Study of statistical parameters to perform a convenient prediction of different endocrine phenotypes in sportsmen based on metabonomic data

Alain Paris, Boris Labrador, François-Xavier Lejeune, Aziz Zoubai, Cécile Canlet, Jerome Molina, Michel Guinot, Armand Megret, Jean-Christophe Thalabard, Michel Rieu

▶ **To cite this version:**

Alain Paris, Boris Labrador, François-Xavier Lejeune, Aziz Zoubai, Cécile Canlet, et al.. Study of statistical parameters to perform a convenient prediction of different endocrine phenotypes in sportsmen based on metabonomic data. 12. The Yearly French Conference on Bioinformatics - JOBIM, Institut Pasteur [Paris]. Labo/service de l'auteur, Paris, FRA., Jun 2011, Paris, France. hal-01191086

HAL Id: hal-01191086
https://hal.science/hal-01191086

Submitted on 3 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Study of Statistical Parameters to Perform a Convenient Prediction of Different Endocrine Phenotypes in Sportsmen Based on Metabonomic Data

Alain PARIS[1], Boris LABRADOR[2], François-Xavier LEJEUNE[1], Aziz ZOUBAÏ[3], Cécile CANLET[3], Jérôme MOLINA[3], Michel GUINOT[4], Armand MEGRET[5], Jean-Christophe THALABARD[6], Michel RIEU[7] and Yves LE BOUC[2]

[1] Unité Mét@risk, INRA, 16, rue Claude Bernard, 75231 Paris, Cedex 05, France
aparis@paris.inra.fr

[2] INSERM U 938, Centre de Recherche Saint-Antoine, Hôpital Saint-Antoine, Bâtiment Kourilsky, 184, rue du Faubourg Saint-Antoine, 75012 Paris, France

[3] UMR1089 – Xénobiotiques, INRA, 180, chemin de Tournefeuille, BP 93173, 31027 Toulouse, Cedex 3, France

[4] Médecine du Sport, Antenne Médicale de Prévention du dopage Rhône Alpes, Pôle Rééducation et Physiologie, Hôpital Sud, Grenoble, France

[5] FFC, 5, rue de Rome, 93561 Rosny sous Bois, France

[6] Unité Gynécologie-Endocrinologie, Hôtel-Dieu, Assistance Publique-Hôpitaux de Paris, 1, place du Parvis de Notre-Dame, 75004 Paris, France

[7] Agence française de lutte contre le dopage, 229, boulevard Saint-Germain, 75007 Paris, France

Keywords Metabonomics, discrimination, multivariate statistics, prediction.

1 Background

The longitudinal endocrine follow-up of sportsmen achieved by conventional methods allows detection of clinical abnormalities that may be related to some prohibited doping practices. Indeed, some disturbed physical performance can be explained by atypical physiological deviations. However, recent events revealing doping cases have shown some limitations of the principles of anti-doping control currently prevailing in the establishment of doping practices.

Screening of pharmacological substances, which use is prohibited (and some of them are often designed to improve athletic performance), basically consists in detecting presence in urine or blood of these compounds or their metabolites. To do so, the direct detection of doping involves very sophisticated physicochemical methods. Nevertheless, these methods are expensive. In addition, they specifically target known molecules.

Yet, an alternative to this direct strategy is to measure in the serum concentration of circulating endogenous hormones or their metabolites in order to get a hormonal fingerprint of subjects that may give indirect proofs of doping practices.

In the case of endogenous hormones is raised the problem of definition of what are normal concentrations and how to define clinical thresholds. Indeed, when a hormonal doping practice is used, homeostatic regulation may have some repercussion on the hormonal fingerprint. Observation of this hormonal anomaly is the first step in indirect detection of doping practice. In France, it is currently done in the frame of the medical longitudinal follow-up. Besides, these hormonal variations may induce some metabolic adjustments which can be detected in a global metabolic assessment in biofluids. This metabolic fingerprinting is called metabonomics.

In this context, statistical and computational approaches used by metabonomics may be helpful to solve such a problem designed as a numerical analysis of the multidimensional metabolic response when metabolic fingerprints, which correspond to the large quantification of the general metabolism of an organism, are used in complement of hormonal fingerprints. Since the last decade, metabonomics has been efficiently applied and developed in various biological domains including plant genotype discrimination, toxicological mechanisms, disease aetiology, and drug discovery among others. In our context, we aim at improving the ability to predict the endocrine phenotype of any individual based on metabolic fingerprints with the constraint of avoiding the risk of false-negative. This requires fine modelling of the relationship between

metabolic profiles and the endocrine status given by the determination of the hormonal concentration class for three hormones, *i.e.* cortisol, IGF-1 and testosterone.

2 Results

We have studied a cohort of 655 individuals described with 419 metabolites (variables) obtained by ^{1}H NMR (Nuclear Magnetic Resonance) spectrometry from a fingerprinting of serum. For the whole cohort, we also get in parallel three endocrine phenotypes (cortisol, IGF-1 and testosterone), for which 3 classes have been defined *a priori* for "low", "normal" and "high" concentrations. In the procedure presented here, we consider a classification method of any of these classes, which relies to the metabolic fingerprints.

The core method combines a data regularization step based on orthogonal signal correction to a shrinkage discriminant analysis (SDA) [1], which is well suited to deal with the multicollinearity carried out by the metabolites. Thus, in our situation, SDA outperforms other usual discriminant methods such as LDA, QDA and PLS-DA. However, for all these methods, it is noteworthy that classification is substantially improved when data are pre-processed using orthogonal signal correction based on partial least squares regression [2]. To improve the level of confidence on the prediction, assignment to a given class is then obtained using bootstrap techniques. Using bootstrap, we have also studied how the prediction rates vary depending on cohort size, choice of metabolites (variables) and phenotypes.

With the same protocol, we have displayed in abacus the prediction rates obtained with different sizes for the cohort together with increasing the number of selected metabolites. Thus, it can be observed that, for each phenotype, a well suited choice in parameter modelling is achievable to expect the highest rates of prediction for each given hormonal phenotype.

Extensive calculus in the resampling step also revealed special classes of individuals for which the classification systematically failed. Identification of such classes may indicate new further and distinct investigations to improve their hormonal characterization.

3 Conclusion

The procedure combining the orthogonal signal correction based on PLS regression and the shrinkage discriminant analysis is very promising to better detect endocrine disruptions from a metabolomic fingerprint. Abacus appear instructive to decide of how many statistical units we need to consider and therefore to control the cost of such experiments. It is also important to build decision rules from a cohort which is large enough with a selected set of metabolites. A quantitative analysis of changes in general metabolism performed in this physiological context may indirectly provide some tangible biochemical suspicion of doping. In other words, use of indirect methods to routinely phenotype the endocrine status of sportsmen from their metabolomic fingerprints is very promising to better detect endocrine disruptions.

Acknowledgements

This work was supported by AFLD (Agence Française de Lutte contre le Dopage) and WADA (World Anti-Doping Agency).

References

[1] M. Ahdesmäki and K. Strimmer, Feature selection in omics prediction problems using cat scores and false non-discovery rate control. *Ann. Appl. Stat.* 4:503-519, 2010.

[2] S. Wold, H. Antti, F. Lindgren and J. Ohman. Orthogonal signal correction of near-infrared spectra. *Chemometrics Intell. Lab. Syst.*, 44: 175-185, 1998.