
Recommandation à propos de l'utilisation du Kit Illumina® DNA PCR-Free Prep à partir de tissus tumoraux congelés

Table des matières

1	OBJET	2
2	INTERETS POUR LES PLATES FORMES DU PLAN FRANCE MEDECINE GENOMIQUE	2
3	CHAMP D'APPLICATION DE LA RECOMMANDATION	2
4	PRESENTATION DU KIT.....	2
5	PLAN DE TESTS.....	2
5.1	Echantillons.....	2
5.2	Expériences	3
5.2.1	Extractions.....	3
5.2.2	Librairies.....	3
6	RESULTATS	3
6.1	Qualités des librairies.....	3
6.2	Evaluation de la qualité des runs de séquençage	3
6.3	Validation technique	3
6.3.1	Résultats des critères de performance.....	4
6.3.2	Couverture des régions GC.....	4
6.4	Analyses Biologiques.....	5
6.4.1	Matériels & Méthodes bioinformatiques.....	5
6.4.2	Analyse des Variants	5
6.4.3	Précision, Recall et F1-Score	6
6.4.4	Copy Number Alteration	6
6.4.5	Signatures mutationnelles - Spectrum Mutations/96 mutational profile	7
6.4.6	Génotype Concordance, Hotspots et SVs.....	8
7	LES POINTS DE VIGILANCES.....	8
8	CONCLUSION GENERALE.....	9
9	DOCUMENTS.....	10
10	HISTORIQUE DES MODIFICATIONS	10
11	VALIDATION DU DOCUMENT	10

N°: 20230322_RT09 Version : 01 Date : 22/03/2023	Recommandation Technique PFMG 2025	 France Médecine Génomique 2025
Sujet	Illumina® DNA PCR-Free Prep	
Page 2/10		

1 OBJET

Evaluation du kit Illumina® DNA PCR-Free Prep (TAG), kit de préparation de bibliothèques par fragmentation enzymatique et sans amplification à partir d'ADN d'échantillons tumoraux congelés (FF) à faible quantité de départ (100ng - 300ng) pour du séquençage de génome entier.

Cette analyse est réalisée par comparaison avec le kit TruSeq DNA PCR-Free (TRU) qui est principalement utilisé dans le plan France Médecine Génomique aujourd'hui.

Cette recommandation est la synthèse du document technique n° : 20230315_DT02.

2 INTERETS POUR LES PLATES FORMES DU PLAN FRANCE MEDECINE GENOMIQUE

Ce kit permettrait :

- **D'abaisser la quantité de matrice** de départ nécessaire pour réaliser des bibliothèques, permettant de séquencer les échantillons avec peu d'ADN sans avoir recours à l'amplification par PCR,
- De l'utiliser pour tout type de matrice (tumoraux ou non) permettant de traiter des échantillons ayant de faible quantité de matrice,
- Une meilleure couverture **des régions GC**,
- Un **gain de temps**, lors de la préparation de bibliothèques,
- Un **gain économique**, l'étape de fragmentation ne demandant plus d'appareil spécifique (ni maintenance).

3 CHAMP D'APPLICATION DE LA RECOMMANDATION

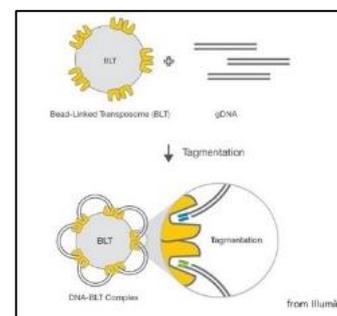
- Cette recommandation fait suite au document rédigé pour les échantillons non tumoraux (20221109_RT7_IlluminaDNA PCR-Free_Prep ref : 20221109_RT07)
- Cette recommandation s'applique aux ADNs extraits à partir de tissus congelés (FF), les échantillons de type FFPE ne sont pas concernés.

4 PRESENTATION DU KIT

Utilisation de transposases accrochées sur des billes magnétiques permettant une coupure des brins d'ADN.

Ces transposases coupent de façon aléatoire l'ADN et ajoutent une partie d'adaptateur sur les extrémités finales. La répartition des transposases sur les billes permet d'obtenir des coupures de tailles homogènes.

La quantité de billes par réaction est fixe, les billes sont saturées dès que la quantité d'ADN atteint les 300 ng.



5 PLAN DE TESTS

5.1 Echantillons

Les expériences ont été réalisées à partir d'échantillons de tissus tumoraux congelés (FF) et leur sang correspondant (PBMC) fournis par la Biobanque CRB-Tumorotheque de Nice BB-0033-00025 et le CLB du centre Léon Bérard.

Echantillon	BREAU *	PATIENT_03	PATIENT_04	PATIENT_08	PATIENT_10	PATIENT_11
Tissu	Sein	Poumon				
Type	-	Adénocarcinome acineux	Adénocarcinome acineux	Adénocarcinome micro-papillaire	Adénocarcinome solide	Adénocarcinome entérique
Cellules tumorales (%)	78	70	80	70	80	70

* BREAU = échantillon cancer sein de la plateforme AURAGEN

N°: 20230322_RT09 Version : 01 Date : 22/03/2023	Recommandation Technique PFMG 2025	 France Médecine Génomique 2025
Sujet	Illumina® DNA PCR-Free Prep	
Page 3/10		

5.2 Expériences

5.2.1 Extractions

Les extractions ont été réalisées en interne avec le kit QIAamp DNA mini kit (Qiagen ref : 51304) pour l'ensemble des échantillons provenant du CRB de Nice. L'ADN extrait a été fourni par le CLB de Lyon pour BREAU.

Les quantités et les qualités des ADNs extraits à partir FF et PBMC sont dans des valeurs attendues (quantité sup à 7 µg, DIN > 7, ...).

5.2.2 Librairies

Les librairies ont été préparées avec le kit Illumina® DNA PCR-Free Prep (TAG) avec une quantité de départ de 100 ng et 300 ng et pour certains en duplicat expérimental à partir des échantillons FF et PBMC. L'ensemble de ceux-ci ont aussi été préparés en TruSeq DNA PCR-Free (TRU) afin de réaliser une analyse par comparaison.

Echantillon	BREAU	PATIENT_03	PATIENT_04	PATIENT_08	PATIENT_10	PATIENT_11
TRU (quantité départ ng)	1000					
TAG (quantité départ ng)	300 D / 100 D	300 D/ 100 D	100	100 D	100 D	100 D

D : duplicat expérimental

L'ensemble des librairies a été déposé sur Novaseq 6000 d'Illumina. Les Flow Cell (FC) sont de type S2 et S4, le dépôt se fait en mode XP ou standard. Le choix des FC et du dépôt s'est fait en fonction du nombre d'échantillons à déposer et de la profondeur désirée.

La profondeur cible est de 80-90X afin de respecter les recommandations du PFMG 2025 pour les échantillons tumoraux et 30-40 X pour les échantillons normaux appariés.

6 RESULTATS

6.1 Qualités des librairies

		Quantité initiale d'ADN (ng)	Attendues (nM) Méd (min-max)	Observées (nM) Méd (min-max)	Résultats
TAG	PBMC	100	2,7 (1,7 - 4,2)*	7,7 (3,5 - 12,3)	✓
	Tissu			7,1 (4,1 - 11,5)	
	PBMC	300	≥ 7 **	9,6 (8,2 - 10,9)	
	Tissu			8,3 (6,6 - 10)	
TRU	PBMC	1000	> 15 *	23,7 (16 - 35,6)	✓
	Tissu			25,1 (13,2 - 41,2)	

* Valeurs obtenues lors des expériences sur des échantillons de référence non tumoraux.

** Valeurs seuils selon Illumina.

Les quantités de librairies obtenues sont dans la fourchette haute par rapport aux attendus. Il y a une plus grande dispersion des concentrations obtenues à 100 ng de matrice de départ par rapport aux concentrations à 300 ng. Les profils de migration, quelle que soit la quantité de départ, sont homogènes.

6.2 Evaluation de la qualité des runs de séquençage

L'analyse des runs obtenus à partir des librairies TAG (4 runs) et TRU (3 runs) ne montrent pas de différence entre eux, les valeurs des métriques (Passing filter, pré phasing, ...) sont dans les valeurs attendues. Il n'y a pas d'impact du kit sur la qualité des runs.

6.3 Validation technique

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'interprétations biologiques éventuelles. Il a été

N°: 20230322_RT09 Version : 01 Date : 22/03/2023	Recommandation Technique PFMG 2025	 France Médecine Génomique 2025
Sujet	Illumina® DNA PCR-Free Prep	
Page 4/10		

choisi d'uniformiser les fichiers fastq des librairies à un sous-échantillonnage à 350 Millions de paires de reads. Des indicateurs de performance propres à la méthode testée ont été définis et permettent une comparaison ciblée entre les différentes librairies.

6.3.1 Résultats des critères de performance

Indicateurs de performance	Seuils recommandés (FF)	FF TRU (1000 ng) Méd (Min-Max)	FF TAG (100 ng) Méd (Min-Max)	FF TAG (300 ng) Méd (Min-Max)
% Couverture 30X	>30	45 (21-58)	34 (27-41)	37,5 (35-40)
Taille insert (pb)	TRU: 380 +/- 50 TAG: 450 +/- 70	382 (341-422)	387 (353-404)	410 (409-412)
IQR	< 12	10 (9-11)	10 (10-11)	10 (10-10)
% Adaptateurs	< 5	0,16 (0,08-0,22)	2,76 (2,16-3,3)	2,29 (2,27-2,3)
% Alignement	>95	97,84 (97,42-98,16)	98 (97,91-98,22)	98,18 (98,15-98,22)
% Chimère	< 5	1,74 (1,51-2,01)	1,36 (1,28-1,52)	1,33 (1,32-1,34)
% Duplicats Optique	NA	11,36 (6,02-19,58)	10,04 (9,04-16,22)	9,05 (8,79-9,31)
% Duplicats Autre	< 10	1,17 (0,49-5,99)	4,39 (2,94-6,59)	5,42 (5,25-5,58)

Aucune déviation critique des indicateurs de performances n'est observée pour le kit TAG et entre les deux inputs (100ng et 300ng) comparé au kit TRU. Seul bémol, au niveau du pourcentage de séquences dupliquées « Autre » (colonne % Duplicats Autre) qui est supérieur à celui observé sur les librairies TRU. Cette observation n'est pas impactante outre mesure mais engendre un rendement plus bas sur le nombre de séquences utiles pour les librairies TAG. Les mêmes mesures ont été compilées sur les échantillons non tumoraux (PBMC) et les conclusions sont identiques à celles des échantillons FF.

6.3.2 Couverture des régions GC

La vérification de l'uniformité de couverture en fonction de la répartition du taux de GC dans le génome est réalisée à l'aide du logiciel Picard [1] (GATK v4.1.7 ; CollectGcBiasMetrics). Les résultats ont été effectués sur les échantillons PBMC et les FF entre les deux kits testés TRU et TAG. La distinction entre les deux inputs du kit TAG est prise en compte.

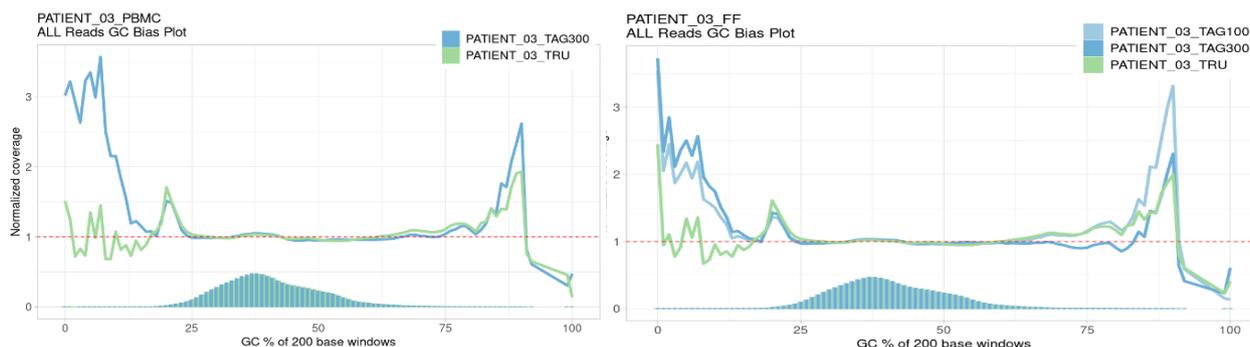


Figure 1: Représentation de l'uniformité de couverture normalisée selon le pourcentage de GC sur le génome pour le PATIENT_03 sur les librairies PBMC et FF.

Kit	AT DROPOUT	GC DROPOUT	NC GC<=20%	NC GC>=80%
TRU	1,24%	0,29%	1,10	3,21
TAG100ng	1,67%	1,85%	1,34	3,37
TAG300ng	1,27%	0,75%	1,37	2,88

Tableau 1: Mesures du AT/GC Dropout sur les échantillons PBMC, ainsi que sur la couverture normalisée (NC) sur les régions pauvres et riches en contenu GC.

Le pourcentage de séquences mal alignées en fonction du contenu des régions AT riches (<50% GC;) ou GC riches (>=50% GC) est représenté dans le tableau n°1, selon les mesures d'Illumina (AT/GC DROPOUT). Les valeurs ont été moyennées suivant les patients présents dans chaque catégorie de librairies (TRU, TAG100 & TAG300). Il semble y avoir un gain d'uniformité de

N°: 20230322_RT09 Version : 01 Date : 22/03/2023	Recommandation Technique PFMG 2025	 France Médecine Génomique 2025
Sujet	Illumina® DNA PCR-Free Prep	
Page 5/10		

couverture dans les régions riches en GC (>=80%) sur les librairies TAG en 300ng comparé aux deux autres types de librairies. Cependant, il n'y a pas d'amélioration visible dans les régions AT riches (<=20%) pour ces dernières. Les librairies TAG à 100ng présentent quant à elles **des valeurs d'uniformités moins efficaces dans les régions AT et GC riches.**

6.4 Analyses Biologiques

6.4.1 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) [2]. Dans un souci d'uniformité, les fichiers d'alignements sont downsampled à une profondeur propre à chaque patient, à l'aide de l'outil samtools (v1.13) [3]. Les profondeurs obtenues sont validées en amont par le logiciel Mosdepth (v0.3.2) [4].

La détection des variants somatiques est effectuée par Mutect2 (GATK v4.1.7.0) [5] et les statistiques et les figures sont générées à l'aide du logiciel R (v4.1.1).

6.4.2 Analyse des Variants

La vérification de la teneur en cellules tumorales des échantillons est réalisée par calcul bioinformatique, à partir des fichiers bam, à l'aide de l'outil FACETS (bioconductor/3.13) [6]. L'estimation de **la ploïdie et de la pureté est similaire entre duplicats expérimentaux, entre inputs différents (100ng & 300ng) et entre les deux versions des kits testés TRU et TAG.**

Avant toute chose, il a été réalisé **une analyse de comparaison entre les échantillons non tumoraux (PBMC)** afin de s'assurer de l'absence de biais liés aux kits TRU et TAG. Les différentes métriques calculées (nombre de variants, F1-Score, % de concordance génotypique, ...) **n'ont montré aucune différence entre les librairies PBMC.**

De même, un travail complet sur l'analyse des duplicats seuls entre les 2 quantités d'input à 100ng et 300ng n'a établi aucune divergence et valide la reproductibilité des résultats entre les duplicats TAG. Reste un point d'attention sur l'analyse des CNAs où les résultats montrent une certaine instabilité. Au vu de ces résultats, seul un duplicat de chaque catégorie d'input est présenté.

Une analyse globale et générale des résultats de Mutect2 est présentée dans le tableau ci-dessous. Le logiciel rvg vcfstats (v.3.12.1) [7] a permis de produire différentes métriques.

Echantillon	Failed Somatiques	Passed Somatiques	TAG/TRU ratio Passed	SNVs Somatiques	MNPs Somatiques	SNV Ti/Tv	Ins/Del ratio
BREAU_TRU	142 868	3 512		2 772	175	1,43	0,75
BREAU_TAG300-1	158 905	4 339	1,2	3 444	327	1,15	0,71
BREAU_TAG100-1	144 293	4 460	1,3	3 482	322	1,22	0,77
PATIENT_03_TRU	133 182	3 036		2 397	204	1,35	0,48
PATIENT_03_TAG300-1	159 072	4 144	1,4	3 379	300	0,96	0,61
PATIENT_03_TAG100-1	135 260	3 974	1,3	3 212	244	1,17	0,77
PATIENT_04_TRU	142 721	19 869		17 902	791	0,46	0,49
PATIENT_04_TAG100-1	134 210	20 651	1,0	18 701	898	0,45	0,49
PATIENT_08_TRU	153 085	153 082		141 425	5 607	0,41	0,38
PATIENT_08_TAG100-1	149 518	150 006	1,0	138 751	5 575	0,40	0,37
PATIENT_10_TRU	146 445	40 803		38 789	717	0,86	0,59
PATIENT_10_TAG100-1	153 794	41 648	1,0	39 557	848	0,85	0,62
PATIENT_11_TRU	140 149	65 453		60 358	2 030	0,32	0,50
PATIENT_11_TAG100-1	142 838	65 575	1,0	60 506	2 135	0,32	0,50

Tableau 2: Comparaison générale des variants somatiques entre les librairies TRU et TAG, dans les régions de haute confiance avec une profondeur normalisée par patient, Les librairies TRU servant de référence pour le calcul du ratio « Passed »

Dans la majorité des comparaisons, **le ratio sur la détection des variants somatiques entre les librairies TRU et TAG est de 1.** Comme, pour les ratios Transition/Transversion (Ti/Tv) et d'insertion/Deletion (Ins/Del). Ces valeurs sont équivalentes selon les deux inputs (100ng & 300ng) pour les librairies TAG. Les p-value des tests de student entre les deux conditions (TRUvsTAG100 & TRUvsTAG300) sont supérieures au seuil de 5%, ainsi, il n'y a pas de différence significative entre les SNVs et Indels pour ces 2 kits.

N°: 20230322_RT09 Version : 01 Date : 22/03/2023	Recommandation Technique PFMG 2025	 France Médecine Génomique 2025
Sujet	Illumina® DNA PCR-Free Prep	
Page 6/10		

6.4.3 Précision, Recall et F1-Score

Les indicateurs de performance Précision et Recall ont été calculés avec des seuils de VAF adaptés à la cellularité tumorale de chaque patient (c.a.d éliminer les SNVs de faible VAF, correspondant essentiellement aux erreurs de séquençage). Les SNVs détectés dans les échantillons TRU servent de référence pour les différents calculs. La Précision donne une indication sur les faux positifs (= FP, c.-à-d. les SNVs trouvés uniquement dans les échantillons TAG) et le Recall sur la part des faux négatifs (=FN, c.-à-d. les SNVs trouvés uniquement dans les échantillons TRU).

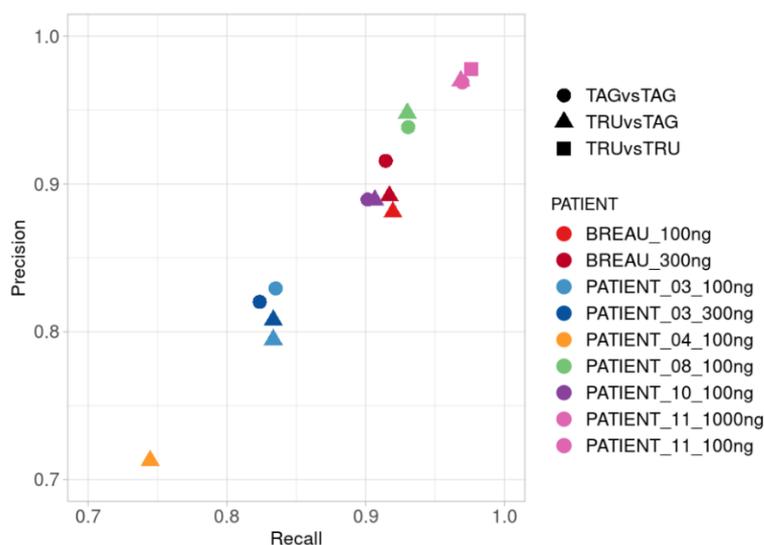


Figure 2: Représentation des valeurs du pourcentage de Précision et Recall

- **Les pourcentages de Recall et de Précision sont satisfaisants** pour l'ensemble des patients, exceptés pour le PATIENT_04 possédant un F1-Score seulement de 73%.
- **Bonne corrélation entre les deux inputs** (100ng & 300ng) pour le kit TAG, visible sur les patients BREAU et PATIENT_03 (représenté par les nuances claires/foncées).
- Performances similaires lors des comparaisons TAGvsTAG (duplicats) et TRUvsTAG pour l'ensemble des patients.
- **Les librairies TRU se comportent comme des duplicats TAG (et inversement)**. Appuyé, par l'ajout du F1-Score du PATIENT_11 réalisé également en duplicat par le kit TRU (TRUvsTRU, représenté par le carré rose).

Présence de nombreux variants communs (>88% et >95%) entre les deux versions de kits et entre inputs TAG 100ng et 300ng. De même, la cohérence dans le pourcentage de variants spécifiques entre chaque librairie, montre que les erreurs de séquençage sont comparables quel que soit le kit testé. Ainsi, une clusterisation et une stabilité des résultats sont obtenues par patient et non en fonction du kit des librairies TRU ou TAG : **les performances sont équivalentes.**

6.4.4 Copy Number Alteration

Les CNAs (Copy Number Alteration) sont obtenues en analysant les paires appariées PBMC-FF_TRU et PBMC-FF_TAG par le logiciel FACETS (bioconductor/3.13) [6], l'échantillon PBMC servant de normal apparié.

Le nombre d'évènements de CNAs entre les librairies TRU et TAG diverge selon les patients. La librairie TRU du PATIENT_11 présente un nombre d'amplifications et de gains plus important (3 fois plus) que pour la librairie TAG. Cette même tendance est visible sur le PATIENT_10. Inversement, la librairie TAG pour le patient BREAU montre un nombre de LOH légèrement supérieur (1,5 fois) mais observé uniquement avec l'input à 300ng.

Pour le PATIENT_11, la librairie TRU a occasionné un signal légèrement plus perturbé dans la détermination du nombre de copies (régions gagnées, encadré jaune sur la figure n°3), la cause principale étant une plus grande dispersion dans les valeurs de log2R que les librairies TAG. Cette variabilité semble avoir une distribution aléatoire et uniforme sur le génome.

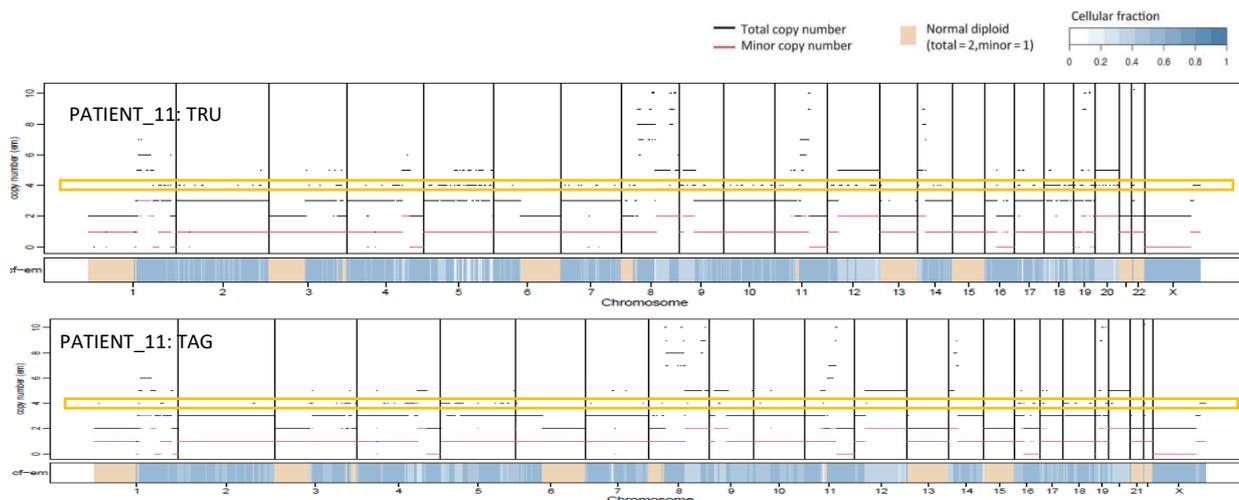


Figure 3: Plot du nombre de copies (total, mineur) et la fraction cellulaire suivant les chromosomes pour le PATIENT_11.

Des investigations supplémentaires devront être menées afin d'éclaircir ce phénomène, une première piste pourrait être la variabilité dans les dépôts des batches de séquençage. Aussi, d'autres logiciels pourront également être testés.

6.4.5 Signatures mutationnelles - Spectrum Mutations/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>) [8] a été utilisé à cet effet.

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.

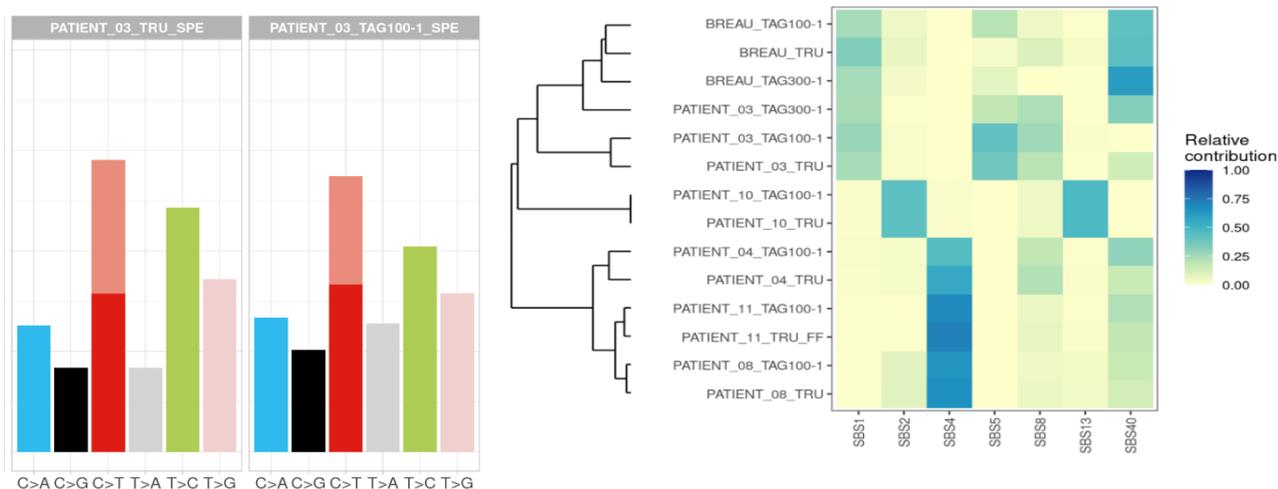


Figure 4: A gauche : Répartition des classes des SNVs spécifiques selon les librairies TRU et TAG sur le PATIENT_03 (mutations somatiques « Passed », régions de haute confiance, bam downsamplé et VAF >0). A droite : Heatmap avec dendrogramme permettant de visualiser la clusterisation des patients en fonction des SBS contributives à hauteur de 10%

La répartition des 6 classes de substitution est similaire entre les librairies TRU et TAG et également entre les 2 inputs (100ng & 300ng). **Le coefficient de similarité de cosinus est supérieur à 0,99** entre les profils des librairies TRU et TAG, et ce pour tous les patients. Les SBS COSMIC sont semblables entre les librairies TRU et TAG et **les patients clusterisent parfaitement ensemble au sein de la heatmap** (figure n°4).

N°: 20230322_RT09 Version : 01 Date : 22/03/2023	Recommandation Technique PFMG 2025	
Sujet	Illumina® DNA PCR-Free Prep	
Page 8/10		

6.4.6 Génotype Concordance, Hotspots et SVs

Le pourcentage de concordance génotypique, calculé sur les variants communs, est identique entre les deux versions de kits de librairie et entre les deux inputs (100ng et 300ng) pour TAG (97% de concordance pour les SNVs et 95% pour les Indels).

La liste de hotspots du catalogue de CancerGenomeInterpreter (2749 mutations; CGI ; <https://www.cancergenomeinterpreter.org/mutations>) est utilisée pour faciliter l'identification des altérations tumorales de patients qui pourrait être cliniquement exploitables. **Pas de perte de hotspots** entre les comparaisons des librairies et sont tous retrouvés comme « Passed » lors de l'analyse Mutect2. **Les VAF sont comparables entre elles**, sauf pour le gène ERBB2 pour le patient BREAU où la VAF varie entre 6% (TAG100ng) et 14% à 16% respectivement pour le kit TRU et TAG300ng. Bien que cette observation soit sporadique à un hotspot, elle devra faire l'objet d'une surveillance.

Le logiciel d'Illumina Manta (v1.0.3) [9] est utilisé pour la détection de variants structurels ou SVs de tous types (délétions, duplications, insertions, translocations et remaniements complexes). La fusion et la comparaison des événements au sein de chaque patient sont réalisées à l'aide du logiciel SURVIVOR (v1.0.3) [10] à partir de fichiers vcf (régions de haute confiance et bam downsamplé). De façon globale, **un ratio de 1 est obtenu après comparaison des événements de SVs** entre les librairies TRU et TAG et entre les différents inputs (100ng vs 300ng) : **aucune disparité n'est visible**.

7 LES POINTS DE VIGILANCES

Les points de vigilances déjà évoqués dans le document n° 20221109_RT07 sont toujours à prendre en considération. Pour rappel :

- le dosage des librairies (libraires en simple brin),
- la création du pool,
- l'influence des index sur le rendement du séquençage,
- le nombre de librairies par Flow Cell, lié à un taux de duplicats plus important,
- la capacité de re-dépôt, liée à une plus faible quantité de librairies générées,

Concernant les régions riches en GC, le CRefIX a noté **un léger gain d'uniformité** avec une quantité de 300 ng par rapport à 100 ng. Malgré une légère baisse de l'uniformité dans le cas des expériences à 100 ng, **cet impact semble minime**.

L'analyse des CNAs étant sujet à plusieurs facteurs (logiciels, design d'expérience, ...) le CRefIX ne statue pas sur l'équivalence entre les deux kits, et attire l'attention des futurs utilisateurs sur ce point. Des analyses supplémentaires sont en cours.

N°: 20230322_RT09 Version : 01 Date : 22/03/2023	Recommandation Technique PFMG 2025	 France Médecine Génomique 2025
Sujet	Illumina® DNA PCR-Free Prep	
Page 9/10		

8 CONCLUSION GENERALE

	TRU	TAG	Commentaires
Quantité	☹️ --	😊 ++	1/10 en moins
Temps	☹️ --	😊 ++	1h30 au lieu de 5h00
Bio-IT	😊 ++	😊 ++	- Aucune disparité visible sur les indicateurs de performances
GC Bias	😊 ++	100ng : 😊 + 300ng : 😊 ++	- Gain d'uniformité sur les régions GC pour le kit TAG à 300ng - Valeurs d'uniformité moins efficaces dans les régions AT et GC riche pour l'input à 100ng pour TAG
Pureté et Ploidie	😊 ++	😊 ++	- Estimation de la pureté tumorale et ploïdie similaires entre les deux versions de kits
Général Overview	😊 ++	😊 ++	- Nombre de variants PASS et FAILED similaires (Ratio=1) entre les deux versions de kits - Différence non significative (P-value > 5%) - Aucune disparité visible sur les différentes métriques
Précision, Recall, F1 score	😊 ++	😊 ++	- La précision, le % de Recall et le F1-score >= à 80% - Performances similaires entre les résultats TAGvsTAG, TRUvsTAG et TRUvsTRU
Signature mutationnelle	😊 ++	😊 ++	- Les profils mutationnels sont identiques entre les deux versions de kits - Coefficients de similarité de cosine sont >=0,99 (~ à 1) - Corrélations des patients prouvées entre les deux versions de kits
Hotspots	😊 ++	😊 ++	- Pas de perte de hotspots - VAF similaires (sauf pour 1 hotspot pour le patient BREAU_TAG100-1)
CNAs	☹️ -	☹️ -	- Profils similaires entre les deux versions de kits - Disparités visibles sur le nombre d'évènements entre les deux kits (Ratio != 1) - Plus d'amplifications/Gains pour le kit TRU et plus de LOH pour le kit TAG
SVs	😊 ++	😊 ++	- Nombre d'évènements similaires entre les deux versions de kits (Ratio ~1)

Au vu des avantages liés à ce kit, le CRefIX considère qu'il peut être utilisé dans le cas d'échantillons tumoraux. Selon la qualité de l'échantillon, il est **possible de réaliser** des librairies avec une quantité de **matrice de 100ng**. Si la quantité initiale n'est pas limitante, **le CRefIX recommande une quantité de 300ng**.

N°: 20230322_RT09 Version : 01 Date : 22/03/2023	Recommandation Technique PFMG 2025	 France Médecine Génomique 2025
Sujet	Illumina® DNA PCR-Free Prep	
Page 10/10		

9 DOCUMENTS

- Illumina DNA PCR-Free Reference Guide :
Comprehensive information on Illumina DNA PCR-Free Library Prep, including a detailed protocol, Ref Document # 100000086922 v03 (February 2021)
- TruSeq DNA PCR-Free :
 - TruSeq DNA PCR-Free Reference Guide Ref Document # 1000000039279 v00 (October 2017)
- Illumina DNA PCR-Free Checklists :
 - Illumina DNA PCR-Free Library Prep Standard Input Checklist Ref Document # 1000000086923 (Jul 21, 2020)
 - Illumina DNA PCR-Free Library Prep Low Input Checklist Ref Document # 1000000130333 (Jul 21, 2020)
 - Illumina DNA PCR-Free Library Prep Hybex Protocol Checklist Ref Document # 1000000130364 (Jul 21, 2020)
- Bruinsma S, et al, Bead-linked transposomes enable a normalization-free workflow for NGS library preparation. BMC Genomics, 2018 Oct 1;19(1):722.
- [1] Chen YC, Liu T, Yu CH, Chiang TY, Hwang CC, Effects of GC bias in next-generation-sequencing data on de novo genome assembly, PLoS One, 2013 Apr 29;8(4):e62856, doi: 10.1371/journal.pone.0062856, PMID: 23638157; PMCID: PMC3639258.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] Calling Somatic SNVs and Indels with Mutect2, David Benjamin, Takuto Sato, Kristian Cibulskis, Gad Getz, Chip Stewart, Lee Lichtenstein bioRxiv 861054; doi: <https://doi.org/10.1101/861054>
- [6] 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,
- [7] RTG Tools: Utilities for accurate VCF comparison and manipulation, <https://github.com/RealTimeGenomics/rtg-tools>
- [8] 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.
- [9] 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.
- [10] 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 HISTORIQUE DES MODIFICATIONS

Version	Date	Nature de la modification
V01	21/03/2023	Revue et correction
V00	01/02/2023	Création

11 VALIDATION DU DOCUMENT

RÉDACTEURS	RELECTEURS
Edouard Turlotte, Mélanie Letexier	Jean François Deleuze, Alain Viari, Violette Turon
PLAN EXPERIMENTAL	Fonction : Direction Unité
Edouard Turlotte, Margaux Gras, Alice Moussy	Date : 22/03/2023
ANALYSE	Visa :
Mélanie Letexier, Marine Rouillon	
Date : 22/03/2023	
Visa :	