



Thomas Walter  
Chef d'équipe  
thomas.walter@curie.fr

De nombreuses technologies nouvelles permettent d'observer les organismes vivant à une échelle jusque-là inexplorée. Par exemple, les techniques de biopuces ou de séquençage massivement parallèle permettent de lire l'information génétique complète d'un échantillon biologique, la spectrométrie de masse permet de quantifier le protéome, la vidéo-microscopie à haute résolution permet de suivre les évolutions de populations de cellules, et les méthodes de criblage à haut débit sont utilisées pour caractériser les effets biologiques de grandes quantités de molécules.

**Toutes ces technologies génèrent de très grandes quantités de données, qu'il est souvent difficile d'interpréter directement. Afin de mieux exploiter ces masses de données et d'en extraire des informations biologiques et médicales utiles, notre équipe développe des méthodes mathématiques et des algorithmes innovants, basés sur notre expertise en modélisation mathématique, statistique, machine learning, bioimage informatics et biologie structurale.**

Nous développons de nouveaux outils pour examiner différentes questions biologiques et médicales, notamment:

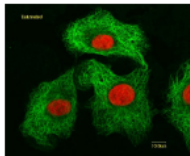
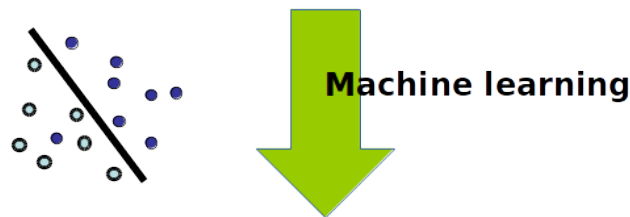
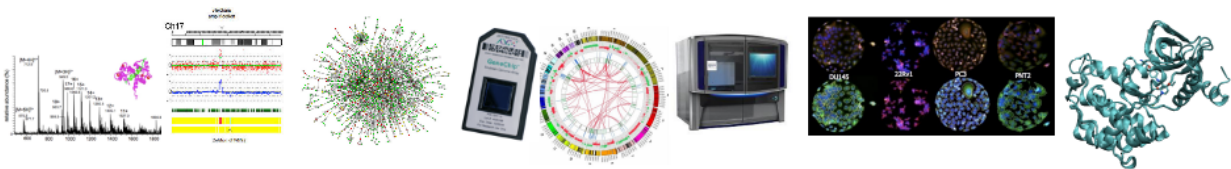
- Biologie et biologie des systèmes *in silico*. Nous développons des approches innovantes pour la reconstruction statistique de réseaux biologiques, l'analyse et la modélisation de la progression tumorale au niveau génomique, transcriptomique et épigénétique, l'annotation automatique de gènes et d'éléments fonctionnels par intégration de données hétérogènes et complexes, en particulier les données provenant de séquençage à haut débit ou de vidéo-microscopie.
- Médecine prédictive et de précision : Nous développons des méthodes pour la classification de tumeurs et l'identification de biomarqueurs diagnostiques, pronostiques et prédictifs de

# Apprentissage statistique et modélisation des systèmes biologiques

## U900 - Cancer et génome : bioinformatique, biostatistiques et épidémiologie

réponse à un traitement. Ces classifications s'appuient sur un grand nombre de paramètres mesurables comme les données cliniques du patient, l'ensemble des mutations somatiques de la tumeur, l'expression de gènes et de transcrits alternatifs, ou les modifications structurales de l'ADN, et fait appel à des méthodes d'apprentissage statistique en grande dimension.

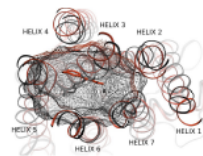
- Développement de médicaments : nous développons de nouvelles méthodes pour le criblage virtuel et la chémogénomique *in silico*, afin d'aider à identifier ou créer de petites molécules capables de se lier spécifiquement à des cibles thérapeutiques par modélisation informatique. Ces méthodes combinent des modélisations 3D des molécules et des procédures statistiques innovantes, et peuvent par exemple aider à prédire les profils d'activité ou de toxicité de candidats médicaments.



**Bioimage informatics**  
*Image segmentation*  
*Classification*  
*High-Content Screening*



**Systems Biology**  
*Network inference*  
*Gene prioritization*  
*Data integration*



**Drug Design**  
*Virtual screening*  
*Chemo-genomics*  
*Side effect prediction*



**Precision medicine**  
*Patient stratification*  
*Prognostic & Predictive*  
*NGS data processing*

Le machine learning permet d'utiliser de nombreux types de données pour des applications en imagerie biomédicale, biologie des systèmes, conception de médicaments, et médecine de précision.

## Publications clés

Année de publication : 2019

---

Collier Olivier, Stoven Véronique, Vert Jean-Philippe (2019 Sep 25)

**A Single- and Multitask Machine Learning Algorithm for the Prediction of Cancer Driver Genes**

*Plos Computational Biology*

Slim L., Chatelain C., Azencott C.A., Vert J.P. (2019 Jun 1)

**kernelPSI: a Post-Selection Inference Framework for Nonlinear Variable Selection**

*International Conference on Machine Learning* International Conference on Machine Learning : 5857-5865

Peter Naylor, Marick Lae, Fabien Reyat, Thomas Walter (2019 Feb 5)

**Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map.**

*IEEE Transactions on Medical Imaging* : 448-459 : DOI : [10.1109/TMI.2018.2865709](https://doi.org/10.1109/TMI.2018.2865709)

Année de publication : 2018

---

Aubin Samacoits, Racha Chouaib, Adham Safieddine, Abdel-Meneem Traboulsi, Wei Ouyang, Christophe Zimmer, Marion Peter, Edouard Bertrand, Thomas Walter, Florian Mueller (2018 Nov 4)

**A computational framework to study sub-cellular RNA localization.**

*Nature Communications* : 4584 : DOI : [10.1038/s41467-018-06868-w](https://doi.org/10.1038/s41467-018-06868-w)

Tom Baladi, Jessy Aziz, Florent Dufour, Valentina Abet, Véronique Stoven, François Radvanyi, Florent Poyer, Ting-Di Wu, Jean-Luc Guerquin-Kern, Isabelle Bernard-Pierrot, Sergio Marco Garrido, Sandrine Piguel (2018 Nov 1)

**Design, synthesis, biological evaluation and cellular imaging of imidazo[4,5-b]pyridine derivatives as potent and selective TAM inhibitors.**

*Bioorganic & medicinal chemistry* : 26 : 5510-5530 : DOI : [10.1016/j.bmc.2018.09.031](https://doi.org/10.1016/j.bmc.2018.09.031)

Benoit Playe, Chloé-Agathe Azencott, Véronique Stoven (2018 Oct 5)

**Efficient multi-task chemogenomics for drug specificity prediction.**

*PloS one* : e0204999 : DOI : [10.1371/journal.pone.0204999](https://doi.org/10.1371/journal.pone.0204999)