ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/268789759

Untargeted profiling of urinary steroid metabolites after testosterone ingestion: Opening new perspectives for antidoping testing

Article *in* Bioanalysis · October 2014

DOI: 10.4155/bio.14.200 · Source: PubMed

CITATIONS
24

READS
426

10 authors, including:

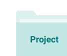
Julien Boccard
University of Geneva
140 PUBLICATIONS **2,745** CITATIONS
SEE PROFILE

Flavia Badoud
Nestlé S.A.
20 PUBLICATIONS **959** CITATIONS
SEE PROFILE

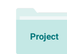
Nicolas Jan
Lausanne University Hospital
9 PUBLICATIONS **96** CITATIONS
SEE PROFILE

Raul Nicoli
Lausanne University Hospital
46 PUBLICATIONS **778** CITATIONS
SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Analytical chemistry View project



Biological monitoring of doping in male and female athletes: a multi-layered –omics approach View project

All content following this page was uploaded by Jean-Luc Veuthey on 23 December 2014.

The user has requested enhancement of the downloaded file.

Research Article
Mini Focus Issue: Systems Biology

Bioanalysis

For reprint orders, please contact reprints@future-science.com

Untargeted profiling of urinary steroid metabolites after testosterone ingestion: opening new perspectives for antidoping testing

Aim: Antidoping procedures are expected to greatly benefit from untargeted metabolomic approaches through the discovery of new biomarkers of prohibited substances abuse. **Results:** Endogenous steroid metabolites were monitored in urine samples from a controlled elimination study of testosterone undecanoate after ingestion. A platform coupling ultra-high pressure LC with high-resolution quadrupole TOF MS was used and high between-subject metabolic variability was successfully handled using a multiblock data analysis strategy. Links between specific subsets of metabolites and influential genetic polymorphisms of the UGT2B17 enzyme were highlighted. **Conclusion:** This exploratory metabolomic strategy constitutes a first step toward a better understanding of the underlying patterns driving the high interindividual variability of steroid metabolism. Promising biomarkers were selected for further targeted study.

Keywords: doping analysis • multiblock data analysis • steroidomics • testosterone • UHPLC-QTOF/MS • urine

Julien Boccard[1], Flavia Badoud[1,2], Nicolas Jan[2], Raul Nicoli[2], Carine Schweizer[2], François Pralong[3], Jean-Luc Veuthey[1], Norbert Baume[2], Serge Rudaz[1] & Martial Saugy*[,2]

[1]School of Pharmaceutical Sciences, University of Geneva, University of Lausanne, Geneva, Switzerland
[2]Swiss Laboratory for Doping Analyses, University Center of Legal Medicine, Epalinges, Switzerland
[3]Service of Endocrinology, Diabetology & Metabolism, University Hospital & Faculty of Biology & Medicine, Lausanne, Switzerland
*Author for correspondence:
Tel.: +41 21 314 73 30
Fax: +41 21 314 70 95
martial.saugy@chuv.ch

Endogenous anabolic androgenic steroids (AAS) include substances structurally related to testosterone (T) as well as biosynthetic precursors and metabolites. Because AAS improve physical performance, but also due to their several adverse effects on human health, they are forbidden in sport. Potential consequences include cardiac disorders, liver toxicity, sexual dysfunction and behavioral alterations [1]. Monitoring steroid metabolites in urine has become a routine practice in antidoping laboratories as an initial testing procedure to detect AAS misuse. So far, the investigation of the **steroid profile** is limited to a series of known endogenous compounds including free and glucuronidated forms of T, epitestosterone (E), androsterone (A), etiocholanolone (Etio), 5α-androstane-3α,17β-diol (Adiol) and 5β-androstane-3α,17β-diol (Bdiol) [2]. Threshold values were defined by antidoping authorities from population studies for both levels of single compounds and ratios, such as the T to E ratio (T/E), a well-known parameter for tracking down AAS administration. Other related compounds, such as dehydroepiandrosterone (DHEA) and dihydrotestosterone (DHT) are also part of the AAS metabolic pathway.

Today, routine screening of AAS misuse is mainly implemented using GC–MS platforms after a deconjugation and derivatization step [3]. Thanks to proper quantitation of parameters from the steroid profile, suspicious urine samples exhibiting unexpected patterns of steroid metabolites can be distinguished from normal ones [4,5]. For confirmation purpose, the former require additional targeted testing by isotope ratio MS (IRMS). Isotopic signature ($^{13}C/^{12}C$ ratio) is necessary for the unambiguous distinction between naturally occurring biosynthesized hormones and artificial synthetic equivalents, and is recognized by the World Anti-Doping Agency (WADA) as the reference confirmatory method for AAS misuse [6].

Up to recently, the control of doping practices was mainly relying on the comparison

FUTURE SCIENCE part of fsg

Key terms

Steroid profile: A subset of known endogenous compounds structurally related to testosterone, its biosynthetic precursors and metabolites, used in antidoping protocols for tracking down anabolic androgenic steroid administration.

Athlete biological passport: The program and methods defined by the World Anti-Doping Agency to collect longitudinal profiles of doping biomarkers and related information for a specific athlete.

Phase II metabolism: Biotransformation mechanisms involving conjugation reactions making substrates more water soluble for the detoxification of xenobiotics and endogenous compounds.

Coclustering: A data mining strategy building groups of similar subsets of variables (biomarkers) and observations simultaneously.

UDP-glucuronosyltransferase UGT2B17: An important enzyme for C19 steroid glucuronidation, including testosterone, reported to be responsible for two-thirds of urinary concentration between-subject variability.

of measured parameters to population-based criteria. This approach is limited by major natural individual and ethnic variations [7,8]. The introduction of the **athlete biological passport** (ABP) [9] has considerably changed the way analytical results are reported and evaluated by antidoping laboratories. This approach takes advantage of the individual follow-up of athletes over time through the regular monitoring of selected biochemical parameters. It is to be noted that the measured markers can be either direct or indirect signs of doping practices, and a steroidal module was implemented in the ABP very recently [2]. These data feed an adaptive Bayesian model [10] accounting for both population statistics and individual results collected in the ABP to evaluate personal reference levels and natural variations. As an extension to the Bayesian model, pattern classification approaches including machine learning schemes, such as support vector machines [11], can be implemented for the evaluation of a score indicating the likelihood of a doping offence. These statistical learning strategies constitute powerful tools but their performance remains somewhat limited by the choice of parameters that are introduced in the model. Therefore, doping control might still greatly benefit from the addition of relevant parameters to the ABP, including new steroid metabolites.

The GC–MS-based screening procedure for AAS misuse has been demonstrated to deliver trustworthy and reproducible results fulfilling sensitivity requirements defined by the WADA [2]. However, it requires a hydrolysis step, whose efficiency depends on several parameters [12], and may miss important information regarding **phase II metabolism**, because AAS are mainly excreted in urine as glucuronide or sulfate conjugates. Improvements of analytical setups coupling LC to MS (LC–MS) have been reported as reliable alternatives to GC–MS, allowing a broad monitoring of analytes in urine, including phase II metabolites [12–15]. The development of ultra-high pressure LC (UHPLC) provides higher chromatographic resolution when maintaining similar analysis time compared with conventional LC. Additionally, current TOF and Orbitrap-based MS devices are able to achieve data acquisition over a broad mass range (e.g., m/z 95–1000) with a wide dynamic range, accurate mass measurements and high sensitivity. Thousands of compounds can be ionized and detected within a single experiment, and this untargeted data acquisition allows the retrospective examination of samples without performing additional analyses [14]. Several applications have demonstrated the relevance of high-resolution MS for doping control purpose [16–22]. Notably, atmospheric pressure ionization modes such as ESI allow phase II metabolites to be directly investigated. These analytical developments strengthened the ability of antidoping laboratories to monitor unexpected changes in urine samples using holistic untargeted approaches and opened the way to the comprehensive monitoring of steroid metabolites accounting for phase II compounds. This opportunity is crucial for the discovery and identification of new metabolites in the perspective to enhance the detection ability and detection windows of existing testing procedures [23,24]. With that end in view, urine samples were collected from a previous controlled elimination study of T undecanoate (TU) after ingestion [25]. The present work refers to the analysis of data produced by UHPLC-QTOF/MS experiments using a state-of-the-art MS device. Untargeted steroidomic profiling was implemented for monitoring steroid metabolites in the context of AAS misuse detection. Data collected from each volunteer were summarized in a specific table and a multiblock data mining strategy was applied for biomarker discovery. Multiblock approaches aim at investigating the underlying relationships between a series of data blocks of possibly related variables [26]. By assessing the relevance of each block, this methodology allowed interindividual metabolic heterogeneity to be taken into account. **Coclustering** was then performed to gain further insight into a subset of highly relevant biomarkers with respect to UGT2B17 genotypes.

Experimental

Materials

Water and acetonitrile of ULC–MS quality were supplied by Biosolve (Chemie Brunschwig, Basel,

Switzerland). Formic acid (FA) and ammonium hydroxide (NH_4OH) were obtained from Fluka (Buchs, Switzerland). Methanol (CH_3OH) was supplied by Merck (Darmstadt, Germany). Oral pills of TU (40 mg, Andriol Testocaps®) were purchased from Essex Chemie AG, Luzern, Switzerland.

Cohort description & urine collection

A clinical trial was carried out to assess the effects of TU administration in healthy male subjects. A cohort of 19 volunteers, aged 19–28, was recruited. All subjects gave signed consent form and the protocol was authorized by the Ethical Commission for the Clinical Research of the Faculty of Biology and Medicine (University of Lausanne, Lausanne, Switzerland) and Swissmedic (protocol n°155/11). **UDP-glucuronosyltransferase UGT2B17** genotyping was achieved, as this enzyme is important for C19 steroid glucuronidation, including testosterone [27]. UGT2B17 deletion polymorphism was reported to be responsible for two-thirds of the between-subject variability [28]. The three possible genotypes formed well-balanced groups with five del/del (homozygote gene deletion), seven ins/del (heterozygote with one functional allele) and seven ins/ins (wild type with two functional alleles) subjects in the cohort. Details about study design and exclusion criteria can be found elsewhere [25]. Urine and blood samples were taken on a regular basis. Control samples were collected for each volunteer during the pretreatment week on Monday and Tuesday (sampling time points: C0 = 0 h, C1 = 2 h, C2 = 4 h, C3 = 8 h, C4 = 12 h, C5 = 24 h). Two oral pills of TU (40 mg, Andriol Testocaps) were taken by all volunteers during the treatment week on Monday and Wednesday. Corresponding urine samples were collected from Monday to Friday (sampling time points: T0 = 0 h, T1 = 2 h, T2 = 4 h, T3 = 8 h, T4 = 12 h, T5 = 24 h, T6 = 48 h, T7 = 60 h, T8 = 72 h, T9 = 96 h), frozen at -20°C and analyzed within 6 months. A series of 304 urine samples corresponding to control and post oral administration time points were considered, while blood samples were used for genotyping. The study design is presented on Figure 1.

Steroids extraction

Urine preparation was carried out using SPE Oasis HLB cartridges 30 mg (30 µm particle size, Waters, MA, USA) in the 96-well plate format [13]. This procedure allowed the selective extraction of glucuro- and sulfo-conjugated steroids. Conditioning and equilibration were achieved with 500 µl of CH_3OH and 500 µl of an aqueous solution containing 2% FA, respectively. The loading solution was composed of 1 ml of urine and 1 ml of 2% FA. A first washing step with 1 ml of a solution of 2% FA and a second one with 1 ml of a mixture 5% NH_4OH-CH_3OH 90:10 (v/v) were applied. The elution step consisted of 500 µl CH_3OH-water 95:5 (v/v). The elution phase was evaporated to dryness under gentle air stream, reconstituted with 100 µl of a mixture CH_3CN-water 70:30 (v/v) and injected in the LC–MS system.

UHPLC–MS analyses

Urine analysis was performed using an Acquity UPLC system (Waters, MA, USA) coupled to a Synapt G2 HDMS mass spectrometer (quadrupole TOF, QTOF, Waters) equipped with an Intellistart fluidic system and an ESI source operated in negative mode. Chromatographic separation was performed at 25°C and a flow rate of 300 µl/min on an Acquity UPLC column (BEH C_{18} 150 × 2.1 mm, 1.7 µm) equipped with a Van Guard precolumn (BEH C_{18} 5 × 2.1 mm, 1.7 µm). The mobile phase was 0.1% FA in water, and 0.1% FA in CH_3CN. The gradient started linearly from 5 to 37% B over 25 min, increased to 95% B for 3 min and was re-equilibrated for 8 min. The injection volume was fixed at 10 µl in the full loop mode and samples were maintained at 4°C in the autosampler.

MS parameters were as follows: capillary voltage at 2.4 kV in ESI negative mode, sampling cone at 50 V, extraction cone at 5 V, source temperature at 120°C, desolvation temperature 360°C, cone gas flow at 20 l/h and desolvation gas flow at 850 l/h. Data acquisition was carried out in the MS^E mode and two functions were recorded simultaneously. A first function was operated at low collision energy (6 eV) over the 100–600 *m/z* range and a second one with a collision energy ramp (35–50 eV) over *m/z* 50–600. The resolution mode operating in V-optics centroid was used for data collection, with a scan time of 0.2 s and an interscan delay of 0.02 s. A solution of leucine–enkephalin (Sigma-Aldrich, Buchs, Switzerland) at 2 ng/ml was infused through the LockSpray probe at 10 µl/min with a scan time of 0.02 s and an interval of 15 s. Data acquisition and instrument control were performed using MassLynx software (Waters).

Data pretreatment

MarkerLynx XS software (Waters) was used for the untargeted handling of raw data from the first MS function acquired at low collision energy. Parameters for automatic peak detection were as follows: time window from 5 to 26 min in the 100–600 *m/z* range, mass tolerance of 0.05 amu and retention time tolerance of 30 s. Ion features detected in the different samples within these limits were considered as a single analyte. A set of 3729 ion features, associated with a specific *m/z* value and chromatographic retention time,

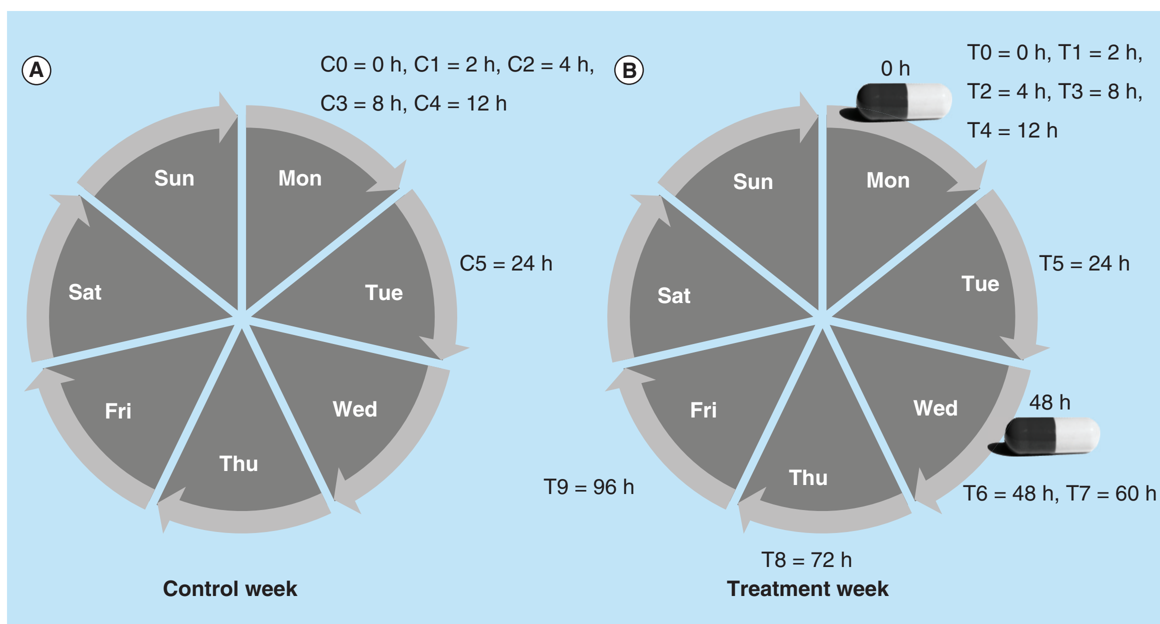


Figure 1. Controlled elimination study design. (A) Control week without treatment. **(B)** Treatment week.

was generated for each sample and peaks areas were normalized using unit norm (scaling to the sum of the total spectrum).

A selection procedure was then carried out to pick up ion features putatively related to steroid metabolites and discard unrelated substances. For that purpose, a previously reported filtering strategy [29] based on the comparison of untargeted acquisition data with reference *m/z* values from the Lipid Maps database [30], was used. Each of the 3729 automatically extracted ion features was compared with reference *m/z* values related to glucuronidated and sulfated forms of endogenous steroid conjugates belonging to the sterol lipids class. Hits were considered within a tolerance interval of ± 0.025 amu. A subset of 436 presumed steroid-related ion features was obtained for each of the 304 urine samples of the study. The steroid structure was confirmed based on the exact mass of the molecular ion from the first function acquired at low energy (6 eV), while the glucuro- and sulfoconjugation was assessed based on the fragmentation pattern provided by the second function (energy ramp, 35–50 eV) as fully described elsewhere [13]. Briefly, sulfoconjugated steroids were identified by the presence of *m/z* 96.96 ($[H_2SO_4]^-$), while glucuroconjugated steroids were confirmed by the presence of common fragments for glucuronides (e.g., *m/z* 157.01, 113.02, 85.03, 75.01 and 71.02). Additionally, a set of 12 known steroid metabolites from the steroid profile was quantified against isotopically labeled standards and reports were generated using TargetLynx software (Waters). It included T glucuronide (TG), T sulfate (TS), E glucuronide (EG), E sulfate (ES), A glucuronide (AG), A sulfate (AS), Etio glucuronide (EtioG), Etio sulfate (EtioS), DHEA glucuronide (DHEAG), DHEA sulfate (DHEAS), 5α-androstane-3β,17β-diol glucuronide (trans-AdiolG) and DHT glucuronide (DHTG).

Data analysis

Data mining was achieved using consensus orthogonal partial least squares discriminant analysis (consensus OPLS-DA) [31] and models were calculated under the MATLAB® 8 environment (The MathWorks, MA, USA) with combinations of toolboxes and in-house functions. Leave-one-out cross-validation was carried out to assess classification performance from model predictions. Global prediction accuracy (ACC), sensitivity (SENS) and specificity (SPEC) were computed based on the number of true positive (TP), true negative (TN), false positive (FP, type I error) and false negative (FN, type II error) cases as follows: ACC = (TP + TN)/(TP + FN + TN + FP), SENS = TP/(TP + FN) and SPEC = TN/(TN + FP). Model validity was verified using permutation tests. Coclustering was carried out with the Bioinformatics Toolbox v4.2 using Euclidean distances and the Ward aggregation method.

Results & discussion

Data structure & multiblock modeling strategy

The monitoring of multiple parameters over time leads to multivariate time-series data, which cannot be analyzed straightforwardly. Because a vector of data is created by the evaluation of *P* variables at a single time point, the temporal follow-up of an individual produces

a data matrix of size *P* variables × *Q* time points. In the context of clinical studies, data collected from a cohort of subjects generates therefore a 3D structure of *N* individuals × *P* variables × *Q* time points. Classical chemometric tools dedicated to multivariate analysis cannot be applied directly to extract relevant patterns from such a high-order data. Nevertheless, these data structure can be seen as a special case in the multiblock framework, with an equal number of observations and variables in each block [32]. Multiblock methods include algorithms that are able to handle a collection of data tables related to the same series of observations and integrate common and/or specific information in a global model. These advanced modeling strategies aim to assess the relevance of each data table and investigate their underlying relationships. This is particularly appropriate when analyzing clinical data, as each individual is characterized by its own genetic background and metabolism. By considering data gathered from each single individual as a specific data table (*P* variables × *Q* time points), natural heterogeneity can be handled suitably by the model. In that perspective, consensus OPLS-DA [31] was applied to gain insight into the collected data. This supervised multiblock method implements data fusion in linear kernel space and combines the advantages of easy interpretability from the OPLS framework and low computing needs, even in the case of highly multivariate data. Such a methodology allows focusing on common information, useful for biomarker detection, but also on individual specificities, such as particular metabolic rates or UGT2B17 genotypes. A view of the data analysis strategy is provided in Figure 2.

In order to highlight metabolic variations related to TU intake, consensus OPLS-DA models were calculated to distinguish urine samples of the control class (CTRL, n = 7 sampling points), collected during the control week (i.e., C0–C5) plus the first sample of the treatment week taken just before pill ingestion (i.e., T0), from the treated class (TESTO, n = 9 sampling points) composed of all the remaining samples of the treatment week (i.e., T1–T9). Autoscaling was applied to each data block before modeling. A first model was computed from the targeted analysis of the steroid profile (12 quantified steroids and the T/E ratio, targeted model, 13 variables), and a second one was calculated from the untargeted profiling data (436 presumed steroid-related ion features, untargeted model). Leave-one-out cross-validation was performed to evaluate the optimal model size and ensure statistical validity. In both cases, a model with two latent variables (one predictive and one orthogonal component) was obtained with satisfactory classification accuracies of 81.3% for the targeted model (specificity 85.7%, sensitivity 77.8%) and 87.5% for the untargeted model (specificity 85.7%, sensitivity 88.9%). Adequate classification results constitute a prerequisite for further model investigation to extract relevant biochemical information. However, due to the limited number of volunteers and sampling time points, prediction performance has to be considered cautiously. It is to be noted that the present study and the associated statistical models were designed to support biomarker discovery but not predictions for daily routine detection of AAS abuse.

Extraction of biochemical information for biomarker discovery

Consensus OPLS-DA generates linear models associated with parameters, such as scores and loadings, allowing straightforward retrieval of the biochemical information. While score plots display the coordinates of the observations in the model subspace accounting for all data blocks, loadings provide information about the contribution of each individual variable. Both consensus OPLS-DA models were characterized by a predictive component designed to separate CTRL from TESTO sampling time points and an orthogonal component summarizing other sources of systematic variability unrelated to the classes. In both cases (Figures 3A & 4A), the score plot highlighted a clear separation between urine collection time points from the control (CTRL) and the treatment week (TESTO) on the predictive component (t_p). The modeling strategy allowed therefore relevant alterations of metabolite levels related to TU intake to be distinguished from confounding factors and natural variations. This information was summarized by the predictive component in a meaningful and compact way. Loadings were then computed for each block of the consensus OPLS-DA models to assess the contribution of the variables, in other words, the measured compounds, to the predictive component (p_p). As each volunteer was related to a particular data table, individual as well as average contributions could be obtained. In order to give insight into the differences between UGT2B17 genotypes, average contributions were calculated for the whole cohort but also for each subgroup of ins/ins, ins/del and del/del volunteers. Such a strategy allowed the impact of each metabolite to be evaluated with respect to natural metabolic variations related to specific UGT2B17 characteristics. It is to be noted that model loadings are not restricted to a variable fold change or a detection window but combine these two important criteria in a multivariate perspective.

Loadings from the targeted model were investigated to assess changes related to known metabolites from the steroid profile. The loading plot derived from the whole cohort (Figure 3B) highlighted increased

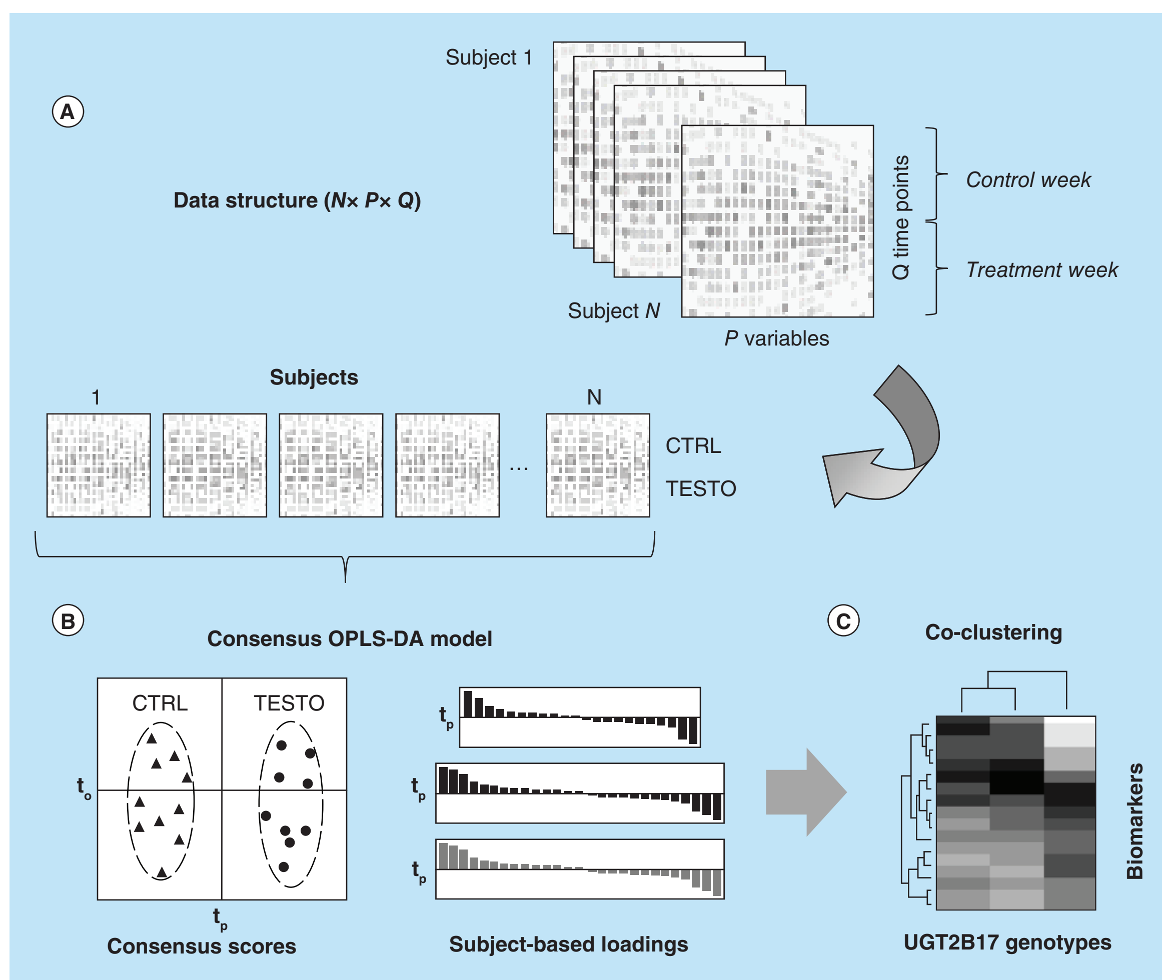


Figure 2. Data analysis workflow. (A) Multiblock data structure, **(B)** orthogonal partial least squares discriminant analysis modeling and **(C)** biomarkers coclustering.
CTRL: Control class; OPLS-DA: Orthogonal partial least squares discriminant analysis; TESTO: Treated class.

levels of EtioG, EtioS, AG, AS, TG, DHTG, trans-AdiolG and to a lesser extent TS, and an increased TG/EG ratio. DHEAG, DHEAS, EG and ES were located at the center of the plot, indicating similar levels before and after TU pills intake. Further biochemical information was gained by the inspection of the average loadings computed for each UGT2B17 genotype (Figure 3C, 3D & 3E). While similar patterns were observed for ins/ins and ins/del subgroups, the del/del genotype presented marked differences. First, DHTG and TG/EG were located near the center of the predictive component (p_p), indicating an absence of detectable alteration due to the TU treatment in the del/del subgroup. Additionally, a moderate increase of EG was observed in the TESTO class. These results are fully in line with prior knowledge of the metabolism of steroids and previous studies [7,33,34], and highlight the current limitations of population-based steroid control protocols regarding metabolic polymorphisms [35]. Notably, the T/E ratio was reported as a relevant parameter when integrating genotypic information in the ABP longitudinal profiles evaluation [36].

Loadings from the untargeted model were then examined to exploit the parallel monitoring of 436 steroid-related ion features provided by the UHPLC-QTOF/MS platform. A subset of 24 putative biomarkers was selected based on their contribution to the predictive component when considering all volunteers (Figure 4B, selected features are symbolized by black circles). Remarkably, an increased classification accuracy of 93.8% was obtained from this subset (specificity 100%, sensitivity 88.9%). The contributions of these compounds were further investigated with respect to specific genotypes. It is to be noted that several biomarkers presented a different pattern according to the UGT2B17 subclass considered (Figure 4C, 4D & 4E). Thanks to the comparison of exact mass and retention time with pure standards injected in the same conditions, ten of these compounds could

be identified as known steroid metabolites (Table 1). Among them, the detection of classical compounds such as TG (*m/z* 463.2333 at 11.5 min) and DHTG (*m/z* 465.2485 at 15.9 min) with average ranks of 11 and 16, respectively, using an untargeted metabolomic approach underlined the relevance of the proposed strategy for AAS monitoring. The importance of these two metabolites was considerably lower for del/del UGT2B17 genotypes (rank 123 and 176, respectively), in accordance with the targeted model and prior biological knowledge. Additionally, several androstanediol isomers could be highlighted at the top of the list, such as AdiolG (*m/z* 467.2634 at 16.5 min), Bdiol-3-O-G (*m/z* 467.2646 at 15.8 min), Bdiol- 17- O-G (*m/z* 467.2647 at 17.4 min) and trans-AdiolG (*m/z* 467.2636 at 10.4 min). The untargeted detection of this subclass of metabolites stresses the potential benefits related to the addition of new androstanediol isomers in routine protocols, as already reported in previous studies [37,38], and the need for future inclusion in the ABP. Moreover, since hydrolysis is not necessary prior to analysis, the metabolic information provided by LC–MS platforms is expected to be biologically more relevant.

Another series of six isobaric ions corresponding to *m/z* 481.2438 detected at distinct retention times contributed strongly to the list. This set included the top-ranked candidate *m/z* 481.2438 at 12.9 min, a compound that was already reported as a relevant biomarker of TU excretion [25,29], potentially related to 6- or 16-hydroxy AG or EtioG [34]. Notably, 6β-hydroxy AG and 6β-hydroxy EtioG were recently described as highly resistant to enzymatic hydrolysis by Fabregat *et al.* [39]. The authors underlined

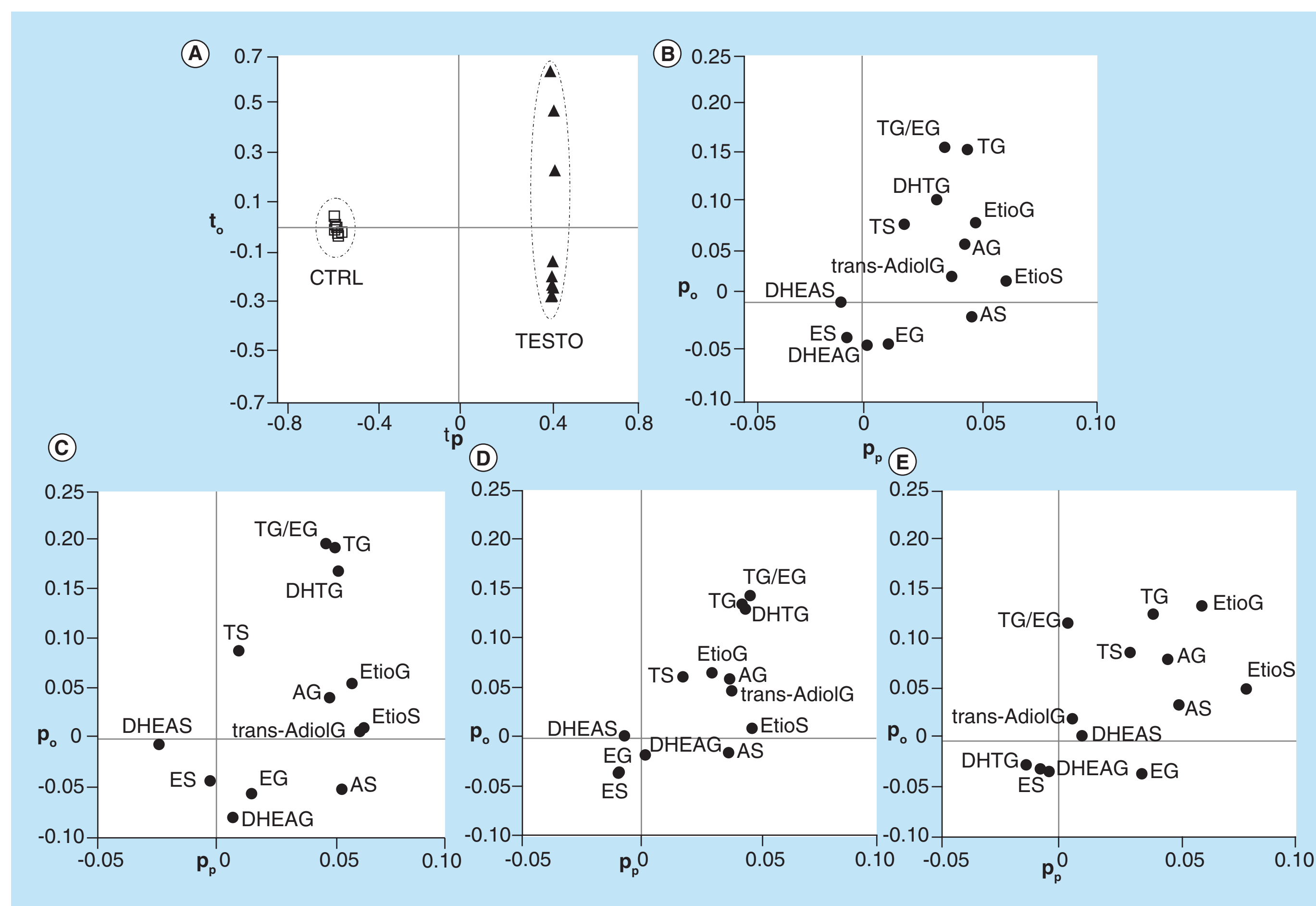


Figure 3. Targeted consensus orthogonal partial least squares discriminant analysis model. (A) Score plot. Sampling time points of the CTRL group are symbolized by white squares and time points of the TESTO class by black triangles. **(B)** Loading plot – all genotypes. **(C)** Loading plot – ins/ins genotype. **(D)** Loading plot – ins/del genotype. **(E)** Loading plot – del/del genotype. Known metabolites from the steroid profile are symbolized by black circles.

AdiolG: 5α-androstane-3β,17β-diol glucuronide; AG: Androsterone glucuronide; AS: Androsterone sulfate; CTRL: Control class; DHEAG: dehydroepiandrosterone glucuronide; DHEAS: Dehydroepiandrosterone sulfate; DHTG: Dihydrotestosterone glucuronide; EG: Epitestosterone glucuronide; EtioG: Etios glucuronide; ES: Epitestosterone sulfate; EtioS: Etio sulfate; TESTO: Treated class; TG: Testosterone glucuronide; trans-TS: Testosterone sulfate.

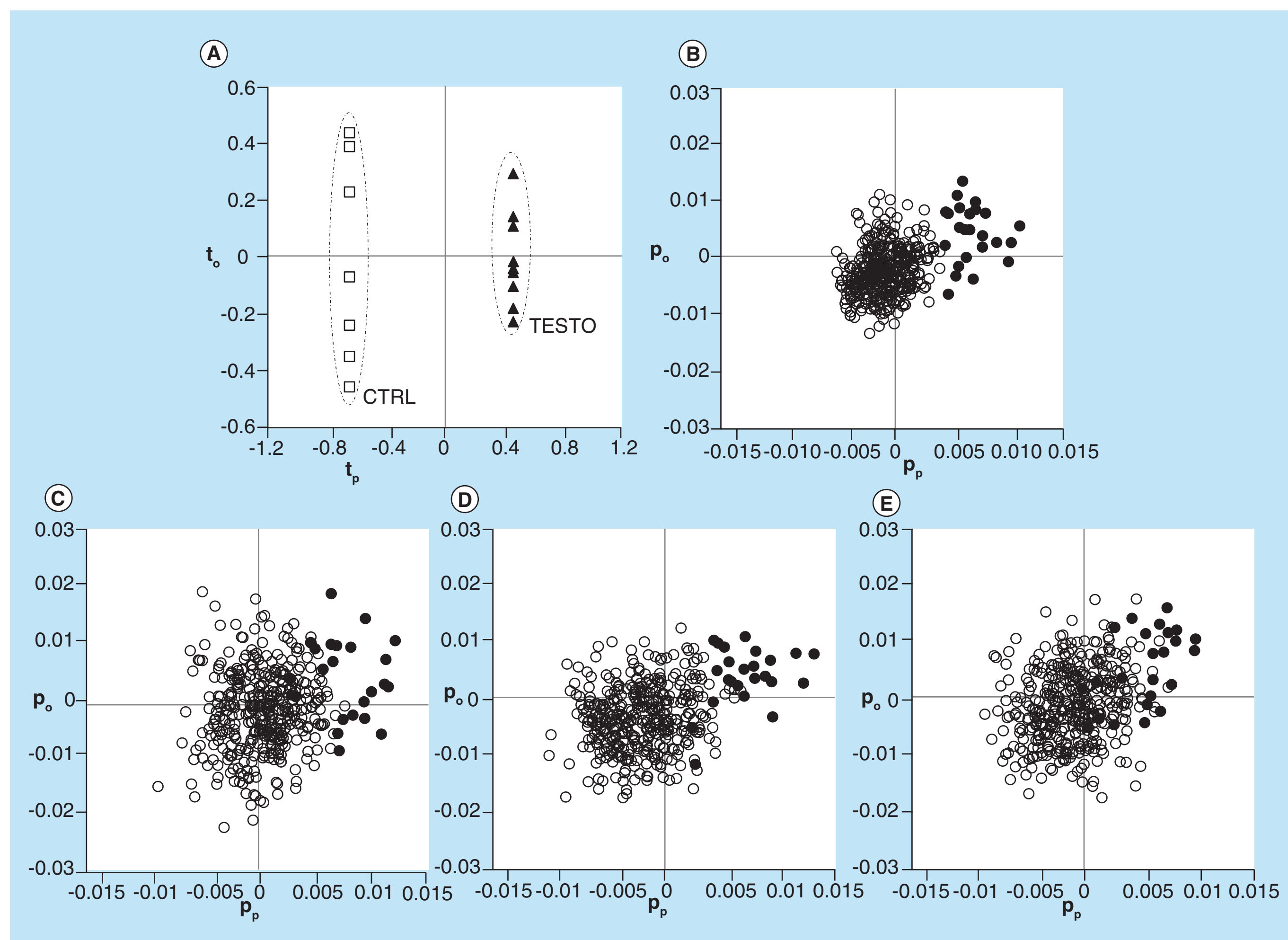


Figure 4. Untargeted consensus orthogonal partial least squares discriminant analysis model. (A) Score plot. Sampling time points of the CTRL group are symbolized by white squares and time points of the TESTO class by black triangles. **(B)** Loading plot – All genotypes. **(C)** Loading plot – ins/ins genotype. **(D)** Loading plot – ins/del genotype. **(E)** Loading plot – del/del genotype. Unknown metabolites are symbolized by circles with nonrelevant compounds in white circles and putative biomarkers in black.
CTRL: Control class; TESTO: Treated class.

the advantages of direct detection methods for the open monitoring of steroid metabolites glucuronide conjugates. This observation can be extended to sulfated metabolites, since another entry of the list (*m/z* 385.1685 at 12.0 min, overall rank 18) could be associated with a possibly sulfated form of hydroxy A or Etio. Despite the absence of formal identification, these candidate compounds further reinforce the hypothesis of possible increased detection ability and prolonged traceability of TU misuse through the use of hydroxylated metabolites compared with TG/EG [25,40]. Because AAS administration could saturate the main metabolic pathways, other metabolic routes associated with minor metabolites may lead to detectable concentrations of relevant biomarkers [41]. It also emphasizes the complementarity between glucuronide and sulfate conjugates for AAS screening [42]. As this series of metabolites was also well-ranked in the del/del UGT2B17 subgroup, it constitutes a very promising subset of parameters in the perspective of future AAS misuse monitoring. An in-depth study of MS patterns did not provide additional clues for unambiguous structural elucidation, due to the large number isomeric compounds. Targeted investigations allowing precise structural determination and dedicated quantitative evaluation remain therefore mandatory before routine implementation. Taken together, these results provide complementary information to a recent study dedicated to the targeted analysis of several T metabolites related to androstadiendione after alkaline treatment [43].

Average loadings calculated for each UGT2B17 genotype were further analyzed using a coclustering approach [44]. This data mining tool searches for clusters in both dimensions of the data table simultaneously and allows links between subsets of

Table 1. Subset of 24 candidate biomarkers extracted from untargeted data acquisition.							
m/z	RetTime	ID	SC	RankALL	RankINS/ INS	RankINS/ DEL	RankDEL/ DEL
481.2438	12.9	Hydroxy AG, hydroxy EtioG or isomer		1	6	1	2
467.2634	16.5	AdiolG	✓	2	2	2	16
467.2646	15.8	Bdiol-3-O-G	✓	3	5	4	1
467.2647	17.4	Bdiol-17-O-G	✓	4	4	3	64
481.2436	6.2	Hydroxy AG, hydroxy EtioG or isomer		5	11	9	3
467.2636	10.4	trans-AdiolG	✓	6	1	7	146
465.2482	18.7	Isomer of AG, EtioG, DHTG		7	9	5	51
463.2333	11.5	TG	✓	8	7	6	117
463.2353	18.4	Isomer of TG, DHEAG, EG		9	22	10	8
369.1737	20.6	EtioS	✓	10	15	12	10
465.2485	15.9	DHTG	✓	11	3	13	168
477.2122	10.1	Methoxyestradiol G, hydroxyandrostenedione G or isomer		12	17	15	11
481.2442	14.5	Hydroxy AG, hydroxy EtioG or isomer		13	14	21	18
465.2492	20.5	EtioG	✓	14	27	20	5
465.2491	21.1	AG	✓	15	21	24	6
483.2557	6.3	Hydroxyandrostane G or isomer		16	20	23	20
481.2435	6.6	Hydroxy AG, hydroxy EtioG or isomer		17	81	8	15
385.1685	12.0	Hydroxy AS, hydroxy EtioS or isomer		18	16	25	21
481.2436	5.7	Hydroxy AG, hydroxy EtioG or isomer		19	45	11	33
369.1737	19.6	AS	✓	20	10	59	19
465.2480	15.2	Isomer of AG, EtioG, DHTG		21	8	51	104
481.2431	9.4	Hydroxy AG, hydroxy EtioG or isomer		22	96	28	4
477.2123	6.4	Methoxyestradiol G, hydroxyandrostenedione G or isomer		23	53	18	63
477.2133	8.8	Methoxyestradiol G, hydroxyandrostenedione G or isomer		24	95	29	9

SC: ID confirmation based on pure standard.
Compounds are ranked according to the average contribution to the model. RankALL: average contributions to the model accounting for all UGT2B17 genotypes. RankINS/INS, RankINS/DEL RankDEL/DEL correspond to average contributions of ins/ins, ins/del and del/del UGT2B17 genotypes, respectively.
AG: Androsterone glucuronide; AS: Androsterone sulfate;DHEAG: Dehydroepiandrosterone glucuronide; DHTG: Dihydrotestosterone glucuronide;
EG: Epitestosterone glucuronide; ES: Epitestosterone sulfate; EtioG: Etio glucuronide; EtioS: Etio sulfate; SC: Standard confirmation; TG: Testosterone glucuronide;
trans-AdiolG: 5α-androstane-3β,17β-diol glucuronide.

variables (biomarkers) with similar patterns and subgroups of observations (UGT2B17 genotypes) to be detected. Hierarchical cluster analysis was applied using Euclidean distances and the Ward aggregation method. Two dendrograms were obtained and displayed together with a heat map summarizing the values taken by each element in Figure 5. The upper dendrogram related to the three UGT2B17 genotypes highlighted similar patterns of biomarkers between ins/ins and ins/del volunteers, indicating a probable compensatory mechanism of the single functional allele. By contrast, the del/del genotype exhibited markedly different levels of AAS metabolites indicating a null or very low compensation of other UGTs for the considered compounds. This observation is consistent with prior knowledge of UGT enzymes [45].

Additionally, four groups of metabolites sharing a similar behavior with respect to UGT2B17 genotypes could be highlighted based on the analysis of the left dendrogram associated to biomarkers clustering, and the heat map patterns:

The first subset of compounds displayed marked increases after TU intake for ins/ins, high to moderate modifications for ins/del and low variations for del/del volunteers. Notably, these biomarkers included TG, DHTG, trans-AdiolG, Bdiol-17-O-G and two isomers of AG, EtioG and DHTG. The occurrence of these metabolites at higher concentrations due to exogenous TU ingestion is therefore strongly related to the glucuronidation activity of the UGT2B17 enzyme.

A second cluster included biomarkers characterized by the highest predictive ability observed in the dataset. Markedly increased levels were detected irrespective of the genotype and these metabolites constituted therefore the most relevant parameters of the study. This small subset involved the biomarkers *m/z* 481.2438 at 12.93 min (6- or 16-hydroxy AG or EtioG), Bdiol-3-O-G and AdiolG. Interestingly, Bdiol-3-O-G was the most predictive metabolite for the del/del subgroup, while Bdiol-17-O-G (group 1) had a rank of 64 indicating a considerably lower predictive ability (Table 1).

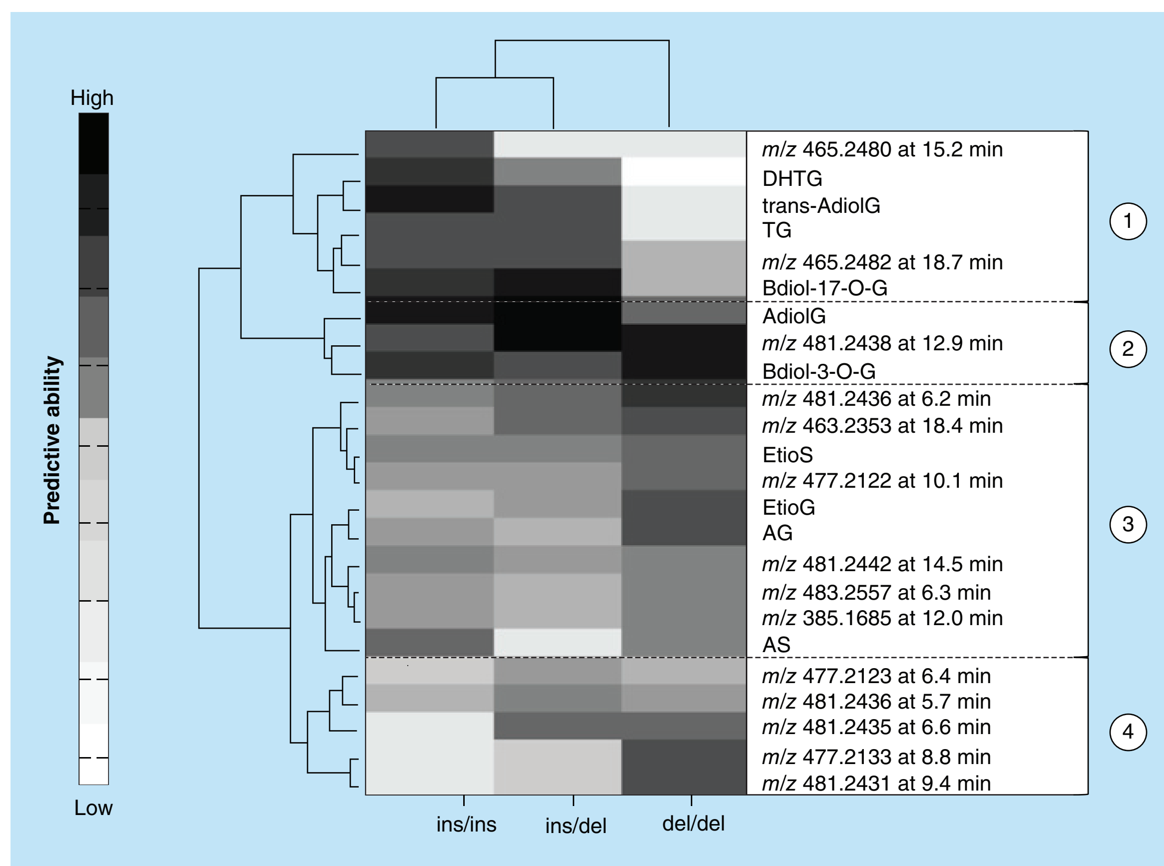


Figure 5. Coclustering using hierarchical cluster analysis and heat map of biomarkers average loadings related to each UGT2B17 genotype based on Euclidean distances and the Ward aggregation method.
AG: Androsterone glucuronide; AS: Androsterone sulfate; DHTG: Dihydrotestosterone glucuronide; EtioG: Etio glucuronide; EtioS: Etio sulfate; TG: Testosterone glucuronide; Trans-AdiolG: 5α-androstane-3β,17β-diol glucuronide.

The third subset involved ten moderately predictive metabolites for all UGT2B17 genotypes, including AG, AS, EtioG, EtioS. Remarkably, these compounds can be derived directly from androstenedione. Additionally, some compounds included in this cluster demonstrated high prediction ability for the group of volunteers bearing the homozygous deletion.

A last group of compounds shared a marked rise in samples corresponding to del/del genotypes, while low-to-moderate increases were observed for ins/ins and ins/del volunteers. This particular behavior may be related to the absence of UGT2B17 enzyme leading to the saturation of other metabolic pathways and the production of specific minor metabolites.

Conclusion

Untargeted metabolomic profiling constitutes an efficient approach for the discovery of new biomarkers to enhance doping control methods. This strategy was applied to the monitoring of steroid metabolites after TU ingestion in urine samples collected during a controlled elimination study. Experiments were carried out using a new generation UHPLC-QTOF/MS device and a series of 436 putative steroid metabolites was extracted from the 3729 ion features detected automatically. Data related to each volunteer were gathered in a specific data table and multivariate modeling accounting for interindividual metabolic heterogeneity was achieved using a multiblock approach originating from the OPLS framework. Two models were compared using 12 identified compounds of the steroid profile plus the T/E ratio on the one hand, and all 436 biomarker candidates on the other hand. Results obtained from the steroid profile monitoring were in accordance with prior biological knowledge, underlining the impact of UGT2B17 polymorphism on AAS metabolism. Comparable profiles were measured for ins/ins and ins/del volunteers suggesting a compensatory mechanism, while the del/del group presented a markedly different pattern. This further emphasized the need for the inclusion of additional parameters in antidoping protocols. In that perspective, a subset of 24 biomarkers with high predictive ability was selected from the steroidomic profiling and further analyzed using a coclustering approach. The latter strategy highlighted additional biochemical information in a very intuitive manner and allowed specific UGT2B17 genotypes to be linked to clusters of biomarkers with similar patterns. Ten of 24 compounds were identified as steroid metabolites based on comparison with reference material. Thanks to the high mass accuracy provided by TOF/MS, strong structural hypotheses could be formulated for the remaining candidates. Further targeted study of these metabolites in each genotype and each subject individually are expected to be highly valuable to doping control.

Executive summary

Experimental

- An extensive monitoring of endogenous steroid metabolites was carried out in urine samples from a controlled elimination study of testosterone undecanoate after ingestion.
- A platform coupling ultra-high pressure LC with high-resolution quadrupole TOF MS was used to assess compounds from the steroid profile classically evaluated in screening protocols, and an untargeted profiling of steroid metabolites.
- High between-subject metabolic variability was successfully handled using a multiblock data analysis strategy.
- A further investigation of the biomarkers contributions was done using coclustering.

Results

- Biochemical information derived from the steroid profile was in agreement with prior knowledge of testosterone metabolism and influential genetic polymorphisms of the UGT2B17 enzyme.
- A panel of 24 biomarker candidates was extracted from the untargeted steroidomic profiling and ten compounds were formally identified.
- Besides known compounds, a series of hydroxylated metabolites and androstanediol isomers were reported as promising parameters for enhanced detection capabilities of anabolic androgenic steroid doping practices.
- The need for a joint monitoring of both glucuronide and sulfate conjugates as complementary sources of metabolic information was underlined.
- Coclustering analysis highlighted links between specific subsets of metabolites and UGT2B17 genotypes.

Conclusion

- These results constitute a first step toward a better understanding of the underlying patterns driving the high interindividual variability of steroid metabolism.
- The exploratory metabolomic strategy allowed promising biomarkers to be selected for further targeted study, including subject-based approaches using individual reference levels, in the perspective of a future inclusion in the steroidal module of the athlete biological passport.

Future perspective

The present study underlined the potential benefits related to the addition of androstanediol isomers and hydroxylated metabolites to parameters traditionally evaluated in existing screening protocols, as well as the complementarity between glucuronide and sulfate conjugates in the perspective of enhanced doping control accounting for metabolic polymorphisms. The relevance of the inclusion of these new biomarkers in longitudinal follow-ups and subject-based strategies using individual references such as the ABP remains to be investigated.

Acknowledgements

The authors wish to thank Waters for the loan of the Synapt G2 HDMS system and their technical support. The authors also wish to thank the nursing staff of the Clinical Research Center at the CHUV Hospital, Lausanne, Switzerland for their help in volunteers' management and biological samples collection. The authors finally acknowledge the Medical Laboratory in the Division of Clinical Pharmacology from the Karolinska Institute in Stockholm (Sweden) for the genotyping of the blood samples and helpful discussion.

Financial & competing interests disclosure

This study was supported by the FIFA Medical Assessment and Research Centre (F-MARC) and by a grant from Antidoping Switzerland. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

References

Papers of special note have been highlighted as:
• of interest; •• of considerable interest

1 Van Amsterdam J, Opperhuizen A, Hartgens F. Adverse health effects of anabolic-androgenic steroids. *Regul. Toxicol. Pharmacol.* 57(1), 117–123 (2010).

2 World Anti-doping Agency (WADA). The world anti-doping code. Endogenous anabolic androgenic steroids, measurement and reporting, technical document TD2014EAAS, Montreal, Canada (2014). www.wada-ama.org/Documents/World_Anti-Doping_Program/WADP-IS-Laboratories/Technical_Documents/WADA-TD2014-EAAS-Endogenous-Anabolic-Androgenic-Steroids-Measurement-and-Reporting-EN.pdf

3 Ayotte C. Detecting the administration of endogenous anabolic androgenic steroids. In: *Doping in Sports: Biochemical Principles, Effects and Analysis.* Thieme D, Hemmersbach P (Eds). Springer, Heidelberg, Berlin, 77–98 (2010).

4 Kaabia Z, Dervilly-Pinel G, Popot MA *et al.* Monitoring the endogenous steroid profile disruption in urine and blood upon nandrolone administration: an efficient and innovative strategy to screen for nandrolone abuse in entire male horses. *Drug Test. Anal.* 6(4), 376–388 (2014).

5 Mareck U, Geyer H, Opfermann G, Thevis M, Schanzer W. Factors influencing the steroid profile in doping control analysis. *J. Mass Spectrom.* 43(7), 877–891 (2008).

•• **A review of factors influencing the steroid profile.**

6 Becchi M, Aguilera R, Farizon Y, Flament MM, Casabianca H, James P. Gas chromatography combustion isotope ratio mass-spectrometry analysis of urinary steroids to detect misuse of testosterone in sport. *Rapid Commun. Mass Spectrom.* 8(4), 304–308 (1994).

7 Peng SH, Segura J, Farre M, Gonzalez JC, De La Torre X. Plasma and urinary markers of oral testosterone undecanoate misuse. *Steroids* 67(1), 39–50 (2002).

8 Jakobsson J, Ekstrom L, Inotsume N *et al.* Large differences in testosterone excretion in Korean and Swedish men are strongly associated with a UDP-glucuronosyl transferase 2B17 polymorphism. *J. Clin. Endocrinol. Metab.* 91(2), 687–693 (2006).

9 Sottas PE, Robinson N, Rabin O, Saugy M. The athlete biological passport. *Clin. Chem.* 57(7), 969–976 (2011).

• **A review of the athlete biological passport.**

10 Sottas PE, Baume N, Saudan C, Schweizer C, Kamber M, Saugy M. Bayesian detection of abnormal values in longitudinal biomarkers with an application to T/E ratio. *Biostatistics* 8(2), 285–296 (2007).

11 Van Renterghem P, Sottas PE, Saugy M, Van Eenoo P. Statistical discrimination of steroid profiles in doping control with support vector machines. *Anal. Chim. Acta* 768, 41–48 (2013).

12 Gomez C, Fabregat, , Pozo OJ, Marcos J, Segura J, Ventura R. Analytical strategies based on mass spectrometric techniques for the study of steroid metabolism. *Trends Analyt. Chem.* 53(0), 106–116 (2014).

• **A review of analytical strategies for steroid monitoring.**

13 Badoud F, Grata E, Boccard J *et al.* Quantification of glucuronidated and sulfated steroids in human urine by ultra-high pressure liquid chromatography quadrupole time-of-flight mass spectrometry. *Anal. Bioanal. Chem.* 400(2), 503–516 (2011).

14 Badoud F, Guillarme D, Boccard J *et al.* Analytical aspects in doping control: Challenges and perspectives. *Forensic Sci. Int.* 213(1–3), 49–61 (2011).

•• **A review of analytical aspects related to doping control.**

15 Kim B, Moon JY, Choi MH *et al.* Global metabolomics and targeted steroid profiling reveal that rifampin, a strong human PXR activator, alters endogenous urinary steroid markers. *J. Proteome Res.* 12(3), 1359–1368 (2013).

16 Virus ED, Sobolevsky TG, Rodchenkov GM. Introduction of HPLC/orbitrap mass spectrometry as screening method for doping control. *J. Mass Spectrom.* 43(7), 949–957 (2008).

17 Peters RJB, Oosterink JE, Stolker AAM, Georgakopoulos C, Nielen MWF. Generic sample preparation combined with high-resolution liquid chromatography-time-of-flight mass spectrometry for unification of urine screening in doping-control laboratories. *Anal. Bioanal. Chem.* 396(7), 2583–2598 (2010).

18 Kolmonen M, Leinonen A, Kuuranne T, Pelander A, Ojanpera I. Generic sample preparation and dual polarity liquid chromatography – time-of-flight mass spectrometry for high-throughput screening in doping analysis. *Drug Test. Anal.* 1(5–6), 250–266 (2009).

19 Vonaparti A, Lyris E, Angelis YS *et al.* Preventive doping control screening analysis of prohibited substances in human urine using rapid-resolution liquid chromatography/high-resolution time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 24(11), 1595–1609 (2010).

20 Touber ME, Van Engelen MC, Georgakopoulus C, Van Rhijn JA, Nielen MWF. Multi-detection of corticosteroids in sports doping and veterinary control using high-resolution liquid chromatography/time-of-flight mass spectrometry. *Anal. Chim. Acta* 586(1–2), 137–146 (2007).

21 Giron AJ, Deventer K, Roels K, Van Eenoo P. Development and validation of an open screening method for diuretics, stimulants and selected compounds in human urine by UHPLC-HRMS for doping control. *Anal. Chim. Acta* 721, 137–146 (2012).

22 Pozo OJ, Van Eenoo P, Deventer K *et al.* Comparison between triple quadrupole, time of flight and hybrid quadrupole time of flight analysers coupled to liquid chromatography for the detection of anabolic steroids in doping control analysis. *Anal. Chim. Acta* 684(1–2), 107–120 (2011).

23 Thevis M, Kuuranne T, Geyer H, Schanzer W. Annual banned-substance review: analytical approaches in human sports drug testing. *Drug Test. Anal.* 5(1), 1–19 (2013).

24 Boccard J, Badoud F, Grata E *et al.* Urine metabolomics by UHPLC-QTOF-MSE: steroid profiling for anti-doping control. *Spectra Analyse* 41(284), 48–54 (2012).

25 Badoud F, Boccard J, Schweizer C, Pralong F, Saugy M, Baume N. Profiling of steroid metabolites after transdermal and oral administration of testosterone by ultra-high pressure liquid chromatography coupled to quadrupole time-of-flight mass spectrometry. *J. Steroid Biochem. Mol. Biol.* 138(0), 222–235 (2013).

26 Boccard J, Rudaz S. Harnessing the complexity of metabolomic data with chemometrics. *J. Chemom.* 28(1), 1–9 (2014).

27 Beaulieu M, Levesque E, Tchernof A, Beatty BG, Belanger A, Hum DW. Chromosomal localization, structure, and regulation of the UGT2B17 gene, encoding a C19 steroid metabolizing enzyme. *DNA Cell Biol.* 16(10), 1143–1154 (1997).

28 Rane A, Ekstrom L. Androgens and doping tests: genetic variation and pit-falls. *Br. J. Clin. Pharmacol.* 74(1), 3–15 (2012).

29 Boccard J, Badoud F, Grata E *et al.* A steroidomic approach for biomarkers discovery in doping control. *Forensic Sci. Int.* 213(1–3), 85–94 (2011).

30 Fahy E, Subramaniam S, Murphy RC *et al.* Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.* 50, S9–S14 (2009).

31 Boccard J, Rutledge DN. A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Anal. Chim. Acta* 769(0), 30–39 (2013).

• **The description of the consensus orthogonal partial-least squares modeling strategy.**

32 Smilde AK, Westerhuis JA, De Jong S. A framework for sequential multiblock component methods. *J. Chemom.* 17(6), 323–337 (2003).

33 Norli HR, Esbensen K, Westad F, Birkeland KI, Hemmersbach P. Chemometric evaluation of urinary steroid profiles in doping control. *J. Steroid Biochem.* 54(1–2), 83–88 (1995).

34 Van Renterghem P, Van Eenoo P, Geyer H, Schanzer W, Delbeke FT. 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 (2010).

35 Schulze JJ, Lorentzon M, Ohlsson C *et al.* Genetic aspects of epitestosterone formation and androgen disposition: influence of polymorphisms in CYP17 and UGT2B enzymes. *Pharmacogenet. Genomics* 18(6), 477–485 (2008).

36 Sottas PE, Saugy M, Saudan C. Endogenous steroid profiling in the athlete biological passport. *Endocrinol. Metab. Clin.* 39(1), 59–73 (2010).

37 Shackleton CHL, Roitman E, Phillips A, Chang T. Androstanediol and 5-androstenediol profiling for detecting exogenously administered dihydrotestosterone, epitestosterone, and dehydroepiandrosterone: potential use in gas chromatography isotope ratio mass spectrometry. *Steroids* 62(10), 665–673 (1997).

38 Van Renterghem P, Van Eenoo P, Delbeke FT. Population based evaluation of a multi-parametric steroid profiling on administered endogenous steroids in single low dose. *Steroids* 75(13–14), 1047–1057 (2010).

39 Fabregat A, Pozo OJ, Marcos J, Segura J, Ventura R. Use of LC–MS/MS for the open detection of steroid metabolites conjugated with glucuronic acid. *Anal. Chem.* 85(10), 5005–5014 (2013).

40 Van Renterghem P, Van Eenoo P, Sottas PE, Saugy M, Delbeke F. A pilot study on subject-based comprehensive steroid profiling: novel biomarkers to detect testosterone misuse in sports. *Clin. Endocrinol.* 75(1), 134–140 (2011).

41 Van Renterghem P, Van Eenoo P, Van Thuyne W, Geyer H, Schanzer W, Delbeke FT. Validation of an extended method for the detection of the misuse of endogenous steroids in sports, including new hydroxylated metabolites. *J. Chromatogr. B* 876(2), 225–235 (2008).

42 Dehennin L, Lafarge P, Dailly P, Bailloux D, Lafarge JP. Combined profile of androgen glucuro- and sulfoconjugates in postcompetition urine of sportsmen: a simple screening procedure using gas chromatography mass spectrometry. *J. Chromatogr. B* 687(1), 85–91 (1996).

43 Fabregat A, Pozo OJ, Marcos J, Segura J, Ventura R. Alternative markers for the long-term detection of oral testosterone misuse. *Steroids* 76(12), 1367–1376 (2011).

44 Bro R, Papalexakis EE, Acar E, Sidiropoulos ND. Coclustering – a useful tool for chemometrics. *J. Chemom.* 26(6), 256–263 (2012).

45 Guillemette C. Pharmacogenomics of human UDP – glucuronosyltransferase enzymes. *Pharmacogenomics J.* 3(3), 136–158 (2003).

View publication stats