

## **Circulating plasma proteins and new-onset diabetes in a population-based study: proteomic and genomic insights from the STANISLAS cohort**

Joao Pedro Ferreira, Zohra Lamiral, Constance Xhaard, Kévin Duarte, Emmanuel Bresso, Marie-Dominique Devignes, Edith Le Floch, Claire Dandine Roulland, Jean-François Deleuze, Sandra Wagner, et al.

► **To cite this version:**

Joao Pedro Ferreira, Zohra Lamiral, Constance Xhaard, Kévin Duarte, Emmanuel Bresso, et al.. Circulating plasma proteins and new-onset diabetes in a population-based study: proteomic and genomic insights from the STANISLAS cohort. *European Journal of Endocrinology, BioScientifica*, 2020, 183 (3), pp.285-295. 10.1530/EJE-20-0246 . hal-02917113

**HAL Id: hal-02917113**

**<https://hal.univ-lorraine.fr/hal-02917113>**

Submitted on 18 Aug 2020

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

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

1 **Circulating plasma proteins and new-onset diabetes in a population-based study: proteomic and**  
2 **genomic insights from the STANISLAS cohort**

3

4 João Pedro Ferreira, MD, PhD<sup>1</sup>; Lamiral Zohra, MSc<sup>1</sup>; Constance Xhaard, PhD<sup>1</sup>; Kévin Duarte, PhD<sup>1</sup>;  
5 Emmanuel Bresso, MSc<sup>2</sup>; Marie-Dominique Devignes, PhD<sup>2</sup>; Edith Lefloch, PhD<sup>3</sup>; Claire Dandine  
6 Roulland, PhD<sup>3</sup>; Jean-François Deleuze, MD, PhD<sup>3</sup>; Sandra Wagner, PhD<sup>1</sup>; Bruno Guerci, MD, PhD<sup>4</sup>;  
7 Nicolas Girerd, MD, PhD<sup>1</sup>; Faiez Zannad, MD, PhD<sup>1</sup>; Jean-Marc Boivin, MD, PhD<sup>1</sup>; Patrick  
8 Rossignol, MD, PhD<sup>1</sup>

9

10 **Affiliations:**

11 <sup>1</sup> Université de Lorraine, INSERM, Centre d'Investigations Cliniques Plurithématique 1433, INSERM  
12 1116, CHRU de Nancy, FCRIN INI-CRCT, Nancy, France.

13 <sup>2</sup> Université de Lorraine, CNRS, Inria, LORIA, Nancy, 54500, France.

14 <sup>3</sup> Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, 91057, Evry,  
15 France.

16 <sup>4</sup> Department of Endocrinology, CHRU de Nancy, Nancy, France

17

18 Word-count: 3027 words (excluding abstract and references)

19

20 **Contact to:**

21 Dr João Pedro Ferreira

22 Centre d'Investigation Clinique 1433 module Plurithématique, CHRU Nancy - Hopitaux de Brabois,

23 Institut Lorrain du Coeur et des Vaisseaux Louis Mathieu

24 4 rue du Morvan, 54500 Vandoeuvre les Nancy

25 Tel : +33 (0) 3 83 15 73 15

26 Fax : +33 (0) 3 83 15 73 24

27 Mail: j.ferreira@chru-nancy.fr

28

29 **Abstract**

30 *Objective:* Determining the factors associated with new-onset pre-diabetes and type 2 diabetes  
31 mellitus (T2D) is important for improving the current prevention strategies and for a better  
32 understanding of the disease.

33 *Design:* To study the factors (clinical, circulating protein and genetic) associated with new onset pre-  
34 diabetes and T2D in an initially healthy (without diabetes) populational familial cohort with a long  
35 follow-up (STANISLAS cohort).

36 *Methods:* 1506 participants that attended both the visit-1 and visit-4, separated by  $\approx 20$  years. Over 400  
37 proteins, GWAS and genetic associations were studied using models adjusted for potential  
38 confounders. Both prospective (V1 to V4) and cross-sectional (V4) analyses were performed.

39 *Results:* People who developed pre-diabetes (n =555) and/or T2D (n =73) were older, had higher  
40 BMI, blood pressure, glucose, LDL cholesterol, and lower eGFR. After multivariable selection,  
41 PAPP-A (pappalysin-1) was the only circulating protein associated with the onset of both pre-diabetes  
42 and T2D with associations persisting at visit-4 (i.e.,  $\approx 20$  years later). FGF-21 (fibroblast growth factor  
43 21) was a strong prognosticator for incident T2D in the longitudinal, but not in the cross-sectional  
44 analysis. The heritability of the circulating PAPP-A was estimated at 44%. In GWAS analysis the  
45 SNP rs634737 was associated with PAPP-A both at V1 and V4. External replication also showed  
46 lower levels of PAPP-A in patients with T2D.

47 *Conclusions:* The risk of developing pre-diabetes and T2D increases with age, and with features of the  
48 metabolic syndrome. Circulating PAPP-A, which has an important genetic component, was associated  
49 with both the development and presence of pre-diabetes and T2D.

50

51 *Key-words:* diabetes mellitus; pre-diabetes; cohort study; pappalysin-1; genetic variants.

52

53

54

55

56

57

## 58 **Introduction**

59 The number of people with type 2 diabetes mellitus (T2D) more than doubled in the past 20 years in  
60 both the developed and developing countries<sup>1</sup>. The latest prediction from the International Diabetes  
61 Federation is that over 380 million people are living with diabetes nowadays, and that by 2035 this  
62 number may rise up to 600 million<sup>2</sup>. Type 2 diabetes was traditionally considered as a disorder of  
63 middle-aged and elderly people; however, the more recent “type 2 diabetes epidemics” is particularly  
64 worrisome among adolescents and young adults, leading to tremendous costs, impairment in the  
65 quality of life and reduced life expectancy, mainly due to associated cardiovascular and renal  
66 complications<sup>1, 3, 4</sup>.

67 Type 2 diabetes results from a complex interaction between genetic/epigenetic systems and  
68 multiple behavioural plus environmental factors<sup>1</sup>. Despite the challenges linked to the gene-  
69 environment interplay, determining the factors associated with new-onset pre-diabetes and T2D is  
70 important not only for improving the current prevention strategies and treatments, but also for a better  
71 understanding of the disease<sup>5, 6</sup>.

72 The STANISLAS (Suivi Temporaire Annuel Non-Invasif de la Santé des Lorrains Assurés  
73 Sociaux) cohort is a single-centre familial prospective cohort comprising 1006 initially healthy  
74 (including absence of diabetes) families (4295 participants) from the Nancy region of France, who  
75 were recruited from 1993 to 1995 to the visit 1 (V1). Participants were then followed each 5 to 10  
76 years, and from the initial visit, 1705 participants returned for the fourth visit (V4), held from 2011 to  
77 2016, which comprised a detailed clinical and laboratorial assessment<sup>7</sup>. Beyond the clinical, imaging  
78 and routine laboratorial measurements, circulating levels of multiple proteins were also assessed at V1  
79 and V4. Additionally, a genome wide association study (GWAS) was performed for the V4  
80 participants. This design allows the assessment of the clinical factors, genetic variants and proteomic  
81 signatures associated with the new-onset of pre-diabetes and/or T2D almost 20 years in advance; the  
82 family structure also allows the study of heritability<sup>8</sup>.

83 In the present analysis, we considered patients without diabetes at V1 (patients entering the  
84 familial STANISLAS study did not have chronic conditions) with the main aims of studying the

85 factors associated with new onset pre-diabetes and T2D at V4, considering the clinical features,  
86 proteomic signatures and genetic variants that might help explaining the main proteomic findings.

87

## 88 **Methods**

### 89 **Study population**

90 A detailed description of the STANISLAS cohort has been previously published<sup>7</sup>. The STANISLAS  
91 cohort was established with the primary objective of investigating gene-gene and gene-environment  
92 interactions in the field of cardiovascular diseases. In order to assess the effect of genetics on the  
93 variability of intermediate phenotypes on the transition toward pathology, the families were deemed  
94 healthy and free of declared acute and/or chronic illness at V1. The implementation of the fourth visit  
95 (V4) enabled a follow-up of 18 to 23 years (mean of 18.5 years). The collected information was  
96 enriched with new circulating proteins and detailed clinical phenotyping. The study protocols for all  
97 examinations were reviewed and approved by the Local Ethics Committee of CPP-Est 3, France. All  
98 participants provided written informed consent to participate in the study.

99 In the present analysis we included 1506 participants that attended both the V1 and V4, and in  
100 whom we had information on fasting glucose, random glucose, glycated hemoglobin, treatments  
101 and/or previous diagnosis that allow the pre-diabetes and T2D classification. The study flow-chart is  
102 shown in the **Figure 1**. The main objective is to assess the circulating proteins that are associated with  
103 the development of diabetes in a prospective analysis (V1 to V4) and to ascertain if the associations  
104 remained significant almost 20 years later at V4 in a cross-sectional analysis.

### 105 **Study design**

106 All participants were observed at the Centre d'Investigation Clinique Plurithématique Pierre Drouin at  
107 Nancy Hospital Center (CIC-P de Nancy) in the morning after a 12- to 14-hour fast. Blood samples  
108 were drawn, medical history, medications, anthropometric parameters, blood pressure, carotid-  
109 femoral pulse-wave velocity (PWV), and carotid intima-media thickness (cIMT) were recorded.

### 110 **Pre-diabetes and diabetes definition**

111 Pre-diabetes was defined as a fasting plasma glucose (FPG) between 100mg/dL and 125mg/dL or a  
112 glycated hemoglobin between 5.7% and 6.4%. Diabetes was defined as a FPG  $\geq$ 126mg/dL or a

113 glycated hemoglobin  $\geq 6.5\%$  or a random plasma glucose  $\geq 200\text{mg/dL}$  with symptoms of  
114 hypoglycemia or previous diagnosis of diabetes or anti-glycemic drug therapy<sup>9</sup>.

### 115 **Circulating proteins and gene-candidate analysis**

116 All samples were collected at the CIC-P de Nancy with minimally traumatic venipuncture.  
117 Standardized sample handling procedures enabled the collection of serum and plasma (EDTA,  
118 heparin) as well as buffy coat fraction. Blood DNA of all the participants to the Stanislas V4 was  
119 extracted using Gentra Puregene Blood Kit (Qiagen, Hilden Germany) and stored at  $-20^{\circ}\text{C}$ .  
120 Genotyping was conducted at CNRGH (Evry, France) using two chips: the Illumina Global Screening  
121 Array (GSA) which is composed of 687 572 intronic and exonic markers and the Illumina Exome  
122 Array, which is constituted of 244 330 SNPs, mostly exonic. All blood-derived bio-samples are stored  
123 in a central biobank facility with temperatures between  $-80^{\circ}\text{C}$  and  $-196^{\circ}\text{C}$  (as required e.g.  $-196^{\circ}\text{C}$  for  
124 plasma and serum).

125 Baseline plasma samples were analysed for proteins by the TATAA-biocenter using the Olink  
126 Proseek® Multiplex cardiovascular (CVD) II, III, cardiometabolic (CM), organ-damage (OD) and  
127 inflammatory (INF) panels, using a proximity extension assay (PEA) technology<sup>10</sup>, where 92  
128 oligonucleotide-labelled antibody probe pairs per panel (total of  $92 \times 5$ ;  $n = 460$  proteins) are allowed  
129 to bind to their respective targets in the sample in 96-well plate format. When binding to their correct  
130 targets, they give rise to new DNA amplicons with each ID-barcoding their respective antigens. The  
131 amplicons are subsequently quantified using a Fluidigm BioMark™ HD real-time PCR platform. The  
132 platform provides  $\log_2$ -normalized protein expression (NPX) data.

133 The strategy for the genetic analyses of the present study was the following. Firstly, we  
134 conducted GWAS on diabetes occurrence, fasting glycemia and glycated haemoglobin. Secondly,  
135 GWAS was conducted on the plasma protein(s) associated with both pre-diabetes and T2D in both  
136 analysis (i.e., longitudinal and cross-sectional). Thereafter, the GWAS top-SNP was tested for  
137 association with study outcome (i.e. pre-diabetes, T2D, fasting glycemia, glycated haemoglobin and  
138 circulating PAPP-A).

139 Our findings were replicated using the baseline data of the HOMAGE (Bioprofiling Response  
140 to Mineralocorticoid Receptor Antagonists for the Prevention of Heart Failure; ClinicalTrials.gov

141 Identifier: NCT02556450) trial, a prospective trial that included 527 people at risk for developing  
142 heart failure randomized to either spironolactone or standard of care for up to 9 months; the  
143 description of this study and baseline characteristics of the participants have been published  
144 elsewhere<sup>11</sup>.

#### 145 **Statistical, genetic and bioinformatical considerations**

146 For the baseline clinical characteristics, continuous variables are expressed as means and respective  
147 standard deviation (SD) and compared using ANOVA or Kruskal-Wallis test. Categorical variables  
148 are presented as frequencies and percentages and compared using Chi-2 test.

149 The main aim of this study was to test the association of multiple circulating proteins with  
150 incident pre-diabetes and T2D. Multinomial logistic regression models were performed (using no pre-  
151 diabetes/T2D as the referent category). Firstly, a stepwise backward model including all the clinical  
152 variables with a  $p < 0.1$  from Table 1 was performed, to select the clinical features with stronger  
153 association with pre-diabetes and T2D. Due to collinearity between body mass index and waist  
154 circumference (correlation  $> 0.9$  in both visits) we included only body mass index in the models  
155 (including waist circumference provided the same results). The clinical features retained in the final  
156 model were age, body mass index, LDL cholesterol and systolic blood pressure. These variables were  
157 used for adjustment in all the subsequent models. Secondly, the potential association of the multiple  
158 circulating proteins with pre-diabetes and T2D was tested 1-by-1 on top of the clinical model (i.e.  
159 adjusted for the clinical features) built in the previous step. The statistical significance level was fixed  
160 at a  $FDR_q < 0.05$ , after applying a Benjamini-Hochberg correction for multiple testing (i.e. false  
161 discovery rate [FDR] set at 5%). Subsequently, the proteins selected at the 5% FDR went through a  
162 multivariable stepwise forward selection model adjusted for the clinical variables. Since the  
163 circulating proteins were measured using NPX values on a  $\log_2$  scale, the odds ratio (OR) for each  
164 protein estimates the increase/decrease in the odds of pre-diabetes/T2D associated with a doubling in  
165 the protein concentration. As exploratory analysis, we have assessed the additive value of the  
166 biomarkers on top of the clinical model for the “prediction” of pre-diabetes and T2D onset (from V1  
167 to V4). These analyses were performed using SAS version 9.1.

168 The genetic analyses were performed using R (version 3.4.1). GWAS were run using a linear

169 mixed model, in order to take into account for the pedigree data, with age and sex used as covariates,  
170 the statistical significance level was fixed at  $10^{-7}$ . For heritability estimation, matrix of pedigree or  
171 genetic relationship were used in a linear mixed model to estimate the variance captured by additive  
172 genetic effects via average information restricted maximum likelihood analysis. To take into account  
173 the genetic information at both V1 and V4, we have added two random effects in the model (one for  
174 individuals and another for visits) in order to assess the variance component resulting from the  
175 repeated measures. Both models were implemented in the R package “gaston”. For the top-SNP  
176 association tests, age and sex were used as covariates and the statistical significance level was fixed at  
177 0.05.

178           The complex network analysis was performed using the FHF-GKB (Fight Heart Failure -  
179 Graph Knowledge Box) resource, representing most available public knowledge about human protein-  
180 protein and protein-pathway relationships is a customized upgraded version of the EdgeBox provided  
181 by the EdgeLeap company (available from: <https://www.edgeleap.com>). FHF-GKB data is extracted  
182 from public data sources resulting in 20,431 protein nodes imported from Uniprot<sup>13</sup>, including all  
183 proteins involved in this study and 2,272 pathway nodes from Reactome (v69)<sup>1</sup> and 47,162 disease  
184 nodes from Disease Ontology and Human Phenotype Ontology<sup>12</sup>. Protein - protein relationships were  
185 retrieved from STRING (v11)<sup>15</sup> or Reactome or WikiPathways<sup>16</sup> or Mentha<sup>17</sup> or BioGrid<sup>18</sup>, protein –  
186 pathway relationships from Reactome and protein – disease relationships from DisGeNET<sup>13</sup>. The  
187 FHF-GKB complex network was queried in order to explore pathways and proteins that connect the  
188 circulating proteins to disease through common mechanistic pathways. Queries were expressed  
189 according to patterns defining a path structure between two nodes, such as “protein-disease” and  
190 “pathway-protein-disease”. The resulting graph is displayed in figure illustrating the main connection  
191 between proteins, pathways, and disease (i.e. T2D [HP:0005978]).

192

## 193 **Results**

### 194 **Patient characteristics**

195 The patients` characteristics at V1 overall (n =1506) and by the pre-diabetes (n =555; 37%) and T2D  
196 (n =73; 4.8%) at V4 are depicted in the **Table 1**. Patients who developed pre-diabetes and/or T2D



197 were older, had higher body mass index (BMI), blood pressure, glucose levels, LDL cholesterol, and  
198 lower estimated glomerular filtration rate (eGFR).

### 199 **Clinical factors associated with pre-diabetes and type 2 diabetes mellitus**

200 The clinical factors independently associated with pre-diabetes onset were older age and higher LDL  
201 cholesterol, and with T2D were older age, higher BMI, higher systolic blood pressure, and higher  
202 LDL cholesterol (tendency). **Table 2.**

### 203 **Circulating proteins associated with diabetes**

204 The full list of proteins, their respective associations with pre-diabetes and T2D, and details on the  
205 Olink® technology are depicted in the **Supplemental Material.**

206 After multivariable selection adjusted for the clinical features, PAPP-A (pappalysin-1) was  
207 the only circulating protein that was associated with lower odds of both pre-diabetes and T2D onset in  
208 the V1 to V4 (prospective) analysis, and also with the presence of pre-diabetes and T2D at V4 (cross-  
209 sectional analysis). CALC-A (calcitonin related polypeptide alpha) and VCAM-1 (vascular cell  
210 adhesion molecule 1) were also associated with lower odds of developing pre-diabetes (but not  
211 diabetes) in the V1 to V4 prospective analysis, but the associations were no longer present at the V4  
212 cross-sectional analysis. On the other hand, FABP2 (fatty acid binding protein 2), LOX-1 (lysyl  
213 oxidase), and THPO (thrombopoietin) were associated with higher odds of developing pre-diabetes  
214 (but not diabetes) from V1 to V4, whereas FGF-21 (fibroblast growth factor 21), MCP-4 (mast-cell  
215 protease 4), SERPINA-12 (serpin family A member 12), and VSIG-2 (V-set and immunoglobulin  
216 domain containing 2) were associated with higher odds of developing diabetes from V1 to V4. None  
217 of these associations remained significant at V4. The details of the associations are fully depicted in  
218 the **Table 3.** In the external replication analysis, patients with diabetes prospectively recruited in the  
219 HOMAGE clinical trial also presented lower levels of PAPP-A. **Supplemental Table 1.**

220 In the exploratory “predictive” analysis, the clinical model for prediabetes presented an AUC  
221 of 0.74, adding PAPP-A to this model improved its predictive capacity to 0.75 (p-value for AUC  
222 comparison =0.003). The clinical model performed even better for diabetes with an AUC of 0.90,  
223 adding PAPP-A to this model did not improve its predictive capacity as the AUC remained at 0.90 (p-  
224 value for AUC comparison =0.32). However, adding FGF-21 (that had a strong association with new-

225 onset diabetes) improved this model to an AUC of 0.92 (p-value for AUC comparison =0.020).

226 Adding more biomarkers lead only to marginal improvements in the models.

### 227 **Genetic associations, heritability estimation, and complex network analyses**

228 No signal was found for GWAS with diabetes, fasting glycemia or glycated haemoglobin. For

229 circulating PAPP-A, the heritability was estimated at 44% for subjects with measurements at both

230 visits (V1 & V4) (**Supplemental Table 2**). Only one SNP, rs634737, was associated with the protein

231 PAPP-A both at V1 and V4 at the GWAS significance level. **Table 4**. Association tests between this

232 SNP and study outcome are shown in the **Table 5**, and suggests that subjects with mutation at

233 rs634737 are more susceptible to develop diabetes, have higher level of fasting glycemia and glycated

234 hemoglobin.

235 Network analysis results are presented in the **Figure 2**.

236

### 237 **Discussion**

238 Our study shows that the risk of developing pre-diabetes and T2D increases with age, and with

239 features of the metabolic syndrome. Furthermore, several circulating proteins were associated with

240 pre-diabetes and T2D; particularly PAPP-A was associated with the development of both pre-diabetes

241 and T2D (i.e. in initially healthy young adults without diabetes, at V1  $\approx$ 19 years before V4), and

242 remained associated with both pre-diabetes and T2D once the disease had been installed (at V4). In

243 this familial cohort, the PAPP-A protein has an important heritability ( $\approx$ 44%) and acts via proteins

244 and a pathway linked to IGFs. These data were externally replicated in the HOMAGE trial cohort, and

245 provide novel insights on the pathophysiology of T2D, generating hypothesis for future research in

246 the field.

247 Beyond the genetic susceptibility, that may be particularly important in specific ethnic

248 groups<sup>14-16</sup>, T2D is essentially caused by an interplay between environmental and behavioural factors,

249 that include sedentary lifestyle, nutritional habits rich in sugars, carbohydrates and processed foods,

250 and obesity<sup>17</sup>. The lack of robust genetic associations with diabetes in our study also supports this

251 notion. Importantly, it has been suggested that the risk of dysglycemia-associated complications

252 commences many years before the onset of clinical diabetes<sup>18</sup>. Our findings also support that the

253 metabolic alterations start many years before the onset of T2D and are also associated with pre-  
254 diabetes, moreover the proteomic profiles of these patients are altered<sup>17, 19-21</sup>. Landmark clinical trials  
255 have shown that lifestyle modification programs focused on weight loss and/or the use of metformin  
256 (although the lifestyle interventions aiming at significant weight loss were the most effective measure)  
257 can delay the onset of T2D<sup>22-24</sup>. Unfortunately, no study has shown that diabetes can be delayed or  
258 prevented in a community setting<sup>19</sup>. This should not demotivate us from continuing the efforts, and  
259 possibly aiming the preventive strategies to individuals at higher risk for developing the disease may  
260 be more effective than an one-size-fits-all strategy<sup>25</sup>. In this regard, people who developed pre-  
261 diabetes and T2D had lower levels of circulating PAPP-A (compared to those who did not developed  
262 pre-diabetes nor diabetes) and continued expressing lower levels of this protein after the disease was  
263 already installed. These findings were replicated in patients prospectively recruited for the HOMAGE  
264 trial, were those with diabetes also expressed lower levels of PAPP-A. PAPP-A increases the IGFs  
265 bioavailability and stimulates IGF-mediated growth through cleavage of the insulin-like growth factor  
266 binding protein-4 (IGFBP-4)<sup>26</sup>. The two central components of the IGF system, IGF-I and IGF-II,  
267 serve as ligands for the ubiquitously expressed IGF-I receptor, through which they facilitate cellular  
268 proliferation and differentiation as well as metabolic effects mimicking those conveyed by insulin<sup>27</sup>.  
269 Our complex network analyses support that PAPP-A interacts directly with IGF I and II through a  
270 pathway linked to IGFs. High concentrations of PAPP-A were initially found in women during  
271 pregnancy, but latter PAPP-A was found to be the primary IGFBP-4 protease in humans and many  
272 other species<sup>28-30</sup>. Under normal circumstances, IGFBPs bind IGFs with high affinity, stabilize the  
273 IGFs and transport them to peripheral tissues, where proteases, such as PAPP-A, cleave IGFBPs into  
274 low binding-affinity fragments to allow release and action of IGFs. Thus, PAPP-A actions occur  
275 within the tissues, and the direct liberation of IGF is a local event. PAPP-A could be a modulator of  
276 the local action of IGF-1, and increased PAPP-A levels may reflect a repair mechanism against  
277 vascular damage<sup>31</sup>. Collectively, whether is glucose, insulin or both that affect PAPP-A/IGFBP-4  
278 levels, and the clinical meaning of the circulating levels of PAPP-A remain speculative<sup>32, 33</sup>. The  
279 association of PAPP-A with diabetes in clinical studies has also been inconsistent<sup>33-35</sup>. It has been  
280 reported that genetic deletion of PAPP-A in mice confers resistance to the development of diabetic

281 nephropathy and induces an overall reduction in age-related diseases<sup>36-38</sup>. The human PAPP-A gene is  
282 located on chromosome 9<sup>39</sup>. In the present study, the SNP rs634737 was strongly associated with the  
283 expression of PAPP-A both at V1 and V4 and may provide valuable information for future studies on  
284 the role of PAPP-A. Moreover, this SNP was also associated with T2D, fasting glycemia, and  
285 glycosylated hemoglobin; thus, reinforcing the strength of our findings. Collectively, given that  
286 experimental and clinical data on PAPP-A have been rather inconclusive so far, much remains to be  
287 elucidated about their roles in metabolic diseases and complications<sup>26</sup>.

288 Other proteins were also associated with the onset and presence of pre-diabetes and T2D,  
289 these may represent several pathways associated with obesity, glucose intolerance, inflammation,  
290 apoptosis, angiogenesis and oxidative stress, for example<sup>40-43</sup>. Particularly, FGF-21 had a strong  
291 association with new-onset diabetes later in life. FGF-21 is an endocrine hormone that regulates  
292 energy homeostasis and insulin sensitivity<sup>44</sup>. Patients with T2D may present higher circulating levels  
293 of FGF-21, which may be influenced by other metabolic factors such as obesity and cholesterol<sup>45</sup>. In  
294 our study, adding FGF-21 to the clinical model for predicting diabetes, improved this model which  
295 already had a very good performance with the clinical variables alone. However, a causal link  
296 between the levels of FGF-21 and T2D has not been established<sup>46</sup>. The proteins (e.g., EPHB4,  
297 VCAM1, CES2) were only present in either one of the visits.

## 298 **Limitations**

299 Some limitations and caution in the interpretation of the results should be acknowledged in the present  
300 study. First, this is an observational study, therefore no causality can be established and these findings  
301 should be regarded as hypothesis generating. Second, PAPP-A and other circulating proteins were  
302 measured with the Olink® technology standardized log<sub>2</sub> NPX values (as described in the methods  
303 section), hence no direct conversion to the standard values is possible. Fourth, the SNP rs634737 is  
304 located in a non-coding region of the genome; therefore, no conclusion can be drawn about its mode  
305 of action and no replication could be performed and these results should be regarded with caution.  
306 Fifth, Mendelian randomization analyses could not be performed due to our “small” sample size  
307 for this type of analysis.

308

309 **Conclusion**

310 The risk of developing pre-diabetes and T2D increases with age, and with features of the metabolic  
311 syndrome. Several circulating proteins were associated with pre-diabetes and T2D. Particularly  
312 circulating PAPP- A, which has an important genetic component, was associated with both the  
313 development and presence of pre-diabetes and T2D. Our data provide novel insights on the  
314 pathophysiology of T2D and generate hypothesis for future research in the field, including the study  
315 of PAPP-A as a potential biotarget for the prevention and/or early treatment of diabetes.

316

317 **Disclosures**

318 None.

319

320 **Acknowledgments**

321 Please see the supplement.

322

323 **Bibliography**

- 324 1. Chen, L.; Magliano, D. J.; Zimmet, P. Z., The worldwide epidemiology of type 2 diabetes  
325 mellitus--present and future perspectives. *Nat Rev Endocrinol* **2011**, *8* (4), 228-36.
- 326 2. Whiting, D. R.; Guariguata, L.; Weil, C.; Shaw, J., IDF diabetes atlas: global estimates of the  
327 prevalence of diabetes for 2011 and 2030. *Diabetes Res Clin Pract* **2011**, *94* (3), 311-21.
- 328 3. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies  
329 with 4.4 million participants. *Lancet* **2016**, *387* (10027), 1513-1530.
- 330 4. Zhang, P.; Zhang, X.; Brown, J.; Vistisen, D.; Sicree, R.; Shaw, J.; Nichols, G., Global  
331 healthcare expenditure on diabetes for 2010 and 2030. *Diabetes Res Clin Pract* **2010**, *87* (3), 293-  
332 301.
- 333 5. Saha, S.; Gerdtham, U. G.; Johansson, P., Economic evaluation of lifestyle interventions for  
334 preventing diabetes and cardiovascular diseases. *Int J Environ Res Public Health* **2010**, *7* (8), 3150-95.
- 335 6. Backholer, K.; Peeters, A.; Herman, W. H.; Shaw, J. E.; Liew, D.; Ademi, Z.; Magliano, D. J.,  
336 Diabetes prevention and treatment strategies: are we doing enough? *Diabetes Care* **2013**, *36* (9),  
337 2714-9.
- 338 7. Ferreira, J. P.; Girerd, N.; Bozec, E.; Merckle, L.; Pizard, A.; Bouali, S.; Eby, E.; Leroy, C.;  
339 Machu, J. L.; Boivin, J. M. et al, Cohort Profile: Rationale and design of the fourth visit of the  
340 STANISLAS cohort: a familial longitudinal population-based cohort from the Nancy region of France.  
341 *Int J Epidemiol* **2018**, *47* (2), 395-395j.
- 342 8. Xhaard, C.; Dandine-Roulland, C.; Villemereuil, P.; Floch, E. L.; Bacq-Daian, D.; Machu, J. L.;  
343 Ferreira, J. P.; Deleuze, J. F.; Zannad, F.; et al, Heritability of a resting heart rate in a 20-year follow-  
344 up family cohort with GWAS data: Insights from the STANISLAS cohort. *Eur J Prev Cardiol* **2019**,  
345 2047487319890763.

- 346 9. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2019.  
347 *Diabetes Care* **2019**, *42* (Suppl 1), S13-s28.
- 348 10. Lundberg, M.; Eriksson, A.; Tran, B.; Assarsson, E.; Fredriksson, S., Homogeneous antibody-  
349 based proximity extension assays provide sensitive and specific detection of low-abundant proteins  
350 in human blood. *Nucleic Acids Res* **2011**, *39* (15), e102.
- 351 11. Pellicori, P.; Ferreira, J. P.; Mariottoni, B.; Brunner-La Rocca, H. P.; Ahmed, F. Z.;  
352 Verdonschot, J.; Collier, T.; Cuthbert, J. J.; Petutschnigg, J.; Mujaj, B.; et al. Effects of  
353 spironolactone on serum markers of fibrosis in people at high risk of developing heart failure:  
354 rationale, design and baseline characteristics of a proof-of-concept, randomised, precision-medicine,  
355 prevention trial. The Heart OMics in AGing (HOMAGE) trial. *Eur J Heart Fail* **2020**.
- 356 12. Kohler, S.; Carmody, L.; Vasilevsky, N.; Jacobsen, J. O. B.; Danis, D.; Gourdine, J. P.;  
357 Gargano, M.; Harris, N. L.; Matentzoglou, N.; McMurry, J. A.; Osumi-Sutherland, D.; et al. Expansion  
358 of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res* **2019**, *47*  
359 (D1), D1018-d1027.
- 360 13. Pinero, J.; Bravo, A.; Queralt-Rosinach, N.; Gutierrez-Sacristan, A.; Deu-Pons, J.; Centeno,  
361 E.; Garcia-Garcia, J.; Sanz, F.; Furlong, L. I., DisGeNET: a comprehensive platform integrating  
362 information on human disease-associated genes and variants. *Nucleic Acids Res* **2017**, *45* (D1), D833-  
363 d839.
- 364 14. Franco, M.; Bilal, U.; Ordunez, P.; Benet, M.; Morejon, A.; Caballero, B.; Kennelly, J. F.;  
365 Cooper, R. S., Population-wide weight loss and regain in relation to diabetes burden and  
366 cardiovascular mortality in Cuba 1980-2010: repeated cross sectional surveys and ecological  
367 comparison of secular trends. *Bmj* **2013**, *346*, f1515.
- 368 15. Maple-Brown, L. J.; Sinha, A. K.; Davis, E. A., Type 2 diabetes in indigenous Australian  
369 children and adolescents. *J Paediatr Child Health* **2010**, *46* (9), 487-90.
- 370 16. Jefferies, C.; Carter, P.; Reed, P. W.; Cutfield, W.; Mouat, F.; Hofman, P. L.; Gunn, A. J., The  
371 incidence, clinical features, and treatment of type 2 diabetes in children <15 yr in a population-based  
372 cohort from Auckland, New Zealand, 1995-2007. *Pediatr Diabetes* **2012**, *13* (4), 294-300.
- 373 17. Zimmet, P.; Alberti, K. G.; Shaw, J., Global and societal implications of the diabetes  
374 epidemic. *Nature* **2001**, *414* (6865), 782-7.
- 375 18. Haffner, S. M.; Stern, M. P.; Hazuda, H. P.; Mitchell, B. D.; Patterson, J. K., Cardiovascular  
376 risk factors in confirmed prediabetic individuals. Does the clock for coronary heart disease start  
377 ticking before the onset of clinical diabetes? *Jama* **1990**, *263* (21), 2893-8.
- 378 19. Kahn, R.; Davidson, M. B., The reality of type 2 diabetes prevention. *Diabetes Care* **2014**, *37*  
379 (4), 943-9.
- 380 20. Bouton, M. E., Why behavior change is difficult to sustain. *Prev Med* **2014**, *68*, 29-36.
- 381 21. Knowler, W. C.; Fowler, S. E.; Hamman, R. F.; Christophi, C. A.; Hoffman, H. J.; Brenneman,  
382 A. T.; Brown-Friday, J. O.; Goldberg, R.; Venditti, E.; Nathan, D. M., 10-year follow-up of diabetes  
383 incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet* **2009**, *374*  
384 (9702), 1677-86.
- 385 22. Knowler, W. C.; Barrett-Connor, E.; Fowler, S. E.; Hamman, R. F.; Lachin, J. M.; Walker, E.  
386 A.; Nathan, D. M., Reduction in the incidence of type 2 diabetes with lifestyle intervention or  
387 metformin. *N Engl J Med* **2002**, *346* (6), 393-403.
- 388 23. Hamman, R. F.; Wing, R. R.; Edelstein, S. L.; Lachin, J. M.; Bray, G. A.; Delahanty, L.;  
389 Hoskin, M.; Kriska, A. M.; Mayer-Davis, E. J.; Pi-Sunyer, X.; et al. Effect of weight loss with lifestyle  
390 intervention on risk of diabetes. *Diabetes Care* **2006**, *29* (9), 2102-7.
- 391 24. Lindstrom, J.; Ilanne-Parikka, P.; Peltonen, M.; Aunola, S.; Eriksson, J. G.; Hemio, K.;  
392 Hamalainen, H.; Harkonen, P.; Keinanen-Kiukaanniemi, S.; Laakso, M.; et al. Sustained reduction in  
393 the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes  
394 Prevention Study. *Lancet* **2006**, *368* (9548), 1673-9.
- 395 25. The 10-year cost-effectiveness of lifestyle intervention or metformin for diabetes  
396 prevention: an intent-to-treat analysis of the DPP/DPPOS. *Diabetes Care* **2012**, *35* (4), 723-30.

397 26. Hjortebjerg, R., IGFBP-4 and PAPP-A in normal physiology and disease. *Growth Horm IGF Res*  
398 **2018**, *41*, 7-22.

399 27. Jones, J. I.; Clemmons, D. R., Insulin-like growth factors and their binding proteins: biological  
400 actions. *Endocr Rev* **1995**, *16* (1), 3-34.

401 28. Conover, C. A.; Mason, M. A.; Bale, L. K.; Harrington, S. C.; Nyegaard, M.; Oxvig, C.;  
402 Overgaard, M. T., Transgenic overexpression of pregnancy-associated plasma protein-A in murine  
403 arterial smooth muscle accelerates atherosclerotic lesion development. *Am J Physiol Heart Circ*  
404 *Physiol* **2010**, *299* (2), H284-91.

405 29. Mazerbourg, S.; Overgaard, M. T.; Oxvig, C.; Christiansen, M.; Conover, C. A.; Laurendeau,  
406 I.; Vidaud, M.; Tosser-Klopp, G.; Zapf, J.; Monget, P., Pregnancy-associated plasma protein-A (PAPP-  
407 A) in ovine, bovine, porcine, and equine ovarian follicles: involvement in IGF binding protein-4  
408 proteolytic degradation and mRNA expression during follicular development. *Endocrinology* **2001**,  
409 *142* (12), 5243-53.

410 30. Byun, D.; Mohan, S.; Yoo, M.; Sexton, C.; Baylink, D. J.; Qin, X., Pregnancy-associated  
411 plasma protein-A accounts for the insulin-like growth factor (IGF)-binding protein-4 (IGFBP-4)  
412 proteolytic activity in human pregnancy serum and enhances the mitogenic activity of IGF by  
413 degrading IGFBP-4 in vitro. *J Clin Endocrinol Metab* **2001**, *86* (2), 847-54.

414 31. Beaudoux, J. L.; Burc, L.; Imbert-Bismut, F.; Giral, P.; Bernard, M.; Bruckert, E.; Chapman,  
415 M. J., Serum plasma pregnancy-associated protein A: a potential marker of echogenic carotid  
416 atherosclerotic plaques in asymptomatic hyperlipidemic subjects at high cardiovascular risk.  
417 *Arterioscler Thromb Vasc Biol* **2003**, *23* (1), e7-10.

418 32. Panagiotou, G.; Anastasilakis, A. D.; Kynigopoulos, G.; Skouvaklidou, E. C.; Saridakis, Z. G.;  
419 Upadhyay, J.; Pagkalidou, E.; Apostolou, A.; Karagiozoglou-Lampoudi, T.; Mantzoros, C. S.,  
420 Physiological parameters regulating circulating levels of the IGFBP-4/Stanniocalcin-2/PAPP-A axis.  
421 *Metabolism* **2017**, *75*, 16-24.

422 33. Hjortebjerg, R.; Laugesen, E.; Hoyem, P.; Oxvig, C.; Stausbol-Gron, B.; Knudsen, S. T.; Kim,  
423 W. Y.; Poulsen, P. L.; Hansen, T. K.; et al. The IGF system in patients with type 2 diabetes:  
424 associations with markers of cardiovascular target organ damage. *Eur J Endocrinol* **2017**, *176* (5),  
425 521-531.

426 34. Pellitero, S.; Reverter, J. L.; Pizarro, E.; Pastor, M. C.; Granada, M. L.; Tassies, D.; Reverter,  
427 J. C.; Salinas, I.; Sanmarti, A., Pregnancy-associated plasma protein-a levels are related to glycemic  
428 control but not to lipid profile or hemostatic parameters in type 2 diabetes. *Diabetes Care* **2007**, *30*  
429 (12), 3083-5.

430 35. Aso, Y.; Okumura, K.; Wakabayashi, S.; Takebayashi, K.; Taki, S.; Inukai, T., Elevated  
431 pregnancy-associated plasma protein-a in sera from type 2 diabetic patients with  
432 hypercholesterolemia: associations with carotid atherosclerosis and toe-brachial index. *J Clin*  
433 *Endocrinol Metab* **2004**, *89* (11), 5713-7.

434 36. Mader, J. R.; Resch, Z. T.; McLean, G. R.; Mikkelsen, J. H.; Oxvig, C.; Marler, R. J.; Conover,  
435 C. A., Mice deficient in PAPP-A show resistance to the development of diabetic nephropathy. *J*  
436 *Endocrinol* **2013**, *219* (1), 51-8.

437 37. Conover, C. A.; Bale, L. K.; Mader, J. R.; Mason, M. A.; Keenan, K. P.; Marler, R. J., Longevity  
438 and age-related pathology of mice deficient in pregnancy-associated plasma protein-A. *J Gerontol A*  
439 *Biol Sci Med Sci* **2010**, *65* (6), 590-9.

440 38. Junnila, R. K.; List, E. O.; Berryman, D. E.; Murrey, J. W.; Kopchick, J. J., The GH/IGF-1 axis in  
441 ageing and longevity. *Nat Rev Endocrinol* **2013**, *9* (6), 366-376.

442 39. Silahatoglu, A. N.; Tumer, Z.; Kristensen, T.; Sottrup-Jensen, L.; Tommerup, N., Assignment  
443 of the human gene for pregnancy-associated plasma protein A (PAPPA) to 9q33.1 by fluorescence in  
444 situ hybridization to mitotic and meiotic chromosomes. *Cytogenet Cell Genet* **1993**, *62* (4), 214-6.

445 40. Sagar, V.; Vatapalli, R.; Lysy, B.; Pamarthy, S.; Anker, J. F.; Rodriguez, Y.; Han, H.; Unno,  
446 K.; Stadler, W. M.; et al. EPHB4 inhibition activates ER stress to promote immunogenic cell death of  
447 prostate cancer cells. *Cell Death Dis* **2019**, *10* (11), 801.

- 448 41. Ruby, M. A.; Massart, J.; Hunerdosse, D. M.; Schonke, M.; Correia, J. C.; Louie, S. M.;  
449 Ruas, J. L.; Naslund, E.; Nomura, D. K.; Zierath, J. R., Human Carboxylesterase 2 Reverses Obesity-  
450 Induced Diacylglycerol Accumulation and Glucose Intolerance. *Cell Rep* **2017**, *18* (3), 636-646.
- 451 42. Ware, C. F., The TNF Superfamily-2008. *Cytokine Growth Factor Rev* **2008**, *19* (3-4), 183-6.
- 452 43. Kong, D. H.; Kim, Y. K.; Kim, M. R.; Jang, J. H.; Lee, S., Emerging Roles of Vascular Cell  
453 Adhesion Molecule-1 (VCAM-1) in Immunological Disorders and Cancer. *Int J Mol Sci* **2018**, *19* (4).
- 454 44. BonDurant, L. D.; Ameka, M.; Naber, M. C.; Markan, K. R.; Idiga, S. O.; Acevedo, M. R.;  
455 Walsh, S. A.; Ornitz, D. M.; Potthoff, M. J., FGF21 Regulates Metabolism Through Adipose-  
456 Dependent and -Independent Mechanisms. *Cell Metab* **2017**, *25* (4), 935-944.e4.
- 457 45. Wang, Y. S.; Ye, J.; Cao, Y. H.; Zhang, R.; Liu, Y.; Zhang, S. W.; Dai, W.; Zhang, Q., Increased  
458 serum/plasma fibroblast growth factor 21 in type 2 diabetes mellitus: a systematic review and meta-  
459 analysis. *Postgrad Med J* **2019**, *95* (1121), 134-139.
- 460 46. Davis, G. R.; Deville, T.; Guillory, J.; Bellar, D.; Nelson, A. G., Relationship between family  
461 history of type 2 diabetes and serum FGF21. *Eur J Clin Invest* **2017**, *47* (11), 853-859.
- 462



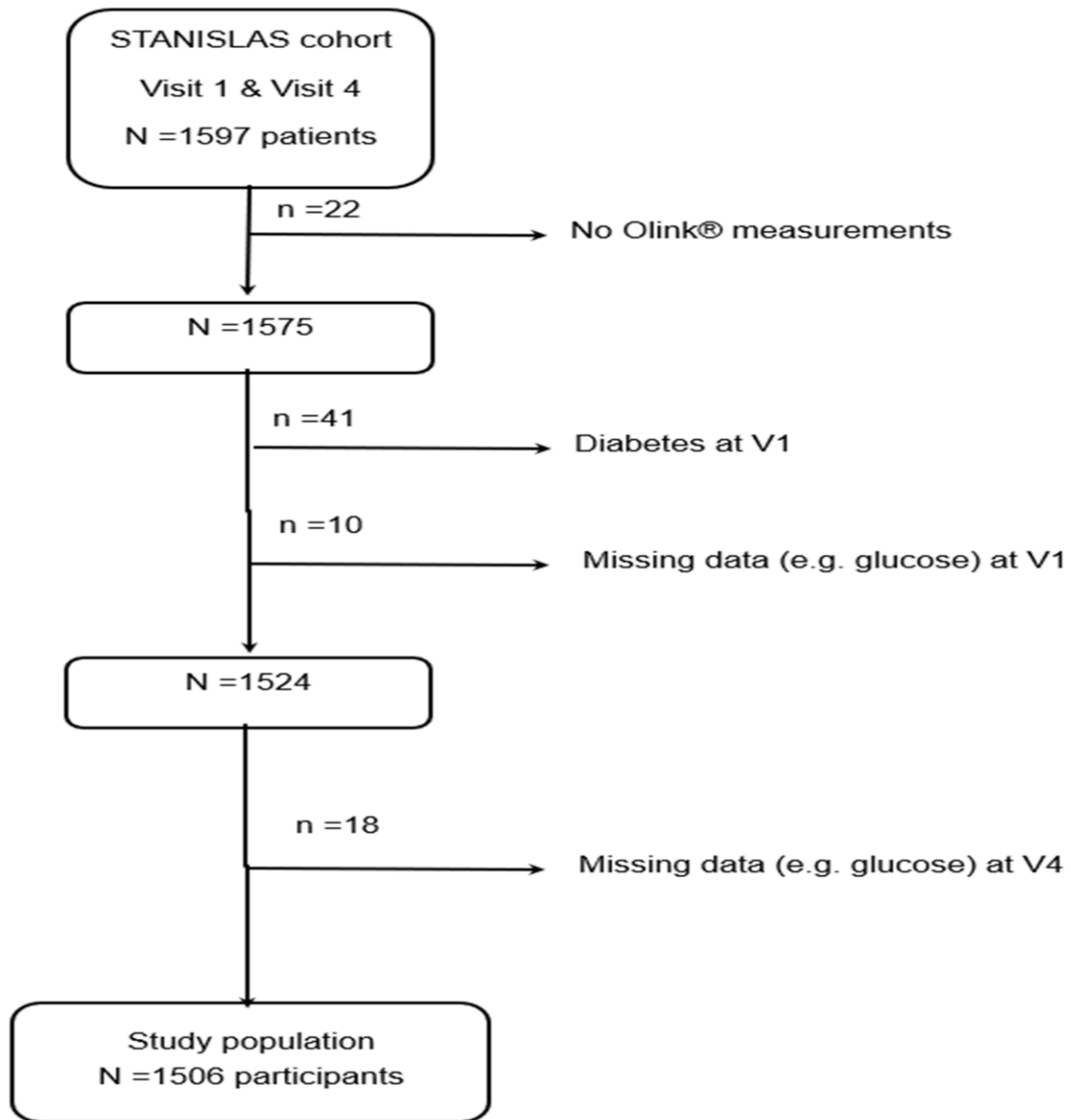


Figure 1. Study flow chart

Figure 2. Pathway (in green) and proteins (grey circles) linking PAPP-A (red circle) to diabetes (blue square).

