An integrated approach for the identification of predictive markers of type 2 diabetes

Estelle Pujos-Guillot, Marion Brandolini-Bunlon, Dhouha Grissa, Yunfei Liu, Mélanie Pééra, Charlotte Joly, Bernard Lyan, Sebastien Czernichow, M Zins, M Goldberg M

AN INTEGRATED APPROACH FOR THE IDENTIFICATION OF PREDICTIVE MARKERS OF TYPE 2 DIABETES

E Pujos-Guillot[a,b], M Brandolini[a], D Grissa[b], Y Liu[b], M Pétéra[a], C Joly[a], B Lyan[a], S Czernichow[c,d,e], M Zins[c], M Goldberg[c,f], B Comte[b*]

[a]INRA, UMR1019, Plateforme d'Exploration du Métabolisme, UNH, F-63000 Clermont-Ferrand, France; [b]INRA, UMR1019, UNH-Mapping, F-63000 Clermont-Ferrand; [c]UMS11 Cohortes en population, INSERM/UVSQ, Villejuif, France; [d]Service de Nutrition, Hôpital A. Paré, APHP, Boulogne-Billancourt, France ; [e]Université Versailles Saint Quentin, Versailles, France, [f]Université Paris Descartes, France.

*E-mail: blandine.comte@clermont.inra.fr

The trajectory and underlying mechanisms of health are determined by a complex interplay between intrinsic and extrinsic factors. There is a need for more accurate assessment of the inputs and their consequences to health (Panagiotou, 2009). Therefore, our objective was to identify accurate and robust multidimensional markers, predictive of type 2 diabetes (T2D). A case-control approach was used within the French cohort GAZEL (n~20,000). Male overweight subjects (n=112, 25≤BMI<30 kg/m², 52-64 y.o.), free of T2D at baseline, were selected and compared for several parameters (clinical, biochemical parameters, and food habits) with Cases defined as having developed T2D at follow-up (5 years later; BMI, age, sex matched). Baseline serum samples were analyzed using mass spectrometry-based untargeted metabolomics. Data mining methods were used to select the best candidate for prediction. Models were built using linear logistic regressions on the reduced dataset and their performances were determined by calculating the area under the receiver operating characteristics curve (AUC) with their 95% confidence intervals (CI), sensitivity and specificity values. Metabolomic data were integrated with the different parameters from the database to determine whether multidimensional models improve prediction. Clinical data and consumption frequencies of vegetables and sugar showed significant differences between Cases and Controls. The metabolomics approach allowed the identification of 5 predictive biomarkers. The resulting metabolite-only model showed better performances than the one built with clinical data: a lower misclassification rate (18% *vs* 26%) and higher AUC (0.82 *vs* 0.74; CI=[0.748-0.892] *vs* [0.659-0.823]). Integration of metabolomic data with the available parameters allowed optimizing prediction performances (10.8% misclassification, AUC=0.89, CI=[0.833-0.950]). Correlation network analyses contributed to explore the links between metabolic, clinical parameters, and food habits. These results show the interest of an integrated approach including untargeted metabolomics in the discovery of predictive biomarkers 5 years before T2D occurrence to better stratify at-risk populations.

KEYWORDS: biomarkers, prediction, cohort