Année 2020, n°03

Date de parution: Juin 2020

Journal TGML

Transcriptomic and Genomic Marseille Luminy

La plateforme « Transcriptomique et Génomique Marseille-Luminy » est intégrée au laboratoire TAGC, unité mixte de recherche Inserm/AMU et est une composante de l'infrastructure distribuée France Génomique.

L'ensemble de l'activité de la plateforme est certifiée ISO 9001:2015 et NFX 50-900:2016. Le système de management de la qualité concerne : Expertise, Recherche et Développement en génomique et bioinformatique dans les sciences du vivant.

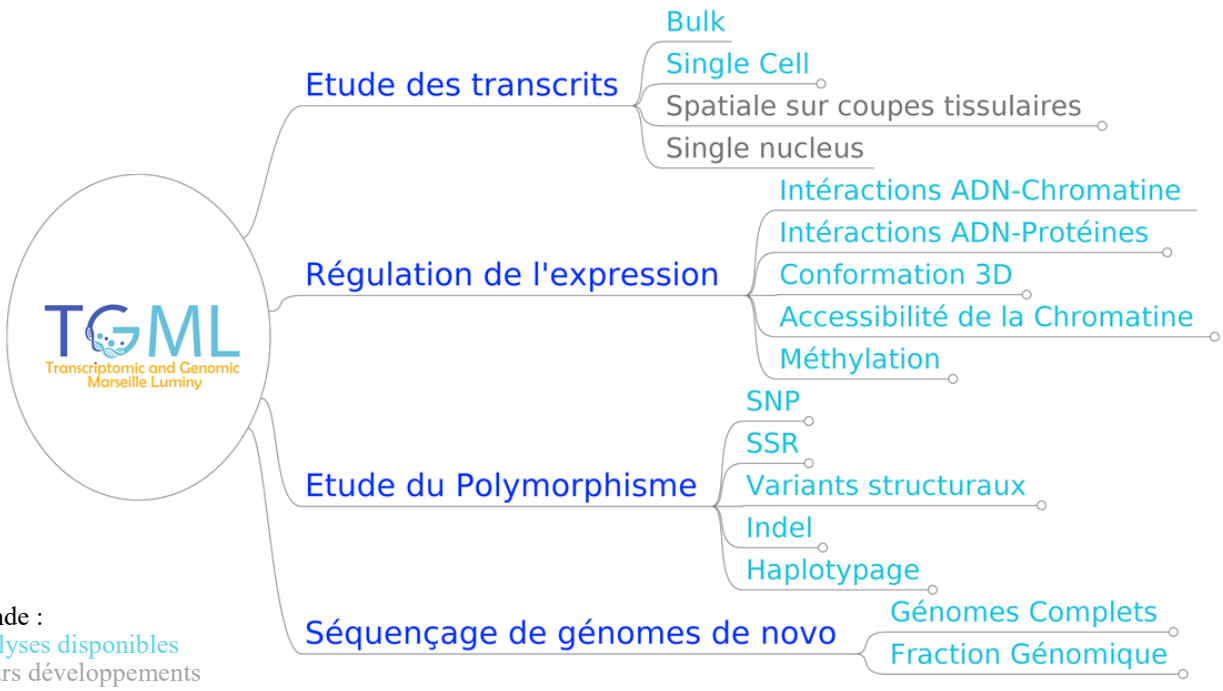
Accessible aux équipes de recherche publique françaises et étrangères, ainsi qu'aux industriels, elle donne accès à une large palette d'outils expérimentaux et bio-informatiques pour les études de génomique fonctionnelle notamment l'analyse transcriptome. L'accès à ces services se fait soit en prestation de service soit en collaboration sur la base d'une tarification adaptée et validée par l'Inserm. Les ingénieurs de la plateforme maîtrisent la procédure complète de traitement des échantillons depuis leur contrôle qualité jusqu'à l'obtention de données analysées.

PLATEFORME TECHNOLOGIQUE AIX-MARSEILLE

La plateforme a obtenu le label « Plateforme technologique Aix-Marseille » en 2016 et son renouvellement en mars 2019 pour une durée de trois ans.

Ce numéro se concentre sur l'aspect bioinformatique de la plateforme, les prestations proposées, les développements récents et à venir.

Prestations bioinformatiques de la plateforme



Légende :
- Analyses disponibles
- Futurs développements

Séquençage du transcriptome en « cell hashing »

- **Principe**

Le « cell hashing » est une méthode permettant de multiplexer les échantillons séquencés en cellule unique afin de réduire les biais d'expérience. Cette technique consiste à marquer les cellules d'un même échantillon avec un HTO (Hashtag Oligo) unique via des anticorps. Ces derniers sont identiques pour tous les échantillons étudiés, et reconnaissent des protéines de surface ubiquitaires. Le séquençage des HTO, différent pour chaque échantillon, permet d'identifier pour chaque cellule son échantillon d'origine et les doublets provenant de plusieurs échantillons.

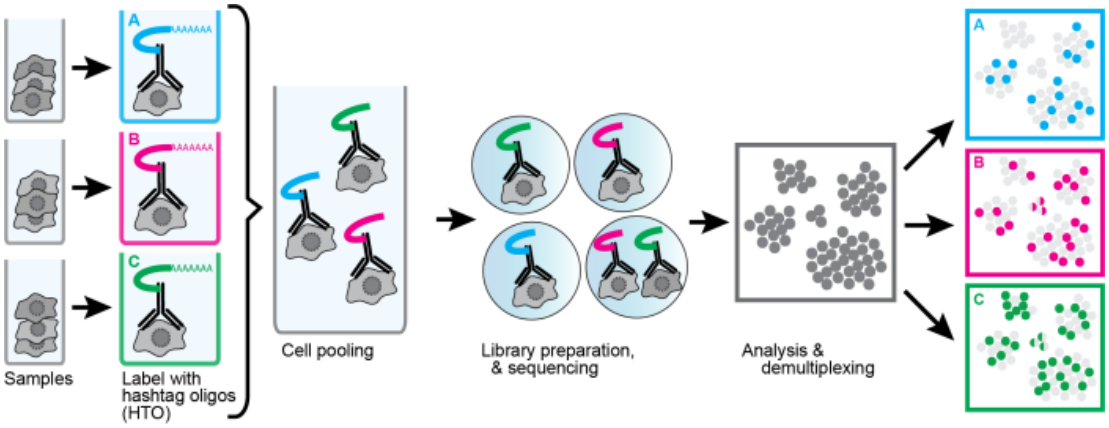
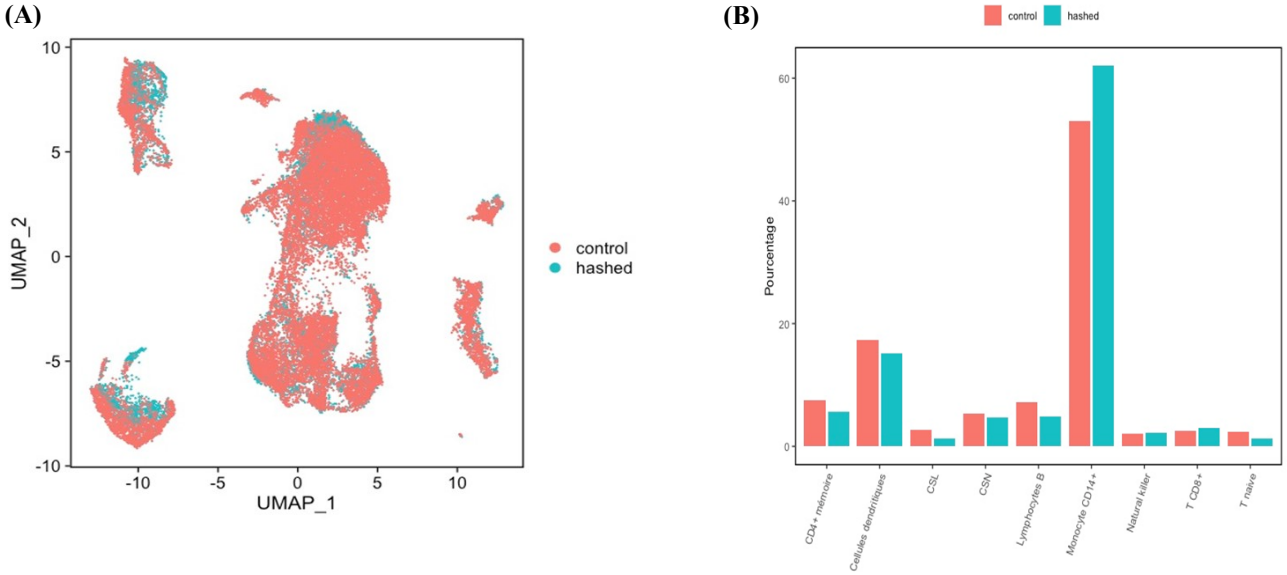


Schéma général du multiplexage d'échantillons scRNA-seq avec la technique du cell hashing.
Stoeckius, M., Zheng, S., Houck-Loomis, B. *et al.* **Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics.** *Genome Biol* 19, 224 (2018).

- **Validation de la technique**

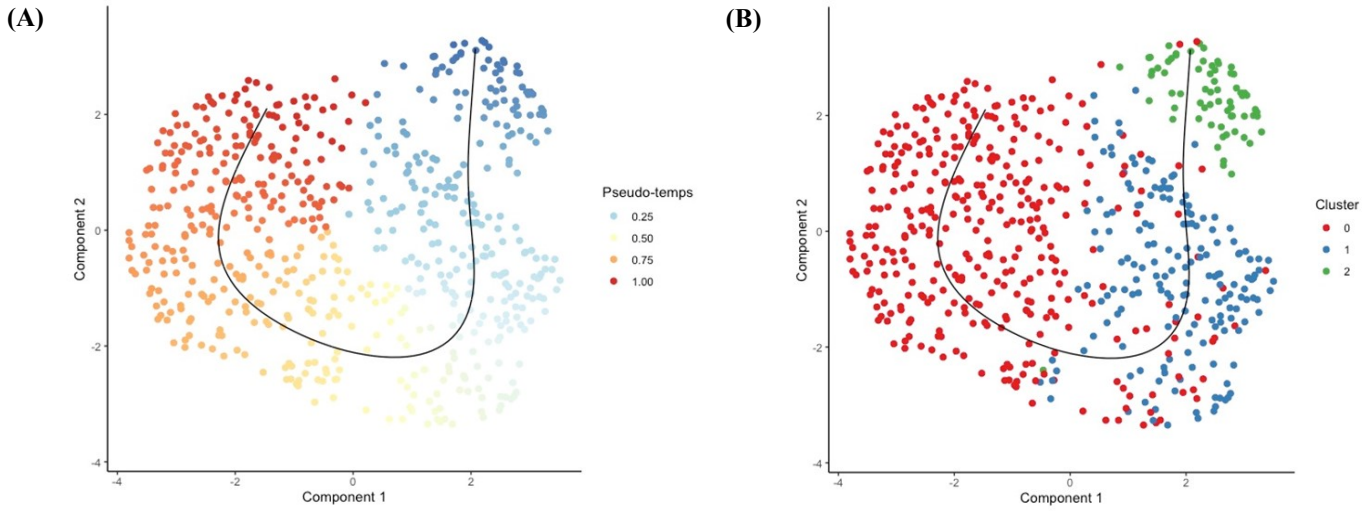
La technique a été validée lors d'une étude comparative entre des données résultant d'un scRNA-seq classique et celles d'un scRNA-seq en multiplexant les échantillons provenant de deux réplicats biologiques issus de PBMC de patients atteints de leucémies aigües myéloïdes (LAM). Le profil transcriptomique observé sur une projection UMAP (Uniform Manifold Approximation and Projection) (**A**) et la proportion des types cellulaires **(B)** sont conservés entre les deux techniques.
Ce projet a fait l'objet d'un stage de master II sur la plateforme.



Analyse comparative de données scRNA-seq multiplexées et non multiplexées. Protocole en scRNA-seq non multiplexé (rouge) et multiplexé (bleu).
(A) Profil transcriptomique de chaque cellule (points) représenté par un UMAP (Uniform Manifold Approximation and Projection).
(B) Représentation de la proportion des types cellulaires.

- **Collaboration avec l'équipe « Cellules souches épithéliales et cancer » du CRCM sur la radiorésistance des cancers du sein (Emmanuelle Charafe-Jauffret, Christophe Ginestier)**

Dans le cancer du sein, une sous population cellulaire appelée « souches cancéreuses mammaires » (bCSC) récapitule l'hétérogénéité de la tumeur initiale et serait à la base des résistances thérapeutiques. La radiothérapie modifie l'homéostasie tumorale en enrichissant la tumeur en bCSC potentiellement responsables des récidives. Parmi les mécanismes expliquant cette radiorésistance, l'irradiation pourrait induire la genèse de bCSC (ibCSC) à partir de cellules cancéreuses plus matures. L'objectif est d'étudier les mécanismes responsables de cette reprogrammation cellulaire par scRNA-seq. Le but final est de cibler la plasticité cellulaire radio-induite afin de radio-sensibiliser les bCSC.



Analyse de la reprogrammation cellulaire par inférence de trajectoire.
(A) L'inférence de trajectoire détermine un processus cellulaire et ordonne les cellules le long de la trajectoire en fonction de leur progression (pseudo-temps) dans le processus.
(B) La trajectoire a pour but de comprendre la dynamique transcriptionnelle responsable de la reprogrammation cellulaire des cellules matures précoces et tardives (cluster 2 et 1) permettant la genèse de ibCSC (cluster 0).

Dernier développement : MetaWorkflow

- **Description**

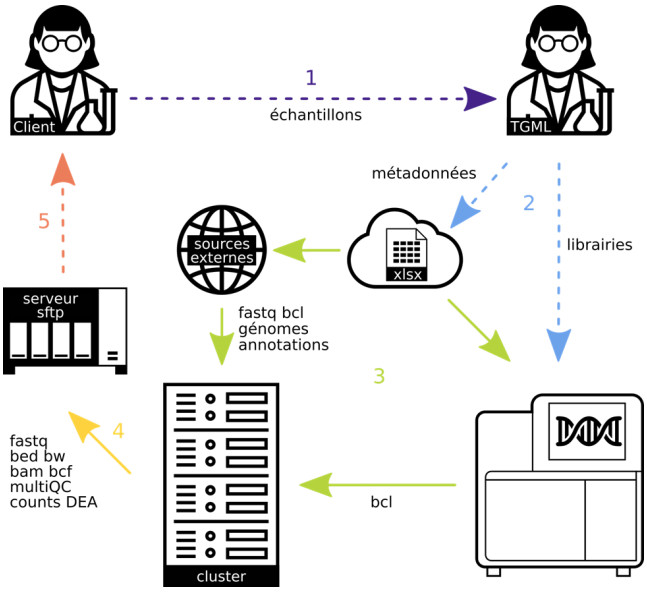
La plateforme s'implique particulièrement dans le développement des approches *FAIR* ("Facile à trouver, Accessible, Interopérable et Réutilisable") pour les données et analyses produites. L'accent est mis sur la mise en place de traitements de données reproductibles en utilisant les technologies basées sur des gestionnaires d'environnements et/ou la conteneurisation.

Depuis peu, la plateforme s'est dotée d'un workflow d'analyses bioinformatiques basé sur des approches innovantes développées dans le cadre des travaux de thèse du Dr Guillaume Charbonnier au TAGC. Cet outil de traitement, nommé *MetaWorkflow*, s'appuie sur une librairie de règles *Snakemake*, *mw-lib*, proposant plusieurs centaines de tâches atomiques permettant la réalisation de traitements unitaires variés. Ce workflow propose une interface simple sous la forme d'une table de métadonnées permettant au biologiste d'appliquer des traitements de routine, tout en améliorant la traçabilité des analyses.

Il est déployé sur un cluster de calcul ayant accès aux données brutes générées par la plateforme et correspond à un *workflow Snakemake* s'exécutant automatiquement à chaque modification de la table de métadonnées. Les données traitées correspondant aux échantillons nouvellement déclarés sont alors produites puis organisées, pour faciliter leur accès, en différentes arborescences parallèles sur la base des métadonnées renseignées (par projet, par client, par run...).

Il intègre de nombreux outils bioinformatiques permettant de réaliser la plupart des analyses primaires et secondaires. Actuellement, un travail est en cours afin de permettre l'ajout de traitement de routine pour les technologies de la cellule unique.

- **Schéma explicatif**



1. Le client fournit les échantillons.
2. L'ingénieur TGML réalise les contrôles qualité, prépare les librairies, complète une table de métadonnées et lance le séquençage.
3. Le workflow exécuté sur le cluster déduit des métadonnées la localisation des données brutes et les traitements à appliquer. Des données de sources externes (telles que GEO, SRA, Encode, Ensembl...) peuvent être intégrées aux analyses.
4. Les différents fichiers produits par le workflow sont synchronisés sur un serveur sftp.
5. Le client dispose d'un mois pour récupérer ses données. L'automatisation du workflow permet toutefois de regénérer les résultats ultérieurement si nécessaire.

Analyse de données de séquençage en routine avec MetaWorkflow. Les processus manuels sont indiqués par des flèches en pointillés. Les processus automatisés sont indiqués par des flèches pleines.

Développements à venir

En plus des développements bioinformatiques en cours, nous voulons mettre en place le séquençage sur noyau ainsi que la transcriptomique spatiale afin d'augmenter l'offre de prestation.
Si vous êtes intéressés par des développements spécifiques, n'hésitez pas à nous *contacter* !

Contact

Adresse: Parc Scientifique de Luminy, Bâtiment TPR2, case 928
163 Avenue de Luminy
13288 MARSEILLE cedex 09
Téléphone : 04.91.82.87.13 / 06.17.25.70.34
Messagerie : tgml.tagc@inserm.fr

Responsables plateforme: *Béatrice Loriod, Denis Puthier*
Représentant de la direction sur la plateforme: *Christophe Chevillard*
Gestionnaire : *Myriam Ramadour*
Personnels: *Hortense Vachon, Lisa Bargier, Alexandre Sarrabay, Sébastien Nin, Charlyne Gard (Stagiaire)*

Collaborateur : *Guillaume Charbonnier*
Chercheurs associés du TAGC : *Pascal Rihet (Directeur d'unité), Salvatore Spicuglia, Nathalie Lalevée*

Retrouvez-nous sur le web !
https://tagc.univ-amu.fr/en/resources/transcriptomic-genomic-platform-marseille-luminy
@TGML_TAGC