Novel protein markers of androgen activity in humans: proteomic study of plasma from young chemically castrated men.

Aleksander Giwercman^{1*}, K Barbara Sahlin^{2,3*}, Indira Pla^{2,3}, Krzysztof Pawlowski^{2,4,6}, Carl Fehniger^{2,3}, Yvonne Lundberg Giwercman⁵, Irene Leijonhufvud¹, Roger Appelqvist^{2,3}, György Marko-Varga^{3,7}, Aniel Sanchez^{2,3†}, and Johan Malm^{2,3†}

¹ Molecular Reproductive Medicine, Department of Translational Medicine, Lund University, Malmö, Sweden

² Section for Clinical Chemistry, Department of Translational Medicine, Lund University, Skåne University Hospital Malmö, 214 28 Malmö, Sweden

³ Clinical Protein Science & Imaging, Biomedical Centre, Department of Biomedical Engineering, Lund University, BMC D13, 221 84 Lund, Sweden

⁴ Department of Experimental Design and Bioinformatics, Faculty of Agriculture and Biology, Warsaw University of Life Sciences SGGW, Warszawa, Poland

⁵ Molecular Genetic Reproductive Medicine, Department of Translational Medicine, Lund University, Malmö, Sweden

⁶ Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX, USA

⁷ First Department of Surgery, Tokyo Medical University, 6-7-1 Nishishinjiku Shinjiku-ku, Japan

* † The authors share the first and last positions.

Short title: Novel protein markers of androgen activity

Key words: testosterone, hypogonadism, proteomics, physiology, biomarker

Corresponding author and to whom reprint requests should be addressed: Aleksander Giwercman. Dept. of Translational Medicine. Lund University. CRC 91-10-058; Jan Waldenströms gata 35; SE 214 28 Malmö, Sweden. E-mail: aleksander.giwercman@med.lu.se

Abstract

Background Reliable biomarkers of androgen activity in humans are lacking. The aim of this study was, therefore, to identify new protein markers of biological androgen activity and test their predictive value in relation to low vs. normal testosterone values and some androgen deficiency linked pathologies.

Methods Blood samples from 30 healthy GnRH-antagonist treated males were collected at three time points: a) before GnRH antagonist administration; b) 3 weeks later, just before testosterone undecanoate injection, and c) after additional 2 weeks. Subsequently they were analysed by mass spectrometry to identify potential protein biomarkers of testosterone activity. Levels of proteins most significantly associated with testosterone fluctuations were further tested in a cohort of 75 hypo- and eugonadal males suffering from infertility. Associations between levels of those markers and cardio-metabolic parameters, bone mineral density as well as androgen receptor CAG repeat lengths, were explored.

Results Using ROC analysis, 4-hydroxyphenylpyruvate dioxygenase (4HPPD), insulin-like growth factor-binding protein 6 (IGFBP6) and fructose-bisphosphate aldolase (ALDOB), as well as a Multi Marker Algorithm, based on levels of 4HPPD and IGFBP6, were shown to be best predictors of low (< 8 nmol/L) vs. normal (> 12 nmol/L) testosterone. They were also more strongly associated with metabolic syndrome and diabetes than testosterone levels. Levels of ALDOB and 4HPPD levels also showed association with AR CAG-repeat lengths.

Conclusions We identified potential new protein biomarkers of testosterone action. Further investigations to elucidate their clinical potential are warranted.

Funding: The work was supported by ReproUnion 2.0 (grant no 20201846), which is funded by the Interreg V EU program.

Introduction

The male sex hormone, testosterone (T), plays an important physiological role in regulating function of both reproductive and non-reproductive organs in males as well as in females. In males, the diagnosis of T deficiency (i.e. male hypogonadism) is based on the presence of low serum T levels combined with clinical symptoms, which, however, are not pathognomonic for this condition (1).

The most common way of assessing T activity is by measuring the total concentration of this hormone in a fasting morning blood sample. However, total T does not accurately reflect biological androgenic activity (BAA), which might be considered a more useful biological and clinical marker.

The association between T levels and BAA is affected by several biological mechanisms such as the concentration of binding proteins, body mass index, certain diseases (e.g. diabetes), androgen receptor (AR) sensitivity (2), and different co-factors (3). So far, no reliable algorithms for translating T levels into BAA are available, but could be useful for example in the diagnosis of male hypogonadism.

A correct hypogonadism diagnosis is important for proper identification of men for whom androgen replacement therapy is warranted. However, the treatment of men with hypogonadism represents a clinical challenge, because the symptoms associated with the condition are highly nonspecific (4). Furthermore, there are limitations in using the level of T in defining testosterone deficiency. Generally, in many clinical guidelines, total T concentration below 8 nmol/L indicates an insufficient hormone concentration, whereas levels above 12 nmol/L are considered normal (5). Apart from the fact that men with testosterone levels between 8 and 12 nmol/L

cannot be assigned to any of these distinct categories, those presenting with a lower or higher hormone concentration may also be misclassified due to an abnormal concentration of sex hormone binding protein (SHBG). Low SHBG levels, as often seen in obese men, may imply low total T despite unaffected BAA. On the other hand, some degree of reduced androgen sensitivity may be associated with decreased BAA despite normal or high testosterone levels (6).

Hypogonadism has been identified as a predictor of several non-communicable chronic diseases as well as premature mortality (7). Understanding the biology of androgen action may therefore contribute to clarifying the pathogenetic mechanisms linking androgen deficiency to comorbid conditions. Thus, the approach based on measuring the protein levels downstream of androgen action is a feasible and logical concept for identifying clinical and biological useful markers of BAA.

Proteomics is a technique aimed to study biological systems based on qualitative and quantitative measuring of proteins and, thereby, integrate the cellular output related to transcription as well as translation. Mapping the quantitative protein response downstream of androgen action may provide new clinically valuable markers of BAA. In order to identify such markers, we compared the protein profile of healthy individuals before and after T deprivation. Subsequently, we assessed the predictive value of the newly identified protein markers in relation to hypogonadism and risk of pathologies related to T deficiency.

Subjects and Methods

Study outline

The study was set up to I) identify new protein markers of BAA in healthy subjects; II) test the markers' predictive values in relation to biochemically diagnosed hypogonadism, metabolic syndrome (MetS), cardiovascular risk lipid profile (CVRLP), diabetes mellitus II (DM) and low bone density (LBD) in infertile men; III) analyse androgen dependence of the identified proteins by assessing how their levels associate with androgen receptor gene CAG-repeat length.

Subjects

All subjects were enrolled with informed written consent. The two studies from which they were recruited were approved by the Swedish Ethical Review Authority (Approval number: DNR 2014/311, date of approval 8 May 2014; DNR 2011/1, date of approval 11 January 2011).

The first part of the study includes plasma samples obtained from 30 healthy men (biological replicates) aged 19-32 years, BMI 19.1-26.9 kg/m². They, underwent chemical castration by subcutaneous administration of 240 mg GnRH-antagonist (Degeralix®, Ferring Pharmaceuticals, Saint-Prex, Switzerland) followed by remediation of testosterone levels by intramuscular injection of 1000 mg testosterone undecanoate (Nebido®, Bayer AG, Leverkusen, Germany) after the duration of three weeks (8,9). Blood samples were collected at baseline (A); three weeks later (B), and, at the end of the study, after two additional weeks (C) (**Figure 1**).

To test the clinical predictive value of the proteins identified in the castrated men, we used a cohort of 75 serum samples from 75 men (biological replication, subject age 32-43 years) previously recruited for a study on hypogonadism among men from infertile couples (10). Eighty-five patients were randomly selected from 213 infertile men and 223 age-matched

controls. The selected patients for the present study had the span of subnormal to upper normal range of T. Out of the 85 patients, ten patients were excluded; seven due to Klinefelter syndrome, one due to missing value of T and the last two were statistical outliers, which were removed after considering the possible causes. One patient had a high level of T (41.6 nmol/L) without androgen replacement therapy (ART) and the other because he was the only one diagnosed with obstructive azoospermia. Background characteristics of these patients can be found in **Table 1a** and **Table 1b**.

Table 1a. Background characteristics of infertile patients.

	Mean (SD)	N
Age at inclusion (years)	37.8 (5.5)	75
BMI	27.2 (4.3)	72
Total testosterone (nmol/L)	12.8 (6.8)	75
FSH (IU/L)	15.8 (14.3)	75
LH (IU/L)	7.5 (5.7)	75
SHBG (nmol/L)	24.0 (4.5-84.5)*	75
Estradiol (pmol/L)	96 (36-321)*	75
Calculated Free testosterone (pmol/L)	260 (50-1360)*	75
ApoB/ApoA1	0.7 (0.2)	68
HOMA-IR	1.6 (0.4-13.9)*	75
DEXA score (lumbar z-score)	-0.5 (1.3)	74
CAG (repeated length)	22 (14-31)*	74

*Characteristics values are expressed as mean (SD), except for those that did not follow a normal distribution (non-Gaussian) and which are shown as median (min-max) *.*

Table 1b. Characteristics of the cohort of infertile patients.

	n (%)
Smoker	9 (12.0)
Current diseases	36 (48.0)
Insulin medication	1 (1.3)
Current ART	8 (10.7)
CVRLP	20 (27.0)
Insulin resistance	20 (27.0)
Diabetes Mellitus 2	4 (5.3)
Metabolic Syndrome (MetS)	14 (20.9)
Low bone density	23 (30.6)

Variables are expressed as frequency (n) and percentage of occurrence (%) - according to the total number of patients (N=75).

ART: Androgen Replacement Therapy;

CVRLP: cardiovascular risk lipid profile

**: 9 with allergy; 8 with local pains; 5 psychiatric disorders, 4 hormonal deficiencies; 2 with diabetes and 8 with mixed conditions*

- 121 The following comorbidities in the cohort of infertile patients were defined MetS, IR, CVRLP,
 122 DM and LBD. MetS was determined according to the criteria defined at the National Cholesterol

Education Program Adult Treatment Panel III 2002 (**Table S1**). Homeostatic Model Assessment of Insulin Resistance (HOMA-IR) was calculated as (glucose x insulin)/22.5 and IR was defined as HOMA-IR > 2.5 (11). CVRLP was defined as the ratio apolipoprotein B/apolipoprotein A1 > 0.9 (12). DM was set at fasting blood glucose > 7 mmol/L (13). LBD was determined based on the DEXA lumbar Z score with the cut off at < -1 (14). The methods for laboratory tests (10), CAG-repeat length (15) and proteomics (16,17) are described in the supplementary Appendix 1.

Statistical analysis

We briefly describe the statistical analyses performed. A full description of the statistical analysis is available in Appendix 1- *Supplementary Statistical analyses*. Proteomics data pre-processing was done using Perseus v1.6.7.0 (18) software and unless other software is specified, the statistical analyzes were performed using R software (19,20).

Healthy human model

Protein intensities were Log2 transformed and standardized by Subtract Median normalization. Differentially expressed proteins were determined by one-way repeated measures ANOVA followed by a pairwise t-test (two-tails and paired). Adjusted *p*-values < 0.05 were considered significant. The power of the candidate biomarkers to discriminate between normal and low T was evaluated by doing receiver operating characteristic (ROC) analysis. Significant proteins (between A-B with significant recovery in B-C) with a) area under the curve (AUC) > 0.80 or b) AUC between 0.75 – 0.80 (21–23) and highly enriched in liver tissues (Human Proteome Map (24) and Kampf C. et. al. (25))(26,27) were considered candidate biomarker. These candidates were included as predictors in a stepwise regression (method: backward) to select the best

combination of markers that predict the odds of being low T. Bootstrap resampling with replacement method was applied to assess consistency. A new variable called Multi Marker Algorithm (MMA) was derived from the predicted log-odds (of being low T) obtained from a binomial logistic regression analysis (see Appendix 1-*Supplementary Statistical analyses*) and it was evaluated together with marker candidates proteins.

Infertile cohort of patients

The normal distribution of the variables that describe background characteristics of the infertile cohort of patients (**Table 1a**) was evaluated by Kolmogorov-Smirnov test. The intensities of the candidate biomarkers were Log2 transformed to achieve normal distributions. In this cohort, MMA variable was created to predict the odds of suffering low T or other medical conditions associated with low T levels.

Changes between two groups were evaluated by two-tailed Student's T-test (p -values < 0.05 were considered significant). Overall changes between more than two groups were evaluated by one-way ANOVA followed by a pairwise FDR correction (28) and adjusted p -values < 0.05 were considered significant. In order to know if the changes in the candidate markers occur with the change in T as observed in the healthy human model, three groups of patients were created based on total T concentration (5)(group 1: low T (LT) ≤ 8 nmol/L (n= 22); group 2: borderline low T (BL_T) between 8 and 12 nmol/L (n=17); group 3: normal T (NT) > 12 nmol/L (n= 36)). Calculated free testosterone (cFT) was determined according to the method described by Vermeulen *et al.* (29). The cut-off level of 220 pmol/L was used to categorise the subjects as having low cFT (L_cFT; n=21) or normal cFT (N_cFT; n=54) (30).

The power of the candidate biomarkers to discriminate between LT and NT (including or not the BL_T) (31–33), or to distinguish patients with medical conditions associated with low T levels (MetS, IR, CVRLP, DM or LBD) was evaluated by a ROC analysis. The same was done to discriminate between L_cFT and N_cFT. The DeLong's test (paired) was used to compare the AUCs.

In order to strengthen the evidence of androgenic dependence of the candidate biomarkers, we looked for associations between their expression and the androgen receptor (AR) CAG repeat length, which was previously reported to have an impact on the activity of the receptor (34–37). Three categories were defined: reference group 1: patients with CAG repeat length 21 and 22 (n= 18); group 2: patients with CAG repeat length < 21 (n= 26) and group 3: patients with CAG repeat length >22). This categorisation was undertaken in order to have three groups of sufficient size and the category including the mean CAG length value of 22 was chosen as reference since this CAG number was previously seen, *in vitro* and *in vivo* to be associated with highest receptor activity (38,39) .

Results

Proteins differentially expressed in chemically castrated men

In total, in the healthy men, the expression level of 31 out of 676 proteins was statistically significantly associated with T concentration (**Table S2**). The levels of 23 proteins changed in the same direction as T, whereas, the remaining eight markers changed in an opposite way. LH and FSH changed significantly in A-B but not in B-C. The protein changes visualised as boxplots can be found in Figure 2—figure supplement 1 (<https://doi.org/10.6084/m9.figshare.14876562>). .

Proteins capable to distinguish between low and normal testosterone

Based on *p*-values for AUC in the ROC analysis, among healthy young men, 90% of the 31 proteins distinguished the low T time point (B) from the normal ones (A and C) with statistical significance (Table S3, **Figure 2a**). ROC-AUC values greater than 0.80 were obtained for the proteins 4-hydroxyphenylpyruvate dioxygenase (4HPPD) and insulin-like growth factor-binding protein 6 (IGFBP6). Additionally, fructose-bisphosphate aldolase (ALDOB) was the only protein enriched in liver tissue with ROC-AUC between 0.75 and 0.80.

The stepwise regression method selected 4HPPD and IGFBP6 as the best markers to be combined to predict the odds of being low T, and thus, they were the basis for the new variable MMA (see Material and Methods). MMA together with 4HPPD, ALDOB and IGFBP6 were selected as potential candidate markers for the diagnosis of BAA (**Figures 2b and 2c**). The expression of the 4HPPD and ALDOB proteins were significantly increased at low T ($p<0.001$; $p<0.001$) and remediated in response to the T treatment, whereas IGFBP6 expression was significantly decreased ($p<0.001$) by castration.

202 *Testing of the candidate biomarkers in infertile men*

203 The three proteins and MMA showed statistically significant differences (4HPPD: $p < 0.001$,
204 ALDOB: $p = 0.003$, IGFBP6: $p = 0.016$, MMA: $p < 0.001$, **Figure 3a**) between the three groups
205 defined according to the total T levels (**Table S2**). 4HPPD, ALDOB, and MMA showed a
206 negative association with T changes, while IGFBP6 displayed a positive association. The three
207 proteins and MMA significantly distinguished the patients with LT from BL_T/NT (4HPPD:
208 AUC=0.75, $p = 0.001$; ALDOB: AUC= 0.70, $p = 0.008$; IGFBP6: AUC= 0.69, $p = 0.01$; MMA:
209 AUC= 0.79, $p < 0.001$) (**Figure 3b**). Additionally, the power to discern low T values improved
210 for all the biomarkers tested when the patients with BL_T were excluded (**Table 2**). Similar
211 results were obtained for discrimination between low and normal FT (Figure 3 d-e; Table 2).

212 **Table 2. Comparison of ROC-Areas Under the Curve for testosterone and the candidate biomarkers in relation to the prediction of**
 213 **hypogonadism and its sequelae in patients.**

Analyte	Low T		Low T*		Low cFT		IR		DM		LBD		CVRLP		MetS	
	(T ≤ 8 nmol/L)		(T ≤ 8 nmol/L)		(cFT < 220 pmol/L)		(HOMA-ir > 2.5)		-		(z-score < -1)		(ApoB/ApoA1 ≥ 0.9)			
	AUC (Sp,Se)	p	AUC (Sp,Se)	p	AUC (Sp,Se)	p	AUC (Sp,Se)	p	AUC (Sp,Se)	p	AUC (Sp,Se)	p	AUC (Sp,Se)	p	AUC (Sp,Se)	p
4HPPD	0.75 (85,59)	<u>8.38E-04</u>	0.77 (86,59)	<u>2.66E-04</u>	0.69 (83,57)	<u>5.14E-03</u>	0.79 (84,70)	<u>1.20E-04</u>	0.89 (93,75)	<u>9.00E-03</u>	0.64 (76,61)	<u>2.24E-02</u>	0.74 (46,90)	<u>5.79E-03</u>	0.74 (95,50)	<u>5.31E-03</u>
ALDOB	0.69 (68,73)	<u>8.25E-03</u>	0.70 (67,73)	<u>5.39E-03</u>	0.66 (67,71)	<u>1.56E-02</u>	0.73 (71,75)	<u>2.85E-03</u>	0.85 (63,100)	<u>1.80E-02</u>	0.57 (98,27)	1.93E-01	0.71 (83,55)	<u>4.64E-02</u>	0.74 (82,64)	<u>6.02E-03</u>
IGFBP6	0.69 (77,59)	<u>1.05E-02</u>	0.70 (81,59)	<u>4.89E-03</u>	0.63 (44,81)	7.24E-02	0.57 (42,80)	3.50E-01	0.59 (39,100)	5.40E-01	0.63 (43,78)	2.75E-01	0.59 (48,80)	1.87E-01	0.65 (45,93)	8.45E-02
Testosterone	-	-	-	-	-	-	0.71 (76,70)	<u>4.96E-03</u>	0.55 (32,100)	7.24E-01	0.75 (74,74)	<u>5.12E-04</u>	0.66 (65,85)	<u>6.55E-03</u>	0.56 (72,57)	5.08E-01
MMA	0.79 (74,82)	<u>9.23E-05</u>	0.80 (72,86)	<u>3.86E-05</u>	0.70 (69,71)	<u>3.90E-03</u>	0.79 (82,70)	<u>1.46E-04</u>	0.92 (84,100)	<u>5.00E-03</u>	0.78 (82,65)	<u>1.40E-02</u>	0.75 (73,75)	<u>3.65E-03</u>	0.78 (63,86)	<u>1.57E-03</u>

Significant p-values are highlighted in bold and underlined. * Excluding patients with testosterone values from the borderline low testosterone ($8 < BL_T \leq 12$).

cFT: calculated free testosterone; IR: insulin resistance; DM: diabetes mellitus type 2; LBD: low bone density; CVRLP: cardiovascular risk lipid profile ; MetS: metabolic syndrome

AUC: area under the curve; Sp: Specificity in %; Se: Sensitivity in %

Ability to distinguish men with abnormal metabolic comorbidities or reduced bone mineral density

The AUCs for 4HPPD, ALDOB and MMA in relation to risk of DM and MetS were statistically significant whereas for T the p-value for AUC was 0.72. The AUCs for 4HPPD, ALDOB and MMA were also statistically significantly larger than this for T (DM: 4HPPD ($p= 0.005$), ALDOB ($p= 0.009$), MMA ($p= 0.002$); MetS: 4HPPD ($p= 0.032$), ALDOB ($p= 0.030$), MMA ($p= 0.002$)), (**Table 2**). Additionally, the AUC values in relation to CVRLP and IR were numerically higher for 4HPPD, ALDOB and MMA than for T, however, the differences between the AUC values were not statistically significant (CVRLP marker vs T: 4HPPD ($p= 0.97$), ALDOB ($p= 0.61$), MMA ($p= 0.87$); IR marker vs T: 4HPPD vs T ($p= 0.30$), ALDOB vs T ($p= 0.88$), MMA ($p= 0.31$)). 4HPPD and MMA statistically significantly distinguished between patients with normal bone density and LBD. The same was true for T, the differences between the AUC for T and those for 4HPPD and MMA not being statistically significant (LBD vs T: 4HPPD ($p= 0.28$), MMA ($p= 0.30$)). No statistical significance, in relation to prediction of LBD was found for IGFBP6 (**Figure 4**)

Association of the candidate biomarkers with androgen receptor CAG-repeat length

Statistically significant inter-CAG-group overall differences were observed for 4HPPD ($p= 0.012$) and ALDOB ($p= 0.008$) (**Figure 5**). Additionally, the protein expressions were significantly higher in the groups with <21 and >22 CAG repeat length as compared with the reference (**Table 3** and **Figure 5**). However, we did not observe any statistically significant association between CAG number and expression of IGFBP6.

237 **Table 3. Ratio between mean concentrations of 4HPPD and ALDOB in men with CAG repeat**
 238 **length < 21 or >22 as compared to the reference group**

<i>Proteins</i>	<i>Overall</i>	<i><21 vs reference</i>		<i>>22 vs reference</i>	
	<i>p-value</i>	<i>ratio (95% CI)</i>	<i>p value*</i>	<i>ratio (95% CI)</i>	<i>p value*</i>
4HPPD	0.012	1.34 (1.02 to 1.76)	0.032	1.62 (1.23 to 2.13)	0.001
ALDOB	0.008	1.35 (1.03 to 1.78)	0.029	1.72 (1.29 to 2.32)	<0.001

Ratio: ratio between mean concentration in <21 or >22 groups divided by the reference group (21 and 22).

** p-value of the post hoc constrain between groups*

Discussion

We have identified three plasma proteins, which are potential markers of BAA. In young healthy men the three markers ALDOB, 4HPPD and IGFB6 were strongly associated with T levels. In a slightly older cohort of infertile men, these markers were indicative of T deficiency i.e. both total and free serum T was low(5). Furthermore, levels of two of the markers, ALDOB and 4HPPD, were more strongly associated with risk of metabolic disturbances than total T. The association seemed to become stronger by creating a combined marker MMA, based on both 4HPPD and IGFBP6 levels. Finally, the androgen dependence of ALDOB and 4HPPD was confirmed by the association between the concentration of those proteins and the length of androgen receptor CAG repeats.

The ALDOB is a glycolytic enzyme, predominantly expressed in liver and kidney, that catalyzes the reversible cleavage of fructose-1,6-bisphosphate into glyceraldehyde 3-phosphate and dihydroxyacetone phosphate. The B isoform of aldolase, e.g. ALDOB, in the liver is under dietary control (40). Ingestion of fructose induces ALDOB mRNA expression in the liver, which is otherwise low in fasting conditions. In humans, absence of functional ALDOB enzyme due to mutations in the ALDOB gene cause hereditary fructose intolerance, characterized by metabolic disturbances that include hypoglycemia, lactic acidosis, and hypophosphatemia (41). An upregulation of ALDOB in human pancreatic β -cells occurs upon the development of hyperglycemia and may contribute to the impairment of insulin secretion in humans (42). In a study on goats, the ALDOB gene was found to be downregulated at the time of post-natal initiation of spermatogenesis (43). This finding is in accordance with our data showing that rising testosterone is inhibiting ALDOB.

Similar to ALDOB, we found that 4HPPD was negatively associated with T levels. This enzyme is involved in the catabolic pathway of tyrosine and catalyzes the conversion of 4-hydroxyphenylpyruvate to homogentisic acid in the tyrosine catabolism pathway (44). The expression of the gene is regulated by hepatocyte-specific and liver-enriched transcription factors, as well as by hormones (24). Tyrosine has previously been reported to be upregulated in hypogonadal men and both tyrosine and phenylalanine levels were suggested as predictors of the risk of developing diabetes many years before manifest disease (45–47). In male tyrosine hydroxylase knock-out mice normal body weight, puberty onset, and basal gonadotropin levels in adulthood were evident, although T was significantly elevated in adult mice (48).

The last marker, IGFBP6 is expressed in most tissues (24) and is one of the binding proteins for Insulin-Like Growth Factor (IGF). The principal function of IGFBP-6 is inhibiting IGF-II actions, whereby IGF-II induced cell proliferation, differentiation, migration and survival is reduced. Serum levels of IGFBP-6 increase gradually with age and are higher in men than in women, but there are conflicting studies of the direct effects of sex steroids on IGFBP-6 expression in different tissues (49). A positive association between the levels of T and IGFBP-6 have previously been found (50) (51). The latter study has a somewhat similar set up as the present study, based on chemical castration with a GnRH agonist, and has also focused on identifying novel markers of BAA, but with candidate markers previously identified as being associated with changes in fat-free mass. The study showed that early increases in IGFBP6 levels in men receiving testosterone were associated with increases in fat-free mass and muscle strength.

Altogether, our findings may not only be clinically valuable in developing new methods of

assessing BAA but also add to our understanding of the biological role of T in human metabolism, regulation of testicular function (ALDOB), as well as muscle strength and body composition (IGFBP6). However, our findings cannot be used as a direct proof of hypogonadism being cause of cardio-metabolic disease but the combined parameter MMA may in this context be an important tool in the detection of long-term morbidity, such as bone mineral density and cardio-metabolic risk, even before clinical diagnosis.

Our study has some strengths and limitations. Using a chemical castration model in young healthy men, we were able to identify proteins influenced by androgens and select those that were most strongly associated with T levels. By utilizing proteomics, we had an explorative approach to identify new markers of BAA without being restricted by previously published findings.

Another strength is the depth of the analysis due to the depletion abundant proteins from plasma. We were able to identify more than 450 proteins, which were identified in the same concentration range as 87 % of FDA-approved biomarkers (52). If depletion is not performed the detection level is dampened by the components from the digested abundant proteins as the proteins removed are of highest abundance in plasma and plasma proteome and exceed some lower abundance proteins by 10 orders of magnitude (53).

Although more clinical testing is needed, we have provided preliminary results showing that these protein markers may also be clinically useful. We have previously shown that median length CAG number is associated with most active AR (54). Thus, the fact that we find the lowest ALDOB and 4HPPD levels in those subjects having AR CAG repeat length close to

304 median, confirms that the candidate markers identified in the present study are androgen
305 dependent.

306 A limitation of our study is that the lack of reliable criteria for clinical hypogonadism, which
307 made it impossible to test the power of the new markers in men in whom androgen replacement
308 is needed. Furthermore, the clinical part of the study was limited, because we do not have
309 sufficient information about potential factors influencing the inter- and intra-individual variation
310 in the levels of these proteins and, thereby, their suitability as clinical markers. Furthermore, the
311 number of subjects included in the BL-T group was not sufficient to clarify whether, in this
312 testosterone concentration interval, the new markers can be useful in discriminating between
313 truly hypogonadal and men being eugonadal.

314 Furthermore, we are not reporting absolute values of quantifications for the potential markers,
315 but the relative quantifications for comparing the protein expression between groups. This kind
316 of comparative proteomics is favorable in research studies, in which preliminary results of
317 protein changes between groups are obtained. Also, the sample processing is complicated putting
318 high demands on the lab. In proteomics, there can be a high variation between labs in reporting
319 absolute concentration proteins in plasma, especially when small sample sizes are reported (55).
320 In order to obtain trustworthy absolute concentration ranges or determine the activity level of the
321 enzymes, it is necessary to analyze the potential markers in large cohorts including both healthy
322 subjects and patients.

323 In this study, we have applied an immunoassay for measuring T levels, although some consider
324 liquid chromatography-tandem mass spectrometry (LC-MS/MS) as gold standard in assessment

of sex hormone levels. However, worldwide the former is most commonly used for T measurements. Additionally, in the concentration range seen in males, there seems to be high correlation between concentration values obtained by immunoassay and by LC-MS/MS (56). Also in identifying men in hypogonadal T range and prediction of cardiometabolic risk, assessment of T by LC-MS/MS was not shown to be superior to that performed by standard methodology (56,57).

For this study, the z-score was employed because of the relatively young age of the subjects. Z-scores, a comparison of an individual's bone density with that of a healthy reference population (NHANES III) of the same age, sex and ethnicity and expressed as standard deviations, were obtained from the DXA-machine. In this study, we defined low BMD as Z-score below -1.0 . The rationale is based primarily on meta-analysis of 12 cohort studies demonstrating significantly increased risk of osteoporotic fractures for men at Z-scores ≤ -1 SD (58) and in addition because it has also been shown that the majority of fragility fractures occur in patients with BMD in the osteopenic range, i.e. T-score between -1 and -2.5 (59). Based on this information, Z-score below -1 can be assumed to imply an increased fracture risk.

In conclusion, we have identified three new potential biomarkers of BAA. Those proteins – alone or in combination – are promising as useful parameters in the clinical diagnosis of male hypogonadism and in the prediction of its long-term sequelae, as well as in studying the biology of androgen action. More extensive testing is vital to elucidate their BAA potential, not only in men but also in women and in prepubertal boys.

Data availability

346 The mass spectrometry proteomics data have been deposited to the ProteomeXchange
347 Consortium via the PRIDE (60) partner repository with the dataset identifier PXD024448.

348 Supplementary tables (datasets):

349 <https://doi.org/10.6084/m9.figshare.14875431>

350 Source data of the Figures can be found on:

351 <https://doi.org/10.6084/m9.figshare.14875431>

352 Supplementary Figure S1 (Figure 2—figure supplement 1):

353 <https://doi.org/10.6084/m9.figshare.14876562>

354 R code:

355 https://github.com/indirapla/TP1_proteins_marker_of_androgen_activity

356

357 **Declaration of conflict of interest:** The authors declare that there is no conflict of interest that
358 could be perceived as prejudicing the impartiality of the research reported.

359

360

Figure Legend

Figure 1. Study design. First, a model of 30 young healthy males was evaluated by proteomics at three time points (A, B and C), where testosterone changes were induced. Identified proteins proposed as candidate biomarkers were then evaluated in a cohort of infertile males. In both steps of the study, the quality of the blood samples was ensured by following an automated workflow for sample aliquoting and storage (-80°C).

Figure 2. Proteins influenced by testosterone in the model of young healthy males. **a)** Top 25 significant proteins selected in the healthy human model (ROC- $p < 0.01$, Table S3). The arrows indicate the direction of change in protein expression in the different conditions. The tissue with highest expression of each protein is indicated in colours. Also, results from the ROC analysis are shown as bar chart (AUC) and heat-map (p -values). **b)** Boxplot (mean (min; max)) of the top three significant proteins proposed as biomarker candidates, able to discriminate between low and normal testosterone (Table S4). The adjusted p -values are specified on top of the comparative horizontal lines. **c)** ROC of the analytes proposed as biomarker candidates, including Multi Marker Algorithm (MMA).

Figure 3. New markers to discern states of different testosterone levels in men investigated for infertility ($n=75$). **a)** Patients grouped by three levels of total testosterone: low testosterone (LT) ≤ 8 nmol/L ($n=22$), border line testosterone (BL_T) between 8 and 12 nmol/L ($n=17$) and normal testosterone (HT) > 12 nmol/L ($n=36$). Each group is represented by the mean and its 95% CI. Horizontal lines indicate significant differences between groups and the adjusted p -values are specified on top of these lines (Table S5). **b)** ROC analysis to discriminate patients with LT in the entire cohort and **c)** in a cohort that excluded patients with borderline testosterone

levels (LT_B). MMA is based on is the combination of levels of the proteins 4HPPD and IGFBP6. **d)** As a, but grouped according to the levels of calculated free testosterone (cFT): low (L_cFT) (n=21) – < 220 pmol/L and normal (N_cFT) (n=54) ≥ 220 pmol/L. **e)** As b and c but for discrimination of L_cFT and N_cFT.

Figure 4 Results from ROC analysis to determine whether the analytes discriminate between the presence of comorbidities or not. Analytes included in the analysis are 4HPPD, IGFBP6, ALDOB and MMA (combination of 4HPPD and IGFBP6). AUC, p-values can be found in Table 2. IR: insulin resistance; DM: type 2 diabetes mellitus; LBD: low bone density; CVRLP: cardiovascular risk lipid profile; MetS: metabolic syndrome.

Figure 5. Association between androgen receptor CAG <21 (n=26), and CAG <22 (n=30) and 4HPPD and ALDOB, respectively, with CAG=21 and 22 (n=18) set as reference.

References

1. Traish AM, Miner MM, Morgentaler A, Zitzmann M. Testosterone deficiency. Am J Med. 2011;124(7):578–87.
2. Zitzmann M. The Role of the CAG Repeat Androgen Receptor Polymorphism in Andrology. In: Advances in the Management of Testosterone Deficiency. Basel: KARGER; 2008. p. 52–61.
3. Furuya K, Yamamoto N, Ohyabu Y, Morikyu T, Ishige H, Albers M, et al. Mechanism of the tissue-specific action of the selective androgen receptor modulator s-101479. Biol

- 403 Pharm Bull. 2013 Mar;36(3):442–51.
- 404 4. Miner M, Barkin J, Rosenberg MT. Testosterone deficiency: myth, facts, and controversy.
405 Can J Urol. 2014 Jun;21 Suppl 2(3):39–54.
- 406 5. Arver S, Lehtihet M. Current guidelines for the diagnosis of testosterone deficiency. Vol.
407 37, Frontiers of Hormone Research. Front Horm Res; 2009. p. 5–20.
- 408 6. Diaz-Arjonilla M, Schwarcz M, Swerdloff RS, Wang C. Obesity, low testosterone levels
409 and erectile dysfunction. Vol. 21, International Journal of Impotence Research. Int J Impot
410 Res; 2009. p. 89–98.
- 411 7. Muraleedharan V, Jones TH. Testosterone and mortality. Clin Endocrinol (Oxf).
412 2014;81(4):477–87.
- 413 8. Sahlin KB, Pla I, Sanchez A, Pawłowski K, Leijonhufvud I, Appelqvist R, et al. Short-
414 term effect of pharmacologically induced alterations in testosterone levels on common
415 blood biomarkers in a controlled healthy human model. Scand J Clin Lab Invest. 2019
416 Nov 18;80(1):1–7.
- 417 9. Pla I, Sahlin KB, Pawłowski K, Appelqvist R, Marko-Varga G, Sanchez A, et al. A pilot
418 proteomic study reveals different protein profiles related to testosterone and gonadotropin
419 changes in a short-term controlled healthy human cohort. J Proteomics. 2020 May
420 30;220:1–3.
- 421 10. Bobjer J, Bogefors K, Isaksson S, Leijonhufvud I, Åkesson K, Giwercman YL, et al. High

prevalence of hypogonadism and associated impaired metabolic and bone mineral status in subfertile men. Clin Endocrinol (Oxf). 2016 Aug;85(2):189–95.

11. Wickramasinghe VP, Arambepola C, Bandara P, Abeysekera M, Kuruppu S, Dilshan P, et al. Insulin resistance in a cohort of 5-15 year old children in urban Sri Lanka. BMC Res Notes. 2017 Jul;10(1).

12. Walldius G, Jungner I. Apolipoprotein B and apolipoprotein A-I: Risk indicators of coronary heart disease and targets for lipid-modifying therapy. Vol. 255, Journal of Internal Medicine. 2004. p. 188–205.

13. Standards of Medical Care in Diabetes-2010. 2010;

14. Isaksson S, Bogefors K, Åkesson K, Egund L, Bobjer J, Leijonhufvud I, et al. Risk of low bone mineral density in testicular germ cell cancer survivors: association with hypogonadism and treatment modality. Andrology. 2017;5(5):898–904.

15. Lundin KB, Giwercman A, Richthoff J, Abrahamsson PA, Giwercman YL. No association between mutations in the human androgen receptor GGN repeat and inter-sex conditions. Mol Hum Reprod. 2003;9(7–8):375–9.

16. Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, et al. Measurement of protein using bicinchoninic acid. Anal Biochem. 1985;150(1):76–85.

17. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics

441 experiments. *Bioinformatics*. 2010 Apr 1;26(7):966–8.

442 18. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus
443 computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*.
444 2016 Sep 27;13(9):731–40.

445 19. RStudio Team. *RStudio: Integrated Development for R*. RStudio. Boston; 2016.

446 20. Team RC. *R: A language and environment for statistical computing*. R Foundation for
447 Statistical Computing. 2016.

448 21. Marshall KW, Mohr S, Khettabi F El, Nossova N, Chao S, Bao W, et al. A blood-based
449 biomarker panel for stratifying current risk for colorectal cancer. *Int J Cancer*. 2009
450 Mar;126(5):1177–86.

451 22. Bowers AJ, Zhou X. Receiver Operating Characteristic (ROC) Area Under the Curve
452 (AUC): A Diagnostic Measure for Evaluating the Accuracy of Predictors of Education
453 Outcomes. *J Educ Students Placed Risk*. 2019 Jan;24(1):20–46.

454 23. Šimundić A-MM. Measures of Diagnostic Accuracy: Basic Definitions. *Ejifcc*. 2009
455 Jan;19(4):203–11.

456 24. Kim M-SS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, et al. A draft map
457 of the human proteome. *Nature*. 2014 May 29;509(7502):575–81.

458 25. Kampf C, Mardinoglu A, Fagerberg L, Hallström BM, Edlund K, Lundberg E, et al. The
459 human liver-specific proteome defined by transcriptomics and antibody-based profiling.

- 460 2014 Jul 1;28(7):2901–14.
- 461 26. Kholodenko I V, Yarygin KN. Cellular Mechanisms of Liver Regeneration and Cell-
462 Based Therapies of Liver Diseases. Biomed Res Int. 2017;v.2017:17 pages.
- 463 27. Schmucker DL, Sanchez H. Liver Regeneration and Aging: A Current Perspective. Curr
464 Gerontol Geriatr Res. 2011;2011.
- 465 28. Benjamini Y, Krieger AM, Yekutieli D. Adaptive linear step-up procedures that control
466 the false discovery rate. Biometrika. 2006 Sep 1;93(3):491–507.
- 467 29. Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the
468 estimation of free testosterone in serum. J Clin Endocrinol Metab. 1999;84(10):3666–72.
- 469 30. Antonio L, Wu FCW, O'Neill TW, Pye SR, Ahern TB, Laurent MR, et al. Low Free
470 Testosterone Is Associated with Hypogonadal Signs and Symptoms in Men with Normal
471 Total Testosterone. J Clin Endocrinol Metab. 2016 Jul 1;101(7):2647–57.
- 472 31. Chan I, Tang Fui MN, Zajac JD, Grossmann M. Assessment and management of male
473 androgen disorders: An update. Aust Fam Physician. 2014;43(5):277–82.
- 474 32. Lunenfeld B, Arver S, Moncada I, Rees DA, Schulte HM. How to help the aging male?
475 Current approaches to hypogonadism in primary care. Vol. 15, Aging Male. Aging Male;
476 2012. p. 187–97.
- 477 33. Zitzmann M, Faber S, Nieschlag E. Association of specific symptoms and metabolic risks
478 with serum testosterone in older men. J Clin Endocrinol Metab. 2006;91(11):4335–43.

- 479 34. Casella R, Maduro MR, Lipshultz LI, Lamb DJ. Significance of the polyglutamine tract
480 polymorphism in the androgen receptor. *Urology*. 2001;58(5):651–6.
- 481 35. Stanworth RD, Kapoor D, Channer KS, Jones TH. Androgen receptor CAG repeat
482 polymorphism is associated with serum testosterone levels, obesity and serum leptin in
483 men with type 2 diabetes. *Eur J Endocrinol*. 2008;159(6):739–46.
- 484 36. Kim JW, Bae YD, Ahn ST, Kim JW, Kim JJ, Moon DG. Androgen Receptor CAG Repeat
485 Length as a Risk Factor of Late-Onset Hypogonadism in a Korean Male Population. *Sex*
486 *Med*. 2018 Sep 1;6(3):203–9.
- 487 37. Ferlin A, Bartoloni L, Rizzo G, Roverato A, Garolla A, Foresta C. Androgen receptor
488 gene CAG and GGC repeat lengths in idiopathic male infertility. *Mol Hum Reprod*.
489 2004;10(6):417–21.
- 490 38. Nenonen H, Björk C, Skjaerpe PA, Giwercman A, Rylander L, Svartberg J, et al. CAG
491 repeat number is not inversely associated with androgen receptor activity in vitro. *Mol*
492 *Hum Reprod*. 2010 Nov 1;16(3):153–7.
- 493 39. Nenonen HA, Giwercman A, Hallengren E, Giwercman YL. Non-linear association
494 between androgen receptor CAG repeat length and risk of male subfertility – a meta-
495 analysis. *Int J Androl*. 2011 Aug 1;34(4pt1):327–32.
- 496 40. Munnich A, Besmond C, Darquy S, Reach G, Vaulont S, Dreyfus JC, et al. Dietary and
497 hormonal regulation of aldolase B gene expression. *J Clin Invest*. 1985;75(3):1045–52.

- 498 41. Hannou SA, Haslam DE, McKeown NM, Herman MA. Fructose metabolism and
499 metabolic disease. Vol. 128, Journal of Clinical Investigation. American Society for
500 Clinical Investigation; 2018. p. 545–55.
- 501 42. Gerst F, Jaghutriz BA, Staiger H, Schulte AM, Lorza-Gil E, Kaiser G, et al. The
502 expression of aldolase B in islets is negatively associated with insulin secretion in humans.
503 J Clin Endocrinol Metab. 2018;103(12):4373–83.
- 504 43. Bo D, Jiang X, Liu G, Xu F, Hu R, Wassie T, et al. Multipathway synergy promotes
505 testicular transition from growth to spermatogenesis in early-puberty goats. BMC
506 Genomics. 2020 May 25;21(1).
- 507 44. Hager SE, Gregerman RI, Kox WE. p-Hydroxyphenylpyruvate oxidase of liver. J Biol
508 Chem. 1957;225(2):935–47.
- 509 45. Wang TJ, Larson MG, Vasan RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles
510 and the risk of developing diabetes. Nat Med. 2011 Apr;17(4):448–53.
- 511 46. Fanelli G, Gevi F, Belardo A, Zolla L. Metabolic patterns in insulin-sensitive male
512 hypogonadism article /101/58. Cell Death Dis. 2018 Jun 1;9(6).
- 513 47. Guasch-Ferré M, Hruby A, Toledo E, Clish CB, Martínez-González MA, Salas-Salvadó J,
514 et al. Metabolomics in prediabetes and diabetes: A systematic review and meta-analysis.
515 Diabetes Care. 2016.
- 516 48. Stephens SBZ, Rouse ML, Tolson KP, Liaw RB, Parra RA, Chahal N, et al. Effects of

517 selective deletion of tyrosine hydroxylase from kisspeptin cells on puberty and
518 reproduction in male and female mice. *eNeuro*. 2017;

519 49. Bach LA, Fu P, Yang Z. Insulin-like growth factor-binding protein-6 and cancer. *Clinical*
520 *Science*. 2013.

521 50. Rooman RPA, Op De Beeck L, Martin M, van Doorn J, Mohan S, Du Caju MVL.
522 Ethinylestradiol and testosterone have divergent effects on circulating IGF system
523 components in adolescents with constitutional tall stature. *Eur J Endocrinol*. 2005
524 Apr;152(4):597–604.

525 51. Huang G, Rocha G V., Pencina KM, Cox K, Krishnan V, Henriksen K, et al. Circulating
526 Biomarkers of Testosterone's Anabolic Effects on Fat-Free Mass. *J Clin Endocrinol*
527 *Metab*. 2019 Apr 11;104(9):3768–78.

528 52. Anderson NL. The Clinical Plasma Proteome: A Survey of Clinical Assays for Proteins in
529 Plasma and Serum. *Clin Chem*. 2010 Feb;56(2):177–85.

530 53. Tu C, Rudnick PA, Martinez MY, Cheek KL, Stein SE, Slebos RJC, et al. Depletion of
531 abundant plasma proteins and limitations of plasma proteomics. *J Proteome Res*. 2010
532 Oct;9(10):4982–91.

533 54. Nenonen H, Björk C, Skjaerpe PA, Giwercman A, Rylander L, Svartberg J, et al. CAG
534 repeat number is not inversely associated with androgen receptor activity in vitro. *Mol*
535 *Hum Reprod*. 2009 Nov;16(3):153–7.

- 536 55. Nanjappa V, Thomas JK, Marimuthu A, Muthusamy B, Radhakrishnan A, Sharma R, et
537 al. Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic*
538 *Acids Res.* 2014 Jan;42(D1):D959–65.
- 539 56. Huhtaniemi IT, Tajar A, Lee DM, O'Neill TW, Finn JD, Bartfai G, et al. Comparison of
540 serum testosterone and estradiol measurements in 3174 European men using platform
541 immunoassay and mass spectrometry; relevance for the diagnostics in aging men. *Eur J*
542 *Endocrinol.* 2012 Jun;166(6):983–91.
- 543 57. Haring R, Baumeister SE, Nauck M, Volzke H, Keevil BG, Brabant G, et al. Testosterone
544 and cardiometabolic risk in the general population - the impact of measurement method on
545 risk associations: A comparative study between immunoassay and mass spectrometry. *Eur*
546 *J Endocrinol.* 2013 Oct;169(4):463–70.
- 547 58. Johnell O, Kanis JA, Oden A, Johansson H, De Laet C, Delmas P, et al. Predictive Value
548 of BMD for Hip and Other Fractures. *J Bone Miner Res.* 2005 Jul 1;20(7):1185–94.
- 549 59. Unnanuntana A, Gladnick BP, Donnelly E, Lane JM. The assessment of fracture risk. *J*
550 *Bone Joint Surg Am.* 2010 Mar 1;92(3):743–53.
- 551 60. Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, et al.
552 The PRIDE database and related tools and resources in 2019: improving support for
553 quantification data. *Nucleic Acids Res.* 2019 Jan 8;47(D1):D442–50.

554

APPENDIX 1. SUPPLEMENTARY DESCRIPTION OF METHODS

Novel protein markers of androgen activity in humans: proteomic study of plasma from young chemically castrated men.

Aleksander Giwercman^{1*}, K Barbara Sahlin^{2,3*}, Indira Pla^{2,3}, Krzysztof Pawlowski^{2,4,6}, Carl Fehniger^{2,3}, Yvonne Lundberg Giwercman⁵, Irene Leijonhufvud¹, Roger Appelqvist^{2,3}, György Marko-Varga^{3,7}, Aniel Sanchez^{2,3†}, and Johan Malm^{2,3†}

Clinical chemistry parameters, bone mineral density and CAG-repeat length

Laboratory tests and dual X-ray absorptiometry

The analysis of the clinical chemistry parameters as well as bone mineral density were conducted as previously described (10) in the study on infertile men. Markers measured in fasting morning plasma samples included: total T, luteinizing hormone (LH), follicle stimulating hormone (FSH), Sex hormone-binding globulin (SHGB), estradiol (E2), glucose, insulin, apolipoprotein A1, and apolipoprotein B. The parameters were analyzed at the Department of clinical chemistry, Skåne University Hospital, Malmö, Sweden. Total testosterone, LH, FSH and SHGB were measured by electrochemiluminescence immunoassay (ECLIA) and an immunofluorometric method (DELFI Estradiol, Wallac OY, Finland) was utilized to measure E2. Glucose was measured by an automated hexokinase method and insulin was measured by immunometric sandwich assay. Bone mineral density (BMD, g/cm²) at total hip (TH) and lumbar spine L1–L4 (LS) was measured using dual X-ray absorptiometry.

CAG-repeat length

Androgen receptor CAG-repeat length was determined in all the infertile men. For the

analysis, DNA was extracted from peripheral leukocytes. The androgen receptor CAG stretch was amplified by polymerase chain reaction (PCR), purified and directly sequenced in a Beckman Coulter CEQ 2000 XL (Beckman Coulter, Bromma, Sweden) as previously described (15).

Proteomics experiments

All chemical reagents for the proteomic analysis not specified in the text were purchased from Sigma Aldrich (St. Louis, MO). Modified porcine trypsin was from Promega (Madison, WI), water from Milli-Q Ultrapure Water system (Millipore, Billerica, MA). Water and organic solvents for LC-MS were of LC-MS quality (Merck, Darmstadt, Germany).

The plasma and serum samples from respective cohort (healthy men and infertile patients) were randomized regarding the order before starting sample preparation. Quantitation of total protein content in the samples was performed using the bicinchoninic acid (BCA) assay (16). Following the instructions provided by the manufacturer, the top seven most-abundant proteins (albumin, IgG, IgA, transferrin, haptoglobin, antitrypsin and fibrinogen) were depleted from approximately 10 μ L of each sample using a MARS7 column (Agilent, Santa Clara, CA). The buffer was exchanged to 1.6% SDC, 50 mM NH_4HCO_3 with Amicon ultra-centrifugal filters (0.5 mL, 10 kDa cut-off, Millipore, Tullagreen, Ireland). The disulphide bonds were reduced by adding DTT to a final concentration of 10 mM and incubated 1 h at 37 $^\circ\text{C}$. The free thiol groups were alkylated by adding iodoacetamide to a final concentration of 25 mM, and the reaction proceeded for an additional 30 min at room temperature in darkness (reduction and alkylation occurred in the Amicon filter). The buffer was exchanged to 50 mM NH_4HCO_3 and the samples were resuspended in 100 μ L 50mM NH_4HCO_3 (30 μ g protein after BCA quantitation) and digested with trypsin at an enzyme-to-substrate mass ratio of 1:30 (w/w) for 16 h at 37 $^\circ\text{C}$. The remaining SDC was precipitated by adding 20% formic acid prior

to filtering the samples through a polypropylene filter plate with a hydrophilic PVDF membrane (mean pore size 0.45 μm , Porvair Filtration Group, Fareham, UK).

A Q-Exactive Plus mass spectrometer connected to an Easy-nLC 1000 liquid chromatography system (Thermo Scientific, San José, CA, USA) was used to analyse the samples for automated Data Dependent Acquisition (DDA) methods. Peptides (2 μL , 1 μg on the column) were loaded onto an Acclaim PepMap 100 precolumn (75 $\mu\text{m} \times 2 \text{ cm}$, Thermo Scientific, San José, CA, USA) and separated on an easy-Spray column (25 $\text{cm} \times 75 \mu\text{m i.d.}$, PepMap C18 2 μm , 100 $\text{\AA}$) at a flow rate of 300 nL/min and a column temperature of 35°C. Solvent A (0.1% formic acid) and solvent B (0.1% formic acid in acetonitrile) were used to create a nonlinear gradient to elute the peptides (90 min).

The MS1 spectra of the peptides were acquired in the orbitrap mass analyser from m/z 400–1600 and a resolution of 70,000 (at m/z 200). The target automated gain control (AGC) value and maximum injection time (IT) were 1e6 and 100 ms, respectively. The ten most intense peaks with charge state ≥ 2 were fragmented by higher-energy collisional dissociation (HCD) with 26% normalised collision energy. Tandem mass spectra were acquired in the orbitrap mass analyser at a resolution of 35,000 (at m/z 200), a target AGC value of 5e4 and a maximum IT of 100 ms. The underfill ratio was 10% and dynamic exclusion was 45 s. The normalised collision energy was 26%.

Selected proteins from the healthy individuals were monitored also in patients by Multiple Reaction Monitoring (MRM) method. A TSQ Quantiva mass spectrometer equipped with an EASY-Spray NG ion source and connected to an EASY n-LC 1000 liquid chromatography system (Thermo Scientific, San José, CA, USA) was used for the MRM method. Peptides (5 μL , 1 μg on the column) were loaded onto an Acclaim PepMap 100 precolumn (100 $\mu\text{m} \times 2 \text{ cm}$, Thermo Scientific, San José, CA, USA) and separated on an easy-Spray column (15 $\text{cm} \times$

75 μm i.d., PepMap C18 3 μm , 100 $\text{\AA}$) at a flow rate of 300 nL/min and a column temperature of 35°C. Solvent A (0.1% formic acid) and solvent B (0.1% formic acid in acetonitrile) were used to create a nonlinear gradient to elute the peptides (60 min). Internal references, isotopically labelled synthetic peptides (UVic-Genome BC Proteomics Centre, BC, Canada) were added in order to determine relative quantifications of proteins. Data was acquired in scheduled MRM mode with 5 min detection windows. SRM transitions were acquired in Q1 and Q3 operated at unit resolution (0.7 full width at half maximum, FWHM), the collision gas pressure in Q2 was 1.5 mTorr. The cycle time was 2 s, and calibrated radio frequency (RF) and S-lens values were used. At least three transition per precursor were monitored and in most cases the ratio is reported between the relative total intensities vs the references.

For peptide and protein identification the Proteome Discoverer v2.2 (PD) software (Thermo Scientific, San José, CA) was used. Peptides were identified using SEQUEST HT against UniProt (<http://www.uniprot.org>) human database (Human 9606, Reviewed, 20 165) integrated into PD. The search was performed with the following parameters: cysteine carbamidomethylation as a static modification, oxidation of methionine as a dynamic modification, 10 ppm precursor tolerance and 0.02 Da fragment tolerance. Up to one missed cleavage for tryptic peptides was allowed. According to the PD software, the filters applied were high- (FDR<0.01) and medium confidence (FDR<0.05) at peptide- and protein levels, respectively. The peptide and protein quantifications were based on the quantification of MS peptide signals (label-free quantification). Label-free quantification used the Minora Feature Detector node in the processing workflow, and the Precursor Ions Quantifier node and the Feature Mapper in the consensus workflow.

To analyse the MRM data, all MS files were imported into Skyline v3.5 (17) (MacCoss Lab Software, Seattle, WA). The peak integration was performed automatically by the software,

using Savitzky-Golay smoothing and to confirm the correct peak detection it was in addition, manually inspected. The relative total intensities vs the references (ratios), were reported as expression values. Only the values that passed the inspected intensity selection criteria according to PD or in Skyline were reported.

Supplementary Statistical analyses

Identification of candidate marker proteins from chemically castrated men

Proteins identified by mass spectrometry techniques in samples from healthy subjects were pre-processed in Perseus v1.6.7.0 (18) software. Intensities were Log2 transformed to follow a normal distribution and standardized by subtracting the median of its respective sample from each value. The statistical analyses were performed using R software (19,20) and SPSS Statistics 21.0 (IBM, Somers, IL, USA). Differentially expressed proteins were determined by doing one-way repeated measures ANOVA (R function: ezANOVA{ez}) to reveal overall differences between conditions (normal (A), low (B) and restored (C) T) and differences between individual conditions were detected by performing a post-hoc test based on pairwise t-test (two-tails and paired) (R function: pairwise.t.test{stats}). The 'pairwise.t.test' function utilized the proteins with significant overall changes (ANOVA p -value < 0.05) to perform pairwise comparisons between conditions while corrected for multiple pairwise testing. Proteins with adjusted p -values (' fdr ' method) < 0.05 following the pairwise t-test were considered significant.

Because gonadotropins (FSH, LH) did not change their levels between conditions B and C, while T did recover its value (8), proteins that changed with statistical significance between A-B and recovered their values significantly between B and C were considered mainly influenced by T changes and, hence, were selected for further analyses.

Based on the outcome of the first selection step of the analysis, we performed receiver operating characteristic (ROC) analysis to select proteins capable to discriminate between normal and low T. Two categories of samples were created: 'low T' that included samples evaluated at time point B and 'normal T' that included samples evaluated at time point A and C. Although there is no consensus about the threshold for sufficient diagnostic accuracy, the area under the curve (AUC) can be interpreted as the following rough classifying system: > 0.90 = excellent; $0.80 - 0.90$ = very good; $> 0.70 - 0.80$ = good; $0.60 - 0.70$ = sufficient and $0.50 - 0.60$ = poor (21-23). Consequently, based on the ROC-AUC values, we created two cut-off values to select the final protein biomarkers: a) $AUC > 0.80$ (i.e. very good) and b) $AUC 0.75 - 0.80$ (i.e. good) and highly enriched in liver tissues (according to the Human Proteome Map (24) and Kampf C. et. al. (25)), which, unlike other organs, appears to be relatively protected from age-related changes (26,27). The Human Proteome Map tool (24) was used to determine the tissue-specific localization of the proteins.

Because the combination of different markers may improve the discriminative power to diagnose hypogonadism and predict its long term sequelae, proteins selected from the ROC-AUC analysis were included as predictors in a stepwise regression (method: backward) to select the best combination of markers that predict the odds of being low T. Bootstrap resampling with replacement method was applied to assess consistency of predictors selected with the stepwise regression. A new variable called Multi Marker Algorithm (MMA) was derived from the predicted log-odds (of being low T) obtained from a binomial logistic regression analysis, in which the dichotomized variable 'testosterone level' (0: normal T (time points A, C) and 1: low T (time point B)) was used as the dependent variable and the intensities of the selected proteins from the stepwise regression were the predictors' variables. Finally, MMA variable was evaluated together with marker candidates proteins. Subsequently, an ROC analysis was performed based on the predicted values. All analyses

described in this part were performed in R software and plotted using ggroc{pROC} function.

Background characteristics of the infertile cohort of patients

To describe background characteristics of the infertile cohort of patients (**Table 1a**), the mean and standard deviation (SD) was calculated for variables that followed a normal distribution or the median and minimum-maximum (min-max) for variables that were not normally distributed (non-Gaussian). The normal distribution was checked by applying a Kolmogorov-Smirnov test.

Testing of the candidate biomarkers in infertile men

To further perform statistical analyses, the quantified intensities of the candidate biomarkers in infertile men were Log2 transformed to achieve normal distributions. Markers selected from the stepwise regression in the healthy human model were also considered in this cohort to create the MMA variable. In this case, MMA was created to predict the odds of suffering low T or other medical conditions associated with low T levels. Unless other software is specified, the statistical analyzes described in the next steps were performed using R software (19,20).

Distinguishing infertile men with different testosterone levels

In order to know if the change in the proteins occur with the change in T as observed in the healthy human model, we created three groups of patients based on total T concentration values commonly described in clinical guidelines (5) to define T deficiency: Group 1 contains patients commonly defined with low T (LT): ≤ 8 nmol/L (n = 22), Group 2 contains patients defined with borderline low T (BL_T): between 8 and 12 nmol/L (n=17), and Group 3 contains patients with normal T (NT): > 12 nmol/L (n = 36). Differences among the three

groups of patients were detected by doing a one-way ANOVA per protein (outcome variable: protein intensity) followed by a pairwise analysis that was corrected to control the FDR provoked by multiple pairwise comparisons (Two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli (28), performed in GraphPad Prism software, version 9.00 for Windows, CA, USA). Adjusted p-values < 0.05 were considered significant. Calculated free testosterone (cFT) was determined according to the method described by Vermeulen *et al.*(29). The cut-off level of 220 pmol/L was used to categorize the subjects as having low cFT (L_cFT; n=21) or normal cFT (N_cFT; n=54) (30). Changes between the two groups were evaluated by two-tailed Student's T-test (*p*-values < 0.05 were considered significant).

An ROC analysis was performed to discriminate the patients with LT from those with BL_T or NT. The same analysis was repeated after excluding the BL_T patients as this group includes males in the gray zone for hypogonadism diagnosis (31-33). ROC analysis was also used to discriminate between L_cFT and N_cFT.

Distinguishing between men with abnormal cardio-metabolic parameters or reduced bone mineral density

In order to evaluate the diagnostic value of our new protein markers in relation to conditions associated with low T levels, we performed a ROC analysis to discriminate patients with MetS, IR, CVRLP, DM or LBD within the patient cohort using the expression level of the candidate biomarkers and T. Based on the ROC curves, the DeLong's test (paired) was applied in R (roc.test {pROC} function) to compare the AUCs of the candidate biomarkers vs the AUC of T levels to discriminate the above mentioned pathological conditions.

Association with Androgen receptor CAG-repeat length

To investigate the androgenic dependence of the levels of the newly defined protein markers, we looked for associations between CAG repeat length and concentration of those markers. The infertile cohort was separated into three groups based on CAG repeat length of the androgen receptor (AR); a group of patients with CAG repeat length 21 and 22 (n=18) was defined as reference group (34-37). Patients with CAG repeat length less than the reference group (< 21) were included in a group (n=26) and patients with CAG repeat length longer than the reference group (>22) made up the third group (n=30). Using the software GraphPad Prism version 9.0.0 for Windows (CA, USA), a one-way ANOVA followed by a pairwise test (two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli (28)) which controls FDR for multiple pairwise comparison was conducted and adjusted p-values < 0.05 were considered significant.

198

HEALTHY YOUNG MEN (n=30)

MALE INFERTILE PATIENTS (n=75)

TESTING CANDIDATE PROTEINS IN PATIENTS

BIOBANK PROCESSING

Candidate markers

- 4HPPD
- ALDOB
- IGFBP6
- Total_testosterone
- MMA(4HPPD+IGFBP6)

4HPPD

ALDOB