

FICHE PRODUIT : CHIP-SEQ

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.

1 Protocoles de préparation de bibliothèques et options de séquençage proposés

1.1 Protocoles de préparation de bibliothèques

Un protocole de préparation de bibliothèques est actuellement proposé par la plateforme.

#	Kit utilisé par la plateforme	Quantité de chromatine	
		Minimale	Optimale
1	Diagenode MicroPlex	2 ng	10 ng

1.2 Options de séquençage

Les bibliothèques seront séquençées à l'aide de la technologie HiSeq 4000 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 longueur de séquençage en fonction des objectifs du projet. Nous encourageons les porteurs de projet à nous contacter afin d'obtenir plus d'informations sur ces différentes options de séquençage si nécessaire.

Projet	Recommandations	
	Type de séquençage	Taille de lecture
Cas général	Single-end	50 pb
Cas particuliers : - Analyse de certaines marques d'histones (ex : H3K27me3) - Analyses de liaison dans les régions répétées	Paired-end	50 pb

Pour des génomes de mammifères, nous recommandons dans le cas général de multiplexer les échantillons par 8 sur une seule ligne de flow cell. En fonction des protéines étudiées, une couverture plus importante peut être nécessaire demandant ainsi un multiplexage avec moins d'échantillons.

1.3 Plan d'expérience

Recommandations pour le design de l'expérience

Contrôles négatifs	Il est fortement recommandé de prévoir un contrôle négatif pour améliorer la détection des régions enrichies spécifiquement par l'IP. Par exemple : <ul style="list-style-type: none"> - ADN de la souche sauvage immunoprécipité avec le même anticorps que l'ADN du mutant étudié. - 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 du même échantillon immunoprécipité sans anticorps (mock).
Réplicats biologiques	Il est recommandé ¹ de réaliser au moins 2 réplicats biologiques (ce nombre est fortement dépendant de la variabilité entre les réplicats). Pour toute question sur le design expérimental, n'hésitez pas à contacter la plateforme avant le début du projet.

2 Services proposés

1. La vérification des échantillons :
 - Quantification par fluorimétrie (Qubit).
2. La préparation des bibliothèques :
 - Préparation des bibliothèques et ligation d'adaptateurs de séquençage possédant un index. Les index sont des séquences d'ADN de taille supérieure ou égale à 6 bases, utilisées pour identifier chaque échantillon. L'utilisation d'index permet de séquencer plusieurs échantillons sur une même piste de 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 HiSeq 4000 d'Illumina :
 - Chargement des bibliothèques sur la flow cell et génération de clusters avec la Cbot (Illumina).
 - Séquençage simple ou paillé avec des tailles de lectures de 50 b selon les options choisies par le porteur de projet.
4. L'analyse primaire des données :
 - Démultiplexage et création des fichiers FASTQ.
 - Suppression des dimères d'adaptateur.
 - Contrôle de la qualité des séquences.
 - Détection d'éventuelles contaminations.
 - Alignement des séquences sur un génome de référence, s'il est disponible.
 - Création d'un rapport synthétisant les méthodes utilisées par le pipeline d'analyse primaire et les résultats obtenus (un rapport détaillé pour chaque échantillon et un rapport global plus synthétique pour chaque projet).
5. L'analyse ultérieure des données (optionnelle, détaillée dans le paragraphe 5).

¹ 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.

3 Préparation des échantillons par le porteur de projet

La préparation des échantillons d'ADN enrichis par CHIP est réalisée par le porteur de projet. L'amplification de l'ADN (WGA, etc.) n'est pas recommandée.

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 (Input ou CHIP)	≥ quantité minimale requise (dépend du protocole ; mesurée au fluorimètre, p. ex. Qubit).
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 à fournir obligatoirement avec les échantillons. Le fichier correspondant peut être chargé dans la partie « <i>share documents</i> » du LIMS de la plateforme (http://ngs-lims.igbmc.fr). 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, sels, solvants, etc.). Une purification supplémentaire, utilisant les billes Agencourt AMPure XP (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.
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 le tube. Les noms des échantillons doivent être clairement identifiés sur les tubes et dans le LIMS de la plateforme.

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>). En cas de problème, le porteur de projet est prévenu par e-mail.

1. Vérification des échantillons	
Quantité après purification sur billes AMPure XP (fluorimètre)	≥ 2 ng
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)
3. Séquençage et analyse primaire des données	

Nombre total de clusters (i.e. nombre de reads en Single-read et nombre de reads ÷ 2 en Paired-end)	≥ 250 x L millions, où L est le nombre total de lignes demandées par le porteur de projet
Scores de qualité (Phred score) > 30	≥ 85% des bases pour des lectures de 50 b.

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 ayant passé le filtre qualité et ne correspondant pas à des dimères d'adaptateur).
- Les fichiers de résultats de l'alignement sur le génome de référence (fichiers BAM et BED).
- Un rapport présentant les contrôles qualité de séquençage ainsi que les statistiques d'alignement (fichier PDF).

Deux fichiers additionnels sont fournis par projet :

- 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 caractères MD5 associée à chaque fichier FASTQ à télécharger. Le porteur de projet peut utiliser ces informations pour vérifier l'intégrité des fichiers après leur téléchargement (voir documentation : <http://genomeast.igbmc.fr/wiki/doku.php?id=help:md5>).

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 que le porteur du projet est responsable de la sauvegarde et de l'archivage de ses données. La plateforme ne s'engage à les conserver que sur une durée limitée de six 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. 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.