

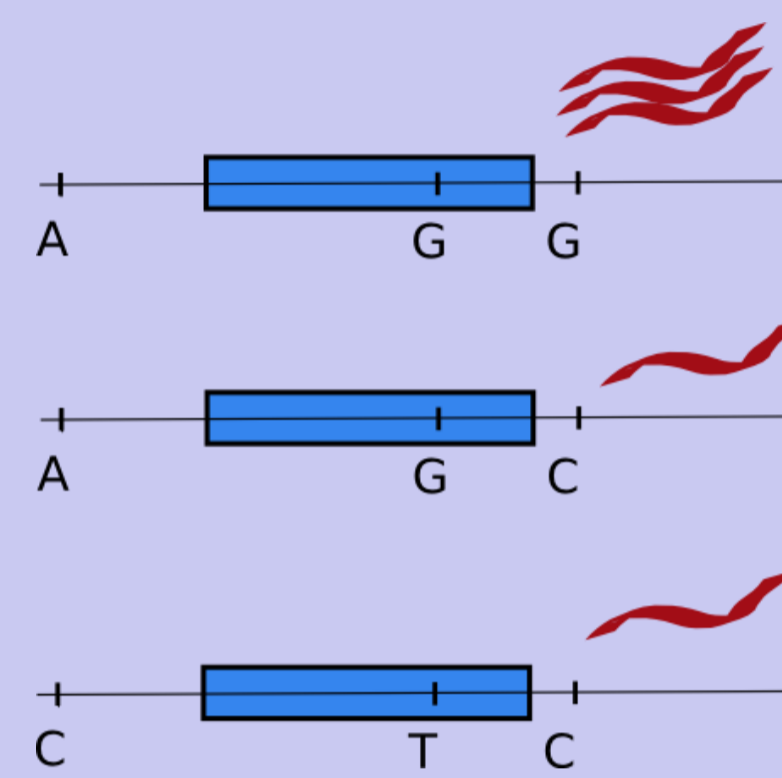
Définitions et Exemple

Haplotype: combinaison d'allèles sur une portion d'ADN

Effet cis : influence d'un allèle ou d'un haplotype sur l'expression d'un gène se situant à proximité

Sonde : Séquence ADN utilisée pour mesurer l'expression d'un gène

Exemple d'un gène dont l'expression est favorablement influencée par l'haplotype AGG



Problématique

- Analyse de l'influence des cis-haplotypes sur l'expression génique
- Etude systématique sur le génome humain

Méthode

- Sélectionner les haplotypes d'au plus 4 SNPs dans une fenêtre de +/- 200 kb du gène
- Pour un haplotype et une sonde donnée, analyse statistique de l'influence de l'haplotype sur l'expression du gène sur la population d'étude avec combinHaplo [1]

Caractéristique du projet

Projet Cardiogenics [2]

Nombre d'individus dans la population: 758

Nombre de SNPs considérés: 346 749

Nombre de sondes utilisées: 19 805

Nombre de gènes étudiés: 15 426

Nombres d'analyses à faire: 2 097 693 183

Projet débuté en septembre 2010

Caractéristique de l'analyse

Durée d'une analyse: de quelques secondes à plusieurs minutes selon le nombre de SNPs dans l'haplotype

Si chaque analyse dure 3 secondes, l'ensemble du projet nécessiterait au moins 10 ans en n'utilisant qu'un seul CPU

Analyse peu gourmande en mémoire

Besoin d'une très grande puissance de calcul

Utilisation d'une grille de calcul (1)

- Cluster local peu adapté: pas assez de puissance de calcul
- Utilisation de la grille avec comme VO biomed
- Offre un grand nombre de processeurs pour les calculs situés en Europe
- fin de l'analyse: juin-juillet 2011

Utilisation d'une grille de calcul (2)

- Développement d'un script en python pour lancer automatiquement une série d'analyses d'association
- Lancement automatique du script
- Problèmes rencontrés:
 - (1) Gestion du flux de tâches à soumettre sur la grille
 - (2) Gestion du renouvellement de la durée des proxy
 - (3) Nécessité de relancer des tâches

Conclusion et perspective

- Utilisation concluante de la grille de calcul
- Meilleure gestion des erreurs ?
- Détection d'un effet haplotypique pour 13,4% des sondes étudiées
- Phase de réplication sur une population indépendante
- Importance lorsque le gène a été trouvé dans la littérature impliqué dans des maladies

Références

[1] Tregouet DA, König IR, Erdmann J, Munteanu A, Braund PS, et al. (2009) Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. Nat Genet 41: 283-285

[2] <http://www.cardiogenics.eu>

www.cardiogenics.eu

<http://www.cardiogenics.eu> www.cardiogenics.eu