

FICHE PRODUIT : EPIGENOMIQUE

L'application de ChIP-seq qui combine l'ImmunoPrécipitation de Chromatine (ChIP) et le séquençage à haut débit, permet de cartographier à l'échelle du génome les sites d'interaction entre l'ADN et des protéines régulant l'expression des gènes tels que les facteurs de transcription ou les modifications post-traductionnelle des histones. Nous proposons des prestations de ChIP-seq standard ou de Cut & Run¹. Il est également possible d'analyser les modifications (méthylation, hydroxy-méthylation,...) des nucléotides via le MeDIP-seq².

1 Protocoles de préparation de bibliothèques et options de séquençage proposés

1.1 Protocoles de préparation de bibliothèques

Plusieurs protocoles de préparation de bibliothèques sont actuellement proposés par la plateforme :

#	Application	Kit utilisé par la plateforme	Quantité de chromatine	
			Minimale	Optimale
1	ChIP-seq	Diagenode MicroPlex	2 ng	10 ng
2	Cut & Run	Diagenode MicroPlex	2 ng	10 ng
3	MeDIP-seq	Takara DNA SMART ChIP-Seq	2 ng	20 ng

1.2 Options de séquençage

Les bibliothèques sont séquençées à l'aide de la technologie NextSeq 2000 d'Illumina. Nous pouvons générer des lectures simples (single-read) ou paires (paired-end) de 50 pb. Le tableau ci-dessous fournit des recommandations de type de séquençage en fonction des objectifs du projet. La profondeur de séquençage dépend de la finalité du projet, de la nature des échantillons et du protocole de préparation de bibliothèques. Nous encourageons les porteurs de projet à nous contacter afin d'obtenir plus d'informations sur ces différentes options de séquençage si nécessaire.

¹ Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*. 6:e21856. doi:10.7554/eLife.21856

² Mohn F, Weber M, Schübeler D, Roloff T-C. Methylated DNA Immunoprecipitation (MeDIP). In: Tost J, ed. *DNA Methylation: Methods and Protocols*. Methods in Molecular Biology. Humana Press; 2009:55-64. doi:10.1007/978-1-59745-522-0_5

Application	Recommandations			
	Type de séquençage	Taille de lecture	Nombre de millions de reads par échantillon	PhiX*
ChIP-seq - cas général	Single-end	50 pb	≥30	-
ChIP-seq - cas particuliers (analyse de certaines marques d'histones comme H3K27me3, analyses de liaison dans les régions répétées, ...)	Paired-end	50 pb	≥30	-
Cut & Run	Paired-end	50 pb	≥10	-
MeDIP*	Single-end	50 pb	≥30	25%

*Les librairies de MeDIP possèdent un adaptateur de 9 nucléotides au début des séquences qui déséquilibre la composition nucléotidique au début du run, au moment de l'identification des clusters. Pour le NextSeq 2000, ceci conduit à la réduction du nombre de clusters détectés et à une baisse de la qualité des séquences. Illumina recommande l'ajout d'une librairie Spike-in (Illumina PhiX control) pour rééquilibrer les bases³.

1.3 Plan d'expérience

Il est recommandé⁴ de réaliser au moins 2 réplicats biologiques (ce nombre est fortement dépendant de la variabilité entre les réplicats). Il faut également prévoir pour chaque condition expérimentale un control négatif indispensable lors de l'analyse des données pour extraire les enrichissements spécifiques par rapport au bruit de fond (voir tableau ci-dessous). De plus, nous recommandons d'essayer dans la mesure du possible de réduire les effets de lots durant la préparation des échantillons. Enfin, nous encourageons les porteurs de projets à nous contacter avant de commencer les expériences pour discuter de la planification expérimentale.

Application	Contrôle négatif recommandé (pour chaque condition)
ChIP-seq	<ul style="list-style-type: none"> – ADN du même échantillon fragmenté mais non immunoprécipité (« Input DNA ») – ADN du même échantillon immunoprécipité avec un anticorps non spécifique (par exemple un anti IgG) – ADN de la souche sauvage immunoprécipité avec le même anticorps que l'ADN du mutant étudié – ADN du même échantillon immunoprécipité sans anticorps (mock)
Cut & Run	– ADN du même échantillon immunoprécipité avec un anticorps non spécifique (par exemple un anti IgG)
MeDIP	

³ <https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/hiseq-phix-control-v3-technical-note.pdf>

⁴ Landt, Stephen G. et al. "ChIP-Seq Guidelines and Practices of the ENCODE and modENCODE Consortia." *Genome Research* 22.9 (2012): 1813–1831. PMC. Web. 13 Mar. 2017.

2 Services proposés

1. La vérification des échantillons :

- Quantification et vérification de l'ADN par fluorimétrie (Qubit ou Varioskan) et électrophorèse capillaire (Bioanalyzer d'Agilent ou Fragment Analyzer d'AATI), lorsque la quantité de matériel de départ le permet.

2. La préparation de bibliothèques :

- Préparation des bibliothèques et ligation d'adaptateurs de séquençage possédant un index. Les index sont des séquences d'ADN utilisées pour identifier chaque échantillon. L'utilisation d'index permet de séquencer plusieurs échantillons sur une même flow cell.
- Quantification et vérification de la qualité des bibliothèques par électrophorèse capillaire (Bioanalyzer d'Agilent ou Fragment Analyzer d'AATI).

3. Le séquençage avec le séquenceur NextSeq 2000 d'Illumina :

- Simple ou paillé avec des tailles de lectures selon les options spécifiées sur le LIMS (<http://ngs-lims.igbmc.fr>) pour chaque projet.

4. L'analyse primaire des données :

- Démultiplexage et création des fichiers FASTQ.
- Contrôle de la qualité des séquences.
- Détection d'éventuelles contaminations.
- Création d'un rapport synthétisant les méthodes utilisées par le pipeline d'analyse primaire et les résultats obtenus.

5. L'analyse ultérieure des données (optionnelle, voir le paragraphe 5 pour plus d'informations).

3 Préparation des échantillons par le porteur de projet

Le porteur de projet réalise la préparation des échantillons d'ADN enrichis. Le succès final de l'expérience est étroitement lié à la qualité des échantillons de départ. Un soin tout particulier doit être pris pour éviter toute trace de contamination ou de dégradation.

Recommandations pour la préparation des échantillons	
Agent de blocage de l'IP	Éviter si possible l'ADN (de sperme de saumon, etc.) comme agent de blocage de l'IP. Utiliser, par exemple, de l'ARNt de levure.
Validation de l'enrichissement de l'IP	Par PCR quantitative sur des cibles connues.

Caractéristiques de l'ADN à fournir	
Quantité de matériel de départ	En fonction du protocole de préparation des bibliothèques.
Volume minimal	10 µl.
Taille des fragments d'ADN	L'image d'électrophorèse sur gel d'agarose ou un profil d'électrophorèse capillaire (p. ex. Bioanalyzer Agilent) de l'ADN fragmenté est souhaitable. Le fichier peut être téléversé sur le LIMS de la plateforme (http://ngs-lims.igbmc.fr) dans le

	projet correspondant. La taille moyenne des fragments doit être inférieure à 500 pb. La plateforme recommande de fragmenter par sonication (p. ex. Covaris).
Qualité	L'ADN doit être dépourvu de contaminants qui pourraient inhiber les réactions enzymatiques de préparation des librairies (protéines, EDTA, sels, solvants, etc.). Une purification supplémentaire, utilisant les billes AMPure XP ou SPRIselect (Beckman Coulter), est indispensable. Si non effectuée par le porteur de projet, cette purification sera automatiquement réalisée par la plateforme avant quantification des échantillons de départ sauf pour les projets de Cut & Run pour préserver les petits fragments d'ADN.
Conditions d'envoi	En solution dans de l'eau dans des tubes « <i>low binding</i> » pour limiter la perte d'ADN par adsorption sur la surface du tube, expédié sur de la carboglace. Les échantillons doivent être enregistrés dans le LIMS puis l'identifiant unique généré pour chaque échantillon par le LIMS doit être clairement indiqué sur son tube.

4 Contrôles qualité réalisés par la plateforme

Les contrôles qualité listés ci-dessous sont effectués par la plateforme. Les contrôles qualité des étapes 1 et 2 sont accessibles sur le LIMS de la plateforme (<http://ngs-lims.igbmc.fr>).

1. Vérification des échantillons	
Quantité après purification sur billes (Fluorimétrie)	≥ quantité minimale requise (dépend du protocole)
2. Préparation des librairies	
Profil de la librairie (électrophorèse capillaire)	Taille moyenne comprise entre 200 et 600 pb.
Pureté de la librairie (électrophorèse capillaire)	Présence minoritaire de dimères d'adaptateur (bande à 120-130 pb) et de primers (bandes à 30-60 pb).
3. Séquençage et analyse primaire des données	
Nombre total de clusters* par projet (incluant le PhiX)	≥ nombre total de clusters indiqué dans le paragraphe « Requested services » de la fiche de soumission (fichier pdf téléchargeable dans l'onglet « Document » du LIMS http://ngs-lims.igbmc.fr , pour chaque projet)
Scores de qualité (Phred score) > 30	≥ 85% des bases

* Nombre de reads en single-read et nombre de reads ÷ 2 en paired-end

5 Livraison des résultats par la plateforme

Pour chaque échantillon, les résultats suivants sont mis à la disposition du porteur du projet :

- Les données brutes de séquençage (séquences nucléotidiques au format FASTQ).
- Un rapport (fichier PDF) présentant le nombre de lectures brutes, le pourcentage de bases ayant un score de qualité Phred supérieur à 30 et la taille de tous les fichiers bruts (FASTQ) à télécharger ;
- Un fichier texte fournissant la chaîne de caractère MD5 associée à chaque fichier FASTQ à télécharger. Le porteur de projet est responsable du téléchargement de ses fichiers, de la vérification de leur intégrité à partir des chaînes de caractères MD5 et de leur stockage. Une documentation est disponible à l'adresse suivante : <http://genomeast.igbmc.fr/wiki/doku.php?id=help:md5>.

Pour le ChIP-seq :

- Les fichiers de résultats de l'alignement sur le génome de référence (fichiers BAM, BED, BIGWIG)

Un e-mail de livraison des données informe le porteur de projet qu'il peut télécharger ses données de séquences en utilisant un login et un mot de passe sur le serveur FTP de la plateforme.

Conformément aux « Conditions Générales de la Plateforme GenomEast », il est rappelé qu'après la livraison de ses données, le porteur du projet est responsable de leur sauvegarde et de leur archivage. Les données seront supprimées du serveur de la Plateforme 6 mois après leur mise à disposition.

6 Analyse ultérieure des données (optionnelle)

L'analyse ultérieure des données n'est pas prise en charge dans la prestation standard. Néanmoins, elle peut être réalisée sous la forme d'une collaboration avec des membres de la plateforme. Par exemple, pour le CHIP-seq, elle peut comprendre :

- La détection des pics : localisation des régions d'ADN liées par les protéines ou les marques d'histone d'intérêt.
- L'annotation des pics : localisation des pics par rapport à des annotations fonctionnelles du génome de référence (distance au TSS, localisation intronique, exonique, intergénique, etc.).
- Découverte de motif de novo ou recherche de motifs à partir de motifs connus dans des bases de données (p. ex. JASPAR).
- Clustering des pics : l'objectif est de rassembler en plusieurs groupes les pics du génome ayant des profils d'enrichissement similaires.

La liste ci-dessus n'est pas exhaustive et nous recommandons aux porteurs de projet qui souhaiteraient initier une collaboration avec la plateforme de nous contacter avant de commencer leur projet pour que nous puissions les aider à définir les analyses qui pourraient répondre au mieux aux questions biologiques posées.