %	76,64 %	85,42 %
	TRA	72	128	1.8	17,95 %	53,85 %	28,21 %	61,11 %
	SUM		228	369	1.6			
Cancer Genes	DEL	11	17	1.5	0 %	35,29 %	64,71 %	100 %
	DUP	8	8	1	0 %	0 %	100 %	100 %
	INV	26	26	1	7,14 %	7,14 %	85,71 %	92,31 %
	TRA	1	5	5	0 %	80 %	20 %	100 %
	SUM		46	56	1.2			
ACMG Genes	DEL	2	3	1.5	0 %	33,33 %	66,67 %	100 %
	DUP	3	3	1	0 %	0	100 %	100 %
	INV	10	10	1	16,67 %	16,67 %	66,67 %	80 %
	SUM		15	16	1.1			


Tableau 5: Répartition des évènements de SV selon 3 catégories de régions génomiques pour le PATIENT_11.
Seuls les SV considérés comme PASS par le logiciel Manta sont représentés ici.

De façon globale, les librairies FFPE des deux patients ont de 1,4 fois à 1,6 fois plus d'évènements que ceux observés dans les librairies FF, respectivement. Ces derniers correspondent essentiellement à des évènements de délétion, jusqu'à presque 3 fois plus. Pour le PATIENT_10, la classe des duplications est aussi plus importante dans la librairie FFPE. Ces délétions supplémentaires possèdent une taille d'insert très petite (<200 pb) et un *somatic score* sensiblement plus faible. En ajustant ce dernier à un seuil de 40, le nombre de faux positifs se réduit drastiquement, et ce pour les deux patients.

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	FFPE en WGS Truseq DNA PCR-free	
Page 19/21		

6 CONCLUSION GENERALE

	Validation FFPE	FFPE vs FF avec le kit TruSeq PCR-free
Validation technique - WetLab	+	<ul style="list-style-type: none"> - Rendement des librairies FFPE largement < celles des FF. - Quantité suffisante pour déposer sur une FC Novaseq en mode XP pour 80x de profondeur. - Qualité identique entre FFPE et FF
Validation technique – Bio-IT	++	<ul style="list-style-type: none"> - Seul le critère de l’homogénéité de couverture dévie des seuils recommandés
Validation biologique - SBS	+	<ul style="list-style-type: none"> - Prépondérance des mutations C>T dans les SNVs spécifiques aux FFPE - Similarité dans les profils mutationnels, le coefficient de cosinus est > à 0.88 et 0.99 - Similarité dans les SBS de COSMIC
Validation biologique - Hotspots	++	<ul style="list-style-type: none"> - Pas de perte de hotspots
Validation biologique – SNVs/Indels	-	<ul style="list-style-type: none"> - 1,5 à 2,5 fois plus de variants pour le tissu FFPE - Le FFPE apporte 15% d’indel en plus
Validation biologique – Precision, Recall, F1 score	+	<ul style="list-style-type: none"> - Recall similaire - Baisse de la précision, possible à corriger par filtrage de VAF
Validation biologique – VAF	+	<ul style="list-style-type: none"> - La VAF des variants communs est similaire - La VAF des variants spécifiques au FFPE est légèrement plus basse que celle du FF
Validation biologique – CNAs	-	<ul style="list-style-type: none"> - Signal perturbé dans la détermination du nombre de copies - Apparition de nombreuses régions perdues sur le génome (jusqu’à 900 fois plus)
Validation biologique - SVs	+/-	<ul style="list-style-type: none"> - Le nombre d’évènements de SVs est 1,4 à 1,6 fois plus élevé pour le FFPE - Induit 3 fois plus de SVs de type DEL pour le tissu FFPE, et retrouvé chez les 2 patients - La majorité des délétions spécifiques à l’échantillon FFPE ont une taille < à 200 pb et un somatic score très bas - Possibilité de réduire le nombre de DEL spécifiques au FFPE en filtrant sur le somatic score (> à 40 ?)

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	FFPE en WGS Truseq DNA PCR-free	
Page 20/21		

7 DOCUMENTS

[1] The full BWA package is distributed under GPLv3 as it uses source codes from BWT-SW which is covered by GPL. Sorting, hash table, BWT and IS libraries are distributed under the MIT license.

[2] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078-9. doi: 10.1093/bioinformatics/btp352. Epub 2009 Jun 8. PMID: 19505943; PMCID: PMC2723002.

[3] Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*. 2018 Mar 1;34(5):867-868. doi: 10.1093/bioinformatics/btx699. PMID: 29096012; PMCID: PMC6030888.

[4] Calling Somatic SNVs and Indels with Mutect2. David Benjamin, Takuto Sato, Kristian Cibulskis, Gad Getz, Chip Stewart, Lee Lichtenstein [bioRxiv 861054](https://doi.org/10.1101/861054); doi: <https://doi.org/10.1101/861054>

[5] Shen R, Seshan VE. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res*. 2016 Sep 19;44(16):e131. doi: 10.1093/nar/gkw520. Epub 2016 Jun 7. PMID: 27270079; PMCID: PMC5027494.

[6] RTG Tools: Utilities for accurate VCF comparison and manipulation. <https://github.com/RealTimeGenomics/rtg-tools>

[7] Manders F, Brandsma AM, de Kanter J, Verheul M, Oka R, van Roosmalen MJ, van der Roest B, van Hoeck A, Cuppen E, van Boxtel R. MutationalPatterns: the one stop shop for the analysis of mutational processes. *BMC Genomics*. 2022 Feb 15;23(1):134. doi: 10.1186/s12864-022-08357-3. PMID: 35168570; PMCID: PMC8845394.

Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med*. 2018 Apr 25;10(1):33. doi: 10.1186/s13073-018-0539-0. PMID: 29695279; PMCID: PMC5922316.


[8] Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, Cox AJ, Kruglyak S, Saunders CT. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016 Apr 15;32(8):1220-2. doi: 10.1093/bioinformatics/btv710. Epub 2015 Dec 8. PMID: 26647377.

[9] Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Jeffares, Daniel C; Jolly, Clemency; Hoti, Mimoza; Speed, Doug; Shaw, Liam; Rallis, Charalampos; Balloux, Francois; Dessimoz, Christophe; Bähler, Jürg; Sedlazeck, Fritz J. *Nature communications*, Vol. 8, 14061, 24.01.2017, p. 1-11. DOI:10.1038/NCOMMS14061; <https://github.com/fritzsedlazeck/SURVIVOR>

[10] Geoffroy V, Guignard T, Kress A, Gaillard JB, Solli-Nowlan T, Schalk A, Gatinois V, Dollfus H, Scheidecker S, Muller J. AnnotSV and knotAnnotSV: a web server for human structural variations annotations, ranking and analysis. *Nucleic Acids Res*. 2021 Jul 2;49(W1):W21-W28. doi: 10.1093/nar/gkab402. PMID: 34023905; PMCID: PMC8262758

8 HISTORIQUE DES MODIFICATIONS

Version	Date	Nature de la modification
V01	15/02/2023	Revue et correction
V00	01/02/2023	Création

N°: 20230201_DT01 Version : 01 Date : 15/02/2023	Document Technique Plan FMG 2025	 France Médecine Génomique 2025
Sujet	FFPE en WGS Truseq DNA PCR-free	
Page 21/21		

9 VALIDATION DU DOCUMENT

RÉDACTEURS	RELECTEURS
Alice Moussy, Mélanie Letexier, Marine Rouillon, Jasmin Cévest	Jean François Deleuze, Alain Viari, Violette Turon
PLAN EXPERIMENTAL	Fonction : Direction Unité
Alice Moussy, Edouard Turlotte, Margaux Gras	Date : 15/02/2023
ANALYSE	Visa :
Mélanie Letexier, Marine Rouillon, Jasmin Cévest	
Date : 15/02/2023	
Visa :	