Devising ATAC-seq methods to support Mature T-Cell Neoplasms classification in clinical settings

Sylvain Mareschal[1], Edith Julia[1,2], Amel Chebel[2], Camille Golfier[1], E. Missiaglia[3], V. Rattina[3], Q. Testard[1], C. Lours[1], S. Hadj-Hamou[2], N. Bissay[2], A. Verney[2], M. Tschantz[2], D. Chartoire[2], S. Sebai[2], D. Sibon[4], A. Marçais[4], S. Carras[5], V. Asnafi[4], L. Payen[1], D. Sanlaville[1], C. Bardel[1], M. Roussel[6], F. Drieux[7], P. Ruminy[7], V. Szablewski[8], L. Martin[9], C. Chassagne-Clément[10], G. Salles[11], L. Baseggio[1], P. Sujobert[1], M. Cheminant[4], C. Laurent[12], A. Traverse-Glehen[1,2], L. De Leval[3], P. Gaulard[13], L. Genestier[2], Emmanuel Bachy[1,2].

1 : Hospices Civils de Lyon (FR), 2 : Lymphoma immunobiology team, CIRI, Lyon (FR), 3 : Lausanne University Hospital (CH), 4 : Hopital Necker-Enfants Malades, AP-HP, Paris (FR), 5 : CHU de Grenoble (FR), 6 : CHU de Rennes (FR),7 : Centre Henri Becquerel, Rouen (FR), 8 : Hopital Saint Eloi, Montpellier (FR), 9 : INSERM UMR 1098, Dijon (FR), 10 : Centre Léon Bérard, Lyon (FR), 11 : Memorial Sloan Kettering Cancer Center, New-York (US), 12 : CHU de Toulouse (FR), 13 : Hopital Henri Mondor, Créteil (FR)

Assay for transposase-accessible chromatin sequencing (ATAC-seq) is a simple technique to profile open chromatin regions, granting the opportunity to classify tumors on their inherited epigenetic landscape rather than a more variable final tumoral state. Following on convincing results obtained by others clustering solid tumors, we applied ATAC-seq to Mature T-cell neoplasms (MTCN), a challenging disease in which **up to 30%** of minor or major **diagnostic reclassifications** have been evidenced following expert review of cases. **Over a thousand hematological samples** of multiple natures and origins were sequenced and a dedicated bioinformatics pipeline was developed, gathering new insights on MTCN pathology and paving the road to fast and accurate classification of these tumors in clinical settings.

Gathering more than epigenetics data from ATAC-seq

Capitalizing on methods published by ATAC-seq inventors and the ENCODE project, we developed a **Singularity-backed Nextflow pipeline** (Figure 1A) which proved **6.5 times faster than ENCODE's** in our setting (from 25.1 h / 717.5 CPUh to 3.9 h / 99.5 CPUh on a 16-sample run of 442e6 read pairs), while providing similar results and additional functionalities.

Extra quality checks include a generic framework to **quantify DNA contamination from other species**, as poor quality samples tended to display bovine contamination from cell culture medium. This framework could be extended to quantify **latent viral and retroviral integration in genomes** (Figures 1B and 1C), which can contribute to classification (7/7 acute T-cell leukemia / lymphomas, a subtype associated with HTLV-1 infection, were positive).

ATAC-seq also provides the opportunity to identify **large copy-number alterations**, using GATK ModelSegments on out-of-peak reads. While sample quality was not always sufficient to obtain clean profiles (Figure 1D), this proved useful to identify iso(7q) (Figure 1E), whose association with hepatosplenic T-cell lymphoma is well-documented.

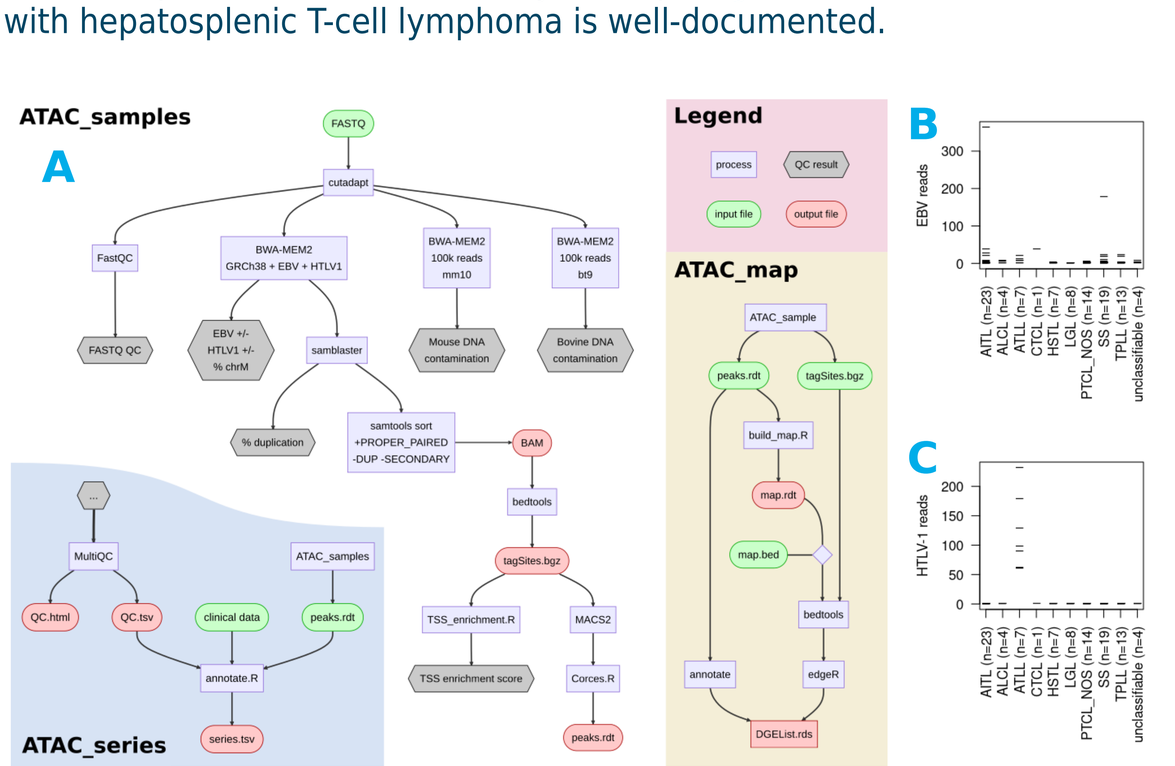


Figure 1 : **A** : Schematic representation of the 3 Nextflow pipelines (inputs in green, outputs in red, processes in purple and QCs in gray). **B** and **C** : ATAC-seq reads mapped to HTLV-1 (Human T-Lymphotropic Virus 1) and EBV (Epstein–Barr Virus) in the 100 FACS-sorted MTCN samples. **D** : DNA copy-number profile derived from ATAC-seq. **E** : Iso-chromosome 7q identified by ATAC-seq.

Semi-supervised clustering and differential analysis

A cohort of **100 FACS-sorted tumor cell samples** from primary MTCN samples was used to better characterize the previously-described MTCN subtypes (FAST-ATAC protocol).

Projection of the 137 469 ATAC peaks was performed using **t-SNE** (Figure 2A), mitigating the arbitrariness of its numerous parameters by using **4 000 sets of parameters** and selecting the more consensual projection (least squares method). The **stability** of each sample could also be assessed, as the proportion of the 4 000 t-SNEs in which it had the same closest neighbors. Clustering was finally performed using an **iterative nearest-neighbor joining strategy**, a method particularly tolerant to clusters of heterogeneous sizes (Figure 2A).

Differential opening analysis between the resulting clusters could pinpoint several markers of interest, such as *EFNB2* in angioimmunoblastic T-cell lymphoma (AITL). The significant peak was actually located at the promoter of an antisens lncRNA (Figure 2B), whose expression was indeed impacted in additional RNA-seq data (Figure 2C).

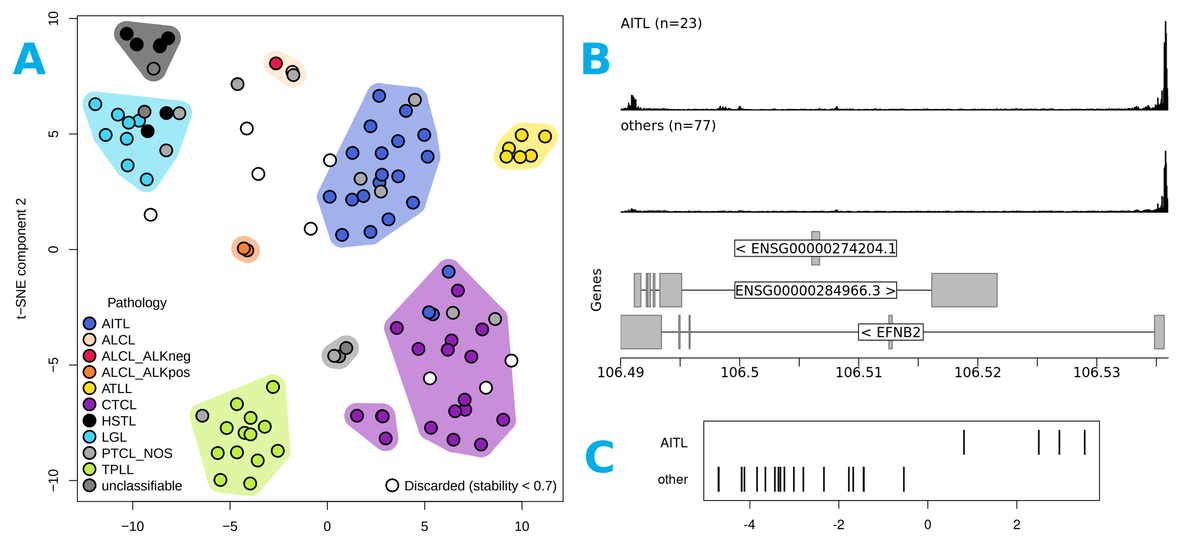


Figure 2 : **A** : t-SNE projection and clustering of 100 FACS-sorted MTCN samples. **B** : Cumulated ATAC-seq profiles of AITL and non-AITL samples along *EFNB2* gene locus. **C** : RNA-seq expression levels of ENSG00000284966.3, a lncRNA antisens to *EFNB2* in a subset of ATAC samples.

Joint clustering with normals reveals cells-of-origin

Joint UMAP projection and clustering with sorted normal cell samples could also fuel the current discussion over possible cells-of-origin for the various MTCN subtypes, notably naive T-cells for prolymphocytic leukemia (TPLL), an entity previously believed to be derived from memory T-cells (Figure 3).

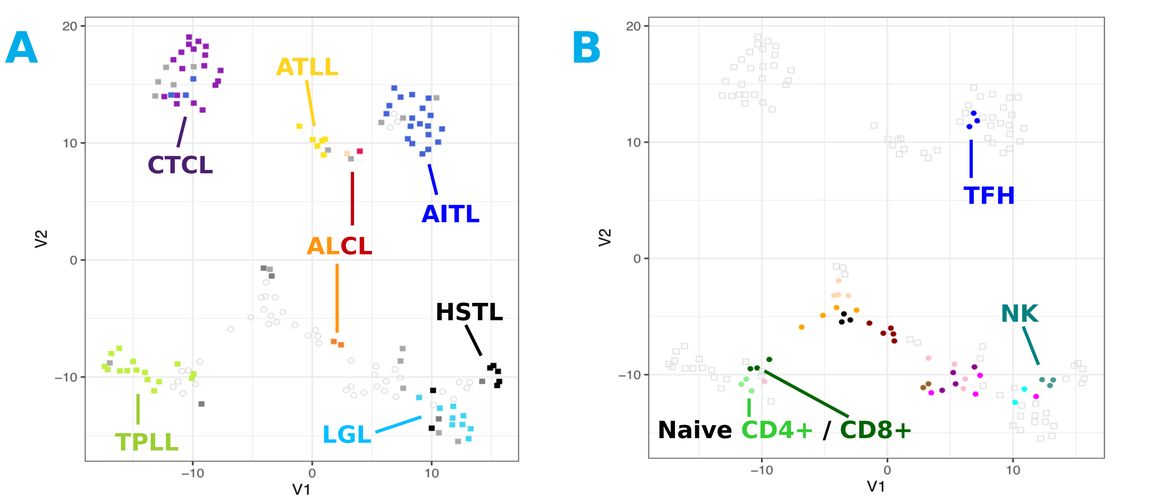
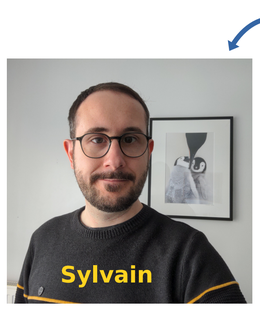


Figure 3 : UMAP projection of 100 FACS-sorted tumor cell samples (**A**) and 73 FACS-sorted normal hematological cell samples (**B**). Samples were projected together and represented separately.

ATAC-seq provided robust classification and novel insights in MTCN, and tools are fast and accurate enough to perform in a clinical setting while benefiting from ATAC-seq simplicity and cost-efficiency. Latest protocols suitable for paraffin-embedded samples, currently under investigation, could make ATAC-seq a key player in MTCN, and other cancers, diagnosis.



Meet the author here at ISMB / ECCB 2023

Download as PDF and stay posted about git code release and publication