

Université de Montréal

LA CARTOGRAPHIE DES SITES DE RÉGULATION GÉNÉTIQUE À PARTIR
DE DONNÉES DE DÉBALANCEMENT ALLÉLIQUE

par

Emilio Damian Vello

Programme de bio-informatique

Faculté de médecine

Mémoire présenté à la Faculté des études supérieures et postdoctorales

en vue de l'obtention du grade de M.Sc.

en bio-informatique

Septembre 2011

© Emilio Damian Vello, 2011

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :
LA CARTOGRAPHIE DES SITES DE RÉGULATION GÉNÉTIQUE À PARTIR
DE DONNÉES DE DÉBALANCEMENT ALLÉLIQUE

présenté par :
Emilio Damian Vello

a été évalué par un jury composé des personnes suivantes :

Dr Nicolas Lartillot
président-rapporteur

Dr Damian Labuda
directeur de recherche

Dr Guillaume Bourque
membre du jury

Résumé

En 1975, Wilson et King ont proposé que l'évolution opère non seulement via des changements affectant la structure des protéines, mais aussi via des mutations qui modifient la régulation génétique. L'étude des éléments régulateurs de l'expression génétique a un rôle important dans la compréhension de l'expression de différentes maladies et de la réponse thérapeutique. Nous avons développé un algorithme bio-informatique qui nous permet rapidement de trouver des sites de régulation génétique à travers tout le génome et pour une grande quantité de gènes. Notre approche consiste à trouver des sites polymorphes (SNPs) qui sont en déséquilibre de liaison avec le débalancement allélique (AI) afin de cartographier la région régulatrice et le site responsable. Notre méthode est avantageuse par rapport à d'autres méthodes, car elle n'a pas besoin des données « phasées ». De plus, les données de débalancement allélique ne sont pas affectées par des facteurs externes étant donné qu'ils sont mesurés dans la même cellule. Nous avons démontré que notre approche est fiable et qu'elle peut détecter des sites loin du gène. De plus, il peut être appliqué à des données de génotypage sans avoir besoin de les « phaser » .

Mots clés: SNPs, Débalancement Allélique, AI, Région Régulatrice, Déséquilibre de Liaison, LD

Abstract

Wilson and King (1975) proposed that evolution frequently operates through mutations affecting genetic regulation. Likewise, it is expected that genetic variation responsible for inter-individual differences will be due to variation in regulatory sites. Identifying such sites is thus important in the genetic and medical research. We have developed a new bioinformatics algorithm to find genome-wide regulatory sites for a big number of genes. Individuals carrying different alleles at a regulatory site will exhibit allelic imbalance(AI) due to differential expression of the two copies the same locus. Our approach consists of searching polymorphic sites (SNPs) in linkage disequilibrium with AI in order to map regulatory regions. We have detected many SNPs associated to the regulation of different genes pointed in previous studies. We have also found regulatory regions far from the transcription start site (TSS). The major advantage of this method is that phased data is not needed. In addition, AI data has the benefit of not being affected by external factors since it is measured in the same cell. The results show that our approach is reliable and it can detect sites far from the gene.

Keywords: SNPs, Allelic Imbalance, AI, Regulatory Region, Linkage Disequilibrium, LD.

Table des matières

Résumé	i
Abstract	ii
Liste des tableaux.....	v
Liste des figures.....	vi
Liste des abréviations.....	vii
Dédicace.....	viii
Remerciements.....	ix
Chapitre 1. Introduction	1
1.1 La transcription génétique.	2
1.2 La régulation génétique.....	3
1.3 Les éléments régulateurs	3
1.4 La régulation en Cis et Trans.....	4
1.5 Méthodes de détection des éléments régulateurs.....	6
1.5.1 Génomique comparative.....	6
1.5.2 Analyses des gènes rapporteurs.....	7
1.5.3 Sites sensibles au clivage par la Dnase I.....	7
1.5.4 Sites de liaison de facteurs de transcriptions.....	8
1.6 Débalancement allélique (AI) et les méthodes de détection.	9
1.7 Problèmes dans la détection et la cartographie de AI.....	11
1.8 Histoire généalogique des haplotypes à deux sites bi-alléliques.....	13
1.9 Mesure de LD	14

1.10 Cartographie de variations régulatrices via LD.....	15
1.11 Objectifs.....	16
Chapitre 2. Méthodologie.....	18
2.1 Tests statistiques.....	18
2.1.1 Test binomial.....	18
2.1.2 Test génotypique de contingence 2x3	19
2.2 Données empiriques.....	20
2.2.1 Les donnée de génotypage.....	20
2.2.2 Les données de AI.....	21
2.2.3 Détermination d'individus en AI.....	23
2.3 Programmation et implémentation informatique.....	24
2.3.1 Recherche des régions régulatrices.....	24
2.3.2 Creation d'un jeux de données pour la détermination de faux positifs....	25
2.3.3 Comparaison de méthodes de LD.....	25
Chapitre 3. Résultats.....	27
3.1 Analyses globales de données et catégorisation.....	27
3.2 Application des tests à travers des exemples.....	33
3.3 Erreur empirique.....	42
3.4 Correspondance avec la littérature.....	43
3.5 Tests génotypiques pour la détection de LD.....	46
Chapitre 4. Discussion et conclusion.....	52
Bibliographie.....	60
Annexe I.....	xvi

Liste des tableaux

Tableau II.I: Format et exemples du fichier d'expression.....	22
Tableau III.I : Analyse globale de l'ensemble des gènes.....	30
Tableau III.II: Estimation de l'erreur empirique.....	42
Tableau III.III: Analyse des gènes détectés par des études précédentes.....	44
Tableau III.IV: Liste de gènes détectés dans des études précédentes.....	45
Tableau III.V. Comparaison de la performance des tests de LD	51

Liste des figures

Figure 1.1: Représentation graphique d'éléments régulateurs.....	5
Figure 1.2: Représentation du débalancement allélique (AI).....	9
Figure 1.3 : Configuration haplotypique.....	13
Figure 1.4: Représentation graphique de la région régulatrice.....	16
Figure 2.1: Histoire génotypique entre le site A et R.	19
Figure 2.2: Association entre le site R, le site A et le AI en cas de LD.....	20
Figure 2.3: Tableaux de contingences pour évaluer le LD.....	26
Figure 3.1: Valeurs de p en fonction du nombre d'individus en AI.....	31
Figure 3.2: Valeurs de p en fonction de la distance.....	32
Figure 3.3: Valeurs de p du test de contingence 2x3 pour le gène LRRIQ3.....	35
Figure 3.4: Cartographie du site de régulation du gène ALDH8A1(1).....	36
Figure 3.5 : Cartographie du site de régulation du gène ALDH8A1(2).....	37
Figure 3.6 : Cartographie du site de régulation du gène AHI1.....	38
Figure 3.7: Cartographie du site de régulation du gène XKR9.....	39
Figure 3.8: Cartographie du site de régulation du gène QDPR.....	40
Figure 3.9: Cartographie du site de régulation du gène NEFH.....	41
Figure 3.10: Corrélacion entre les tests de contingence 2x2 et 2x3	48
Figure 3.11: Corrélacion entre les tests de contingence 2x2 et 3x3.....	49
Figure 3.12: Comparaison des tests LD sur des données réelles.....	50

Liste des abréviations

ADN	Acide DésoxyriboNucléique
AI	Débalancement allélique (« Allelic Imbalance »)
ARN	Acide RiboNucléique
CEPH	Centre d'Étude du Polymorphisme Humain
ChIP	Immunoprécipitation de la chromatine
eQTL	« Expression Quantitative Trait Loci »
LD	Déséquilibre de liaison (« Linkage Disequilibrium »)
SNP	Polymorphisme de simple nucléotide (« Single Nucleotide Polymorphism »)
SH	Sites d'hypersensibilité
SCN	Séquences conservées non codantes

Dédicace

À ma femme Mercedes, mes filles et mes parents

Remerciements

Je tiens tout d'abord à remercier mon directeur, Dr Damian Labuda, qui m'a fait confiance en me donnant l'opportunité de réaliser ce projet de recherche. Je lui exprime aussi ma gratitude pour ses conseils, son soutien et son support inconditionnel tout au long de mes études.

J'aimerais aussi remercier Jean-François Lefebvre d'avoir été à mon côté tout au long de cette aventure, pour ses conseils, ses réflexions et son aide précieuse avec la correction de ce mémoire.

Je remercie aussi Vania et Claudia pour leur aide, leurs conseils et leur esprit d'équipe.

Je veux aussi remercier les membres du jury d'avoir accepté d'évaluer mon mémoire et tous mes professeurs pour leur dévouement. J'aimerais aussi remercier Elaine Meunier pour sa gentillesse et ses conseils administratifs.

Je remercie également l'amour de ma vie, Mercedes, pour son soutien moral, son encouragement et tous ses sacrifices pour que je puisse faire cette maîtrise. Je remercie aussi mes filles pour leur patient et leur compréhension pendant mes études.

Je remercie aussi mes parents pour leurs prières et finalement Dieu le Tout-Puissant pour me donner le courage et la force de réaliser ce travail.

Chapitre 1. Introduction

Avant que la génétique et même la biologie existent comme sciences, l'homme s'est posé la question sur l'hérédité. Les peuples anciens ont amélioré les plantes cultivées et les animaux domestiques en sélectionnant les meilleurs individus pour la reproduction à chaque génération. L'être humain s'est toujours interrogé « pourquoi les enfants ressemblent à leurs parents » [J.F.Griffiths et al., 2001]. La génétique a vu le jour en 1860, lorsque Gregor Mendel a publié ses expériences sur la transmission des caractères chez les pois comestibles. À partir de ses observations, il a énoncé les trois lois principales de la génétique classique, la loi de dominance, la loi de ségrégation et la loi de la distribution indépendante.

Mendel a mis en évidence le concept fondamental de la génétique, le gène. Ce dernier peut être défini comme les éléments de l'hérédité qui sont transmis des parents à leurs descendants lors de la reproduction qui déterminent les caractères héréditaires. En 1975, King et Wilson ont étudié la distance génétique entre l'être humain et le chimpanzé et ils ont constaté que cette distance semble trop courte par rapport à leurs différences anatomiques et à leurs styles de vie. Ils ont suggéré que les différences biologiques observées entre les deux espèces sont dues aux mutations affectant la régulation génétique [King et al. 1975]. Aujourd'hui, il est aussi connu que l'altération du processus de régulation génétique est la cause de plusieurs maladies communes, telles que le cancer [Diaw et al., 2007]. La connaissance des variations génétiques dans les sites régulateurs est essentielle pour identifier les causes des ma-

ladies associées aux polymorphismes régulateurs [Pastinen et al., 2004]. C'est pourquoi les études de ces variations ont une importance fondamentale en médecine moderne, ainsi que dans la compréhension de l'évolution humaine.

Nous avons développé une nouvelle approche d'analyse de données de déséquilibre allélique pour cartographier des régions régulatrices contenant des variations génétiques responsables de ce déséquilibre. Nous l'avons appliquée à tout le génome en étudiant tous les gènes dont l'information nous était disponible. Cette méthode a deux avantages importants. La mesure de déséquilibre allélique n'est pas affectée par des facteurs externes à la cellule et notre approche n'a pas besoin de l'information haplotypique.

1.1 La transcription génétique.

Lorsque la cellule a besoin d'une protéine spécifique ou bien d'un ARN fonctionnel, l'ADN est copié en ARN. Tout d'abord, l'enzyme ARN polymérase identifie une région spécifique appelée promoteur et s'y attache formant un complexe avec l'ADN. Ensuite, elle commence la transcription à partir d'un site proche du promoteur appelé « site d'initiation de la transcription ». Puis, elle synthétise l'ARN en « se déplaçant » sur le brin d'ADN qui sert de matrice. Finalement, la séquence de terminaison arrête la polymérase qui ensuite libère la molécule d'ARN. Le produit immédiat de la transcription est appelé transcrit et il va subir plusieurs changements avant de devenir un ARN fonctionnel ou un « programme » à la production de protéines.[Lewin, 2008]

1.2 La régulation génétique.

Le contrôle de l'expression génique affecte à la plupart des fonctions biologiques. Les gènes s'expriment à un moment déterminé, dans un type de cellule spécifique et en quantité adéquate. De plus, les processus biologiques requièrent l'expression coordonnée de milliers de gènes. Ce processus de contrôle est appelé la régulation génétique. Le mécanisme de régulation n'est pas complètement compris, et spécialement, où ce mécanisme est codé; nous savons cependant qu'il existe plusieurs régions dans le génome qui affectent l'expression des gènes. Britten et Davison ont proposé que l'abondance des éléments régulateurs chez les eucaryotes supérieurs et la complexité du réseau de régulation conduisent à la complexité biologique [Noonan et al., 2010].

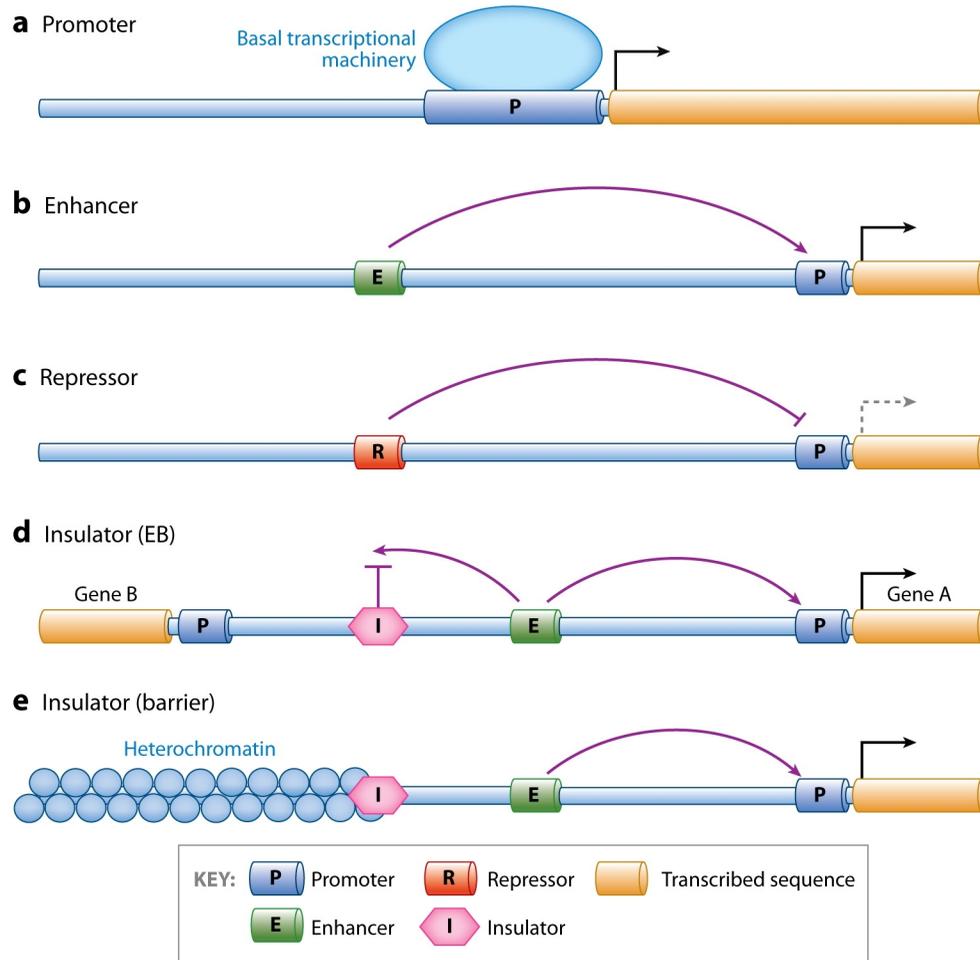
1.3 Les éléments régulateurs

Il existe différents types d'éléments régulateurs (figure 1.1). Les promoteurs sont des éléments régulateurs situés typiquement à quelques dizaines de paires de bases du gène. La machinerie de la transcription utilise les promoteurs pour s'assembler et pour déterminer l'orientation et l'origine de la transcription. Les amplificateurs (en anglais « enhancers») et les silenceurs sont des éléments régulateurs qui produisent une régulation positive ou négative respectivement, en interagissant avec les promoteurs. Ils peuvent être situés près ou loin du promoteur du gène cible [Noonan et al., 2010]. Les insulateurs sont des éléments régulateurs qui peuvent bloquer l'activité des amplificateurs ou bien jouer un rôle de barrière. Ce blocage consiste à obstruer l'effet d'activation de l'amplificateur quand l'insulateur est placé entre l'amplificateur

et le gène. La fonction de barrière consiste à séparer la chromatine active et inactive. Les régions de contrôle contiennent un programme transcriptionnel pour plusieurs gènes. Ces régions contiennent plusieurs séquences régulatrices comme les amplificateurs et silenceurs [Noonan et al., 2010]. Les amplificateurs, les silenceurs, les insulateurs et les régions de contrôle peuvent être localisés à plusieurs kb de distance du gène qu'ils régulent.

1.4 La régulation en *Cis* et *Trans*.

La régulation en *cis* désigne les éléments régulateurs qui se trouvent sur le même chromosome que le gène qu'ils régulent. Ils affectent seulement la copie du gène localisé sur la même copie du chromosome. Ces éléments sont généralement situés près du site d'initiation de la transcription. Cependant, ils peuvent se trouver à quelques milles paires de bases de leur gène. Les amplificateurs, les silenceurs et les insulateurs sont en général des éléments en *cis*. À l'inverse, la régulation en *Trans* désigne les éléments régulateurs qui se trouvent sur d'autre chromosome. Ces éléments influent les deux copies parentales du gène régulé [Pastinen et al., 2006].



AR Noonan JP, McCallion AS. 2010.
Annu. Rev. Genomics Hum. Genet. 11:1–23

Figure 1.1: Représentation graphique des différents types d'éléments régulateurs. a) promoteur, b) amplificateur, c) silencieux, d) insulateur en fonction de blocage et e) insulateur en fonction de barrière. Cette figure a été tirée de [Noonan et al., 2010]

1.5 Méthodes de détection des éléments régulateurs.

Dans la littérature, nous pouvons trouver plusieurs approches pour détecter une région régulatrice à l'échelle génomique. Dans toutes ces approches, il y a une étape expérimentale fournissant les données biologiques suivie d'une étape d'analyse bio-informatique. Certaines méthodes sont utilisées à l'échelle génomique, alors que d'autres méthodes visent des régions régulatrices pour des gènes spécifiques [Noonan et al., 2010].

1.5.1 Génomique comparative

Dans le génome, des séquences fonctionnelles suivent souvent un processus de sélection purificatrice. En effet, les mutations dans un gène ou dans une région régulatrice qui altèrent la fonction et/ou la quantité de protéines ont un effet négatif dans l'organisme et par conséquent, elles sont éliminées de la population. Par contre, les mutations se produisant dans des régions « non fonctionnelles » n'ont pas de conséquences négatives pour les organismes. C'est ainsi que ces mutations peuvent être fixées grâce à la dérive génétique. Si l'on compare les génomes de deux espèces, la similarité entre eux dans des régions « non fonctionnelles » diminue au fur et à mesure que la distance évolutive augmente. Par contre, dans des régions où la fonction est maintenue chez les deux espèces, les séquences génomiques sont conservées et les changements observés sont beaucoup moins importants. Typiquement, les éléments régulateurs sont situés dans des séquences conservées non codantes (SCN) et plusieurs études de génomique comparative ont été faites pour déterminer la distribu-

tion de SCN dans le génome humain et ainsi localiser des séquences régulatrices [Hardison, 2003].

1.5.2 Analyses des gènes rapporteurs

Cette méthode expérimentale est beaucoup utilisée pour détecter des séquences régulatrices dans des études qui ciblent un gène ou un groupe de gènes particuliers. Cette approche peut également être utilisée dans des analyses à l'échelle génomique. Dans cette technique, la région d'ADN ciblée est transférée sur un plasmide en amont d'un gène rapporteur. Un gène rapporteur est un gène qui produit une protéine qui est facilement observable en laboratoire, par exemple, la protéine fluorescente verte (GFP). Le plasmide est donc transféré à des cellules en culture et l'activité du gène rapporteur est mesurée pour déterminer s'il y a des éléments qui altèrent son expression dans la région testée. Afin d'effectuer des études à l'échelle génomique, les régions d'ADN à tester sont choisies au hasard en utilisant des enzymes ou d'autres techniques de clivage de l'ADN. [Maston et al., 2006]

1.5.3 Sites sensibles au clivage par la Dnase I

Dans les cellules, il y a beaucoup de processus qui demandent une interaction entre diverses protéines et l'ADN. Cette interaction peut être réalisée dans des régions où la chromatine est accessible. C'est ainsi que la Dnase I¹ clive l'ADN dans des régions libres de nucléosomes. Ces « fenêtres ouvertes » dans la chromatine sont appelées sites d'hypersensibilité (SH). Dans les *Saccharomyces cerevisiae*, les SH se

1 Dnase I est une endonuclease qui clive l'ADN non-spécifique

trouvent dans des centromères, dans des silenceurs, dans des séquences propices à la recombinaison, dans des origines de la réplication de l'ADN et dans des promoteurs. Tous ces éléments ont une activité en *cis*. Pour détecter les SH, une des premières méthodes a été la digestion de l'ADN par la DNase I suivie d'un marquage indirect de l'ADN résultant. [Gross and Garrard, 1988]. Aujourd'hui, il y a plusieurs méthodes qui utilisent le même principe. Par exemple, Sabo et ses collègues [Sabo et al., 2006] ont développé une méthode appelée DNase-array pour examiner les SH à l'échelle génomique. Brièvement, ils ont fait la digestion de l'ADN en utilisant la Dnase I, pour ensuite employer une puce à ADN pour analyser les morceaux d'ADN résultants.

1.5.4 Sites de liaison de facteurs de transcriptions.

Le but de cette méthode est la détermination des séquences génomiques où les facteurs de transcription se lient à l'ADN. De façon générale, la méthode consiste à faire l'immunoprécipitation de la chromatine (ChIP) du facteur de transcription d'intérêt pour isoler l'ADN cible. Ensuite, les séquences d'ADN sont identifiées en utilisant soit une puce d'ADN (ChIP-chip) avec les régions significatives du génome, soit le séquençage direct (ChIP-seq). Dans le premier cas, la puce d'ADN est composée de régions génomiques spécifiques telles que des régions promotrices, de grands segments génomiques, des chromosomes entiers ou même le génome au complet. Cependant, il faut toujours avoir un échantillon de référence, contrairement à la deuxième méthode. Les séquences génomiques obtenues par la ChIP sont séquencées

par une des technologies modernes de séquençage. [Euskirchen et al., 2007]

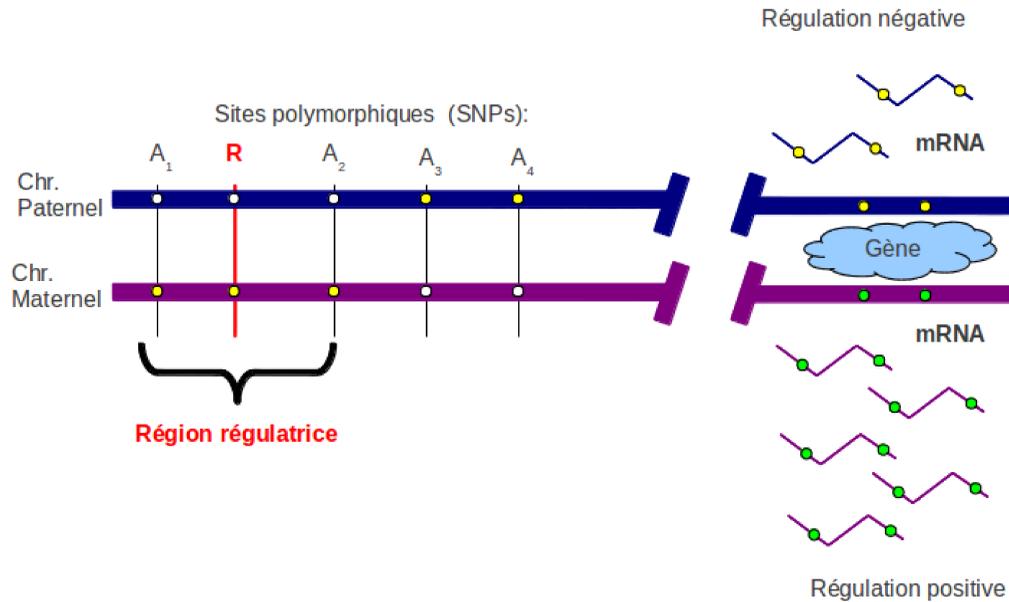


Figure 1.2: Représentation du déséquilibre allélique (AI). En rouge, la variation régulatrice qui produit le phénotype d'AI dans le gène. À droite, nous avons l'illustration de régulation positive (plus quantité de transcrits) et régulation négative (moins quantité de transcrits).

1.6 Déséquilibre allélique (AI) et les méthodes de détection.

Dans la régulation génétique, il existe un phénomène assez courant qui s'appelle le déséquilibre allélique (AI pour « Allelic Imbalance » en anglais) (figure 1.2). Nous pouvons définir l'AI comme la différence dans le niveau d'expression d'un gène entre les deux copies d'un chromosome. En d'autres termes, les quantités de transcrits qui proviennent de chaque copie d'un gène sont différentes. Cet événement est causé

par l'hétérozygotie du site régulateur. Imaginons que nous avons les allèles Rr en un site régulateur, et que les facteurs de transcriptions reconnaissent ce site plus facilement quand ils trouvent l'allèle R. Cela va produire un déséquilibre de l'expression mesurable entre les deux allèles [Pastinen and Hudson, 2004]. Différentes techniques ont été utilisées pour détecter le AI telles que SnapShot [Bray et al., 2003], séquençage [Tuupanen et al., 2008], pyroséquençage [Loeuillet et al., 2007], iPLEX [Ding et al., 2004], intensité des bandes dans le gel [Heighway et al., 2005], entre autres. Aujourd'hui, les deux méthodes les plus répandues sont la puce à ADN et le séquençage de deuxième génération. Ces deux techniques sont fortement utilisées, car elles peuvent être utilisées à l'échelle génomique. Les autres techniques sont employées à des fins de validation ou pour analyser des loci spécifiques. La première méthode consiste à traiter l'ARN messager ou l'ADN codant avec une puce à ADN qui contient des sondes avec les transcrits les plus fréquents. Chaque transcrit est reconnu par un ou plusieurs SNPs hétérozygotes appelés marqueurs. En comparant l'intensité de chaque sonde et normalisant avec l'ADN génomique, un ratio de AI peut être obtenu [Ge et al., 2009]. La technique de séquençage de deuxième génération consiste en un séquençage quantitatif de l'ARN messager. Le nombre de transcrits provenant de chaque copie chromosomique est compté en utilisant aussi des marqueurs [Montgomery et al., 2010]. En général, pour pouvoir quantifier le nombre de transcrits d'un gène, il faut avoir au moins un marqueur dans le gène en question.

1.7 Problèmes dans la détection et la cartographie de AI.

Il existe différents problèmes dans la détection de l'AI. Premièrement, même si un individu présente de l'AI pour un gène particulier, si ses marqueurs sont homozygotes, l'AI ne pourra pas d'être détecté. On ne pourrait pas différencier la provenance des transcrits. Cet individu sera alors classé comme non-AI par défaut. Cela réduira la visibilité de la variation régulatrice pour les tests statistiques. Le deuxième problème qui peut arriver est l'incohérence dans l'information que donne chaque marqueur, c'est-à-dire, lorsque pour un gène spécifique chez un individu spécifique, un marqueur indique que la copie paternelle du gène est régulée positivement alors que le marqueur à côté indique le contraire. Ainsi, Ge et collègues [Ge et al., 2009] ont utilisé un système de fenêtres avec un nombre minimal de marqueurs pour assurer la présence de AI. Cette incohérence est causée par de grandes variations dans l'intensité du signal entre différents marqueurs d'un même individu [Wagner et al., 2010].

Pour cartographier l'origine des variations régulatrices, la détection de l'AI n'est pas suffisante. Il faut faire des études d'association entre des sites polymorphes dans le génome et le phénotype de AI. La plupart du temps, ces études utilisent l'information haplotypique; cependant, la reconstruction incorrecte d'haplotypes peut conduire à l'obtention des faux positifs ou faux négatifs dans les études d'association. Il existe plusieurs algorithmes pour la détermination de la phase haplotypique, mais aucun n'est précis à 100 %. De plus, ce processus devient plus compliqué avec des indivi-

des non apparentés, et la précision des algorithmes de reconstruction haplotypique (phasage) est plus faible au fur et à mesure que la distance augmente [Higasa et al., 2009] [Fan et al., 2011] [Marchini et al., 2006]. Dans son étude, [Ge et al., 2009] a utilisé une régression linéaire entre l'intensité de AI et le génotype ordonné (chromosome 1 / chromosome 2) du SNP à tester. Ils ont seulement analysé une région de 250 kb autour de chaque gène sur des données phasées de HapMap II. De cette façon, ils ont essayé d'éviter les possibles problèmes de phasage qui arrivent en utilisant des méthodes basées sur l'information haplotypique. Ainsi, la résolution correcte des haplotypes ajoute un autre problème à ceux propres à la détection de l'AI. Ces trois problèmes rendent difficile la détection de variations régulatrices, spécialement, celles qui se trouvent loin du gène.

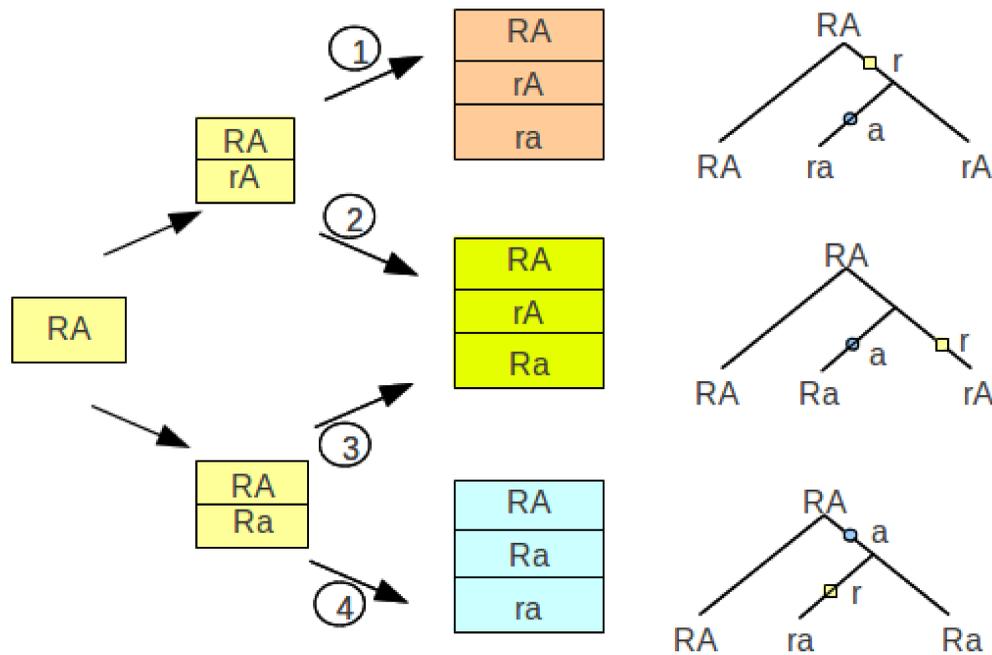


Figure 1.3 : Les quatre chemins possibles mutuellement exclusifs qui conduisent à trois trios d'haplotype. Dans la partie gauche, la séquence des mutations donne trois trio d'haplotypes commençant par l'haplotype ancestral. Dans la partie droite, les arbres généalogiques pour les différents chemins sont montrés. Nous assumons qu'il n'y a pas de mutations récurrentes.

1.8 Histoire généalogique des haplotypes à deux sites bi-alléliques

La succession de deux mutations dans un segment génomique crée deux nouveaux haplotypes portant des allèles dérivés (figure 1.3). Sans recombinaison ces deux sites polymorphes restent en déséquilibre de liaison (LD pour « linkage disequilibrium »). En absence de recombinaison, le LD est complet; il n'existe que trois haplotypes dont un « non-muté » portant deux allèles ancestraux. Tel que présenté dans la figure 1.3,

il existe quatre chemins pour y arriver, dont deux (chemin 2 et 3) donnant le même trio d'haplotypes. Suivant le chemin 1, c'est la mutation « R » à « r » qui arrive en premier i.e.: en amont de la mutation « A » à « a ». Suivant le chemin 4 c'est la mutation « R » à « r » qui arrive en deuxième, i.e.: en aval de la mutation « A » à « a ». Les chemins 2 et 3 décrivent la situation où les deux mutations arrivent en parallèle sur des branches séparées de l'arbre généalogique. En principe, sachant l'état ancestral de chacun des sites polymorphes impliqués ainsi que l'ensemble des haplotypes présents dans la population, on devrait être capable de distinguer entre les différents chemins qui ont mené aux haplotypes observés.

Avec le temps, le LD entre ces sites décroît à cause de la recombinaison. Le système de quatre haplotypes va éventuellement atteindre l'équilibre, où la fréquence des haplotypes reflète la fréquence des deux allèles. La force du LD entre deux sites indique l'origine récente et/ou la proximité physique entre ces sites, dépendamment de la distance génétique qui les sépare.

1.9 Mesure de LD

Le niveau de LD entre deux sites bi-alléliques peut être évalué en tenant compte des fréquences haplotypiques par rapport aux fréquences alléliques. Les deux paramètres de LD, le plus souvent utilisés sont D' et r^2 [Flint-Garcia et al., 2003]. Une autre approche est l'évaluation de la dépendance entre ces allèles, autrement dit l'association allélique, dans un tableau de contingence 2×2 . Ces approches requièrent l'inférence préalable des haplotypes à partir des génotypes. Cependant, il n'est pas

toujours possible de le faire avec précision [Tewhey et al., 2011], spécialement quand l'information familiale est absente. Il y a d'autres approches qui utilisent seulement l'information génotypique. Par exemple, la méthode de la mesure de LD composé. ((Weir, 1979 — attribué à P. M. Burrows) [Sved 2009]. Ces méthodes s'avèrent très utiles pour estimer le LD quand la reconstruction haplotypique est difficile.

1.10 Cartographie de variations régulatrices via LD

La cartographie génétique sert à localiser et ensuite identifier les variations génétiques responsables d'un phénotype particulier. Ici, le phénotype mesuré est celui du AI. Pour identifier cette variation, il existe deux possibilités. La première est d'utiliser des analyses de liaison où les liens familiaux sont utilisés pour prédire la correspondance entre le génotype et le phénotype. Cependant, ces études sont difficiles à réaliser car il faut avoir de grandes familles qui présentent le phénotype recherché. La deuxième possibilité est d'utiliser des analyses d'association via le LD. Ici, la diversité naturelle est sondée dans la population pour identifier des sites polymorphes associés à la variation phénotypique d'intérêt. Les événements de recombinaison font ressortir principalement les SNPs en LD avec la région responsable du phénotype [Flint-Garcia et al., 2003]. Étant donné que les SNPs en LD se trouvent généralement proche, la région responsable du phénotype peut être cartographiée. Avec ce type d'études, le gène de la fibrose kystique [Kerem et al, 1989], de la dystrophie dysplasie et un des gènes le plus important associé à la maladie d'Alzheimer [Corder et al.,

1994] ont pu être identifiés [Flint-Garcia et al., 2003]. Ici, les variations régulatrices sont cartographiées en évaluant le LD entre le phénotype de AI et des sites polymorphes connus donnant comme résultat les sites qui sont en LD avec le site régulateur (figure 1.4). De cette façon la région régulatrice contenant la variation génétique R responsable de l'AI peut être localisée.

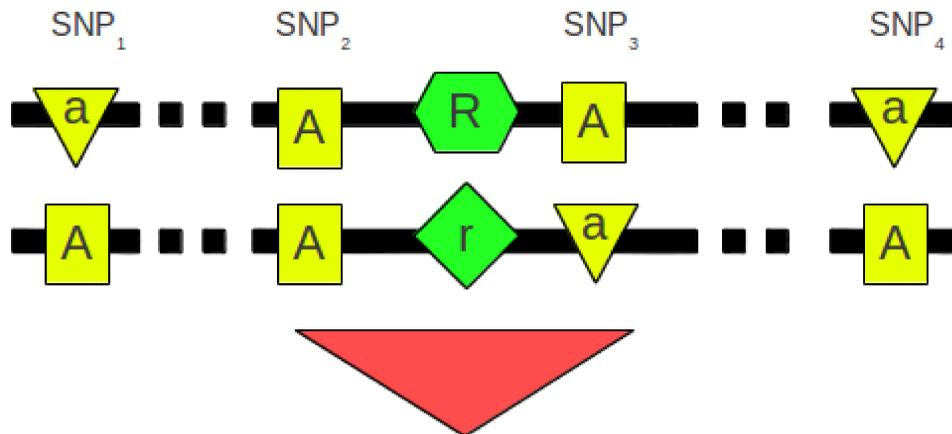


Figure 1.4: Représentation graphique de la région régulatrice cherchée. En vert, la variation génétique responsable d'AI. En jaune, différents sites polymorphes connus (SNPs). Le triangle rouge représente le LD entre le site régulateur et des SNPs connus. Cette association est identifiée en évaluant le LD entre les SNPs et l'AI.

1.11 Objectifs

L'objectif principal de ce travail est le développement d'une approche pour cartographier des régions régulatrices en utilisant les données de AI. Une approche utilisable à l'échelle génomique qui n'est pas affectée par les erreurs de phasage et qui

peut détecter des sites régulateurs distants. L'avantage d'utiliser des données de AI vient du fait que la différence d'expression est mesurée dans la même cellule et cette mesure n'est donc pas influencée par des facteurs externes comme c'est le cas dans les autres données d'expression. La possibilité d'utiliser des données génotypiques sans connaître les haplotypes donnent à notre approche un avantage important. Fan et al. donnent un très bel exemple qui illustre la limitation de l'interprétation des données ou pire encore, la possibilité de tirer des conclusions erronées si des haplotypes avec des erreurs de phasage sont utilisés [Fan et al., 2011]. Dans les données de HapMap II qui vont être introduites dans la méthodologie, on estime qu'il y a environ une erreur de phasage tous les 3 à 8 Mb en travaillant avec des trios familiaux. De façon générale, quand l'analyse se fait sur des individus en absence d'information familiale, une erreur de phasage se produit tous les 300 kb. C'est pourquoi, cet atout permet de voir des éléments régulateurs qui se trouvent à grande distance du gène qu'ils régulent. Nous avons aussi étudié deux méthodes pour regarder le débalancement de liaison (LD), car la plupart des méthodes sont aussi influencées par les erreurs de phasage. Cependant, ce dernier est un objectif secondaire dérivé de l'analyse de notre but principal.

Chapitre 2. Méthodologie

2.1 Tests statistiques

Pour évaluer le LD entre le site régulateur R et le site polymorphe A, nous introduisons les deux tests suivants qui exploitent l'information partielle sur les fréquences alléliques et génotypiques données par le AI sur le site R.

2.1.1 Test binomial

L'observation du phénotype de AI indique que le site R est hétérozygote. Les allèles R et r sont associés à la régulation positive ou à la régulation négative. Si la phase haplotypique est connue, un allèle de n'importe quel site hétérozygote A sur le même chromosome peut être assigné à la régulation positive ou négative selon le chromosome où il se trouve. En absence de LD entre le site R et le SNP A testé, les allèles de ce dernier seront assignés aléatoirement entre régulation positive et négative avec la même probabilité. Par contre, si le site R et A sont en LD, nous observerons toujours le même allèle du site A associé soit à la régulation positive soit à la régulation négative (figure 2.2 A). La signification du LD peut être évaluée en utilisant la distribution binomiale avec une probabilité $p = 0.5$ d'observer un allèle de A lié à la régulation positive ou à la régulation négative. Ce test peut être seulement utilisé avec les individus en AI et hétérozygotes pour le site A analysé.

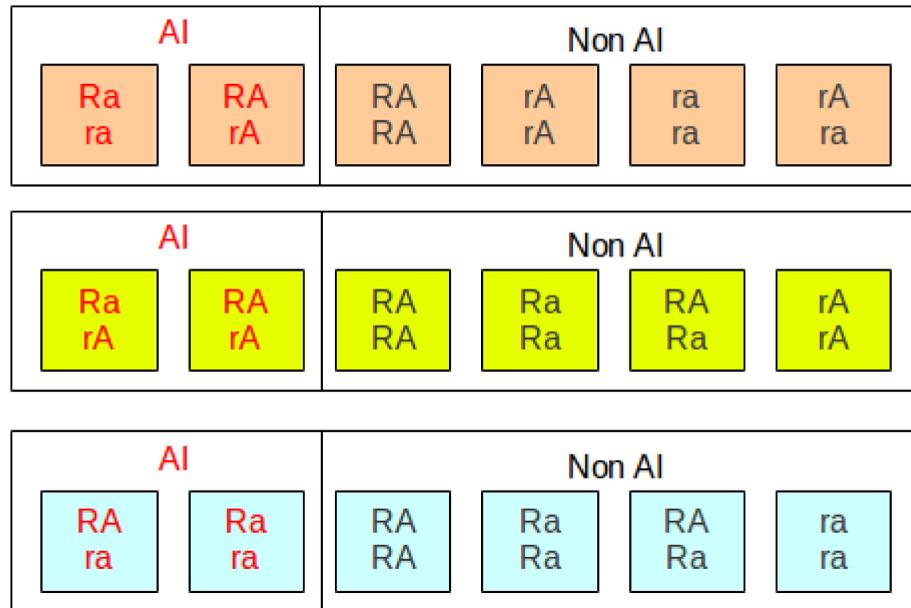


Figure 2.1: Trois ensembles de possibles génotypes selon la position de la mutation entre le site A et R. En complet LD, il y a toujours deux génotypes qui caractérisent les individus en AI. Il y a seulement une type d'homozygote présente soit AA ou aa.

2.1.2 Test génotypique de contingence 2x3

Si le site régulateur R et n'importe quel site polymorphe A sont en LD, la répartition des génotypes de A ne sera pas la même entre les individus en AI et les autres. Nous observerons un seul type de A homozygote (AA ou aa) (figure 2.1) chez les individus qui sont en AI. Si R et A ne sont pas en LD, Aa, AA et aa devraient être distribués de façon équivalente chez les individus en AI et les autres. Nous utilisons le test exacte de Fisher pour tester cette dépendance génotypique (figure 2.2 B). La fonction R `fisher.test()` est utilisée pour calculer les valeurs de p. Ce test est avanta-

geux pour ne pas être affecté par les erreurs de phasage.

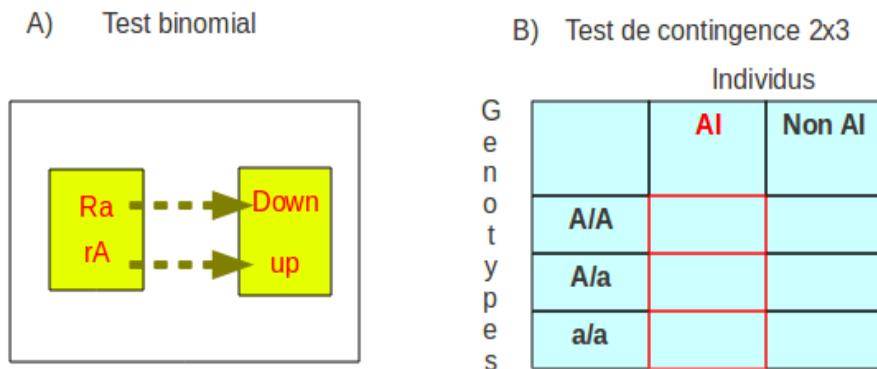


Figure 2.2: Représentation de l'association entre le site R, le site A et le AI en cas de LD pour le test binomial haplotypique (A) et pour le test génotypique de contingence 2x3 (B)

2.2 Données empiriques

2.2.1 Les données de génotypage

Les données de génotypage ont été obtenues de la phase II de HapMap (HapMap II). Cette base de données a plus de 3.1 millions de SNPs de 270 individus provenant de diverses populations du monde. Dans ce projet, nous avons utilisé 54 résidents d'Utah originaires d'Europe du Nord et de l'ouest non apparentés de la collection du Centre d'Étude du Polymorphisme Humain. (CEPH). Pour être plus précis, c'est la version 22 du 22 décembre 2007 de HapMap II où les génotypes sont phasés. Nous avons donc accès à l'information haplotypique.

2.2.2 Les données de AI

Les données de AI proviennent de Ge et al. 2009 et elles ont été obtenues à partir de 54 lignées cellulaires lymphoblastiques appartenant aux mêmes individus de HapMap II mentionnés plus haut. Brièvement, plusieurs marqueurs ont été utilisés à travers le génome pour évaluer le AI. Les intensités spécifiques d'ARN (ou ADN codant) et d'ADN génomique (pour normaliser) ont été mesurées dans une puce d'ADN illumina Infinium Human 1 M ou Human 1 M-Duo SNP [Ge et al., 2009]. L'index de AI a été mesuré par la différence d'intensité entre les marqueurs informatifs. Ces derniers sont des sites hétérozygotes dans les transcrits où le niveau d'expression " $R = \log(X_{raw} + Y_{raw}) > 1000$ ". "X" et "Y" représentent l'intensité de chaque version de transcrit.

Nous avons considéré qu'un marqueur indique la présence de différence d'expression si l'index de AI $> 0,1$. Cette valeur est la même qui a été utilisé par [Ge et al., 2009]. Dans le tableau II.I, il y a un exemple de l'information de AI utilisée.

Tableau II.I: Format et exemples du fichier d'expression reçu de [Ge et al., 2009]. Chaque ligne représente un marqueur. Le nom du gène, le SNP, le chromosome, la localisation et le génotype de chaque marqueur sont trouvés entre la colonne A et E. Ensuite, il y a trois colonnes pour chaque individu avec le génotype (F), le niveau d'expression (G) et le AI (H).

(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(F)	(G)	(H)	(F)	(G)	(H)
#genes	snp	chrom	location	AB	NA07357	NA07357	NA07357	NA11829	NA11829	NA11829	NA11830	NA11830	NA11830
-	rs10076494	5	91139 TC	CT		0	0 TT		0	0 CC		0	0
-	rs10039735	5	101534 TC	TC		0	0 CC		0	0 TT		0	0
-	rs1507712	5	111298 AG	GA		0	0 AA		0	0 GG		0	0
-	rs341267	5	171406 TC	CC		0	0 CC		0	0 TT		0	0
-	rs6867447	5	192702 TC	TC		0	0 CC		0	0 CC		0	0
PLEKHG4B	rs11949577	5	196197 AG	AG		0	0 GG		0	0 GG		0	0
PLEKHG4B	rs11745772	5	208137 AG	AG		1094	-0,2686 GG		0	0 GG		0	0
PLEKHG4B	rs9312845	5	208367 TC	TC		1077	-0,2607 CC		0	0 CC		0	0
PLEKHG4B	rs10076077	5	208671 TC	TC		1819	-0,5071 CC		0	0 CC		0	0
PLEKHG4B	rs1108867	5	222902 TC	CC		0	0 CC		0	0 CC		0	0
PLEKHG4B	rs4956987	5	227106 AG	GA		0	0 AG		0	0 GA		0	0
PLEKHG4B	rs2059860	5	233325 TG	GG		0	0 GG		0	0 GG		0	0
PLEKHG4B	rs7717970	5	234660 TG	TG		1353	0,109 GG		0	0 GG		0	0
PLEKHG4B	rs11133847	5	234762 AG	AG		0	0 GG		0	0 GG		0	0
PLEKHG4B	rs4956990	5	236134 AG	AG		4083	0,3455 GA		0	0 GG		0	0
PLEKHG4B	rs13176160	5	236401 AG	AG		0	0 GG		0	0 GG		0	0
PLEKHG4B	rs4956991	5	236516 AG	AG		0	0 GG		0	0 GG		0	0
PLEKHG4B	rs1053477	5	237393 TC	CT		0	0 TT		0	0 TT		0	0
PLEKHG4B	rs4956993	5	237702 AG	AG		0	0 GG		0	0 GG		0	0
PLEKHG4B	rs1053478	5	237750 AG	AG		0	0 GG		0	0 GG		0	0
PLEKHG4B	rs1053479	5	237760 AG	AG		0	0 GG		0	0 GG		0	0
PLEKHG4B	rs7727015	5	238498 TC	CT		0	0 TC		0	0 CT		0	0
PLEKHG4B	rs39788	5	240302 AG	GA		0	0 AA		0	0 AA		0	0

2.2.3 Détermination d'individus en AI

Étant donné que notre méthode a besoin de savoir si un individu présente ou pas de l'AI et si oui, à quel chromosome est associé la régulation positive, nous avons développé un algorithme qui permet de l'évaluer. Autrement dit, c'est un algorithme qui transforme des données quantitatives d'expression en des données qualitatives. Un transcrit contient plusieurs marqueurs. Ces derniers peuvent être classés en deux catégories : des marqueurs informatifs, correspondant aux sites hétérozygotes pour le transcrit, et des marqueurs significatifs qui sont les sites hétérozygotes avec un niveau d'expression plus élevé que 1000. Même valeur utilisée par [Ge et al., 2009]. Bien entendu, tous les marqueurs significatifs sont des marqueurs informatifs; mais, l'inverse n'est pas vraie. Un marqueur indique la présence d'AI si le niveau d'expression est supérieur à 1000 et son index est $> 0,1$ [Ge et al., 2009]. Pour déterminer si un individu présente de l'AI, nous calculons le pourcentage de marqueurs significatifs sur le total de marqueurs informatifs. Si ce pourcentage est supérieur à 20 %, nous estimons la moyenne de l'index de AI de tous les marqueurs significatifs seulement, sinon, tous les marqueurs informatifs sont utilisés. Ensuite, l'index global d'AI doit être $> 0,1$. Nous avons utilisé la moyenne pour éviter les problèmes de marqueurs contradictoires dans le jeu de données. Nous différencions aussi le type de marqueur à utiliser dans la moyenne pour éviter de mettre en AI les individus qui ont peu de marqueurs significatifs avec un signal faible.

2.3 Programmation et implémentation informatique

Tous les programmes informatiques ont été développés en java (openJDK version 1.6) et en R [RDevelopmentCoreTeam 2009] ainsi que Rserve [Urbanek 2011] comme interface entre les deux. Les résultats ont été stockés dans une base de données mySQL (version 14.14 Distrib 5.1.41). Cette dernière nous a permis d'analyser une grande quantité des données à une grande vitesse en conjonction avec R. Toute cette implémentation a été faite sous Linux (Ubuntu 10.04) et nous avons utilisé des scripts bash pour coordonner tous les programmes. Les noms des logiciels développés ont été choisis arbitrairement.

2.3.1 Recherche des régions régulatrices

Le développement et l'implémentation de notre méthode ont été faits en trois étapes. Dans la première étape, nous avons développé le logiciel « AIcoalescent » qui effectue la classification des individus en AI et qui calcule le test binomial et le test génotypique de contingence 2x3. « AIcoalescent » reçoit un fichier hapmap avec l'information de génotypage de la région à analyser et un fichier d'expression du gène d'intérêt avec les mêmes individus (tableau II.I). La sortie de « AIcoalescent » est un fichier avec les résultats du test binomial et du test génotypique 2x3 ainsi que la distribution des génotypes en AI et en non-AI, entre autres. Ensuite, nous avons analysé avec « AIcoalescent » quelques gènes. La région analysée autour de chaque gène a été d'environ 200 kb. Dans la deuxième étape, nous avons développé les scripts nécessaires pour analyser un groupe de 250 gènes avec « AIcoalescent ». Nous avons

regardé une région de 1000 kb autour de chaque gène. Dans la troisième étape, nous avons amélioré « AIcoalescent » pour analyser tous les SNPs du chromosome pour chaque gène disponible dans nos fichiers de [Ge et al., 2009] (tableau II.I). Étant donné la grande quantité de données, nous avons aussi construit une base de données « genomeAI » pour garder et gérer les résultats produits par « AIcoalescent » précédemment modifié. Bien entendu, nous avons développé plusieurs autres programmes pour interagir avec « AIcoalescent » et coordonner les différentes manipulations que nous avons faites.

2.3.2 Creation d'un jeux de données pour la détermination de faux positifs.

Pour évaluer le taux de faux positifs et sa distribution, nous avons tiré au hasard 5, 10, 15, 20, 25² et 29 individus et nous les avons assignés comme des individus en AI. Cette procédure a été faite 20 fois. Ensuite, nous avons utilisé « AIcoalescent » pour tester le génome au complet pour chaque groupe en gardant les résultats dans la base de données « genomeAI ».

2.3.3 Comparaison de méthodes de LD

Pour comparer les méthodes de LD, nous avons utilisé la même méthodologie informatique que pour les régions régulatrices (2.3.1). Nous avons développé le logiciel « LDStart » qui calcule le test haplotypique de contingence 2x2, normalement

2 Ces résultats n'ont pas été gardés dans la base de données à cause de problèmes d'ordre techniques.

utilisé pour l'évaluation de LD et le test génotypique 3x3 qui étend le test de contingence 2x3 pour utiliser les 3 génotypes pour les deux sites à tester (figure 2.3). Nous avons aussi inclu l'algorithme pour calculer le test génotypique 2x3. Le « LDStart » stocke les résultats dans la base de données « genomeAI ».

A)	2x3	B/b	B/B b/b
	A/A		
	A/a		
	a/a		

B)	3x3	B/B	B/b	b/b
	A/A			
	A/a			
	a/a			

Figure 2.3: Représentation des tableaux de contingences pour évaluer le LD avec les tests génotypiques A) 2x3 et B) 3x3. Supposons que nous avons SNP A avec allèles A,a et SNP B avec allèles B,b.

Chapitre 3. Résultats

3.1 Analyses globales de données et catégorisation.

Pour chaque gène disponible dans les données de [Ge et al., 2009] et tel que de l'AI est détecté chez au moins un individu, tous les SNPs de HapMap II se trouvant sur le même chromosome que le gène ont été testés en utilisant le test binomial et le test de contingence 2x3 décrits dans la méthodologie. Dans le reste du texte, nous utiliserons simplement l'expression « test de contingence » pour parler du test de contingence 2x3. Pour gérer ces résultats, nous avons créé une grande base de données d'un terabyte d'information contenant 3800 millions de résultats de ces tests. De plus, nous avons également créé une base de données pour évaluer la distribution des faux positifs, contenant 249 millions de résultats. L'analyse des données a été faite en utilisant des filtres de plus en plus restrictifs, appliqués de façon cumulative (tableau III.I). Dans les données expérimentales de [Ge et al., 2009], 22 366 gènes ont été testés pour la présence de AI. Après l'application de l'algorithme de détermination des individus en AI, 15 818 gènes ont au moins un individu en AI, soit 70 %. Dans la littérature, nous retrouvons les mêmes proportions [Wagner et al., 2010]. Ensuite, nous avons évalué le nombre de gènes avec au moins un SNP montrant une association significative avec AI à 1 % ($p < 0,01$). Selon le test binomial, nous avons observé 4380 gènes et 11 839 dans le test de contingence. Pour éviter les SNPs solitaires qui sortent significatifs dû à la chance, nous avons sélectionné les gènes qui

ont au moins 2 SNPs à moins de 5 kb de distance avec la valeur de $p < 0,001$. Cette règle que nous appelons « 2 en 5 » permet de filtrer les erreurs de type I plus efficacement. Le test de contingence n'a pas été beaucoup affecté (seulement 13 % de diminution). Par contre, dans le test binomial, 59 % des gènes ne passent pas ce filtre. Cela montre que le nombre minimum de SNPs et la distance entre eux affectent différemment les deux tests. Nous voyons que la plupart des SNPs les plus significatifs, sortent déjà en « clusters » pour le test de contingence. L'autre paramètre que nous avons évalué est le nombre d'individus en AI. Quand nous avons ajouté la restriction d'avoir au moins 8 individus en AI aux filtres précédents, le nombre de gènes pour lesquels le test binomial a identifié des SNPs significatifs est resté égal, alors que celui pour le test de contingence a diminué de plus de 50 %. Cependant, si nous regardons l'intersection entre les deux groupes, nous voyons que 99,49 % des gènes dans le groupe du test binomial se retrouve dans celui du test de contingence.

La figure 3.1 présente la comparaison entre la distribution de la valeur de p la plus significative par gène du test binomial et celle du test de contingence en fonction du nombre d'individus en AI. Les données de la figure 3.1 et 3.2 ont été filtrées parallèlement en appliquant le filtre « 2 en 5 » et elles ne montrent que le SNP le plus significatif par gène. Premièrement, nous observons qu'il faut au moins 11 individus en AI pour passer le filtre « 2 en 5 » pour le test binomial. Par contre, pour le test de contingence, seuls deux individus en AI sont nécessaires. Aussi, les valeurs de p obtenues avec le test de contingence sont plus significatives. Par exemple, avec 20 individus en AI, nous avons obtenu une valeur de p minimale de 10^{-14} pour le test de

contingence, contre un p minimal de 10^{-10} pour le test binomial (figure 3.1). Deuxièmement, pour ce qui est des faux positifs, tous nombres d'individus en AI confondus, les p minimaux pour le test de contingence et le test binomial sont 10^{-8} et 10^{-6} respectivement.

La figure 3.2 présente la distribution de ces mêmes valeurs de p en fonction de la distance entre le SNP le plus significatif et le site d'initiation de la transcription du gène régulé. Nous constatons que les SNPs avec les valeurs de p les plus significatives, se trouvent près de leur gène. Si nous prenons la section agrandie de chaque distribution, nous constatons que les SNPs avec les valeurs minimales tombent à l'intérieur de 200 à 300 kb. Cette distribution a aussi été constatée par [Idaghdour et al., 2010].

Tableau III.I : Analyse globale de l'ensemble des gènes disponibles dans les données d'expression différentielle de Ge. et al. 2009 avec les données de génotypage de HapMap II. Chaque colonne contient le nombre de gènes résultant de l'utilisation de filtres progressivement restrictifs appliqués de façon cumulative.

Nombre de gènes	Total	Au moins 1 Individu en AI	Au moins 1 SNPs avec p < 0,01	Au moins 2 SNPs avec p < 0,001 à moins de 5 kb de distance	Au moins 8 individus en AI
Test binomial	22366	15818	4380	1796	1796
% / colonne précédente.		71	28	41	100
% / total.		71	20	8	8
Test de contingence 2x3	22366	15818	11839	10367	4831
% / colonne précédente.		71	75	88	47
% / total.		71	53	46	22
Test de contingence 2x3 et binomial	22366	15818	4380	1787	1787
% / colonne précédente.		71	28	41	100
% / total.		71	20	8	8

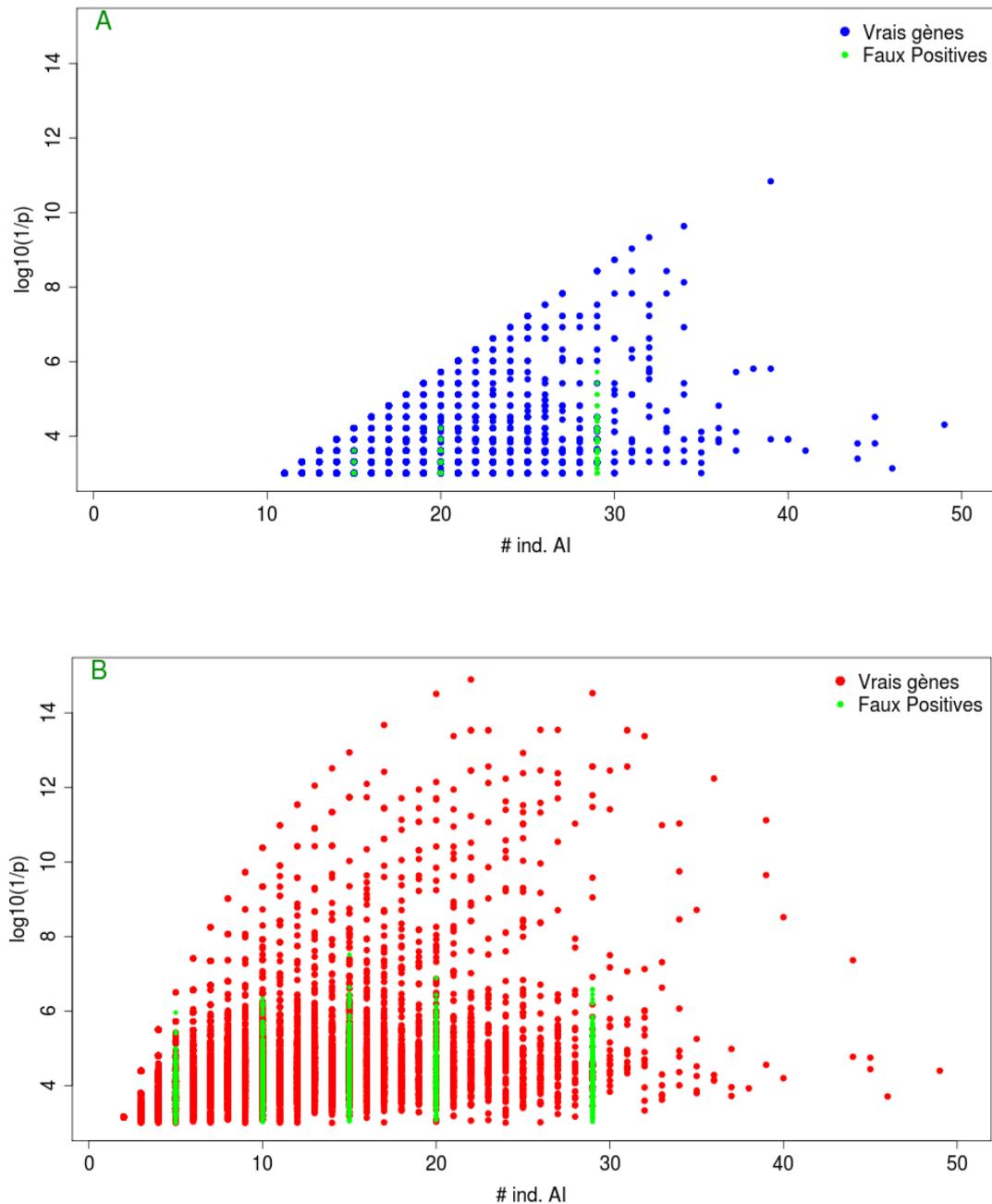


Figure 3.1: Distribution des valeurs de p du test binomial (A) et du test de contingence 2×3 (B) en fonction du nombre d'individus en AI. Pour chaque gène des données de Ge et al(2009), on a testé l'ensemble des SNPs de HapMap II résidant sur le même chromosome que le gène analysé. Seuls les SNPs avec un $p < 0,001$ et ayant au moins un autre SNP avec un $p < 0,001$ à moins de 5 Kb, ont été gardés. Les points bleus (test binomial) et rouge (test de contingence 2×3) représentent la valeur de p la plus significative de ces SNPs pour chaque gène. Tels qu'indiqué dans la méthodologie, les SNPs de HapMap II de tout le génome ont été testés pour chaque simulation avec 10, 15, 20 et 29 individus choisis et mis aléatoirement en AI. Les points verts, correspondant à des faux-positifs, représentent les mêmes résultats pour les données simulées par chromosome et par simulation.

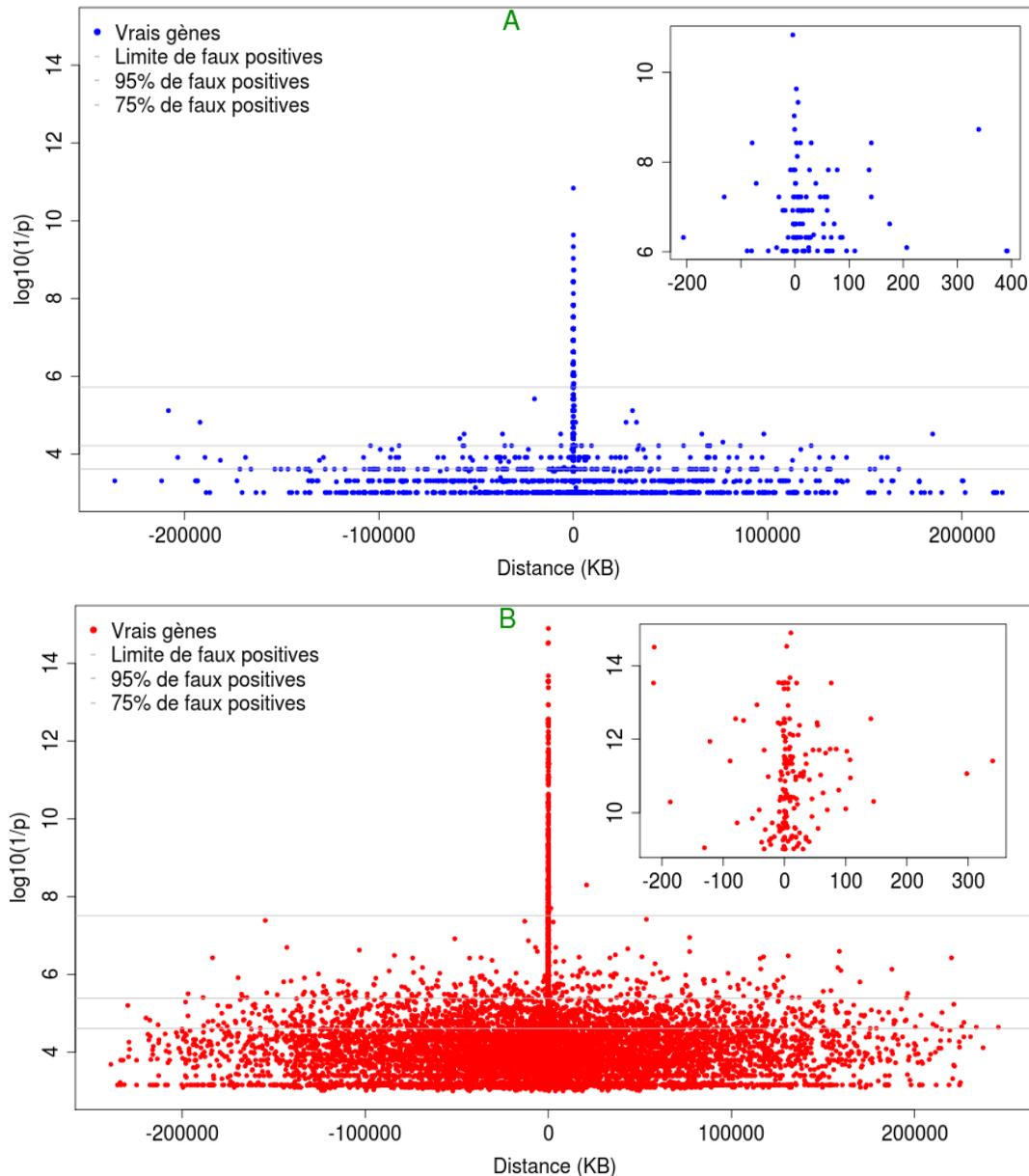


Figure 3.2: Distribution de des valeurs de p du test binomial (A) et du test de contingence 2×3 (B) en fonction de la distance au gène (site d'initiation de la transcription). Pour chaque gène des données de Ge et al(2009), on a testé les SNPs de HapMap II résidant sur le même chromosome que le gène analysé. Seuls les SNPs avec un $p < 0,001$ et ayant au moins un autre SNP avec un $p < 0,001$ à moins de 5 Kb, ont été gardés. Les points bleus (test binomial) et rouge (test de contingence 2×3) représentent la valeur de p la plus significative de ces SNPs pour chaque gène. Tels qu'indiqué dans la méthodologie, les SNPs de HapMap II de tout le génome ont été testés pour chaque simulation avec 10, 15, 20 et 29 individus choisis et mis aléatoirement en AI. Les résultats significatifs des simulations, correspondant à des faux-positifs, sont représentés par les lignes grises qui indiquent, de la plus haute à la plus basse, le maximum, le top 5% et le top 25%. La figure intégrée est une amplification du pic de la distribution.

3.2 Application des tests à travers des exemples.

La figure 3.3 montre les résultats du test de contingence pour le gène *LRRIQ3* à l'échelle génomique (tous les chromosomes autosomaux). Le pic sur le chromosome 1 représente une région autour du gène. Ce pic contient 56 SNPs étalés sur une distance de 230 kb et ayant un $p < 10^{-6}$. Un deuxième pic, visiblement élevé est observé sur le chromosome 15. Cependant, il ne représente qu'un seul SNP. Les résultats sur le chromosome 6 du test binomial et du test de contingence pour le gène *ALDH8A1* sont présentés dans la figure 3.4. Les résultats de faux positifs sur le même chromosome et avec le même nombre d'individus en AI sont également présentés. Clairement, les pics dans la figure 3.4 A et C révèlent la même région régulatrice candidate qui n'apparaît pas dans les faux positifs (figure 3.4 B et D). La figure 3.5 représente cette région en détail et nous pouvons constater qu'elle représente un seul bloc de LD (figure 3.5 C). De plus, 29 des SNPs qui sortent significatifs dans notre analyse ont déjà été relevés dans des études précédentes. (Tableau III.IV)[Myers et al., 2007] [Zeller et al., 2010] [Stranger et al., 2007]. La figure 3.6 montre le même genre de résultats que la figure 3.5 mais pour le gène *AHI1*. Nous y observons aussi une région régulatrice candidate contenant 13 SNPs qui ont également été détectés dans des études précédentes (tableau III.IV) [Veyrieras et al., 2008] [Zeller et al., 2010]. Dans ces deux figures (3.5 et 3.6), nous retrouvons la même région génomique de 1500 kb sur le chromosome 6 et une région bien définie pour chaque gène. Cela montre la spécificité de notre approche pour détecter des régions candidates uniques. Dans la figure 3.7, nous voyons les résultats de deux tests sur un

segment de 1500kb du chromosome 8 pour le gène XKR9. Le pic autour du gène représente une région régulatrice candidate unique. En regardant de plus près la région autour de XKR9, nous observons que tous les SNPs significatifs sont dans un bloc de LD d'environ 500 kb. Cependant, les SNPs les plus significatifs pour le test de contingence 2x3 sortent à 200 kb de distance du gène et s'étalent sur environ 75 kb. Dans la région de 500 kb, nous trouvons 169 eQTLs qui ont été rapportés par des études précédentes (tableau III.IV) [Veyrieras et al., 2008]. Cela confirme une bonne performance de notre approche pour détecter la région régulatrice.

Afin d'illustrer un autre aspect de notre méthode, les figure 3.8 et 3.9 montrent la cartographie des régions régulatrices pour les gènes QDPR et NEFH respectivement. Dans le premier cas, nous voyons un bloc de SNPs significatifs pour le test de contingence (fig. 3.8 B) à environ 850 kb du gène, alors que dans le test binomial (fig. 3.8 A), aucun site ne sort significatif. Le triangle de LD (figure 3.8 C) montre un grand fractionnement de la région qui sépare le gène de la région régulatrice candidate. Ce fractionnement s'explique également par un haut taux de recombinaison observé tel que présenté en bas de la figure 3.8 C. Dans le deuxième cas, nous identifions une région candidate également à 1000 kb de distance (fig. 3.9 A et B) avec un signal concordant pour les deux tests, quoique plus faible pour le test binomial. Le triangle de LD (fig. 3.9 C) montre un plus faible taux de recombinaison avec un seul fractionnement.

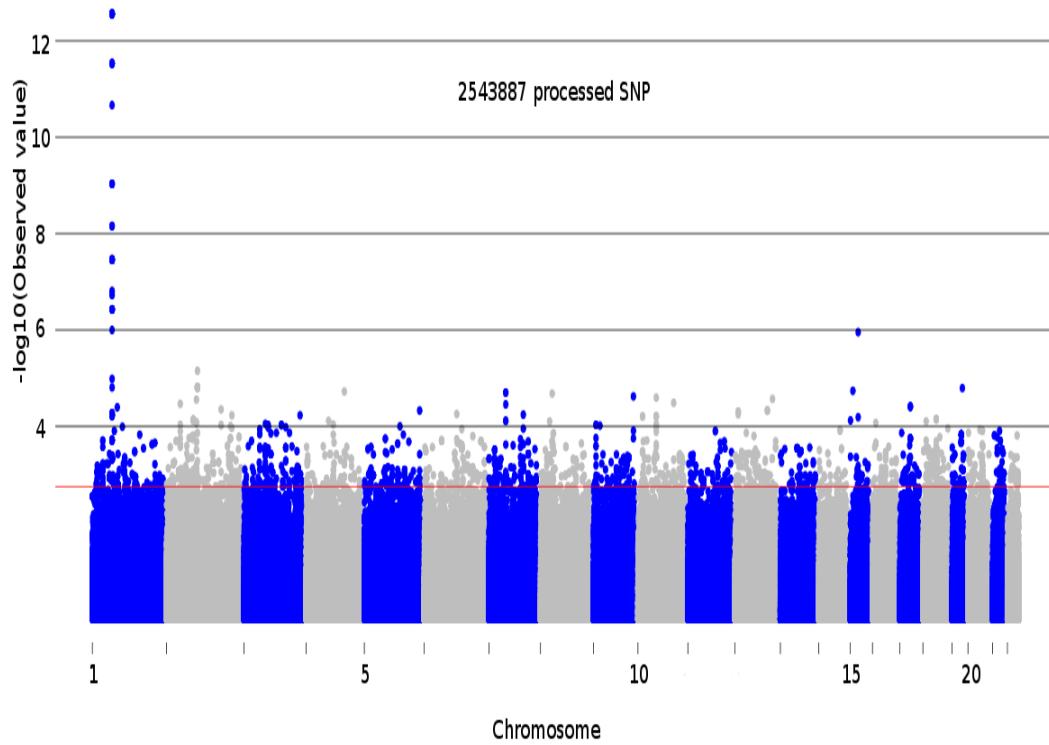


Figure 3.3: Graphique de valeurs de p du test de contingence 2x3 pour le gène LRR1Q3. Tous les autosomes ont été testés en utilisant les données de HapMap II et de Ge et al. 2009. 29 individus en AI.

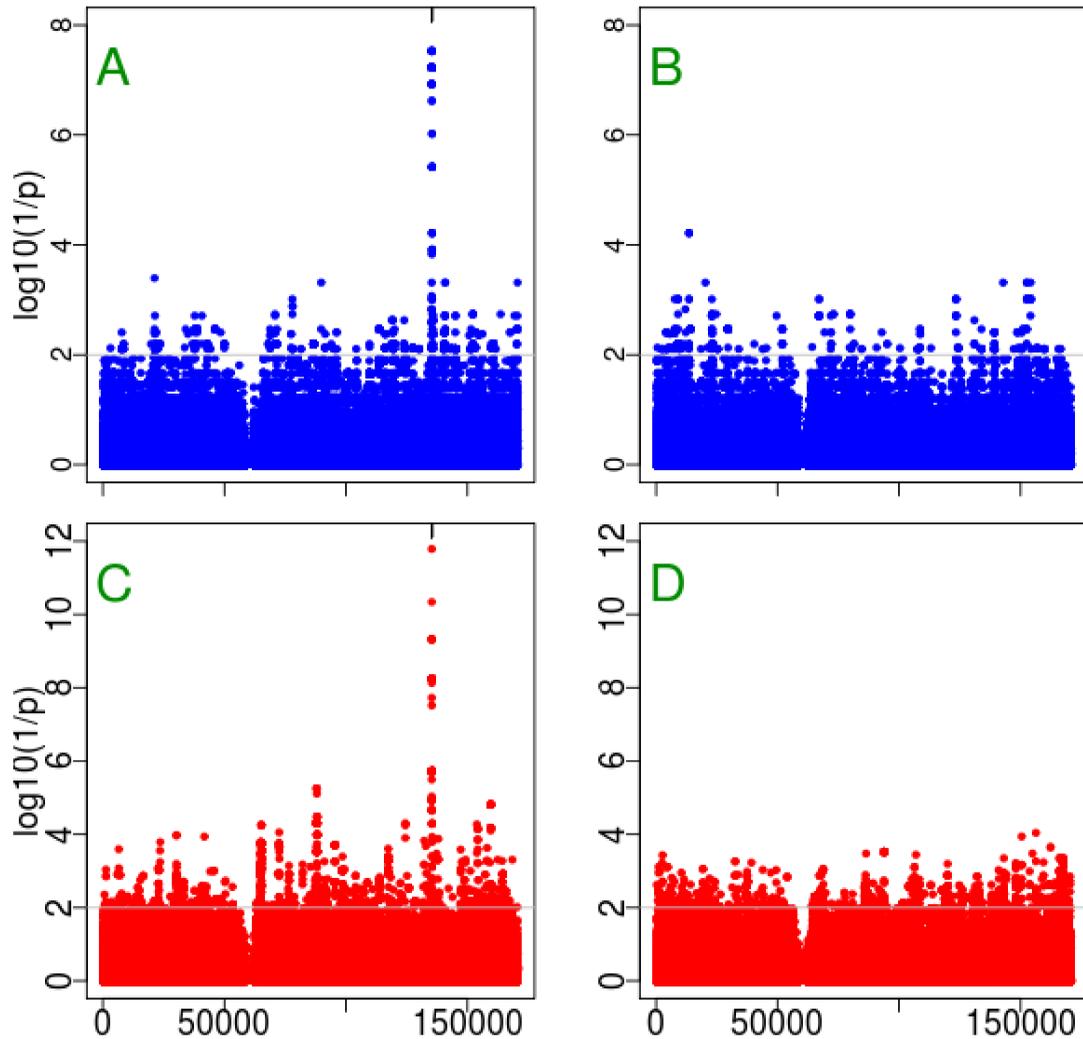


Figure 3.4: Cartographie du site de régulation du gène *ALDH8A1* sur le chromosome 6 et faux positifs sur le même chromosome. Valeurs de p du test binomial (A) et du test de contingence 2x3 (C) appliqués aux SNPs de HapMap II du chromosome 6 avec les données de AI du gène *ALDH8A1* de Ge et al.(2009). Les traits noirs représentent la position des marqueurs utilisés pour ce gène. Pour les mêmes SNPs et choisissant aléatoirement le même nombre d'individus en AI (29), les résultats du test binomial (B) et du test de contingence 2x3 (D) représentent les faux positifs.

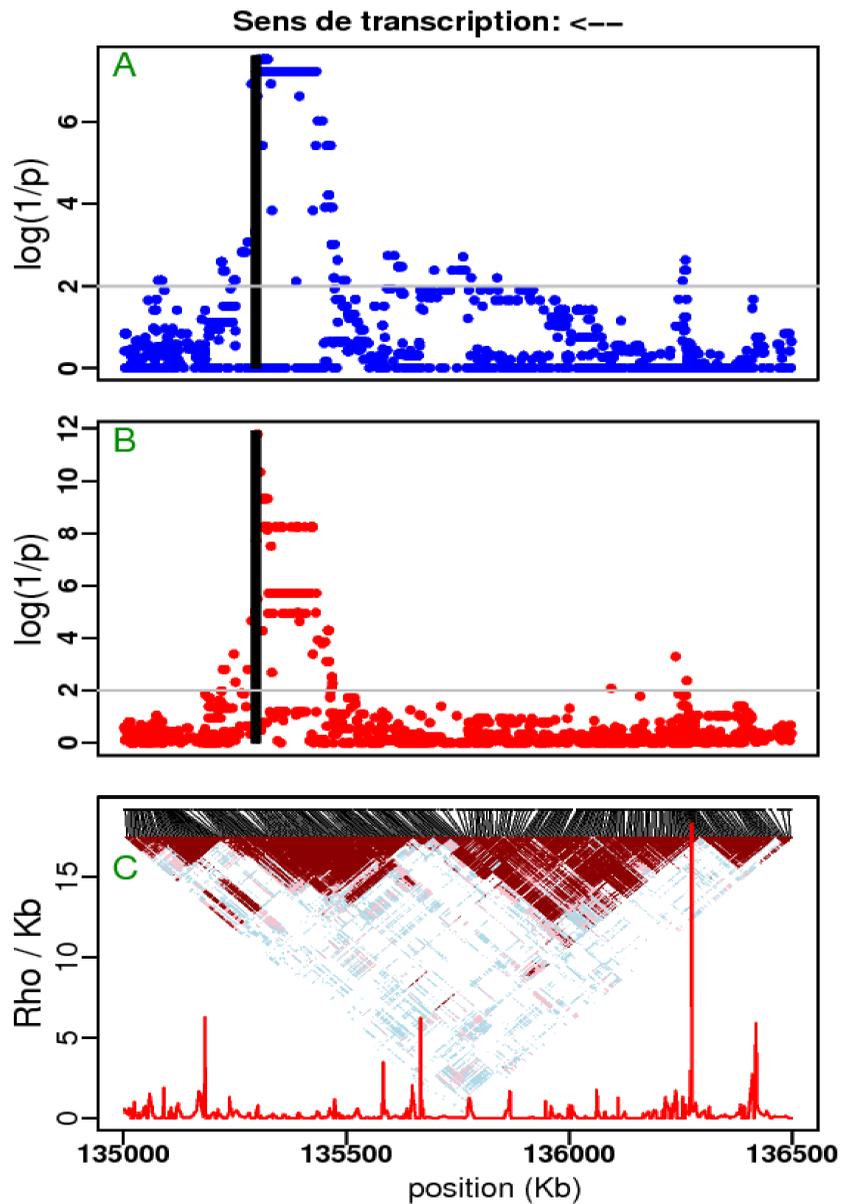


Figure 3.5 : Cartographie du site de régulation du gène ALDH8A1 (chromosome 6). Résultats des tests binomial (A) et de contingence 2x3 (B) utilisant les SNPs de HapMap II avec les données de AI de Ge et al.(2009). Les lignes verticales représentent les marqueurs utilisés pour déterminer le AI. Le triangle de LD calculé par le test exact de Fisher 2x2 (rouge: $p < 0,001$; rose: $p < 0,01$; bleu: $p < 0,05$; blanc: $p \geq 0,05$) et le taux de recombinaison calculé le long du segment avec InfRec [Lefebvre and Labuda, 2008], sont en (C). Les positions des SNPs dans le triangle sont données par les lignes obliques. 29 individus en AI.

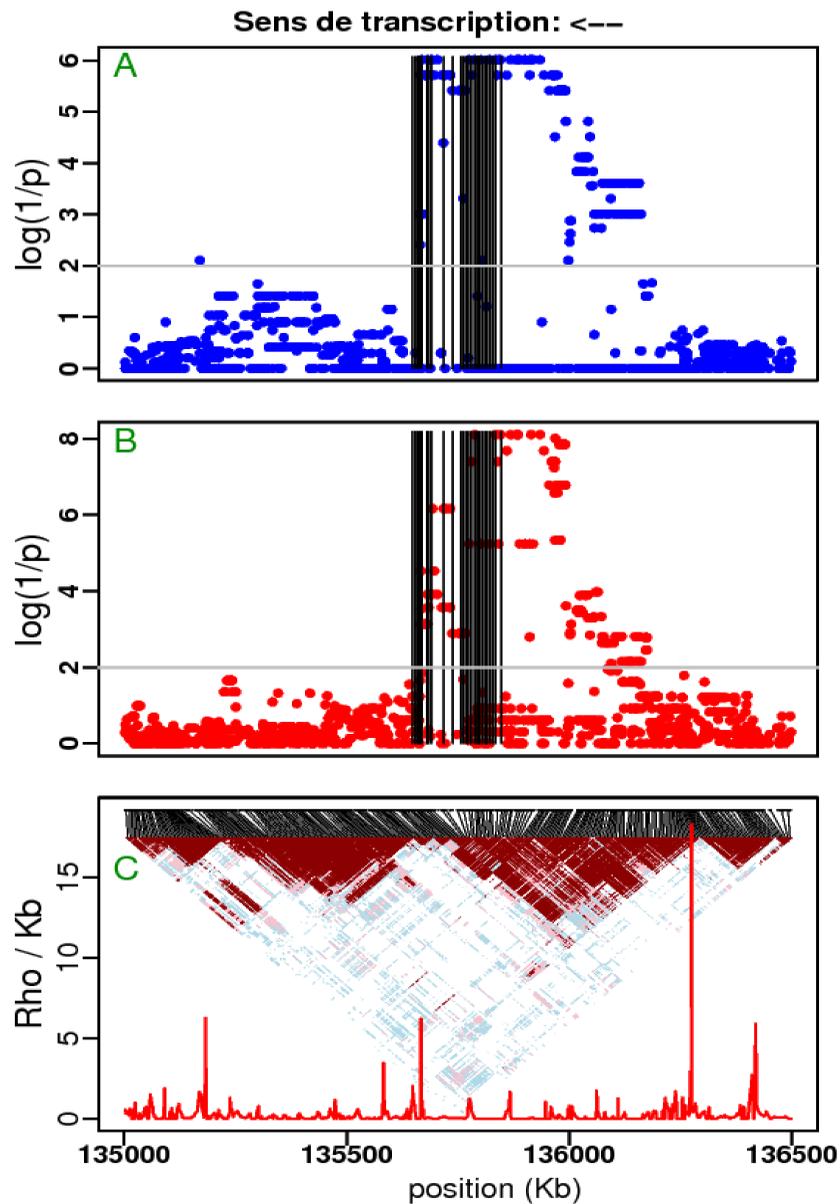


Figure 3.6 : Cartographie du site de régulation du gène AHI1 (chromosome 6). Résultats des tests binomial (A) et de contingence 2x3 (B) utilisant les SNPs de HapMap II avec les données de AI de Ge et al.(2009). Les lignes verticales représentent les marqueurs utilisés pour déterminer le AI. Le triangle de LD calculé par le test exact de Fisher 2x2 (rouge: $p < 0,001$; rose: $p < 0,01$; bleu: $p < 0,05$; blanc: $p > = 0,05$) et le taux de recombinaison calculé le long du segment avec InfRec [Lefebvre and Labuda, 2008], sont en (C). Les positions des SNPs dans le triangle sont données par les lignes obliques. 24 individus en AI.

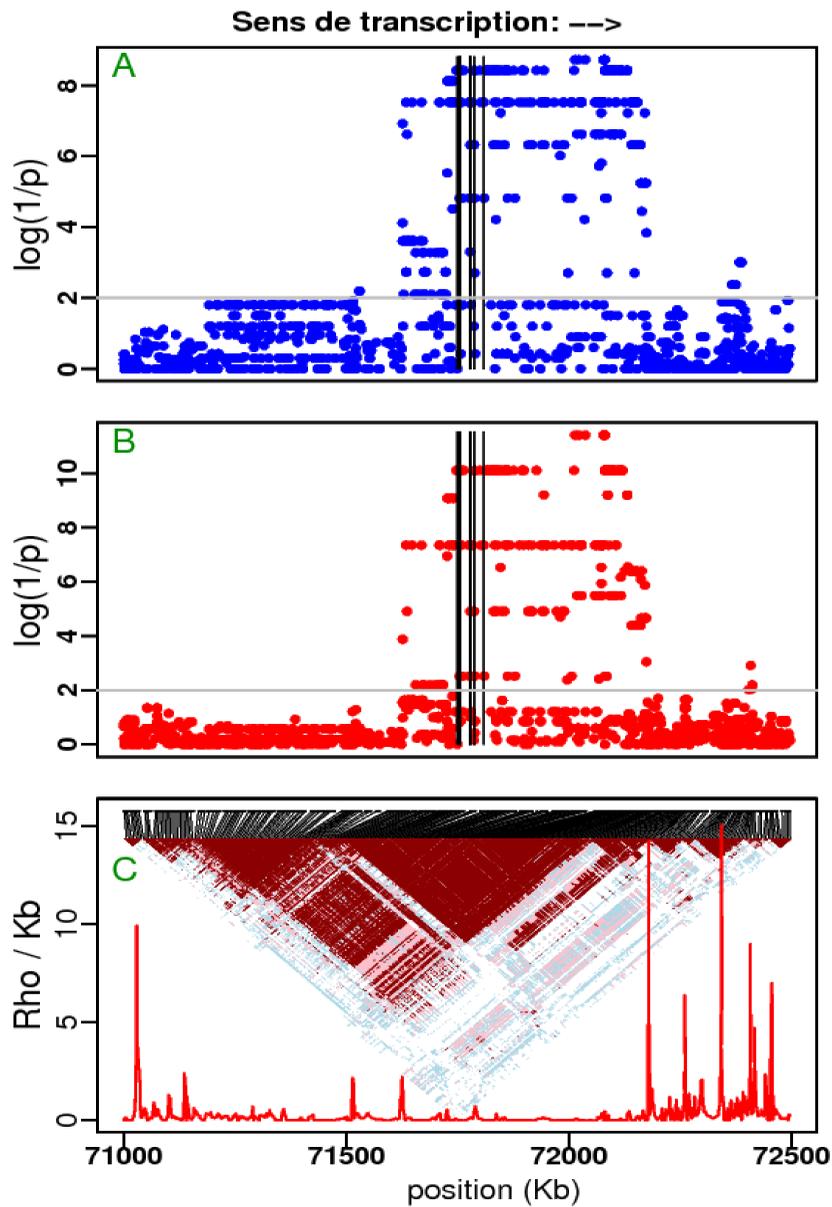


Figure 3.7: Cartographie du site de régulation du gène XKR9 (chromosome 8). Résultats des tests binomial (A) et de contingence 2x3 (B) utilisant les SNPs de HapMap II avec les données de AI de Ge et al.(2009). Les lignes verticales représentent les marqueurs utilisés pour déterminer le AI. Le triangle de LD calculé par le test exact de Fisher 2x2 (rouge: $p < 0,001$; rose: $p < 0,01$; bleu: $p < 0,05$; blanc: $p \geq 0,05$) et le taux de recombinaison calculé le long du segment avec InfRec [Lefebvre and Labuda, 2008], sont en (C). Les positions des SNPs dans le triangle sont données par les lignes obliques. 30 individus en AI.

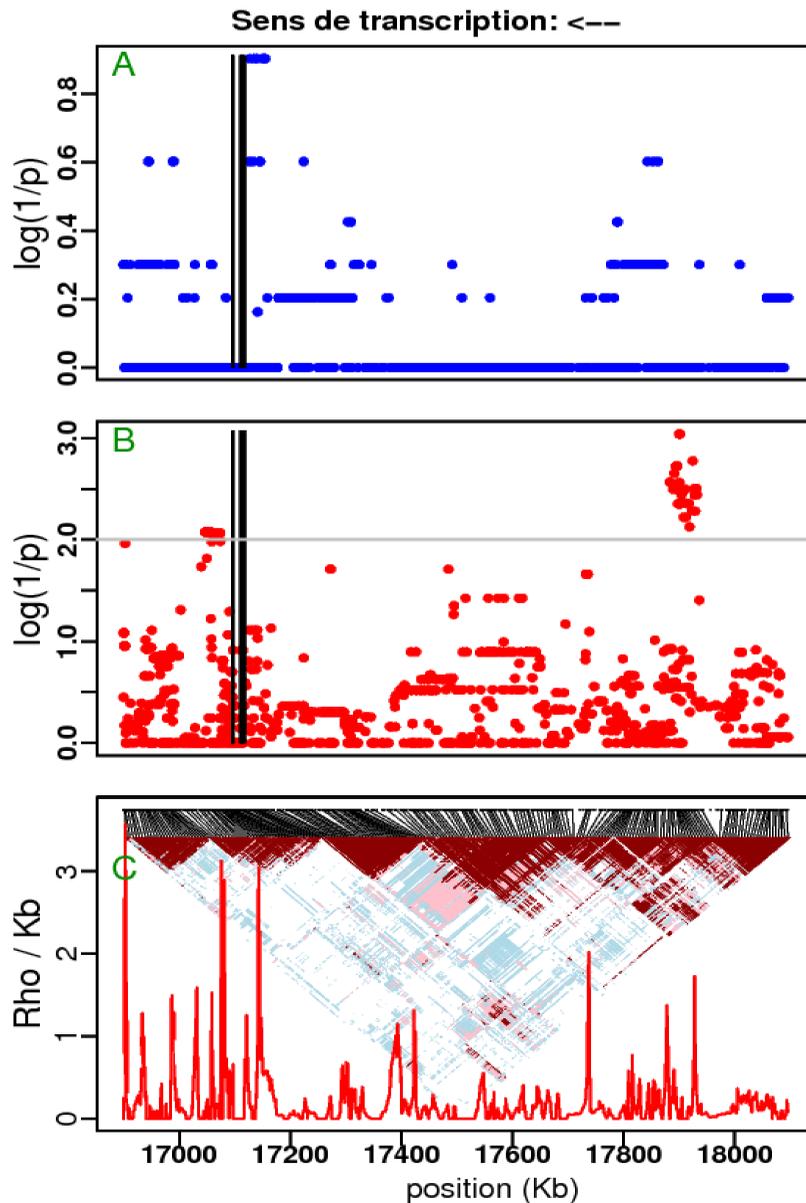


Figure 3.8: Cartographie du site de régulation du gène QDPR (chromosome 4). Résultats des tests binomial (A) et de contingence 2x3 (B) utilisant les SNPs de HapMap II avec les données de AI de Ge et al.(2009). Les lignes verticales représentent les marqueurs utilisés pour déterminer le AI. Le triangle de LD calculé par le test exact de Fisher 2x2 (rouge: $p < 0,001$; rose: $p < 0,01$; bleu: $p < 0,05$; blanc: $p \geq 0,05$) et le taux de recombinaison calculé le long du segment avec InfRec [Lefebvre and Labuda, 2008], sont en (C). Les positions des SNPs dans le triangle sont données par les lignes obliques. 8 individus en AI.

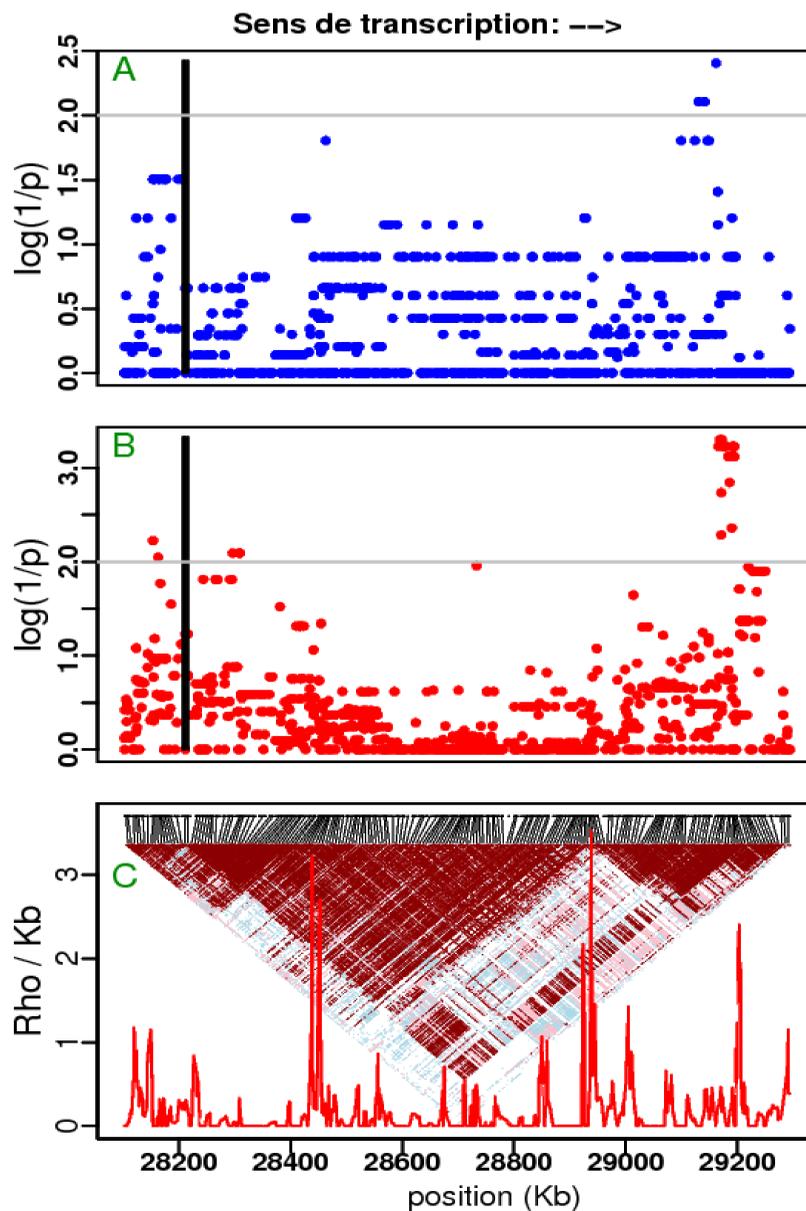


Figure 3.9: Cartographie du site de régulation du gène NEFH (chromosome 22). Résultats des tests binomial (A) et de contingence 2x3 (B) utilisant les SNPs de HapMap II avec les données de AI de Ge et al.(2009). Les lignes verticales représentent les marqueurs utilisés pour déterminer le AI. Le triangle de LD calculé par le test exact de Fisher 2x2 (rouge: $p < 0,001$; rose: $p < 0,01$; bleu: $p < 0,05$; blanc: $p >= 0,05$) et le taux de recombinaison calculé le long du segment avec InfRec [Lefebvre and Labuda, 2008], sont en (C). Les positions des SNPs dans le triangle sont données par les lignes obliques. 13 individus en AI.

3.3 Erreur empirique.

Dans le tableau III.II, nous voyons le résultat de l'estimation de l'erreur empirique obtenu en mettant aléatoirement 10,15,20 et 25 individus en AI pour ensuite tester le génome au complet avec chaque groupe. Cette procédure a été répétée 20 fois. En général, le taux de faux positifs ne dépasse pas 1 %. De plus, les valeurs de p les plus significatives sont d'environ 10^{-8} et il s'agit de SNPs solitaires la plupart du temps. Fait à noter, le taux de faux positifs augmente 20 fois plus entre 10 et 25 individus en AI pour le test binomial, alors qu'il reste constant pour le test de contingence.

Tableau III.II: Estimation de l'erreur empirique. Pourcentage de SNPs du génome au complet qui montrent des valeurs de p plus petits que 0,01 après avoir assigné au hasard 10, 15, 20 et 25 individus en AI 20 fois.

AI individual	10	15	20	25
Binomial	0,01%	0,09%	0,16%	0,21%
Contingency 2x3	0,65%	0,69%	0,73%	0,73%

3.4 Correspondance avec la littérature.

Pour étudier la correspondance avec la littérature, nous avons utilisé une base de données publiée dans le site du laboratoire du Dr Pritchard de l'Université de Chicago (eQTL resources @ the pritchard lab, <http://eqtl.uchicago.edu/Home.html>). Cette base de données que nous appellerons « BD eQTL » contient des eQTLs qui ont été détectés dans une ou plusieurs études pour différents gènes. Le tableau III.III compile les gènes dont le AI a été démontré à la fois dans « BD eQTL » et dans cette étude. Nous avons appliqué le même protocole de filtrage au gènes de « BD eQTL » que pour l'analyse globale (tableau III.I) afin d'utiliser la même mesure de comparaison. Du total de gènes de « BD eQTL », 9 % pour le test binomial et 22 % pour le test de contingence passent le plus haut niveau de filtrage. Ces deux pourcentages sont équivalents à ceux présentés dans le tableau III.I. Le tableau III.IV dresse une liste résumée qui contient pour chaque gène le nombre d'eQTLs signalés à la fois dans la « BD eQTL » et dans notre étude pour le test de contingence. Dans l'annexe I, il y a un tableau détaillé avec l'identification des SNPs et l'étude.

Tableau III.III: Analyse de l'ensemble des gènes qui ont été détectés par des études précédentes (base de donnée de eQTLs du laboratoire du Dr Pritchard , Université de Chicago) et dans cette étude pour chaque filtre progressivement restrictif appliqué de façon cumulative.

Nombre de gènes	Total	Au moins 1 Individu en AI	Au moins 1 SNPs avec p < 0,01	Au moins 2 SNPs avec p < 0,001 à moins de 5 kb de distance	Au moins 8 individus en AI
Test binomial	6052	4310	1236	552	552
% / colonne précédente.		71	29	45	100
% / total.		71	20	9	9
Test de contingence 2x3	6052	4310	3159	2727	1346
% / colonne précédente.		71	73	86	49
% / total.		71	52	45	22

Tableau III.IV: Liste de gènes avec le nombre d'eQTLs qui ont été détectés dans des études précédentes (base de donnée de eQTLs du laboratoire du Dr Pritchard , Université de Chicago) et dans cette étude pour le même gène. (Test de contingence 2x3)

Gène	eqtls	Gène	eqtls
XKR9-chr8	169	FLJ45422-chr6	8
BTN3A2-chr6	90	GNMT-chr6	8
HEBP2-chr6	73	LILRA2-chr19	7
TSGA10-chr2	69	STOX1-chr10	7
CHURC1-chr14	59	BLOC1S2-chr10	6
CP110-chr16	40	RPS23-chr5	6
RIBC2-chr22	33	C9orf128-chr9	5
ALDH8A1-chr6	29	CLIC6-chr21	5
DTNB-chr2	24	ECHDC2-chr1	5
GUCA1B-chr6	24	VNN3-chr6	5
MPHOSPH6-chr16	24	C7orf25-chr7	3
WBSCR27-chr7	24	ZNF502-chr3	3
FLJ22374-chr7	22	ANKRD42-chr11	2
CHI3L2-chr1	21	APIP-chr11	2
ZFP57-chr6	17	C21orf63-chr21	2
NT5E-chr6	16	GBP3-chr1	2
TUBB6-chr18	16	PGBD4-chr15	2
RLN1-chr9	14	SEPT10-chr2	2
TBKBP1-chr17	14	SLC5A11-chr16	2
AHI1-chr6	13	C16orf88-chr16	1
GGTA1-chr9	13	CATSPER2-chr15	1
LOC400642-chr18	13	CCL14-chr17	1
TGM5-chr15	13	DHX58-chr17	1
TLR1-chr4	13	FAM164A-chr8	1
DPYSL4-chr10	11	GCSH-chr16	1
APOC2-chr19	10	SLFN5-chr17	1
SIGLEC12-chr19	10	VENTX-chr10	1
ANKDD1A-chr15	9	WARS-chr14	1
DNASE1L3-chr3	9	XRRA1-chr11	1
C13orf18-chr13	8	ZNF391-chr6	1

3.5 Tests génotypiques pour la détection de LD.

Nous avons réalisé que le test de contingence 2x3 (LD2x3) peut aussi servir à détecter le LD entre deux SNPs sans connaître la phase. Nous avons aussi étendu ce test de 2x3 jusqu'au 3x3 (LD3x3) permettant d'exploiter les trois génotypes pour les deux sites entre les individus testés. Ces deux tests, LD2x3 et LD3x3, ne nécessitent pas d'information haplotypique, seulement génotypique. Nous avons comparé ces deux tests avec le test de contingence 2x2 (LD2x2), habituellement utilisé pour la détection de LD [Hudson 2001]. L'évaluation du LD a été faite sur des données simulées et sur les données de HapMap II. Nous avons fait quatre simulations avec le programme ms [Hudson 2002] de 100 séquences de 100 kb avec $\Theta = 100$ (taux de mutation) et $\rho = 50$ (taux de recombinaison). La figure 3.10 et 3.11 présentent la corrélation des valeurs de p entre le test haplotypique (LD2x2) et les deux tests génotypiques (LD2x3 et LD3x3) respectivement, pour des données simulées. La régression linéaire ($r^2 = 0,8493$; $r^2 = 0,9429$; $p < 2.2e^{-16}$) de chaque corrélation montre l'existence et la force de l'association qui existe entre les deux types de tests. Cependant, nous voyons clairement que le test LD3x3 a une efficacité supérieure comparé au test LD2x3. La figure 3.12 montre la comparaison entre les trois tests appliqués à des données réelles de HapMap II le long d'un segment du chromosome 5. Dans les trois cas, les mêmes blocs de LD sont distinguables avec différentes intensités, la plus basse étant pour LD2x3. De plus, les triangles de LD2x2 et LD3x3 sont presque identiques. Le tableau III.V présente la comparaison analytique des trois tests pour la même région génomique à différents p critiques. Le test LD2x2 donne 48% de

paires testés en LD, le test LD3x3 donne 3 % de moins et le test LD2x3 donne 12 % de moins à un p critique de 0,01. Cet écart augmente avec la diminution du p critique. Par contre, presque 100 % des paires détectées par le test LD2x3 et LD3x3 ont été détecté avec le test LD2x2. Cela montre les différents degrés d'efficacité et un taux de faux positifs presque nul dans les tests génotypiques.

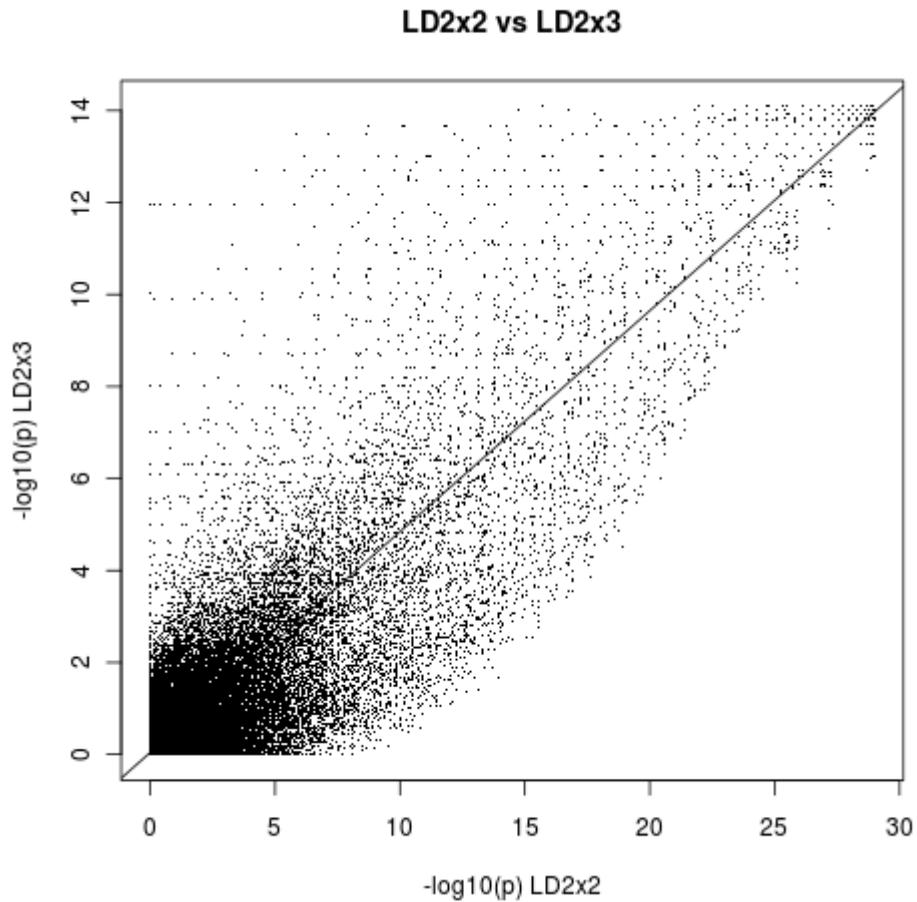


Figure 3.10: Corrélation des valeurs de p entre le test de contingence 2x2 haplotypique et le test de contingence 2x3 génotypique dans l'évaluation de LD. Les données ont été simulées avec le logiciel ms [Hudson 2001]. (4 simulations de 100 séquences de 100 kb. $\Theta = 100$ et $\rho = 50$. La ligne noire représente la régression linéaire : $r^2 = 0.8493$, $p < 2.2e^{-16}$)

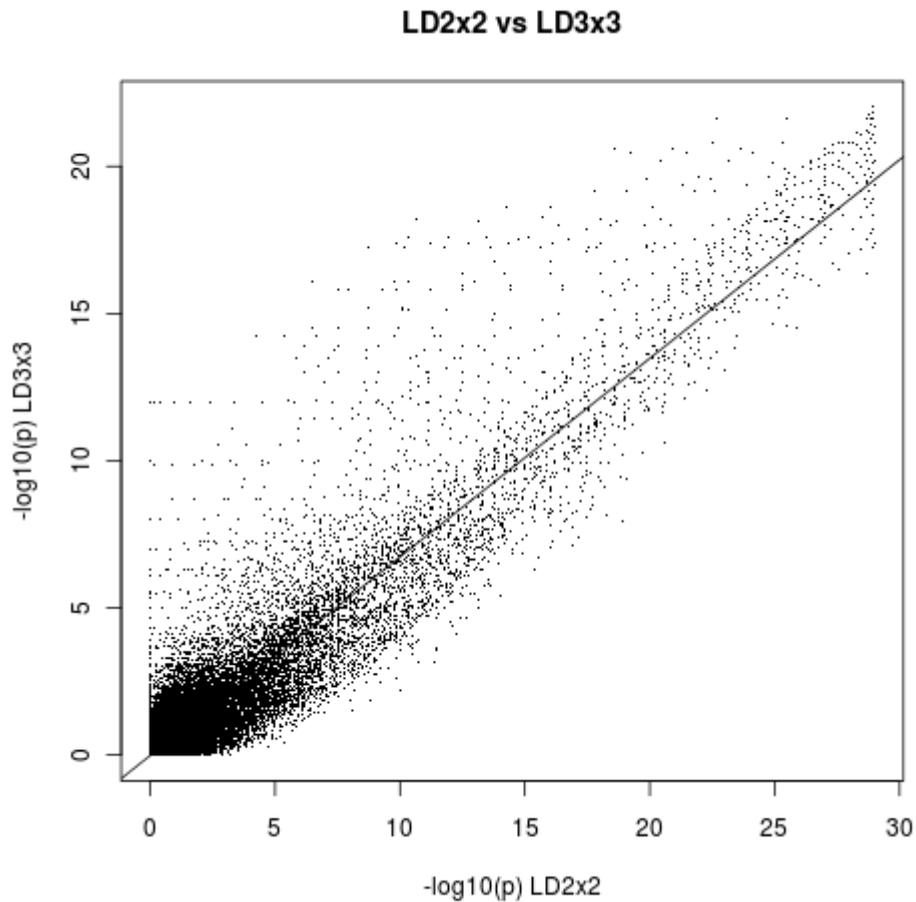


Figure 3.11: Corrélation des valeurs de p entre le test de contingence 2x2 haplotypique et le test de contingence 3x3 génotypique dans l'évaluation de LD. Les données ont été simulées avec le logiciel ms[Hudson 2001]. (4 simulations de 100 séquences de 100 kb. $\Theta = 100$ et $\rho = 50$. La ligne noire représente la régression linéaire : $r^2 = 0.9429$, $p < 2.2e^{-16}$)

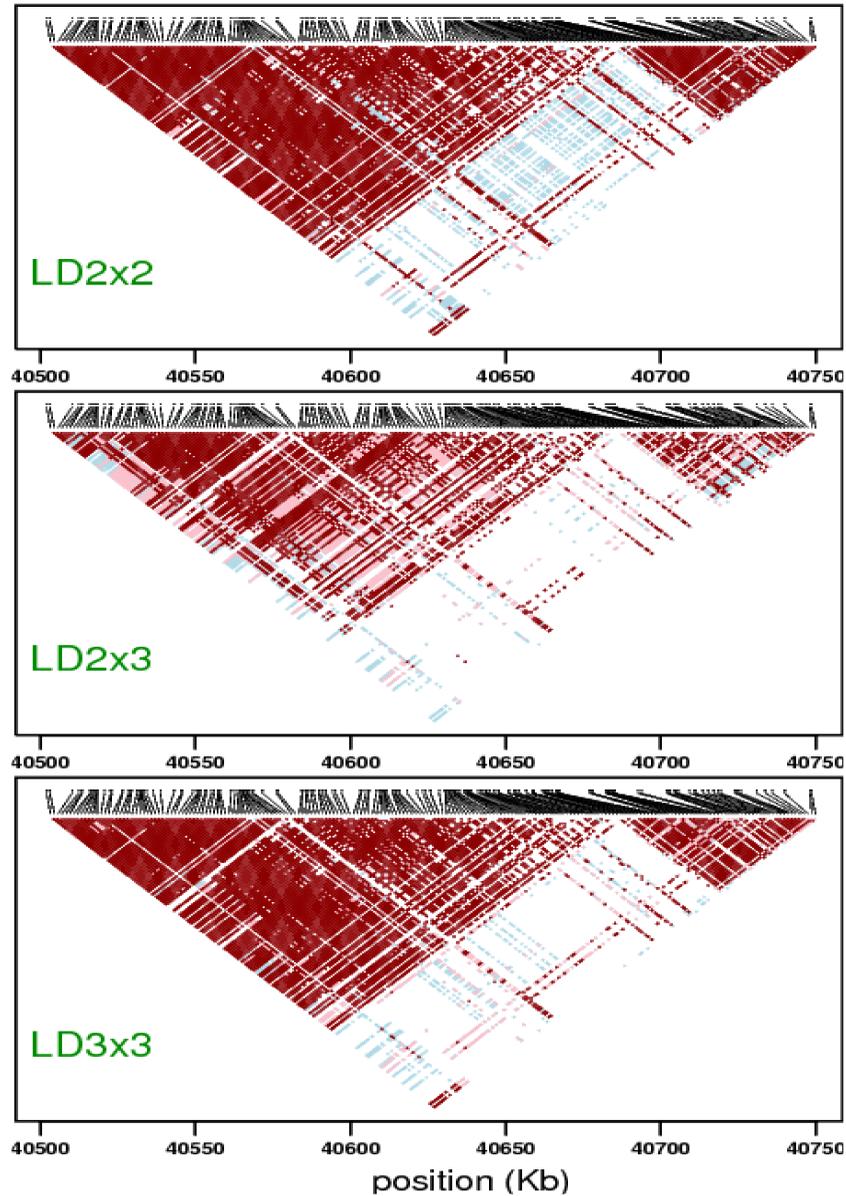


Figure 3.12: Comparaison des tests LD sur des données réelles. Triangles de LD d'un segment génomique sur le chromosome 5 calculés avec le test de contingence 2x2 (LD2x2), le test de contingence 2x3 (LD2x3) et le test de contingence 3x3 (LD3x3). Les données proviennent de HapMap II. Le rouge indique $p < 0,001$, le rose $p < 0,01$ et le vert $p < 0,05$.

Tableau III.V. Comparaison de la performance des tests de LD pour un segment de 250 kb sur le chromosome 5 (HapMap II) entre le test haplotypique de contingence 2x2 (traditionnel), le test génotypique de contingence 2x3 et le test génotypique de contingence 3x3. Différents p-critique ont été utilisés pour comparer l'exactitude entre les tests. Total de paires testés est 19701.

	p-critique	0,01	0,001	0,0001
LD2x2		9 523	9 109	8 458
LD2x3		7 162	4 994	4 281
LD3x3		8 876	8 095	7 313
LD2x2 et LD2x3		7 084	4 993	4 280
LD2x2 et LD3x3		8 711	8 093	7 309

En pourcentage (%)

	p-critique	0,01	0,001	0,0001
% LD2x2 / Total		48	46	43
% LD2x3 / Total		36	25	22
% LD3x3 / Total		45	41	37
% (LD2x2 et LD3x3) / LD3x3		98	100	100
% (LD2x2 et LD2x3) / L2x3		99	100	100

Chapitre 4. Discussion et conclusion

Les variations génétiques qui causent l'expression différentielle des transcrits alléliques sont très répandues [Lee et al., 2006]. La grande quantité de données génotypiques plus les données de AI qui commencent à être disponibles exigent d'avoir des méthodes rapides et efficaces pour identifier ces variations. De plus, l'énorme masse de résultats à gérer et comparer oblige l'utilisation de bases de données informatiques. Ainsi, notre approche utilise un système de filtrage pour assurer la qualité des données de AI, un test génotypique pour cartographier la région régulatrice sans connaître la phase, un test haplotypique complémentaire et une base de données pour gérer l'ensemble des données et des résultats.

À l'issue de nos premières analyses, nous avons constaté qu'un grand nombre de loci présentent au moins un individu en AI. Ceci a été confirmé par [Wagner et al., 2010] qui a observé le même phénomène. Cependant, les données expérimentales de AI n'étant pas parfaites, nous avons développé un système de filtres pour contrôler la qualité de nos résultats. Dans l'analyse globale (tableau III.I), nous constatons que nous perdons beaucoup moins de gènes à chaque étape de filtrage avec le test de contingence qu'avec le test binomial. Aussi, après chaque filtre, le sous-ensemble de gènes détectés par le test binomial se retrouve dans celui des gènes identifiés par le test de contingence. En appliquant la règle de « 2 en 5 », nous gardons la moitié des gènes suite à l'utilisation du test de contingence contre un dixième après le test binomial. De plus, si nous comparons le nombre de gènes avant et après l'application

de la règle « 2 en 5 », nous voyons que ce filtre n'a pas beaucoup d'effet sur la quantité de gènes retenus avec le test de contingence. Ce n'est pas le cas du test binomial où seulement moins de la moitié des gènes passent le filtre. Cela signifie que les SNPs les plus significatifs sortent déjà en « cluster » pour le test de contingence. L'application de la dernière étape de filtrage exigeant au moins 8 individus en AI fait que le nombre de gènes gardés suite à l'application du test de contingence est réduit presque à moitié. Par contre, l'ensemble des gènes identifiés par le test binomial n'est pas affecté par ce filtre. De cette observation, nous trouvons l'explication dans la figure 3.1 qui montre qu'il faut avoir au moins 11 individus en AI pour obtenir un résultat significatif avec le test binomial à un seuil de 0,001. Par contre, le test de contingence donne déjà des résultats significatifs avec deux individus en AI pour le même seuil. Le test de contingence est donc moins dépendant du nombre d'individus en AI que le test binomial qui n'utilise que les individus en AI. De plus, nous voyons qu'avec le même nombre d'individus en AI, nous pouvons obtenir des valeurs de p plus significatives avec le test de contingence qu'avec le test binomial. Au fur et à mesure que le nombre d'individus en AI augmente, les valeurs de p pour le test de contingence diminuent plus rapidement jusqu'à des valeurs qui ne sont pas atteintes avec le test binomial. Si l'on prend ensuite les valeurs de p des faux positifs, nous observons que pour le test de contingence, elles restent stables peu importe le nombre d'individus en AI. Ainsi, l'écart entre la valeur minimale de p des faux positifs et celle des gènes en AI augmente considérablement avec le nombre d'individus en AI pour devenir maximal aux environs de 22 individus. Par contre, nous observons que pour le test binomial les valeurs minimales de p des faux positifs diminuent avec

l'augmentation du nombre d'individus en AI. Ainsi, l'écart entre les valeurs de p minimales des faux positifs et celles des gènes en AI reste presque constante. Dans la figure 3.2, nous voyons bien cet écart, peu importe le nombre d'individus en AI. Le pic des valeurs de p minimales est beaucoup plus élevé pour le test de contingence que pour le test binomial. De plus, la différence entre la valeur de p minimale des faux positifs et celle des gènes en AI est trois fois plus grande pour le test de contingence que pour le test binomial. De ces observations, nous pouvons constater que le test de contingence a plus de pouvoir que le test binomial. De plus, le nombre d'individus en AI a un effet plus important sur le test binomial que sur le test de contingence.

Dans plusieurs cas, une région régulatrice unique peut être détectée clairement en regardant seulement les pics de valeurs de p comme les gènes *LRR1Q3*, *XKR9*, *ALDH8A1* et *AHI1* (figures 3.3 à 3.9). Ces derniers gènes, le *ALDH8A1* et le *AHI1*, sont situés sur le même chromosome à 500 kb l'un de l'autre. Malgré la courte distance qui les sépare, nous avons détecté une région régulatrice spécifique pour chacun. Dans le cas du *LRR1Q3*, pour lequel tout le génome a été testé avec le test de contingence, nous voyons un pic qui signale une région régulatrice candidate dans le chromosome 1 à côté du gène. Si nous regardons le gène *ALDH8A1* dans la figure 3.4, le bruit de fond pour les deux tests atteint les mêmes valeurs de p , par contre le pic signalant la région candidate régulatrice dans le test de contingence atteint des valeurs de p plus basses par rapport au bruit de fond que dans le test binomial. Il faut remarquer que ces gènes (*LRR1Q3*, *XKR9*, *ALDH8A1* et *AHI1*) ont entre 24 et 30

individus en AI et comme nous l'avons mentionné précédemment, c'est là où les tests performant le mieux. Ces quatre exemples montrent la spécificité et l'exactitude de notre approche malgré la différence de pouvoir entre les deux tests.

Quand la région régulatrice est plus loin du gène, elle est plus difficile à localiser [Pastinen et al., 2006]. Dans le cas du gène QDPR (figure 3.8) le test de contingence détecte une région à 850 kb du gène régulé alors que le test binomial ne le détecte pas. Le triangle de LD montre une grande fragmentation entre le gène et la région régulatrice. Le taux élevé de recombinaison rend l'inférence de la phase moins fiable. C'est aussi vrai pour les données de HapMap II qui semblent très bien phasées parce que leurs données proviennent de trios (père, mère et enfant). De plus, le nombre d'erreurs de phasage augmente avec la distance. Selon [Fan et al., 2011], dans le cas du trio familial de HapMap, une distance de 3 à 8 megabases sépare les erreurs de phasage entre eux; mais, dans les cas généraux, ces erreurs se trouvent à chaque 300 kilobases de distance. Cela montre la magnitude du problème de phasage. Dans le cas du gène NEFH, le test de contingence détecte une région régulatrice à 1000 kb de la localisation du gène, mais, dans ce cas ici, le test binomial montre quelques SNPs qui dépassent à peine le seuil de signification de 0,01 dans la même région signalée par le test de contingence. Contrairement au cas précédent, la région qui sépare le gène de sa région régulatrice est beaucoup moins fragmentée facilitant l'inférence de la phase. Même si le nombre d'individus en AI est un peu plus élevé dans le derniers cas, nous pensons que l'échec du test binomial dans le cas du gène QDPR est dû aux problèmes de phasage. En plus, la distance physique et

génétiq ue rend la détermination de la phase plus difficile. Par rapport au test de contingence, nous voyons que les valeurs de p ne sont pas hautement significatives. Cependant, ce qui caractérise la région régulatrice dans les deux cas est un groupement de plusieurs SNPs significatifs situés à très courte distance l'un de l'autre. En fait, nous cherchons un « cluster » de SNPs significatifs pour déterminer la région régulatrice. Cela est très utile quand les valeurs de p ne sont pas excessivement significatives pour identifier la région candidate. Dans l'analyse de la distribution de la distance entre la région régulatrice et le gène (figure 3.2), nous validons l'observation précédente avec le gène QDPR et NEFH. Le test de contingence est capable de détecter des régions significatives à plus grande distance que le test binomial. Ainsi, le test de contingence s'avère plus utile pour cartographier des régions régulatrices indépendamment de la distance et sans besoin d'avoir l'information haplotypique.

La figure 3.2 montre aussi que la plus haute concentration de SNPs avec les valeurs les plus significatives se trouve près du gène. Cette observation a été faite par [Idaghdour et al., 2010] et [Myers et al., 2007] qui ont montré que la plupart des variations régulatrices se trouvent près du gène qu'elles régulent, mais qu'il y a des cas où des éléments régulateurs en *cis* peuvent se localiser à plus grande distance. Cela confirme que nos résultats coïncident avec la littérature par rapport à la distance qui sépare la variation régulatrice et le gène régulé.

De façon générale, l'erreur empirique pour les deux tests est très basse avec un taux de faux positifs de 0,1%. Il a été calculé avec un $p < 0,01$. (Tableau III.II) Cela nous a permis de valider qu'un seuil de 0,01 est adéquat pour l'ensemble des tests.

Cependant, nous avons utilisé un seuil de significativité de 0,001 au niveau génomique pour plusieurs comparaisons afin de s'assurer de la validité des associations. La variation observée dans le taux du test binomial et la constance du test de contingence peuvent être expliquées par la nature de chaque test. Le test binomial est plus sensible au nombre d'individus en AI.

La comparaison avec la base de données « BD eQTL » (provenant du laboratoire du Dr. Pritchard), nous a permis de comparer notre méthode avec d'autres études. Nous avons vu que la correspondance parmi ces dernières est déjà faible. Par exemple, le gène APOC2 a été rapporté seulement par [Veyrieras et al., 2008] et le gène APIP seulement par [Zeller et al., 2010]. En tenant compte de cette situation, avoir trouvé 22% des gènes simultanément dans les données de [Ge et al., 2009] et dans la base de données « BD eQTLs » avec le plus haut niveau de filtrage, est très appréciable (tableau III.III). De plus, nous montrons 60 exemples de gènes (tableau III.IV) pour lesquels nous avons identifié des SNPs signalés dans « BD eQTLs » pour le même gène. C'est le cas du gène XKR9, pour lequel nous avons identifié 169 SNPs en commun (tableau III.IV) [Veyrieras et al., 2008]. Cela montre que notre méthode détecte bien les régions régulatrices.

Un objectif secondaire, mais aussi important, que nous avons poursuivi était l'utilisation des tests génotypiques de contingence 2x3 (LD2x3) et 3x3 (LD3x3) pour évaluer le LD entre deux SNPs sans connaître la phase haplotypique. Ces deux tests génotypiques sont idéaux pour éviter la reconstruction haplotypique et donc les erreurs de phasage. Nous avons comparé le test haplotypique de contingence 2x2

(LD2x2) [Hudson 2001] utilisé couramment pour évaluer le LD comme base de comparaison. Le test LD2x3 a une puissance inférieure au test LD3x3. D'ailleurs, ce dernier a une puissance quasi égale au test LD2x2 avec un seuil de 0,01. Cependant, la diminution du seuil de signification creuse l'écart entre la puissance de ces tests. Il faut également remarquer que l'erreur des tests LD2x3 et LD3x3 est presque nulle. Plus de 98 % des paires identifiées en LD par le test LD2x3 et LD3x3 sont aussi identifiées par le test LD2x2. (Tableau III.V). De plus, nos résultats montrent une corrélation plus forte entre le test LD2x2 et LD3x3 qu'entre LD2x3 et LD2x2 au niveau des valeurs de p . (figure 3.10 et 3.11). Les différents niveaux d'efficacité s'expliquent en partie par le degré d'information de chaque test. Le test LD2x2 possède l'information haplotypique, ce qui permet d'évaluer le LD directement. Par contre, les tests génotypiques évaluent le LD indirectement, en regardant la distribution génotypique. Le test LD3x3 est plus performant que le test LD2x3 parce que ce dernier possède l'information génotypique d'un seul SNP contrairement au LD3x3.

Nous avons démontré que notre approche est très efficace pour détecter des régions régulatrices. Nous avons présenté un test génotypique qui peut détecter des régions loin du gène et qui n'a pas besoin de données phasées. Cela réduit énormément le temps d'analyse parce que les algorithmes les plus précis comme PHASE peuvent prendre des jours, même des mois pour reconstruire des haplotypes [Scheet and Stephens, 2006] et [Delaneau et al., 2008]. De plus, les variations régulatrices éloignées sont les plus difficiles à détecter [Pastinen et al., 2006]. Nous avons aussi introduit avec succès le test binomial complémentaire au test génotypique à utiliser quand

l'information haplotypique est disponible. Nous avons aussi démontré l'efficacité des tests génotypiques 2x3 et 3x3 pour évaluer le LD entre deux SNPs sans connaître la phase.

Ce travail ouvre la porte à des futures recherches. La base de données créée pourrait être intégrée à des bases de données d' eQTLs , des sites régulateurs ou d'information de déséquilibre allélique déjà existant. De plus, la grande quantité de données générée est une opportunité unique pour analyser la co-régulation. L'analyse de « clustering » pourrait apporter plus d'efficacité pour trouver des régions régulatrices où les valeurs de p ne sont pas excessivement significatives.

Bibliographie

- J.F.Griffiths et al., 2001:** J.F. Griffiths, William M. Gelbart, Jeffrey H. Miller, Richard C. Lewontin. Traduction de la 1re. édition américaine par Chrystelle Sanlaville. *Analyse Génétique Moderne*. 2001. De Boeck Universit r  s.a. 676.
- King et al. 1975:** M. C. King and A. C. Wilson. Evolution at two levels in humans and chimpanzees. *Science* 1975; 188(4184):107-116
- Diaw et al., 2007:** Lena Diaw, Karen Woodson and John W Gillespie. Prostate cancer epigenetics: a review on gene regulation. *Gene Regul Syst Bio* 2007; 1:313-325
- Pastinen et al., 2004:** Tomi Pastinen, Robert Sladek, Scott Gurd, Alya'a Sammak, Bing Ge, Pierre Lepage, Karine Lavergne, Amelie Villeneuve, Tiffany Gaudin, Helena Br ndstr m, Allon Beck, Andrei Verner and Jade Kingsley, Eef Harmsen, Damian Labuda, Kenneth Morgan, Marie-Claude Vohl, Anna K Naumova, Daniel Sinnott and Thomas J Hudson. A survey of genetic and epigenetic variation affecting human gene expression. *Physiol Genomics* 2004; 16(2):184-193
- Lewin, 2008:** Benjamin Lewin, *Genes IX*, 2008. Jones and Bartlett Publishers. 892.
- Noonan et al., 2010:** James P Noonan and Andrew S McCallion. Genomics of long-range regulatory elements. *Annu Rev Genomics Hum Genet* 2010; 11:1-23
- Pastinen et al., 2006:** Tomi Pastinen, Bing Ge and Thomas J Hudson. Influence of human genome polymorphism on gene expression. *Hum Mol Genet* 2006; 15 Spec No 1:R9-16

- Hardison, 2003:** Ross C Hardison. Comparative genomics. *PLoS Biol* 2003; 1(2):E58
- Maston et al., 2006:** Glenn A Maston, Sara K Evans and Michael R Green. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* 2006; 7:29-59
- Gross and Garrard, 1988:** D. S. Gross and W. T. Garrard. Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem* 1988; 57:159-197
- Sabo et al., 2006:** Peter J Sabo, Michael S Kuehn, Robert Thurman, Brett E Johnson, Ericka M Johnson, Hua Cao, Man Yu, Elizabeth Rosenzweig, Jeff Goldy, Andrew Haydock, Molly Weaver, Anthony Shafer, Kristin Lee, Fidencio Neri, Richard Humbert, Michael A Singer, Todd A Richmond, Michael O Dorschner, Michael McArthur, Michael Hawrylycz, Roland D Green, Patrick A Navas, William S Noble and John A Stamatoyannopoulos. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods* 2006; 3(7):511-518
- Euskirchen et al., 2007:** Ghia M Euskirchen, Joel S Rozowsky, Chia-Lin Wei, Wah Heng Lee, Zhengdong D Zhang, Stephen Hartman, Olof Emanuelsson, Viktor Stolc, Sherman Weissman, Mark B Gerstein, Yijun Ruan and Michael Snyder. Mapping of transcription factor binding regions in mammalian cells by CHIP: comparison of array- and sequencing-based technologies. *Genome Res* 2007; 17(6):898-909
- Pastinen and Hudson, 2004:** Tomi Pastinen and Thomas J Hudson. Cis-acting regulatory variation in the human genome. *Science* 2004; 306(5696):647-650
- Bray et al., 2003:** Nicholas J Bray, Paul R Buckland, Nigel M Williams and Hywel J Williams, Nadine Norton, Michael J Owen and Michael C O'Donovan. A haplotype

implicated in schizophrenia susceptibility is associated with reduced COMT expression in human brain. *Am J Hum Genet* 2003; 73(1):152-161

Tuupanen et al., 2008: Sari Tuupanen, Iina Niittymäki, Kari Nousiainen, Sakari Vanharanta, Jukka-Pekka Mecklin, Kyösti Nuorva, Heikki Järvinen, Sampsa Hautaniemi, Auli Karhu and Lauri A Aaltonen. Allelic imbalance at rs6983267 suggests selection of the risk allele in somatic colorectal tumor evolution. *Cancer Res* 2008; 68(1):14-17

Loeuillet et al., 2007: Corinne Loeuillet, Michael Weale, Samuel Deutsch, Margalida Rotger, Nicole Soranzo, Josiane Wyniger, Guillaume Lettre, Yann Dupré, Delphine Thuillard, Jacques S Beckmann, Stylianos E Antonarakis, David B Goldstein and Amalio Telenti. Promoter polymorphisms and allelic imbalance in ABCB1 expression. *Pharmacogenet Genomics* 2007; 17(11):951-959

Ding et al, 2004: Chunming Ding, Esther Maier, Adelbert A Roscher, Andreas Braun and Charles R Cantor. Simultaneous quantitative and allele-specific expression analysis with real competitive PCR. *BMC Genet* 2004; 5:8

Heighway et al., 2005: J. Heighway and N. L. Bowers, S. Smith and D. C. Betticher and M. F Santibáñez Koref. The use of allelic expression differences to ascertain functional polymorphisms acting in cis: analysis of MMP1 transcripts in normal lung tissue. *Ann Hum Genet* 2005; 69(Pt 1):127-133

Ge et al., 2009: Bing Ge, Dmitry K Pokholok, Tony Kwan, Elin Grundberg, Lisanne Morcos, Dominique J Verlaan, Jennie Le, Vonda Koka, Kevin C L Lam, Vincent Gagné, Joana Dias, Rose Hoberman, Alexandre Montpetit, Marie-Michele Joly, Edward J Harvey, Daniel Sinnett, Patrick Beaulieu, Robert Hamon, Alexandru

Graziani, Ken Dewar, Eef Harmsen, Jacek Majewski, Harald H H Göring, Anna K Naumova, Mathieu Blanchette, Kevin L Gunderson and Tomi Pastinen. Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. *Nat Genet* 2009; 41(11):1216-1222

Montgomery et al., 2010: Stephen B Montgomery, Micha Sammeth, Maria Gutierrez-Arcelus, Radoslaw P Lach, Catherine Ingle, James Nisbett, Roderic Guigo and Emmanouil T Dermitzakis. Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 2010; 464(7289):773-777

Higasa et al., 2009: Koichiro Higasa, Yoji Kukita, Kiyoko Kato, Norio Wake, Tomoko Tahira and Kenshi Hayashi. Evaluation of haplotype inference using definitive haplotype data obtained from complete hydatidiform moles, and its significance for the analyses of positively selected regions. *PLoS Genet* 2009; 5(5):e1000468

Fan et al., 2011: H. Christina Fan, Jianbin Wang, Anastasia Potanina and Stephen R Quake. Whole-genome molecular haplotyping of single cells. *Nat Biotechnol* 2011; 29(1):51-57

Marchini et al., 2006: Jonathan Marchini, David Cutler, Nick Patterson, Matthew Stephens, Eleazar Eskin, Eran Halperin, Shin Lin, Zhaohui S Qin, Heather M Munro, Goncalo R Abecasis, Peter Donnelly and International HapMap Consortium. A comparison of phasing algorithms for trios and unrelated individuals. *Am J Hum Genet* 2006; 78(3):437-450

Flint-Garcia et al., 2003: Sherry A Flint-Garcia, Jeffry M Thornsberry and Edward S Buckler. Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* 2003;

54:357-374

Tewhey et al., 2011: Ryan Tewhey, Vikas Bansal, Ali Torkamani, Eric J Topol and Nicholas J Schork. The importance of phase information for human genomics. *Nat Rev Genet* 2011; 12(3):215-223

Sved 2009: John A Sved. Linkage disequilibrium and its expectation in human populations. *Twin Res Hum Genet* 2009; 12(1):35-43

Kerem et al, 1989: B. Kerem, J. M. Rommens, J. A. Buchanan, D. Markiewicz, T. K. Cox, A. Chakravarti, M. Buchwald and L. C. Tsui. Identification of the cystic fibrosis gene: genetic analysis. *Science* 1989; 245(4922):1073-1080

Corder et al., 1994: E. H. Corder, A. M. Saunders, N. J. Risch, W. J. Strittmatter, D. E. Schmechel, P. C. Gaskell, J. B. Rimmler, P. A. Locke, P. M. Conneally and K. E. Schmechel. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nat Genet* 1994; 7(2):180-184

RDevelopmentCoreTeam 2009: R Development Core Team, *R: A Language and Environment for Statistical Computing*, 2009

Urbanek 2011: Simon Urbanek, *Rserve: Binary R server*, 2011

Wagner et al., 2010: James R Wagner, Bing Ge, Dmitry Pokholok, Kevin L Gunderson, Tomi Pastinen, Mathieu Blanchette. Computational analysis of whole-genome differential allelic expression data in human. *PLoS Comput Biol* 2010; 6(7):e1000849

Idaghdour et al., 2010: Youssef Idaghdour, Wendy Czika and Kevin V Shianna, Sang H Lee, Peter M Visscher, Hilary C Martin, Kelci Miclaus, Sami J Jadallah, David B Goldstein, Russell D Wolfinger and Greg Gibson. Geographical genomics

of human leukocyte gene expression variation in southern Morocco. *Nat Genet* 2010; 42(1):62-67

Myers et al., 2007: Amanda J Myers, J. Raphael Gibbs, Jennifer A Webster, Kristen Rohrer, Alice Zhao, Lauren Marlowe, Mona Kaleem, Doris Leung, Leslie Bryden, Priti Nath, Victoria L Zismann, Keta Joshipura, Matthew J Huentelman, Diane Hurlin, Keith D Coon, David W Craig, John V Pearson, Peter Holmans, Christopher B Heward, Eric M Reiman, Dietrich Stephan and John Hardy. A survey of genetic human cortical gene expression. *Nat Genet* 2007; 39(12):1494-1499

Zeller et al., 2010: Tanja Zeller, Philipp Wild, Silke Szymczak, Maxime Rotival, Arne Schillert, Raphaelae Castagne, Seraya Maouche, Marine Germain, Karl Lackner, Heidi Rossmann, Medea Eleftheriadis, Christoph R Sinning, Renate B Schnabel, Edith Lubos, Detlev Mennerich, Werner Rust, Claire Perret, Carole Proust, Viviane Nicaud, Joseph Loscalzo, Norbert Hübner, David Tregouet, Thomas Münzel, Andreas Ziegler, Laurence Turet, Stefan Blankenberg and François Cambien. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* 2010; 5(5):e10693

Stranger et al., 2007: Barbara E Stranger, Alexandra C Nica, Matthew S Forrest, Antigone Dimas, Christine P Bird, Claude Beazley, Catherine E Ingle, Mark Dunning, Paul Flicek, Daphne Koller, Stephen Montgomery, Simon Tavaré, Panos Deloukas and Emmanouil T Dermizakis. Population genomics of human gene expression. *Nat Genet* 2007; 39(10):1217-1224

Veyrieras et al., 2008: Jean-Baptiste Veyrieras, Sridhar Kudaravalli, Su Yeon Kim, Emmanouil T Dermizakis, Yoav Gilad, Matthew Stephens and Jonathan K Pritchard.

High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet* 2008; 4(10):e1000214

Lefebvre and Labuda, 2008: J-F. Lefebvre and D. Labuda. Fraction of informative recombinations: a heuristic approach to analyze recombination rates. *Genetics* 2008; 178(4):2069-2079

Hudson 2001: R.R. Hudson, Linkage Disequilibrium and Recombination. In D.J. Balding, M.Bishop and C. Cannings, *Handbook of Statistical Genetics*, 2001, John Wiley & sons, LTD, 864.

Hudson 2002: Richard R Hudson. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 2002; 18(2):337-338

Lee et al., 2006: Peter D Lee, Bing Ge, Celia M T Greenwood, Donna Sinnett, Yannick Fortin, Sebastien Brunet, Anny Fortin, Marina Takane, Emil Skamene, Tomi Pastinen, Michael Hallett, Thomas J Hudson and Robert Sladek. Mapping cis-acting regulatory variation in recombinant congenic strains. *Physiol Genomics* 2006; 25(2):294-302

Scheet and Stephens, 2006: Paul Scheet and Matthew Stephens. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* 2006; 78(4):629-644

Delaneau et al., 2008: Olivier Delaneau, Cédric Coulonges and Jean-François Zagury. Shape-IT: new rapid and accurate algorithm for haplotype inference. *BMC Bioinformatics* 2008; 9:540

Pickrell et al., 2010: Joseph K Pickrell, John C Marioni, Athma A Pai, Jacob F Degner, Barbara E Engelhardt, Everlyne Nkadori, Jean-Baptiste Veyrieras, Matthew

Stephens, Yoav Gilad and Jonathan K Pritchard. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 2010; 464(7289):768-772

Schadt et al., 2008: Eric E Schadt, Cliona Molony, Eugene Chudin, Ke Hao, Xia Yang, Pek Y Lum, Andrew Kasarskis, Bin Zhang, Susanna Wang, Christine Suver, Jun Zhu, Joshua Millstein, Solveig Sieberts, John Lamb, Debraj GuhaThakurta, Jonathan Derry, John D Storey, Iliana Avila-Campillo, Mark J Kruger, Jason M Johnson, Carol A Rohl, Atila van Nas, Margarete Mehrabian, Thomas A Drake, Aldons J Lusis, Ryan C Smith, F. Peter Guengerich, Stephen C Strom, Erin Schuetz, Thomas H Rushmore and Roger Ulrich. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* 2008; 6(5):e107

Annexe I

Liste détaillée d'eQTLs qui ont été détectés dans des études précédente (base de donnée de eQTLs du laboratoire du Dr Pritchard , Université de Chicago) et dans cette étude pour le même gène. (Test de contingence 2x3)

gene	SNPid	pvalue	source
AHI1	rs17064702	9,74E-009	VeyrierasEQTL
AHI1	rs11154810	1,41E-008	Zeller10_CIS_QTL
AHI1	rs6925161	1,41E-008	Zeller10_CIS_QTL
AHI1	rs9494308	1,41E-008	Zeller10_CIS_QTL
AHI1	rs9321512	3,97E-008	Zeller10_CIS_QTL
AHI1	rs13202115	5,79E-008	VeyrierasEQTL
AHI1	rs9376115	1,64E-007	Zeller10_CIS_QTL
AHI1	rs6934889	1,64E-007	Zeller10_CIS_QTL
AHI1	rs4896156	2,64E-007	Zeller10_CIS_QTL
AHI1	rs7769840	2,64E-007	Zeller10_CIS_QTL
AHI1	rs6941396	4,62E-006	VeyrierasEQTL
AHI1	rs6941396	4,62E-006	Zeller10_CIS_QTL
AHI1	rs9402729	2,44E-004	Zeller10_CIS_QTL
ALDH8A1	rs4440481	4,76E-010	Zeller10_CIS_QTL
ALDH8A1	rs4472368	4,76E-010	MyersEQTL
ALDH8A1	rs4472368	4,76E-010	Zeller10_CIS_QTL
ALDH8A1	rs4440481	4,76E-010	MyersEQTL
ALDH8A1	rs6923512	5,71E-009	Zeller10_CIS_QTL
ALDH8A1	rs4289677	5,71E-009	Zeller10_CIS_QTL
ALDH8A1	rs4896118	5,71E-009	Zeller10_CIS_QTL
ALDH8A1	rs4896128	5,71E-009	Zeller10_CIS_QTL
ALDH8A1	rs3817776	1,89E-008	StrangerEQTL
ALDH8A1	rs4445070	1,92E-006	StrangerEQTL
ALDH8A1	rs6923765	1,92E-006	Zeller10_CIS_QTL
ALDH8A1	rs7741515	1,92E-006	Zeller10_CIS_QTL
ALDH8A1	rs7741515	1,92E-006	MyersEQTL
ALDH8A1	rs2150681	1,92E-006	Zeller10_CIS_QTL
ALDH8A1	rs1590975	1,92E-006	Zeller10_CIS_QTL
ALDH8A1	rs9321481	1,92E-006	MyersEQTL
ALDH8A1	rs1590975	1,92E-006	MyersEQTL
ALDH8A1	rs7750300	1,92E-006	Zeller10_CIS_QTL
ALDH8A1	rs12526055	1,92E-006	Zeller10_CIS_QTL

ALDH8A1	rs9321481	1,92E-006	Zeller10_CIS_QTL
ALDH8A1	rs2150681	1,92E-006	MyersEQTL
ALDH8A1	rs12661423	1,92E-006	Zeller10_CIS_QTL
ALDH8A1	rs6923765	1,92E-006	MyersEQTL
ALDH8A1	rs7750300	1,92E-006	MyersEQTL
ALDH8A1	rs2072827	9,33E-006	StrangerEQTL
ALDH8A1	rs11754265	1,15E-005	Zeller10_CIS_QTL
ALDH8A1	rs11754265	1,15E-005	MyersEQTL
ALDH8A1	rs1014021	1,15E-005	Zeller10_CIS_QTL
ALDH8A1	rs1014021	1,15E-005	MyersEQTL
ANKDD1A	rs1719264	2,72E-008	VeyrierasEQTL
ANKDD1A	rs1684054	1,79E-007	VeyrierasEQTL
ANKDD1A	rs1684054	1,79E-007	PickrelEQTL
ANKDD1A	rs12916644	1,79E-007	VeyrierasEQTL
ANKDD1A	rs1628955	1,79E-007	VeyrierasEQTL
ANKDD1A	rs1719266	1,79E-007	VeyrierasEQTL
ANKDD1A	rs1684034	1,79E-007	VeyrierasEQTL
ANKDD1A	rs4238397	1,79E-007	VeyrierasEQTL
ANKDD1A	rs1719273	1,79E-007	VeyrierasEQTL
ANKRD42	rs552997	1,50E-006	Zeller10_CIS_QTL
ANKRD42	rs506321	2,14E-004	Zeller10_CIS_QTL
APIP	rs17711539	2,38E-008	Zeller10_CIS_QTL
APIP	rs17783998	2,38E-008	Zeller10_CIS_QTL
APOC2	rs2288911	2,54E-010	VeyrierasEQTL
APOC2	rs3760629	2,54E-010	VeyrierasEQTL
APOC2	rs2239375	2,54E-010	VeyrierasEQTL
APOC2	rs2288912	2,54E-010	VeyrierasEQTL
APOC2	rs7257476	2,54E-010	VeyrierasEQTL
APOC2	rs7257916	1,47E-009	VeyrierasEQTL
APOC2	rs11669173	1,47E-009	VeyrierasEQTL
APOC2	rs204468	3,61E-004	VeyrierasEQTL
APOC2	rs204467	3,61E-004	VeyrierasEQTL
APOC2	rs204474	3,61E-004	VeyrierasEQTL
BLOC1S2	rs17668159	6,64E-006	Zeller10_CIS_QTL
BLOC1S2	rs7073610	6,64E-006	Zeller10_CIS_QTL
BLOC1S2	rs17112791	6,64E-006	Zeller10_CIS_QTL
BLOC1S2	rs6584356	1,77E-005	Zeller10_CIS_QTL
BLOC1S2	rs7924284	1,77E-005	Zeller10_CIS_QTL
BLOC1S2	rs2242052	3,98E-005	Zeller10_CIS_QTL
BTN3A2	rs16891725	1,25E-010	Zeller10_CIS_QTL
BTN3A2	rs16891725	1,25E-010	VeyrierasEQTL
BTN3A2	rs9348712	1,25E-010	VeyrierasEQTL
BTN3A2	rs2073531	1,25E-010	VeyrierasEQTL
BTN3A2	rs1977	1,25E-010	VeyrierasEQTL
BTN3A2	rs13216828	1,25E-010	VeyrierasEQTL

BTN3A2	rs9393707	1,25E-010	VeyrierasEQTL
BTN3A2	rs9393706	1,25E-010	VeyrierasEQTL
BTN3A2	rs9393708	1,25E-010	VeyrierasEQTL
BTN3A2	rs9379851	1,25E-010	VeyrierasEQTL
BTN3A2	rs9393705	1,25E-010	VeyrierasEQTL
BTN3A2	rs9348709	1,25E-010	VeyrierasEQTL
BTN3A2	rs12174631	1,25E-010	VeyrierasEQTL
BTN3A2	rs12176317	1,25E-010	VeyrierasEQTL
BTN3A2	rs9379851	1,25E-010	Zeller10_CIS_QTL
BTN3A2	rs2073529	1,25E-010	VeyrierasEQTL
BTN3A2	rs1978	1,25E-010	Zeller10_CIS_QTL
BTN3A2	rs9348716	1,25E-010	VeyrierasEQTL
BTN3A2	rs1977	1,25E-010	Zeller10_CIS_QTL
BTN3A2	rs2073529	1,25E-010	Zeller10_CIS_QTL
BTN3A2	rs12174602	1,25E-010	VeyrierasEQTL
BTN3A2	rs9358932	1,25E-010	VeyrierasEQTL
BTN3A2	rs9379855	1,25E-010	VeyrierasEQTL
BTN3A2	rs9358932	1,25E-010	Zeller10_CIS_QTL
BTN3A2	rs9379855	1,25E-010	Zeller10_CIS_QTL
BTN3A2	rs9358935	1,25E-010	VeyrierasEQTL
BTN3A2	rs1978	1,25E-010	VeyrierasEQTL
BTN3A2	rs1979	1,25E-010	VeyrierasEQTL
BTN3A2	rs9379864	1,25E-010	VeyrierasEQTL
BTN3A2	rs9366653	1,25E-010	VeyrierasEQTL
BTN3A2	rs12173854	1,25E-010	VeyrierasEQTL
BTN3A2	rs9379859	1,25E-010	VeyrierasEQTL
BTN3A2	rs9393708	1,25E-010	Zeller10_CIS_QTL
BTN3A2	rs9379856	1,25E-010	VeyrierasEQTL
BTN3A2	rs9379857	1,25E-010	VeyrierasEQTL
BTN3A2	rs9393714	1,25E-010	VeyrierasEQTL
BTN3A2	rs9379858	1,25E-010	VeyrierasEQTL
BTN3A2	rs9393713	1,25E-010	VeyrierasEQTL
BTN3A2	rs9393713	1,25E-010	Zeller10_CIS_QTL
BTN3A2	rs9358937	1,25E-010	VeyrierasEQTL
BTN3A2	rs9358934	1,25E-010	VeyrierasEQTL
BTN3A2	rs3799380	3,80E-009	Zeller10_CIS_QTL
BTN3A2	rs9358945	3,80E-009	Zeller10_CIS_QTL
BTN3A2	rs7773938	3,80E-009	Zeller10_CIS_QTL
BTN3A2	rs3823158	4,95E-009	VeyrierasEQTL
BTN3A2	rs13195509	4,95E-009	VeyrierasEQTL
BTN3A2	rs13195509	4,95E-009	Zeller10_CIS_QTL
BTN3A2	rs9366655	4,95E-009	VeyrierasEQTL
BTN3A2	rs3734544	1,23E-007	VeyrierasEQTL
BTN3A2	rs853681	1,31E-007	Zeller10_CIS_QTL
BTN3A2	rs2799079	1,24E-006	Zeller10_CIS_QTL

BTN3A2	rs1778508	1,24E-006	Zeller10_CIS_QTL
BTN3A2	rs200956	1,24E-006	Zeller10_CIS_QTL
BTN3A2	rs200991	1,24E-006	Zeller10_CIS_QTL
BTN3A2	rs17693963	1,34E-006	Zeller10_CIS_QTL
BTN3A2	rs17720293	2,01E-006	Zeller10_CIS_QTL
BTN3A2	rs201002	6,33E-006	Zeller10_CIS_QTL
BTN3A2	rs4712969	6,33E-006	Zeller10_CIS_QTL
BTN3A2	rs200995	6,33E-006	Zeller10_CIS_QTL
BTN3A2	rs1892253	7,40E-006	Zeller10_CIS_QTL
BTN3A2	rs9380069	1,67E-005	Zeller10_CIS_QTL
BTN3A2	rs202906	2,18E-005	Zeller10_CIS_QTL
BTN3A2	rs1892252	2,46E-005	Zeller10_CIS_QTL
BTN3A2	rs9348775	3,03E-005	Zeller10_CIS_QTL
BTN3A2	rs9468317	3,46E-005	Zeller10_CIS_QTL
BTN3A2	rs9393909	3,46E-005	Zeller10_CIS_QTL
BTN3A2	rs200994	4,46E-005	Zeller10_CIS_QTL
BTN3A2	rs9468220	5,27E-005	Zeller10_CIS_QTL
BTN3A2	rs1237875	5,88E-005	Zeller10_CIS_QTL
BTN3A2	rs1233704	5,88E-005	Zeller10_CIS_QTL
BTN3A2	rs1233701	5,88E-005	Zeller10_CIS_QTL
BTN3A2	rs1233699	5,88E-005	Zeller10_CIS_QTL
BTN3A2	rs4712972	7,94E-005	Zeller10_CIS_QTL
BTN3A2	rs10484403	1,05E-004	Zeller10_CIS_QTL
BTN3A2	rs203888	1,05E-004	Zeller10_CIS_QTL
BTN3A2	rs9295756	1,05E-004	Zeller10_CIS_QTL
BTN3A2	rs9393881	1,05E-004	Zeller10_CIS_QTL
BTN3A2	rs9295755	1,05E-004	Zeller10_CIS_QTL
BTN3A2	rs1119211	1,67E-004	Zeller10_CIS_QTL
BTN3A2	rs17301128	1,67E-004	Zeller10_CIS_QTL
BTN3A2	rs200988	2,08E-004	Zeller10_CIS_QTL
BTN3A2	rs6933583	3,17E-004	Zeller10_CIS_QTL
BTN3A2	rs4712980	3,17E-004	Zeller10_CIS_QTL
BTN3A2	rs2073528	3,17E-004	VeyrierasEQTL
BTN3A2	rs6933583	3,17E-004	VeyrierasEQTL
BTN3A2	rs4712981	3,17E-004	VeyrierasEQTL
BTN3A2	rs9358933	3,17E-004	VeyrierasEQTL
BTN3A2	rs9379870	3,17E-004	VeyrierasEQTL
BTN3A2	rs4712980	3,17E-004	VeyrierasEQTL
BTN3A2	rs1233708	8,76E-004	Zeller10_CIS_QTL
C13orf18	rs2477601	7,90E-006	Zeller10_CIS_QTL
C13orf18	rs9534398	3,46E-005	Zeller10_CIS_QTL
C13orf18	rs9534399	3,46E-005	Zeller10_CIS_QTL
C13orf18	rs259728	3,47E-005	Zeller10_CIS_QTL
C13orf18	rs259737	4,85E-005	Zeller10_CIS_QTL
C13orf18	rs259734	1,22E-004	Zeller10_CIS_QTL

C13orf18	rs2765619	1,22E-004	Zeller10_CIS_QTL
C13orf18	rs1216719	2,07E-004	Zeller10_CIS_QTL
C16orf88	rs8044607	9,48E-011	PickrelEQTL
C21orf63	rs2833834	2,15E-008	Zeller10_CIS_QTL
C21orf63	rs2833839	2,15E-008	Zeller10_CIS_QTL
C7orf25	rs667209	3,76E-007	Zeller10_CIS_QTL
C7orf25	rs591875	9,41E-007	Zeller10_CIS_QTL
C7orf25	rs647618	1,26E-004	Zeller10_CIS_QTL
C9orf128	rs7873145	7,34E-010	Zeller10_CIS_QTL
C9orf128	rs2145923	7,34E-010	Zeller10_CIS_QTL
C9orf128	rs12342037	5,13E-008	Zeller10_CIS_QTL
C9orf128	rs6476491	5,13E-008	Zeller10_CIS_QTL
C9orf128	rs3793543	5,13E-008	Zeller10_CIS_QTL
CATSPER2	rs2278860	7,32E-006	Zeller10_CIS_QTL
CCL14	rs10491118	6,78E-006	SchadtEQTL
CHI3L2	rs942697	3,18E-008	Zeller10_CIS_QTL
CHI3L2	rs942697	3,18E-008	StrangerEQTL
CHI3L2	rs6691256	7,77E-007	Zeller10_CIS_QTL
CHI3L2	rs7514853	7,77E-007	Zeller10_CIS_QTL
CHI3L2	rs6691497	7,77E-007	Zeller10_CIS_QTL
CHI3L2	rs11102223	7,77E-007	Zeller10_CIS_QTL
CHI3L2	rs2494004	9,12E-007	Zeller10_CIS_QTL
CHI3L2	rs2182114	9,12E-007	Zeller10_CIS_QTL
CHI3L2	rs2477578	9,12E-007	Zeller10_CIS_QTL
CHI3L2	rs6537686	9,12E-007	Zeller10_CIS_QTL
CHI3L2	rs942694	9,12E-007	Zeller10_CIS_QTL
CHI3L2	rs2820086	9,12E-007	Zeller10_CIS_QTL
CHI3L2	rs2477579	9,12E-007	Zeller10_CIS_QTL
CHI3L2	rs2764546	9,41E-006	Zeller10_CIS_QTL
CHI3L2	rs12048900	9,54E-006	StrangerEQTL
CHI3L2	rs12024633	9,54E-006	StrangerEQTL
CHI3L2	rs17699274	2,49E-005	Zeller10_CIS_QTL
CHI3L2	rs2147789	7,02E-005	Zeller10_CIS_QTL
CHI3L2	rs2255089	7,02E-005	Zeller10_CIS_QTL
CHI3L2	rs684559	7,28E-004	Zeller10_CIS_QTL
CHI3L2	rs2477574	8,70E-004	Zeller10_CIS_QTL
CHURC1	rs11848113	1,09E-009	VeyrierasEQTL
CHURC1	rs12433639	1,09E-009	VeyrierasEQTL
CHURC1	rs1951487	1,09E-009	VeyrierasEQTL
CHURC1	rs1951488	1,09E-009	VeyrierasEQTL
CHURC1	rs1957421	1,09E-009	VeyrierasEQTL
CHURC1	rs1064108	4,51E-009	VeyrierasEQTL
CHURC1	rs2737844	4,51E-009	Zeller10_CIS_QTL
CHURC1	rs2737844	4,51E-009	VeyrierasEQTL
CHURC1	rs12878229	4,51E-009	VeyrierasEQTL

CHURC1	rs4902334	4,51E-009	VeyrierasEQTL
CHURC1	rs8016174	4,51E-009	VeyrierasEQTL
CHURC1	rs10132858	1,07E-008	VeyrierasEQTL
CHURC1	rs10133054	1,07E-008	VeyrierasEQTL
CHURC1	rs10133290	1,07E-008	VeyrierasEQTL
CHURC1	rs10138534	1,07E-008	VeyrierasEQTL
CHURC1	rs10139595	1,07E-008	VeyrierasEQTL
CHURC1	rs10141986	1,07E-008	VeyrierasEQTL
CHURC1	rs10143017	1,07E-008	VeyrierasEQTL
CHURC1	rs10148687	1,07E-008	VeyrierasEQTL
CHURC1	rs10149617	1,07E-008	VeyrierasEQTL
CHURC1	rs10151701	1,07E-008	Zeller10_CIS_QTL
CHURC1	rs10151701	1,07E-008	VeyrierasEQTL
CHURC1	rs2412065	1,07E-008	VeyrierasEQTL
CHURC1	rs2412065	1,07E-008	Zeller10_CIS_QTL
CHURC1	rs17102298	1,07E-008	VeyrierasEQTL
CHURC1	rs6745	1,07E-008	VeyrierasEQTL
CHURC1	rs4902332	1,07E-008	VeyrierasEQTL
CHURC1	rs4902336	1,07E-008	VeyrierasEQTL
CHURC1	rs4902345	1,07E-008	VeyrierasEQTL
CHURC1	rs4902346	1,07E-008	VeyrierasEQTL
CHURC1	rs4902333	1,07E-008	VeyrierasEQTL
CHURC1	rs4902335	1,07E-008	VeyrierasEQTL
CHURC1	rs4899152	1,07E-008	VeyrierasEQTL
CHURC1	rs7143432	1,07E-008	VeyrierasEQTL
CHURC1	rs7144811	1,07E-008	VeyrierasEQTL
CHURC1	rs4902343	1,07E-008	VeyrierasEQTL
CHURC1	rs7143764	1,07E-008	VeyrierasEQTL
CHURC1	rs7153990	1,07E-008	VeyrierasEQTL
CHURC1	rs7154406	1,07E-008	VeyrierasEQTL
CHURC1	rs10132742	1,07E-007	VeyrierasEQTL
CHURC1	rs11623886	4,73E-007	Zeller10_CIS_QTL
CHURC1	rs1886506	1,92E-006	VeyrierasEQTL
CHURC1	rs3742599	1,92E-006	VeyrierasEQTL
CHURC1	rs4902352	4,43E-006	VeyrierasEQTL
CHURC1	rs17752452	1,60E-005	VeyrierasEQTL
CHURC1	rs11621450	1,60E-005	VeyrierasEQTL
CHURC1	rs11621450	1,60E-005	Zeller10_CIS_QTL
CHURC1	rs4902348	1,60E-005	VeyrierasEQTL
CHURC1	rs2277502	1,60E-005	VeyrierasEQTL
CHURC1	rs4902348	1,60E-005	Zeller10_CIS_QTL
CHURC1	rs4902349	1,60E-005	VeyrierasEQTL
CHURC1	rs4902350	1,60E-005	VeyrierasEQTL
CHURC1	rs4902351	1,60E-005	VeyrierasEQTL
CHURC1	rs11622077	2,83E-005	VeyrierasEQTL

CHURC1	rs12881353	2,83E-005	VeyrierasEQTL
CHURC1	rs17767960	7,37E-005	VeyrierasEQTL
CHURC1	rs11623705	7,37E-005	VeyrierasEQTL
CHURC1	rs17767942	7,37E-005	VeyrierasEQTL
CHURC1	rs4902331	7,37E-005	VeyrierasEQTL
CLIC6	rs2070367	5,14E-008	VeyrierasEQTL
CLIC6	rs2236610	1,03E-006	VeyrierasEQTL
CLIC6	rs2834586	1,03E-006	VeyrierasEQTL
CLIC6	rs7281533	5,22E-005	VeyrierasEQTL
CLIC6	rs2834574	5,22E-005	VeyrierasEQTL
CP110	rs12447418	5,38E-008	VeyrierasEQTL
CP110	rs12447418	5,38E-008	StrangerEQTL
CP110	rs2285659	5,82E-005	VeyrierasEQTL
CP110	rs12447447	5,82E-005	VeyrierasEQTL
CP110	rs7191012	5,82E-005	VeyrierasEQTL
CP110	rs7189880	5,82E-005	VeyrierasEQTL
CP110	rs12927277	5,82E-005	VeyrierasEQTL
CP110	rs7190666	5,82E-005	VeyrierasEQTL
CP110	rs11642738	5,82E-005	VeyrierasEQTL
CP110	rs2072087	5,82E-005	VeyrierasEQTL
CP110	rs12925996	5,82E-005	VeyrierasEQTL
CP110	rs12926048	5,82E-005	VeyrierasEQTL
CP110	rs12919821	5,82E-005	VeyrierasEQTL
CP110	rs2285660	5,82E-005	VeyrierasEQTL
CP110	rs12929603	5,82E-005	VeyrierasEQTL
CP110	rs2256152	5,82E-005	VeyrierasEQTL
CP110	rs2074167	5,82E-005	VeyrierasEQTL
CP110	rs2353912	5,82E-005	VeyrierasEQTL
CP110	rs8057847	5,82E-005	VeyrierasEQTL
CP110	rs11074400	5,82E-005	VeyrierasEQTL
CP110	rs7184710	5,82E-005	VeyrierasEQTL
CP110	rs11640464	5,82E-005	VeyrierasEQTL
CP110	rs2521795	5,82E-005	VeyrierasEQTL
CP110	rs16972219	5,82E-005	VeyrierasEQTL
CP110	rs8057980	6,41E-005	StrangerEQTL
CP110	rs11640077	6,41E-005	StrangerEQTL
CP110	rs2285658	6,41E-005	StrangerEQTL
CP110	rs2856608	6,41E-005	StrangerEQTL
CP110	rs2074169	6,41E-005	StrangerEQTL
CP110	rs8057970	6,41E-005	StrangerEQTL
CP110	rs2353913	6,41E-005	StrangerEQTL
CP110	rs2353696	6,41E-005	StrangerEQTL
CP110	rs1475963	6,41E-005	StrangerEQTL
CP110	rs2074172	6,41E-005	StrangerEQTL
CP110	rs2074169	6,41E-005	VeyrierasEQTL

CP110	rs7135	6,41E-005 VeyrierasEQTL
CP110	rs2009787	6,41E-005 StrangerEQTL
CP110	rs2856613	6,41E-005 StrangerEQTL
CP110	rs7135	6,41E-005 StrangerEQTL
CP110	rs3794697	6,41E-005 StrangerEQTL
DHX58	rs2074166	1,17E-010 PickrelleQTL
DNASE1L3	rs3732631	1,40E-006 VeyrierasEQTL
DNASE1L3	rs13098114	1,40E-006 VeyrierasEQTL
DNASE1L3	rs17059007	1,40E-006 VeyrierasEQTL
DNASE1L3	rs4681824	1,40E-006 Zeller10_CIS_QTL
DNASE1L3	rs4681824	1,40E-006 VeyrierasEQTL
DNASE1L3	rs4681823	1,40E-006 VeyrierasEQTL
DNASE1L3	rs4681823	1,40E-006 Zeller10_CIS_QTL
DNASE1L3	rs3772989	2,86E-006 VeyrierasEQTL
DNASE1L3	rs4681668	1,99E-005 VeyrierasEQTL
DPYSL4	rs902627	1,60E-008 VeyrierasEQTL
DPYSL4	rs2637629	2,31E-006 VeyrierasEQTL
DPYSL4	rs2637629	2,31E-006 StrangerEQTL
DPYSL4	rs923661	5,65E-006 VeyrierasEQTL
DPYSL4	rs1532186	2,02E-004 VeyrierasEQTL
DPYSL4	rs1532186	2,02E-004 StrangerEQTL
DPYSL4	rs2637632	2,02E-004 VeyrierasEQTL
DPYSL4	rs2814155	5,43E-004 VeyrierasEQTL
DPYSL4	rs2637630	5,43E-004 VeyrierasEQTL
DPYSL4	rs2637631	5,43E-004 VeyrierasEQTL
DPYSL4	rs2637630	5,43E-004 StrangerEQTL
DTNB	rs7583617	3,52E-010 StrangerEQTL
DTNB	rs7583617	3,52E-010 VeyrierasEQTL
DTNB	rs17046986	3,52E-010 VeyrierasEQTL
DTNB	rs17046986	3,52E-010 StrangerEQTL
DTNB	rs17046986	3,52E-010 PickrelleQTL
DTNB	rs1369704	3,52E-010 StrangerEQTL
DTNB	rs6546148	3,52E-010 StrangerEQTL
DTNB	rs10206012	3,52E-010 StrangerEQTL
DTNB	rs10206012	3,52E-010 VeyrierasEQTL
DTNB	rs2384232	3,52E-010 VeyrierasEQTL
\$'Tes	rs7586918	3,52E-010 VeyrierasEQTL
DTNB	rs6546148	3,52E-010 VeyrierasEQTL
DTNB	rs7586918	3,52E-010 StrangerEQTL
DTNB	rs2384232	3,52E-010 StrangerEQTL
DTNB	rs7571570	3,52E-010 VeyrierasEQTL
DTNB	rs1369704	3,52E-010 VeyrierasEQTL
DTNB	rs7571570	3,52E-010 StrangerEQTL
DTNB	rs10189164	1,42E-005 StrangerEQTL
DTNB	rs10189164	1,42E-005 VeyrierasEQTL

DTNB	rs6546149	1,86E-005	StrangerEQTL
DTNB	rs6546149	1,86E-005	VeyrierasEQTL
DTNB	rs10180663	1,33E-004	StrangerEQTL
DTNB	rs6731296	9,11E-004	VeyrierasEQTL
DTNB	rs6731296	9,11E-004	StrangerEQTL
ECHDC2	rs514881	5,91E-006	Zeller10_CIS_QTL
ECHDC2	rs481440	4,80E-005	Zeller10_CIS_QTL
ECHDC2	rs515857	7,09E-005	Zeller10_CIS_QTL
ECHDC2	rs499195	1,44E-004	Zeller10_CIS_QTL
ECHDC2	rs512723	1,44E-004	Zeller10_CIS_QTL
FAM164A	rs7017898	8,98E-007	PickrelEQTL
FLJ22374	rs7790537	1,54E-008	VeyrierasEQTL
FLJ22374	rs11972565	1,80E-008	VeyrierasEQTL
FLJ22374	rs2391919	1,05E-007	VeyrierasEQTL
FLJ22374	rs739981	1,05E-007	VeyrierasEQTL
FLJ22374	rs7781249	1,57E-007	StrangerEQTL
FLJ22374	rs7781249	1,57E-007	VeyrierasEQTL
FLJ22374	rs7803974	1,57E-007	VeyrierasEQTL
FLJ22374	rs7785179	1,57E-007	StrangerEQTL
FLJ22374	rs7785179	1,57E-007	VeyrierasEQTL
FLJ22374	rs7803974	1,57E-007	StrangerEQTL
FLJ22374	rs10278285	5,29E-007	VeyrierasEQTL
FLJ22374	rs10282606	5,29E-007	VeyrierasEQTL
FLJ22374	rs10282606	5,29E-007	StrangerEQTL
FLJ22374	rs10278285	5,29E-007	StrangerEQTL
FLJ22374	rs6979305	5,29E-007	StrangerEQTL
FLJ22374	rs6979305	5,29E-007	VeyrierasEQTL
FLJ22374	rs1990010	5,31E-007	VeyrierasEQTL
FLJ22374	rs1017012	6,05E-007	VeyrierasEQTL
FLJ22374	rs12690796	4,39E-005	VeyrierasEQTL
FLJ22374	rs12690795	4,39E-005	VeyrierasEQTL
FLJ22374	rs4723018	4,39E-005	VeyrierasEQTL
FLJ22374	rs12701041	4,39E-005	VeyrierasEQTL
FLJ45422	rs9295843	8,13E-007	Zeller10_CIS_QTL
FLJ45422	rs1473102	8,13E-007	Zeller10_CIS_QTL
FLJ45422	rs2523747	8,13E-007	Zeller10_CIS_QTL
FLJ45422	rs6924453	1,47E-005	Zeller10_CIS_QTL
FLJ45422	rs6912771	1,47E-005	Zeller10_CIS_QTL
FLJ45422	rs9357097	1,39E-004	Zeller10_CIS_QTL
FLJ45422	rs9295844	8,10E-004	Zeller10_CIS_QTL
FLJ45422	rs9378145	8,10E-004	Zeller10_CIS_QTL
GBP3	rs17434342	6,35E-009	Zeller10_CIS_QTL
GBP3	rs2624	2,56E-006	Zeller10_CIS_QTL
GCSH	rs16954582	7,63E-008	SchadtEQTL
GGTA1	rs3816094	5,27E-009	Zeller10_CIS_QTL

GGTA1	rs4837862	5,27E-009	Zeller10_CIS_QTL
GGTA1	rs10985274	9,48E-008	Zeller10_CIS_QTL
GGTA1	rs3860151	3,92E-007	Zeller10_CIS_QTL
GGTA1	rs10818548	6,95E-006	Zeller10_CIS_QTL
GGTA1	rs10818555	6,95E-006	Zeller10_CIS_QTL
GGTA1	rs17417696	6,95E-006	Zeller10_CIS_QTL
GGTA1	rs7045245	6,95E-006	Zeller10_CIS_QTL
GGTA1	rs7873990	3,96E-005	Zeller10_CIS_QTL
GGTA1	rs10985284	3,96E-005	Zeller10_CIS_QTL
GGTA1	rs4837866	3,96E-005	Zeller10_CIS_QTL
GGTA1	rs10985285	3,96E-005	Zeller10_CIS_QTL
GGTA1	rs10985287	1,35E-004	Zeller10_CIS_QTL
GNMT	rs2296804	8,09E-008	VeyrierasEQTL
GNMT	rs2296804	8,09E-008	StrangerEQTL
GNMT	rs2395943	8,09E-008	VeyrierasEQTL
GNMT	rs2395943	8,09E-008	StrangerEQTL
GNMT	rs3818554	2,78E-007	Zeller10_CIS_QTL
GNMT	rs2274517	2,78E-007	Zeller10_CIS_QTL
GNMT	rs9986447	2,90E-006	Zeller10_CIS_QTL
GNMT	rs6941212	3,03E-006	Zeller10_CIS_QTL
GUCA1B	rs9381157	5,86E-011	VeyrierasEQTL
GUCA1B	rs6912658	1,73E-010	VeyrierasEQTL
GUCA1B	rs2894443	1,73E-010	VeyrierasEQTL
GUCA1B	rs2395803	1,73E-010	VeyrierasEQTL
GUCA1B	rs9394885	1,73E-010	VeyrierasEQTL
GUCA1B	rs2180172	1,73E-010	Zeller10_CIS_QTL
GUCA1B	rs9369346	1,73E-010	Zeller10_CIS_QTL
GUCA1B	rs2180174	1,73E-010	Zeller10_CIS_QTL
GUCA1B	rs4714578	1,73E-010	VeyrierasEQTL
GUCA1B	rs9357390	1,73E-010	VeyrierasEQTL
GUCA1B	rs6912658	1,73E-010	Zeller10_CIS_QTL
GUCA1B	rs2395804	1,73E-010	VeyrierasEQTL
GUCA1B	rs2295277	1,73E-010	VeyrierasEQTL
GUCA1B	rs9369346	1,73E-010	VeyrierasEQTL
GUCA1B	rs9369350	1,73E-010	VeyrierasEQTL
GUCA1B	rs2180174	1,73E-010	VeyrierasEQTL
GUCA1B	rs4714576	1,73E-010	VeyrierasEQTL
GUCA1B	rs2144444	1,73E-010	VeyrierasEQTL
GUCA1B	rs9369347	1,73E-010	VeyrierasEQTL
GUCA1B	rs2180172	1,73E-010	VeyrierasEQTL
GUCA1B	rs3749921	1,73E-010	VeyrierasEQTL
GUCA1B	rs6925368	1,73E-010	VeyrierasEQTL
GUCA1B	rs13217993	1,73E-010	VeyrierasEQTL
GUCA1B	rs4714577	1,73E-010	VeyrierasEQTL
HEBP2	rs9495054	2,71E-007	VeyrierasEQTL

HEBP2	rs9495052	2,71E-007	VeyrierasEQTL
HEBP2	rs9484151	2,71E-007	StrangerEQTL
HEBP2	rs6918995	2,71E-007	VeyrierasEQTL
HEBP2	rs17053774	2,71E-007	StrangerEQTL
HEBP2	rs11966257	2,71E-007	StrangerEQTL
HEBP2	rs9495054	2,71E-007	StrangerEQTL
HEBP2	rs9495052	2,71E-007	StrangerEQTL
HEBP2	rs9495064	2,71E-007	VeyrierasEQTL
HEBP2	rs9484152	2,71E-007	VeyrierasEQTL
HEBP2	rs9495066	2,71E-007	VeyrierasEQTL
HEBP2	rs7753628	2,71E-007	VeyrierasEQTL
HEBP2	rs9495065	2,71E-007	StrangerEQTL
HEBP2	rs9484151	2,71E-007	VeyrierasEQTL
HEBP2	rs9495066	2,71E-007	StrangerEQTL
HEBP2	rs9495065	2,71E-007	VeyrierasEQTL
HEBP2	rs17053774	2,71E-007	VeyrierasEQTL
HEBP2	rs7748815	2,71E-007	VeyrierasEQTL
HEBP2	rs11966257	2,71E-007	VeyrierasEQTL
HEBP2	rs9495063	2,71E-007	VeyrierasEQTL
HEBP2	rs11970756	2,71E-007	VeyrierasEQTL
HEBP2	rs9495064	2,71E-007	StrangerEQTL
HEBP2	rs7748815	2,71E-007	StrangerEQTL
HEBP2	rs7765991	2,71E-007	StrangerEQTL
HEBP2	rs7753628	2,71E-007	StrangerEQTL
HEBP2	rs9495063	2,71E-007	StrangerEQTL
HEBP2	rs2076272	2,71E-007	VeyrierasEQTL
HEBP2	rs2076273	2,71E-007	VeyrierasEQTL
HEBP2	rs7765991	2,71E-007	VeyrierasEQTL
HEBP2	rs11970756	2,71E-007	StrangerEQTL
HEBP2	rs2076273	2,71E-007	StrangerEQTL
HEBP2	rs12181239	2,71E-007	VeyrierasEQTL
HEBP2	rs2076272	2,71E-007	StrangerEQTL
HEBP2	rs6918995	2,71E-007	StrangerEQTL
HEBP2	rs12181239	2,71E-007	StrangerEQTL
HEBP2	rs9484152	2,71E-007	StrangerEQTL
HEBP2	rs6570228	1,12E-005	StrangerEQTL
HEBP2	rs9495068	1,12E-005	StrangerEQTL
HEBP2	rs9495068	1,12E-005	VeyrierasEQTL
HEBP2	rs6570228	1,12E-005	VeyrierasEQTL
HEBP2	rs9495070	1,12E-005	VeyrierasEQTL
HEBP2	rs9495070	1,12E-005	StrangerEQTL
HEBP2	rs9495051	8,06E-004	VeyrierasEQTL
HEBP2	rs10223504	8,06E-004	StrangerEQTL
HEBP2	rs17067480	8,06E-004	VeyrierasEQTL
HEBP2	rs9689169	8,06E-004	StrangerEQTL

HEBP2	rs9495053	8,06E-004	VeyrierasEQTL
HEBP2	rs9689169	8,06E-004	VeyrierasEQTL
HEBP2	rs9495051	8,06E-004	StrangerEQTL
HEBP2	rs11966129	8,06E-004	VeyrierasEQTL
HEBP2	rs9484143	8,06E-004	StrangerEQTL
HEBP2	rs7752490	8,06E-004	StrangerEQTL
HEBP2	rs9495050	8,06E-004	VeyrierasEQTL
HEBP2	rs9495045	8,06E-004	VeyrierasEQTL
HEBP2	rs11966129	8,06E-004	StrangerEQTL
HEBP2	rs9495053	8,06E-004	StrangerEQTL
HEBP2	rs17067480	8,06E-004	StrangerEQTL
HEBP2	rs7756927	8,06E-004	StrangerEQTL
HEBP2	rs17067484	8,06E-004	StrangerEQTL
HEBP2	rs17067484	8,06E-004	VeyrierasEQTL
HEBP2	rs9495050	8,06E-004	StrangerEQTL
HEBP2	rs9495030	8,06E-004	StrangerEQTL
HEBP2	rs9495045	8,06E-004	StrangerEQTL
HEBP2	rs6933880	8,06E-004	VeyrierasEQTL
HEBP2	rs6933880	8,06E-004	StrangerEQTL
HEBP2	rs9495030	8,06E-004	VeyrierasEQTL
HEBP2	rs9495025	8,06E-004	VeyrierasEQTL
HEBP2	rs9484143	8,06E-004	VeyrierasEQTL
HEBP2	rs9495025	8,06E-004	StrangerEQTL
HEBP2	rs17067484	8,06E-004	Zeller10_CIS_QTL
HEBP2	rs7756927	8,06E-004	VeyrierasEQTL
HEBP2	rs7752490	8,06E-004	VeyrierasEQTL
HEBP2	rs10223504	8,06E-004	VeyrierasEQTL
LILRA2	rs12983808	3,16E-006	Zeller10_CIS_QTL
LILRA2	rs12459217	1,58E-005	Zeller10_CIS_QTL
LILRA2	rs12461246	1,77E-004	Zeller10_CIS_QTL
LILRA2	rs12980993	2,21E-004	Zeller10_CIS_QTL
LILRA2	rs8106636	2,21E-004	Zeller10_CIS_QTL
LILRA2	rs2555687	2,21E-004	Zeller10_CIS_QTL
LILRA2	rs16985743	2,21E-004	Zeller10_CIS_QTL
LOC400642	rs8092760	2,96E-007	StrangerEQTL
LOC400642	rs4429380	2,96E-007	StrangerEQTL
LOC400642	rs17379984	2,96E-007	StrangerEQTL
LOC400642	rs11877479	2,96E-007	StrangerEQTL
LOC400642	rs11877315	2,96E-007	StrangerEQTL
LOC400642	rs12352	2,96E-007	StrangerEQTL
LOC400642	rs8089069	3,70E-006	StrangerEQTL
LOC400642	rs8092448	6,63E-005	StrangerEQTL
LOC400642	rs4798330	6,63E-005	StrangerEQTL
LOC400642	rs1062837	7,42E-005	StrangerEQTL
LOC400642	rs8576	2,63E-004	StrangerEQTL

LOC400642	rs8084872	2,63E-004	StrangerEQTL
LOC400642	rs9948693	2,63E-004	StrangerEQTL
MPHOSPH6	rs4889476	8,49E-012	Zeller10_CIS_QTL
MPHOSPH6	rs2967370	8,35E-011	Zeller10_CIS_QTL
MPHOSPH6	rs2288050	2,88E-010	Zeller10_CIS_QTL
MPHOSPH6	rs2911391	4,72E-010	Zeller10_CIS_QTL
MPHOSPH6	rs1862820	4,72E-010	Zeller10_CIS_QTL
MPHOSPH6	rs918761	1,69E-007	Zeller10_CIS_QTL
MPHOSPH6	rs2967381	1,70E-007	Zeller10_CIS_QTL
MPHOSPH6	rs918763	5,28E-007	Zeller10_CIS_QTL
MPHOSPH6	rs2967355	4,67E-004	VeyrierasEQTL
MPHOSPH6	rs7192444	4,67E-004	Zeller10_CIS_QTL
MPHOSPH6	rs2967374	4,67E-004	VeyrierasEQTL
MPHOSPH6	rs11641736	4,67E-004	VeyrierasEQTL
MPHOSPH6	rs2967355	4,67E-004	Zeller10_CIS_QTL
MPHOSPH6	rs6564996	4,67E-004	VeyrierasEQTL
MPHOSPH6	rs12597171	9,61E-004	Zeller10_CIS_QTL
MPHOSPH6	rs7205534	9,61E-004	Zeller10_CIS_QTL
MPHOSPH6	rs12597534	9,61E-004	VeyrierasEQTL
MPHOSPH6	rs12597171	9,61E-004	VeyrierasEQTL
MPHOSPH6	rs10514532	9,61E-004	Zeller10_CIS_QTL
MPHOSPH6	rs12597534	9,61E-004	SchadtEQTL
MPHOSPH6	rs2967359	9,61E-004	VeyrierasEQTL
MPHOSPH6	rs12597534	9,61E-004	Zeller10_CIS_QTL
MPHOSPH6	rs7205534	9,61E-004	VeyrierasEQTL
MPHOSPH6	rs10514532	9,61E-004	VeyrierasEQTL
NT5E	rs4487548	4,95E-009	VeyrierasEQTL
NT5E	rs1570311	4,95E-009	VeyrierasEQTL
NT5E	rs7772233	4,95E-009	VeyrierasEQTL
NT5E	rs12662033	4,95E-009	VeyrierasEQTL
NT5E	rs7772233	4,95E-009	PickrelEQTL
NT5E	rs7754559	4,95E-009	VeyrierasEQTL
NT5E	rs9362213	4,95E-009	VeyrierasEQTL
NT5E	rs9359654	4,95E-009	VeyrierasEQTL
NT5E	rs9344525	4,95E-009	VeyrierasEQTL
NT5E	rs1000633	4,95E-009	VeyrierasEQTL
NT5E	rs4144167	4,95E-009	VeyrierasEQTL
NT5E	rs9344521	4,95E-009	VeyrierasEQTL
NT5E	rs1322401	4,95E-009	VeyrierasEQTL
NT5E	rs4400191	4,95E-009	VeyrierasEQTL
NT5E	rs9362218	4,95E-009	VeyrierasEQTL
NT5E	rs9344528	2,91E-008	VeyrierasEQTL
PGBD4	rs4780215	1,08E-006	Zeller10_CIS_QTL
PGBD4	rs4780216	1,08E-006	Zeller10_CIS_QTL
RIBC2	rs717295	7,58E-011	VeyrierasEQTL

RIBC2	rs1022480	4,22E-010	VeyrierasEQTL
RIBC2	rs1022481	4,22E-010	VeyrierasEQTL
RIBC2	rs1022477	4,22E-010	VeyrierasEQTL
RIBC2	rs738227	4,22E-010	VeyrierasEQTL
RIBC2	rs2281013	4,22E-010	VeyrierasEQTL
RIBC2	rs5764725	4,22E-010	VeyrierasEQTL
RIBC2	rs5764729	4,22E-010	VeyrierasEQTL
RIBC2	rs9754402	4,22E-010	VeyrierasEQTL
RIBC2	rs5765356	4,22E-010	VeyrierasEQTL
RIBC2	rs5765359	4,22E-010	VeyrierasEQTL
RIBC2	rs9614665	4,22E-010	VeyrierasEQTL
RIBC2	rs738229	4,22E-010	VeyrierasEQTL
RIBC2	rs2072770	4,22E-010	VeyrierasEQTL
RIBC2	rs6007031	4,22E-010	VeyrierasEQTL
RIBC2	rs5765341	4,22E-010	VeyrierasEQTL
RIBC2	rs5765352	4,22E-010	VeyrierasEQTL
RIBC2	rs6007030	4,22E-010	VeyrierasEQTL
RIBC2	rs9614468	9,05E-010	VeyrierasEQTL
RIBC2	rs5764701	1,71E-008	VeyrierasEQTL
RIBC2	rs5764698	1,71E-008	VeyrierasEQTL
RIBC2	rs5765275	1,71E-008	VeyrierasEQTL
RIBC2	rs1012493	1,71E-008	VeyrierasEQTL
RIBC2	rs5765261	1,71E-008	VeyrierasEQTL
RIBC2	rs5765273	1,71E-008	VeyrierasEQTL
RIBC2	rs13057793	1,71E-008	VeyrierasEQTL
RIBC2	rs2064068	1,71E-008	VeyrierasEQTL
RIBC2	rs9614457	1,71E-008	VeyrierasEQTL
RIBC2	rs9614462	1,71E-008	VeyrierasEQTL
RIBC2	rs719925	1,71E-008	VeyrierasEQTL
RIBC2	rs2272804	2,44E-008	PickrelleQTL
RIBC2	rs2072769	6,20E-008	VeyrierasEQTL
RIBC2	rs10343	1,14E-006	VeyrierasEQTL
RLN1	rs4742081	1,85E-012	VeyrierasEQTL
RLN1	rs2273782	1,85E-012	VeyrierasEQTL
RLN1	rs1885498	1,85E-012	VeyrierasEQTL
RLN1	rs7858028	9,43E-011	VeyrierasEQTL
RLN1	rs4740813	2,57E-005	VeyrierasEQTL
RLN1	rs7873062	2,57E-005	VeyrierasEQTL
RLN1	rs3758240	7,56E-005	VeyrierasEQTL
RLN1	rs4237161	6,54E-004	VeyrierasEQTL
RLN1	rs16923036	6,54E-004	VeyrierasEQTL
RLN1	rs4742083	8,46E-004	VeyrierasEQTL
RLN1	rs7048887	8,46E-004	VeyrierasEQTL
RLN1	rs1885499	8,46E-004	VeyrierasEQTL
RLN1	rs1322220	8,46E-004	VeyrierasEQTL

RLN1	rs7859081	8,46E-004 VeyrierasEQTL
RPS23	rs4703879	4,16E-007 Zeller10_CIS_QTL
RPS23	rs12523353	2,41E-004 Zeller10_CIS_QTL
RPS23	rs1864181	2,41E-004 Zeller10_CIS_QTL
RPS23	rs10036937	2,41E-004 Zeller10_CIS_QTL
RPS23	rs1864183	7,46E-004 Zeller10_CIS_QTL
RPS23	rs2195448	7,46E-004 Zeller10_CIS_QTL
SEPT10	rs9330320	2,81E-005 Zeller10_CIS_QTL
SEPT10	rs12998808	2,81E-005 Zeller10_CIS_QTL
SIGLEC12	rs6509544	3,14E-008 Zeller10_CIS_QTL
SIGLEC12	rs3752135	3,14E-008 Zeller10_CIS_QTL
SIGLEC12	rs11084083	1,69E-005 Zeller10_CIS_QTL
SIGLEC12	rs10405214	2,00E-005 Zeller10_CIS_QTL
SIGLEC12	rs2864105	6,54E-004 Zeller10_CIS_QTL
SIGLEC12	rs2864104	7,25E-004 Zeller10_CIS_QTL
SIGLEC12	rs9304718	7,25E-004 Zeller10_CIS_QTL
SIGLEC12	rs9304718	7,25E-004 SchadtEQTL
SIGLEC12	rs7253637	7,25E-004 Zeller10_CIS_QTL
SIGLEC12	rs2009362	8,71E-004 Zeller10_CIS_QTL
SLC5A11	rs1465423	8,98E-007 Zeller10_CIS_QTL
SLC5A11	rs11639856	8,98E-007 Zeller10_CIS_QTL
SLFN5	rs883416	7,63E-012 PickrellEQTL
STOX1	rs17470053	2,62E-007 VeyrierasEQTL
STOX1	rs12243354	3,91E-007 Zeller10_CIS_QTL
STOX1	rs7069305	3,91E-007 Zeller10_CIS_QTL
STOX1	rs10998374	3,98E-006 Zeller10_CIS_QTL
STOX1	rs7897916	3,98E-006 Zeller10_CIS_QTL
STOX1	rs11818897	1,91E-005 VeyrierasEQTL
STOX1	rs11818897	1,91E-005 Zeller10_CIS_QTL
TBKBP1	rs9913503	2,32E-011 VeyrierasEQTL
TBKBP1	rs8078309	2,32E-011 VeyrierasEQTL
TBKBP1	rs4794053	2,32E-011 VeyrierasEQTL
TBKBP1	rs6503807	2,32E-011 VeyrierasEQTL
TBKBP1	rs1962307	1,56E-010 Zeller10_CIS_QTL
TBKBP1	rs1962307	1,56E-010 VeyrierasEQTL
TBKBP1	rs11079783	1,05E-009 Zeller10_CIS_QTL
TBKBP1	rs7214799	3,92E-008 Zeller10_CIS_QTL
TBKBP1	rs4267364	4,11E-008 Zeller10_CIS_QTL
TBKBP1	rs4267364	4,11E-008 VeyrierasEQTL
TBKBP1	rs8070463	1,48E-007 Zeller10_CIS_QTL
TBKBP1	rs11079784	4,30E-007 Zeller10_CIS_QTL
TBKBP1	rs4794047	7,37E-007 Zeller10_CIS_QTL
TBKBP1	rs4794048	6,42E-006 Zeller10_CIS_QTL
TGM5	rs531910	9,85E-012 StrangerEQTL
TGM5	rs572959	9,85E-012 StrangerEQTL

TGM5	rs574065	9,85E-012	StrangerEQTL
TGM5	rs524547	9,85E-012	StrangerEQTL
TGM5	rs493177	9,85E-012	StrangerEQTL
TGM5	rs554001	9,85E-012	StrangerEQTL
TGM5	rs471122	9,85E-012	StrangerEQTL
TGM5	rs504417	9,85E-012	StrangerEQTL
TGM5	rs555001	9,85E-012	StrangerEQTL
TGM5	rs501884	9,85E-012	StrangerEQTL
TGM5	rs12909095	1,28E-006	StrangerEQTL
TGM5	rs12911132	1,28E-006	StrangerEQTL
TGM5	rs748404	1,28E-006	StrangerEQTL
TLR1	rs5743565	7,19E-013	VeyrierasEQTL
TLR1	rs5743557	7,19E-013	VeyrierasEQTL
TLR1	rs5743595	7,19E-013	VeyrierasEQTL
TLR1	rs11725309	7,19E-013	VeyrierasEQTL
TLR1	rs5743592	7,19E-013	VeyrierasEQTL
TLR1	rs5743563	7,19E-013	VeyrierasEQTL
TLR1	rs5743592	7,19E-013	SchadtEQTL
TLR1	rs11466617	1,49E-009	VeyrierasEQTL
TLR1	rs11466640	1,49E-009	VeyrierasEQTL
TLR1	rs11466645	1,49E-009	VeyrierasEQTL
TLR1	rs4274855	1,49E-009	VeyrierasEQTL
TLR1	rs11096956	3,49E-008	StrangerEQTL
TLR1	rs7653908	3,49E-008	VeyrierasEQTL
TSGA10	rs1133977	7,87E-008	Zeller10_CIS_QTL
TSGA10	rs1133977	7,87E-008	StrangerEQTL
TSGA10	rs13033012	7,87E-008	StrangerEQTL
TSGA10	rs2516835	7,87E-008	SchadtEQTL
TSGA10	rs2516835	7,87E-008	StrangerEQTL
TSGA10	rs11683188	7,87E-008	StrangerEQTL
TSGA10	rs6542866	7,87E-008	StrangerEQTL
TSGA10	rs12613624	7,87E-008	Zeller10_CIS_QTL
TSGA10	rs12613624	7,87E-008	StrangerEQTL
TSGA10	rs17759607	7,87E-008	StrangerEQTL
TSGA10	rs11688004	7,87E-008	StrangerEQTL
TSGA10	rs4850897	7,87E-008	StrangerEQTL
TSGA10	rs720229	7,87E-008	StrangerEQTL
TSGA10	rs1453563	7,87E-008	StrangerEQTL
TSGA10	rs2516830	7,87E-008	StrangerEQTL
TSGA10	rs4851188	7,87E-008	StrangerEQTL
TSGA10	rs2632283	7,87E-008	StrangerEQTL
TSGA10	rs6733285	7,87E-008	StrangerEQTL
TSGA10	rs13008171	7,87E-008	StrangerEQTL
TSGA10	rs6727749	7,87E-008	StrangerEQTL
TSGA10	rs1563108	3,21E-006	Zeller10_CIS_QTL

TSGA10	rs1563107	3,21E-006	Zeller10_CIS_QTL
TSGA10	rs1563107	3,21E-006	StrangerEQTL
TSGA10	rs1563108	3,21E-006	StrangerEQTL
TSGA10	rs2516832	3,50E-006	StrangerEQTL
TSGA10	rs10865029	3,50E-006	StrangerEQTL
TSGA10	rs11123765	3,50E-006	StrangerEQTL
TSGA10	rs8179690	3,50E-006	Zeller10_CIS_QTL
TSGA10	rs11692699	3,50E-006	StrangerEQTL
TSGA10	rs11689265	3,50E-006	StrangerEQTL
TSGA10	rs13035311	3,50E-006	StrangerEQTL
TSGA10	rs2632277	3,50E-006	Zeller10_CIS_QTL
TSGA10	rs11695379	3,50E-006	StrangerEQTL
TSGA10	rs11883676	3,50E-006	StrangerEQTL
TSGA10	rs2516827	3,50E-006	StrangerEQTL
TSGA10	rs2516829	3,50E-006	StrangerEQTL
TSGA10	rs4851187	3,50E-006	StrangerEQTL
TSGA10	rs2516834	3,50E-006	StrangerEQTL
TSGA10	rs2171359	3,50E-006	StrangerEQTL
TSGA10	rs12995473	3,50E-006	StrangerEQTL
TSGA10	rs6542864	3,50E-006	Zeller10_CIS_QTL
TSGA10	rs2632260	3,50E-006	StrangerEQTL
TSGA10	rs11689265	3,50E-006	Zeller10_CIS_QTL
TSGA10	rs13008496	3,50E-006	StrangerEQTL
TSGA10	rs13030961	3,50E-006	StrangerEQTL
TSGA10	rs4340581	3,50E-006	StrangerEQTL
TSGA10	rs6708576	3,50E-006	StrangerEQTL
TSGA10	rs11695379	3,50E-006	Zeller10_CIS_QTL
TSGA10	rs6741772	3,50E-006	StrangerEQTL
TSGA10	rs2632277	3,50E-006	StrangerEQTL
TSGA10	rs7577477	3,50E-006	StrangerEQTL
TSGA10	rs6751927	3,50E-006	StrangerEQTL
TSGA10	rs6542864	3,50E-006	StrangerEQTL
TSGA10	rs4851187	3,50E-006	Zeller10_CIS_QTL
TSGA10	rs6542865	3,50E-006	StrangerEQTL
TSGA10	rs8179690	3,50E-006	StrangerEQTL
TSGA10	rs6753063	3,50E-006	StrangerEQTL
TSGA10	rs6756650	3,50E-006	StrangerEQTL
TSGA10	rs10181344	3,01E-005	StrangerEQTL
TSGA10	rs12470471	3,01E-005	Zeller10_CIS_QTL
TSGA10	rs12470471	3,01E-005	StrangerEQTL
TSGA10	rs2123686	3,01E-005	StrangerEQTL
TSGA10	rs17708262	3,01E-005	StrangerEQTL
TSGA10	rs6737257	3,01E-005	StrangerEQTL
TSGA10	rs9710648	3,01E-005	StrangerEQTL
TSGA10	rs6754843	3,01E-005	StrangerEQTL

TSGA10	rs6750876	3,01E-005	StrangerEQTL
TSGA10	rs6756880	3,01E-005	StrangerEQTL
TSGA10	rs2045057	5,96E-005	PickrelEQTL
TUBB6	rs2298542	5,00E-007	VeyrierasEQTL
TUBB6	rs9956171	5,00E-007	VeyrierasEQTL
TUBB6	rs9303770	5,00E-007	VeyrierasEQTL
TUBB6	rs9956171	5,00E-007	Zeller10_CIS_QTL
TUBB6	rs2298543	5,00E-007	VeyrierasEQTL
TUBB6	rs4797680	5,00E-007	VeyrierasEQTL
TUBB6	rs4797677	5,89E-007	VeyrierasEQTL
TUBB6	rs7235296	1,80E-006	VeyrierasEQTL
TUBB6	rs7235296	1,80E-006	Zeller10_CIS_QTL
TUBB6	rs11662523	1,80E-006	VeyrierasEQTL
TUBB6	rs8097342	1,80E-006	VeyrierasEQTL
TUBB6	rs7407640	1,80E-006	VeyrierasEQTL
TUBB6	rs11080572	1,80E-006	VeyrierasEQTL
TUBB6	rs7231304	2,21E-006	VeyrierasEQTL
TUBB6	rs9966470	2,21E-006	VeyrierasEQTL
TUBB6	rs11664216	2,68E-004	VeyrierasEQTL
VENTX	rs6537595	1,16E-004	Zeller10_CIS_QTL
VNN3	rs7752016	7,28E-006	Zeller10_CIS_QTL
VNN3	rs4895943	7,28E-006	Zeller10_CIS_QTL
VNN3	rs6924414	7,28E-006	Zeller10_CIS_QTL
VNN3	rs6569837	1,24E-004	Zeller10_CIS_QTL
VNN3	rs6942306	1,24E-004	Zeller10_CIS_QTL
WARS	rs11624738	7,35E-007	PickrelEQTL
WBSCR27	rs1073909	1,26E-011	StrangerEQTL
WBSCR27	rs4304218	1,26E-011	VeyrierasEQTL
WBSCR27	rs13232463	1,26E-011	StrangerEQTL
WBSCR27	rs1127155	1,26E-011	Zeller10_CIS_QTL
WBSCR27	rs1073909	1,26E-011	VeyrierasEQTL
WBSCR27	rs13232463	1,26E-011	VeyrierasEQTL
WBSCR27	rs4255023	1,26E-011	StrangerEQTL
WBSCR27	rs4255023	1,26E-011	VeyrierasEQTL
WBSCR27	rs8629	1,26E-011	Zeller10_CIS_QTL
WBSCR27	rs1127155	1,26E-011	VeyrierasEQTL
WBSCR27	rs4304218	1,26E-011	StrangerEQTL
WBSCR27	rs8629	1,26E-011	VeyrierasEQTL
WBSCR27	rs1127155	1,26E-011	StrangerEQTL
WBSCR27	rs8629	1,26E-011	StrangerEQTL
WBSCR27	rs11316	3,61E-011	Zeller10_CIS_QTL
WBSCR27	rs11316	3,61E-011	StrangerEQTL
WBSCR27	rs11316	3,61E-011	VeyrierasEQTL
WBSCR27	rs6946037	2,15E-009	VeyrierasEQTL
WBSCR27	rs6946037	2,15E-009	PickrelEQTL

WBSCR27	rs6946037	2,15E-009	StrangerEQTL
WBSCR27	rs7455225	1,03E-006	VeyrierasEQTL
WBSCR27	rs7455225	1,03E-006	StrangerEQTL
WBSCR27	rs10949830	2,75E-004	StrangerEQTL
WBSCR27	rs10949830	2,75E-004	VeyrierasEQTL
XKR9	rs2639901	3,89E-012	VeyrierasEQTL
XKR9	rs1493198	3,89E-012	VeyrierasEQTL
XKR9	rs2732087	3,89E-012	VeyrierasEQTL
XKR9	rs3098866	3,89E-012	VeyrierasEQTL
XKR9	rs3098867	3,89E-012	VeyrierasEQTL
XKR9	rs3110239	3,89E-012	VeyrierasEQTL
XKR9	rs2732091	3,89E-012	VeyrierasEQTL
XKR9	rs2732090	3,89E-012	VeyrierasEQTL
XKR9	rs10089583	7,81E-011	VeyrierasEQTL
XKR9	rs10091471	7,81E-011	VeyrierasEQTL
XKR9	rs1389206	7,81E-011	VeyrierasEQTL
XKR9	rs2639898	7,81E-011	VeyrierasEQTL
XKR9	rs2639899	7,81E-011	VeyrierasEQTL
XKR9	rs1471114	7,81E-011	VeyrierasEQTL
XKR9	rs10504481	7,81E-011	VeyrierasEQTL
XKR9	rs2639953	7,81E-011	VeyrierasEQTL
XKR9	rs1596566	7,81E-011	VeyrierasEQTL
XKR9	rs3098874	7,81E-011	VeyrierasEQTL
XKR9	rs11993453	7,81E-011	VeyrierasEQTL
XKR9	rs3098879	7,81E-011	VeyrierasEQTL
XKR9	rs12542300	7,81E-011	VeyrierasEQTL
XKR9	rs1908035	7,81E-011	VeyrierasEQTL
XKR9	rs1908036	7,81E-011	VeyrierasEQTL
XKR9	rs12680166	7,81E-011	VeyrierasEQTL
XKR9	rs12680988	7,81E-011	VeyrierasEQTL
XKR9	rs13256250	7,81E-011	VeyrierasEQTL
XKR9	rs13258584	7,81E-011	VeyrierasEQTL
XKR9	rs2732104	7,81E-011	VeyrierasEQTL
XKR9	rs2732106	7,81E-011	VeyrierasEQTL
XKR9	rs13271442	7,81E-011	VeyrierasEQTL
XKR9	rs2732110	7,81E-011	VeyrierasEQTL
XKR9	rs1348535	7,81E-011	VeyrierasEQTL
XKR9	rs1353057	7,81E-011	VeyrierasEQTL
XKR9	rs1353058	7,81E-011	VeyrierasEQTL
XKR9	rs2732115	7,81E-011	VeyrierasEQTL
XKR9	rs2732116	7,81E-011	VeyrierasEQTL
XKR9	rs2639896	7,81E-011	VeyrierasEQTL
XKR9	rs2639897	7,81E-011	VeyrierasEQTL
XKR9	rs3098876	7,81E-011	VeyrierasEQTL
XKR9	rs6472538	7,81E-011	VeyrierasEQTL

XKR9	rs1493203	7,81E-011	VeyrierasEQTL
XKR9	rs2639952	7,81E-011	VeyrierasEQTL
XKR9	rs2639954	7,81E-011	VeyrierasEQTL
XKR9	rs3098873	7,81E-011	VeyrierasEQTL
XKR9	rs2732108	7,81E-011	VeyrierasEQTL
XKR9	rs3098877	7,81E-011	VeyrierasEQTL
XKR9	rs1908033	7,81E-011	VeyrierasEQTL
XKR9	rs1908034	7,81E-011	VeyrierasEQTL
XKR9	rs3098883	7,81E-011	VeyrierasEQTL
XKR9	rs3110244	7,81E-011	VeyrierasEQTL
XKR9	rs2732094	7,81E-011	VeyrierasEQTL
XKR9	rs3110253	7,81E-011	VeyrierasEQTL
XKR9	rs2732095	7,81E-011	VeyrierasEQTL
XKR9	rs2732096	7,81E-011	VeyrierasEQTL
XKR9	rs2732114	7,81E-011	VeyrierasEQTL
XKR9	rs2732109	7,81E-011	VeyrierasEQTL
XKR9	rs6472543	7,81E-011	VeyrierasEQTL
XKR9	rs2732112	7,81E-011	VeyrierasEQTL
XKR9	rs2732113	7,81E-011	VeyrierasEQTL
XKR9	rs4434651	7,81E-011	VeyrierasEQTL
XKR9	rs4504667	7,81E-011	VeyrierasEQTL
XKR9	rs4571761	7,81E-011	VeyrierasEQTL
XKR9	rs2956610	7,81E-011	VeyrierasEQTL
XKR9	rs3098868	7,81E-011	VeyrierasEQTL
XKR9	rs4738105	7,81E-011	VeyrierasEQTL
XKR9	rs2639957	7,81E-011	VeyrierasEQTL
XKR9	rs2639958	7,81E-011	VeyrierasEQTL
XKR9	rs2639951	7,81E-011	VeyrierasEQTL
XKR9	rs6472561	7,81E-011	VeyrierasEQTL
XKR9	rs3098880	7,81E-011	VeyrierasEQTL
XKR9	rs2639956	7,81E-011	VeyrierasEQTL
XKR9	rs2732102	7,81E-011	VeyrierasEQTL
XKR9	rs2732103	7,81E-011	VeyrierasEQTL
XKR9	rs6993115	7,81E-011	VeyrierasEQTL
XKR9	rs2732093	7,81E-011	VeyrierasEQTL
XKR9	rs3110240	7,81E-011	VeyrierasEQTL
XKR9	rs7012404	7,81E-011	VeyrierasEQTL
XKR9	rs3110255	7,81E-011	VeyrierasEQTL
XKR9	rs3098871	7,81E-011	VeyrierasEQTL
XKR9	rs4074909	7,81E-011	VeyrierasEQTL
XKR9	rs4078067	7,81E-011	VeyrierasEQTL
XKR9	rs3110251	7,81E-011	VeyrierasEQTL
XKR9	rs6998786	7,81E-011	VeyrierasEQTL
XKR9	rs6993170	7,81E-011	VeyrierasEQTL
XKR9	rs7819332	7,81E-011	VeyrierasEQTL

XKR9	rs6991329	7,81E-011	VeyrierasEQTL
XKR9	rs3110250	7,81E-011	VeyrierasEQTL
XKR9	rs7826483	7,81E-011	VeyrierasEQTL
XKR9	rs3110252	7,81E-011	VeyrierasEQTL
XKR9	rs3110258	7,81E-011	VeyrierasEQTL
XKR9	rs3110256	7,81E-011	VeyrierasEQTL
XKR9	rs7017914	7,81E-011	VeyrierasEQTL
XKR9	rs7014012	7,81E-011	VeyrierasEQTL
XKR9	rs7825446	7,81E-011	VeyrierasEQTL
XKR9	rs2639955	6,17E-010	VeyrierasEQTL
XKR9	rs2732107	6,17E-010	VeyrierasEQTL
XKR9	rs10504479	8,18E-010	VeyrierasEQTL
XKR9	rs1993862	8,18E-010	VeyrierasEQTL
XKR9	rs2639911	4,45E-008	VeyrierasEQTL
XKR9	rs1444339	4,45E-008	VeyrierasEQTL
XKR9	rs2639937	4,45E-008	VeyrierasEQTL
XKR9	rs2639938	4,45E-008	VeyrierasEQTL
XKR9	rs2639945	4,45E-008	VeyrierasEQTL
XKR9	rs10504487	4,45E-008	VeyrierasEQTL
XKR9	rs1596565	4,45E-008	VeyrierasEQTL
XKR9	rs11987501	4,45E-008	VeyrierasEQTL
XKR9	rs1845217	4,45E-008	VeyrierasEQTL
XKR9	rs2732089	4,45E-008	VeyrierasEQTL
XKR9	rs12681420	4,45E-008	VeyrierasEQTL
XKR9	rs2017035	4,45E-008	VeyrierasEQTL
XKR9	rs2035924	4,45E-008	VeyrierasEQTL
XKR9	rs13256093	4,45E-008	VeyrierasEQTL
XKR9	rs2732118	4,45E-008	VeyrierasEQTL
XKR9	rs2732140	4,45E-008	VeyrierasEQTL
XKR9	rs2732144	4,45E-008	VeyrierasEQTL
XKR9	rs2732146	4,45E-008	VeyrierasEQTL
XKR9	rs2639924	4,45E-008	VeyrierasEQTL
XKR9	rs2639939	4,45E-008	VeyrierasEQTL
XKR9	rs2639941	4,45E-008	VeyrierasEQTL
XKR9	rs2639948	4,45E-008	VeyrierasEQTL
XKR9	rs1493199	4,45E-008	VeyrierasEQTL
XKR9	rs3098865	4,45E-008	VeyrierasEQTL
XKR9	rs17760766	4,45E-008	VeyrierasEQTL
XKR9	rs3110268	4,45E-008	VeyrierasEQTL
XKR9	rs2279128	4,45E-008	VeyrierasEQTL
XKR9	rs2035925	4,45E-008	VeyrierasEQTL
XKR9	rs2035929	4,45E-008	VeyrierasEQTL
XKR9	rs3110263	4,45E-008	VeyrierasEQTL
XKR9	rs2732111	4,45E-008	VeyrierasEQTL
XKR9	rs4532626	4,45E-008	VeyrierasEQTL

XKR9	rs6981149	4,45E-008	VeyrierasEQTL
XKR9	rs2639949	4,45E-008	VeyrierasEQTL
XKR9	rs2639950	4,45E-008	VeyrierasEQTL
XKR9	rs3098859	4,45E-008	VeyrierasEQTL
XKR9	rs3098858	4,45E-008	VeyrierasEQTL
XKR9	rs6993326	4,45E-008	VeyrierasEQTL
XKR9	rs7822474	4,45E-008	VeyrierasEQTL
XKR9	rs2956611	4,45E-008	VeyrierasEQTL
XKR9	rs6986479	4,45E-008	VeyrierasEQTL
XKR9	rs950903	4,45E-008	VeyrierasEQTL
XKR9	rs7017481	4,45E-008	VeyrierasEQTL
XKR9	rs2732141	2,94E-007	VeyrierasEQTL
XKR9	rs10504492	4,06E-007	VeyrierasEQTL
XKR9	rs17699084	4,06E-007	VeyrierasEQTL
XKR9	rs11987409	4,06E-007	VeyrierasEQTL
XKR9	rs12674838	4,06E-007	VeyrierasEQTL
XKR9	rs6472564	4,06E-007	VeyrierasEQTL
XKR9	rs17699685	4,06E-007	VeyrierasEQTL
XKR9	rs17699710	4,06E-007	VeyrierasEQTL
XKR9	rs17778232	4,06E-007	VeyrierasEQTL
XKR9	rs4526394	4,06E-007	VeyrierasEQTL
XKR9	rs9650238	4,06E-007	VeyrierasEQTL
XKR9	rs7825815	4,06E-007	VeyrierasEQTL
XKR9	rs11779364	6,82E-007	VeyrierasEQTL
XKR9	rs2639947	1,16E-006	VeyrierasEQTL
XKR9	rs3098878	3,24E-006	VeyrierasEQTL
XKR9	rs2732117	3,24E-006	VeyrierasEQTL
XKR9	rs2732120	3,24E-006	VeyrierasEQTL
XKR9	rs1973364	3,24E-006	VeyrierasEQTL
XKR9	rs2008517	3,24E-006	VeyrierasEQTL
XKR9	rs2732124	3,24E-006	VeyrierasEQTL
XKR9	rs2732127	3,24E-006	VeyrierasEQTL
XKR9	rs3098872	3,24E-006	VeyrierasEQTL
XKR9	rs3098869	3,24E-006	VeyrierasEQTL
XKR9	rs7838421	1,23E-005	VeyrierasEQTL
XKR9	rs7013495	1,23E-005	VeyrierasEQTL
XKR9	rs6991708	1,23E-005	VeyrierasEQTL
XKR9	rs7012796	1,23E-005	VeyrierasEQTL
XKR9	rs13268251	4,15E-005	VeyrierasEQTL
XRR1	rs12577127	2,42E-011	Zeller10_CIS_QTL
ZFP57	rs387642	1,97E-008	Zeller10_CIS_QTL
ZFP57	rs3129063	5,54E-007	Zeller10_CIS_QTL
ZFP57	rs1610601	3,54E-006	Zeller10_CIS_QTL
ZFP57	rs1633063	3,54E-006	Zeller10_CIS_QTL
ZFP57	rs916570	3,54E-006	Zeller10_CIS_QTL

ZFP57	rs3117295	3,54E-006	Zeller10_CIS_QTL
ZFP57	rs1611701	5,18E-006	Zeller10_CIS_QTL
ZFP57	rs1610742	2,68E-005	Zeller10_CIS_QTL
ZFP57	rs3129046	2,68E-005	Zeller10_CIS_QTL
ZFP57	rs1264702	6,37E-005	Zeller10_CIS_QTL
ZFP57	rs1736915	1,75E-004	Zeller10_CIS_QTL
ZFP57	rs2517861	2,52E-004	Zeller10_CIS_QTL
ZFP57	rs5013088	2,52E-004	Zeller10_CIS_QTL
ZFP57	rs1614309	4,60E-004	Zeller10_CIS_QTL
ZFP57	rs1737031	4,60E-004	Zeller10_CIS_QTL
ZFP57	rs1633030	4,60E-004	Zeller10_CIS_QTL
ZFP57	rs1611350	6,28E-004	Zeller10_CIS_QTL
ZNF391	rs10080569	1,11E-004	PickrelEQTL
ZNF502	rs2372828	1,15E-008	Zeller10_CIS_QTL
ZNF502	rs4429669	1,36E-005	Zeller10_CIS_QTL
ZNF502	rs4234448	1,36E-005	Zeller10_CIS_QTL

Source:

Zeller10_CIS_QTL = [Zeller et al., 2010]

PickrelEQTL = [Pickrell et al., 2010]

VeyrierasEQTL = [Veyrieras et al., 2008]

StrangerEQTL = [Stranger et al., 2007]

SchadtEQTL = [Schadt et al., 2008]

MyersEQTL = [Myers et al., 2007]