



HAL
open science

Metabolomic signatures in Elite cyclists: differential characterization of a seeming normal endocrine status regarding three serum hormones

Alain Paris, Boris Labrador, François-Xavier Lejeune, Cécile Canlet, Jérôme Molina, Michel Guinot, Armand Mégret, Michel Rieu, Jean-Christophe Thalabard, Yves Le Bouc

► To cite this version:

Alain Paris, Boris Labrador, François-Xavier Lejeune, Cécile Canlet, Jérôme Molina, et al.. Metabolomic signatures in Elite cyclists: differential characterization of a seeming normal endocrine status regarding three serum hormones. *Metabolomics*, 2021, 17 (7), 10.1007/s11306-021-01812-4 . hal-03257323

HAL Id: hal-03257323

<https://hal.science/hal-03257323v1>

Submitted on 10 Jun 2021

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.

Metabolomic signatures in Elite cyclists: differential characterization of a seeming normal endocrine status regarding three serum hormones

Alain Paris¹, Boris Labrador^{2,‡}, François-Xavier Lejeune^{2,‡}, Cécile Canlet³, Jérôme Molina^{3,4}, Michel Guinot⁵, Armand Mégret^{6,†}, Michel Rieu⁷, Jean-Christophe Thalabard⁸, Yves Le Bouc⁹

Corresponding author: alain.paris@mnhn.fr

¹ Unité Molécules de Communication et Adaptation des Microorganismes (MCAM), Muséum national d'Histoire naturelle, CNRS, Paris, France

² Institut du Cerveau et de la Moelle épinière (ICM), Sorbonne Université, Inserm U 1127, CNRS UMR 7225, Hôpital Pitié Salpêtrière, Paris, France

³ Axiom, Toxalim, INRAE, ENVT, INPT-EI Purpan, Université Paul Sabatier, Toulouse, France

⁴ Dynamiques et écologie des paysages agriforestiers (DYNAFOR), INRAE, INPT-ENSAT, INPT-EI Purpan, Centre de recherche Occitanie-Toulouse, Auzeville, Castanet-Tolosan Cedex, France

⁵ CHU Grenoble-Alpes, UM Sports et Pathologies, Grenoble, France. Université Grenoble-Alpes, INSERM U 1042, Hypoxia and Pathophysiology Unit, Grenoble, France. UM Sports et Pathologies, CHU Sud, Echirolles, France

⁶ Fédération française de Cyclisme, 1 rue Laurent Fignon, France

⁷ Agence Française de Lutte contre le Dopage (AFLD), Paris, France

⁸ MAP5, Université de Paris, UMR CNRS 8145, Paris, France

⁹ Sorbonne Université, INSERM, UMR S 938, Centre de Recherche Saint-Antoine (CRSA), Paris, France

[‡] BL and FXL contribute equally to the study.

[†] This paper is dedicated to the memory of Dr Armand Mégret who died in September 2019. He was pioneering at the very end of nineties in systematic serum biobanking to help to fight doping practices in sport, particularly in cycling.

Abstract

•Introduction

Serum phenotyping of elite cyclists regarding cortisol, IGF1 and testosterone is a way to detect endocrine disruptions possibly explained by exercise overload, non-balanced diet or by doping. This latter disruption-driven approach is supported by fundamental physiology although without any evidence of any metabolic markers.

•Objectives

Serum samples were distributed through *Low*, *High* or *Normal* endocrine classes according to hormone concentration. A ^1H NMR metabolomic study of 655 serum obtained in the context of the longitudinal medical follow-up of 253 subjects was performed to discriminate the three classes for every endocrine phenotype.

•Methods

An original processing algorithm was built which combined a partial-least squares-based orthogonal correction of metabolomic signals and a shrinkage discriminant analysis (SDA) to get satisfying classifications. An extended validation procedure was used to plan in larger size cohorts a minimal size to get a global prediction rate (GPR), *i.e.* the product of the three class prediction rates, higher than 99.9%.

•Results

Considering the 200 most SDA-informative variables, a sigmoidal fitting of the GPR gave estimates of a minimal sample size to 929, 2346 and 1408 for cortisol, IGF1 and testosterone, respectively. Analysis of outliers from cortisol and testosterone *Normal* classes outside the 97.5%-confidence limit of score prediction revealed possibly i) an inadequate protein intake for outliers or ii) an intake of dietary ergogenics, glycine or glutamine, which might explain the significant presence of heterogeneous metabolic profiles in a supposedly normal cyclists subgroup.

•Conclusion

In a next validation metabolomics study of a so-sized cohort, anthropological, clinical and dietary metadata should be recorded in priority at the blood collection time to confirm these functional hypotheses.

running title:

¹H NMR metabolomics-assisted doping detection of cyclists

keywords:

Cortisol, IGF1, Testosterone, Doping, Diet, SDA

abbreviations:

AAS: Androgenic Anabolic Steroids

ABP: Athletics Biological Passport

AFLD : Agence Française de Lutte contre le Dopage – French Anti-doping Agency

ANOVA: ANalysis Of VAriance

BATMAN: Bayesian AuTomed Metabolite Analyser for NMR spectra

cat scores: correlation-adjusted t-scores

CPMG: Carr-Purcell-Meiboom-Gill NMR sequence

E: Epitestosterone

FFC: French Federation of Cycling

GH: Growth Hormone

GPR: Global Prediction Rate

IC: Independent Component

ICDA: Independent Component – Discriminant Analysis

IGF1: Insulin-like Growth Factor 1

LS: Learning Size (number of samples selected in the dataset)

MAGIC: Markov Affinity-based Graph Imputation of Cells

MANOVA: Multivariate ANalysis Of VAriance

MDS: MultiDimensional Scaling

NV: Number of selected ordered Variables

nls: non-linear squares fitting

OSC: Orthogonal Signal Correction

PCA: Principal Component Analysis

PHATE: mapping using Potential of Heat-diffusion Affinity-based Transition Embedding

PLS-DA: Partial Least Squares-Discriminant Analysis

PLSR: Partial Least Squares Regression

PR: Prediction Rate

SDA: Shrinkage Discriminant Analysis

T: Testosterone

UCI: Union Cycliste Internationale – International Cycling Union

WADA: World Anti Doping Agency

Author contributions

AP, MG, JCT and YLB supervised this work. JCT, YLB, MR, MG and AP contributed to design the experiment. AM and MG contributed to sample handling and collection and sample biobanking. MG achieved endocrine phenotyping and curation. CC and JM performed metabolomic analyses. FXL, BL and AP performed statistical analyses. CC did an expert-based assignment of putative biomarkers. AP performed the data interpretation and wrote the manuscript. JCT, YLB, MG, MR, CC, FXL and BL revised the manuscript. All authors have read and approved the manuscript.

Acknowledgements

Authors thank sportsmen enrolled in the cohort of the French Federation of Cycling for having accepted to be anonymously involved in this exploratory phenotyping study. The French Federation of Cycling is acknowledged for his involvement in the study and providing samples. This study was supported by the French Agency for Doping Control (AFLD, contract Inserm n° R09223DD) and the World Anti-Doping Agency (WADA, contract n° 006D1442 – Inserm n° R07157DD). B Labrador was supported by AFLD and WADA and FX Lejeune by the French Agency of Research.

Compliance with ethical standards

Conflict of interest

The authors have no conflict of interest to declare.

Ethical approval

The samples concerning metabolomic research were carried out during the mandatory medical and biological monitoring of the French Federation of Cycling (FFC) according to French legislation (article L 3621 of the French Public Health Code concerning protection of the health of sportsmen and women). The aliquots of blood samples reserved for this project were anonymized with a specific consent signed by those who gave their consent (WADA research grant and project submitted by WADA to an independent ethic review, reference number 06D14YL).

Informed consent

Informed consent was obtained from all participants included in the study.

Supplementary material

Metabolomics_AParis_Revised_Final_SI_05072021.pdf

Data available at www.ebi.ac.uk/metabolights/MTBLS2799

1. Introduction

Usage suspicions, clinical or analytical proofs and continuous improvements of doping usage in sportsmen milieu have progressed in parallel to the development of new health technologies for the four last decades (Robinson et al. 2017). Even though new detection methods have arisen to control fraudulent use of doping agents, there is still a large discrepancy between the doping prevalence compared to the true analytical or clinical detection statistics (de Hon & van Bottenburg 2017). This has encouraged the use of indirect methods based on the longitudinal individual follow-up of either biochemical markers (Sottas et al. 2008) or physical performances (Schumacher & Pottgiesser 2009; Iljukov & Schumacher 2017; Montagna & Hopker 2018). Biochemical markers have been at the origin of the Athletics Biological Passport (ABP) (Robinson et al. 2007; Sottas et al. 2008, 2011) which progressively has integrated different modules (Robinson et al. 2017). World Anti-Doping Agency (WADA) adopted the ABP concept in 2009 through the Haematological Module and since this date defined regularly updated guidelines (WADA 2020). This module aims blood doping detection (Sottas et al. 2010a) thanks now to more than 10 haematological markers; the International Cycling Union (UCI), has first implemented it in 2008 to detect blood doping in elite sportsmen on different road cycling events (Zorzoli & Rossi 2010). Androgenic anabolic steroids (AAS), glucocorticoids (in competition), and growth hormone (GH) are banned substances in agreement with antidoping rules (Sottas et al. 2010b). Suspicions of AAS doping are supported by the Steroidal Module (Sottas et al. 2008), included in 2014 in the ABP, which consists in the individual longitudinal follow-up of urinary testosterone (T), epitestosterone (E), a minor metabolite of T, the ratio T/E, and four other T metabolites and related ratios used as associated covariates (WADA 2020). Doping with GH, insulin growth factor-1 (IGF1) and other growth factors can be detected thanks to the Endocrine Module of ABP, which still needs further validation before implementation in laboratories accredited by WADA (Alladio et al. 2016). Glucocorticoids are prohibited in sports competition and a list of forbidden substances is periodically updated (WADA 2020). Either exogenous cortisol or synthetic glucocorticoids are detected in urine. A depleted cortisol production in a context of physiologically depressed adrenal functions has been well characterized in multiple clinical situations (Duclos et al. 2007; Cadegiani et al. 2019). So, a deeply altered regulation of the hypothalamo-pituitary-adrenals axis (Younes & Younes 2017) in response to use of exogenous glucocorticoids (Broersen et al. 2015) may compromise recovery from injury and thus may be a risky metabolic situation for the concerned sportsmen.

All doping detection methods introduced in ABP modules are based on a univariate approach requiring a Bayesian modelling of the target biomarker (Sottas et al. 2007). More precisely, such a modelling is integrated in a Bayesian network which integrates covariates; a longitudinal monitoring of the target biomarker results in a reduced variability with acceptable range adjusted at a 99- or 99.9%-credibility value for every subject (Robinson et al. 2017). Therefore, for every athlete, after a few samplings, the measured biomarker at time t is compared to the estimated limits of the personalized acceptable range calculated from previous samplings. Hence, points beyond this acceptable range are suspected to result from an abnormal physiological situation; this provides a first objective detection of an abnormal blood profile. A multivariate approach of the detection of abnormal blood profiles was attempted using exclusively a limited set of functionally related variables either for blood doping (Sottas et al. 2006) or androgenic/anabolic steroids (Van Renterghem et al. 2013; Alladio et al. 2016; Ayotte et al. 2017; Amante et al. 2019) detection. In the latter case, this was possible thanks to a deep knowledge of metabolism of androgens (Pozo et al. 2010; Fabregat et al. 2016) and their concentration normally found in control populations (Van Renterghem et al. 2010; Thevis et al. 2007). Therefore, these targeted variables represent a restricted set of *a priori* highly correlated variables; extended sets of steroids can be acquired now by steroidomics-based methods (Jeanneret et al. 2016; Ponzetto et al. 2018, 2019), hence extending the number of putative steroidal biomarkers.

Since late 2000s, metabolomics has been repeatedly invoked as a promising approach to reinforce doping detection (Reichel 2011; Pitsiladis et al. 2014; Robinson et al. 2017; Wang et al. 2017). Most of recent metabolomics studies published in sports describe physiological changes, more precisely homeostatic adjustments of metabolism resulting from energy use in a sports context of simulated altitude (Lawler et al. 2019) or physical exercise charge in a context of strenuous long distance exercise like in endurance (Shi et al. 2020) or ultra-endurance sports (Howe et al. 2018). But, to our knowledge, none of these published studies has dealt with the doping detection question. However, in a pioneering physio-metabolomic study, metabolic disruptions induced by a controlled administration of steroidal anabolics in a farm animal model has proved its

interest to get discrimination of the different groups (Dumas et al. 2002a, 2002b), either treated or not, and, at the same time, detection of fully comprehensive biomarkers (Dumas et al. 2005) which were supported by numerous targeted studies performed for more than 40 years, since the late fifties, on a plethora of documented biochemical anabolic effects. Therefore, by using some indirect endocrine biomarkers, it is virtually possible to get a classification of sportsmen into two different classes, either normal or abnormal, taking into account a given endocrine biomarker and, at the same time, a large set of metabolic variables supporting this classification. Yet, as shown hereafter in a sportsmen cohort study, we had to face methodological problems related to i) the large imbalance between normal and abnormal classes inside a given endocrine phenotype, ii) the relative part of metabolomic information shared between the endocrine phenotype-related part and the unrelated one, iii) the quality of the resulting discrimination and, then, the validation of the classification obtained thanks to discriminants, and last iv) the homogeneity status of the subgroups of subjects enrolled in a seemingly normal class regarding every endocrine phenotype studied here. All these questions were examined from a methodological side in a cyclists cohort prior submitted to endocrine phenotyping of the three following hormones: cortisol, IGF1 and testosterone. In the present pilot study, these questions were explored using an extended multivariate statistical approach described in four main blocks (Figure 1). From this extended datamining approach and because we used a generic metabolomic one, which is based by principle on measurement of endogenous metabolic biomarkers, and from the significant detection of some metabolic deviations of putative biomarkers in few athletes whose endocrine profiles were kept *a priori* as normal, we were able to characterize a seeming outlier situation which may indicate also that other environmental causes like exercise overload or non-balanced diets should be considered to explain metabolic deviations which could be detrimental to get better predictions of some normal individuals and, therefore, may falsely superimpose to potent doping practices. Indeed, this metabolomic exploration focused on three endocrine phenotypes was rendered possible by using an in-depth datamining strategy consisting in i) an original discriminant analysis associating a prior partial least squares-based orthogonal signal correction applied once to the raw dataset and a shrinkage discriminant analysis, and ii) a bootstrap procedure to repeatedly study variation of global prediction rates according to the size of learning datasets and the number of prior selected variables used to build these datasets (Figure 1).

2. The cohort of elite cyclists

The present study corresponds to a retrospective exploratory cohort study performed on sera collected prospectively from French elite or professional French cyclists participating in national or international competitions. These cyclists, like all cyclists in French teams, were mandatory requested to undergo 3-4 times a year a physical examination including a blood collection according to a standardised protocol (Guinot et al. 2007). Samples were stored in a biobank built in the frame of this large longitudinal mandatory medical follow-up of cyclists placed under the responsibility of the French Federation of Cycling (FFC).

Since the objective of the metabolomic study was to detect anomalous metabolic profiles related to a suspected use of three types of banned substances as revealed from a prior analysis of abnormal variations of cortisol, IGF1 or testosterone concentrations in serum, the biobank used in the study was built to merge as numerous disrupted hormonal profiles as possible. Different groups of individuals according to serum values which were either lower, inside or higher than normal concentration limits were built as shown in the Supplementary information (see § 1). For every endocrine phenotype parameter, each subgroup of samples was considered as belonging to a homogeneous endocrine class. By convention, *Normal* was designating a hormone concentration value placed inside the normal hormonal concentration range whereas *Low* and *High* corresponded to hormonal concentrations which were respectively lower and higher than normally awaited in healthy young men.

The cohort comprising 655 samples was built from 253 subjects (see Supplementary information § 1). All samples were obtained anonymously according to the French legislation concerning the scientific follow-up of human populations. Every subject was asked for a complementary express consent concerning the use of an anonymized aliquot of blood sample for this specific study, out of the initially collected blood samples for his regular medical follow-up.

A preliminary analysis of variance (ANOVA) restricted to the blood samples belonging to the *Normal* groups did not show any significant effect of age on hormonal concentrations for the three endocrine traits (Figure S1 and Supplementary information) except for IGF1, which is known to decrease with age. For any of the three

endocrine traits studied, no significant interaction between any class (*Low*, *Normal*, *High*) of the three endocrine factors (cortisol, IGF1, testosterone) was detected to explain the distribution through classes of serum concentrations of any of the three hormones according to the two others (not shown). In addition, when discretizing the year period (months) at the time of blood collection, no sub structuring of metabolomic data could be observed from a PCA, neither nor any cyclist discipline (Figure S3). Therefore, every hormonal profile was processed independently from the two others on every sample collected on all subjects enrolled in the cohort, hence increasing the power of statistical analyses. For the ease of interpretation of results, the term ‘subject’ was referring to every cyclist enrolled in the cohort when the terms ‘individuals’ and ‘statistical individuals’ were strictly referring to samples analyzed in the study. The term ‘sample’ was reserved only for parts of the text dealing with pure analytical points.

3. Materials and Methods

3.1. ¹H NMR fingerprinting

All ¹H NMR spectra were acquired on a Bruker Avance DRX-600 spectrometer (Bruker SA, Wissembourg, France) operating at 600.13 MHz for ¹H resonance frequency, and which was equipped with a pulsed field gradients *z* system, an inverse ¹H-¹³C-¹⁵N cryoprobe attached to a cryoplatfrom (the preamplifier cooling unit), and a temperature control unit maintaining the sample temperature at 300 ± 0.1 K. Fingerprints were obtained using the CPMG NMR sequence and raw data was processed as detailed in Supplementary information, § 2.1. The global metabolomic exploration of the cohort was done into two sessions, two years apart, with 368 and 287 sera phenotyped, respectively.

3.2. Statistical analyses

3.2.1. Pre-processing and non-supervised data explorations

Eight hundred eleven quantitative variables corresponding to ¹H NMR chemical shift buckets were obtained for every spectrum (Figure 1, Block 1). Among these variables, a subset of the most informative variables (*n* = 419) was selected using multidimensional scaling (MDS) repeatedly performed on the transposed matrix as published previously (Rohart et al. 2012). Dataset was then log₁₀-transformed.

To study in an unsupervised way the metabolic phenotypes related to the different endocrine traits and the inherent information structure which was then accessible at the metabolome level, Principal Component Analyses (PCA) and mapping using Potential of Heat-diffusion Affinity-based Transition Embedding (PHATE) were performed using R 3.6.0 (R Core Team 2019) with the respective R packages mixOmics (Rohart et al. 2017) and phateR (Moon et al. 2019). PHATE was applied on MAGIC (Markov Affinity-based Graph Imputation of Cells) imputed data thanks to the R package Rmagic (van Dijk et al. 2018) as detailed in Supplementary information, § 2.2.1 (Figure 1, Block 1).

3.2.2. Discrimination procedures

A robust statistical procedure, efficient enough to minimize the prediction error in the assignment of unknown subjects to a given hormonal class (either *Low*, *Normal* or *High*) of a specific endocrine phenotype was achieved from metabolomic profiles. Metabolites were given in a 655 × 419 spectral matrix *X* for the whole cohort. The *Y* response vector represented the concentration classes of cortisol, IGF1 or testosterone assayed in either the *Low*, *Normal* or *High* class.

For the three endocrine phenotypes, the classification of every unknown sample from his metabolomic profile into one of the three classes could be computed according to a prior discrimination rule established from a learning (training) set of documented samples. In this aim, we first compared partial least squares-discriminant analysis (PLS-DA), a classification method based on partial least squares regression (PLSR, Barker & Rayens 2003), to a shrinkage discriminant analysis (SDA, Ahdesmäki & Strimmer 2010) for their respective ability to handle the problem of multicollinearity, which typically arises when the explanatory variables are numerous (Figure 1, Block 2, see Supplementary information § 2.2.2). Besides, a single prior Orthogonal Signal Correction (OSC) was used to remove on the complete matrix *X* the variation orthogonal to the *Y* response vector (Figure 1, Block 2, see Supplementary information § 2.2.2). The final algorithm used in further analyses consisted then in the combined use of OSC and SDA.

3.2.3. Validation procedure

Further validation of the classification process was done considering i) the Number of ordered and selected Variables (NV) according to the cat score (Zuber & Strimmer 2009) given by SDA and ii) the Learning Size (LS, number of statistical individuals selected in the dataset) (Figure 1, Block 3). For every combination of NV and LS parameters, discrimination rules and prediction of the test dataset were obtained at each step of a bootstrap procedure (see Supplementary information, § 2.2.3). The bootstrap-training set was selected from the original sample population and remaining samples were used as a test set. Performance validation was based on the rate of correct classification of every individual to its respective endocrine class as estimated from the whole resampling. Therefore, a prediction rate (PR) was calculated for every class for a given combination of NV and LS parameters and a global prediction rate (GPR) was then obtained as a stringent criterion from multiplication of the three PR values.

To further explore what should be the lower cohort size limit required to get a GPR fixed at a value β , with $\beta \leq 1$, we performed a fitting analysis of the GPR curves as a function of the sample size using either 200 or 400 variables included in the training dataset (Figure 1, Block 3). To take into account misclassified individuals (with lower classification rates across the bootstraps) in the determination of β , the curve fitting was applied to the 5% quantile values (Q5) of GPRs. A sigmoid function was used to fit the sigmoidal variation between the Q5 value of GPR (y) as a function of the training cohort size (x) within a 50-600 size range. Details of fitting according to different sigmoidal models are given in Supplementary information, § 2.2.3.

3.2.4. Relative quantification of targeted metabolites

The relative quantification of metabolites was performed using the R package BATMAN (Hao et al. 2014), an acronym for Bayesian AuTomed Metabolite Analyser for NMR spectra, which deconvolutes peaks from 1-D ^1H NMR spectra to automatically assign a chemical shift to specific metabolites from a target list of 109 molecules and then to estimate their respective relative concentrations. This was done as published previously (Sotton et al. 2017) and a final set of 58 metabolites quantified from ^1H NMR files (Supplementary information, Table S4) was submitted to multivariate statistics (Figure 1, Block 4). Empirically, from our past experience, we noticed that using a number of candidate metabolites higher than the size of the final metabolite dataset is a simple way to increase for some chemical shift windows the specificity of quantification of metabolites.

3.2.5. Complementary modelling methods

Search of sets of variables describing contrasts according to a single independent component (IC) between two distinct sub-populations was carried out using the Independent Component – Discriminant Analysis (ICDA) as published previously (Habchi et al. 2017) (Figure 1, Block 4). Pathway enrichment analysis was performed with MetExplore, a web server freely available at <https://metexplore.toulouse.inra.fr/metexplore2/> using *Homo sapiens* KEGG pathways, which allows mapping of data from metabolomic experiments onto genome-scale metabolic networks (Cottret et al. 2018). Putative metabolic biomarkers were used to extract pathways enriched into influential or differential metabolites from lists of biomarkers previously selected by cross-validation procedures (Figure 1, Block 4).

4. Results

Serum samples from anonymous elite cyclists were analyzed by ^1H NMR through two different NMR sessions to get a dataset comprising 655 samples collected from 253 cyclists, some being collected at different dates in a longitudinal follow-up frame. This dataset constituted both a training dataset for a varying proportion of statistical individuals and a validation one for the remaining part; training dataset was repeatedly sampled through 1000 bootstraps. A representative metabolic fingerprint of serum collected on one anonymous cyclist is shown in Figure S2. Some chemical shifts were assigned thanks to a prior knowledge obtained on the normal biochemical composition of human serum accessible to ^1H NMR detection. Amino-acids like L-alanine, L-lysine, L-valine, L-glutamine and L-glutamate, and lipids or sugars were detected among all assigned analytes.

Since i) no explicit information structure regarding the different subgroups of endocrine phenotypes built from merging classes of the three hormones was accessible by either PCA, MDS or MAGIC-PHATE analysis (for details see Supplementary information § 3.1 and Figures S3 & S4) and ii) all statistical individuals were considered as independently selected from a larger population, even though a significant proportion was

collected from subjects placed in a medical follow-up frame (see Supplementary information, § 3.1, for details), separation of sample classes was achieved only by a discrimination procedure linking OSC-correction of matrix X to SDA as efficient algorithm (see details in Supplementary information, § 3.2 and Figure S5). In addition, when a permutation test ($n = 1000$) was applied to training datasets containing the 200 most informative variables built from 400 individuals, the F-values resulting from a MANOVA test applied to SDA components in the conditions of a random distribution of class assignment were significantly lower ($p < 2 \times 10^{-16}$) than true values (Figure S6). Therefore, an extensive validation done with a bootstrap procedure ($n = 1000$) examining the respective weights of the size LS of the learning dataset and the number NV of variables selected from their cat score was then performed.

4.1. Variation of the prediction rate according to the training set size

Firstly, for a given endocrine phenotype, the validation of the algorithm was performed according to the size LS of the training set. At each iteration, performance classification inside the three classes was calculated. At this step, no prior selection of variables was done. Evidently, for every hormonal trait, as the size of the training set increased, prediction error rates decreased whatever the class considered was (see Figure S7). However, for every hormonal trait, a higher prediction rate was obtained for the *Normal* class, even for very low LS sizes, than for the two other classes. This was easily explained by the relative higher proportion this class had in the cohort. The same situation was found for the class displaying a higher than normal serum cortisol concentration (class *High*) for which the initial LS size was nearly the same as the *Normal* one (Figure S7A). More interestingly was the regular increase of prediction rates for the two abnormal classes in the different hormonal traits studied. Nearly 300 statistical individuals were necessary to significantly lower these two error rates (error rates less than 40%, *i.e.* prediction rates above 60%) for every hormonal trait. The same qualitative variation in the prediction rate was shown with IGF1 for which a prior correction was done using four orthogonal components instead one (Figure S7D).

The overall prediction performance estimate of the present classification method was calculated through the global prediction rate (GPR), which took into account class sizes; classes were extremely imbalanced and this led to much contrasted error prediction rates between classes. A 90%-confidence interval of GPR could be easily estimated thanks to the prior resampling procedure used ($n = 1000$). Clearly, this GPR became maximal for cortisol, IGF1 and testosterone for a LS value of nearly 400, 600 and 600, respectively (Figures S7A, S7B & S7C). When GPR was evaluated for IGF1 considering a prior correction of the dataset with four orthogonal components, it became maximal with a LS value estimated to 450 individuals (Figure S7D). In addition, we could notice for IGF1 and testosterone (Figures S7B & S7C) a decrease in the median GPR values between 600 and 650 samples. This corresponded to the presence of an over-informed situation involving the improper presence of non-informative (noisy) variables as, at this step, none prior variables selection was done. In addition, a part of the over-information is also explained by the unique prior OSC step applied to all statistical individuals. Therefore, as shown hereafter, the validation procedure was completed by taking into account a varying number of variables (parameter NV) that were previously selected on their cat score.

4.2. Variation of GPR according to the number of selected variables

The cat score parameter calculated prior to SDA is useful to sort the different variables according to the relevant information they bring to the classification procedure. Most informative variables according to their cat score were privileged for constituting the working metabolomic dataset. So, GPR was calculated according to both LS, the training set size varying from 50 to 650 statistical individuals, and NV, the number of selected variables varying from 50 to 400 variables (Figure 2). For every hormonal trait, considering only 50 variables was insufficient to get remarkable GPR. Only 200 variables measured on 300 to 350 individuals were sufficient to get a median GPR higher than 99% for cortisol (Figure 2A). For both response factors Y of IGF1 and testosterone used to prior correct the metabolomic dataset with a one-component OSC, maximal median GPR were obtained with 400 variables and 600-650 individuals; and were estimated to 91 and 93%, respectively (Figures 2B & 2C). A more important correction using four orthogonal components used to increase the prediction performance of the IGF1 phenotype could lower the training set size to about 350 individuals with 400 variables or to 300 individuals with only 200 variables (Figure 2D). Again, the over-informed situation was well reflected in the median GPR curves obtained with 400 variables, which were lower than those obtained with only 200 variables, and/or when the training dataset had less than 300-350 individuals, but higher when the

training dataset had more than 300-350 individuals. This indicated clearly that a minimal size of the training cohort was necessary if we wanted to get a median GPR higher than 99% and to use most of informative variables for a given hormonal trait.

4.3. Estimation of a minimal cohort size to get a GPR above 99.9%

The necessary training size of datasets built by using 200 or 400 variables preselected according to their cat score could be tentatively estimated directly from data as in the case of cortisol to get a 5%-quantile (Q5) value of GPR defining the lower limit of the GPR range, which would be above 99.9% (see above). However, for the two other endocrine phenotypes, estimation of a minimal cohort size to get a GPR higher than 99.9% required an estimation based on a robust statistical fitting (Figure 1, Block 3). To minimize prediction error rates, we used curves built from the Q5 values and tried to predict the minimal cohort size for which Q5 GPR values predicted from the fitted sigmoidal curves were above 99.9% (Table 1).

With most informative learning datasets prepared from the 400 preselected variables, the non-linear least squares modelling of sigmoidal curves was efficient enough to get GPR higher than 99.9% for only cortisol with 1 OSC and IGF1 with 4 OSC components used to prior correct datasets (Table 1). For cortisol with 1 OSC component, the size of the present cohort used here ($n = 646$) was higher than the lower estimated limit, *i.e.* 433 and 503 depending on the model used, respectively models (3) and (5) with for this latter a polynomials degree equal to 1 (Table 1). For IGF1 with 4 OSC components, the minimal size requirement based on the sigmoidal curve model (4) was estimated to 603. When it was based on model (5) with polynomial degrees equal to 1 or 3, the minimal size was estimated to 836 or 820, respectively (Table 1). With testosterone and IGF1 datasets submitted to only 1 OSC correction, no direct sigmoidal curve fitting using models (1) to (4) displayed a GPR prediction higher than 99.9% (Table 1). Nevertheless, prediction of a minimal size could be obtained indirectly from a multiple regression between $\text{logit}(y)$ and polynomials x^i using the model (5). For testosterone, the minimal size varied from 963 to 1342 depending on the degree of the polynomials used to fit the curve, respectively 3 and 1 (Table 1). For IGF1, the minimal size estimated grew up to 1913 with a polynomials degree of 1 (Table 1).

In brief, except for cortisol, the present size of the cohort giving a dataset, which was built from the 400 most informative variables and corrected with 1 OSC, was far from being sufficient to repeatedly get a GPR higher than 99.9% since required cohort sizes were estimated to 1913 and 963 statistical individuals for IGF1 and testosterone, respectively. For the dataset, which was built from the 200 most informative variables and corrected with 1 OSC, at least 929, 2346 and 1408 statistical individuals were necessary to repeatedly get a GPR higher than 99.9% for cortisol, IGF1 and testosterone, respectively.

4.4. Factorial exploration of the predictability of classes inside a given endocrine phenotype

As anticipated from the study of median PR inside every class in a given endocrine phenotype, the GPR variation was highly depending on the size LS of the learning dataset, but also on the disequilibrium between numbers of individuals per class (Figure 3). But this global pattern did not give any indication on specific behaviour of some subsets of individuals which at this step were not yet considered as possible outliers. Nevertheless, the bootstrap procedure used here gave possibility to analyse at the individual scale the variation of the prediction performance when the parameter LS varied from 50 to 600, considering here the dataset prepared from the 200 most informative variables preselected on their cat score. Because the prediction performance was fully documented for every statistical individual all along the different prediction tests done by varying the size LS, a PCA could be easily applied (Figure 3).

For the four datamining situations corresponding to cortisol, testosterone and IGF1 datasets corrected with 1 OSC component and to IGF1 dataset corrected with 4 OSC components, the *Normal* class was projected with low absolute values of scores, both for the principal components 1 and 2 (PC1, PC2). This denoted that this class was predicted with a lower error risk by using relatively low-sized learning datasets. This meant also that with highly imbalanced classes, low-sized classes needed a larger learning dataset containing the necessary atypical metabolomic information to get higher quality predictions. This was true for testosterone and IGF1 phenotypes for which classes *High* and *Low* were equally under-represented (Figures 3B, 3C & 3D). For cortisol, only the *Low* class was under-represented (Figure 3A). Translated at the score plot level, these under-represented classes were projected on higher range values of PC1 and PC2. But, what was interesting with this synthetic datamining procedure was the possibility to detect, even in the most abundant classes like *Normal* ones, some statistical individuals that could still 'resist' to a satisfying prediction contrary to other individuals of the same membership.

This might attest existence of some probable metabolomic peculiarities for these outliers that were not shared by most of individuals belonging to these *Normal* classes. These outliers were outside the minimal 97.5% confidence limit, whatever the endocrine trait considered.

4.5. Anomalous metabolic profiles inside seemingly normal endocrine populations

For every endocrine phenotype, from performances of PCA applied to classification statistics, it was therefore possible to graphically select a subset of statistical individuals belonging to the *Normal* class that were considered as outliers (see above and Figure 3). Consequently, it was possible to select some common ^1H NMR variables characterizing these subgroups that might explain this atypical predicted endocrine trait arising inside the *Normal* endocrine phenotype. This was obtained using the ICDA procedure based on an independent component (IC) analysis joined to a discriminant analysis (Habchi et al. 2017) (Figure 1). Then, a correlation was calculated between the resulting IC and metabolites which relative concentrations were estimated independently by the BATMAN procedure from raw ^1H NMR fingerprints. The resulting BATMAN datasets were restricted to integrated metabolites which displayed a non-null variance and a mean above 0.001 as expressed in relative units. They covered 58 metabolites for the three endocrine phenotypes (Table S4). They were then OSC-corrected by taking into account two orthogonal components; the remaining part of variance corresponded to 53.2, 60.3 and 53.6% of the total initial variance for cortisol, IGF1 and testosterone metabolite datasets, respectively. Then, a bootstrap procedure was performed ($n = 1000$) from sampling the same number of normal individuals as outliers. This was a way to get repeatedly i) correlation between the variables present in metabolite datasets and the IC_λ calculated from the ^1H NMR bucket dataset, for which the one-way ANOVA between the normal and subgroups of outliers was maximal, ii) $F_{1,p}$ and iii) related p-values in the ANOVA of every 58 BATMAN-integrated metabolites. Quantiles in the 0.05-0.95 confidence range were obtained from bootstrapped data (Table 2).

From the *Normal* class, we detected for the cortisol, IGF1 and testosterone phenotypes, respectively, 8, 9 and 9 statistical individuals who could be considered as being non-representative enough to get a similar classification behaviour according to the progressive increase in the learning dataset size (Figure 3). Therefore, they could be used as atypical subsets to detect among them some common putative metabolic biomarkers that could explain significantly the construction obtained by ICDA of a single IC_λ which explained a significant differentiation between true *Normal* individuals and the subset of outliers inside the *Normal* class. From the bootstrap-based validation procedure, a unique IC_λ was obtained which represented the best contrast between the two balanced subsets of samples as revealed by a simple one-way ANOVA whatever the endocrine phenotype considered. A list of highly significant δ variables with a statistical score equal to or higher than 900 or 883 over 1000 was obtained for cortisol and testosterone, respectively (Table 2). By contrast, only three δ variables with a statistical score ranging from 359 to 455 over 1000 were detected for the construction of a unique IC_λ characterizing the differentiation between normal individuals and outliers for IGF1 (Table 2). For cortisol, bucketed δ variables, which displayed scores equal to or higher than 900 over 1000, were tentatively assigned to L-valine, L-glutamine, lipids, glucose and citrate using an in-house metabolic database containing chemical shifts for most important metabolites (Table 2). For testosterone, bucketed δ variables displaying scores equal to or higher than 883 over 1000 were tentatively assigned to L-threonine, L-arginine, choline and α -glycerylphosphorylcholine (Table 2).

Interestingly, when using, with default modelling parameters, a PHATE-supported factorial analysis performed on the *Normal* metabolomics subsets submitted to one OSC, outliers detected in the cortisol *Normal* class were classified in a unique subclass (Figure S8A) among the four possible classes of metabolomic heterogeneity that were used in the model, when outliers for testosterone and IGF1 were distributed among two and four classes, respectively (Figures S8B & S8C). The number of four classes describing the metabolomic heterogeneity inside the *Normal* class corresponded to the larger number of heterogeneous classes possibly used for cortisol trait for which outliers were projected in a unique subclass.

Concerning the metabolites integrated by the BATMAN procedure from raw ^1H NMR datasets for the three endocrine phenotypes, which were correlated to IC_λ and were calculated at the previous step, only the cortisol trait could be characterized by a significant subset of two putative metabolites, glycine and 7-methylxanthine, which displayed a significant statistical score equal to or higher than 902 over 1000 (Table 2). For testosterone, three metabolites were detected with a very low statistical score (less than 18 over 1000). This meant that less

than 2% of samplings gave a significant correlation between IC_λ and these assigned metabolites. Clearly, for testosterone, even though an IC could be repeatedly calculated from δ variables, no metabolite present in the BATMAN-based dataset was correlated to a subset of δ variables. For IGF1, no metabolite present in the BATMAN-based dataset was correlated to IC_λ and none could be directly used to calculate a significant IC (Table 2).

Finally, from this bootstrap-supported validation of the possible existence of outliers in the *Normal* subset of individuals, only for the cortisol endocrine phenotype, two metabolites generated by the BATMAN-integration of raw ^1H NMR data were considered as statistically significant to base on them a possible analysis of the deviation of metabolism in these outliers. These were glycine and 7-methylxanthine for which a complementary unbalanced ANOVA could be further done and gave highly significant results (Table 2).

5. Discussion

5.1. Parameters influencing GPR

A high complementarity emerges from combination of PLS-based correction (OSC) of ^1H NMR data and shrinkage in Mahalanobis distance calculation which results in a higher discrimination efficiency than by using PLS-DA, and hence on GPR. Indeed, this chemometric property had never been noticed before. Clearly, the noisy information, which corresponds to the part of information discarded by a PLS-based OSC, is highly detrimental to obtain high classification performances by SDA when no correction was applied to the initial dataset. This efficiency is mainly explained by the James-Stein shrinkage estimate of correlations, variances and priors used together to calculate Mahalanobis distances (Ahdesmäki & Strimmer 2010), and hence probabilities to predict a final class attribution of unknowns (see details given in Supplementary information, § 4.1). In addition, James-Stein shrinkage used in SDA not only improves classification performance but also ensures the interpretability of fitted models by setting some parameters to zero (Samworth 2012). Yet, in an extended validation procedure applied to a larger cohort, a systematic OSC would be achieved prior to SDA in every iteration of the bootstrap procedure to better control some sources of over-information. Classification performances were higher for cortisol than for testosterone and, lastly, for IGF1. For this latter endocrine phenotype, four OSC were necessary to get a satisfying GPR; this showed for IGF1 phenotype a more ambiguous differential metabolic signature between classes than for the two other endocrine phenotypes. Probably, the IGF1 phenotype (Holt & Sönksen 2008) could be now better characterized by using simultaneously and complementarily two other biomarkers of GH action, *i.e.* procollagen III peptide (Holt et al. 2010; Lehtihet et al. 2019) and fibronectin 1 (Sieckmann et al. 2020). Moreover the qualitative status of either hypo- or hyper-secretion of cortisol, IGF1 and testosterone, respectively, is also probably not defined with the same qualitative clinical consensus. Indeed, the three endocrine phenotypes were highly contrasted in their respective minimal cohort size which was estimated to be at least 433, 1913 and 963 for cortisol, IGF1 and testosterone, respectively, with 400 variables to get a GPR higher than 99.9%. A longitudinal follow-up of every cyclist, as for the constitution of the longitudinal ABP (Zorzoli & Rossi 2010, 2012), should bring more precision on how to consider a supposedly ‘normal’ hormonal status by contrast with ‘abnormal’ ones considering every cyclist as his own reference. Until now, no longitudinal omics follow-up has been assessed for a possible integration in ABP, even this promising approach has been claimed for more than a decade (Reichel 2011; Pitsiladis et al. 2014; Robinson et al. 2017; Wang et al. 2017).

The bootstrapped procedure used to study the variability of prediction rules built at each iteration, which are depending on the learning samples and their size, ends up in a variability in the prediction calculation of the *Normal* class unknowns. It was well represented through the two first components of a PCA, which were useful to detect at the same time some outliers of *Normal* classes in the three endocrine phenotypes outside to 97.5%-confidence limit. Thus, their existence may question the status of the quantitative variable, *i.e.* concentrations of every hormone assayed in serum, used as a unique covariate to get distribution of samples into the three classes. These outliers were well predicted in the *Normal* class only when the size of the learning dataset was sufficient; otherwise their prediction was random. This indicated that a constitutive metabolic heterogeneity might exist in the class of normal subjects that could have detrimental effect in the GPR of this group if it was not well represented in the sampled learning dataset. Such a heterogeneity was well revealed by the PHATE analysis of normal subjects in the three endocrine phenotypes, which indicated existence of some distinct subgroups (Figure

S8). PHATE gave a low-dimension embedding which preserved both global, as PCA, but also local non-linear informational structure and, therefore, could be used to get access to the intrinsic metabolic heterogeneity, present in a high-dimension dataset (Moon et al. 2019). For cortisol, normal class outliers were included inside a unique larger size subgroup among four possible. This was not the case for outliers of testosterone neither nor for IGF1 ones for which the metabolic outlier situation is questioned. This is again in favor of a requirement for larger cohorts to get more robust discriminants and to better reveal intra class heterogeneity.

5.2. Putative biomarkers of heterogeneity in *Normal* endocrine classes

5.2.1. General considerations

As underlined above, no systematic bias was detected for every *Normal* outlier subgroup that would be explained by a classification in one given false class. Besides, as revealed by a PHATE procedure, the accessible metabolomic information supported a substructure of the *Normal* class across four subsets in which only one subset contained all *Normal* outliers for the cortisol phenotype (Figure S8A), whereas two subsets were necessary to classify *Normal* outliers for testosterone one. Therefore, only one IC as given by ICDA would be *a priori* necessary to describe *Normal* cortisol outliers, when at least two IC would characterize *Normal* testosterone outliers. A large number of bucketed δ variables documenting IC were repeatedly found with high statistical scores equal to or higher than 883 and 900 over 1000 for these two *Normal* endocrine phenotypes, respectively (Table 2). Yet, for the testosterone dataset (Table 2), even some putative biomarkers could be ‘manually’ assigned from bucketed δ variables which regularly arose by bootstrap, hence characterizing robustly these outliers, none of these bucketed δ variables was clearly related to any of the metabolites of the BATMAN-integrated dataset, probably because the targeted set of metabolites did not cover a sufficiently large δ region (see details in Supplementary information, § 4.2.1).

5.2.2. Testosterone

Relative concentrations of all manually assigned candidate biomarkers explaining IC_λ with a high statistical score, *i.e.* choline, α -glycerylphosphorylcholine, L-threonine and L-arginine, were decreased in outliers of the *Normal* class of the testosterone phenotype (Table 2). α -Glycerylphosphorylcholine has been recognized recently as an emerging ergogenics (Bellar et al. 2015; Marcus et al. 2017) which could favour biosynthesis of choline (Kawamura et al. 2012), which depletion was thought to play a possible role in fatigue (Conlay et al. 1992) and which is a precursor of acetylcholine, a neurotransmitter, also involved in enhancement of muscular power and strength (Bellar et al. 2015; Zeigenfuss et al. 2008). An oral choline supplementation might only increase endurance performance (Penry & Manore 2008). Here, in testosterone *Normal* outliers, both choline, α -glycerylphosphorylcholine and probably the related acyl-derivatives of α -glycerylphosphorylcholine were decreased when compared to values of subjects from the true *Normal* class (Figure 4A). This situation could result in this outliers’ subgroup from strenuous effort training that could thereby affect endurance characteristics; indeed, as shown by Marcus et al. (2017), a daily supplementation for 7 days of α -glycerylphosphorylcholine to young men improved maximum velocity and maximum mechanical power on the counter movement jump. This same hypothesis could be also invoked to explain the relative decrease in L-threonine and L-arginine observed in serum of this subgroup of cyclists. Kingsbury et al. (1998) have shown that, when compared to a group of field athletes, plasmatic concentrations of L-threonine, a glucogenic amino acid, and L-arginine were significantly decreased in a judo sportsmen group submitted to a higher intensity training revealing that an inadequate protein intake would be a privileged causative factor.

5.2.3. Cortisol

The two significant candidate BATMAN-processed biomarkers putatively assigned to glycine and 7-methylxanthine were related to a unique metabolic trait which would explain metabolic heterogeneities detected inside the *Normal* cortisol group; their serum concentration would be significantly increased in the outliers’ subgroup (respectively, $p = 2.5 \times 10^{-3}$ and $p = 2.8 \times 10^{-2}$, Table 2). A metabolic network filtered from the Kegg *Homo sapiens* one summarizes these two metabolites correlated to some other metabolites tentatively assigned from chemical shifts that also explained construction of IC_λ with a high statistical score (Table 2), *i.e.* L-valine, glucose, citrate, L-glutamine and lipids (Figure 4B).

Diet variation would explain the existence of glycine and 7-methylxanthine as putative biomarkers. First, presence of 7-methylxanthine can sign high consumption of caffeine (1,3,7-trimethylxanthine), which is sequentially metabolized by hepatic cytochrome P450 CYP1A2, alone or with CYP2E1, to get N-1- and N-3-demethylation reactions (Carrillo & Benitez 1994, 1996; Kim et al. 2019). Caffeine can be supposedly used as pharmacological ergogenics, mostly in competition periods, to enhance performance in endurance sports by stimulating the central nervous system (Costill et al. 1978; Erickson et al. 1987; Spriet et al. 1992; Maughan et al. 2018). Thus, in elite cycling sportsmen, caffeine may improve endurance-based time trial by 3-7% (Spriet et al. 1992; Schneiker et al. 2006). But, here, samples were not specifically collected during competition periods and, as a consequence, the probability of the detection of caffeine as ergogenics would be low. More probably, a possible genetic variability of the CYP1A2 may relate this group to a specific high caffeine metabolizer phenotype (Nehlig 2018). Yet, in absence of any personal dietary recording at the blood collection time, this functional hypothesis cannot be validated.

This specific outliers' subgroup may use also glycine in designed nutritional supplements as it can be supplied for numerous endurance athletes (Buford & Koch 2004; Jacobs et al. 2009; Wax et al. 2013). In human, the normal glycine concentration in plasma results from dietary intake which varies from 1.5 to 3 g/day depending on protein intake, *de novo* synthesis from serine, sarcosine (N-methylglycine), and glyoxalate on one side and the glycine biotransformation into serine, the glycine oxidation and urea synthesis, but also some minor conversions of glycine into glyoxalate, glycoamine or sarcosine, and last from conjugation to benzoate and hydroxybenzoate on the other side, as reviewed recently (Alves et al. 2019). A high intake of glycine and more generally a higher bioavailability of glycine may lower insulin resistance index (Gao et al. 2017). Here, as already anticipated in the general population regarding effect of glycine and as reinforced by a sustained physical exercise as it was shown in Type 2 Diabetes Mellitus patients (Umpierre et al. 2011; Colberg et al. 2016; Zheng et al. 2020), this targeted subgroup of cyclists would probably display some very specific metabolic properties regarding a very efficient energy regulation and a very low insulin resistance index as compared to the remaining part of the *Normal* group of the cortisol trait. Of note, the health benefit of glycine is also accompanied by an improvement in the biosynthesis of glutathione and, hence, by an increased protection towards pro-oxidative compounds (Díaz-Flores et al. 2013; McCarty et al. 2018).

Last, glutamine, another putative biomarker of the outliers' subgroup in the *Normal* class, is a key anaplerotic precursor of the Krebs cycle and a non-essential glucogenic amino acid in normal energy expenditure conditions. But it is considered as conditionally essential in case of metabolic stress (Earnest & Rasmussen 2015). This explains its large use as nutritional supplement with its supposed anti-proteolytic effect on skeletal muscle as well as anti-catabolic effect on branch-chain amino acids (BCAAs) (Raizel et al. 2016). A too low glutamine availability may define a specific outlier metabolic picture in case of elite sportsmen at overreaching physiological risk (Rowbottom et al. 1996) as it could be the case for this outlier *Normal* cortisol subgroup which displayed glutamine concentrations in serum significantly lower than those found in the true *Normal* group. In connection with this atypical metabolic picture displayed by this outlier subgroup, glucose and citrate concentrations are also lower than in the true normal one (Figure 4B), which is coherent with an apparent reduced energy availability in this subgroup, also in line with the global decrease in ¹H NMR lipids markers. Again, at this step, no available anthropometric, dietary or clinical metadata could confirm these putative biomarkers of such a heterogeneity found inside the *Normal* cortisol class.

6. Conclusion

The present pilot metabolomic epidemiological study of the disrupted hormonal status in sportsmen, regarding circulating cortisol, IGF1 and testosterone, thanks to the intensive data modelling approach based on a combined OSC of raw ¹H NMR signals and a SDA used to discriminate *Normal*, *High* and *Low* classes of every endocrine phenotype, was useful to get some critical parameters to size a fully informative cohort. Therefore, the number of statistical individuals should be above 929, 2346 and 1408 for cortisol, IGF1 and testosterone traits, respectively, to get a global discrimination error risk lower than 0.1% when considering a dataset built from the 200 most informative variables only, this error risk being in total agreement with the credibility value of 99.9% considered by WADA for the Athletics Biological Passport.

In addition, such an intensive statistical modeling gives access to a fine characterization of a sub heterogeneity in the true *Normal* classes. For cortisol and testosterone phenotypes, few statistical individuals may be considered

as apparent outliers. Diet is probably the main causative factor. Functional hypotheses regarding i) the inadequate protein intake when considering a possible strenuous exercise to which these outliers could be submitted, and ii) intake of ergogenic nutritional supplements, as glycine or glutamine, might explain heterogeneous metabolic profiles in a supposedly *Normal* cyclist class.

In the present study, neither correlative dietary nor clinical nor anthropometric metadata was recorded at the blood collection time to get any further confirmation of these hypotheses. Beside the cohort size, this represents a clear limitation. To fully validate this present pioneering metabolomic approach in sport, future follow-up of cycling sportsmen should record these metadata. In conjunction with ^1H NMR fingerprinting used here, a better access to some very specific markers of diet or physical overload would be brought by LC-MS platforms.

Nevertheless, the present results are in favor of the systematic use of metabolomic scores to complete the ABP-supported assessment of doping. An extended validation of coordinated metabolic deviations supporting major endocrine changes between *Normal* classes and either *Low* or *High* ones is planned in future to get a specific statistical modelling of negative hormonal feedbacks.

Tables

Table 1. Fitting the lower limit value of Q5 values of GPR for the different endocrine phenotypes considering either 200 or 400 variables to predict the minimal sample size to get GPR higher than 99.9%. The Q5 curves were built thanks to a bootstrapped approach with LS varying between 50 and 600 statistical individuals as shown in Figure S7. Two types of models were used to fit the sigmoidal curves, either directly with a family of different sigmoid equations (models 1 to 4), or indirectly (model 5) with a multiple regression between $\logit(y)$, y corresponding to the Q5 GPR values, and polynomials of x , which is the size of the cohort, using degrees ≤ 7 . Rows in light blue: maximal reference values; rows in light green: defective fitting models with predicted maximal GPR lower than 99.9%; rows in white: selected fitting models.

Hormones	OSC ¹	Sigmoidal models	200 variables					400 variables				
			Model number	Degree of polynomials	GPR Q5 value	Predicted cohort size ²	Residual variance	Model number	Degree of polynomials	GPR Q5 value	Predicted cohort size	Residual variance
cortisol	1	no			97.4	500				100.0	500	
		nls ³	(3)		97.8	508	0.980	(3)		99.9	433	1.105
		polynomials	(5)	1	99.9	929	7.320	(5)	1	99.9	503	7.368
testosterone	1	no			74.6	600				80.6	600	
		nls	(1)		75.2	1000	0.870	(2)		84.2	746	1.622
		polynomials	(5)	1	99.9	1408	2.485	(5)	1	99.9	1342	2.385
									3	99.9	963	2.372
IGF1	1	no			41.0	600				46.8	600	
		nls	(1)		44.7	2000	0.382	(1)		58.4	2000	0.471
		polynomials	(5)	1	99.9	2346	0.626	(5)	1	99.9	1913	0.642
	4	no			98.7	600				98.8	600	
		nls	(4)		99.9	643	1.209	(4)		99.9	603	1.072
		polynomials	(5)	1	99.9	908	7.129	(5)	1	99.9	836	6.356
3	99.9			750	6.988	3	99.9		820	6.319		

¹ Number of orthogonal signal components calculated from raw ¹H NMR data prior fitting the GPR Q5 curves.

² Size of population estimated for a Q5 value of GPR fixed at 99.9%.

³ Non-linear least squares fitting of Q5 values of GPR.

Table 2. Putative metabolites correlated to the IC describing the common traits explaining heterogeneities inside every normal endocrine group which were detected for the validation procedure by a bootstrapped-based classification.

Endocrine traits (# outliers vs # normal samples)	Bootstrap results (n = 1000)									
	Chemical shifts-based datasets			BATMAN-processing datasets						
	Variance explained by IC_{λ} [Q5-Q95] (%)	δ variables (ppm), variation and putative assignment between brackets	Statistical scores of δ variables (%)	Metabolite variables significantly correlated to IC_{λ}	Quantiles [Q5-Q95]			Statistical scores of metabolite variables (%)	p-values for imbalanced t-test	
				correlations to IC_{λ}	$F_{1,p}$	p-values				
cortisol (8 vs 252)	5.92 – 21.00	<ul style="list-style-type: none"> • 2.285*** (↓), 2.275*** (↓), 2.265*** (↓) [L-valine] • 3.435*** (↓) [glucose] • 2.745*** (↓), 2.725*** (↓) [lipids] • 2.695*** (↓), 2.715*** (↓) [citrate] • 2.435* (↓), 2.445* (↓) [L-glutamine] 	<ul style="list-style-type: none"> • 945, 953, 931 • 902 • 900, 937 • 954, 901 • 929, 940 	7-methylxanthine* (↑)	0.25 – 0.88	14.81 – 59.20	$1.0 \cdot 10^{-4} - 2.3 \cdot 10^{-2}$	984	$2.8 \cdot 10^{-2}$	
				glycine** (↑)	0.23 – 0.77	7.74 – 33.44	$1.3 \cdot 10^{-3} - 9.3 \cdot 10^{-2}$	902	$2.5 \cdot 10^{-3}$	
IGF1 (9 vs 413)	4.63 – 16.50	<ul style="list-style-type: none"> • 1.505, 1.495, 1.515, 1.155 	<ul style="list-style-type: none"> • 455, 388, 359, 224 	No significant metabolite correlated to IC_{λ}						
				<i>Direct correlation to IC_{λ} from BATMAN-supported metabolites dataset</i>	7-methylxanthine	0.09 – 0.78	1.99 – 11.40	$1.7 \cdot 10^{-1} - 9.2 \cdot 10^{-1}$	66	nd ¹
					L-lactic acid	0.07 – 0.75	0.83 – 10.64	$2.1 \cdot 10^{-1} - 9.6 \cdot 10^{-1}$	46	nd
testosterone (9 vs 368)	6.48 – 16.22	<ul style="list-style-type: none"> • 4.075*** (↓), 4.065*** (↓) [choline] • 4.205*** (↓), 4.215*** (↓) [?] • 4.255*** (↓), 4.265*** (↓), 4.285*** (↓), 4.275*** (↓) [L-threonine] • 4.295*** (↓), 4.305*** (↓) [?] • 4.325*** (↓), 4.315*** (↓), 4.335*** (↓), 4.345*** (↓) [α-glycerylphosphorylcholine] • 1.615*** (↓), 1.625*** (↓), 1.635*** (↓), 1.645*** (↓), 1.655*** (↓), 1.665*** (↓) [L-arginine] 	<ul style="list-style-type: none"> • 959, 974 • 909, 919 • 988, 975, 975, 963 • 978, 978 • 917, 976, ?, 923 • ?, 883, 900, 910, 910, ? 	methyl-isobutyl ketone	0.035 – 0.543	0.13 – 9.01	$3.6 \cdot 10^{-1} - 9.8 \cdot 10^{-1}$	17	nd	
				L-threonine	0.015 – 0.482	0.01 – 6.98	$5.2 \cdot 10^{-1} - 9.9 \cdot 10^{-1}$	14	nd	
				deoxycholic acid	0.077 – 0.721	0.04 – 7.70	$4.4 \cdot 10^{-1} - 9.9 \cdot 10^{-1}$	11	nd	

¹ nd: not done.

* p < 0.05, ** p < 0.01, *** p < 0.001

Figure captions

Figure 1. Flow-chart of the different statistical analyses sequentially used to mine the ^1H NMR dataset and data subsets used to predict the endocrine phenotypes of cyclist sportsmen. It was divided in four blocks corresponding to parts of results indicated between brackets as following: [3.2.1 & 4.1], [4.2], [4.3 & 4.4] and [4.7]. Objectives of these specific parts are indicated in conclusive items colored in light violet. Specific statistical tools are indicated in bold black on green background.

Figure 2. Abacuses describing the variation of the median GPR and the confidence interval at 90% according to LS (50-650) and NV, the number of prior selected informative variables (50, 100, 150, 200 and 400) thanks to the cat scores. The confidence interval limits (in dotted lines) were colored as their respective median curve. Only one orthogonal component was used to correct the dataset for cortisol (A), IGF1 (B) and testosterone (C). Four orthogonal components were used to significantly increase the GPR performances for IGF1 (D).

Figure 3. Score plots and frequency distributions along the two first PCs for cortisol (A), IGF1 (B) and testosterone (C) with one orthogonal component (1 OSC), and IGF1 with 4 orthogonal components (4 OSC) (D). The 3 classes are given in green, violet and red for classes *Normal*, *Low* and *High* of every endocrine phenotype. Light grey squares marked with a green line focus on the apparent outliers of the *Normal* class excluded from the ellipses drawn at a confidence limit above 97.5%.

Figure 4. Metabolic modules containing the putative biomarkers characterizing the outliers of the testosterone (A) and cortisol (B) *Normal* classes and connected metabolites as summarized in the respective sub-networks defined from a significant enrichment according to a Benjamini-Hochberg correction of the prior selected biomarkers using the MetExplore web tool. 1-Methylhistidine (meat consumption) also detected as marker for cortisol *Normal* class was not enclosed in the prior selection and enrichment analysis of targeted metabolic biomarkers.

Supplementary figure captions

Figure S1. Distribution of cortisol (A), IGF1 (B) and testosterone (C) concentrations of the cyclist cohort as a function of age. A distinction was done between the groups they belonged to, *i.e.* the hormonally normal group in green, the hormonally higher than normal group (group *High*) in red and the hormonally lower than normal group (group *Low*) in violet. For every subgroup inside a hormonal trait, a linear regression between serum concentration and age [20-34 yrs.] was performed. Although a very significant variation of IGF1 concentration with age was observed only in the normal IGF1 group studied here, the slope value was not sufficient to conclude that age would be considered as a possible confounding factor when older cycling athletes with very low IGF1 concentration in serum would be still considered as normal ones.

Figure S2. Example of a ^1H NMR metabolic fingerprint obtained at 600.13 MHz on a serum collected on an anonymous subject. Some assignments to known chemical shifts were done from those usually recorded for analytes detected in serum or plasma. Glu, Gln and Val corresponds to L-glutamate, L-glutamine and L-valine, respectively.

Figure S3. Score plot of the different statistical individuals from a PCA analysis of the raw dataset comprising bucketed ^1H NMR fingerprints ($p = 419$ variables) acquired on two different NMR sessions (1 and 2). Three types of factors were encoded differently according to the color of plot symbols for cycling discipline, their shape for season where blood samples were collected, and their size for the NMR session of samples fingerprinting.

Figure S4. Score plot of the different statistical individuals from a PHATE analysis of raw dataset comprising bucketed ^1H NMR fingerprints ($p = 419$ variables) acquired on two different NMR sessions (1 and 2) appearing in filled circles and triangles, respectively. PHATE analysis was performed on a data subset ($n = 633$) for which all samples were completely phenotyped for the three endocrine traits with no non-phenotyped (NP) data. Circled numbers give the number of samples collected in a longitudinal follow-up for every subject and which are projected on a barycenter corresponding to the mean score point calculated from these different points acquired respectively on the 1st (red circles), 2nd (green circles) or both NMR sessions (blue circles). Most of mean points with a number of samplings per cyclist above one are projected in the center part of the factorial map. Lines connect respective barycenters to the different points corresponding to samples collected for every cyclist and are colored according to the different subjects. Using MANOVA between the two coordinates of the factorial map on one part and, on another part, the set of following factors, *i.e.* the '*NMR session*' factor, the number of samples in a longitudinal series, the Euclidean distance calculated from barycenters to samples inside a longitudinal series, all these factors being joined to the dummy matrix built from subjects phenotyped more than one time, shows in addition to a highly significant '*NMR session*' effect and a '*barycenter-to-sample distance*' effect, an effect of some rare subjects whose barycenters are projected close to the extreme parts of the 2nd component of the factorial map and are indicated by the following labels: ● when $p < 0.10$, * when $p < 0.05$, ** when $p < 0.01$ and *** when $p < 0.05$. The 21 specific endocrine phenotypes appear as differently colored from dark blue to light yellow for disrupted samples and in grey for controls (or equivalent to the combined normal endocrine phenotype '*Cortisol N / IGF1 N / Testo N*'). Insert gives only the PHATE scores of the different samples on the plan 1×2 with circles corresponding to samples analyzed on the 1st NMR session and signs + for samples analyzed on the 2nd one.

Figure S5. Progressive building of a generic classification algorithm. Partial least squares-discriminant analysis (PLS-DA) without (A) or with a prior PLS-based orthogonal signal correction with one orthogonal component (1 OSC) (B) considering the cortisol grouping factor. Shrinkage discriminant analysis (SDA) without (C) or with a prior OSC with 1 component (D). Mahalanobis distance plots used to calculate probabilities to predict belonging of one individual to a given phenotypic subgroup by SDA without (E) or with a prior OSC with 1 component (F). Legend: the different phenotypic subgroups are given with normal cortisol concentrations in serum in green, lower cortisol concentrations than normal in violet, and higher cortisol concentrations than normal in red.

Figure S6. Density plots of the distribution log₁₀-transformed F-values coming from a MANOVA test (Hotelling-Lawley criterion) applied to SDA components repeatedly calculated in a bootstrap procedure using at each iteration a permutation test (line curves in red). True values for cortisol (A), IGF1 (B) and testosterone (C) datasets, all corrected by OSC with one orthogonal component, and for IGF1 dataset prior submitted to 4 OSC components (D), were indicated by a narrow distribution obtained by bootstrap ($n = 1000$, line curves in black). All these specific distributions were significantly different from those resulting from random permutations ($p < 2 \cdot 10^{-16}$).

Figure S7. Variation of the prediction rates of classification of individuals in one of the different phenotypic subgroups for either cortisol (A), IGF1 (B and D) or testosterone (C) assayed in serum according to the size of the training set, which varied from 50 to 650. One PLS-based correction component (1 OSC) was used to get a prior correction of the dataset for the different hormonal traits. Influence of a correction with 3 supplementary orthogonal components (4 OSC) was also tested on the prediction rates of the IGF1 status. For every hormonal trait, prediction rates were presented for the different subgroups. A global prediction rate (GPR) corresponding to the multiplication of the three specific ones for a given hormonal trait was done at every training set size. Thanks to the resampling procedure used ($n = 1000$), a confidence interval (in light grey) was calculated in the 5%-95% quantile range (dotted lines).

Figure S8. PHATE analysis of the distribution of individuals belonging to the class *Normal* of the following endocrine phenotypes: cortisol (A), IGF1 (B) and testosterone (C) through 4 groups. This number of groups in the *Normal* class corresponded to the maximal one for which all outliers of the cortisol *Normal* class were assigned to a unique subgroup. The same number of subgroups was used to display the metabolomic heterogeneity inside IGF1 and

testosterone *Normal* classes and then to check the distribution of outliers inside the respective different subgroups. Outliers prior detected by a PCA analysis of the bootstrap supported cross-validation data are marked with larger size points (circle or triangle). Only for the normal cortisol trait, outliers belonged to only one group.

Figure S9. Clustering of bucketed δ variables involved in the building of the most important independent component discriminating outliers from true normal individuals for the testosterone endocrine phenotype. All selected variables displayed a score above 700. A: Heatmap focused on selected δ variables, B: spectrum plotting of selected δ variables, and C: expert-based putative assignment of chemical shifts.

Figure S10. Clustering of bucketed δ variables involved in the building of the most important independent component discriminating outliers from true normal individuals for the cortisol endocrine phenotype. All selected variables displayed a score above 899. A: Heatmap focused on selected δ variables, B: spectrum plotting of selected δ variables, and C: expert-based putative assignment of chemical shifts.

References

- Ahdesmäki, M., & Strimmer, K. (2010). Feature selection in omics prediction problems using cat scores and false non-discovery rate control. *Annals of Applied Statistics* 4:503-519. doi: 10.1214/09-AOAS277
- Alladio, E., Caruso, R., Gerace, E., Amante, E., Salomone, A., & Vincenti, M. (2016). Application of multivariate statistics to the Steroidal Module of the Athlete Biological Passport: a proof of concept study. *Analytica Chimica Acta* 922:19-29. doi: 10.1016/j.aca.2016.03.051.
- Alves, A., Bassot, A., Bulteau, A.L., Pirola, L., & Morio, B. (2019). Glycine metabolism and its alterations in obesity and metabolic diseases. *Nutrients* 11(6):1356. doi: 10.3390/nu11061356.
- Amante, E., Pruner, S., Alladio, E., Salomone, A., Vincenti, M., & Bro, R. (2019). Multivariate interpretation of the urinary steroid profile and training-induced modifications. The case study of a Marathon runner. *Drug Testing and Analysis* 11:1556-1565. doi: 10.1002/dta.2676.
- Ayotte C., Miller J., Thevis M. (2017). Challenges in Modern Anti-Doping Analytical Science. *Medicine & Sport Science* 62:68–76. doi: 10.1159/000460701.
- Barker, M.L., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics* 17:166-173. doi: 10.1002/cem.785
- Bellar, D., LeBlanc, N.R., & Campbell, B. (2015). The effect of 6 days of alpha glycerylphosphorylcholine on isometric strength. *Journal of International Society Sports Nutrition* 12:42. doi: 10.1186/s12970-015-0103-x.
- Broersen, L.H., Pereira, A.M., Jørgensen, J.O., & Dekkers, O.M. (2015). Adrenal insufficiency in corticosteroids use: systematic review and meta-analysis. *Journal of Clinical Endocrinology and Metabolism* 100:2171-2180. doi: 10.1210/jc.2015-1218.
- Buford, B.N., & Koch, A.J. (2004). Glycine-arginine-alpha-ketoglutaric acid improves performance of repeated cycling sprints. *Medicine & Science in Sports & Exercise* 36:583-587. doi:10.1249/01.
- Cadegiani, F.A., Kater, C.E., & Gazola, M. (2019). Clinical and biochemical characteristics of high-intensity functional training (HIFT) and overtraining syndrome: findings from the EROS study (The EROS-HIFT). *Journal of Sports Sciences* 37(11):1296-1307. doi: 10.1080/02640414.2018.1555912.
- Carrillo, J.A., & Benítez, J. (1994). Caffeine metabolism in a healthy Spanish population: N-acetylator phenotype and oxidation pathways. *Clinical Pharmacology & Therapeutics* 55(3):293-304. doi: 10.1038/clpt.1994.30.
- Carrillo, J.A., & Benitez, J. (1996). CYP1A2 activity, gender and smoking, as variables influencing the toxicity of caffeine. *British Journal of Clinical Pharmacology* 41(6):605-8. doi: 10.1046/j.1365-2125.1996.35418.x.
- Colberg, S.R., Sigal, R.J., Yardley, J.E., Riddell, M.C., Dunstan, D.W., Dempsey, P.C., et al. (2016). Physical activity/exercise and diabetes: a position statement of the American diabetes association. *Diabetes Care* 39:2065-2079. doi: 10.2337/dc16-1728.
- Conlay, L.A., Sabounjian, L.A., & Wurtman, R.J. (1992). Exercise and neuromodulators: choline and acetylcholine in marathon runners. *International Journal of Sports Medicine* 13 Suppl 1:S141-142. doi: 10.1055/s-2007-1024619.
- Costill, D.L., Dalsky, G.P., & Fink, W.J. (1978). Effects of caffeine ingestion on metabolism and exercise performance. *Medicine and Science in Sports* 10(3):155-158.
- Cottret, L., Frainay, C., Chazalviel, M., Cabanettes, F., Gloaguen, Y., Camenen, E., et al. (2018). MetExplore: collaborative edition and exploration of metabolic networks. *Nucleic Acids Research* 46(W1):W495-W502. doi: 10.1093/nar/gky301.
- de Hon, O., & van Bottenburg, M. (2017). True dopers or negligent athletes? An analysis of anti-doping rule violations reported to the world anti-doping agency 2010-2012. *Substance Use & Misuse* 52(14):1932-1936. doi: 10.1080/10826084.2017.1322105.
- Díaz-Flores, M., Cruz, M., Duran-Reyes, G., Munguia-Miranda, C., Loza-Rodríguez, H., Pulido-Casas, E., et al. (2013). Oral supplementation with glycine reduces oxidative stress in patients with metabolic syndrome, improving their systolic blood pressure. *Canadian Journal of Physiology and Pharmacology* 91(10):855-860. doi: 10.1139/cjpp-2012-0341.

- Duclos, M., Guinot, M., Colsy, M., Merle, F., Baudot, C., Corcuff, J. B., & Lebouc, Y. (2007). High risk of adrenal insufficiency after a single articular steroid injection in athletes. *Medicine & Science in Sports & Exercise* 39(7), 1036–1043. doi: 10.1249/mss.0b013e31805468d6
- Dumas, M.E., Canlet, C., André, F., Vercauteren, J., & Paris, A. (2002a). Metabonomic assessment of physiological disruptions using ^1H - ^{13}C HMBC-NMR spectroscopy combined with pattern recognition procedures performed on filtered variables. *Analytical Chemistry* 74:2261-2273. doi: 10.1021/ac0156870.
- Dumas, M.E., Debrauwer, L., Beyet, L., Lesage, D., André, F., Paris, A., & Tabet, J.C. (2002b). Analyzing the physiological signature of anabolic steroids in cattle urine using pyrolysis/metastable atom bombardment mass spectrometry and pattern recognition. *Analytical Chemistry* 74:5393-5404. doi: 10.1021/ac025656k.
- Dumas, M.E., Canlet, C., Vercauteren, J., André, F., & Paris, A. (2005). Homeostatic signature of anabolic steroids in cattle using ^1H - ^{13}C HMBC NMR metabonomics. *Journal of Proteome Research* 4:1493-1502. doi: 10.1021/pr0500556.
- Earnest, C.P., & Rasmussen, C. (2015). Nutritional supplements for endurance athletes. In M. Greenwood, M.B. Cooke, T. Ziegenfuss, D.S. Kalman, & J. Antonio (Eds.), *Nutritional Supplements in Sports and Exercise* (pp. 253–272). Springer International Publishing, Switzerland. doi: 10.1007/978-3-319-18230-8_11.
- Erickson, M.A., Schwarzkopf, R.J., & McKenzie, R.D. (1987). Effects of caffeine, fructose, and glucose ingestion on muscle glycogen utilization during exercise. *Medicine & Science in Sports & Exercise* 19(6):579-583.
- Fabregat, A., Marcos, J., Segura, J., Ventura, R., Pozo, O.J. (2016). Factors affecting urinary excretion of testosterone metabolites conjugated with cysteine. *Drug Testing Analysis* 8(1):110-119. doi: 10.1002/dta.1801.
- Gao, X., Wang, Y., & Sun, G. (2017). High dietary choline and betaine intake is associated with low insulin resistance in the Newfoundland population. *Nutrition* 33:28-34. doi: 10.1016/j.nut.2016.08.005.
- Guinot, M., Duclos, M., Idres, N., Souberbielle, J.-C., Megret, A., Le Bouc, Y. (2007). Value of basal serum cortisol to detect corticosteroid-induced adrenal insufficiency in elite cyclists. *European Journal of Applied Physiology* 99:205-216. doi: 10.1007/s00421-006-0332-4.
- Habchi, B., Alves, S., Jouan-Rimbaud Bouveresse, D., Moslah, B., Paris, A., Lécluse, Y., Gauduchon, P., Lebailly, P., Rutledge, D.N., & Rathahao-Paris, E. (2017). An innovative chemometric method for processing direct introduction high resolution mass spectrometry metabolomic data: independent component–discriminant analysis (IC–DA). *Metabolomics* 13, 45. doi: 10.1007/s11306-017-1179-x
- Hao, J., Liebeke, M., Astle, W., De Lorio, M., Bundy, J.G., & Ebbels, T.M. (2014). Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. *Nature Protocol* 9:1416-1427. doi: 10.1038/nprot.2014.090.
- Holt, R.I., Erotokritou-Mulligan, I., McHugh, C., Bassett, E.E., Bartlett, C., Fityan, A., Bacon, J.L., Cowan, D.A., & Sönksen, P.H. (2010). The GH-2004 project: the response of IGF1 and type III pro-collagen to the administration of exogenous GH in non-Caucasian amateur athletes. *European Journal of Endocrinology* 163(1):45-54. doi: 10.1530/EJE-09-0978.
- Holt, R.I., & Sönksen, P.H. (2008). Growth hormone, IGF-I and insulin and their abuse in sport. *British Journal of Pharmacology* 154(3):542-556. doi: 10.1038/bjp.2008.99.
- Howe, C.C.F., Alshehri, A., Muggeridge, D., Mullen, A.B., Boyd, M., Spendiff, O., Moir, H.J., & Watson, D.G. (2018). Untargeted metabolomics profiling of an 80.5 km simulated treadmill ultramarathon. *Metabolites* 8:14. doi: 10.3390/metabo8010014.
- Iljukov, S., & Schumacher, Y.O. (2017). Performance profiling-perspectives for anti-doping and beyond. *Frontiers in Physiology* 8:1102. doi: 10.3389/fphys.2017.01102.
- Jeanneret, F., Tonoli, D., Rossier, M.F., Saugy, M., Boccard, J., & Rudaz, S. (2016). Evaluation of steroidomics by liquid chromatography hyphenated to mass spectrometry as a powerful analytical strategy for measuring human steroid perturbations. *Journal of Chromatography A* 1430:97-112. doi: 10.1016/j.chroma.2015.07.008.
- Jacobs, P.L., Goldstein, E.R., Blackburn, W., Orem, I., Hughes, J.J. (2009). Glycine propionyl-L-carnitine produces enhanced anaerobic work capacity with reduced lactate accumulation in resistance trained males. *Journal International Society Sports Nutrition* 6:9.

- Kawamura, T., Okubo, T., Sato, K., Fujita, S., Goto, K., Hamaoka, T., & Iemitsu, M. (2012). Glycerophosphocholine enhances growth hormone secretion and fat oxidation in young adults. *Nutrition* 28(11-12):1122-1126. doi: 10.1016/j.nut.2012.02.011.
- Kim, H.J., Choi, M.S., Rehman, S.U., Ji, Y.S., Yu, J.S., Nakamura, K., & Yoo, H.H. (2019). Determination of urinary caffeine metabolites as biomarkers for drug metabolic enzyme activities. *Nutrients* 11(8):1947. doi: 10.3390/nu11081947.
- Kingsbury, K.J., Kay, L., & Hjelm, M. (1998). Contrasting plasma free amino acid patterns in elite athletes: association with fatigue and infection. *British Journal of Sports Medicine* 32(1):25-32. doi: 10.1136/bjism.32.1.25.
- Lawler, N.G., Abbiss, C.R., Gummer, J.P.A., Broadhurst, D.I., Govus, A.D., Fairchild, T.J., Thompson, K.G., Garvican-Lewis L.A., Gore, C.J., Maker, G.L., Trengove, R.D., & Peiffer, J.J. (2019). Characterizing the plasma metabolome during 14 days of live-high, train-low simulated altitude: A metabolomic approach. *Experimental Physiology* 104:81-92. doi: 10.1113/EP087159.
- Lehtihet, M., Bhuiyan, H., Dalby, A., Ericsson, M., & Ekström, L. (2019). Longitudinally monitoring of P-III-NP, IGF-I, and GH-2000 score increases the probability of detecting two weeks' administration of low-dose recombinant growth hormone compared to GH-2000 decision limit and GH isoform test and micro RNA markers. *Drug Testing and Analysis* 11:411-421. doi: 10.1002/dta.2506.
- Marcus, L., Soileau, J., Judge, L.W., & Bellar, D. (2017). Evaluation of the effects of two doses of alpha glycerylphosphorylcholine on physical and psychomotor performance. *Journal of International Society Sports Nutrition* 14:39. doi: 10.1186/s12970-017-0196-5.
- Maughan, R.J., Burke, L.M., Dvorak, J., Larson-Meyer, D.E., Peeling, P., Phillips, S.M., Rawson, E.S., Walsh, N.P., Garthe, I., Geyer, H., Meeusen, R., van Loon, L.J.C., Shirreffs, S.M., Spriet, L.L., Stuart, M., Vernec, A., Currell, K., Ali, V.M., Budgett, R.G., Ljungqvist, A., Mountjoy, M., Pitsiladis, Y.P., Soligard, T., Erdener, U., & Engebretsen, L. (2018). IOC consensus statement: dietary supplements and the high-performance athlete. *British Journal of Sports Medicine* 52(7):439-455. doi: 10.1136/bjsports-2018-099027.
- McCarty, M.F., O'Keefe, J.H., & DiNicolantonio, J.J. (2018). Dietary glycine is rate-limiting for glutathione synthesis and may have broad potential for health protection. *Ochsner Journal* 18:81-87.
- Montagna, S., & Hopker, J. (2018). A Bayesian approach for the use of athlete performance data within anti-doping. *Frontiers in Physiology* 9:884. doi: 10.3389/fphys.2018.00884.
- Moon, K.R., van Dijk, D., Wang, Z., Gigante, S., Burkhardt, D.B., Chen, W.S., Yim, K., Elzen, A.V.D., Hirn, M.J., Coifman, R.R., Ivanova, N.B., Wolf, G., & Krishnaswamy, S. (2019). Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology* 37(12):1482-1492. doi: 10.1038/s41587-019-0336-3. Erratum in: *Nature Biotechnology* 38(1):108 (2020).
- Nehlig, A. (2018). Interindividual differences in caffeine metabolism and factors driving caffeine consumption. *Pharmacological Reviews* 70:384-411. doi: 10.1124/pr.117.014407.
- Penry, J.T., & Manore, M.M. (2008). Choline: an important micronutrient for maximal endurance-exercise performance? *International Journal of Sport Nutrition and Exercise Metabolism* 18:191-203. doi: 10.1123/ijsnem.18.2.191
- Pitsiladis, Y.P., Durussel, J., & Rabin, O. (2014). An integrative 'omics' solution to the detection of recombinant human erythropoietin and blood doping. *British Journal of Sports Medicine* 48:856-861. doi: 10.1136/bjsports-2014-093529.
- Ponzetto, F., Boccard, J., Nicoli, R., Kuuranne, T., Saugy, M., & Rudaz, S. (2018). UHPLC-HRMS Analysis for Steroid Profiling in Serum (Steroidomics). *Methods in Molecular Biology (Clifton, N.J.)* 1738:261-278. doi: 10.1007/978-1-4939-7643-0_18.
- Ponzetto, F., Boccard, J., Nicoli, R., Kuuranne, T., Saugy, M., & Rudaz, S. (2019). Steroidomics for highlighting novel serum biomarkers of testosterone doping. *Bioanalysis* 11(12):1171-1187. doi: 10.4155/bio-2019-0079.
- Pozo, O.J., Marcos, J., Ventura, R., Fabregat, A., Segura, J. (2010). Testosterone metabolism revisited: discovery of new metabolites. *Analytical Bioanalytical Chemistry* 398(4):1759-1770. doi: 10.1007/s00216-010-4082-0.

- R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Raizel, R., Leite, J.S.M., Hypólito, T.M., Coqueiro, A.Y., Newsholme, P., Cruzat, V.F., & Tirapegui, J. (2016). Determination of the anti-inflammatory and cytoprotective effects of l-glutamine and l-alanine, or dipeptide, supplementation in rats submitted to resistance exercise. *British Journal of Nutrition* 116, 470–479.
- Reichel, C. (2011). OMICS-strategies and methods in the fight against doping. *Forensic Science International* 213:20-34. doi: 10.1016/j.forsciint.2011.07.031.
- Robinson, N., Sottas, P.E., Mangin, P., & Saugy, M. (2007). Bayesian detection of abnormal hematological values to introduce a no-start rule for heterogeneous populations of athletes. *Haematologica* 92(8):1143-1144. doi: 10.3324/haematol.11182.
- Robinson, N., Sottas, P.E., & Schumacher, Y.O. (2017). The athlete biological passport: how to personalize anti-doping testing across an athlete's career? *Medicine and Sport Science* 62:107-118. doi: 10.1159/000460722.
- Rohart, F., Gautier, B., Singh, A., & Lê Cao, K.A. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Computational Biology* 13(11):e1005752. doi: 10.1371/journal.pcbi.1005752
- Rohart, F., Paris, A., Laurent, B., Canlet, C., Molina, J., Mercat, M.J., Tribout, T., Muller, N., Iannuccelli, N., Villa-Vialaneix, N., Liaubet, L., Milan, D., & San Cristobal, M. (2012). Phenotypic prediction based on metabolomic data for growing pigs from three main European breeds. *Journal of Animal Science* 90:4729-4740. doi: 10.2527/jas.2012-5338.
- Rowbottom, D.G., Keast, D., & Morton, A.R. (1996). The emerging role of glutamine as an indicator of exercise stress and overtraining. *Sports Medicine* 21(2):80-97. doi: 10.2165/00007256-199621020-00002.
- Samworth, R.J. (2012) Stein's paradox. *Eureka* 62:38–41.
- Schneiker, K.T., Bishop, D., Dawson, B., & Hackett, L.P. (2006). Effects of caffeine on prolonged intermittent-sprint ability in team-sport athletes. *Medicine & Science in Sports & Exercise* 38(3):578-585. doi: 10.1249/01.mss.0000188449.18968.62.
- Schumacher, Y.O., & Pottgiesser, T. (2009). Performance profiling: a role for sport science in the fight against doping? *International Journal Sports Physiology & Performance* 4(1):129-133. doi: 10.1123/ijssp.4.1.129.
- Shi, R., Zhang, J., Fang, B., Tian, X., Feng, Y., Cheng, Z., Fu, Z., Zhang, J., & Wu, J. (2020). Runners' metabolomic changes following marathon. *Nutrition & Metabolism* 17:19. doi: 10.1186/s12986-020-00436-0. Erratum in: *Nutrition & Metabolism* 17:53 (2020).
- Sieckmann, T., Elmongy, H., Ericsson, M., Bhuiyan, H., Lehtihet, M., & Ekström, L. (2020). Longitudinal studies of putative growth hormone (GH) biomarkers and hematological and steroidal parameters in relation to 2 weeks administration of human recombinant GH. *Drug Testing and Analysis* 12(6), 711–719. doi: 10.1002/dta.2787.
- Sottas, P.E., Baume, N., Saudan, C., Schweizer, C., Kamber, M., & Saugy, M. (2007). Bayesian detection of abnormal values in longitudinal biomarkers with an application to T/E ratio. *Biostatistics* 8(2):285-296. doi: 10.1093/biostatistics/kxl009.
- Sottas, P.E., Robinson, N., Giraud, S., Taroni, F., Kamber, M., Mangin, P., & Saugy, M. (2006). Statistical classification of abnormal blood profiles in athletes. *The International Journal of Biostatistics* 2:1-21. doi: 10.2202/1557-4679.1011.
- Sottas, P.E., Robinson, N., Rabin, O., & Saugy, M. (2011). The athlete biological passport. *Clinical Chemistry* 57(7):969-976. doi: 10.1373/clinchem.2011.162271.
- Sottas, P.E., Robinson, N., & Saugy M. (2010a). The athlete's biological passport and indirect markers of blood doping. *Handbook Experimental Pharmacology* 195:305-326. doi: 10.1007/978-3-540-79088-4_14.
- Sottas, P.E., Saudan, C., Schweizer, C., Baume, N., Mangin, P., & Saugy, M. (2008). From population- to subject-based limits of T/E ratio to detect testosterone abuse in elite sports. *Forensic Science International* 174(2-3):166–172. doi: 10.1016/j.forsciint.2007.04.001.
- Sottas, P.E., Saugy, M., & Saudan, C. (2010b). Endogenous steroid profiling in the athlete biological passport. *Endocrinology and Metabolism Clinics of North America* 39(1):59-73, viii-ix. doi: 10.1016/j.ecl.2009.11.003.

- Sotton, B., Paris, A., Le Manach, S., Blond, A., Lacroix, G., Millot, A., Duval, C., Huet, H., Qiao, Q., Labrut, S., Chiappetta, G., Vinh, J., Catherine, A., & Marie, B. (2017). Metabolic changes in Medaka fish induced by cyanobacterial exposures in mesocosms: an integrative approach combining proteomic and metabolomic analyses. *Scientific Reports* 7:4051. doi: 10.1038/s41598-017-04423-z. Erratum in: *Scientific Reports*, 8:4270 (2018).
- Spriet, L.L., MacLean, D.A., Dyck, D.J., Hultman, E., Cederblad, G., & Graham, T.E. (1992). Caffeine ingestion and muscle metabolism during prolonged exercise in humans. *American Journal of Physiology* 262(6 Pt 1):E891-898. doi: 10.1152/ajpendo.1992.262.6.E891.
- Thevis, M., Geyer, H., Mareck, U., Sigmund, G., Henke, J., Henke, L., Schänzer, W. (2007). Detection of manipulation in doping control urine sample collection: a multidisciplinary approach to determine identical urine samples. *Analytical Bioanalytical Chemistry* 388(7):1539-1543. doi: 10.1007/s00216-006-1112-z.
- Umpierre, D., Ribeiro, P.A., Kramer, C.K., Leitão, C.B., Zucatti, A.T., Azevedo, M.J., Gross, J.L., Ribeiro, J.P., & Schaan, B.D. (2011). Physical activity advice only or structured exercise training and association with HbA1c levels in type 2 diabetes: a systematic review and meta-analysis. *JAMA* 305(17):1790-1799. doi: 10.1001/jama.2011.576.
- van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., Bieri, B., Mazutis, L., Wolf, G., Krishnaswamy, S., & Pe'er, D. (2018). Recovering gene interactions from single-cell data using data diffusion. *Cell* 174(3):716-729.e27. doi: 10.1016/j.cell.2018.05.061.
- Van Renterghem, P., Van Eenoo, P., Geyer, H., Schänzer, W., Delbeke, F.T. (2010). Reference ranges for urinary concentrations and ratios of endogenous steroids, which can be used as markers for steroid misuse, in a Caucasian population of athletes. *Steroids* 75(2):154-163. doi: 10.1016/j.steroids.2009.11.008.
- Van Renterghem, P., Sottas, P.E., Saugy, M., & Van Eenoo, P. (2013). Statistical discrimination of steroid profiles in doping control with support vector machines. *Analytica Chimica Acta* 768:41-48. doi: 10.1016/j.aca.2013.01.003.
- WADA. The World Anti-Doping Code. (2020). The 2016 prohibited list international standard. Available at: https://www.wada-ama.org/sites/default/files/wada_2020_english_prohibited_list_0.pdf [accession January 2020].
- Wang, G., Karanikolou, A., Verdouka, I., Friedmann, T., & Pitsiladis, Y. (2017). Next generation "omics" approaches in the "fight" against blood doping. *Medicine and Sport Science* 62:119-128. doi: 10.1159/000470919.
- Wax, B., Kavazis, A.N., Brown, S.P., & Hilton, L. (2013). Effects of supplemental GAKIC ingestion on resistance training performance in trained men. *Research Quarterly for Exercise and Sport* 84(2):245-251. doi: 10.1080/02701367.2013.784845.
- Younes, A.K. & Younes, N.K. (2017). Recovery of steroid induced adrenal insufficiency. *Translational Pediatrics* 6(4):269-273. doi: 10.21037/tp.2017.10.01.
- Ziegenfuss, T., Landis, J., & Hofheins, J. (2008). Acute supplementation with alpha-glycerylphosphorylcholine augments growth hormone response to, and peak force production during, resistance exercise. *Journal of International Society Sports Nutrition* 5(Suppl 1):15. doi: 10.1186/1550-2783-5-S1-P15.
- Zheng, X., Qi, Y., Bi, L., Shi, W., Zhang, Y., Zhao, D., Hu, S., Li, M., & Li, Q. (2020). Effects of exercise on blood glucose and glycemic variability in type 2 diabetic patients with dawn phenomenon. *Biomedical Research International* 2020:6408724. doi: 10.1155/2020/6408724.
- Zorzoli, M., & Rossi, F. (2010). Implementation of the biological passport: the experience of the International Cycling Union. *Drug Testing and Analysis* 2(11-12):542-547. doi: 10.1002/dta.173.
- Zorzoli, M., & Rossi, F. (2012). Case studies on ESA-doping as revealed by the biological passport. *Drug Testing and Analysis* 4(11):854-858. doi: 10.1002/dta.1340.
- Zuber, V., & Strimmer, K. (2009). Gene ranking and biomarker discovery under correlation. *Bioinformatics* 25(20):2700-2707. doi: 10.1093/bioinformatics/btp460.