

1 **Skeletal muscle and intermuscular adipose tissue gene expression profiling identifies new**  
2 **biomarkers with prognostic significance for insulin resistance progression and intervention**  
3 **response**

4  
5 Dominik Lutter<sup>1,2,\*</sup>, Stephan Sachs<sup>3</sup>, Marc Walter<sup>2,3</sup>, Anna Kerege<sup>4</sup>, Leigh Perreault<sup>4</sup>, Darcy E  
6 Kahn<sup>4</sup>, Amare D Wolide<sup>1,2,5</sup>, Maximilian Kleinert<sup>2,6,7</sup>, Bryan C Bergman<sup>4</sup>, Susanna Hofmann<sup>2,3,8,\*</sup>

7  
8 <sup>1</sup>Computational Discovery Research, Institute for Diabetes and Obesity (IDO), Helmholtz Diabetes  
9 Center (HDC), Helmholtz Zentrum München – German Research Center for Environmental Health,  
10 Neuherberg, Germany

11 <sup>2</sup>German Center for Diabetes Research (DZD), Neuherberg, Germany

12 <sup>3</sup>Institute for Diabetes and Regeneration (IDR-H), Helmholtz Zentrum München, German Research  
13 Center for Environmental Health, Neuherberg, Germany;

14 <sup>4</sup>University of Colorado Anschutz Medical Campus, Aurora, Colorado;

15 <sup>5</sup>Division of Metabolic Diseases, Department of Medicine, Technische Universität München, TUM

16 <sup>6</sup> Drug Development Unit, Institute for Diabetes and Obesity (IDO), Helmholtz Diabetes Center  
17 (HDC), Helmholtz Zentrum München – German Research Center for Environmental Health,  
18 Neuherberg, Germany

19 <sup>7</sup> Dife, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany

20 <sup>8</sup>Medizinische Klinik and Poliklinik IV, Ludwig-Maximilians University, Munich, Germany

21

22

23

24 \*Corresponding authors:

25 Dominik Lutter, PhD

26 Helmholtz Zentrum München

27 Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH)

28 Ingolstädter Landstraße 1 · D-85764 Neuherberg

29 [dominik.lutter@helmholtz-muenchen.de](mailto:dominik.lutter@helmholtz-muenchen.de)

30

31 Susanna Hofmann, MD

32 Helmholtz Zentrum München

33 Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH)

34 Ingolstädter Landstraße 1 · D-85764 Neuherberg

35 [susanna.hofmann@helmholtz-muenchen.de](mailto:susanna.hofmann@helmholtz-muenchen.de)

## 36 **Summary**

37 Although insulin resistance often leads to Type 2 Diabetes Mellitus (T2D), its early stages remain  
38 often unrecognized thus reducing the probability of successful prevention and intervention.  
39 Moreover, treatment efficacy is affected by the genetics of the individual patient. We used gene  
40 expression profiles from a cross-sectional study to identify potential candidate genes for the  
41 prediction of diabetes risk and intervention response. Using a multivariate regression model, we  
42 linked gene expression profiles of human skeletal muscle and intermuscular adipose tissue (IMAT)  
43 to fasting glucose (FG) and glucose infusion rate (GIR). Predictive potential of identified candidate  
44 genes was validated with muscular gene expression data from a longitudinal intervention study. We  
45 found that genes with a strong association to clinical measures clustered into three distinct  
46 expression patterns. Their predictive values for insulin resistance varied strongly between muscle  
47 and IMAT. Moreover, we discovered that individual gene expression based classifications may  
48 differ from those classifications based predominantly on clinical parameters indicating a potential  
49 incomplete patient stratification. Out of the 15 top hit candidate genes, we identified *ST3GAL2*,  
50 *AASS*, *ARF1* and the transcription factor *SIN3A* as novel candidates for a refined diabetes risk and  
51 intervention response prediction. Our results confirm that disease progression and a successful  
52 intervention depend on individual genetics. We anticipate that our findings may lead to a better  
53 understanding and prediction of the individual diabetes risk and may help to develop individualized  
54 intervention strategies.

55

56

## 57 **Introduction**

58 Obesity is a frequent precondition for the development of chronic metabolic diseases like insulin  
59 resistance (IR) and type 2 diabetes (T2D). Based on the recently published results from the 2017–  
60 2018 National Health and Nutrition Examination Survey (NHANES) 42.5% of U.S. adults are  
61 currently obese and are thus all individuals at high risk for developing T2D and its complications  
62 [1]. Moreover, the International Diabetes Federation (IDF) predicts an 51% increase from 463 to  
63 700 million individuals with diabetes worldwide by 2045 and indicates that 1 in 2 adults with  
64 diabetes remain undiagnosed at presence [2]. Although current assessment of diabetes and  
65 prediabetes is based on purely glycemetic indicators it is important to emphasize that the risk for  
66 developing diabetes is also dependent on age, sex, fat tissue distribution, genetic, environmental,  
67 and ethnic characteristics. Depending on these individual risk factors and on the inclusion criteria of  
68 the studied cohorts a wide heterogeneity in the progression from prediabetes to diabetes has been  
69 observed. Emerging evidence from a population-based study with 381,363 participants indicates  
70 that even people referred to as metabolically healthy obese are at a substantially higher risk to

71 develop diabetes and its complications [3]. Although interventions, medically or by change of  
72 lifestyle (diet, exercise) reduce the risk for severe complications evidence is emerging in  
73 population-based cohorts that treatment efficacy also depends on individual's genetics [4-6].  
74 Meaning that patients treated with glucose-lowering interventions vary in their response with some  
75 gaining a considerable benefit, others no benefit; and some getting limiting side effects. The  
76 prevalent opinion in the field is that there are many factors impacting this variation, but one key  
77 factor is a heritable trait [7]. Taken together, it is becoming increasingly clear that the current  
78 clinical standards to define the metabolic health status of an individual is not adequate and new  
79 strategies for the effective prevention of diabetes have become critically important to reduce the  
80 impact of this disease burden. A deeper understanding of the individual features and a precise  
81 phenotyping of prediabetes may improve stratification of disease risk and optimize the benefit/risk  
82 ratio and cost-effectiveness of any therapeutic approach for the prevention, intervention and  
83 improvement of T2D.

84 Given that skeletal muscle is responsible for more than 85% of insulin-stimulated whole-body  
85 glucose disposal [8], and that any dysfunction impairing glucose metabolism in this tissue will  
86 affect whole-body glucose homeostasis ultimately contributing to the development of diabetes [9]  
87 mechanistic studies mostly focus on this very tissue to elucidate metabolic adaptation and its  
88 regulation. More recently, evidence is pointing to intermuscular adipose tissue (IMAT)  
89 accumulation as another local regulator of muscular IR and the progression to diabetes [10, 11]. We  
90 thus hypothesized in this study that tissue specific gene expression profiling of muscular and/or  
91 IMAT can achieve a more specific and detailed characterization and classification of individual  
92 physiological states than circulating parameters alone. We further presumed that the expression of  
93 individual genes might (i) allow for the prediction of individual disease related states, (ii) identify  
94 individuals with a high or low risk for diabetes, and (iii) assess beforehand the potential response of  
95 a given individual to a specific treatment strategy.

96 To this end, we investigated herein interactions between gene expression and clinical diabetes  
97 markers in skeletal muscle and IMAT from overweight individuals with and without diagnosed  
98 T2D. We used multivariate regression to model the tissue specific gene expression impact on the  
99 two key IR markers, glucose infusion rate (GIR) during a hyperinsulinemic/euglycemic clamp and  
100 fasting glucose (FG). Out of the 65 top predictive genes, we identified three distinct gene clusters.  
101 Gene expression-based classification of our study participants led to a refined view on an  
102 individual's metabolic state that partly differs from the clinical classification. In an independent  
103 lifestyle and exercise intervention study with diabetic patients, we could validate some selected  
104 candidates and identify four genes with the potential to predict individual intervention responses.

105

106

## 107 **Methods**

### 108 **Human transcriptional profiling dataset**

109 Human muscle and IMAT transcriptional profiles were taken from a cross-sectional study  
110 previously published by Sachs et al. [12]. To identify features suitable for characterization of  
111 individual pre-diabetic states and potentially predictive for disease progression we selected all 16  
112 participants with obesity (OB) and diabetes mellitus type two (T2D) for which paired samples were  
113 available. All participants were clinically characterized by determining insulin sensitivity via  
114 glucose infusion rate (GIR, mg/kg/min) during hyperinsulinemic-euglycemic clamp, fasting glucose  
115 (FG, mg/dl), age, BMI, relative fat mass (RelFat, % kg), fat free mass (FFM, kg), height (in), body  
116 weight (BW, kg) (Table 1, Figure 1AB, S1AB).

117

### 118 **Longitudinal Intervention dataset**

119 Seventeen individuals with obesity with and without pre-diabetes were recruited for this  
120 study. Subjects were asked to refrain from planned physical activity for 48 hours before the  
121 metabolic study and were given a standardized diet for 7 days prior to the first and second  
122 metabolic study. After overnight fasting a basal muscle biopsy was taken and followed by metabolic  
123 profiling including a 3-hour hyperinsulinemic euglycemic clamp. After the first metabolic study,  
124 volunteers entered a 12-week supervised weight loss and exercise training intervention. After the 3-  
125 month intervention, subjects transitioned to a 2-week weight maintenance diet. Volunteers continued to  
126 exercise during the weight stabilization period. After completing the intervention and 2-week weight  
127 maintenance period volunteers then repeated the metabolic study with the muscle biopsy and 3-hour  
128 hyperinsulinemic euglycemic clamp (Table 2). Changes in metabolic parameters pre to post  
129 intervention were estimated using a paired t-test. Pre and post intervention biopsies were used for gene  
130 expression analysis. Due to the problem that RNA isolation of all IMAT samples tissue material was  
131 not sufficient or RNA RIN values did not match quality requirements for gene expression analysis  
132 we were forced to remove all IMAT samples and used only the remaining muscle samples for gene  
133 expression analysis. Pre to post differential gene expression was estimated using one way ANOVA.

134

### 135 **Models and Statistics**

136 To estimate the impact of gene expression on FG and GIR we utilized linear multivariate regression  
137 models. Thus we created a predictive model for each gene, simultaneously predicting clinical  
138 parameters based on gene expression in muscular and IMAT tissue. Models can be formalized in  
139 matrix notation as:  $\mathbf{Y} = \beta\mathbf{X}_j + \epsilon$ , where  $\mathbf{Y}$  is a matrix of the sampled response variables GIR (g) and  
140 FG (f) and  $\mathbf{X}$  a matrix of the predictor values, the expression of gene  $j$  in muscle (m) and IMAT (i).

141  $\beta$  forms the 2x2 matrix of the four estimated regression coefficients  $\beta_{mg}$ ,  $\beta_{mf}$ ,  $\beta_{ig}$  and  $\beta_{if}$  describing  
142 the four relationships between tissue specific gene expression and response variables. The residues  
143 or errors are formed in  $\epsilon$ . Our model can be interpreted as a mixture model that allows us to jointly  
144 estimate four coefficients to predict insulin sensitivity and glucose homeostasis from gene  
145 expression in muscle and IMAT. Genes that contributed the most to insulin sensitivity and glucose  
146 homeostasis were then scored based on log-likelihood and negative log-likelihood of the single  
147 regression models (S1). Analysis was done using Matlab R2020a.

148 Participant kNN-Networks were generated on cluster wise gene expression profiles with  $k = 3$   
149 nearest neighbors, using Euclidean distance metric. Predictive classification score was calculated as  
150 ratio of each participant direct neighbor's clinical classification.

151

152 Further information, methods and protocols for genetic work and statistics are described in the  
153 Supplementary File S1.

154

155

## 156 **Results**

### 157 **Multivariate regression unravels tissue specific gene expression patterns correlating with** 158 **insulin resistance**

159 We first compared demographic and metabolic parameter of participants. As expected, we found  
160 gender specific differences for RelFat ( $p < 0.05$ ), FFM ( $p < 0.001$ ) and height ( $p < 0.01$ ) and for  
161 participants with obesity and T2D we found significant differences in GIR ( $p < 0.05$ ) and FG  
162 ( $p < 0.001$ ) (Table 1, Figure 1AB, S1A). Additionally, as expected, we found that low GIR values  
163 correlate with high FG levels (Figure S1B). As shown in Figure 1B, individuals classified as OB  
164 ( $BMI > 30$  &  $FG < 125$  mg/dl) exhibited a wide range of insulin sensitivity (GIR 0.8 – 11.1  
165 mg/kg/min). Some individuals with strong progressed insulin sensitivity ( $< 3$  mg/kg/min) were still  
166 able to maintain FG levels below 125 mg/dl. In contrast, some diabetic individuals exhibited a  
167 better GIR with levels up to 4 mg/kg/min while being diagnosed as diabetic. To explore whether  
168 transcriptional changes in muscle and IMAT tissue at the transition from obesity with compensated  
169 insulin resistance (normoglycemia) to T2DM (hyperglycemia) may reflect the inconsistency  
170 between FG and GIR we performed a multivariate regression analysis to identify genes with a  
171 strong expressional link to insulin resistance and glucose homeostasis.

172 After multivariate regression analysis we selected the 59 top genes (Figure S1C) contributing to  
173 GIR and FG and used k-means to cluster them into three clusters of 14, 23 and 22 genes with  
174 distinct expression patterns in muscle and IMAT (Figure 1C, S1D, Table S1). We compared the  
175 gene cluster corresponding  $\beta$  values and observed three distinct patterns (Figure 1D, S1E): Cluster 1

176 (blue) combines genes whose expression is positively associated with GIR and negatively with FG  
177 in both tissues, muscle and IMAT, thus correlating with a healthy glucose metabolism. Cluster 2  
178 contains genes that positively correlate with GIR and negatively with FG in muscle similarly to  
179 cluster 1, while no such effect was observed for these genes in IMAT. Cluster 3 to the contrary,  
180 contains genes with an opposing effect to GIR and FG exclusively in muscle while again no such  
181 impact was observed in IMAT.

182 Our results suggest that these largely different gene expression profiles of muscle and IMAT are  
183 associated with varying impact on glucose metabolism changes. In particular, expression of *PDK4*,  
184 which has been linked to diabetes and glucose metabolism previously [13], shows a high correlation  
185 with GIR and FG in muscle but almost none in IMAT (Figure 1D). We also identified genes with  
186 opposing effects on glucose metabolism in muscle compared to IMAT, such as *UBTD1* and  
187 *ST3GAL2*: Both genes show strong positive coefficients with FG and negative coefficients with GIR  
188 in muscle whereas in IMAT we observe negative  $\beta$  values with FG and positive with GIR (Figure  
189 1D). This effect was mainly associated with cluster 3. In contrast, *NAPB* from cluster 2, shows  
190 exactly opposite associations. In a third observation we found genes, here represented by *SIN3A*,  
191 that seem to have a relatively high predictive value for FG but low or none for GIR in muscle but  
192 completely opposing values in IMAT (high on GIR low on FG).

193 Taken together, the here identified genes in three clusters show a striking association of muscle gene  
194 expression with GIR and FG, while only genes in cluster 1 also associate IMAT gene expression  
195 with glucose homeostasis and insulin sensitivity (Figure 1D, S1E). These results suggest that  
196 muscular gene expression profiles allow for a more specific and detailed characterization and  
197 classification of individual physiological states than serum based physiological parameters alone are  
198 able to.

199

## 200 **Gene expression based classification enables a refined view on the individual physiological** 201 **state of obese patients.**

202 To test our hypothesis that gene expression patterns are superior in categorizing individual insulin  
203 resistance states compared to conventional clinical markers we performed a k-nearest-neighbor  
204 (kNN) classification for each tissue and gene cluster solely based on expression profiles. Thus, we  
205 generated six nearest neighbor networks (NNNs) representing expression based participant  
206 similarities for all 16 individuals (Figure 2A). Based on direct network neighbors we then  
207 calculated a predictive classification score for each individual (Figure 2A, S2). For muscle, we  
208 found for five of the 16 participants (two OB: Pb029, Pb043; three T2D: Pb034, PB053, Pb032) a  
209 non-unique classification over all three NNN. After averaging over the three clusters, one  
210 participant with obesity was classified as T2D and two participants with T2D were classified as OB.



211 For the IMAT tissue derived NNNs we found for seven participants a predicted classification that  
212 differs from the clinical (three OB: Pb048, Pb028, Pb043; four T2D: Pb034, Pb053, Pb033, Pb032).  
213 IMAT gene expression averaged over the clusters classified two participants different from clinical  
214 classification: Pb043 (OB) and Pb033 (T2D). Averaging over both tissues, we identified two  
215 participants with divergent classification: Pb043 and Pb053. Over all, classification for OB  
216 classified participants was more consistent than for T2D with only two participants consistently  
217 classified as T2D (Pb036, Pb052).

218 When comparing NNN based classification to metabolic parameters, we found that the estimated  
219 probability of being predicted as T2D correlates with decreasing GIR rates (Figure 2B) for OB  
220 participants in both, muscle and IMAT tissue. In contrast, classification for hyperglycemic  
221 participants clinically classified as diabetic ( $FG > 125\text{mg/dl}$ ), did not correlate with either GIR or  
222 FG in both tissues. These results suggest that beyond a binary clinical classification of T2D with FG  
223 levels  $> 125\text{ mg/dl}$ , there is a continuous development from insulin resistance to diabetes that  
224 follows individual traces with an individual diagnostic potential to predict a high or low risk for  
225 diabetes or the response interventions.

226

### 227 ***ST3GAL2*, *SIN3A*, *ARF1* and *AASS* mRNA levels in muscle tissue predict intervention** 228 **response**

229 To evaluate whether gene expression profiles within muscle and/or IMAT define individual health  
230 states with predictive potential for disease progression or modulation of insulin sensitivity, we  
231 analyzed 17 human individuals with obesity, with and without prediabetes, undergoing a combined  
232 weight loss and exercise training intervention study (Table 2). Clinical parameters such as GIR, FG,  
233 BW, RelFat, FFM and BMI were measured pre and post intervention. Almost all individuals showed  
234 an increase of GIR ( $p = 2.2 \cdot 10^{-5}$ ) after the intervention and a decrease of BMI ( $p = 2.3 \cdot 10^{-8}$ ), BW  
235 ( $p = 6.1 \cdot 10^{-8}$ ), RelFat ( $p = 2.7 \cdot 10^{-6}$ ) and FFM ( $p = 1.4 \cdot 10^{-6}$ ). A general change of FG levels upon  
236 intervention could not be observed ( $p = 0.12$ ) (Figure 3A). However, when correlating the relative  
237 pre/post change ( $\Delta\%$ ) between all clinical parameters we found, a change of GIR did only  
238 significantly correlate with a change in BW (Figure S3). A decrease of FG in turn, was significantly  
239 correlated with a relative decrease in BMI, FFM and BW.

240 Subsequently, to test if these physiological changes are linked to individual gene expression in  
241 muscle, biopsies taken before and after intervention were used for RNA expression analysis. Six pre  
242 and eight post intervention muscle samples did not match quality requirements and were removed  
243 from subsequent analyses. Out of the previously selected 59 top genes initially identified in our first  
244 patient cohort, we combined various criteria to select 15 candidate genes from all clusters for  
245 validation. Gene expression in muscle tissue was measured with quantitative RT-PCR (Table S1)

246 within this second independent intervention trial. Among those *SIN3A*, *UBTD1*, *ST3GAL2* and  
247 *NAPB* showed notable  $\beta$  value profiles (Figure 1D). *AASS*, *DBNDD1*, *PDK4*, *PIGA*, *POLR3GL*,  
248 *SNAP23*, *SPCS2*, *SSU72* and *UBTD1* could be linked to diabetes associated SNPs identified with  
249 the TD2M portal [14] and *ARF1*, *BCAT2* and *LDHD* could be linked to skeletal muscle lipid  
250 metabolism and insulin resistance [15-17]. *PDK4* was included as a well described marker for  
251 muscle insulin resistance and as a potential therapeutic target [18]. From these 15 genes, five  
252 (*LDHD*, *ARF1*, *NAPB*, *POLR3GL* and *SNAP23*) showed a significant change in expression between  
253 pre and post intervention (Figure 3B). Since we hypothesized that distinct gene expression states  
254 may refer to individual disease states, we tested selected genes on their predictive potential for  
255 individual intervention response. To this end, we correlated individual pre-intervention gene  
256 expression with the relative change (pre to post) of the clinical parameters BMI, BW, RelFat, FFM,  
257 FG and GIR (Figure 3C).  $\Delta$ GIR and  $\Delta$ RelFat mass could not be significantly with any of the genes  
258 tested. A change of the remaining parameters could be significantly predicted by the genes  
259 *ST3GAL2* (FG, BW, FFM and BMI), *SIN3A* (FG, FFM and BMI), *ARF1* (FG) and *AASS* (FFM)  
260 (Figure 1C, S4, Table S2). In contrast to the five genes that change expression after intervention  
261 (*LDHD*, *ARF1*, *NAPB*, *POLL3GL*, *SNAP23*), none of the four genes identified with predictive  
262 character appeared to be differentially expressed between pre- and post-intervention (Figure 3B).  
263 Together, these findings indicate that individual susceptibility to exercise intervention for the  
264 improvement of glucose homeostasis is independent of the individual clinical parameters, but  
265 correlates with individual gene expression profiles prior to intervention. We next compared these  
266 four identified genes with *PDK4*, a well-described muscle marker for insulin resistance. To our  
267 surprise, *PDK4* was not significant associated with any intervention-induced change of metabolic  
268 parameters (Figure S4).

269 Next, we found that low expression levels of three out of the four genes (*AASS*, *ARF1* and *SIN3A*)  
270 were associated with a good health prognosis. In particular, *ARF1* showed a significant decrease of  
271 expression upon exercise intervention. In turn, *ST3GAL2* was the only gene for which increased  
272 expression levels in muscle tissue increases likeliness of an effective intervention. In summary,  
273 within this independent intervention trial we were able to validate our proposed predictive potential  
274 of muscle gene expression profiles for individual insulin resistance states.

275

## 276 Discussion

277 In this work, we showed that human transcriptional profiles of muscle and IMAT tissue from obese  
278 and diabetic individuals are differentially coupled to insulin resistance and glucose homeostasis. We  
279 identified predictive gene clusters that mirror genetic states reflecting a continuous progression  
280 from early insulin resistance to T2D following individual traits. From a subset, we identified the



281 genes *AASS*, *ARF1*, *SIN3A* and *ST3GAL2* to predict individual exercise intervention response to  
282 improve impaired glucose metabolism.

283 We started our analysis with the observation that there is a wide range of GIR measurements that  
284 overlap between OB and T2D. We hypothesized the binary clinical classification of T2D does not  
285 reflect individual underlying genetics and that specific gene expression patterns of skeletal muscle  
286 and/or IMAT tissue might have the potential to identify and predict individuals with a high or low  
287 risk to develop diabetes or to predict individual susceptibility to interventions.

288 Our multivariate regression analysis, revealed a strong association of muscle to GIR and FG for all  
289 three gene clusters, while IMAT was associated with glucose homeostasis and insulin resistance  
290 only with the genes in cluster 1. Together with the observation that estimates for  $\beta$  coefficients of  
291 cluster 2 were contrary in muscle and IMAT we concluded that IMAT and muscle contribute  
292 differentially to glucose metabolism. However, we see higher variance in IMAT gene expression  
293 [12] which might also weaken a detectable correlation. Reason's for IMAT's increased variability  
294 may arise from technical difficulties to dissect IMAT from muscle resulting in less material for  
295 RNA extraction or a higher heterogeneity in IMAT tissue itself being composed of multiple cell  
296 types such as pre-adipocytes, adipocytes, adipocyte-like cells, myoblasts, stromal and vascular cells.  
297 Participant classification based on kNN-networks revealed that insulin sensitivity could be  
298 accurately predicted for non-diabetic subjects from gene expression patterns, whereas  
299 hyperglycemic subjects were scored differently from clinical classification in several cases. These  
300 observations are consistent with hyperglycemia usually being a consequence of pancreatic beta-cell  
301 failure, and insulin sensitivity is associated with multiple organ malfunctions and, in particular, the  
302 skeletal muscle as the primary organ for glucose uptake [19]. Although GIR measured with  
303 hyperinsulinemic-euglycemic clamp, is still the gold standard to directly measure insulin resistance  
304 it is highly invasive and time consuming and with very limited predictive potential. FG levels by  
305 themselves are unlikely to identify pre-diabetic or obese individuals with impaired insulin  
306 sensitivity, but rather individuals at late stages with an increased risk for irreversible damages of  
307 tissues and organs [20]. We thus conclude, that both parameters are not suitable for reliable for early  
308 diagnosis and disease progression prognosis. In contrast, with the here identified gene expression  
309 profiles that only represent a muscle specific state of individual insulin resistance, we could identify  
310 their predictive potential for the characterization of the individual insulin sensitivity.

311 This predictive potential was finally tested on muscle tissue from an additional independent cohort  
312 of 18 individuals with impaired glucose metabolism undergoing 12 weeks of combined weight loss  
313 and exercise training. By correlating the pre-intervention expression levels of our candidate genes  
314 with the relative change of the clinical parameters upon intervention we identified 4 genes with  
315 significant predictive value: *AASS*, *ARF1*, *SIN3A* and *ST3GAL2*. *AASS*, *ARF1* and *SIN3A* indicated

316 a positive prognosis with a lower level of expression. *AASS* encodes for the enzyme *Amino adipate-*  
317 *Semialdehyde Synthase* that is involved in mammalian lysine degradation and in hyperlysinemia  
318 [21], but was not described in the context of impaired glucose metabolism, insulin resistance or  
319 diabetes yet. Beside its predictive potential we also found that ADP Ribosylation Factor 1 (*ARF1*)  
320 expression was significantly reduced after intervention. The *Ink4/Arf* locus was recently linked to  
321 insulin resistance in mice [22] and ADP-ribosylation factor was linked to insulin signaling in  
322 *Drosophila* [23]. The transcription factor SIN3 Transcription Regulator Family Member A (*SIN3A*)  
323 was recently linked to glucose metabolism in murine  $\beta$ -cells [24]. It could be further shown that  
324 *Sin3a* is an insulin sensitive FOXO1 corepressor of glucokinase in murine livers [25]. Very recently,  
325 *SIN3A* was shown to negatively regulate insulin receptor (*Insr*) mRNA in mice and human muscle  
326 [26]. Finally, ST3 Beta-Galactoside Alpha-2,3-Sialyltransferase 2 (*ST3GAL2*) was the only gene  
327 that was identified to positively predict exercise response with a high expression. Mice lacking the  
328 ST3Gal-II protein, have been shown to develop obesity and insulin resistance after 7-9 month of  
329 age [27]. In summary, 3 of the 4 predictive genes identified by us have already been linked to  
330 insulin resistance and diabetes but their predictive potential has not been explored yet.

331 We herein identified novel markers to predict impaired insulin sensitivity in the human muscle and  
332 found four markers that predict individual exercise intervention response in diabetic patients. These  
333 findings may help to classify and characterize obese, pre-diabetic or diabetic individuals more  
334 precisely than using FG alone. Clamp GIR measurements are highly invasive and may be partly  
335 replaced by use of our marker genes. Additionally we anticipate these findings may also help to  
336 develop precise and individualized interventions strategies for diabetic and obese risk patients.

337

### 338 **Contributors**

339 DL, BCB and SH conceived the research question and designed and planned the study. DL, SS,  
340 MW, AK, LP, DEK, AW and MK prepared the data. DL analyzed the data. DL, BCB and SH wrote  
341 the manuscript. All authors critically revised the manuscript. DL had full access to all the data in the  
342 study and had final responsibility to submit for publication. BCB and SH had access to and verified  
343 the study data.

344

### 345 **Declaration of interests**

346 S.S. is an employee of Cellarity, Inc and has stake-holder interests. The present work was carried  
347 out as an employee of the Helmholtz Zentrum Muenchen, HMGU.

348

349

### 350 **Data sharing**

351 Gene expression profiles and clinical data can be shared upon request to BCB. Statistical code as  
352 not provided in the manuscript or supplementary information is available upon request to DL  
353

354 **Acknowledgements**

355 We gratefully thank Ricardo Berutti for bioinformatics support provided at the Institute of Human  
356 Genetics at the Helmholtz Zentrum München and Shohreh Esmaeili for technical assistance.  
357  
358

359

360

361 **Tables**

362 Table 1. Subject demographics: Human transcriptional Profiling (N = 16)

<b>ClinPar</b>	<b>OB</b>	<b>T2D</b>
<b>N subjects</b>	10	6
<b>Age (years)</b>	40.5 ± 2.44	45.67 ± 2.54
<b>BMI (Kg/m<sup>2</sup>)</b>	36.65 ± 1.63	34.83 ± 1.73
<b>Fasting Glucose (mg/dl)</b>	87.8 ± 3.18	183.17 ± 16.27*
<b>Insulin Sensitivity (mg/kg/min)</b>	5.13 ± 0.96	1.66 ± 0.68*
<b>RelFat (%)</b>	38.23 ± 2.59	37.82 ± 2.68
<b>Fat free mass (kg)</b>	71.92 ± 4.81	62.95 ± 4.52
<b>Height (in)</b>	70.08 ± 1.6	67.15 ± 1.29
<b><u>Body Weight (kg)</u></b>	116.78 ± 7.39	101.42 ± 6.4

363 Values are means±SEM. \* = significantly different between OB and T2D, p<0.05

364

365 Table 2. Subject demographics: Longitudinal Intervention Study (N = 21)

	<b><u>Pre-Intervention</u></b>	<b><u>Post-Intervention</u></b>
<b><u>Age (yrs)</u></b>	46.5±2.2	
<b><u>BMI (kg/m<sup>2</sup>)</u></b>	34.7±1.0	30.7±1.0 *
<b><u>Body Weight (kg)</u></b>	96.9±2.7	85.9±2.6*
<b><u>RelFat (%)</u></b>	41.3±1.5	38.0±1.8 *
<b><u>Fat free mass (kg)</u></b>	56.7±1.8	52.9±1.5 *
<b><u>Fasting glucose (mg/dl)</u></b>	93.2±2.7	90.5±2.6
<b><u>VO<sub>2</sub>peak (L/min)</u></b>	2.2±0.1	2.5±0.1 *
<b><u>Insulin Sensitivity (mg/kg/min)</u></b>	3.5±0.4	5.4±0.5 *

366 Values are means±SEM. \* = significantly different from pre-intervention, p<0.05.

367

368 **Figure Captions**

369 **Figure 1: Multivariate regression analysis unravels tissue specific gene expression patterns**  
370 **correlating with insulin resistance.**

371 (A, B) Boxplot comparing Fasting Glucose and Insulin sensitivity (GIR mg/kg/min) distributions  
372 between subjects of different gender (A) and classifications (B). OB: obese, T2D: type 2 diabetic F:  
373 female, M: male. Red marked dots indicate outliers. (C) Heatmaps of muscle and IMAT genes  
374 correlating with insulin resistance identified by multivariate regression. Colors of left dendrogram  
375 refer to identified clusters 1-3. Bottom color bars indicate individual disease classification, Glucose  
376 Infusion Rate (GIR), Fasting Glucose levels and sex. Vertical color bars show the four estimated  
377 regression coefficients for each gene indicating four relationships between tissues and response  
378 variables (GIR & FGI). (D) Scatterplots comparing gene specific regression coefficients  $\beta$  for GIR  
379 and FGI for muscle and IMAT tissue. Dot colors refers to gene cluster assignment.

380

381 **Figure 2: Gene expression based participant classification reveals a refined view on**  
382 **physiological states.**

383 (A) k-nearest-neighbor networks for the three cluster and two tissues respectively. Nodes refer to  
384 individual subjects. Node outer color refers to assigned clinical classification OB or T2D. Node  
385 inner color refers to estimated disease state based on connected individuals. (B) Scatter plot  
386 displays clinical parameter FG and GIR for all subjects. Node shape refers to assigned clinical  
387 classification OB (diamond) or T2D (dot). Node inner color refers to estimated disease state across  
388 all three gene cluster for muscle (left), IMAT (middle) and a combination of both tissues (right).

389

390 **Figure 3: ST3GAI2, SIN3A, ARF1 and AASS mRNA expression in muscle predict intervention**  
391 **response.**

392 (A) Clinical parameters pre and post intervention. Significance levels refer to paired t test. (B)  
393 mRNA levels of selected genes pre and post intervention. Significant differences in pre/post  
394 expression was estimated using 1 way ANOVA. (C) Correlation volcano plots for pre-intervention  
395 gene expression with relative change of clinical parameters between pre and post intervention.  
396 Significantly correlated mRNAs are colored orange.

397

398

399 **References**

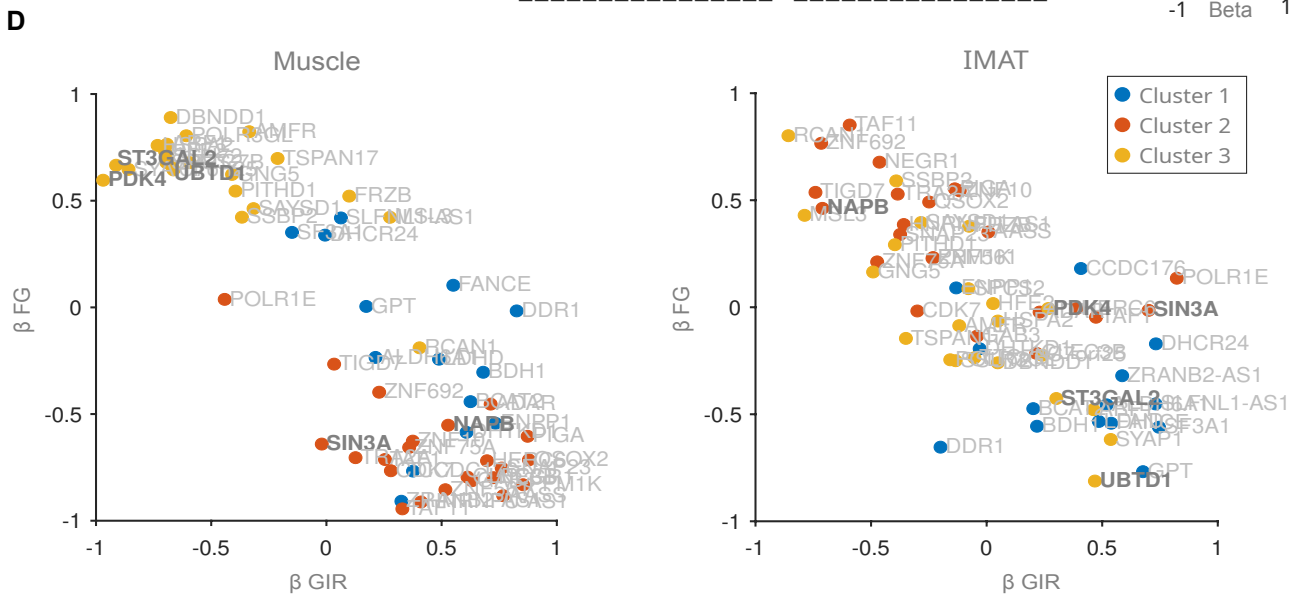
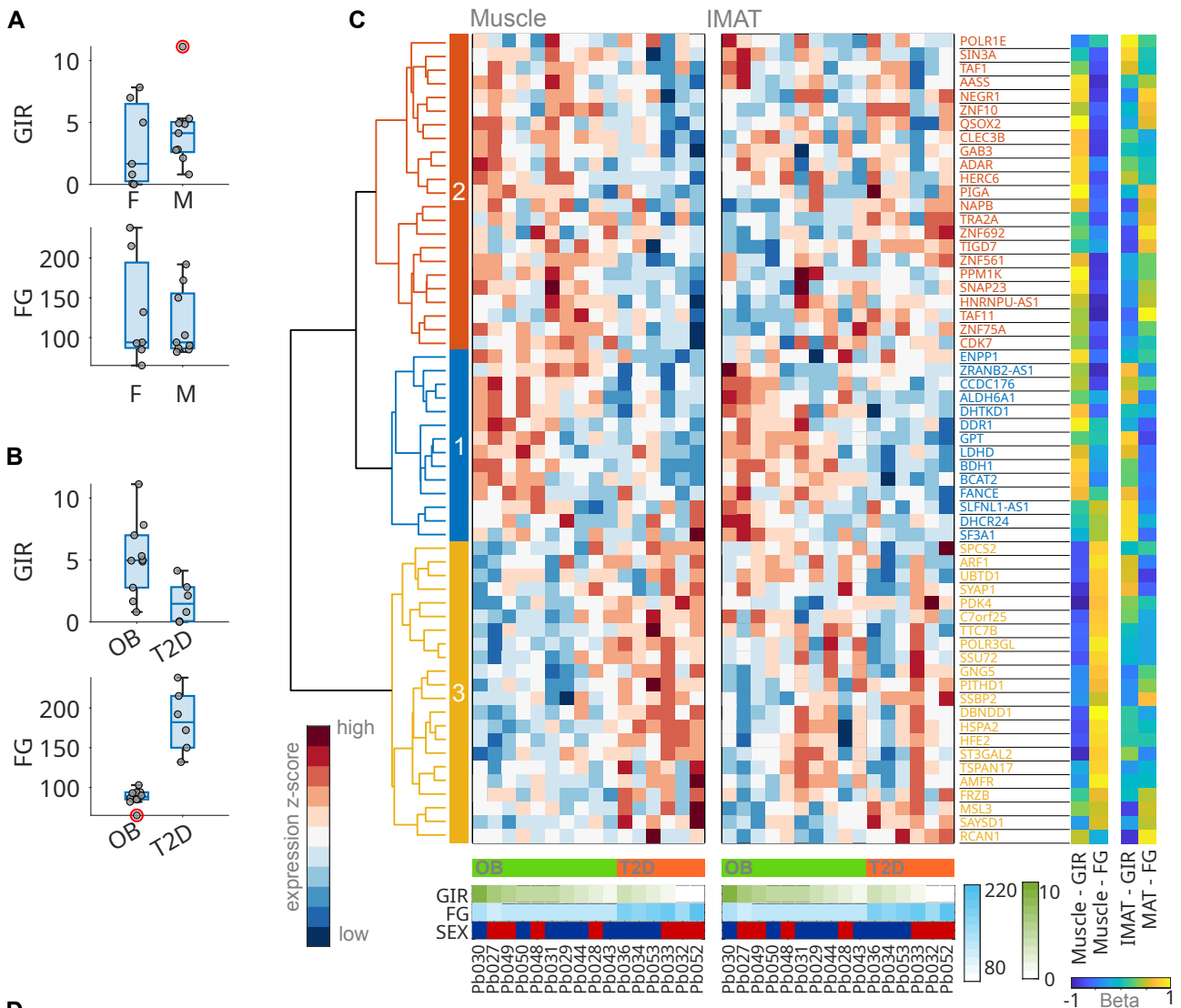
400

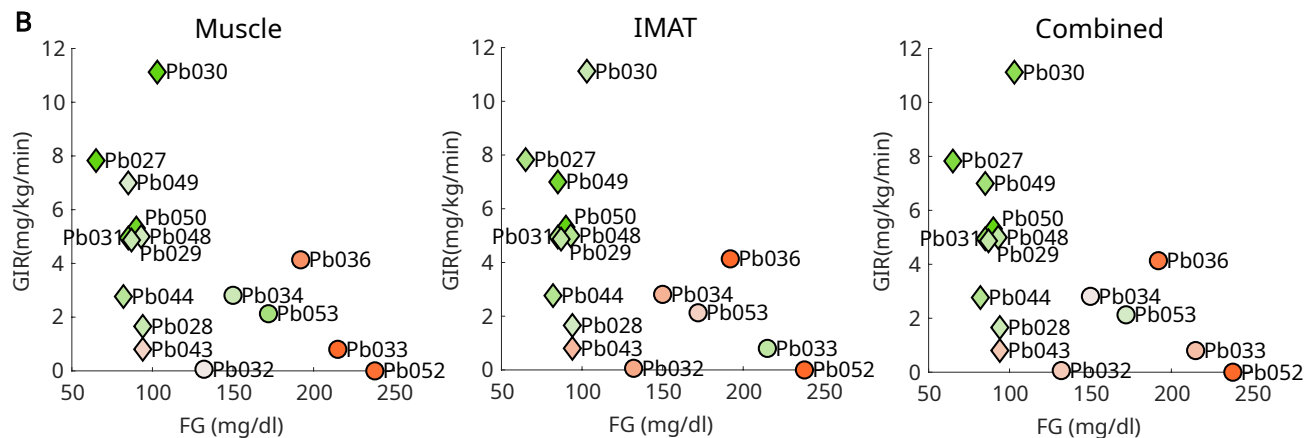
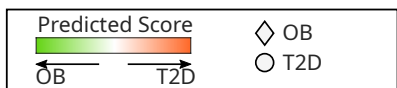
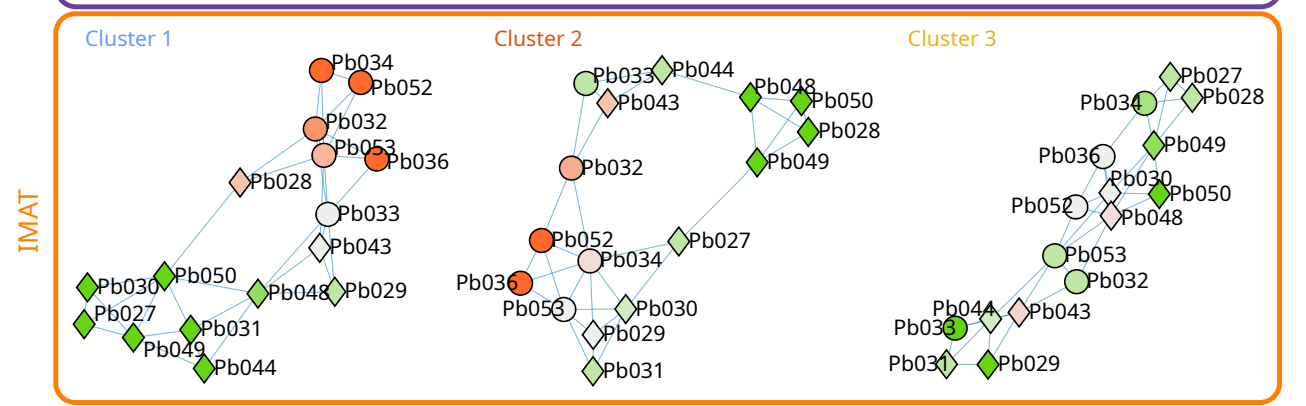
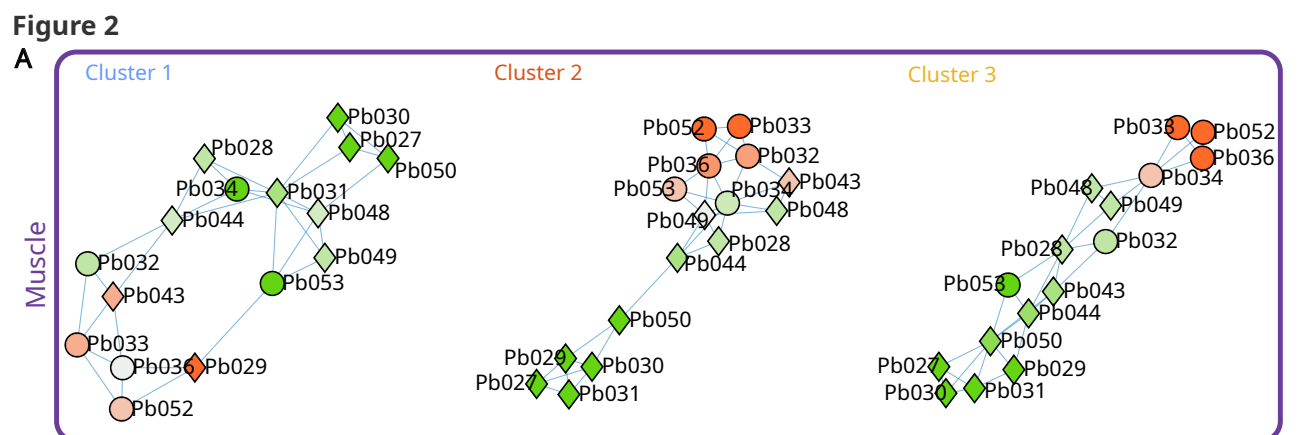
- 401 1. Fryar, C.D., M.D. Carroll, and J. Afful, *Prevalence of overweight, obesity, and severe*  
402 *obesity among adults aged 20 and over: United States, 1960–1962 through 2017–2018.*  
403 NCHS Health E-Stats, 2020.

- 404 2. Federation, I.D., *IDF Diabetes Atlas 9th Edition 2019*. 2019.
- 405 3. Zhou, Z., et al., *Are people with metabolically healthy obesity really healthy? A prospective*  
406 *cohort study of 381,363 UK Biobank participants*. *Diabetologia*, 2021: p. 1-10.
- 407 4. Goni, L., et al., *Effect of the interaction between diet composition and the PPM1K genetic*  
408 *variant on insulin resistance and beta cell function markers during weight loss: results from*  
409 *the Nutrient Gene Interactions in Human Obesity: implications for dietary guidelines*  
410 *(NUGENOB) randomized trial*. *Am J Clin Nutr*, 2017. **106**(3): p. 902-908.
- 411 5. Huls, A., et al., *Polygenic risk for obesity and its interaction with lifestyle and*  
412 *sociodemographic factors in European children and adolescents*. *Int J Obes (Lond)*, 2021.  
413 **45**(6): p. 1321-1330.
- 414 6. Kriska, A.M., et al., *The Impact of Physical Activity on the Prevention of Type 2 Diabetes:*  
415 *Evidence and Lessons Learned From the Diabetes Prevention Program, a Long-Standing*  
416 *Clinical Trial Incorporating Subjective and Objective Activity Measures*. *Diabetes Care*,  
417 2021. **44**(1): p. 43-49.
- 418 7. Pearson, E.R., *Dorothy Hodgkin Lecture 2021: Drugs, genes and diabetes*. *Diabet Med*,  
419 2021. **38**(12): p. e14726.
- 420 8. DeFronzo, R.A., et al., *Synergistic interaction between exercise and insulin on peripheral*  
421 *glucose uptake*. *The Journal of clinical investigation*, 1981. **68**(6): p. 1468-1474.
- 422 9. Garneau, L. and C. Aguer, *Role of myokines in the development of skeletal muscle insulin*  
423 *resistance and related metabolic defects in type 2 diabetes*. *Diabetes Metab*, 2019. **45**(6): p.  
424 505-516.
- 425 10. Correa-de-Araujo, R., et al., *Myosteatosis in the Context of Skeletal Muscle Function*  
426 *Deficit: An Interdisciplinary Workshop at the National Institute on Aging*. *Front Physiol*,  
427 2020. **11**: p. 963.
- 428 11. Boettcher, M., et al., *Intermuscular adipose tissue (IMAT): association with other adipose*  
429 *tissue compartments and insulin sensitivity*. *J Magn Reson Imaging*, 2009. **29**(6): p. 1340-5.
- 430 12. Sachs, S., et al., *Intermuscular adipose tissue directly modulates skeletal muscle insulin*  
431 *sensitivity in humans*. *Am J Physiol Endocrinol Metab*, 2019. **316**(5): p. E866-E879.
- 432 13. Majer, M., et al., *Insulin downregulates pyruvate dehydrogenase kinase (PDK) mRNA:*  
433 *potential mechanism contributing to increased lipid oxidation in insulin-resistant subjects*.  
434 *Molecular genetics and metabolism*, 1998. **65**(2): p. 181-186.
- 435 14. T2DKP, *Type 2 Diabetes Knowledge Portal*. 2018, Nov 15th:  
436 <http://www.type2diabetesgenetics.org>.
- 437 15. Covington, J.D., et al., *Skeletal muscle perilipin 3 and coatomer proteins are increased*  
438 *following exercise and are associated with fat oxidation*. *PLoS One*, 2014. **9**(3): p. e91675.
- 439 16. Zhao, X., et al., *The Relationship between Branched-Chain Amino Acid Related*  
440 *Metabolomic Signature and Insulin Resistance: A Systematic Review*. *J Diabetes Res*, 2016.  
441 **2016**: p. 2794591.
- 442 17. Xu, P., et al., *1640-P: Common and Ethnic-Specific Molecular Networks and Drivers of*  
443 *Insulin Resistance in African and European Ancestry Individuals*. *Diabetes*, 2020.  
444 **69**(Supplement\_1).
- 445 18. Jeon, J.H., et al., *Loss of metabolic flexibility as a result of overexpression of pyruvate*  
446 *dehydrogenase kinases in muscle, liver and the immune system: Therapeutic targets in*  
447 *metabolic diseases*. *J Diabetes Investig*, 2021. **12**(1): p. 21-31.
- 448 19. Goldstein, B.J., *Insulin resistance as the core defect in type 2 diabetes mellitus*. *Am J*  
449 *Cardiol*, 2002. **90**(5A): p. 3G-10G.
- 450 20. Stratton, I.M., et al., *Association of glycaemia with macrovascular and microvascular*  
451 *complications of type 2 diabetes (UKPDS 35): prospective observational study*. *BMJ*, 2000.  
452 **321**(7258): p. 405-12.
- 453 21. Sacksteder, K.A., et al., *Identification of the alpha-amino adipic semialdehyde synthase*  
454 *gene, which is defective in familial hyperlysinemia*. *Am J Hum Genet*, 2000. **66**(6): p. 1736-  
455 43.



- 456 22. Vinue, A., et al., *Ink4/Arf locus restores glucose tolerance and insulin sensitivity by reducing*  
457 *hepatic steatosis and inflammation in mice with impaired IRS2-dependent signalling.*  
458 *Biochim Biophys Acta*, 2015. **1852**(9): p. 1729-42.
- 459 23. Fuss, B., et al., *The cytohesin Steppke is essential for insulin signalling in Drosophila.*  
460 *Nature*, 2006. **444**(7121): p. 945-8.
- 461 24. Yang, X., et al., *Coregulator Sin3a Promotes Postnatal Murine beta-Cell Fitness by*  
462 *Regulating Genes in Ca(2+) Homeostasis, Cell Survival, Vesicle Biosynthesis, Glucose*  
463 *Metabolism, and Stress Response.* *Diabetes*, 2020. **69**(6): p. 1219-1231.
- 464 25. Langlet, F., et al., *Selective Inhibition of FOXO1 Activator/Repressor Balance Modulates*  
465 *Hepatic Glucose Handling.* *Cell*, 2017. **171**(4): p. 824-835 e18.
- 466 26. Cen, H.H., et al., *Transcriptomic analysis of human and mouse muscle during*  
467 *hyperinsulinemia demonstrates insulin receptor downregulation as a mechanism for insulin*  
468 *resistance.* *bioRxiv*, 2021: p. 556571.
- 469 27. Lopez, P.H., et al., *Mice lacking sialyltransferase ST3Gal-II develop late-onset obesity and*  
470 *insulin resistance.* *Glycobiology*, 2017. **27**(2): p. 129-139.
- 471

**Figure 1**



**Figure 3**