SINBIO

(Projet de Recherche 2003)

ACI IMPBIO

Statistique et INformatique intégrées à la BIOlogie des traits complexes

http://ecgene.net/Sinbio/

CONTEXTE

Etablir le lien entre la variabilité naturelle du génome humain et les traits complexes constitue l'un des enjeux majeurs de la recherche post-génomique. A la différences des maladies monogéniques dont la caractérisation génétique a connu un succès considérable dans les dernières années, les avancées dans le domaine des maladies multifactorielles ont été beaucoup plus ténues et les stratégies pour identifier les gènes de susceptibilité à ces maladies doivent être reconsidérées. L'approche traditionnelle consistant à étudier les gènes un par un et à mettre en relation quelques polymorphismes avec la maladie a montré ses limites. Il apparaît de plus en plus évident que l'étude des maladies complexes requiert non seulement d'analyser la variabilité complète des gènes candidats, mais aussi d'aborder la variabilité génétique des systèmes biologiques dans leur ensemble.

L'exploration moléculaire de la variabilité génétique ne constitue plus une étape limitante (du moins d'un point de vue technologique) grâce au développement d'outils de séquençage et de génotypage à haut débit. Le challenge actuel réside dans le développement d'outils statistiques et informatiques capables de traiter la grande quantité d'information générée par la biologie moléculaire et de modéliser au mieux la relation génotype/phénotype.

OBJECTIFS

L'objectif est de développer de nouveaux outils statistiques et informatiques destinés à relier la variabilité du génome humain aux traits complexes, et d'appliquer ces outils à de grandes études épidémiologiques sur les pathologies cardiovasculaires et l'asthme. Un aspect important du projet concerne également la valorisation et la diffusion des outils développés. Ce projet repose sur une interaction étroite entre les biologistes moléculaires qui caractérisent la variabilité génétique des systèmes candidats, et les épidémiologistes/ statisticiens qui développent de nouveaux outils mathématiques pour exploiter les données moléculaires ainsi générées dans de grandes études épidémiologiques. Plusieurs logiciels d'analyse génétique, THESIAS pour les analyses haplotypiques, DICE pour la fouille de données, FINESSE pour les modèles de ségrégation/linkage, ont déjà été conçus par les équipes participant au projet.

PROJET

- Définition de systèmes biologiques candidats (molécules de l'inflammation, molécules d'adhésion, métabolisme des lipides…)
- Criblage moléculaire des gènes du système
- Génotypage des polymorphismes dans de grandes études épidémiologiques (Athero*Gene*, MORGAM, EGEA)
- Analyse de la relation génotype/phénotype par des logiciels spécifiquement développés pour l'analyse des traits complexes
- Extension des logiciels existants
- Création d'interfaces conviviales pour les logiciels et mise à disposition sur le site Internet GeneCanvas de l'U525 (**http://www.genecanvas.org/**)

RESULTATS

- THESIAS (Logiciel d'analyse haplotypique) : Extension à l'analyse de survie [1] et à l'analyse cas/témoin appariée [2]
- THESIAS : Création d'une interface JAVA et mise à disposition du logiciel sur le site web GeneCanvas
- FINESSE (Logiciel d'analyse de ségrégation/linkage basée sur les modèles régressifs) : Création d'une interface.
- DICE (Logiciel de fouille de données) : Application à l'analyse du gène de l'ApoB en relation avec les taux plasmatiques d'ApoB [3]
- Analyse génétique du système PAF-AH dans l'étude Athero*Gene* [4]
- Analyse haplotypique des polymorphismes du gène *ABCA1* dans l'étude ECTIM [5]
- Analyse génétique du système de l'IL-18 dans l'étude Athero*Gene* [6]
- Analyse haplotypique du gène de la caspase-1 dans l'étude Athero*Gene*

LABORATOIRES IMPLIQUES

INSERM U 525 "Génétique Epidémiologique et Moléculaire des Pathologies Cardiovasculaires" - Paris
- Equipe "Méthodes en Epidémiologie Génétique" (Laurence Tiret)
- Equipe "Polymorphismes des Gènes Candidats" (François Cambien)

INSERM EMI 0006 "Méthodologie Statistique et Epidémiologie Génétique des Maladies Multifactorielles - Evry (Florence Demenais)

MOTS CLES

Polymorphisme génétique, gène candidat, méthode statistique, logiciel

PUBLICATIONS PRINCIPALES

•1. Tregouet D, Tiret L. Cox proportional hazards survival regression in haplotype-based association analysis using the Stochastic-EM algorithm. *Eur J Hum Genet.* 2004

•2. Morange PE, Tregouet DA, Frère C, et al . TAFI gene haplotypes, TAFI plasma levels and future risk of coronary heart disease : The PRIME Study. *J Thromb Haemost* 2005.

•3. Tahri-Daizadeh N, Tregouet D, Nicaud V, Poirier O, Cambien F, Tiret L. Exploration of multilocus effects in a highly polymorphic gene, the apolipoprotein (APOB) gene, in relation to plasma apoB levels. *Ann Hum Genet.* 2004

•4. Ninio E, Tregouet DA, Carrier JL, et al. Platelet-activating factor-acetylhydrolase (PAF-AH) and PAF-receptor gene haplotypes in relation to future cardiovascular event in patients with coronary artery disease. *Hum Mol Genet.* 2004

•5. Tregouet D, Ricard S, Nicaud V, et al. In depth haplotype analysis of ABCA1 gene polymorphisms in relation to plasma ApoA1 levels and myocardial infarction. *Arterioscler Thromb Vasc Biol.* 2004

•6. Tiret L, Godefroy T, Lubos E, et al. Genetic analysis of the interleukin-18 system highlights the role of the interleukin-18 gene in cardiovascular disease. *Circulation* 2005.

THESIAS – INTERFACE JAVA