

Title:

Analysis of transcriptional networks of pediatric leukemia

Lab:

Gustave Roussy Institute - INSERM U1170
Team 3: Genetics and modeling of childhood leukemia
39 rue Camille Desmoulins, 94800 Villejuif
Email: thomas.mercher [at] inserm.fr
Tel: + 33-1-42-11-44-83

Context:

Cancer development results from accumulation of genetic alterations whose consequences generally converge towards the alteration of the transcriptional regulation and the gene expression. The identification of transcriptional regulators essential for the maintenance of cancer cell proliferation and survival may allow the development of new therapeutic approaches. In leukemia, chromosomal alterations frequently lead to the creation of fusion genes, associated with the expression of chimeric fusion oncoproteins involving transcription factors. Therefore, it is important to understand the consequences of their expression on the organization of chromatin and on transcriptional regulation. The ETO2-GLIS2 fusion, identified in the laboratory, is associated with acute megakaryoblastic leukemia in children with poor prognosis. We generated several sets of data [ChIPseq, single cell RNAseq (scRNAseq), ATACseq from patient cells and an experimental mouse model for this fusion]^{1,2}. On the other hand, we started to model *in silico* transcriptional networks active in normal and tumor cells (ARACNe, VIPER tools)^{3,4}.

Objectives:

In order to identify the transcriptional regulators whose activity is modified by the expression of the ETO2-GLIS2 fusion, we will follow an integrated bioinformatics / experimental approach in order to:

- 1) Rebuild transcriptional networks of early hematopoietic progenitors and identify transcription factors presenting a change in activity with fusion.
- 2) Integrate the expression results (scRNAseq) and the chromatin state (ATAC-seq and chromatin marks).

Prospects are to correlate these data with experimental data on methylation of DNA and chromatin organization (3C / Hi-Seq) in childhood leukemia.

The student will be in charge of the bioinformatics part including analyzes of data generated in the laboratory as well as analysis of available data (scRNAseq, ChIPseq, ATACseq). The student will benefit from a co-supervision in biology and bioinformatics as well as proximity to bioinformaticians and the bioinformatics platform of Gustave Roussy.

An extension of the internship by a thesis or a recruitment is possible.

Technical skills:

Linux, R and a programming language like perl or python; concepts of statistics and analysis of genomic data.

Bibliographical references:

1. Thiollier, C. et al. Characterization of novel genomic alterations and therapeutic approaches using acute megakaryoblastic leukemia xenograft models. *J. Exp. Med.* 209, 2017-2031 (2012).
2. Thirant, C. et al. ETO2-GLIS2 Hijacks Transcriptional Complexes to Cellular Identity Drive and Self-Renewal in Pediatric Acute Megakaryoblastic Leukemia. *Cancer Cell* 31, 452-465 (2017).
3. Alvarez, M.J. et al. Functional characterization of somatic mutations in cancer using a network-based inference of protein activity. *Nat. Biomed.* 48, 838-847 (2016).
4. Lachmann, A., Giorgi, F.M., Lopez, G. & Califano, A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinforma. Oxf. Engl.* 32, 2233-2235 (2016).

Keywords:

Single cell transcriptome, Genomics, Epigenetics, Transcription Factors Network, Leukemia, Pediatrics

Titre du stage :

Analyse des réseaux transcriptionnels des leucémies pédiatriques

Laboratoire :

Institut Gustave Roussy - INSERM U1170
Equipe 3: Génétique et modélisation des leucémies de l'enfant
39 rue Camille Desmoulins, 94800 Villejuif
Email: thomas.mercher [at] inserm.fr
Tel: +33-1-42-11-44-83

Contexte :

Les transformations cancéreuses résultent d'altérations génétiques dont les conséquences convergent généralement vers l'altération de la régulation transcriptionnelle et de l'expression génique. L'identification des régulateurs transcriptionnels essentiels au maintien de la prolifération et survie des cellules cancéreuses peut permettre le développement de nouvelles approches thérapeutiques ciblées. Dans les leucémies, les altérations chromosomiques entraînent fréquemment la création de gènes de fusion, conduisant à l'expression de protéines chimériques entre deux facteurs de transcription. Dans ces situations, il est important de comprendre la conséquence de leur expression sur l'organisation de la chromatine et la transcription. La fusion ETO2-GLIS2, identifiée au laboratoire, est associée aux leucémies aiguës mégacaryoblastiques de l'enfant de mauvais pronostic. Nous avons généré plusieurs sets de data [ChIPseq, single cell RNAseq (scRNAseq), ATACseq à partir de cellules de patients et d'un modèle expérimental murin]^{1,2}. D'autre part, nous avons commencé à modéliser les réseaux transcriptionnels actifs dans les cellules normales et les cellules tumorales (ARACNe, VIPER tools)^{3,4}.

Objectifs :

Dans le but d'identifier les régulateurs transcriptionnels dont l'activité est modifiée par l'expression de la fusion ETO2-GLIS2, nous suivrons une approche intégrée bioinformatique/expérimentale afin de :

- 1) Reconstruire les réseaux transcriptionnels des progéniteurs hématopoïétiques précoces et identifier les facteurs de transcription montrant un changement d'activité avec la fusion.
- 2) Intégrer les résultats d'expression (scRNAseq) et de l'état chromatinien (ATAC-seq et marques de chromatine).

Les perspectives sont de corrélérer ces données à des données expérimentales de méthylation et d'organisation de la chromatine (3C/Hi-Seq) dans les leucémies de l'enfant.

L'étudiant sera en charge de la partie bioinformatique incluant des analyses de données générées dans le laboratoire ainsi que de l'analyse de données disponibles (scRNAseq, ChIPseq, ATACseq). L'étudiant bénéficiera d'un co-encadrement en biologie et en bioinformatique ainsi que de la proximité avec les bioinformaticiens et la plateforme de bioinformatique de Gustave Roussy.

Une prolongation du stage par une thèse ou un CDD est possible.

Compétences techniques recherchées :

Linux, R et un langage de programmation type perl ou python ; notions de statistiques et d'analyse de données génomiques.

Références bibliographiques :

1. Thiollier, C. *et al.* Characterization of novel genomic alterations and therapeutic approaches using acute megakaryoblastic leukemia xenograft models. *J. Exp. Med.* **209**, 2017–2031 (2012).
2. Thirant, C. *et al.* ETO2-GLIS2 Hijacks Transcriptional Complexes to Drive Cellular Identity and Self-Renewal in Pediatric Acute Megakaryoblastic Leukemia. *Cancer Cell* **31**, 452–465 (2017).
3. Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).
4. Lachmann, A., Giorgi, F. M., Lopez, G. & Califano, A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinforma. Oxf. Engl.* **32**, 2233–2235 (2016).

Mots clés: Single cell transcriptome, Génomique, Épигénétique, Réseau de facteurs de transcription, Leucémie, Pédiatrie