

Centre de bio-informatique

(MINES ParisTech – CBIO)

Directeur : Jean-Philippe VERT

Responsable communication :
Isabelle Schmitt

Téléphone 01 64 69 47 81

Courriel cbio@cbio.mines-paristech.fr

Web et publications

<http://www.mines-paristech.fr/Fr/CBIO>

Enseignants chercheurs 4

Autres personnels 3

Doctorants MINES ParisTech 4

Doctorant autre établissement 1



Formation

Le CBIO intervient dans la formation des ingénieurs civils des Mines en assurant les cours d'option, les enseignements spécialisés en biotechnologie et en participant au MIG *L'ingénieur et la santé*. Il contribue également aux enseignements de 2^e année du master *Mathématiques, vision et apprentissage* de l'École normale supérieure de Cachan, en assurant le cours *Apprentissage statistique par méthodes à noyaux*. Les personnels du CBIO interviennent par ailleurs ponctuellement pour enseigner dans des universités et instituts de recherche divers, comme par exemple à l'Institut génomique de Malaisie, en novembre 2010.



Recherche

L'activité principale de recherche au CBIO consiste à développer des méthodes mathématiques et informatiques innovantes pour l'analyse et la modélisation de données biologiques et chimiques. En 2008 nous avons noué un partenariat privilégié avec l'Institut Curie et l'unité mixte INSERM U900 *Cancer et Génome : bioinformatique, biostatistiques et épidémiologie d'un système complexe*. Le CBIO constitue l'équipe *Apprentissage statistique et modélisation des systèmes biologiques* de cette unité, dédiée à la recherche en biostatistique et bioinformatique pour la recherche contre le cancer.

Vers une médecine prédictive et personnalisée

L'apparition de nouvelles technologies permettant de caractériser les tumeurs au niveau moléculaire, telles les puces à ADN pour mesurer l'expression des gènes, ou l'hybridation génomique comparative (CGH) pour détecter les aberrations génomiques, ouvre la voie à nouvelles méthodes pour le diagnostic, le pronostic sur l'évolution probable de la maladie, et la prédiction d'efficacité des traitements. Ces nouveaux outils impactent la prise en charge des malades, par une meilleure prise en compte des spécificités de chaque cancer et l'arrivée d'une médecine de plus en plus personnalisée et efficace. Dans cette optique nous avons proposé de nouvelles méthodes pour l'estimation de signatures moléculaires et avons mis en lumière les limites de méthodes couramment utilisées en termes de robustesse et de précision. Nous avons développé une nouvelle méthode pour la détection rapide d'aberrations fréquentes dans des ensembles de profils de nombre de copie d'ADN, qui peut notamment être utile pour la détection de facteurs diagnostiques et pronostiques dans la structure de l'ADN.

Enfin, nous avons développé une nouvelle méthode basée sur le formalisme des réseaux Bayésiens pour identifier les mutations génomiques à partir de données de séquençage à haut débit. Ces développements méthodologiques s'accompagnent de nombreuses collaborations avec des médecins et chercheurs de l'Institut Curie et d'autres instituts, notamment sur le cancer du sein, de la vessie, de la prostate, et sur le neuroblastome. Nous avons également entamé un projet européen (NADINE) avec 16 partenaires académiques et privés, dans le but de développer de nouveaux biomarqueurs pour le diagnostic précoce de maladies neuro-dégénératives, notamment la maladie d'Alzheimer.

Inférence de réseaux biologiques

De nombreuses voies de signalisation, régulation, et métaboliques, impliquant des interactions entre de nombreux gènes, jouent un rôle critique dans l'initiation et le développement des tumeurs. Notre connaissance de ces systèmes reste cependant très parcellaire. Il semble possible de combler ces lacunes

en exploitant les grandes quantités de données générées par les différentes technologies en génomique et protéomique. Afin de reconstruire *in silico* les informations manquantes sur ces réseaux, nous avons continué à développer un cadre général d'inférence de graphe à partir de données génomiques hétérogènes, en nous appuyant sur des nouveaux développements en apprentissage statistique. Nous avons poursuivi notre collaboration avec William Noble (University of Washington), sur la prédiction à grande échelle du réseau d'interaction protéine-protéine à partir de structures 3D des protéines, et avec plusieurs institutions japonaises (universités de Kyoto et Tokyo, Institut technologique de Tokyo, CBRC) sur la reconstruction de réseaux métaboliques. Nous avons établi un nouveau lien méthodologique entre le problème de reconstruction de réseaux biologiques et le problème d'apprentissage statistique multi-instance, qui ouvre la voie à de nouvelles stratégies d'inférence. Nous avons développé une nouvelle méthode pour la priorisation de gènes d'intérêt par intégration de données hétérogènes, qui surpasse les approches existantes en performance, ainsi qu'une nouvelle méthode pour la reconstruction de réseaux de régulation à partir de données d'expression, basée sur des outils d'apprentissage statistique pour la sélection de variables.

Bio-informatique structurale, interactions moléculaires et criblage virtuel

Nous avons combiné nos expertises en biologie structurale et en apprentissage statistique pour prédire les interactions entre protéines et petites molécules, à l'aide de modèles statistiques utilisant des informations sur les structures 3D des molécules. La prédiction de ces interactions est utile, à la fois pour comprendre la fonction de protéines ayant un intérêt thérapeutique, et pour identifier des molécules susceptibles d'inhiber des cibles thérapeutiques connues. Nous avons ainsi poursuivi nos recherches en chémogénomique *in silico*, visant à prédire systématiquement les interactions entre une banque de petites molécules et une famille de protéines, dans le cadre d'une collaboration avec l'Université de Kyoto. Nous avons notamment développé une nouvelle mesure de similarité entre poches de fixation de ligands sur des structures 3D de protéines. Ces développements méthodologiques trouvent leurs applications pour la recherche de nouveaux traitements contre divers

cancers, et pour la prédiction des effets secondaires de molécules candidates. Nous travaillons notamment sur la recherche d'inhibiteurs de la protéine phosphatase calcineurine et sur la tyrosine kinase TYRO3, en collaboration avec des chercheurs de l'Institut Curie.

Criblage multi-cellulaire à haut débit

L'utilisation de cribles à haut débit, basés sur le phénotypage cellulaire par microscopie, est une approche expérimentale pour tester l'activité et les effets phénotypiques de petites molécules qui va devenir une force d'innovation majeure pour la biologie des décennies à venir. Cette technologie est de plus en plus utilisée en recherche bio-médicale, et a le potentiel de devenir un outil versatile pouvant être utile dans de nombreux domaines en biologie cellulaire, biologie du développement, et génétique. La modélisation et l'analyse des données produites par cette technologie, qui caractérise le phénotype de chaque cellule au sein de chaque population à l'aide d'un grand nombre de paramètres, reste cependant un défi à relever. Nous poursuivons le développement du logiciel CellRA pour l'analyse statistique de ces données, en particulier pour l'identification de relations cause-effet entre un traitement appliqué aux cellules et le phénotype de la population. Nous avons poursuivi notre collaboration avec le CEA et l'entreprise IMSTAR pour identifier des kinases et des micro-ARN susceptibles de devenir des cibles thérapeutiques contre le cancer de la prostate, et avec les entreprises Pierre Fabre, ADCIS et le CNRS pour la recherche de nouvelles molécules anti-mitotiques. Nous avons également initié des collaborations avec des chercheurs de l'Institut Curie sur la biophysique de la migration cellulaire, ainsi que plusieurs projets visant à cribler des banques moléculaires sur différents phénotypes d'intérêt.



Faits marquants

Nous avons participé à la compétition internationale DREAM5, sur la reconstruction *in silico* de réseaux de régulations, et avons obtenu une mention honorable décernée lors d'une conférence à l'Université de Columbia à New-York, en septembre 2010. Deux étudiants du CBIO ont soutenu leur thèse en 2010 : Mikhail Zaslavskiy, sur l'alignement de graphes et ses applications en vision et bio-informatique, et Fantine Mordelet, sur les méthodes d'apprentissage à partir d'exemples positifs et sans label et leurs applications en bio-informatique.