

1 Title: The Effect of Metformin Treatment on the Circulating Proteome

2

3 Ben Connolly*, Laura McCreight*, Roderick C Slieker*, Khaled F Bedair, Louise Donnelly,

4 Juliette A de Klerk, JWJ Beulens, PM Elders, IMI-DIRECT, IMI-RHAPSODY, Göran

5 Bergström, Mun-Guan Hong, Robert W. Koivula, Paul W. Franks, Leen 't Hart†, Jochen M

6 Schwenk†, Anders Gummesson†, Ewan R Pearson†

7

8 BC- b.w.connolly@dundee.ac.uk, LMcC- laura.mccreight@nhs.scot, KFB-

9 k.f.e.bedair@dundee.ac.uk, LD- l.y.donnelly@dundee.ac.uk, ERP-

10 e.z.pearson@dundee.ac.uk: Division of Population Health & Genomics, School of Medicine,

11 University of Dundee, UK.

12 MGH- mun-gwan.hong@scilifelab.se, JMS- jochen.schwenk@scilifelab.se: Department of

13 Protein Science, SciLifeLab, KTH Royal Institute of Technology, Stockholm, Sweden

14 RCS- r.c.slieker@lumc.nl: Department of Cell and Chemical Biology, Leiden University

15 Medical Center, Leiden, the Netherlands

16 LtH- lmthart@lumc.nl: Department of Cell and Chemical Biology, Leiden University

17 Medical Center, Leiden, the Netherlands

18 Department of Epidemiology and Data Sciences, Amsterdam University Medical Center,

19 Location Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

20 Health Behaviors & Chronic Diseases Research Program & Personalised Medicine Research

21 Program, Amsterdam Public Health, Amsterdam, The Netherlands

22 Department of Biomedical Data Sciences, Section Molecular Epidemiology, Leiden

23 University Medical Center, Leiden, The Netherlands

24

25 RK- robert.koivula@med.lu.se: Oxford Centre for Diabetes, Endocrinology and Metabolism,

26 Radcliffe Department of Medicine, University of Oxford, Oxford, UK.

27 Genetic and Molecular Epidemiology Unit, Lund University Diabetes Centre, Department of

28 Clinical Medicine, Lund University, Malmo, Sweden.

29

30 JWB- j.beulens@amsterdamumc.nl: Department of Epidemiology & Data Science,

31 Amsterdam UMC, location Vrije Universiteit, Amsterdam, The Netherlands

32 Amsterdam Public Health, Amsterdam Cardiovascular Sciences research institutes,

33 Amsterdam, The Netherlands

34 JAdK- j.a.de_klerk@lumc.nl: Department of Cell and Chemical Biology, Leiden University

35 Medical Center, Leiden, the Netherlands / Department of Internal Medicine (Nephrology),

36 Leiden University Medical Center, Leiden, the Netherlands

37 AG- anders.gummesson@vgregion.se: Region Västra Götaland, Sahlgrenska University

38 Hospital, Department of Clinical Genetics and Genomics, Gothenburg, Sweden.

39 GB- goran.bergstrom@hjl.gu.se: Region Västra Götaland, Sahlgrenska University Hospital,

40 Department of Clinical Physiology, Gothenburg, Sweden.

41 AG, GB: Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska

42 Academy, University of Gothenburg, Gothenburg, Sweden.

43

44 *joint first author; †joint senior author

45 Adress for correspondence

46 Ewan Pearson
47 Division of Population Health & Genomics
48 School of Medicine
49 University of Dundee
50 DD1 9SY
51 +44 1382 383387
52 e.z.pearson@dundee.ac.uk

53 Word Count- 3986, 4 Figures

54

55 Highlights

- 56 • In the most comprehensive proteomic analysis of metformin exposure to date, we
57 showed 97 proteins to be associated with metformin exposure in at least one study.
- 58 • 14 proteins were consistently associated with metformin exposure in 2 or more
59 platforms or studies.
- 60 • Gene enrichment analysis shows that the strongest protein set is of intestinal origin.
- 61 • These data provide further insight into the mechanism of action of metformin,
62 potentially identify novel targets for diabetes treatment and highlight the need to
63 account for metformin exposure in proteomic studies and where protein biomarkers
64 are used for clinical care.

65

66 Objective

67 Metformin is one of the most used drugs worldwide. However, its mechanism of action
68 remains uncertain. Given the potential to reveal novel insights into the pleiotropic effects of

69 metformin treatment, we aimed to undertake a comprehensive analysis of circulating
70 proteins.

71 [Research Design and Methods](#)

72 We analysed 1195 proteins using the SomaLogic platform in 1175 participants, using cross-
73 sectional data from the GoDARTS and DCS cohorts; 450 proteins using the Olink platform in
74 784 participants, using cross-sectional data from IMI-DIRECT; and combined longitudinal
75 data from the IMPOCT, RAMP and S3WP-T2D cohorts with 372 proteins in 98 participants
76 using the Olink platform. Finally, we performed systems level analysis on the longitudinal
77 OLINK data to identify any possible relationships for the proteins changing concentration
78 following metformin exposure.

79 [Results](#)

80 Overall, 97 proteins were associated with metformin exposure in at least one of the studies
81 ($P_{adj} < 0.05$), and 10 proteins (EpCAM, SPINK1, t-PA, Gal-4, TFF3, TF, FAM3C, COL1A1,
82 SELL, CD93) were associated in two independent studies. Four proteins, REG4, GDF15,
83 REG1A, and OMD were consistently associated across all studies and platforms. Gene-set
84 enrichment analysis revealed that the effect of metformin exposure was on intestinal tissues.
85 In the longitudinal analysis 18% of proteins were significantly altered by metformin.

86 [Conclusions](#)

87 These data provide further insight into the mechanism of action of metformin, potentially
88 identifying novel targets for diabetes treatment, and highlight the need to account for
89 metformin exposure in proteomic studies and where protein biomarkers are used for clinical
90 care where metformin treatment will generate false positive results.

91

92

93

94

95 Metformin works by several mechanisms known to have a positive impact on inflammation
96 and metabolism; it does this by primarily acting on the liver and the gut(1), however the exact
97 molecular mechanisms remain uncertain. The increasing availability of deep molecular
98 phenotyping in patients treated with metformin, including genomic, transcriptomic,
99 metabolic, proteomic and metagenomic data, offers the opportunity to gain further
100 mechanistic insight into the mechanism of action of metformin in humans. Genome-wide
101 association studies (GWAS) have provided some insight into the mechanism of action of
102 metformin in people with type 2 diabetes(2). In GWAS studies, two genetic variants have
103 been reproducibly associated with glycaemic response to metformin – rs11212617 at a locus
104 including *ATM*(3) and rs8192675, intronic in *SLC2A2*, associated with altered expression of
105 GLUT2(4). Other studies have reported on epigenetic markers(5), the transcriptome(6), the
106 metabolome(7) and the microbiome(8) altering glycaemic response, weight change or
107 intolerance in people with diabetes.

108 There are limited proteomic studies of metformin exposure, and these have largely been
109 targeted or using small panels. A reproducible robust association has been described between
110 metformin exposure and serum Growth differentiation factor 15 (GDF15) concentrations.

111 This was first identified in a Luminex panel measuring 237 proteins from the ORIGIN
112 study(9). This has been subsequently replicated with mechanistic rodent studies establishing
113 that metformin associated increase in GDF15 resulted in a reduction in food intake and body
114 weight and that the origin of GDF15 associated with metformin exposure was the
115 intestine(10). More recently, Gummesson et al carried out a more comprehensive proteomic

116 analysis following metformin treatment. This further showed that GDF15 was increased
117 following metformin treatment in addition to identifying other proteins significantly altered
118 by metformin exposure; for example, EpCAM was reduced in those treated with
119 metformin(11). Given the potential for proteomic signatures after metformin exposure to
120 inform on its mechanism of action and identify novel diabetes drug targets, here we extend
121 these analyses, including the study by Gummesson et al. but greatly increasing the number of
122 individuals included to 2,057 and increasing the number of studies to incorporate both cross-
123 sectional and longitudinal studies of metformin exposure, using two commonly used
124 proteomic methods – Olink and SomaLogic.

125 [Research Design and Methods](#)

126

127 **Cohorts**

128 **GoDARTS:** The Genetics of Diabetes Audit and Research Tayside Study (GoDARTS) is a
129 cohort of ~8,000 individuals with T2D(12). Laboratory measurements were non-fasted. For
130 SomaLogic analysis, samples from 599 patients were selected age >35 years, GAD antibody
131 negative, with blood sampled close to diagnosis (median diabetes duration 1.4 years).

132 **DCS:** The Hoorn Diabetes Care System (DCS) cohort is a prospective cohort with currently
133 over 14,000 individuals with routine care data. In 2008–2014, additional blood sampling was
134 done in 5500 participants, who provided written informed consent. These samples were used
135 for this study. For SomaLogic analysis, samples from 576 patients were selected age >35
136 years, GAD antibody negative, with blood sampled close to diagnosis (median diabetes
137 duration 2.6 years)(13).

138 **DIRECT:** This cohort included 784 patients with recently diagnosed type 2 diabetes. The
139 mean age at inclusion was 62 years with the youngest 35 years at baseline, which should
140 exclude any individuals with MODY. Participants were diagnosed within two years before
141 recruitment, were on lifestyle and/or metformin treatment only, and had glycated
142 haemoglobin (HbA_{1c}) < 60.0 mmol/mol (< 7.6%) within previous three months(14).

143 **S3WP-T2D:** This study was carried out to elucidate the changes in the proteome in the early
144 stages of diabetes and how the proteome is affected by diabetes treatment including
145 metformin(11). 52 previously undiagnosed patients were identified as having type 2 diabetes
146 from a screening program and as a result, were recruited for the study. Patients were excluded
147 if they had a pre-existing disease which would affect their ability to participate, severe
148 hyperglycaemia needing hospital attention or immediate insulin therapy, or a major surgical
149 procedure or trauma within 4 weeks. Included patients were treated for diabetes via first line
150 therapy; weight management and exercise with or without metformin which was decided by a
151 doctor. Protein levels in the blood were measured at baseline, one month and 3 months. Of
152 the 52 participants, 51 completed the 3 month follow up visit and for 3 patients plasma
153 samples were not available for the 1 month visit. This left data for 48 patients to be analysed.

154 **IMPOCT:** The IMPOCT study was designed to investigate the impact of the OCT1 genotype
155 and OCT1 inhibiting drugs on an individual's ability to tolerate metformin. For our analysis,
156 only the data from when individuals were treated with metformin and placebo was utilised,
157 and not data from individuals on OCT1 inhibiting drugs. 38 healthy participants without
158 diabetes were recruited for this study. They were on metformin for 4 weeks, titrated to a max
159 dose of 1000mg BD which they took for the final week of the study. Protein levels in the
160 blood were measured at baseline and after the 4 weeks of metformin treatment.

161 **RAMP:** The RAMP study was designed to investigate the response of individuals with ataxia
162 telangiectasia to metformin and pioglitazone. For our analysis, we only utilised data from
163 control patients (without ataxia telangiectasia) on metformin (not pioglitazone). 12 non-
164 diabetic, healthy controls, who had never been on metformin before were started on the drug.
165 They were treated with metformin for 8 weeks, titrated to a maximum dose of 1000mg BD
166 which they were on for the final 4 weeks of the study. Protein levels in the blood were
167 measured at baseline and after the 8 weeks of metformin treatment.

168 **Proteomics assays**

169 We used two complementary affinity proteomics approaches to determine the relative levels
170 of circulating proteins in blood samples(15). Each technology is capable to measure
171 thousands of proteins, but a list of 500 proteins have been described to correlate with high
172 confidence between the two platforms. Cross-sectional data from GoDARTS and DCS was
173 analysed using SomaLogic; 1195 proteins were measured and included in the analysis after
174 standardized QC. Olink panels were used as follows:

175 **DIRECT:** The proteins were measured on five Olink panels: Cardiometabolic,
176 Cardiovascular II, Cardiovascular III, Development, Metabolism. After proteins were
177 removed following quality control, this left 450 proteins to be analysed.

178 **S3WP-T2D:** The proteins were measured on eleven Olink panels (Cardiometabolic, Cell
179 Regulation, Cardiovascular II, Cardiovascular III, Development, Immune Response,
180 Oncology II, Inflammation, Metabolism, Neurology, and Organ Damage).

181 **IMPOCT.** The proteins were measured on five Olink panels (Cardiometabolic,
182 Cardiovascular II, Cardiovascular III, Development and Metabolism).

183 RAMP. These proteins were measured on the same five Olink panels as the IMPOCT study
184 (Cardiometabolic, Cardiovascular II, Cardiovascular III, Development and Metabolism).

185 After proteins were removed following quality control and to ensure each protein was
186 included in all three studies, 372 proteins were analysed in the combined analysis of S3WP-
187 T2D, IMPOCT and RAMP.

188 **Statistical methods**

189 **RHAPSODY: GoDARTS and DCS**

190 We undertook a linear regression for the biomarker as dependent variable, with metformin
191 exposure (Y/N), adjusted for age and gender. This was done for both DCS and GoDARTS
192 and then data were combined using random effects meta-analysis. A Bonferroni correction
193 was applied for the 1195 assays included in the analysis.

194 In both cohorts we then analysed the protein levels of the proteins significantly associated
195 with metformin exposure, in relation to the daily metformin dose used by the participants.
196 Protein levels were used as endpoints in linear regression analyses and metformin dose as
197 predictor with adjustment for gender, age, BMI and HbA1c levels. Persons not using
198 metformin were excluded prior to this analysis. Metformin dose was binned per 500mg to
199 ease the interpretation of the data. i.e. the beta is the change in protein level per extra 500mg
200 tablet of metformin. Similar results were obtained in analyses where metformin dose was
201 included as a continuous variable.

202 **DIRECT**

203 A linear mixed model was applied using the lmer function of the R package lme4. In this
204 model, the Olink NPX data was adjusted by information related to the donor (age at
205 sampling, sex) the sampling event (date, centre) as well as technical aspects (assay plate).

206 **Combined analysis: S3WP-T2D, IMPOCT and RAMP**

207 The longitudinal data from the S3WP-T2D study described by Gummesson et al(11) were
208 combined with the longitudinal data from two Dundee studies; IMPOCT and RAMP. In these
209 studies, Olink panels were used to measure proteins before and after metformin exposure.

210 Statistical analysis was performed using R Studio version 4.1.2. Proteins were analysed using
211 linear mixed models, with the R package LmerTest. The metformin dose and study name
212 were used as a fixed effect and the study individual was used as a random effect. Proteins
213 were removed so that the all the proteins present in the combined study data were analysed in
214 each of the three studies. This left 372 proteins to be analysed in the combined study data.
215 The metformin dose was simplified and allocated a 0 if the individual was not on metformin
216 and a 1 if the individual was on metformin treatment, regardless of the dose. P values were
217 adjusted using the Bonferroni method.

218 **Adjusting for BMI with metformin exposure in S3WP-T2D and RAMP**

219 It has been shown that the proteome can vary in response to weight gain and weight loss(16),
220 and that metformin is associated with weight loss. Consequently, a further linear mixed
221 model analysis was performed in which BMI was included as a covariate, in two of the three
222 longitudinal cohorts where weight was measured before and after metformin initiation.

223 The longitudinal data from S3WP-T2D and RAMP was combined, and the same 372 proteins
224 were analysed as before using a linear mixed model. Again, R package LmerTest was used
225 for this statistical analysis on R Studio version 4.1.2. Metformin dose, study name and BMI
226 were used as fixed effects, with the study individual as a random effect. P values were also
227 adjusted using the Bonferroni method.

228 **Tissue specific gene expression and gene set analysis for proteins altered by metformin**
229 **exposure**

230 Using Genotype Tissue Expression (GTEx) analysis we explored in what tissues the genes
231 encoding the proteins altered by metformin were expressed. We then used the enrichR
232 package in R to evaluate which tissues were enriched for, based upon the change in proteins
233 in response to metformin treatment in the combined analysis of SCAPIS, IMPOCT and
234 RAMP. Proteins were converted to gene symbols and upregulated and downregulated
235 proteins were tested separately. An adjusted P-value smaller than 0.05 was considered
236 significant.

237 **Causal inference- pQTL**

238 Protein Quantitative Trait Loci (pQTL) analysis was carried out on 14 proteins that were
239 significantly associated with metformin exposure in two or more studies. pQTLs were
240 obtained from Sun et al. (BioRxiv, 2022). pQTLs *in cis* were filtered based on the UniProt
241 ID. pQTLs associated with proteins were compared to available traits in the OpenGWAS
242 database, but eQTLs, cancer-, peptide-, unknown metabolite traits were excluded. pQTLs and
243 identified traits were harmonised using the *harmonise_data* function in the TwoSampleMR
244 package. eQTLs were obtained from GTEx v8. EQTLs were considered significant if the p-
245 value was below GTEx's P-value threshold.

246 **Results**

247 **Cross sectional metformin exposure and the SomaLogic platform.** We first assessed
248 differences in protein levels measured using the SomaLogic platform in metformin treated
249 and untreated individuals from two cross-sectional studies as part of the IMI-RHAPSODY
250 consortium, where proteomic analysis was undertaken on two populations with type 2

251 diabetes close to diagnosis. The baseline characteristics of the two included cohorts,
252 GoDARTS and DCS cohort are shown in Supplementary Table 1. In the meta-analysis of
253 these two datasets, 1195 proteins were analysed in 1175 subjects. After Bonferroni
254 correction, levels of 34 proteins were significantly associated with metformin exposure
255 (Supplementary Table 2). The proteins where metformin exposure was associated with the
256 largest difference in levels were REG4 (Beta=0.698, SE=0.084), GDF15 (Beta=0.657, SE
257 0.098), PYY (Beta=0.662, SE=0.147) and FGF19 (Beta=-0.519, SE=0.115). Given that this
258 analysis was cross-sectional, and associations may not be causally associated with metformin
259 exposure we investigated the effect of metformin dose on protein concentrations, as a dose
260 effect would support a causal relationship between metformin exposure and protein
261 expression. Of the 34 proteins whose concentration was associated with metformin exposure
262 adjusted for gender, age, BMI and HbA1c, a nominally significant ($p < 0.05$) dose effect was
263 seen for most proteins (25/34). After Bonferroni correction ($p < 0.05/34$) increased metformin
264 dose was associated with increased REG4, GDF15, CDH6 and PYY concentrations and
265 decreased FGF19 and OMD concentrations (Supplementary Table 3).

266 **Cross Sectional metformin exposure and the Olink platform.** We then assessed
267 differences in protein levels measured using the Olink platform in individuals from the IMI-
268 DIRECT cohort of patients with type 2 diabetes diagnosed within the prior 2 years. Baseline
269 characteristics of the IMI-DIRECT cohort are shown in Supplementary Table 1. In this
270 study, 450 proteins were analysed in 784 subjects. After Bonferroni correction, 13 proteins
271 were significantly different between metformin users and non-users (Supplementary Table 4).
272 The largest signal was seen for a reduction in EpCAM, followed by an increase in SPINK1,
273 REG4 and GDF15).

274 **Longitudinal metformin exposure and the Olink platform.** We then analysed protein
275 concentrations longitudinally in individuals before metformin treatment and after metformin
276 initiation in 3 clinical trials. Baseline characteristics of included cohorts are shown in
277 Supplementary Table 1. Individual level data from the S3WP-T2D(11), IMPOCT and
278 RAMP studies were combined, and 372 proteins were analysed in 98 subjects using Olink
279 panels before and during metformin treatment. After Bonferroni correction, 68 proteins (18%
280 of measured proteins) were found to be significantly changed with metformin treatment
281 (Supplementary Table 5) and are represented in the volcano plot (Figure 1). The top 8 most
282 significant proteins are labelled in the figure and are as follows from most significant to least
283 significant: REG4, GDF-15, EpCAM, SPINK1, REG1A, LDL receptor, IGFBP-2 and t-PA.
284 We analysed the longitudinal data in males and females separately (Supplementary Table 10)
285 and show a consistent signal by sex for the top 6 most significant proteins. However, there
286 were differences, with 30 proteins including IGFBP-2 and t-PA being significantly changed
287 following metformin in males but not females. Similarly, CDH5, FAM3C, HAOX1 and
288 CCL15 were significantly changed after metformin in females but not males.

289 As metformin is associated with weight loss, we then evaluated whether the change in protein
290 concentration with metformin exposure was attenuated by BMI change in two of the
291 longitudinal cohorts where BMI was measured before and after metformin initiation (S3WP-
292 T2D and RAMP). Complete attenuation of the change in protein concentration would
293 suggest that the difference is secondary to, or causal for, BMI change. As a positive control,
294 LEP (Leptin) is significantly reduced by metformin treatment, but adjusting for metformin
295 associated BMI change attenuates the effect by 61%, with loss of significance ($p=0.15$).
296 Supplementary Table 9 provides the results for the impact of metformin on protein
297 concentration with and without adjustment for BMI. Adjusting for BMI reduces the number
298 of Bonferroni significant proteins from 31 out of 372 proteins analysed across these two

299 studies to 21 proteins. Other than for leptin, the largest attenuation by adjusting for BMI was
300 seen with FUCA1 (84.2% attenuation); GUSB (29.5% attenuation); SELE (26.3%
301 attenuation), IGFBP2 (21% attenuation) and FCN2 (15% attenuation). Interestingly there
302 was no attenuation (0%) for GDF15, a protein previously reported to potentially mediate the
303 weight change caused by metformin(10).

304

305 Given the large number of proteins identified to be altered in our longitudinal analysis we
306 used GTEX(17), HPA(18), and STRING(19) to obtain system-level insights about possible
307 relationships for the 68 proteins identified. The tissue expression (mRNA and protein) from
308 GTEX of the top 68 proteins altered by metformin in the longitudinal analysis is shown in
309 Figure 2. Note that the expression reported here is tissue specific expression and does not
310 relate to metformin exposure. Expression of genes for proteins most changed by metformin
311 exposure were seen predominantly in the pancreas and intestine. In a gene set enrichment
312 analysis (Supplementary Table 7), upregulated proteins were enriched for colon (OR=20.4,
313 $P_{adj} = 6.49 \times 10^{-5}$), based on overlap with REG4, REG1A, GDF15, TFF, SPINK1, CCL15,
314 PIGR and LGALS4. Downregulated proteins were enriched for omentum (OR=5.2, $P_{adj} =$
315 2.93×10^{-6}) and liver (OR = 4.84, $P_{adj} = 7.52 \times 10^{-6}$). We also explored possible protein
316 interactions using the default functions of the STRING database(19). This revealed that many
317 proteins have known or predicted interactions with each other (Figure 3). Among them, a
318 group of proteins involved in cell adhesion (KEGG pathway, $P=2.8 \times 10^{-8}$) were particularly
319 connected based on experimental evidence.

320

321 **Proteomic signatures of metformin across study and platform.** Across the three studies,
322 there were 14 proteins where metformin exposure was associated with protein concentration
323 in at least two studies; the direction and effect sizes for these associations are shown in Figure

324 4. Four proteins were consistently associated with metformin exposure in the 3 studies
325 (including 2 cross-sectional and 1 longitudinal design) and across the two platforms (Olink
326 and SomaLogic); these were **REG4, GDF15, REG1A and OMD**. 8 additional proteins
327 were consistently associated with metformin exposure across the two studies using Olink;
328 these were **EpCAM, SPINK1, t-PA, Gal-4, TFF3, TF, FAM3C, COL1A1**. There were 2
329 proteins associated with metformin exposure common between the cross-sectional
330 SomaLogic and longitudinal Olink studies; these were **SELL** and **CD93**. Protein Quantitative
331 Trait Loci (pQTL) analysis was carried out for these 14 proteins. 11 proteins had at least one
332 associated pQTL, while Ep-CAM, SELL and CD93 did not (Supplementary table 8). There
333 were limited informative pQTL: trait associations that could be linked to metformin
334 exposure. A few examples include: where metformin and a pQTL both cause an increase in
335 GDF-15, neutrophil counts are increased, and monocyte counts are decreased; where
336 metformin lowers OMD, the A allele at rs35209758 which is also associated with lower
337 OMD is associated with an increase in Asporin and Hematopoietic progenitor cell antigen
338 CD34; where metformin increases FAM3C, the A allele at rs36198735 that is associated with
339 higher FAM3C is associated with higher heel bone mineral density.

340

341 **Conclusions**

342 We have undertaken the most comprehensive proteomic analysis of metformin exposure in
343 people with and without diabetes to date. Our analysis spans different proteomic approaches
344 and large cross-sectional studies and longitudinal studies with measures before and during
345 metformin treatment. Overall, 97 proteins were associated with metformin exposure in at
346 least one study. The concentration of 4 proteins (REG4, GDF15, REG1A and OMD) were
347 associated with metformin exposure across all platforms and studies, and a further 10 proteins
348 were consistently associated with metformin exposure in two independent studies.

349 Enrichment analysis shows that the strongest protein-set is of intestinal origin, consistent with
350 the very high concentrations of metformin seen in intestinal epithelial cells. An increase in
351 GDF-15(9) and a decrease in EpCAM(11) after metformin has been previously described and
352 our results have confirmed these findings.

353 Our data add to the already robust literature that metformin increases serum GDF-15. GDF-
354 15 is a protein that increases in concentration due to cellular stress caused by mitochondrial
355 dysfunction, hypoxia, and exercise(20). It has been previously shown that the intestine
356 (particularly the lower small intestine and colon) was a main site of increased GDF-15
357 expression following metformin treatment(10). Although GDF-15 is associated with adverse
358 outcomes such as increasing age, cancer and cellular stress, pharmacologically increasing
359 GDF-15 could be beneficial. In wild-type mice treated with a high fat diet, metformin
360 prevented weight gain – an effect not seen in mice lacking GDF15 or lacking the GDF
361 receptor (GFRAL1)(10). These results establish in mice that the weight benefit observed
362 with metformin treatment was mediated by metformin associated increase in GDF15. In the
363 CAMERA trial of 74 non-diabetic participants there was a weak correlation between serum
364 GDF15 concentrations and weight loss in metformin treated individuals(10). However, our
365 data do not support a role for GDF15 in mediating the weight benefits of metformin as,
366 unlike for leptin, we show no attenuation of the GDF15 association with metformin when
367 adjusting for weight change.

368 Our tissue enrichment analysis identified a set of 8 proteins originating from the intestine as
369 the strongest tissue contributing to the metformin signature: **REG4, GDF15, REG1A,**
370 **IGFBP2, TFF3, SPINK1, Gal-4, PlgR,** with all but IGFBP2 and PlgR identified in two or
371 more studies. Two are Regenerating gene (REG) proteins – REG4 and REG1A – which are a
372 part of the calcium-dependent (C-type) lectin superfamily(21). These proteins have been

373 shown to be responsible for triggering cellular proliferation and are associated with some
374 malignancies such as colorectal cancer(22). REG4 can act as a marker for both
375 enteroendocrine cells and Paneth cells in the small intestine(23) and deep crypt secretory
376 cells (the colon equivalent of Paneth cells) in the colon(24), and has been shown to modulate
377 intestinal inflammation and is associated with ulcerative colitis and Crohn's disease(21).
378 Trefoil Factor Peptide 3 (TFF3) has a role in colonic epithelial homeostasis and response to
379 gastrointestinal inflammation and mucosal injury(25). Increased TFF3 in mouse hepatocytes
380 has been shown to cause inhibition of genes involved in gluconeogenesis such as PEPCK,
381 G6pc and PGC-1 α , reducing hepatic glucose output(26). Moreover, adenoviral
382 overexpression of TFF3 was shown to improve glucose tolerance and insulin sensitivity in
383 diabetic mice(26). In addition, TFF3 has been demonstrated to increase beta cell mass in rat
384 pancreatic islets(27). Serine Protease Inhibitor Kazal Type 1 (SPINK 1) is produced by
385 pancreatic acinar cells and has two main functions; acting as a trypsin inhibitor which acts to
386 protect the pancreas and acting as a cell growth and survival factor which leads to tumour
387 progression(28). Its role in pancreas protection is very important as mutations in the SPINK1
388 gene are associated with different forms of chronic pancreatitis(29).

389 The strong association of these intestinal-related proteins with metformin treatment may
390 simply reflect the high exposure of intestinal epithelial cells to metformin and does not
391 necessarily implicate these proteins as mediating any beneficial or potentially harmful effects
392 of metformin. A pQTL association with a trait could help inform on any causal benefit or
393 harm, although a pQTL in a non-metformin exposed state may differ from a pQTL under
394 metformin treatment. For example, metformin increases GDF-15 and a pQTL SNP
395 associated with increased serum GDF15 (in population level data) was associated with
396 increased neutrophil and lower lymphocytes. This is consistent with GDF15 being associated
397 with adverse conditions (age, cancer, cellular stress) but not consistent with the known effect

398 of metformin to lower neutrophil:lymphocyte ratio(30). Thus, although we do find pQTL
399 associated traits, these need to be interpreted with caution. However, whilst we cannot
400 conclude that the intestinal signature for the metformin proteome mediates metformin, it is
401 important to be aware of these strong associations as they could potentially be major
402 confounders in any proteomic analysis where some people may be metformin treated, and
403 clinically where a protein concentration is being used as a clinical biomarker, such as a
404 tumour marker. For example, REG4 is a postulated tumour marker for pancreatic
405 adenocarcinoma(31), gastric and colorectal cancer(32), EpCAM is a well-known tumour
406 marker associated with many cancers including colorectal, ovarian and breast cancers(33) and
407 REG1A has been recently associated with the development of pancreatic cancer(34).

408 The use of two proteomic platforms in both cross sectional and interventional studies,
409 totalling 2057 participants is a major strength of our study. However, we recognise there are
410 limitations. Firstly, newer proteomic panels include substantially more proteins (e.g. Olink
411 Explore HT measures >5300 proteins, and Somascan measures 7000 proteins). Secondly
412 whilst we combine 3 studies that measure proteins before and after metformin initiation, the
413 number of participants in these longitudinal studies remains small, especially when
414 considering the impact of metformin on BMI change. Finally, whilst we establish a large
415 number of robust signals and for some infer causal association, we don't establish a causal
416 mechanism for any of the associations. This will require further work with, for example,
417 mouse models as has been demonstrated for the mechanistic contribution of GDF15 to
418 metformin action(10).

419 In conclusion, we have carried out a comprehensive study on changes to the proteome
420 following metformin treatment. We identified many proteins that are increased or decreased
421 with metformin treatment, with enrichment for proteins originating from the intestine.
422 Overall, we have shown that the proteomic signature of metformin provides further insight

423 into the mechanism of action, as well as highlighting the potential for false positive signals in
424 human proteomic studies if metformin treatment is not considered as an exposure.

425 **Acknowledgements**

426 We thank Ragna Häussler, Matilda Dale, the Affinity Proteomics Unit at SciLifeLab in
427 Stockholm for generating the Olink data. This project has received funding from the
428 Innovative Medicines Initiative Joint Undertaking 2, under grant agreement no. 115881
429 (RHAPSODY) and from the Innovative Medicines Initiative Joint Undertaking under grant
430 agreement no. 115317 (DIRECT), resources of which are composed of financial contribution
431 from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA
432 companies in kind contribution. There are no relevant conflicts of interest to disclose.

433

434 **Author contributions**

435 BC, LM, RS, JMS, AG, LtH, ERP designed the study and wrote the manuscript. BC, KB,
436 LD, JK, RS, LtH analysed the data. MH, RK, PF and all co-authors critically reviewed the
437 final manuscript. ERP is the guarantor of this work and, as such, had full access to all the data
438 in the study and takes responsibility for the integrity of the data and the accuracy of the data
439 analysis.

440

441 **Data Availability**

442 All summary results for all 3 studies are provided in the supplementary excel file. The
443 generated individual level proteomic data and linked clinical data in RHAPSODY (DCS,
444 GoDARTS), DIRECT, S3WP-T2D, IMPOCT and RAMP are considered sensitive patient
445 data and cannot be made publicly available in compliance with the European privacy

446 regulations governed by GDPR and according to limitations included in the informed
447 consents signed by the study participants. Data are available by request to the corresponding
448 authors. Requests should include name and contact details of the person requesting the data,
449 which molecular data and clinical variables are requested and the purpose of requesting the
450 data.

451

452 **Figure Legends**

453 **Figure 1 Effect of metformin on plasma protein levels.** Volcano plot showing if protein
454 concentrations are significantly increased or decreased following metformin treatment in the
455 longitudinal Olink analysis. Estimate (beta coefficient) is plotted on the x axis and $-\log_{10}$ of
456 the unadjusted p-value (calculated from the linear mixed model) is plotted on the y axis.
457 Proteins with an adjusted p value (Bonferroni method) of less than 0.05 are represented by a
458 yellow dot and all other non-significantly changed proteins are represented by a grey dot.
459 Proteins which have an increased concentration following metformin treatment have a
460 positive effect size whereas proteins which have a decreased concentration following
461 metformin treatment have a negative effect size.

462

463 **Figure 2: GTEx Analysis Showing the Tissues of Origin of Proteins significantly altered**

464 **by metformin exposure.** The plot on the left shows the tissues where the mRNA
465 corresponding to the significant proteins are expressed. The plot on the right shows the tissue
466 of origin of the significant proteins. The plot in the middle visualises the significant proteins
467 based upon the longitudinal Olink analysis and whether they are increased or decreased
468 following metformin treatment alongside their tissue of origin. Proteins/gene expression in
469 this figure are ranked in decreasing order of significance. Only 59 of the 68 significant
470 proteins could be included in this analysis due to some proteins missing in the proteomics

471 data. The 9 proteins missing are SEMA7A, GAL-4, LEP, SELE, ADGRG2, CCL15,
472 CD300LG, TIMD4 and PGF.

473 **Figure 3. A. Protein relationships for the 68 proteins associated with metformin**
474 **exposure in the longitudinal Olink study based on StringDB.** A large number of proteins
475 showed experimentally validated interactions shown with with pink lines, were co-expressed
476 (black lines) or were mentioned together (green lines). Experimentally validated interactions
477 are shown in thicker lines (pink, light blue).

478

479 **Figure 4. Comparison of effect size across the three studies** (Only proteins that are shared
480 in at least two studies are shown). X-axis, effect size; y-axis, protein.

481

482

483

484

485 **References**

- 486 1. Rena G, Hardie DG, Pearson ER. The mechanisms of action of metformin.
487 Diabetologia. 2017 Sep 3;60(9).
- 488 2. Florez JC. The pharmacogenetics of metformin. Diabetologia. 2017 Sep;60(9):1648–
489 55.
- 490 3. GoDARTS and UKPDS Diabetes Pharmacogenetics Study Group, Wellcome Trust
491 Case Control Consortium 2, Zhou K, Bellenguez C, Spencer CCA, Bennett AJ, et al.
492 Common variants near ATM are associated with glycemic response to metformin in
493 type 2 diabetes. Nat Genet. 2011 Feb;43(2):117–20.
- 494 4. Zhou K, Yee SW, Seiser EL, van Leeuwen N, Tavendale R, Bennett AJ, et al.
495 Variation in the glucose transporter gene SLC2A2 is associated with glycemic
496 response to metformin. Nat Genet. 2016;48(9):1055–9.
- 497 5. García-Calzón S, Perfilyev A, Martinell M, Ustinova M, Kalamajski S, Franks PW, et
498 al. Epigenetic markers associated with metformin response and intolerance in drug-
499 naïve patients with type 2 diabetes. Sci Transl Med. 2020 Sep 16;12(561).
- 500 6. Ustinova M, Anson L, Silamikelis I, Rovite V, Elbere I, Silamikele L, et al. Whole-
501 blood transcriptome profiling reveals signatures of metformin and its therapeutic
502 response. PLoS One. 2020;15(8):e0237400.
- 503 7. Xiao S, Li VL, Lyu X, Chen X, Wei W, Abbasi F, et al. Lac-Phe mediates the effects
504 of metformin on food intake and body weight. Nat Metab. 2024 Mar 18;
- 505 8. Zhang Q, Hu N. <p>Effects of Metformin on the Gut Microbiota in Obesity and Type
506 2 Diabetes Mellitus</p>. Diabetes Metab Syndr Obes. 2020 Dec;Volume 13:5003–14.
- 507 9. Gerstein HC, Pare G, Hess S, Ford RJ, Sjaarda J, Raman K, et al. Growth
508 Differentiation Factor 15 as a Novel Biomarker for Metformin. Diabetes Care. 2017
509 Feb 1;40(2):280–3.

- 510 10. Coll AP, Chen M, Taskar P, Rimmington D, Patel S, Tadross JA, et al. GDF15
511 mediates the effects of metformin on body weight and energy balance. *Nature*. 2020
512 Feb 20;578(7795):444–8.
- 513 11. Gummesson A, Björnson E, Fagerberg L, Zhong W, Tebani A, Edfors F, et al.
514 Longitudinal plasma protein profiling of newly diagnosed type 2 diabetes.
515 *EBioMedicine*. 2021 Jan;63:103147.
- 516 12. Hébert HL, Shepherd B, Milburn K, Veluchamy A, Meng W, Carr F, et al. Cohort
517 Profile: Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS). *Int*
518 *J Epidemiol*. 2018 Apr 1;47(2):380–381j.
- 519 13. van der Heijden AA, Rauh SP, Dekker JM, Beulens JW, Elders P, ‘t Hart LM, et al.
520 The Hoorn Diabetes Care System (DCS) cohort. A prospective cohort of persons with
521 type 2 diabetes treated in primary care in the Netherlands. *BMJ Open*. 2017 May
522 6;7(5):e015599.
- 523 14. Koivula RW, Forgie IM, Kurbasic A, Viñuela A, Heggie A, Giordano GN, et al.
524 Discovery of biomarkers for glycaemic deterioration before and after the onset of type
525 2 diabetes: descriptive characteristics of the epidemiological studies within the IMI
526 DIRECT Consortium. *Diabetologia*. 2019 Sep 15;62(9):1601–15.
- 527 15. Eldjarn GH, Ferkingstad E, Lund SH, Helgason H, Magnusson OT, Gunnarsdottir K,
528 et al. Large-scale plasma proteomics comparisons through genetics and disease
529 associations. *Nature*. 2023 Oct;622(7982):348–58.
- 530 16. Piening BD, Zhou W, Contrepois K, Röst H, Gu Urban GJ, Mishra T, et al. Integrative
531 Personal Omics Profiles during Periods of Weight Gain and Loss. *Cell Syst*. 2018
532 Feb;6(2):157-170.e8.
- 533 17. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-
534 Tissue Expression (GTEx) project. *Nat Genet*. 2013 Jun 29;45(6):580–5.

- 535 18. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al.
536 Proteomics. Tissue-based map of the human proteome. *Science*. 2015 Jan
537 23;347(6220):1260419.
- 538 19. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The
539 STRING database in 2021: customizable protein-protein networks, and functional
540 characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*. 2021 Jan
541 8;49(D1):D605–12.
- 542 20. Wang D, Day EA, Townsend LK, Djordjevic D, Jørgensen SB, Steinberg GR. GDF15:
543 emerging biology and therapeutic applications for obesity and cardiometabolic disease.
544 *Nat Rev Endocrinol*. 2021 Oct 11;17(10):592–607.
- 545 21. Tsuchida C, Sakuramoto-Tsuchida S, Taked M, Itaya-Hironaka A, Yamauchi A, Misu
546 M, et al. Expression of REG family genes in human inflammatory bowel diseases and
547 its regulation. *Biochem Biophys Rep*. 2017 Dec;12:198–205.
- 548 22. Zheng H chuan, Sugawara A, Okamoto H, Takasawa S, Takahashi H, Masuda S, et al.
549 Expression Profile of the *REG* Gene Family in Colorectal Carcinoma. *Journal of*
550 *Histochemistry & Cytochemistry*. 2011 Jan 23;59(1):106–15.
- 551 23. Grün D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, et al. Single-cell
552 messenger RNA sequencing reveals rare intestinal cell types. *Nature*. 2015 Sep
553 10;525(7568):251–5.
- 554 24. Sasaki N, Sachs N, Wiebrands K, Ellenbroek SIJ, Fumagalli A, Lyubimova A, et al.
555 *Reg4*⁺ deep crypt secretory cells function as epithelial niche for *Lgr5*⁺ stem cells in
556 colon. *Proceedings of the National Academy of Sciences*. 2016 Sep 13;113(37).
- 557 25. Aihara E, Engevik KA, Montrose MH. Trefoil Factor Peptides and Gastrointestinal
558 Function. *Annu Rev Physiol*. 2017 Feb 10;79(1):357–80.

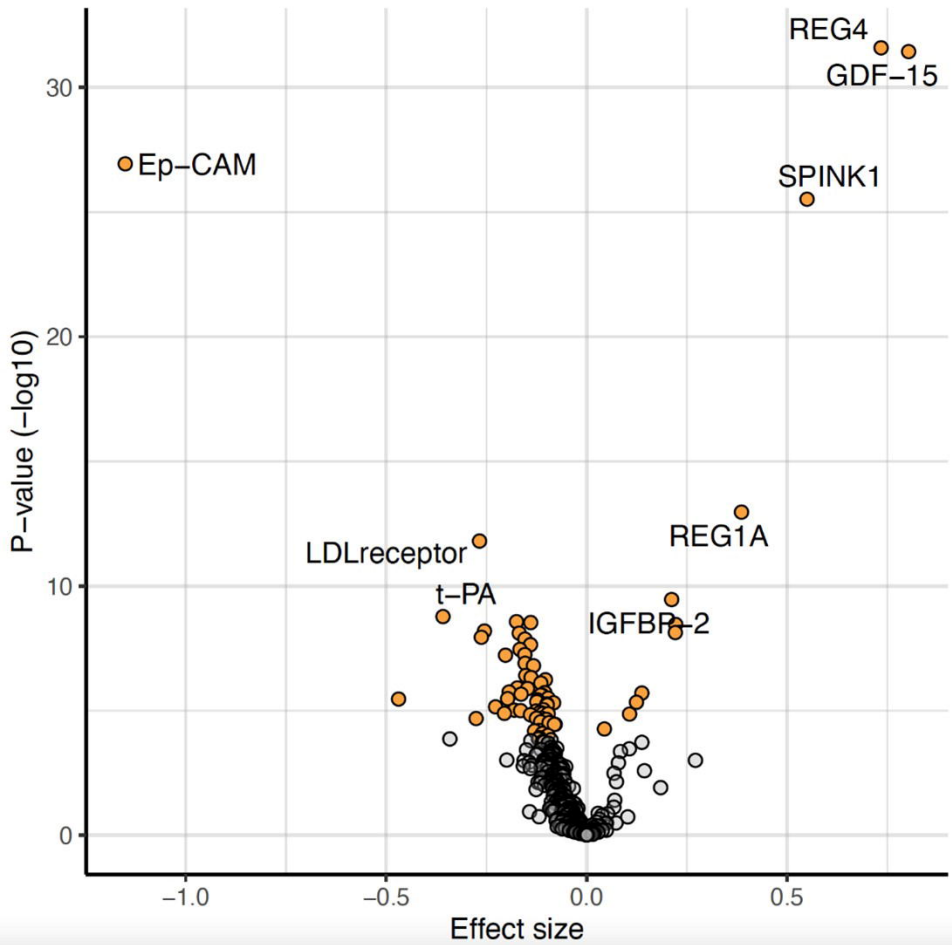
- 559 26. Xue Y, Shen L, Cui Y, Zhang H, Chen Q, Cui A, et al. Tff3, as a Novel Peptide,
560 Regulates Hepatic Glucose Metabolism. *PLoS One*. 2013 Sep 23;8(9):e75240.
- 561 27. Fueger PT, Schisler JC, Lu D, Babu DA, Mirmira RG, Newgard CB, et al. Trefoil
562 Factor 3 Stimulates Human and Rodent Pancreatic Islet β -Cell Replication with
563 Retention of Function. *Molecular Endocrinology*. 2008 May 1;22(5):1251–9.
- 564 28. Mehner C, Radisky ES. Bad Tumors Made Worse: SPINK1. *Front Cell Dev Biol*.
565 2019 Feb 4;7.
- 566 29. Pfützer RH, Barmada MM, Brunskill APJ, Finch R, Hart PS, Neoptolemos J, et al.
567 SPINK1/PSTI polymorphisms act as disease modifiers in familial and idiopathic
568 chronic pancreatitis. *Gastroenterology*. 2000 Sep;119(3):615–23.
- 569 30. Rena G, Mordi IR, Lang CC. Metformin: still the sweet spot for CV protection in
570 diabetes? *Curr Opin Pharmacol*. 2020 Oct;54:202–8.
- 571 31. Takayama R, Nakagawa H, Sawaki A, Mizuno N, Kawai H, Tajika M, et al. Serum
572 tumor antigen REG4 as a diagnostic biomarker in pancreatic ductal adenocarcinoma. *J*
573 *Gastroenterol*. 2010 Jan 30;45(1):52–9.
- 574 32. Hu Y, Pan C, Hu J, Zhang S. The role of Reg IV in colorectal cancer, as a potential
575 therapeutic target. *Współczesna Onkologia*. 2015;4:261–4.
- 576 33. Mohtar M, Syafruddin S, Nasir S, Low TY. Revisiting the Roles of Pro-Metastatic
577 EpCAM in Cancer. *Biomolecules*. 2020 Feb 7;10(2):255.
- 578 34. Lyu J, Jiang M, Zhu Z, Wu H, Kang H, Hao X, et al. Identification of biomarkers and
579 potential therapeutic targets for pancreatic cancer by proteomic analysis in two
580 prospective cohorts. *Cell Genomics*. 2024 May;100561.

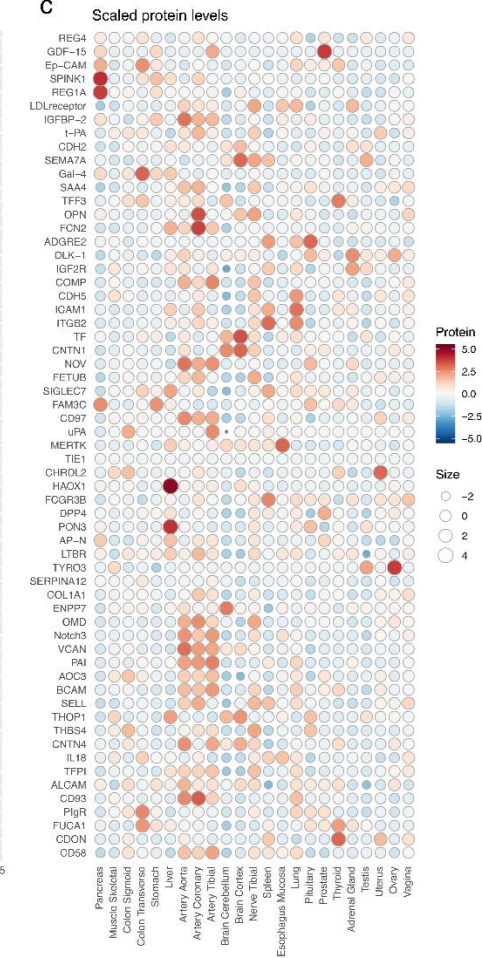
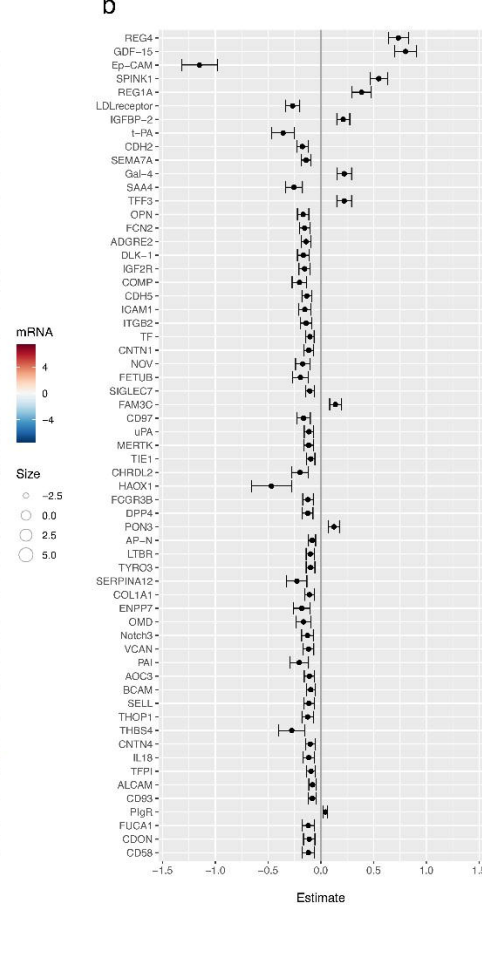
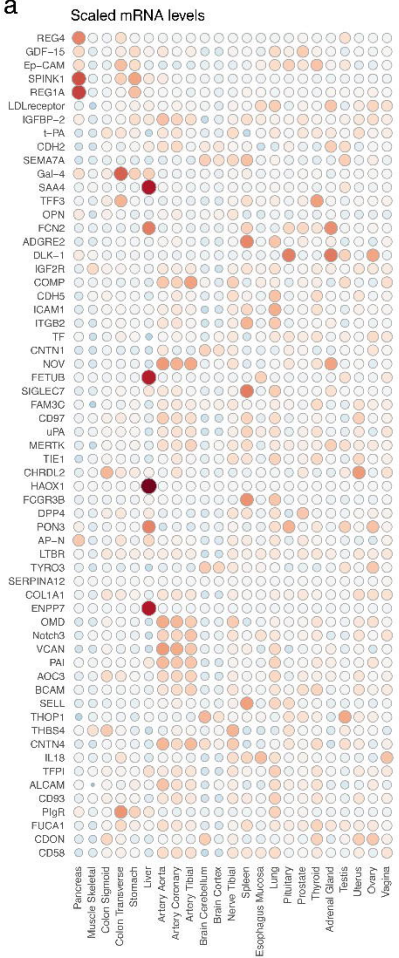
581

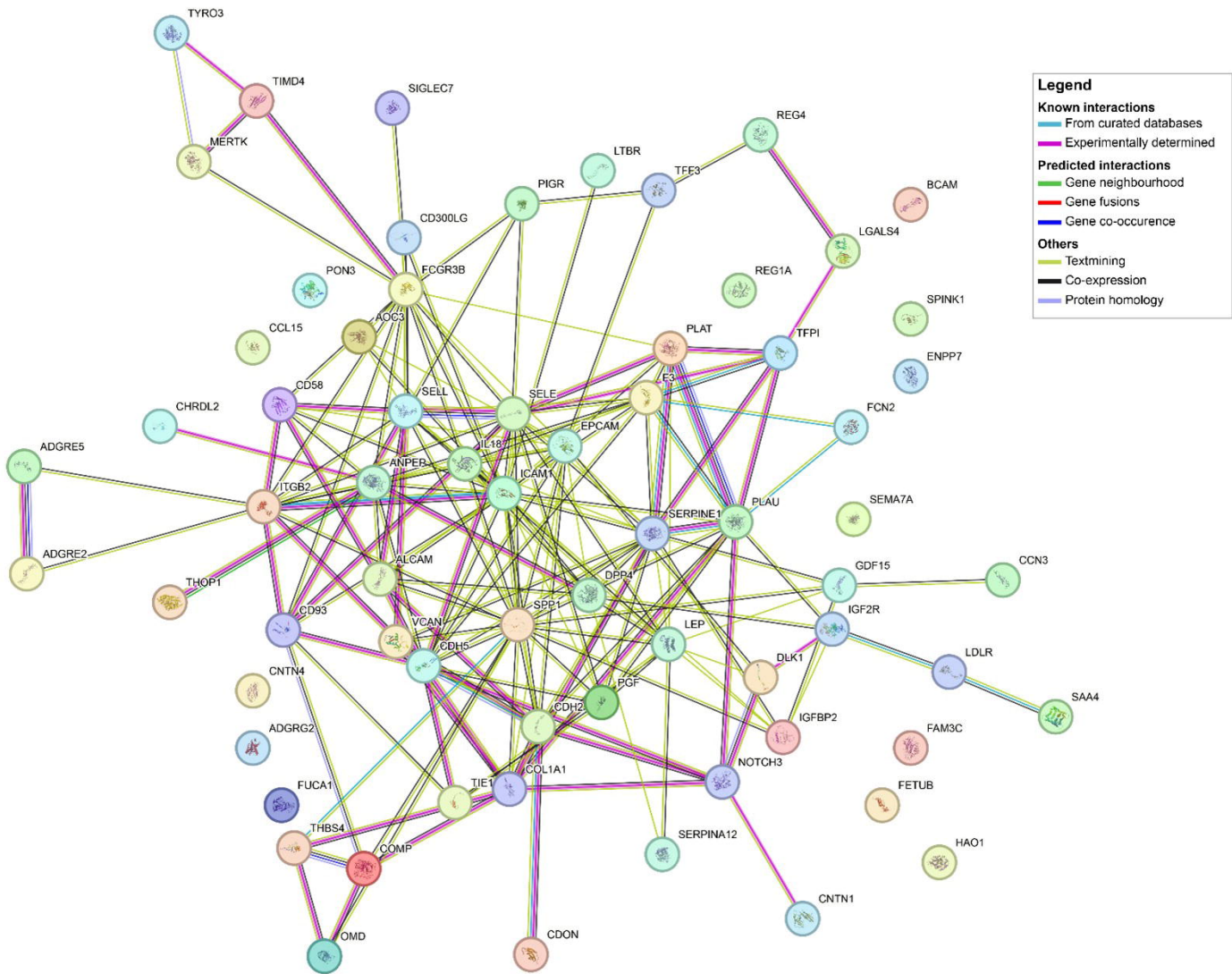
582

583

584







Protein

