Association of genetic risk and physical activity with incident type 2 diabetes

Xuan Zhou[1], Germán D. Carrasquilla[1], Malene R. Christiansen[1], Roelof A.J. Smit[1,2], Lars Ängquist[1], Torben Hansen[1], Ruth J.F. Loos[1,2], Niels Grarup[1], Allan Linneberg[1,3,4], Tuomas O. Kilpeläinen[1], Jordi Merino[1,5]

Affiliations:

[1] Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark,

[2] Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[3] Center for Clinical Research and Prevention, Copenhagen University Hospital - Bispebjerg and Frederiksberg, The Capital Region, Copenhagen, Denmark

[4] Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[5] Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA.

Running title: Gene-physical activity interaction in diabetes

Word count: 3,648 words

Corresponding authors' contact information:
Jordi Merino, Ph.D. Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, 665 Blegdamsvej 3B, DK-2200 Copenhagen, Denmark. Phone: +45 51 42 40 55
Email: jordi.merino@sund.ku.dk

Abstract

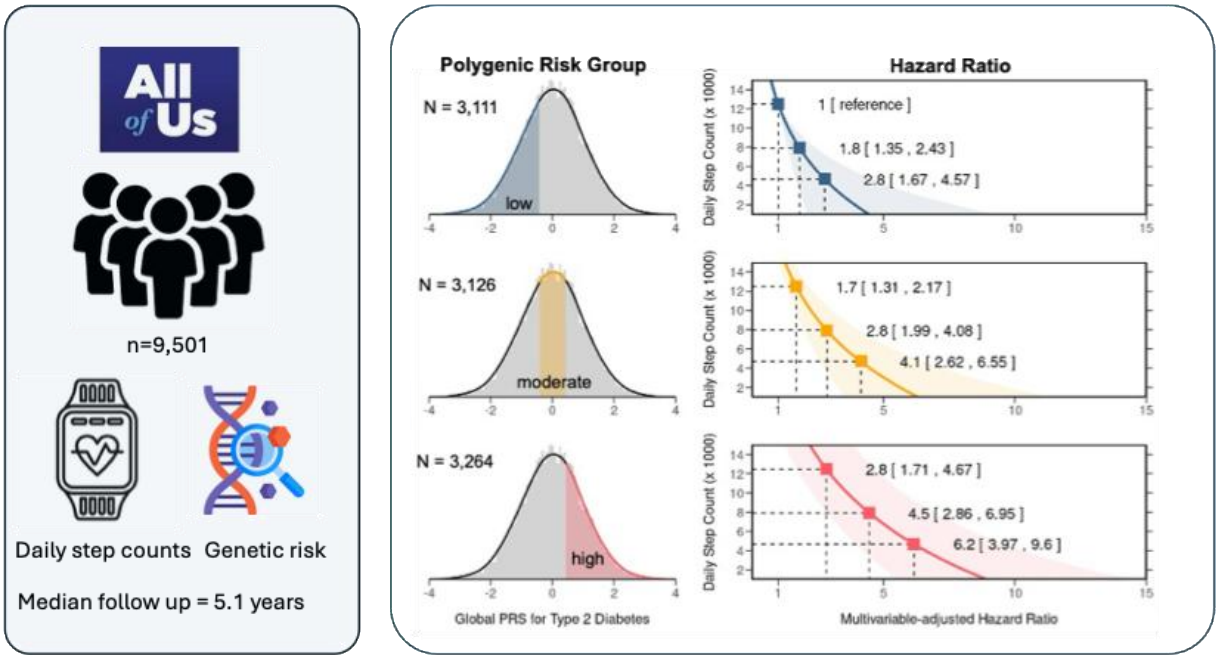
Objective: To assess whether daily step counts and genetic risk interact to influence the risk of developing type 2 diabetes.

Research Design and Methods: We analyzed data from 9,501 participants in the *All of Us* Research Program with both genetic and wearable device–derived physical activity data and without diabetes at baseline. Physical activity was quantified using daily step counts. Genetic risk was assessed using a global polygenic score. Incident type 2 diabetes was identified using electronic health record–linked diagnostic codes. Multivariable Cox proportional hazards models estimated hazard ratios (HRs) for type 2 diabetes across genetic risk and physical activity levels. We tested for additive interaction using the relative excess risk due to interaction (RERI). In secondary analyses we used physical-activity intensity measures using wearable-derived and self-reported intensity levels.

Results: Type 2 diabetes incidence rates ranged from 4.1 per 1,000 person-years (95% CI, 2.5–5.7) in individuals with high physical activity and low genetic risk to 18.4 (95% CI, 15.2–21.6) in those with low physical activity and high genetic risk (HR, 6.2 (95% CI: 3.97, 9.6)). A significant additive interaction was observed (RERI, 0.20; 95% CI, 0.04–0.36; P = .007), with 15% (95% CI, 2–27) of excess risk attributed to the interaction. Similar interaction patterns were found using device-based intensity metrics and self-reported physical activity measures.

Conclusions: These findings provide evidence of additive interactions between genetic risk and physical activity, underscoring the potential value of integrating genomic and device-derived data to identify individuals who would more likely benefit from increasing physical activity.

48 **Graphical abstract**



49

50

51

52

53

54

55

56

57 **Article Highlights**

58 **Why did we undertake this study?**

59 Despite longstanding interest in gene–environment interactions influencing the development of
60 type 2 diabetes, identifying such interactions in human populations remains challenging, partially
61 due to difficulties in accurately measuring environmental exposures. Step count is an objective,
62 interpretable measure of daily ambulatory physical activity that offers a novel opportunity to
63 evaluate these interactions. To date, no studies have specifically investigated how daily step counts
64 may interact with genetic susceptibility to influence type 2 diabetes risk.

65

66 **What is the specific question(s) we wanted to answer?**

67 To what extent do genetic risk and physical activity, measured via daily step count, independently
68 and jointly contribute to the risk of developing type 2 diabetes?

69

70 **What did we find?**

71 We observed a gradient of increasing type 2 diabetes risk associated with high genetic
72 susceptibility and low step count. This gradient deviated from additivity, with the combination of
73 high genetic risk and low physical activity accounting for substantial proportion of excess type 2
74 diabetes risk. The evidence of additive interactions between genetic risk and physical activity
75 was replicated in an independent cohort using self-reported physical activity data and longer-
76 term follow-up.

77 **What are the implications of our findings?**

78 Our results highlight the joint impact of genetic risk and physical inactivity on type 2 diabetes
79 development. These findings support the integration of genetic and behavioral risk factors in risk
80 stratification and prevention strategies and may inform targeted public health interventions aimed
81 at reducing the global burden of type 2 diabetes.

82

INTRODUCTION

The risk of developing type 2 diabetes varies considerably among individuals, reflecting differences in susceptibility to the often collectively called environmental factors (1,2). This variability underpins the significance of gene-environment interaction, wherein an individual's genetic profile, together with its downstream molecular effectors, interact with environmental exposures to uniquely modulate disease risk (3–6). While gene-environment interactions are pervasive in nature, shaping phenotypes such as plant growth (7), bird migration (8), and grasshopper adaptation to new environments (9), identifying such interactions in the context of human traits and diseases remains challenging (1), with often conflicting findings in the literature (10–13). This challenge is partly due to the limited precision and depth of environmental exposure assessments, particularly when self-reported (14,15).

Step count provides a simple and interpretable metric of daily ambulatory activity, offering advantages over self-reported physical activity measures or less intuitive accelerometer-derived intensity classifications (16). Prior studies have shown that higher step counts are associated with lower type 2 diabetes risk (17,18), and that accelerometer-measured intensity-specific physical activity may attenuate type 2 diabetes genetic risk (19). However, previous studies of gene–physical activity interactions have generally relied on single time-point physical activity assessments and have not accounted for the dynamic nature of physical activity over time. Moreover, no prior study has directly investigated how daily step counts interact with genetic risk on the development of type 2 diabetes.

In this study, we analyzed longitudinal data from a population-based cohort in the United States, comprising 9,501 adults with objectively measured physical activity over time, to test the

105 hypothesis that physical activity, measured via daily step count, interacts with genetic risk in the
106 development of type 2 diabetes. We further evaluated whether similar patterns were observed
107 using additional activity measures, including device-based intensity levels and self-reported
108 physical activity.

109

110 **METHODS**

111 **Study design and population**

112 We used data collected from participants in the *All of Us* Research Program. Details on the
113 design and execution of the *All of Us* have been published previously (20,21). In this study, we
114 used Curated Data Repository (CDR) version 7 for Controlled Tier (C2022Q4R9). This dataset
115 included information on physical measurements and vital signs collected at enrollment, surveys,
116 electronic health record (EHR), Fitbit data, and genetic information from participants enrolled
117 between May 6, 2018, and July 1, 2022. Our analyses focused on participants who owned a
118 Fitbit and agreed to share their Fitbit, EHR, and genetic information. We excluded participants
119 who did not wear a Fitbit for at least 6 months. Only the authorized authors who completed *All*
120 *of Us* Research Program Responsible Conduct of Research Training accessed the deidentified
121 data from the Researcher Workbench (a secured cloud-based platform). Since the authors were
122 not directly involved with the participants, institutional review board review was exempted in
123 compliance with *All of Us* Research Program policy.

124 In a secondary analysis, we used data from Inter99 (21). Inter99 is a population-based pre-
125 randomized lifestyle intervention study aiming to prevent ischemic heart disease and type 2
126 diabetes. Baseline data were collected from March 1999 until January 2001 and there were 6,784
127 participants. Follow-up was conducted after one, three, five, ten, and twenty years for a health
128 examination, completion of questionnaires and risk assessment.

129

130 **Physical activity assessment**

131 Activity tracking data for this study came from the Bring Your Own Device program that
132 allowed individuals who already owned a tracking device (Fitbit, Inc) to consent to link their
133 activity data with other data in the *All of Us* cohort. By registering their personal device on the
134 patient portal, participants could share all activity data collected since the creation of their
135 personal device account. We used mean daily steps calculated on a monthly basis for each
136 participant. Consistent with prior physical activity data curation approaches (17), days with less
137 than 10 hours of wear time, less than 100 steps, or greater than 45 000 steps or for which the
138 participant was younger than 18 years were removed. In addition, we removed months with
139 fewer than 15 valid days of monitoring. To capture additional dimensions of physical activity,
140 we derived device-based measures of intensity, including daily lightly active minutes, moderate
141 active minutes, and very active minutes []. We also analyzed self-reported physical activity data
142 from the Inter99 study, assessed via validated questionnaires covering work and leisure-time
143 activity (22), at baseline and at one-, three-, and five-year follow-ups. Reported activity levels
144 were converted into minutes per week by summing responses based on a five-day working week.

145

146 **Type 2 diabetes polygenic scores**

147 We generated a global polygenic score for type 2 diabetes that captures overall genetic burden
148 using genetic variants and weights identified in a previous study that did not include participants
149 from this study (23). We followed standard procedures to filter genetic data to include only
150 variants that had passed initial quality control (24). We used PRS–continuous shrinkage (PRScs)
151 (25) to infer posterior SNP effect sizes under continuous shrinkage priors, with the scaling

152 parameter auto-estimated (25). Scores were computed with ancestry-specific variants and
153 weights.

154

155 **Ascertainment of type 2 diabetes**

156 Incident type diabetes cases were identified using any incident billing code in EHR. We
157 excluded any new diagnoses coded during the first six months of monitoring, assuming that such
158 conditions were likely prevalent, but not yet recognized clinically. The EHR data from different
159 participating sites were mapped and harmonized using the Observational Medical Outcomes
160 Partnership common data model (26). In Inter99, we used data from separate nationwide health
161 registries to define type 2 diabetes events.

162

163 **Statistical analysis**

164 A flow diagram describes how many participants were excluded based on the criteria used to
165 create the analytical dataset (Supplementary Figure 1). Descriptive statistics for participant's
166 demographic and clinical characteristics were presented by median and IQR for continuous
167 variables and frequency for categorical variables.

168 We used Cox proportional hazard models to examine associations of genetic risk and step counts
169 with type 2 diabetes risk. For these analyses, we censored cases that occurred 6 months after
170 completing enrolment to mitigate potential bias due to time-varying confounding. Person-time
171 for each participant was calculated from the six months after completing baseline assessment to

172 the diagnosis of type 2 diabetes, death, loss to follow-up, or the end of the follow-up period,
173 whichever came first. We modelled polygenic scores and step counts as continuous variables.
174 We averaged daily step counts within each month and subsequently treated step counts as a time-
175 varying covariate in the Cox models to account for month-to-month variations in physical
176 activity during the follow-up period. We also classified participants according to categories of
177 genetic risk and physical activity (nine categories based on thirds of genetic risk and step counts,
178 with low genetic risk and high physical activity as reference) and conducted stratified analyses.
179 We adjusted the multivariable models for age (in years, continuous), sex (male, female) and
180 genetic ancestry inferred from the first 10 principal components of genomic relationship matrix.
181 We verified the proportional hazards assumption of the Cox model by using the Schoenfeld
182 residuals technique (p values for the global test > 0.1). Analyses were conducted for complete
183 cases given the small number of people with missing covariates. In secondary analyses, we
184 further adjusted our models for BMI (quintiles of kg/m^2). The rationale for not including BMI in
185 the main model was because BMI can act as a collider or a mediator (27).

186 We tested for additive and multiplicative interactions of genetic risk and physical activity with
187 type 2 diabetes incidence. For additive interactions we assessed the relative excess risk due to
188 interaction (RERI) and further examined the decomposition of the joint effect—that is, the
189 population attributable proportions to genetic risk alone, physical activity alone, and to their
190 interaction(28). For these analyses, we used the following formula (RERI = RR11 - RR10 -
191 RR01 + 1) and (attributable proportion (AP) = RERI/ RR11) (28). CIs for each of the interaction
192 measures were calculated using the delta method described by Hosmer and Lemeshow. We
193 tested for multiplicative interactions using the log-likelihood ratio test to compare the goodness
194 of fit of a multivariable-adjusted model with and without the cross-product interaction term(29).

195 For these analyses, we considered genetic risk and physical activity as continuous variables and
196 modeled the effect per 1SD.

197 In secondary interaction analyses, we derived device-based measures of physical activity
198 intensity, including daily lightly active minutes, fairly active minutes, and very active minutes.
199 These intensity levels were defined using metabolic equivalents (METs): 1.5–3.0 METs for
200 lightly active, 3.0–6.0 METs for fairly active, and >6.0 METs for very active minutes (30,31).
201 We also incorporated total self-reported physical activity data from the Inter99 study,
202 categorizing participants as having high (7–12 h/week) or low (0–2 h/week) activity. Additional
203 analyses were conducted stratifying participants by ethnic group (European, African, East Asian,
204 and South Asian) and by sex.

205 Two-sided tests were used to assess whether polygenic risk for T2D and physical activity are
206 associated with T2D incidence. A one-sided test was used to assess whether RERI is greater than
207 0, i.e., testing the hypothesis that the combined effects of polygenic risk for T2D and physical
208 activity exceeds their individual effects. All statistical analyses were performed using R
209 software, version 4.0.3 (R Foundation). We followed the Better Precision-data Reporting of
210 Evidence from Clinical Intervention Studies & Epidemiology (BePRECISE) reporting guideline
211 (32).

212 **RESULTS**

213 Of 14,200 participants with available Fitbit data at the time of our analysis, 9,501 had valid
214 physical activity data, genetic information, and no diabetes at baseline (Supplementary Figure 1).
215 The median age was 56 years (interquartile range (IQR) 42–66), the median body mass index
216 (BMI) was 28.3 kg/m^2 (IQR 24.8–33.3), and most participants were women (72%), of European
217 ancestry (88%), and held a college degree (72%; Table 1). Participants contributed 11.6 million
218 person-days of Fitbit monitoring over a median of 5.7 years (IQR, 3.6–7.7), yielding 97.9 billion
219 recorded steps. Median daily step count was 7,891 (IQR, 5,748–10,471), and step count showed
220 a normal distribution (Supplementary Figure 2). Compared to non-include All of US
221 participants, included individuals were more likely to be younger, female, of European ancestry,
222 with lower BMI and higher educational attainment (Supplementary Table 1).

223 During 53,672 person-years of follow-up, 494 incident cases of type 2 diabetes were
224 documented. Crude type 2 diabetes incidence rates per 1,000 person-year according to daily step
225 counts tertiles were 13.5 (95% CI: 11.9-15.2), 7.5 (95% CI: 6.3-8.8), and 5.9 (95% CI: 4.7-7.0)
226 for the lowest, intermediate, and highest tertile, respectively (multivariable-adjusted HR of 0.40
227 (95% CI: 0.26, 0.63; Table 2). Each standard deviation (SD) increase in step count was
228 associated with lower risk of type 2 diabetes, with an adjusted HR of 0.65 (95% CI, 0.57–0.73).

229 Similarly, type 2 diabetes incidence increased across tertiles of polygenic risk, with incidence
230 rates per 1,000 person-years of 5.9 (95% CI: 4.8, 7.0) in the lowest, 7.9 (95% CI: 6.6, 9.2) in the
231 intermediate, and 14.0 (95% CI: 12.3, 15.8) in the highest tertile. In adjusted models, participants
232 in the highest tertile had 2.50 times the risk of type 2 diabetes compared with those in the lowest

233 (95% CI, 2.04–3.06; Table 2). When modelled in the continuous scale, each SD increase in
234 polygenic score was associated with an adjusted HR of 1.52 (95% CI, 1.38–1.67).

235 When examining joint associations, crude incidence rates ranged from 4.1 (95% CI, 2.5–5.7) per
236 1,000 person-years among participants with low genetic risk and high physical activity to 18.4
237 (95% CI, 15.2–21.6) among those with high genetic risk and low physical activity (Figure 1).
238 The corresponding multivariable-adjusted HR was 6.20 (95% CI, 3.97–9.60). Within the high
239 genetic risk group, individuals with low activity (~5,000 steps/day) had nearly double the
240 incidence of those with high activity (~13,000 steps/day; HR, 1.92; 95% CI, 1.36–2.69). The
241 joint association was higher than the sum of the risk associated with each factor alone, with a
242 relative excess risk due to interaction (RERI) of 0.20 (95% CI: 0.04, 0.36); p=0.007).
243 Decomposition of excess type 2 diabetes risk attributed 40% to genetic risk (95% CI: 31, 49),
244 45% to low physical activity (95% CI: 29, 62), and 15% to their interaction (95% CI: 3, 27;
245 Table 3). No evidence of multiplicative interaction was observed (p>0.05).

246 We conducted a series of sensitivity and subgroup analyses. When physical activity was
247 modeled by intensity, similar additive interactions were observed. The RERI was 0.21 (95% CI,
248 0.02–0.41; p=0.015) for moderate activity minutes, with 19% of excess risk attributable to
249 interaction, and 0.40 (95% CI, 0.21–0.59; p<0.001) for highly active minutes, with 31%
250 attributable to interaction (Supplementary Table 2). Adjustment for BMI did not materially
251 change these estimates (Supplementary Table 2). Using total physical activity assessed by self-
252 reported physical activity in Inter99, the RERI was 0.53 (95% CI, 0.30–0.76; p=0.007), with
253 44% of excess risk attributable to interaction (Supplementary Table 3). Adjustment for BMI
254 attenuated this proportion to 38% (95% CI, –6 to 82; Supplementary Table 3).

255 In ancestry-stratified analyses, no significant associations between genetics and physical activity
256 or interactions were observed in the African ancestry group (68 cases among 583 participants;
257 Supplementary Table 4). Due to small sample sizes, models for South Asian (5 cases among 85
258 participants) and East Asian (9 cases among 207 participants) did not converge. Sex-stratified
259 analyses showed additive interactions in both sexes. Among men, the RERI was 0.2 (95% CI:
260 0.01, 0.42; p=0.03), and 22% of excess risk was attributed to interaction. Among women, RERI
261 was 0.2 (95% CI: -0.04, 0.43; p=0.050), with an attributable proportion of 11% (Supplementary
262 Table 5).

263

264

265 **DISCUSION**

266 Our findings provide consistent evidence of an additive interaction between genetic risk and
267 physical activity in the development of type 2 diabetes. Specifically, individuals with both high
268 genetic risk and low step counts had higher risk than expected from the sum of their individual
269 effects, with approximately one-fifth of excess risk attributable to this interaction. Similar
270 patterns were observed when physical activity was assessed using device-based intensity metrics
271 or self-reported total physical activity. Since additive interactions reflect differences in absolute
272 risk across populations (28), these observations may have potential public health relevance,
273 suggesting that reducing the overlap of these risk factors could contribute to lowering type 2
274 diabetes incidence. However, replication in other cohorts and settings will be important to better
275 understand the robustness and generalizability of these results.

276 Our study builds on prior research investigating gene-environment interactions in type 2
277 diabetes, which has often yielded inconsistent findings (10–13). One likely contributor to these
278 discrepancies is error in exposure assessment due to reliance on self-reported methods (14,15).
279 Using accelerometer-based physical activity measures, a recent UK Biobank study demonstrated
280 an additive effect between physical activity and polygenic scores on T2D risk, with greater
281 absolute risk reductions observed among individuals at high genetic risk who engaged in
282 moderate-to-vigorous physical activity (19). Our findings extend this evidence base in three
283 important ways. First, unlike the UK Biobank study, which relied on single time-point physical
284 activity assessment, we modelled physical activity as a time-varying exposure, capturing
285 behavioural changes over a median follow-up of 5 years. This approach more accurately reflects
286 behavioral dynamics leading up to diabetes onset (33). Second, we used daily step counts as the
287 primary exposure. Compared with intensity-based classifications such as "moderate-to-

288 vigorous," which may vary in interpretation and are less intuitive, step counts provide a simple,
289 interpretable, and actionable metric for public health messaging (16). Third, our polygenic scores
290 were derived from an external discovery dataset that excluded participants from this study,
291 thereby mitigating potential inflation of type I error due to sample overlap, a methodological
292 limitation that may have influenced the UK Biobank analyses. Together, our study strengthens
293 the evidence that objectively measured physical activity interacts with genetic risk on the
294 development of type 2 diabetes.

295 Physical activity remains central to public health recommendations for type 2 diabetes
296 prevention (34), yet current guidelines do not account for heterogeneity in individual responses.
297 Factors such as genetic susceptibility, body composition, and diet can all influence the
298 effectiveness of physical activity (35). For example, in the ARIC cohort, the protective effects of
299 physical activity were attenuated in women with higher genetic predisposition to insulin
300 resistance (36), suggesting that greater physical activity may be needed to offset genetic risk.
301 Similarly, a prior analysis from the All of Us Research Program estimated that individuals at
302 high genetic risk for obesity would need to walk an additional ~2,300 steps per day to match the
303 obesity risk of those at lower risk (37). In our study, participants at high genetic risk who walked
304 ~5,000 steps per day had nearly double the incidence of type 2 diabetes compared to those
305 walking ~13,000 steps per day. Our models estimated that 15% of type 2 diabetes cases could
306 have been prevented by eliminating one of the two risk factors, emphasizing the potential impact
307 of increasing activity levels, particularly among genetically susceptible individuals.

308 Although the proportion of risk attributable to gene–physical activity interactions was similar
309 when using step counts or device-based intensity measures, higher estimates were observed in
310 the Inter99 cohort for total physical activity. This discrepancy may reflect differences in follow-

311 up time, population characteristics such as age and obesity prevalence, or potential inflation due
312 to self-reported activity measures. Since population-attributable risk depends on both exposure
313 prevalence and effect size (38), exposure misclassification, especially with self-reported physical
314 activity, could partly explain these differences. While these excess risk estimates due to the
315 interaction may change over time with shifting diabetes incidence, our findings underscore the
316 ongoing relevance of gene–physical activity interactions in type 2 diabetes.

317 This study has several limitations. First, as an observational study, participants were not
318 randomized to high- or low-physical activity groups, limiting causal inference interpretations.
319 Recall-by-genotype studies could help test these hypotheses under experimental conditions.
320 Second, incident type 2 diabetes was defined solely based on any incident billing codes from the
321 electronic health record, which may introduce misclassification. However, this approach is
322 widely used in biobank research, including other All of Us analyses (17,39). Third, the number
323 of type 2 diabetes cases was relatively small, reflecting limited follow-up and the subset of
324 participants with linked wearable data. Fourth, while we used comprehensive polygenic scores
325 for type 2 diabetes, other genetic risk factors (i.e., for obesity) were not included. Although we
326 adjusted for BMI in secondary analyses, residual confounding, collider bias, or mediation effects
327 could still exist. Larger studies with longer follow-up and detailed modelling of BMI trajectories
328 may help address this. Finally, other molecular, clinical, or behavioral factors such as diet, body
329 composition, or gut microbiome were not explicitly included in our analyses, but could affect
330 risk or modify gene–environment interactions.

331 In conclusion, by quantifying the combined contributions of genetic risk and physical activity to
332 type 2 diabetes risk, our results underscore the potential value of integrating genomic data with
333 objective, device-derived measures of physical activity to better identify those who are most

334 likely to benefit from targeted interventions. Future studies with larger, diverse populations and
335 longer follow-up are needed to confirm these interactions and to evaluate how they can be
336 translated into effective public health and clinical interventions.

337

338 **REFERENCES**

339 1. Franks PW, McCarthy MI. Exposing the exposures responsible for type 2 diabetes and
340 obesity. Science. 2016 Oct 7;354(6308):69–73.

341 2. McCarthy MI. Painting a new picture of personalised medicine for diabetes. Diabetologia.
342 2017 Feb 7.

343 3. Franks PW, Pearson E, Florez JC. Gene-environment and gene-treatment interactions in
344 type 2 diabetes: progress, pitfalls, and prospects. Diabetes Care. 2013 May 1;36(5):1413–
345 21.

346 4. Franks PW, Merino J. Gene-lifestyle interplay in type 2 diabetes. Curr Opin Genet Dev.
347 2018;50:35–40.

348 5. DJ H. Gene-environment interactions in human diseases. Nat Rev Genet. 2005;6(4).

349 6. Herrera-Luis E, Benke K, Volk H, Ladd-Acosta C, Wojcik GL. Gene–environment
350 interactions in human health. Nat Rev Genet. 2024 Nov 1;

351 7. Napier JD, Heckman RW, Juenger TE. Gene-by-environment interactions in plants:
352 Molecular mechanisms, environmental drivers, and adaptive plasticity. Vol. 35, Plant
353 Cell. 2023.

354 8. Bourret A, Garant D. Candidate gene-environment interactions and their relationships
355 with timing of breeding in a wild bird population. Ecol Evol. 2015;5(17).

356 9. John B, Lewis KR. Chromosome variability and geographic distribution in insects.
357 Science (1979). 1966;152(3723).

358 10. Merino J, Guasch-Ferré M, Li J, Chung W, Hu Y, Ma B, et al. Polygenic scores, diet
359 quality, and type 2 diabetes risk: An observational study among 35,759 adults from 3 US
360 cohorts. PLoS Med. 2022;19(4).

361 11. Langenberg C, Sharp SJ, Franks PW, Scott RA, Deloukas P, Forouhi NG, et al. Gene-
362 Lifestyle Interaction and Type 2 Diabetes: The EPIC InterAct Case-Cohort Study. PLoS
363 Med. 2014 May 20;11(5):e1001647.

364 12. Wang T, Heianza Y, Sun D, Huang T, Ma W, Rimm EB, et al. Improving adherence to
365 healthy dietary patterns, genetic risk, and long term weight gain: gene-diet interaction
366 analysis in two prospective cohort studies. BMJ. 2018 Jan 10;360:j5644.

367 13. Kilpeläinen TO, Qi L, Brage S, Sharp SJ, Sonestedt E, Demerath E, et al. Physical activity
368 attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults
369 and 19,268 children. Lewis C, editor. PLoS Med. 2011 Nov 1;8(11):e1001116.

370 14. Hebert JR, Ebbeling CB, Matthews CE, Hurley TG, MA Y, Druker S, et al. Systematic
371 errors in middle-aged women's estimates of energy intake: Comparing three self-report
372 measures to total energy expenditure from doubly labeled water. Ann Epidemiol.
373 2002;12(8).

374 15. Olds TS, Gomersall SR, Olds ST, Ridley K. A source of systematic bias in self-reported
375 physical activity: The cutpoint bias hypothesis. J Sci Med Sport. 2019 Aug 1;22(8):924–8.

376 16. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step Counting: A Review of
377 Measurement Considerations and Health-Related Applications. Sports Medicine. 2017 Jul
378 1;47(7):1303–15.

379 17. Master H, Annis J, Huang S, Beckman JA, Ratsimbazafy F, Marginean K, et al.
380 Association of step counts over time with the risk of chronic disease in the All of Us
381 Research Program. Nat Med. 2022;28(11).

382 18. Ding D, Nguyen B, Nau T, Luo M, del Pozo Cruz B, Dempsey PC, et al. Daily steps and
383 health outcomes in adults: a systematic review and dose-response meta-analysis. Lancet
384 Public Health. 2025 Aug 1;10(8):e668–81.

385 19. Luo M, Yu C, Del Pozo Cruz B, Chen L, Ding D. Accelerometer-measured intensity-
386 specific physical activity, genetic risk and incident type 2 diabetes: A prospective cohort
387 study. Br J Sports Med. 2023;57(19).

388 20. The “All of Us” Research Program. New England Journal of Medicine. 2019 Aug
389 15;381(7):668–76.

390 21. Jørgensen T, Borch-John sen K, Thomsen TF, Ibsen H, Glümer C, Pisinger C. A
391 randomized non-pharmacological intervention study for prevention of ischaemic heart
392 disease: Baseline results Inter99 (1). Eur J Prev Cardiol. 2003;10(5).

393 22. Aadahl M, Jørgensen T. Validation of a new self-report instrument for measuring physical
394 activity. Med Sci Sports Exerc. 2003;35(7).

395 23. Suzuki K, Hatzikotoulas K, Southam L, Taylor HJ, Yin X, Lorenz KM, et al. Genetic
396 drivers of heterogeneity in type 2 diabetes pathophysiology. Nature. 2024;627(8003).

397 24. All of Us Genomic Quality Report – User Support. Available from:
398 https://support.researchallofus.org/hc/en-us/articles/27633757470228-All-of-Us-
399 Genomic-Quality-Report

400 25. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian
401 regression and continuous shrinkage priors. Nat Commun. 2019;10(1).

402 26. Observational Medical Outcomes Partnership (OMOP) common data model (CDM) –
403 User Support. Available from: https://support.researchallofus.org/hc/en-
404 us/articles/360039685271-Observational-Medical-Outcomes-Partnership-OMOP-
405 common-data-model-CDM

406 27. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider scope: when
407 selection bias can substantially influence observed associations. Int J Epidemiol. 2018 Feb
408 1;47(1):226–35.

409 28. VanderWeele TJ, Tchetgen Tchetgen EJ. Attributing effects to interactions. Epidemiology
410 [Internet]. 2014 Sep [cited 2020 Jul 21];25(5):711–22.

411 29. Van Der Weele TJ, Knol MJ. A tutorial on interaction. Epidemiol Methods. 2014;3(1).

412 30. Fitbit Web API Data Dictionary.

413 31. Bailey CP, Dodd KW, McClain JJ, Seo I, Wheeler W, Wolff-Hughes DL. Fitbit Physical
414 Activity and Sleep Data in the All of Us Research Program: Data Exploration and
415 Processing Considerations for Research. Med Sci Sports Exerc. 2025. In Press.

416 32. Lim SS, Semnani-Azad Z, Morieri ML, Ng AH, Ahmad A, Fitipaldi H, et al. Reporting
417 guidelines for precision medicine research of clinical relevance: the BePRECISE
418 checklist. Nature Medicine 2024 30:7. 2024 Jul 19;30(7):1874–81.

419 33. Hilden P, Schwartz JE, Pascual C, Diaz KM, Goldsmith J. How many days are needed?
420 Measurement reliability of wearable device data to assess physical activity. PLoS One.
421 2023 Feb 1;18(2):e0282162.

422 34. 3. Prevention or Delay of Diabetes and Associated Comorbidities: Standards of Care in
423 Diabetes—2024. Diabetes Care. 2024;47.

424 35. Martiskainen T, Lamidi ML, Venojärvi M, Tikkanen H, Laatikainen T. Factors affecting
425 the effectiveness of the physical activity counselling intervention implemented in primary
426 health care in adults with type 2 diabetes. BMC Endocr Disord. 2023 Dec 1; 23(1).

427 36. Yann C, Zhao K, Amit Arora C, Hsu CH. Association of physical activity with lower type
428 2 diabetes incidence is weaker among individuals at high genetic risk. Diabetologia.
429 2014;57(12).

430 37. Brittain EL, Han L, Annis J, Master H, Hughes A, Roden DM, et al. Physical Activity and
431 Incident Obesity Across the Spectrum of Genetic Risk for Obesity. JAMA Netw Open.
432 2024 Mar 4;7(3):e243821–e243821.

433 38. Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions.
434 Am J Public Health. 1998;88(1):15–9.

435 39. Eastwood S V., Mathur R, Atkinson M, Brophy S, Sudlow C, Flaig R, et al. Algorithms
436 for the capture and adjudication of prevalent and incident diabetes in UK Biobank. PLoS
437 One. 2016 Sep 1;11(9).

438

Funding/Acknowledgements:

We gratefully acknowledge *All of Us* and Inter99 participants for their contributions, without whom this research would not have been possible. The *All of Us* Research Program is supported by grants 1 OT2 OD026549, 1 OT2 OD026554, 1 OT2 OD026557, 1 OT2 OD026556, 1 OT2 OD026550, 1 OT2 OD 026552, 1 OT2 OD026553, 1 OT2 OD026548, 1 OT2 OD026551, 1 OT2 OD026555, IAA AOD21037, AOD22003, AOD16037, and AOD21041 (regional medical centers); grant HHSN 263201600085U (federally qualified health centers); grant U2C OD023196 (data and research center); 1 U24 OD023121 (Biobank); U24 OD023176 (participant center); U24 OD023163 (participant technology systems center); grants 3 OT2 OD023205 and 3 OT2 OD023206 (communications and engagement); and grants 1 OT2 OD025277, 3 OT2 OD025315, 1 OT2 OD025337, and 1 OT2 OD025276 (community partners) from the National Institutes of Health (NIH).

This study was also supported by the Novo Nordisk Foundation grant NNF23SA0084103 and the European Union. JM was additionally supported by grants from the European Commission (HORIZON-EIC-2023-PATHFINDERCHALLENGES-01 - 101161509) and the EFSD/NNF Future Leaders Award (#0094134). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Innovation Council and SMEs Executive Agency (EISMEA). Neither the European Union nor the granting authority can be held responsible for them. RL was additionally supported by personal grants from the Novo Nordisk Foundation (Laureate award no. NNF20OC0059313) and the Danish National Research Fund (Chair DNRF161). TOK and MRC and were additionally supported by grants from the Novo Nordisk Foundation (NNF22OC0074128) and EFSD/Lilly European Diabetes Research Programme.

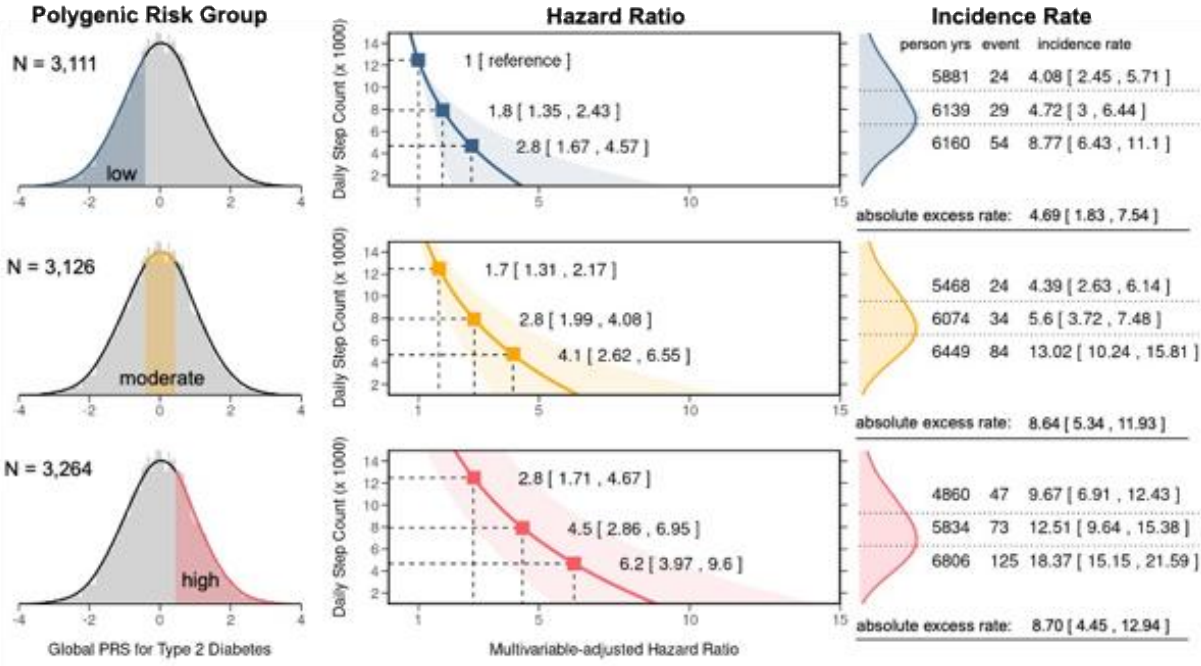
462 **Ethics declarations:**

463 The authors declare no competing interests. JM is the guarantor of this work, and, as such, had
464 full access to all the data in the study and takes responsibility for the integrity of the data and the
465 accuracy of the data analysis.

466

Figures and Tables

Figure 1: Risk of incident type 2 diabetes by global polygenic risk score for type 2 diabetes, physical activity.



<u>Figure Legend:</u> Risk is indexed using crude incidence rate and hazard ratio adjusted for age, sex and genetic ancestry. HR hazard ratio. Dots correspond to averages of daily step count tertiles, which are ~ 5k, ~ 8k and ~13k. Histograms along the y-axis (to the right) show the density of daily step counts. Incidence rates are computed for tertiles of daily step counts.

Table 1. Baseline characteristics of the included *All of US* research program participants.

Person-y[ϕ]	53,672
Age, years	56 (42-66)
Sex, Women, No. (%)	6,863 (72)
BMI, kg/m^2	28.3 (24.8-33.3)
Family history of diabetes, No. (%)	2,994 (32)
Smoker, No. (%)	3,136 (33)
Self-reported race/ethnicity[ε]	
White, No. (%)	8132 (86)
African, No. (%)	583 (6)
East Asian, No. (%)	207 (2)
South Asian, No. (%)	85 (1)
Education	
College degree, No. (%)	6,856 (72)
Some/No college, No. (%)	2,587 (27)

Table Legend: Values are medians (interquartile range) for continuous variables; numbers and (percentages) for categorical variables.

Table 2: Adjusted hazard ratios of T2D risk according to tertiles of daily step counts and genetic risk.

	Low step count	**Intermediate step count**	**High step count**	**P for trend**
No. of events/person-years	263 / 19415	136 / 18,046	95 / 16,210	—
Incidence rate (1,000 person-years; 95% CI)	13.5 (11.9-15.2)	7.5 (6.3-8.8)	5.9 (4.7-7.0)	—
Multivariable-adjusted model	1.00 (Ref)	0.68 (0.57-0.82)	0.40 (0.26-0.63)	<0.001

	Low genetic risk	**Intermediate genetic risk**	**High genetic risk**	**P for trend**
No. of events/person-years	107/ 18,180	142/ 17,991	245/ 17,500	—
Incidence rate (1,000 person-years; 95% CI)	5.9 (4.8-7)	7.9 (6.6-9.2)	14 (12.2-15.8)	—
Multivariable-adjusted model	1.00 (Ref)	1.59 (1.43-1.76)	2.50 (2.04-3.06)	<0.001

Table Legend: Hazards ratios and 95% CI for T2D risk. Daily step counts and genetic risk were categorized in three categories based on the distribution of physical activity and polygenic score. Cox proportional hazards models were adjusted for age (in years, continuous), sex (male, female), and the first 10 principal components of the genomic relationship matrix.

Table 3 Additive interactions between physical activity and global polygenic risk of T2D.

Main effects	
Physical activity†	1.53 (1.32, 1.74)
Polygenic score‡	1.61 (1.32, 1.74)
Joint effect	2.34 (1.87, 2.81)
Relative excess risk due to interaction	
Relative excess risk due to interaction	0.2 (0.04, 0.36)
P value	0.007
Attributable risk proportion, %	
Daily step count	39.8 (31.0, 48.5)
Polygenic score	45.4 (28.6, 62.1)
Additive interaction	14.8 (2.5, 27.2)

Table Legend: Multivariable adjusted risk of type 2 diabetes estimated from Cox proportional hazards models adjusted for age, sex, and the first ten principal components of genomic relationship matrix. The joint analysis was conducted to quantitatively estimate the combined association of physical activity and genetic risk simultaneously with the incidence of type 2 diabetes. The relative excess risk due to interaction (RERI) is based on joint effects of a one-standard-deviation increase in PRS and a one-standard-deviation decrease in daily step count using the following formula (RERI = RR11 - RR10 - RR01 + 1). The model's attributable risk proportion due to the interaction was calculated as AP = RERI/ RR11(28).

† per SD decrease in daily step counts.

‡ per SD increase in the global polygenic scores.