

Développement d'une méthodologie d'analyse intégrant données de puce et réseaux génétiques ; application à la pathologie cancéreuse

<http://www.bioinfo.curie.fr/projets/kernelchip>

CONTEXTE

Le problème de l'identification des voies de régulation impliquées dans la pathologie cancéreuse est central dans l'étude du cancer, non seulement pour en comprendre les mécanismes et proposer des cibles thérapeutiques, mais aussi pour affiner la sous-classification des différents types de cancers et orienter la prise en charge thérapeutique du patient. L'utilisation de données de puces est une approche courante, mais difficile d'interprétation étant donné le grand nombre de gènes à considérer et la difficulté à construire des hypothèses biologiques. Une partie de cette difficulté peut être levée si on met les données d'expression en corrélation avec d'autres données (réseaux génétiques, données cliniques).

OBJECTIFS

Nous avons pour but de proposer une méthodologie pour identifier ces voies de régulation et analyser les données de puces en intégrant l'information a priori de structure des réseaux de régulation de l'expression (par exemple en classification supervisée ou non).

Nous nous basons sur un formalisme rigoureux déjà proposé. Ce formalisme est basé sur la théorie des noyaux définis positifs et sur l'apprentissage statistique dans des espaces de Hilbert à noyaux auto-reproduisants. L'idée principale consiste à représenter les informations sur les gènes par des noyaux (fonctions de similarités symétriques et définies positives) permettant de définir de manière implicite de nouveaux espaces dans lesquels les représentations sont simples à manipuler. On peut alors appliquer des algorithmes puissants pour la classification, la régression ou la recherche de similarités.

Nous étendons ce formalisme en nous basant sur la représentation des réseaux proposée dans SHARP. Nous avons choisi cette représentation car elle est issue des délibérations d'un consortium international et que la base qui a été établie dans SHARP est une base orientée vers le cancer. Cependant, nous comptons intégrer aussi les données de KEGG et d'autres bases de données

Il nous faut affiner le choix des noyaux en analysant les résultats issus de données de cancer dont les voies de régulation sont (en partie) connues. Dans un deuxième temps, nous étendrons le formalisme pour la détection de nouvelles voies métaboliques et de régulation.

Il est à noter que si nous nous consacrons principalement à l'étude de la pathologie cancéreuse, notre méthode peut être appliquée à tout organisme (pas forcément l'homme) et à toute pathologie ou processus faisant intervenir l'expression des gènes et les réseaux de régulation qui les gouvernent

PROJET

Proposer un formalisme rigoureux permettant de mettre en corrélation données de puces transcriptome et données de réseaux de régulation.
Affiner le formalisme à l'aide des résultats de tests sur des exemples précis.
Intégrer des covariables phénotypiques à l'analyse (données cliniques)
Etendre le formalisme pour détecter de nouvelles voies métaboliques et/ou de régulation.

RESULTATS

- Construction d'un réseau génétique intégrant les données de SHARP et KEGG ainsi que celles d'autres bases
- Elaboration de noyaux intégrant l'information des voies de régulation)
- Intégration de données cliniques

MOTS CLES

Réseaux métaboliques, modélisation, puces transcriptome, méthodes à noyaux, cancer

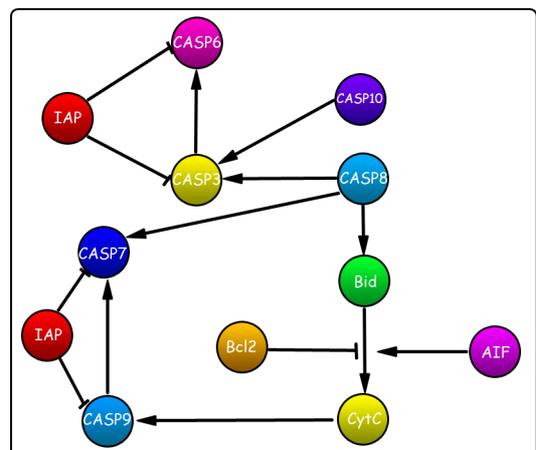
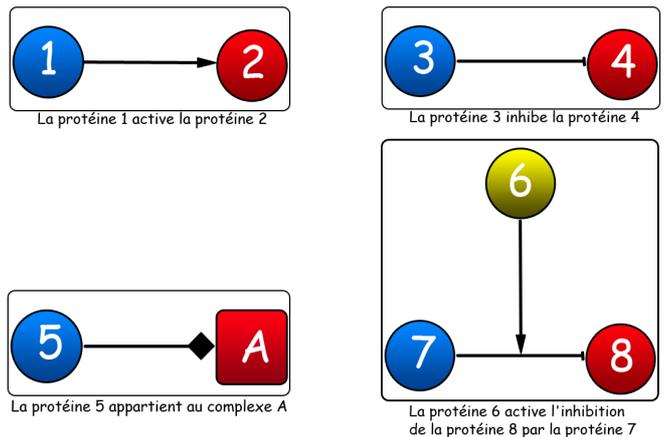
LABORATOIRES IMPLIQUES

Institut Curie / Service de Bioinformatique / Paris
ENSMP / Computational Biology Group / Fontainebleau

CALENDRIER

2004 Novembre (Paris) Début du projet
2005 Une réunion par mois (Paris ou Fontainebleau). Participants récurrents : Franck Rapaport, Emmanuel Barillot, Jean-Philippe Vert, Andrei Zynoviev (Institut Curie, Service de Bioinformatique)
Intervenants : François Radvanyi (institut Curie), Pierre Neuvial (institut Curie, Service de Bioinformatique)

REPRESENTATION DES RESEAUX



Exemple de représentation de réseau : extrait de voies liées à l'apoptose (homo sapiens)

La représentation de SHARP est très simple : même si elle n'est pas la plus complète, elle a l'avantage de présenter un formalisme défini de façon suffisamment rigoureuse pour pouvoir être analysé facilement par des outils automatiques. Dans SHARP, un réseau est représenté par ses protéines et les gènes associés, ses complexes et les relations qui les unissent. Ces relations, (dirigées) peuvent être de trois types : entre une protéine et un complexe (appartenance, inhibition ou activation), entre deux protéines (activation ou inhibition) ou entre une protéine et une relation (activation ou inhibition) déjà établie (cf. dessin ci-dessus)