

## Infinity Flow: phenotypage quasi-exhaustif de tissus complexes par cytométrie et machine learning

Etienne Becht<sup>1,2</sup> (ebecht@fredhutch.org), Evan Newell<sup>2</sup>, Raphael Gottardo<sup>2</sup>, Mark Headley<sup>2</sup>

<sup>1</sup> Carte d'Identité des Tumeurs, Ligue Nationale Contre le Cancer, Paris, France

<sup>2</sup> Fred Hutchinson Cancer Research Center, Seattle, États-Unis

**Introduction:** Des outils capables d'identifier les interactions cellulaires complexes sont nécessaires à l'étude de la physiologie et des pathologies. Les cytomètres modernes peuvent quantifier ~20 protéines, permettant de révéler une grande variété de phénotypes cellulaires, mais qu'il est difficile d'annoter. Les panels d'anticorps ayant une taille limitée, il est de plus nécessaire d'arbitrer entre marqueurs phénotypiques, fonctionnels ou exploratoires. Des alternatives techniques existent, comme la cytométrie de masse ou les anticorps conjugués à des oligonucléotides, mais ces technologies sont limitées en nombre de cellules, plus chères et moins accessibles. Nous avons développé une méthode combinant la cytométrie en flux et l'apprentissage machine (machine learning) permettant de dépasser ces limitations.

**Méthodes:** Expérimentalement, les cellules sont marquées par un panel de 10-20 anticorps dit "backbone". L'échantillon est ensuite aliquoté dans des centaines de puits distincts, chacun contenant un anticorps "exploratoire" de spécificité unique. L'acquisition se fait par cytométrie en flux.

Informatiquement, un modèle de régression non-linéaire et multivarié (par exemple réseaux de neurones) est entraîné pour chaque puit. Ces modèles visent à prédire le niveau d'expression des anticorps "exploratoires" à partir des anticorps "backbone", au niveau de chaque cellule unique. Leur application permet d'obtenir une matrice d'expression protéique pour des millions de cellules et des centaines de protéines.

**Résultats:** Nous avons appliqué cette méthode aux poumons dissociés de souris C57/B6, et obtenu une matrice de données contenant 2,660,000 événements et 266 protéines. Les patterns imputés d'expression et de co-expression sont fiables, permettant une annotation quasiment exhaustive de chaque population cellulaire (32 sur 33 des clusters, soit 96,7% des cellules). Cette approche nous permet d'identifier les sous-populations lymphoïdes, myéloïdes, épithéliales, endothéliales, mésenchymales, en une seule expérience et par cytométrie conventionnelle. Nous avons obtenu une meilleure granularité par cette méthode comparée à une analyse computationnelle "naïve" limitée au backbone.

**Conclusion:** En permettant l'analyse de réseaux cellulaires complexes au niveau protéique, cet outil est bien adapté à la caractérisation de systèmes immunologiques complexes. Etant donnée la grande quantité de données qu'il génère et son accessibilité, nous pensons qu'il sera largement adopté.

**Mots clés :** Machine learning - High-dimensional analysis - Parallel cytometry - Single cell - Protein expression.

Références :

1. Dutertre et al, 2019, Immunity (<https://doi.org/10.1016/j.immuni.2019.08.008>)
2. Kwok et al, 2020, Immunity (<https://doi.org/10.1016/j.immuni.2020.06.005>)
3. Becht et al, 2020, preprint (<https://www.biorxiv.org/content/10.1101/2020.06.17.152926v2>)