

Génomique **I**ntégrative et **M**odélisation des **M**aladies **M**étaboliques
UMR 8199 (CNRS / Université de Lille 2 / Institut Pasteur de Lille)

CNRS UMR 8199 - Institut de Biologie de Lille
1 Rue du Professeur Calmette
BP 245
F-59019 LILLE CEDEX

Développement et Application de Méthodologies Statistiques pour Etudes Longitudinales d'Association Génétique

Comité de Suivi de Thèse: deuxième année

Mickaël Canouil

Doctorant en Biostatistique
(mickael.canouil@cnrs.fr)

Directeur de thèse:

Pr. Philippe Froguel

Co-directeur de thèse:

Dr. Ghislain Rocheleau (Lille 2)

Membres du CST:

Dr. Hélène Jacqmin-Gadda (INSERM U 897, Bordeaux)

Pr. Cristian Preda (Ecole Polytechnique, Lille)

26 septembre 2016



Table des matières

1	Introduction et contexte	2
2	Objectif de la thèse	3
3	Données phénotypiques et génotypiques	5
4	Méthodes	6
4.1	Modèle Joint	6
4.2	Simulation	8
4.3	Puissance statistique	8
5	Résultats d'application	9
6	Perspectives	13
7	Formations	14
7.1	Futures	14
7.2	Passées	14
	Références	17

1 Introduction et contexte

La prévalence mondiale du diabète de type 2 (DT2) est en constante augmentation et induit de ce fait la recherche de nouveaux marqueurs génétiques impliqués dans le risque de survenue de la maladie [[World Health Organization, 2015](#)]. Ces dernières années, les études d'association pangénomique (GWAS) ont identifié 65 loci associés à la susceptibilité au DT2 [[Morris et al., 2012](#)]. Les GWAS ont notamment identifié 36 variants associés à la glycémie à jeun chez des individus normoglycémiques [[Dupuis et al., 2010](#); [Scott et al., 2012](#)]. Bien que le rôle fonctionnel de la plupart de ces loci reste à déterminer, ils ont tout de même permis de mettre en lumière des liens intéressants entre la concentration de glucose sanguin et le DT2, et souligné quelques différences entre l'homéostasie du glucose dans la population générale et les niveaux physiopathologiques observés chez les diabétiques. La majorité de ces loci ont été trouvés par une méta-analyse de GWAS, malgré la très grande hétérogénéité des études et des cohortes incluses dans la méta-analyse.

La grande majorité des GWAS publiées a utilisé un design transversal plutôt qu'un design longitudinal. Cependant, une étude prospective longitudinale offre la possibilité de mesurer une même variable plusieurs fois au cours de l'étude, permettant également une description de ses fluctuations temporelles. La puissance permettant de détecter des variants génétiques associés à ces variables devrait être augmentée grâce aux multiples mesures prises à plusieurs moments, puisqu'elles donnent globalement plus d'informations qu'une seule mesure prise à un moment donné. De plus, un design longitudinal permet d'étudier les variants associés aux changements (trajectoires) de ces mesures à travers le temps, tout en augmentant la puissance permettant de détecter des différences entre les groupes (définis par leur génotype, dans un contexte génétique). Tout porte à croire qu'une bonne modélisation de ces trajectoires optimiserait les tests d'association et permettrait une meilleure exploitation des phénotypes disponibles.

2 Objectif de la thèse

Cette thèse propose de développer et implémenter des approches pangénomiques basées sur des modèles de régression, tels les modèles linéaires mixtes (LMM), les équations estimantes généralisées (GEE) [Sitlani et al., 2015], et les modèles joints (JM) liant de façon statistique la variation temporelle d'une variable (p.ex. le glucose sanguin à jeun FG) avec la survenue d'une maladie (p.ex. le DT2). Ces approches pourraient mener à l'identification de nouveaux loci. De plus, ces méthodologies permettront de confronter la vision actuelle, selon laquelle les gènes impliqués dans le contrôle de la glycémie chez les individus normoglycémiques diffèrent de ceux observés chez les individus atteints de DT2 [Yaghootkar and Frayling, 2013]. Aussi, ces approches statistiques appliquées à un contexte génétique pourraient permettre d'expliquer pourquoi certains individus en légère hyperglycémie ne développent pas de DT2. Les marqueurs ainsi identifiés renforceront du même coup l'utilité clinique de la stratification génétique du risque dans la population générale [Pal and McCarthy, 2013].

Néanmoins, les approches LMM, GEE et JM nécessitent de grandes quantités de ressources informatiques lorsqu'appliquées sur des millions de loci et des milliers d'individus. Ces approches sont le plus souvent utilisées sur un sous-ensemble de loci dans des régions d'intérêts de la pathologie étudiée [Beyene and Hamid, 2014; Wu et al., 2014]. Des approches dites "rapides" ont été développées pour pallier au problème du temps de calcul, présents notamment sur les implémentations actuelles des LMM (`n1me` et `lme4` pour R), et se basent sur des méthodes d'approximation des LMM [Sikorska et al., 2013, 2015; Sikorska, 2014]. A ce jour, deux implémentations des JM existent sur la plateforme CRAN pour le logiciel R [R Core Team, 2014]. La première est développée par Dimitris Rizopoulos (JM) [Rizopoulos, 2010] et la seconde est développée par Pete Philipson (`joiner`) [Philipson et al., 2012].

La cohorte prospective D.E.S.I.R (Données Épidémiologiques sur le Syndrome d'Insulino-Résistance) [Balkau, 1996], comprend 5 212 individus suivis pendant neuf années au cours desquelles plusieurs mesures phénotypiques ont été prises à quatre occasions uniformément réparties durant la période de suivi (0, 3, 6 et 9 ans). En exploitant l'approche par GWAS dans la cohorte D.E.S.I.R, une percée majeure a été réalisée dans la génétique des traits métaboliques menant, d'une part, à la découverte de nouveaux loci associés au DT2, et d'autre part, à l'identification de loci impliqués dans le contrôle de la glycémie d'individus normoglycémiques [Dupuis et al., 2010; Voight et al., 2010]. La base de données phénotypiques et génotypiques existante de D.E.S.I.R offre l'opportunité unique de modéliser la trajectoire temporelle de plusieurs phénotypes dont la glycémie à jeun, et peut-être même permettra de découvrir de nouveaux loci simultanément associés à l'incidence du DT2 et à un taux élevé de glucose

sanguin.

L'étude des données a été réalisée avec le logiciel R sur un serveur de calcul de 80 coeurs et 500 Go de mémoire vive, accessible dans le laboratoire de l'unité UMR 8199 dirigée par le Pr. Philippe Froguel.

3 Données phénotypiques et génotypiques

Le génotypage a été réalisé à l'aide de la puce MetaboChip (Infinium iSelect Custom Genotyping Arrays)[[Voight et al., 2012](#)], via la technologie Illumina HiScan et le logiciel GenomeStudio (Illumina, San Diego, CA, 19 USA), pour .. Les individus de cette cohorte ont été suivis pendant neuf ans, où des données phénotypiques ont été recueillies à différentes occasions : à l'inclusion, trois ans, six ans et neuf ans. Le contrôle qualité a été réalisée au moyen du logiciel PLINK (1.90 beta)[[Chang et al., 2015](#); [Purcell and Chang, 2015](#)].

Un "call rate" (qualité du génotypage) de 95% a été utilisé pour filtrer aussi bien les individus que les SNPs (Single Nucleotide Polymorphism), de même seuls les SNPs ne présentant pas de déviation à l'équilibre de Hardy-Weinberg ($p < 1 \times 10^{-3}$) et dont la fréquence de l'allèle mineur était supérieure à 1% ont été conservés, soit 124 095 SNPs et 4 884 individus.

Une analyse en composantes principales a été réalisée sur un jeu de données combinant les 4 884 individus de la cohorte D.E.S.I.R. et les 2 504 individus disponibles dans la base de données publique du projet 1 000 Génomes [[The 1000 Genomes Project Consortium, 2015](#)], et réduit aux SNPs présents dans les deux jeux de données et non palindromique (107 591 SNPs). Le premier plan factoriel était suffisant pour discriminer l'origine ethnique des individus et a permis de réduire le nombre d'individus à 4 565 individus Caucasiens. De plus, 86 individus ont été exclus, en raison de phénotypes manquants (sexe, Indice de masse Corporelle, glycémie à jeun), ne permettant pas d'établir le statut DT2 de ces individus. Les cas prévalents de DT2 (glycémie supérieure à 7mmol/L) ont également été retirés de l'analyse (115 individus).

Le jeu de données final utilisé pour les analyses incluait 4 364 individus (dont 179 DT2 incidents) et 124 095 SNPs.

4 Méthodes

4.1 Modèle Joint

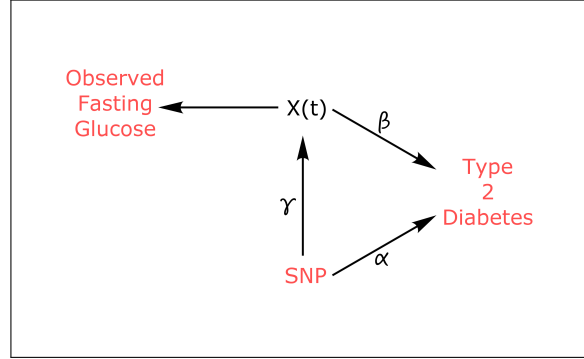


Figure 1: Diagramme général d'un modèle joint pour le T2D (adapté de Ibrahim et al. [2010])

$X(t)$: trajectoire de la glycémie à jeun inférée des données longitudinales observées ; α : effet du SNP sur le diabète ; γ : effet du SNP sur la trajectoire de la glycémie à jeun ; β : effet de la trajectoire de la glycémie à jeun sur le diabète.

En utilisant, l'approche de modèle joint implémentée dans l'extension JM [Rizopoulos, 2010] du logiciel R (version 3.2.3)[R Core Team, 2015], 124 095 SNPs de la MetaboChip ont été testés simultanément pour leur association avec la glycémie à jeun et le risque de DT2.

La formulation standard du modèle joint implique deux composantes, d'une part, une composante longitudinale pour modéliser la trajectoire de la variable étudiée, et d'autre part, une composante de survie pour modéliser la survenue de l'événement étudié. La composante longitudinale consiste typiquement à l'application d'un modèle linéaire mixte :

$$Y_{ij} = X_{ij} + \epsilon_{ij}, \quad (1)$$

où Y_{ij} est la valeur observée et X_{ij} la vraie (non-observée) valeur de la variable longitudinale. Le terme ϵ_{ij} est le terme d'erreur aléatoire supposé être distribué selon la Loi Normale :

$$\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2) \quad (2)$$

La quantité X_{ij} (ou $X(t)$) est la fonction de la trajectoire et est définie usuellement comme une fonction linéaire (ou quadratique) du temps t .

Des covariables peuvent être incluses dans la fonction de la trajectoire, comme l'âge, le sexe ou l'IMC. Dans notre étude, Y_{ij} représente les valeurs mesurées la glycémie à jeun au temps t_{ij} , Z_i désigne le génotype du SNP analysé pour l'individu i et W_i désigne les covariables selon le modèle suivant :

$$Y_{ij} = X_{ij} + \epsilon_{ij} = \theta_{0i} + \theta_{1i} \times t_{ij} + \gamma \times Z_i + \delta \times W_i + \epsilon_{ij} \quad (3)$$

Pour simplifier, le terme $\delta \times W_i$ sera omis dans la suite.

Les paramètres θ_{0i} et θ_{1i} sont supposés être distribués selon une distribution normale multivariée :

$$\theta \sim \mathcal{N}_2(\mu, \Sigma) \quad (4)$$

Le paramètre γ évalue l'effet additif du SNP (Z_i) sur la fonction de la trajectoire. Pour tenir compte éventuellement de pentes variant entre les génotypes, un terme d'interaction entre le SNP et le temps peut être inclus dans la fonction de la trajectoire. Le terme d'interaction n'a pas été inclus dans notre étude.

La composante de survie (survenue du DT2) se compose généralement d'un modèle paramétrique (p.ex. exponentielle ou Weibull) ou semi-paramétrique (p.ex. risques proportionnels de Cox) avec :

$$h_i(t) = h_0(t) \exp(\beta X_i(t) + \alpha Z_i), \quad (5)$$

où $h_i(t)$ est la fonction de risque au temps t pour l'individu i et $h_0(t)$ est la fonction de risque de base non spécifiée, supposée être une constante par morceaux avec deux noeuds placés à des temps intermédiaires (c'est-à-dire, à trois et six ans sur les neuf ans du suivi). Le coefficient α mesure l'effet du SNP sur le délai d'apparition du DT2, alors que le coefficient β mesure l'association entre la trajectoire du niveau de la glycémie à jeun et le temps d'apparition du DT2.

4.2 Simulation

Des études de simulation ont été menées pour examiner la puissance statistique et l'erreur de type 1 des SNPs trouvés comme nominalement associé (à 5%) en utilisant le modèle joint, comme implémenté par Rizopoulos [2010]. Notre objectif principal était de déterminer le gain ou la perte de puissance du modèle joint par rapport aux approches classiques transversales (p.ex. régression logistique ou linéaire, modèle de Cox) pour détecter l'effet d'un SNP, sur la glycémie à jeun et le statut DT2 dans notre étude. Les jeux de données de simulation, qui ont été réalisés avec R, suivent les Equations 1 à 5, avec la fonction de risque de base fixée ($h_0(t) = \lambda$) de façon à obtenir une incidence équivalente à celle de la cohorte D.E.S.I.R (environ 5%), durant la période de suivi de neuf ans. Les temps d'événements ont été générés selon une distribution exponentielle dans le cadre du modèle de Cox à risque proportionnel [Austin, 2012].

$$H(T) = \int_0^T \lambda \exp(\beta \times X(t) + \alpha \times Z) dt \quad (6)$$

$$T = \frac{1}{\beta\theta_1} \log \left(-\frac{\beta\theta_1 \times \log(1-u)}{\lambda \exp(\beta\theta_0 + (\beta\gamma + \alpha)Z)} + 1 \right) \quad (7)$$

4.3 Puissance statistique

La puissance statistique a été calculée, 1) au niveau du modèle joint, à l'aide de la formule de Chen et al. [2011] :

$$z_{\tilde{\beta}} = \pm \sqrt{Df(1-f)(\beta\gamma + \alpha)^2} + z_{1-\tilde{\alpha}/2}, \quad (8)$$

avec D , le nombre de DT2 incidents et f , la fréquence de l'allèle à risque, 2) au niveau des paramètres γ et α respectivement pour l'effet du SNP sur la trajectoire de la glycémie à jeun et sur le risque de DT2.

5 Résultats d'application

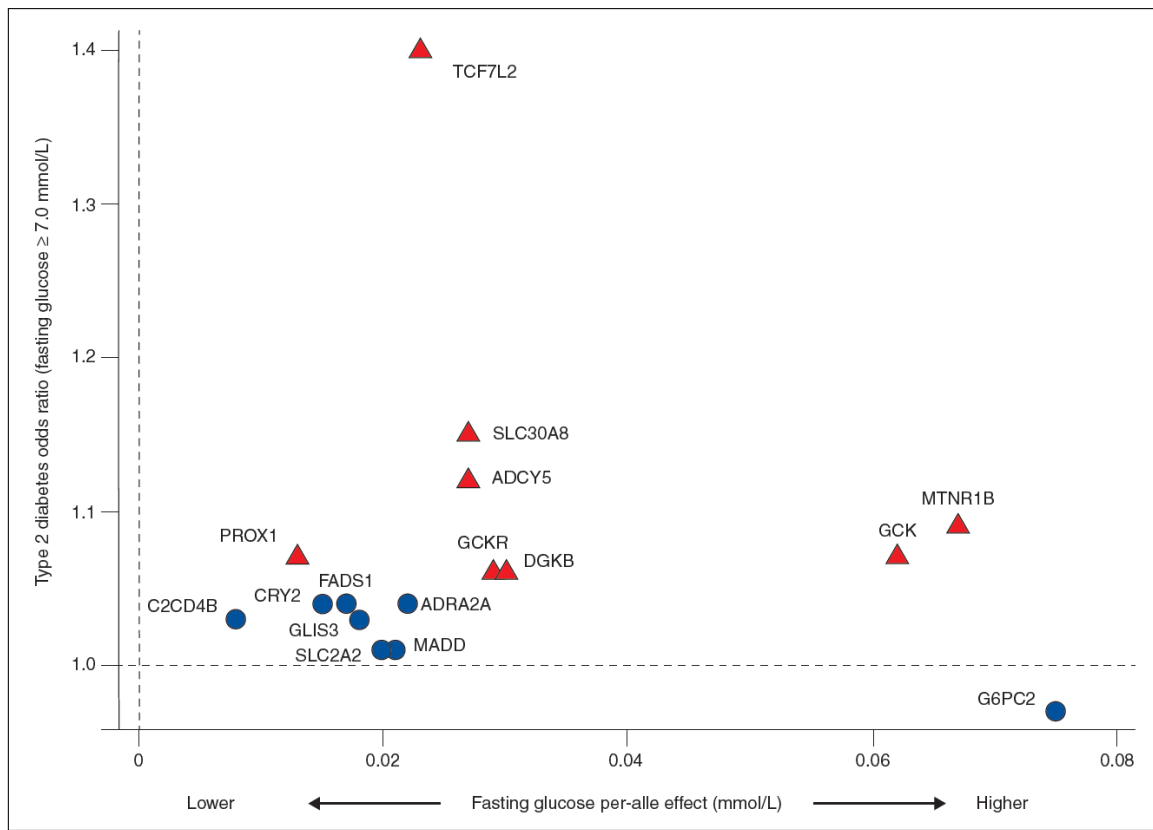


Figure 2: Résultat d'association des allèles à risque pour le T2D et/ou la glycémie à jeun par GWAS [Yaghootkar and Frayling, 2013]. Cercle bleu : l'association à la glycémie à jeun pour $p < 5 \times 10^{-8}$; Triangle rouge : l'association au T2D et à la glycémie à jeun $p < 5 \times 10^{-8}$.

Les approches classiques de GWAS ont pu mettre en évidence entre 80 et 100 loci associés au DT2 et au FG, notamment via des méta-analyses [Yaghootkar and Frayling, 2013; Dupuis et al., 2010; Vaxillaire et al., 2014], dont certains de ces loci (gènes) sont montrés sur la Figure 2.

Dans le cadre des données génotypiques et phénotypiques disponibles dans la cohorte D.E.S.I.R. (Section 3), nous avons analysé l'ensemble des 124 095 SNPs disponibles à l'aide d'un modèle joint (Section 4), nous permettant ainsi d'identifier 145 loci associés à la fois à la trajectoire de la glycémie à jeun et également à la survenue du DT2 (au seuil statistique de 5%). Ces loci sont présentés dans la Figure 3, avec la puissance calculée selon les formules développées par Chen et al. [2011].

Parmi les 145 loci obtenus, 13 ont déjà été identifiés dans des GWAS [Welter et al., 2014] pour être associés à la glycémie à jeun et/ou au DT2. Les 132 loci restants ne pourraient être

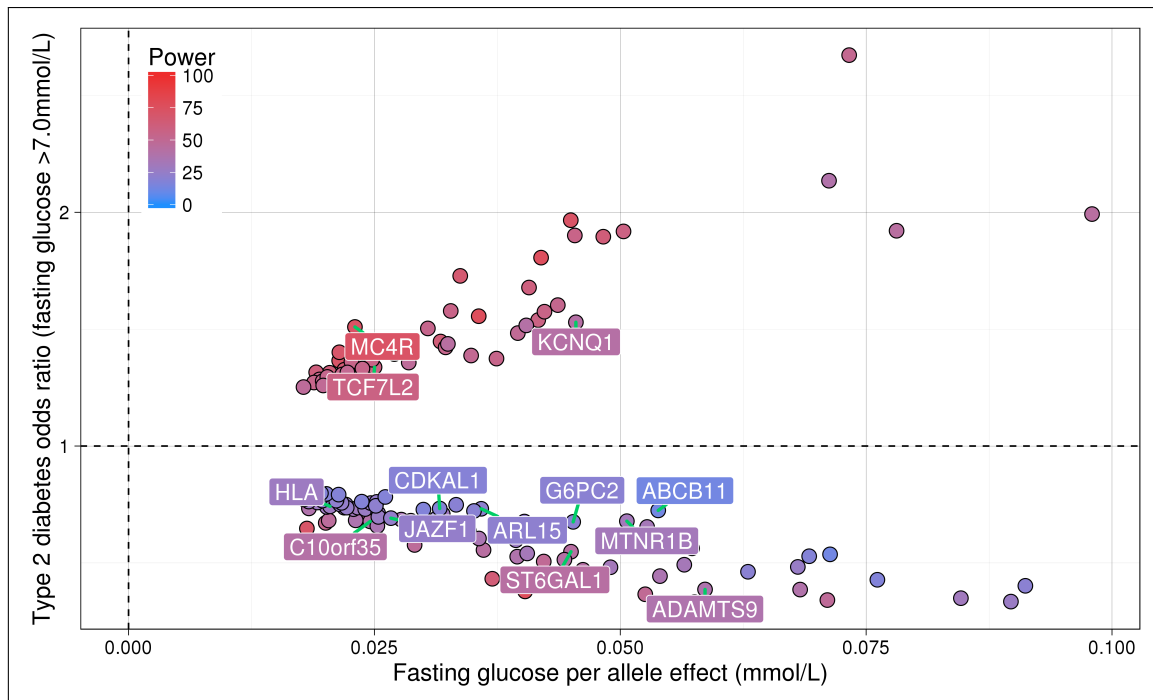


Figure 3: Résultat d'association des allèles à risque pour le T2D et/ou la glycémie à jeun à partir du modèle joint.

que des faux positifs en raison du seuil peu stringent utilisé ici (5%) et du nombre de tests effectués (124 095). Ainsi, une validation de ces résultats devra être effectués à l'aide d'une cohorte de réplification par exemple.

Cependant, au regard des résultats obtenus par l'approche de modélisation jointe de la trajectoire de la glycémie à jeun et la survenue du DT2, notamment au niveau des 13 loci répertoriés dans le GWAS catalog [Welter et al., 2014] (Table 1 et Figure 3), nous pouvons noter des effets cohérents par rapport à ceux reportés dans la littérature. À cela, le gène MTNR1B fait exception, comme le montre la localisation de ce gène sur la Figure 2 et la Figure 3, suggérant un effet opposé sur le risque de survenue du DT2 au cours du suivi. En effet, MTNR1B a été reporté, notamment par Bouatia-Naji et al. [2009] et Lyssenko et al. [2009] comme associé à une élévation de la glycémie et une augmentation du risque de DT2. Ce phénomène pourrait être expliqué par une certaine spécificité de la cohorte, puisque le gène MTNR1B a été identifié via méta-analyse. D'autre part, comme le souligne Dr. Dimitris Rizopoulos [Rizopoulos, 2012], les cas incidents de DT2 pourraient ne pas être homogènes de par leurs caractéristiques cliniques et plus généralement phénotypiques.

“ In some occasions the typical assumption that the current value of the time-dependent covariate affects the current risk for an event may lead to medically illogical conclusions. For example, this has been observed by Cavender et al. (1992) who in a study on patients with coronary artery disease noted that the current smoking status decreased the risk for death (although not statistically significantly). The reason behind this surprising result was that most of those who died were smokers, but many had stopped smoking at the last follow-up before their death. Thus, many of the patients who died had just quit smoking, whereas some of the patients who were still alive were still smoking, leading to the surprising result. ”

— Dimitris Rizopoulos

Une approche dites à classes latentes pourrait être utilisée [Proust-Lima et al., 2009] pour prendre en compte une possible hétérogénéité dans la population et ainsi pourrait permettre d'expliquer le résultat obtenu pour le gène MTNR1B.

SNP	α	γ	β	Power
rs1942873_G (MC4R)	0.412	0.023	3.15	72.70
rs55899248_C (TCF7L2)	0.291	0.025	3.49	57.60
rs17301514_T (ST6GAL1)	-0.602	0.045	3.21	43.10
rs61871514_T (KCNQ1)	0.425	0.0455	3.17	41.50
rs7072870_T (C10orf35)	-0.389	0.0246	3.2	41.00
rs7432858_T (ADAMTS9)	-0.951	0.0586	3.22	37.70
rs10830962_G (MTNR1B)	-0.388	0.0507	3.24	31.50
rs114508985_C (HLA)	-0.299	0.0205	3.56	28.50
rs17155987_C (JAZF1)	-0.369	0.0267	3.57	24.20
rs4311394_T (ARL15)	-0.313	0.0359	3.19	21.80
rs12475693_C (G6PC2)	-0.392	0.0452	3.17	20.30
rs4712580_C (CDKAL1)	-0.311	0.0316	3.2	18.80
rs17540154_C (ABCB11)	-0.322	0.0539	3.19	13.00

Table 1: Résultat d'analyse pour une sélection de SNPs connus pour être associés au T2D et/ou à la glycémie à jeun. En **bleu** : $p < 0.05$, en **rouge** : $p < 5 \times 10^{-8}$. Puissance selon la formule de Chen et al. [2011].

L'un des objectifs de l'utilisation d'un modèle joint, en dehors de la modélisation jointe du processus de survie et du processus longitudinal, était de mettre à profit les mesures répétées pouvant être disponibles dans l'objectif d'accroître la puissance statistique, pour détecter un effet d'un SNP sur le ou les traits étudiés, en comparaison des approches classiques et transversales des GWAS (Table 2).

SNP	α	γ
rs1942873_G (MC4R)	47.7 / 71.6	64.9 / 52.8
rs55899248_C (TCF7L2)	66.6 / 35.3	91.6 / 81.0
rs17301514_T (ST6GAL1)	83.6 / 23.8	99.9 / 99.9
rs61871514_T (KCNQ1)	44.7 / 73.6	83.3 / 70.1
rs7072870_T (C10orf35)	65.7 / 36.5	73.3 / 60.9
rs7432858_T (ADAMTS9)	94.2 / 17.6	99.9 / 99.9
rs10830962_G (MTNR1B)	55.1 / 32.1	60.8 / 47.1
rs114508985_C (HLA)	92.6 / 75.7	63.8 / 48.2
rs17155987_C (JAZF1)	35.6 / 58.5	66.0 / 52.1
rs4311394_T (ARL15)	52.4 / 23.6	79.0 / 65.7
rs12475693_C (G6PC2)	76.2 / 56.5	56.7 / 44.5
rs4712580_C (CDKAL1)	37.9 / 59.8	62.4 / 49.6
rs17540154_C (ABCB11)	98.4 / 89.9	53.2 / 40.6

Table 2: Puissance statistique obtenu par simulation pour les paramètres α (effet du SNP sur le statut) et γ (effet du SNP sur la trajectoire de la glycémie à jeun). En bleu, lorsque le modèle joint (JM) est plus puissant que le modèle transversal à un temps donné (CM), présenté sous la forme JM / CM. (Erreur de type I : $5.81\% \pm 0.83$ / $5.47\% \pm 0.64$ pour JM et CM.)

6 Perspectives

1. Ecriture d'un rapport scientifique (Bourse SFD-Lilly : 21 Mai 2017) et d'un article sur l'application du modèle joint à la cohorte DESIR
2. Validation des SNPs/gènes mis en évidence par le modèle joint (p.ex. cohorte de réplication)
3. Etude d'autre trait tel que l'HbA1c (hémoglobine glyquée)
4. Inclusion des individus incident pour l'IFG (Impaired Fasting Glucose ; $FG > 6.1mM/L$) en plus des individus DT2 ($FG > 7mM/L$)

7 Formations

7.1 Futures

- *Congrès IGES 2016* (International Genetic Epidemiology Society) : Présentation poster "Single Nucleotide Polymorphisms Associated With Fasting Blood Glucose Trajectory And Type 2 Diabetes Incidence : A Joint Modelling Approach"
- *Congrès de la SFD 2017* (Société Francophone du Diabète)
- *Congrès de la SFdS 2017* (Société Française de Statistique)
- *Rencontres R 2017*
- *useR 2017*
- ...

7.2 Passées

Nombre de crédits obtenus à ce jour : 47/60 crédits.



Nom / Prénom: CANOUIL / Mickael
ED : ED BSL
Université : LILLE 2
Doctorant en 2 année

Inscriptions déjà prises :

Recherche documentaire et publication	statut de l'inscription	Crédits
Veille et stratégie de recherche documentaire <u>session</u> : Session ED BSL (Lille - Campus CHR)) : vendredi 20 février 2015 matin	Définitivement Validée	1/1
Maîtriser l'information scientifique : de la recherche à la publication	statut de l'inscription	Crédits
Gérer efficacement sa documentation avec ZOTERO <u>session</u> : Session Sciences (SCD Lille1) : mercredi 08 avril 2015 matin	Définitivement Validée	1/1
Formations spécifiques à l'ED BSL	statut de l'inscription	Crédits
Implication dans les activités de l'Ecole Doctorale (2014-2015) <u>session</u> : Comité d'organisation JAV	Définitivement Validée	0/0
Journées nationales du DEveloppement logiciel de l'Enseignement Supérieur et Recherche (JDEV2015)	statut de l'inscription	Crédits
<u>session</u> : 30/06/2015 - 03/07/2015	Définitivement Validée	7/7
Développement et valorisation des compétences	statut de l'inscription	Crédits
Mener sa thèse en mode projet <u>session</u> : 3ème session : 5/01/2016 ; retour d'exp 20/01/2016 matin	Définitivement Validée	10/10
Développement et valorisation des compétences	statut de l'inscription	Crédits
Les fondamentaux du management d'équipe <u>session</u> : 1ère session : 14 et 15/01/2016	Définitivement Validée	7/7

	statut de l'inscription	Crédits
Lille Genomic Institute - First Symposium, Breakthroughs in Personalized Medicine <u>session</u> : mercredi 13 janvier 2016	Définitivement Validée	2/2
<hr/>		
Formations spécifiques à l'ED BSL	statut de l'inscription	Crédits
Implication dans les activités de l'Ecole Doctorale (2014-2015) <u>session</u> : Autres	Définitivement Validée	6/6
<hr/>		
	statut de l'inscription	Crédits
SMPGD 2016: Statistical Methods for Post Genomic Data <u>session</u> : jeudi 11 février et vendredi 12 février	Définitivement Validée	6/6
<hr/>		
	statut de l'inscription	Crédits
Big Data: Modeling, Estimation and Selection <u>session</u> : jeudi 9 juin et vendredi 10 juin.	Définitivement Validée	3/3
<hr/>		
	statut de l'inscription	Crédits
Introduction to the statistical analysis of genome-wide association studies <u>session</u> : du 07/07/2016 au 08/07/2016	Définitivement Validée	4/4
<hr/>		
TOTAL : 47 / 60		

Références

- Austin, P. C. (2012). Generating survival times to simulate Cox proportional hazards models with time-varying covariates. *Statistics in Medicine*, 31(29) :3946–3958.
- Balkau, B. (1996). An epidemiologic survey from a network of French Health Examination Centres, (D.E.S.I.R.) : epidemiologic data on the insulin resistance syndrome. *Revue D'épidémiologie Et De Santé Publique*, 44(4) :373–375.
- Beyene, J. and Hamid, J. S. (2014). Longitudinal Data Analysis in Genome-Wide Association Studies. *Genetic Epidemiology*, 38(S1) :S68–S73.
- Bouatia-Naji, N., Bonnefond, A., Cavalcanti-Proença, C., Sparsø, T., Holmkvist, J., Marchand, M., Delplanque, J., Lobbens, S., Rocheleau, G., Durand, E., De Graeve, F., Chèvre, J.-C., Borch-Johnsen, K., Hartikainen, A.-L., Ruokonen, A., Tichet, J., Marre, M., Weill, J., Heude, B., Tauber, M., Lemaire, K., Schuit, F., Elliott, P., Jørgensen, T., Charpentier, G., Hadjadj, S., Cauchi, S., Vaxillaire, M., Sladek, R., Visvikis-Siest, S., Balkau, B., Lévy-Marchal, C., Pattou, F., Meyre, D., Blakemore, A. I. F., Jarvelin, M.-R., Walley, A. J., Hansen, T., Dina, C., Pedersen, O., and Froguel, P. (2009). A variant near MTNR1b is associated with increased fasting plasma glucose levels and type 2 diabetes risk. *Nature Genetics*, 41(1) :89–94.
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK : rising to the challenge of larger and richer datasets. *Giga-Science*, 4 :7.
- Chen, L. M., Ibrahim, J. G., and Chu, H. (2011). Sample size and power determination in joint modeling of longitudinal and survival data. *Statistics in Medicine*, 30(18) :2295–2309.
- Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A. U., Wheeler, E., Glazer, N. L., Bouatia-Naji, N., Gloyn, A. L., Lindgren, C. M., Mägi, R., Morris, A. P., Randall, J., Johnson, T., Elliott, P., Rybin, D., Thorleifsson, G., Steinthorsdottir, V., Henneman, P., Grallert, H., Dehghan, A., Hottenga, J. J., Franklin, C. S., Navarro, P., Song, K., Goel, A., Perry, J. R. B., Egan, J. M., Lajunen, T., Grarup, N., Sparsø, T., Doney, A., Voight, B. F., Stringham, H. M., Li, M., Kanoni, S., Shrader, P., Cavalcanti-Proença, C., Kumari, M., Qi, L., Timpson, N. J., Gieger, C., Zabena, C., Rocheleau, G., Ingelsson, E., An, P., O'Connell, J., Luan, J., Elliott, A., McCarroll, S. A., Payne, F., Roccascaccia, R. M., Pattou, F., Sethupathy, P., Ardlie, K., Ariyurek, Y., Balkau, B., Barter, P., Beilby, J. P., Ben-Shlomo, Y., Benediktsson, R., Bennett, A. J., Bergmann, S., Bochud, M., Boerwinkle, E., Bonnefond, A., Bonnycastle, L. L., Borch-Johnsen, K., Böttcher, Y., Brunner, E., Bumpstead, S. J., Charpentier, G.,

Chen, Y.-D. I., Chines, P., Clarke, R., Coin, L. J. M., Cooper, M. N., Cornelis, M., Crawford, G., Crisponi, L., Day, I. N. M., de Geus, E., Delplanque, J., Dina, C., Erdos, M. R., Fedson, A. C., Fischer-Rosinsky, A., Forouhi, N. G., Fox, C. S., Frants, R., Franzosi, M. G., Galan, P., Goodarzi, M. O., Graessler, J., Groves, C. J., Grundy, S., Gwilliam, R., Gyllensten, U., Hadjadj, S., Hallmans, G., Hammond, N., Han, X., Hartikainen, A.-L., Hassanali, N., Hayward, C., Heath, S. C., Hercberg, S., Herder, C., Hicks, A. A., Hillman, D. R., Hingorani, A. D., Hoffman, A., Hui, J., Hung, J., Isomaa, B., Johnson, P. R. V., Jørgensen, T., Jula, A., Kaakinen, M., Kaprio, J., Kesaniemi, Y. A., Kivimaki, M., Knight, B., Koskinen, S., Kovacs, P., Kyvik, K. O., Lathrop, G. M., Lawlor, D. A., Le Bacquer, O., Lecoœur, C., Li, Y., Lyssenko, V., Mahley, R., Mangino, M., Manning, A. K., Martínez-Larrad, M. T., McAteer, J. B., McCulloch, L. J., McPherson, R., Meisinger, C., Melzer, D., Meyre, D., Mitchell, B. D., Morken, M. A., Mukherjee, S., Naitza, S., Narisu, N., Neville, M. J., Oostra, B. A., Orrù, M., Pakyz, R., Palmer, C. N. A., Paolisso, G., Pattaro, C., Pearson, D., Peden, J. F., Pedersen, N. L., Perola, M., Pfeiffer, A. F. H., Pichler, I., Polasek, O., Posthuma, D., Potter, S. C., Pouta, A., Province, M. A., Psaty, B. M., Rathmann, W., Rayner, N. W., Rice, K., Ripatti, S., Rivadeneira, F., Roden, M., Rolandsson, O., Sandbaek, A., Sandhu, M., Sanna, S., Sayer, A. A., Scheet, P., Scott, L. J., Seedorf, U., Sharp, S. J., Shields, B., Sigurðsson, G., Sijbrands, E. J. G., Silveira, A., Simpson, L., Singleton, A., Smith, N. L., Sovio, U., Swift, A., Syddall, H., Syvänen, A.-C., Tanaka, T., Thorand, B., Tichet, J., Tönjes, A., Tuomi, T., Uitterlinden, A. G., van Dijk, K. W., van Hoek, M., Varma, D., Visvikis-Siest, S., Vitart, V., Vogelzangs, N., Waeber, G., Wagner, P. J., Walley, A., Walters, G. B., Ward, K. L., Watkins, H., Weedon, M. N., Wild, S. H., Willemsen, G., Witteman, J. C. M., Yarnell, J. W. G., Zeggini, E., Zelenika, D., Zethelius, B., Zhai, G., Zhao, J. H., Zillikens, M. C., Borecki, I. B., Loos, R. J. F., Meneton, P., Magnusson, P. K. E., Nathan, D. M., Williams, G. H., Hattersley, A. T., Silander, K., Salomaa, V., Smith, G. D., Bornstein, S. R., Schwarz, P., Spranger, J., Karpe, F., Shuldiner, A. R., Cooper, C., Dedoussis, G. V., Serrano-Ríos, M., Morris, A. D., Lind, L., Palmer, L. J., Hu, F. B., Franks, P. W., Ebrahim, S., Marmot, M., Kao, W. H. L., Pankow, J. S., Sampson, M. J., Kuusisto, J., Laakso, M., Hansen, T., Pedersen, O., Pramstaller, P. P., Wichmann, H. E., Illig, T., Rudan, I., Wright, A. F., Stumvoll, M., Campbell, H., Wilson, J. F., Hamsten, A., Bergman, R. N., Buchanan, T. A., Collins, F. S., Mohlke, K. L., Tuomilehto, J., Valle, T. T., Altshuler, D., Rotter, J. I., Siscovick, D. S., Penninx, B. W. J. H., Boomsma, D., Deloukas, P., Spector, T. D., Frayling, T. M., Ferrucci, L., Kong, A., Thorsteinsdottir, U., Stefansson, K., van Duijn, C. M., Aulchenko, Y. S., Cao, A., Scuteri, A., Schlessinger, D., Uda, M., Ruukonen, A., Jarvelin, M.-R., Waterworth, D. M., Vollenweider, P., Peltonen, L., Mooser, V., Abecasis, G. R., Wareham, N. J., Sladek, R., Froguel, P., Watan-

- abe, R. M., Meigs, J. B., Groop, L., Boehnke, M., McCarthy, M. I., Florez, J. C., and Barroso, I. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature genetics*, 42(2) :105–116.
- Ibrahim, J. G., Chu, H., and Chen, L. M. (2010). Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data. *Journal of Clinical Oncology*, 28(16) :2796–2801.
- Lyssenko, V., Nagorny, C. L. F., Erdos, M. R., Wierup, N., Jonsson, A., Spégel, P., Bugliani, M., Saxena, R., Fex, M., Pulizzi, N., Isomaa, B., Tuomi, T., Nilsson, P., Kuusisto, J., Tuomilehto, J., Boehnke, M., Altshuler, D., Sundler, F., Eriksson, J. G., Jackson, A. U., Laakso, M., Marchetti, P., Watanabe, R. M., Mulder, H., and Groop, L. (2009). Common variant in MTNR1b associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nature Genetics*, 41(1) :82–88.
- Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segrè, A. V., Steinthorsdottir, V., Strawbridge, R. J., Khan, H., Grallert, H., Mahajan, A., Prokopenko, I., Kang, H. M., Dina, C., Esko, T., Fraser, R. M., Kanoni, S., Kumar, A., Lagou, V., Langenberg, C., Luan, J., Lindgren, C. M., Müller-Nurasyid, M., Pechlivanis, S., Rayner, N. W., Scott, L. J., Wiltshire, S., Yengo, L., Kinnunen, L., Rossin, E. J., Raychaudhuri, S., Johnson, A. D., Dimas, A. S., Loos, R. J. F., Vedantam, S., Chen, H., Florez, J. C., Fox, C., Liu, C.-T., Rybin, D., Couper, D. J., Kao, W. H. L., Li, M., Cornelis, M. C., Kraft, P., Sun, Q., van Dam, R. M., Stringham, H. M., Chines, P. S., Fischer, K., Fontanillas, P., Holmen, O. L., Hunt, S. E., Jackson, A. U., Kong, A., Lawrence, R., Meyer, J., Perry, J. R., Platou, C. G., Potter, S., Rehnberg, E., Robertson, N., Sivapalaratnam, S., Stančáková, A., Stirrups, K., Thorleifsson, G., Tikkanen, E., Wood, A. R., Almgren, P., Atalay, M., Benediktsson, R., Bonnycastle, L. L., Burt, N., Carey, J., Charpentier, G., Crenshaw, A. T., Doney, A. S. F., Dorkhan, M., Edkins, S., Emilsson, V., Eury, E., Forsen, T., Gertow, K., Gigante, B., Grant, G. B., Groves, C. J., Guiducci, C., Herder, C., Hreidarsson, A. B., Hui, J., James, A., Jonsson, A., Rathmann, W., Klopp, N., Kravic, J., Krjutškov, K., Langford, C., Leander, K., Lindholm, E., Lobbens, S., Männistö, S., Mirza, G., Mühleisen, T. W., Musk, B., Parkin, M., Rallidis, L., Saramies, J., Sennblad, B., Shah, S., Sigurðsson, G., Silveira, A., Steinbach, G., Thorand, B., Trakalo, J., Veglia, F., Wennauer, R., Winckler, W., Zabaneh, D., Campbell, H., van Duijn, C., Uitterlinden, A. G., Hofman, A., Sijbrands, E., Abecasis, G. R., Owen, K. R., Zeggini, E., Trip, M. D., Forouhi, N. G., Syvänen, A.-C., Eriksson, J. G., Peltonen, L., Nöthen, M. M., Balkau, B., Palmer, C. N. A., Lyssenko, V., Tuomi, T., Isomaa, B., Hunter, D. J., Qi, L., Shuldiner, A. R., Roden, M., Barroso, I., Wilsgaard, T., Beilby, J., Hovingh, K., Price, J. F., Wilson, J. F., Rauramaa, R., Lakka, T. A., Lind, L., Dedoussis,

- G., Njølstad, I., Pedersen, N. L., Khaw, K.-T., Wareham, N. J., Keinanen-Kiukkaanniemi, S. M., Saaristo, T. E., Korpi-Hyövähti, E., Saltevo, J., Laakso, M., Kuusisto, J., Metspalu, A., Collins, F. S., Mohlke, K. L., Bergman, R. N., Tuomilehto, J., Boehm, B. O., Gieger, C., Hveem, K., Cauchi, S., Froguel, P., Baldassarre, D., Tremoli, E., Humphries, S. E., Saleheen, D., Danesh, J., Ingelsson, E., Ripatti, S., Salomaa, V., Erbel, R., Jöckel, K.-H., Moebus, S., Peters, A., Illig, T., de Faire, U., Hamsten, A., Morris, A. D., Donnelly, P. J., Frayling, T. M., Hattersley, A. T., Boerwinkle, E., Melander, O., Kathiresan, S., Nilsson, P. M., Deloukas, P., Thorsteinsdottir, U., Groop, L. C., Stefansson, K., Hu, F., Pankow, J. S., Dupuis, J., Meigs, J. B., Altshuler, D., Boehnke, M., and McCarthy, M. I. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics*, 44(9) :981–990.
- Pal, A. and McCarthy, M. I. (2013). The genetics of type 2 diabetes and its clinical relevance. *Clinical Genetics*, 83(4) :297–306.
- Philipson, P., Diggle, P., Sousa, I., Kolamunnage-Dona, R., Williamson, P., and Henderson, R. (2012). joiner : Joint modelling of repeated measurements and time-to-event data.
- Proust-Lima, C., Joly, P., Dartigues, J.-F., and Jacqmin-Gadda, H. (2009). Joint modelling of multivariate longitudinal outcomes and a time-to-event : A nonlinear latent class approach. *Computational Statistics & Data Analysis*, 53(4) :1142–1154.
- Purcell, S. and Chang, C. (2015). PLINK v1.90b3.36.
- R Core Team (2014). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- R Core Team (2015). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rizopoulos, D. (2010). JM : An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, 35(9) :1–33.
- Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data : With applications in R*. CRC Press.
- Scott, R. A., Lagou, V., Welch, R. P., Wheeler, E., Montasser, M. E., Luan, J., Mägi, R., Strawbridge, R. J., Rehnberg, E., Gustafsson, S., Kanoni, S., Rasmussen-Torvik, L. J., Yengo, L., Lecoeur, C., Shungin, D., Sanna, S., Sidore, C., Johnson, P. C. D., Jukema, J. W., Johnson, T., Mahajan, A., Verweij, N., Thorleifsson, G., Hottenga, J.-J., Shah, S., Smith, A. V., Sennblad,

B., Gieger, C., Salo, P., Perola, M., Timpson, N. J., Evans, D. M., Pourcain, B. S., Wu, Y., Andrews, J. S., Hui, J., Bielak, L. F., Zhao, W., Horikoshi, M., Navarro, P., Isaacs, A., O'Connell, J. R., Stirrups, K., Vitart, V., Hayward, C., Esko, T., Mihailov, E., Fraser, R. M., Fall, T., Voight, B. F., Raychaudhuri, S., Chen, H., Lindgren, C. M., Morris, A. P., Rayner, N. W., Robertson, N., Rybin, D., Liu, C.-T., Beckmann, J. S., Willems, S. M., Chines, P. S., Jackson, A. U., Kang, H. M., Stringham, H. M., Song, K., Tanaka, T., Peden, J. F., Goel, A., Hicks, A. A., An, P., Müller-Nurasyid, M., Franco-Cereceda, A., Folkersen, L., Marullo, L., Jansen, H., Oldehinkel, A. J., Bruinenberg, M., Pankow, J. S., North, K. E., Forouhi, N. G., Loos, R. J. F., Edkins, S., Varga, T. V., Hallmans, G., Oksa, H., Antonella, M., Nagaraja, R., Trompet, S., Ford, I., Bakker, S. J. L., Kong, A., Kumari, M., Gigante, B., Herder, C., Munroe, P. B., Caulfield, M., Antti, J., Mangino, M., Small, K., Miljkovic, I., Liu, Y., Atalay, M., Kiess, W., James, A. L., Rivadeneira, F., Uitterlinden, A. G., Palmer, C. N. A., Doney, A. S. F., Willemsen, G., Smit, J. H., Campbell, S., Polasek, O., Bonnycastle, L. L., Hercberg, S., Dimitriou, M., Bolton, J. L., Fowkes, G. R., Kovacs, P., Lindström, J., Zemunik, T., Bandinelli, S., Wild, S. H., Basart, H. V., Rathmann, W., Grallert, H., Maerz, W., Kleber, M. E., Boehm, B. O., Peters, A., Pramstaller, P. P., Province, M. A., Borecki, I. B., Hastie, N. D., Rudan, I., Campbell, H., Watkins, H., Farrall, M., Stumvoll, M., Ferrucci, L., Waterworth, D. M., Bergman, R. N., Collins, F. S., Tuomilehto, J., Watanabe, R. M., de Geus, E. J. C., Penninx, B. W., Hofman, A., Oostra, B. A., Psaty, B. M., Vollenweider, P., Wilson, J. F., Wright, A. F., Hovingh, G. K., Metspalu, A., Uusitupa, M., Magnusson, P. K. E., Kyvik, K. O., Kaprio, J., Price, J. F., Dedoussis, G. V., Deloukas, P., Meneton, P., Lind, L., Boehnke, M., Shuldiner, A. R., van Duijn, C. M., Morris, A. D., Toenjes, A., Peyser, P. A., Beilby, J. P., Körner, A., Kuusisto, J., Laakso, M., Bornstein, S. R., Schwarz, P. E. H., Lakka, T. A., Rauramaa, R., Adair, L. S., Smith, G. D., Spector, T. D., Illig, T., de Faire, U., Hamsten, A., Gudnason, V., Kivimäki, M., Hingorani, A., Keinanen-Kiukkaanniemi, S. M., Saaristo, T. E., Boomsma, D. I., Stefansson, K., van der Harst, P., Dupuis, J., Pedersen, N. L., Sattar, N., Harris, T. B., Cucca, F., Ripatti, S., Salomaa, V., Mohlke, K. L., Balkau, B., Froguel, P., Pouta, A., Jarvelin, M.-R., Wareham, N. J., Bouatia-Naji, N., McCarthy, M. I., Franks, P. W., Meigs, J. B., Teslovich, T. M., Florez, J. C., Langenberg, C., Ingelsson, E., Prokopenko, I., and Barroso, I. (2012). Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature genetics*, 44(9) :991–1005.

Sikorska, K. (2014). *Computationally Fast Approaches to Genome-Wide Association Studies*. PhD thesis.

Sikorska, K., Montazeri, N. M., Uitterlinden, A., Rivadeneira, F., Eilers, P. H., and Lesaffre,

- E. (2015). GWAS with longitudinal phenotypes : performance of approximate procedures. *European Journal of Human Genetics*.
- Sikorska, K., Rivadeneira, F., Groenen, P. J., Hofman, A., Uitterlinden, A. G., Eilers, P. H., and Lesaffre, E. (2013). Fast linear mixed model computations for genome-wide association studies with longitudinal data. *Statistics in Medicine*, 32(1) :165–180.
- Sitlani, C. M., Rice, K. M., Lumley, T., McKnight, B., Cupples, L. A., Avery, C. L., Noordam, R., Stricker, B. H. C., Whitsel, E. A., and Psaty, B. M. (2015). Generalized estimating equations for genome-wide association studies using longitudinal phenotype data. *Statistics in Medicine*, 34(1) :118–130.
- The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature*, 526(7571) :68–74.
- Vaxillaire, M., Yengo, L., Lobbens, S., Rocheleau, G., Eury, E., Lantieri, O., Marre, M., Balkau, B., Bonnefond, A., and Froguel, P. (2014). Type 2 diabetes-related genetic risk scores associated with variations in fasting plasma glucose and development of impaired glucose homeostasis in the prospective DESIR study. *Diabetologia*, 57(8) :1601–1610.
- Voight, B. F., Kang, H. M., Ding, J., Palmer, C. D., Sidore, C., Chines, P. S., Burt, N. P., Fuchsberger, C., Li, Y., Erdmann, J., Frayling, T. M., Heid, I. M., Jackson, A. U., Johnson, T., Kilpeläinen, T. O., Lindgren, C. M., Morris, A. P., Prokopenko, I., Randall, J. C., Saxena, R., Soranzo, N., Speliotes, E. K., Teslovich, T. M., Wheeler, E., Maguire, J., Parkin, M., Potter, S., Rayner, N. W., Robertson, N., Stirrups, K., Winckler, W., Sanna, S., Mulas, A., Nagaraja, R., Cucca, F., Barroso, I., Deloukas, P., Loos, R. J. F., Kathiresan, S., Munroe, P. B., Newton-Cheh, C., Pfeuffer, A., Samani, N. J., Schunkert, H., Hirschhorn, J. N., Altshuler, D., McCarthy, M. I., Abecasis, G. R., and Boehnke, M. (2012). The MetaboChip, a Custom Genotyping Array for Genetic Studies of Metabolic, Cardiovascular, and Anthropometric Traits. *PLoS Genetics*, 8(8) :e1002793.
- Voight, B. F., Scott, L. J., Steinthorsdottir, V., Morris, A. P., Dina, C., Welch, R. P., Zeggini, E., Huth, C., Aulchenko, Y. S., Thorleifsson, G., McCulloch, L. J., Ferreira, T., Grallert, H., Amin, N., Wu, G., Willer, C. J., Raychaudhuri, S., McCarroll, S. A., Langenberg, C., Hofmann, O. M., Dupuis, J., Qi, L., Segrè, A. V., Hoek, M. v., Navarro, P., Ardlie, K., Balkau, B., Benediktsson, R., Bennett, A. J., Blagieva, R., Boerwinkle, E., Bonnycastle, L. L., Boström, K. B., Bravenboer, B., Bumpstead, S., Burt, N. P., Charpentier, G., Chines, P. S., Cornelis, M., Couper, D. J., Crawford, G., Doney, A. S. F., Elliott, K. S., Elliott, A. L., Erdos, M. R., Fox,

- C. S., Franklin, C. S., Ganser, M., Gieger, C., Grarup, N., Green, T., Griffin, S., Groves, C. J., Guiducci, C., Hadjadj, S., Hassanali, N., Herder, C., Isomaa, B., Jackson, A. U., Johnson, P. R. V., Jørgensen, T., Kao, W. H. L., Klopp, N., Kong, A., Kraft, P., Kuusisto, J., Lauritzen, T., Li, M., Lieverse, A., Lindgren, C. M., Lyssenko, V., Marre, M., Meitinger, T., Midthjell, K., Morken, M. A., Narisu, N., Nilsson, P., Owen, K. R., Payne, F., Perry, J. R. B., Petersen, A.-K., Platou, C., Proença, C., Prokopenko, I., Rathmann, W., Rayner, N. W., Robertson, N. R., Rocheleau, G., Roden, M., Sampson, M. J., Saxena, R., Shields, B. M., Shrader, P., Sigurdsson, G., Sparsø, T., Strassburger, K., Stringham, H. M., Sun, Q., Swift, A. J., Thorand, B., Tichet, J., Tuomi, T., Dam, R. M. v., Haeflten, T. W. v., Herpt, T. v., Vliet-Ostaptchouk, J. V. v., Walters, G. B., Weedon, M. N., Wijmenga, C., Witteman, J., Investigators, T. M., Consortium, T. G., Bergman, R. N., Cauchi, S., Collins, F. S., Gloyn, A. L., Gyllenstein, U., Hansen, T., Hide, W. A., Hitman, G. A., Hofman, A., Hunter, D. J., Hveem, K., Laakso, M., Mohlke, K. L., Morris, A. D., Palmer, C. N. A., Pramstaller, P. P., Rudan, I., Sijbrands, E., Stein, L. D., Tuomilehto, J., Uitterlinden, A., Walker, M., Wareham, N. J., Watanabe, R. M., Abecasis, G. R., Boehm, B. O., Campbell, H., Daly, M. J., Hattersley, A. T., Hu, F. B., Meigs, J. B., Pankow, J. S., Pedersen, O., Wichmann, H.-E., Barroso, I., Florez, J. C., Frayling, T. M., Groop, L., Sladek, R., Thorsteinsdottir, U., Wilson, J. F., Illig, T., Froguel, P., Duijn, C. M. v., Stefansson, K., Altshuler, D., Boehnke, M., and McCarthy, M. I. (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genetics*, 42(7) :579–589.
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., and Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research*, 42(D1) :D1001–D1006.
- World Health Organization (2015). Diabetes Fact sheet No 312.
- Wu, Z., Hu, Y., and Melton, P. E. (2014). Longitudinal Data Analysis for Genetic Studies in the Whole-Genome Sequencing Era. *Genetic Epidemiology*, 38(S1) :S74–S80.
- Yaghootkar, H. and Frayling, T. M. (2013). Recent progress in the use of genetics to understand links between type 2 diabetes and related metabolic traits. *Genome Biology*, 14(3) :203.