Lower hemoglobin levels associate with lower body mass index and healthier metabolic profile

Juha Auvinen[1,2§], Joona Tapio[3,4§], Ville Karhunen[1,5§], Johannes Kettunen[3,6,7], Raisa Serpi[3,4], Elitsa y. Dimova[3,4], Pasi Soininen[6,7], Tuija Tammelin[8], Juha Mykkänen[9], Katri Puukka[10], Mika Kähönen[11], Emma Raitoharju[12], Terho Lehtimäki[12], Mika Ala-Korpela[3,6,7,13,14,15], Olli T. Raitakari[9,16], Sirkka Keinänen-Kiukaanniemi[1,2], Marjo-Riitta Järvelin[1-3,17,18*] & Peppi Koivunen[3,4*]

[§]These authors contributed equally to the work

[*]These authors jointly supervised the work

[1]Faculty of Medicine, Center for Life Course Health Research, University of Oulu, 90014 Oulu, Finland. [2]Unit of Primary Care, Oulu University Hospital, 90220 Oulu, Finland. [3]Biocenter Oulu, [4]Faculty of Biochemistry and Molecular Medicine, Oulu Center for Cell-Matrix Research, University of Oulu, 90014 Oulu, Finland. 90014 Oulu, Finland. [5]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, Norfolk Place, W2 1PG, London, UK. [6]Computational Medicine, Faculty of Medicine, University of Oulu, 90014 Oulu, Finland. [7]NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, 70210 Kuopio, Finland. [8]LIKES Research Center for Physical Activity and Health, 40700 Jyväskylä, Finland. [9]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, 20520 Turku, Finland. [10]NordLab Oulu, Medical Research Center Oulu, Oulu University Hospital and Department of Clinical Chemistry, University of Oulu, 90014 Oulu, Finland. [11]Department of Clinical Physiology, Tampere University Hospital and Faculty of Medicine and

Life Sciences, University of Tampere. [12]Department of Clinical Chemistry, Fimlab Laboratories and Finnish Cardiovascular Research Center-Tampere, Faculty of Medicine and Life Sciences, University of Tampere. [13]Population Health Science, Bristol Medical School, University of Bristol, Clifton BS8 2BN, UK. [14]Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Clifton BS8 2BN, UK. [15]Systems Epidemiology, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia. [16]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, 20520 Turku, Finland. [17]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK. [18]Department of Life Sciences, College of Health and Life Sciences, Brunel University London, Kingston Lane, Uxbridge, Middlesex UB8 3PH, UK.

Running title: Hemoglobin and metabolism

Correspondence to: Peppi Koivunen, MD, PhD, Faculty of Biochemistry and Molecular Medicine, University of Oulu, Aapistie 7, P.O. Box 5400, FIN-90014, Finland, Email peppi.koivunen@oulu.fi,Tel. +358-8-5375822 and Marjo-Riitta Järvelin, MD, PhD, Faculty of Medicine, University of Oulu, Aapistie 7, P.O. Box 5000, FIN-90014, Finland, or Department of Epidemiology and Biostatistics, Imperial College London, Norfolk Place, London W2 1PG, UK, and Department of Life Sciences, College of Health and Life Sciences, Brunel University London, Kingston Lane, Uxbridge, Middlesex UB8 3PH, UK. Email m.jarvelin@imperial.ac.uk, Tel. +44 (0)20 7594 3345

Hemoglobin (Hb) is the main carrier of oxygen. In general, high-end Hb levels within the normal range are considered beneficial for health[1]. However, activation of the hypoxia response has been shown to protect mice against metabolic dysfunction[2-4]. We used Hb levels as an indicator for oxygenation status and studied its association with >170 anthropometric or metabolic parameters in two Finnish birth cohorts (total n = 7,175) in a longitudinal design. Here we show a remarkable strong positive linear association between Hb levels and body mass index (BMI). Subjects with the lower Hb levels also had better glucose tolerance, lower cholesterol and blood pressure levels, less adverse metabolite profiles and lower inflammatory load. Notably, these associations were not only mediated by the lower BMI, and the effect size of many of them increased with age. Polygenic risk score analyses indicated shared genetic determinants between Hb levels and BMI and insulin, triglyceride and HDL cholesterol levels. Manipulation of Hb levels by venesection in mice confirmed causal effects on body weight and metabolic parameters. Our findings suggest that lower-end normal Hb levels are favorable for and maintenance of healthy metabolism involving mild chronic activation of the hypoxia response. Therefore modulation of Hb levels could serve as a novel strategy towards treatment of metabolic syndrome.

Hb, an iron-containing metalloprotein in red blood cells, is the main carrier of oxygen. Its levels directly affect arterial oxygen concentration and thereby tissue oxygenation[5, 6]. Hb levels are regulated genetically and environmentally, and they vary by sex, race, age and altitude[1, 7]. Individual's Hb levels during adult life are, however, very stable and used *e.g.* in the athlete's biological passport to detect doping[8]. When tissues encounter reduced oxygen levels as a key transcriptional response the hypoxia-inducible factor (HIF) becomes stabilized. It upregulates genes that induce oxygen delivery and reduce its usage, such as those regulating energy metabolism[9]. The stability of HIF is governed by HIF prolyl 4-hydroxylases (HIF-P4Hs), which require oxygen for catalysis and targeting HIF for destruction[9]. Recent studies show that inhibition of HIF-P4Hs, which stabilizes HIF in normoxia, protects mice from obesity, metabolic dysfunction and associated diseases[2-4, 10, 11]. To investigate oxygenation status with regards to metabolic health in human population at a larger scale, we explored the possibility to use the normal variation in Hb levels as a surrogate measure. To test this we first evaluated the association of Hb levels with key metabolic parameters in C57Bl/6 mice. We found positive associations between Hb levels and body weight, glucose tolerance and homeostatic model assessment-insulin resistance (HOMA-IR) scores, respectively (Figs. 1a-1c), suggesting that Hb levels can influence metabolic health.

We then examined association of Hb levels in humans with anthropometric and metabolic parameters in cross-sectional and longitudinal design from 31 to 46 years in the population based, sea-level Northern Finland Birth Cohort 1966 (NFBC1966)[12, 13] (n = 5,351), and searched for a replication in another sea-level cohort, Cardiovascular Risk in Young Finns Study (YFS)[14] (n = 1,824). The statistical analyses were adjusted for potential confounding factors, and as appropriate, for BMI (see Methods and Extended data Figure 1 and Table 1 for characteristics of the subjects and flow chart of the study). There was evidence for a strong positive linear

association between Hb levels and BMI in NFBC1966 at age of 46 (Fig. 1d, effect size, log-transformed 0.296, $p = 4.73 \times 10^{-58}$, Extended data Fig. 2a), and this association replicated in YFS (Fig. 1e, Extended data Table 2). The increase of 10 g/l in Hb corresponded to an increase of 3.5% in BMI (Fig. 1b, Extended data Tables 2 and 3). Hb levels also associated positively with waist circumference, hip circumference, waist-hip ratio, body fat percentage and visceral fat area and negatively with body muscle percentage (Fig. 1b, Extended data Tables 2 and 3). Thus the subjects with the lower Hb levels appeared less abdominally obese and having more muscle tissue. Hematocrit levels and red blood cell counts also showed positive associations with BMI (Extended data Fig. 2 and Table 4).

We next showed by associating Hb and fasting glucose and insulin levels, and the derived insulin resistance indexes (HOMA-IR, HOMA-B and Matsuda[15]) from an oral glucose tolerance test that the subjects with lower Hb levels had better glucose tolerance and insulin sensitivity than those with higher Hb levels (Fig. 2a, Extended data Table 5). There was also evidence for positive association between Hb and levels of systolic and diastolic blood pressure, serum total cholesterol, LDL cholesterol, triglycerides and serum high sensitive C-reactive protein (CRP), respectively, and negative association between Hb and HDL cholesterol levels (Fig. 2a, Extended data Table 5). The effect sizes of these associations weakened when additionally adjusted for BMI, but all except that for CRP remained significant (Fig. 2a, Extended data Table 5).

To evaluate in a longitudinal set up the association of Hb levels with anthropometric and metabolic measures, we compared the effect sizes of these associations in subjects of NFBC1966 that had been examined for them at age of 31 and 46 years (n = 3,624). As expected, the mean Hb levels had stayed unchanged from 31 to 46 years (Extended data Figure 3). In general, the effect sizes of the associations increased with age or stayed the same (Fig. 2b), the most pronounced increases being found for the association of Hb levels with levels of triglycerides and insulin,

HOMA-IR and waist-hip-ratio, respectively (Fig. 2c, Extended data Tables 6 and 7).

We then evaluated the association between Hb levels and systemic metabolite profiles of 150 variables analyzed by nuclear magnetic resonance (NMR) metabolomics. Evidence for associations of Hb levels was found with lipoprotein subclass particle concentrations and particle sizes, and levels of ApoB, triglycerides and cholesterol in lipoproteins, fatty acids (FAs), lactate, glycerol, branched-chain amino acids isoleucine, leucine and valine, aromatic amino acids phenylalanine and tyrosine, the inflammatory glycoprotein acetyls, creatinine, albumin and ketone bodies, respectively, these associations being mostly positive (Fig. 3, Extended data Table 8). High levels of branched-chain and aromatic amino acids, glycoprotein acetyls and lactate, associated with the higher Hb levels here (Fig. 3, Extended data Table 8), have previously been associated with insulin resistance[16], and the first two with adiposity[17]. Moreover, high phenylalanine levels and high ratio of monounsaturated FAs to all FAs, associated with higher Hb levels here (Fig. 3, Extended data Table 8), have earlier been associated with increased risk for cardiovascular events[18]. The magnitudes of cross-sectional and longitudinal effect estimates of the Hb level difference from 31 to 46 years on the NMR metabolite levels in NFBC1966 were consistent (Extended data Figure 4). Furthermore, the estimated slope for regression of longitudinal effects on cross-sectional effects (0.97, 95% CI 0.89 to 1.05) indicated similar effects sizes for both the between-subject and the within-subject effects (Extended data Figure 4). Altogether, lower Hb levels associated with less adverse metabolite profiles.

We next analyzed in a subpopulation of NFBC1966[19] association of Hb levels with oxygen consumption at rest, and found evidence for positive association of it and for those of oxygen consumption and body fat percentage and Hb levels and body fat percentage, respectively (Fig. 4a, Extended data Fig. 1 and Table 9). Thus individuals with lower Hb levels and less fat consumed less oxygen.

HIF-P4Hs act as cellular oxygen sensors[20]. Even slight reductions in tissue oxygenation, such as those stemming from the normal variation in Hb levels, may result via HIF-P4H inhibition to upregulation of hypoxia-induced genes. To test this hypothesis we analyzed in a subpopulation of YFS[14] (n=1,636) association of Hb levels with the expression of a set of hypoxia-induced genes (Extended data Table 10) in whole-blood genome wide-expression data. After adjustment for confounding factors, we found evidence for geneset level transcriptional activation of hypoxia-induced genes in individuals in the lowest Hb quartile (Hb < 132 g/l) compared to the highest quartile (Hb > 152 g/l) (Figure 4b).

To examine for causality of Hb levels on BMI or the key metabolic measures bidirectional MR analyses were carried out using the European ancestry summary statistics from genome-wide meta-analysis studies[7, 21] for these variables. Neither inverse variance weighting (IVW) nor MR-Egger regression method gave substantial evidence for causality on any parameter to either direction (Extended data Table 11 and Figures 5 and 6). However, in classical MR using the genome wide single nucleotide polymorphisms (SNP) the instrument is weak. We therefore secondly optimized the instrument and carried out polygenic risk score analyses that revealed evidence for shared genetic determinants between Hb and BMI and levels of insulin, triglycerides and HDL cholesterol. However, with the currently available data sets the direction of causation remained unclear and its untangling would almost certainly require a larger Hb genome wide association study material (Extended data Table 12). We therefore manipulated Hb levels in C57Bl/6 mice by venesection, and showed that in two weeks the induced erythropoiesis increased their Hb levels which associated with increased body weight, insulin resistance and cholesterol levels (Fig. 4c, Extended data Figure 7).

While no association of Hb levels with overall metabolic health has been reported before, high Hb or hematocrit levels have earlier been associated in individual epidemiological studies

with increased insulin resistance, hypertension, hypercholesterolemia or hypertriglyceridemia[22-25]. Although the underlying mechanisms have been poorly understood hyperviscosity or changes in plasma volume have been suggested as mediators of these associations. The HIF pathway was only discovered after these analyses. One of the key adjustments in the hypoxia response is to reprogram energy metabolism transcriptionally to less oxygen consuming while less ATP producing, more inefficient mode[9, 26] (Fig. 4d). The data presented here suggest that the lower Hb levels activated the hypoxia response which mediated the beneficial effects on metabolic health. Individuals being environmentally exposed to hypoxia by living at high altitude have been reported to have a lower fasting glycemia and better glucose tolerance compared to those at near sea level, and demographic studies associate living at high altitude with protection against obesity and diabetes[27, 28]. Further, clinical trials with HIF-P4H inhibitors for treatment of anemia and ischemia reported ≥20% reduced cholesterol levels in patients receiving these therapeutics[29, 30]. While more mechanistic studies are needed, our data show that Hb levels contribute significantly to metabolic health, and that variation in Hb levels should be considered in clinical practice as potential indicator of risk of metabolic disadvantage.

Online Content Methods, along with any additional Extended Data display items are available in the online version of the paper; references unique to these sections appear only in the online paper.

References

1. Patel, K. V. Variability and heritability of hemoglobin concentration: an opportunity to improve understanding of anemia in older adults. *Haematologica* **93**, 1281-1283 (2008).

2. Matsuura, H. *et al*. Prolyl hydroxylase domain protein 2 plays a critical role in diet-induced obesity and glucose intolerance. *Circulation* **127**, 2078-2087 (2013).

3. Rahtu-Korpela, L. *et al*. HIF prolyl 4-hydroxylase-2 inhibition improves glucose and lipid metabolism and protects against obesity and metabolic dysfunction. *Diabetes* **63**, 3324-3333 (2014).

4. Michailidou, Z. *et al*. Adipocyte Pseudohypoxia Suppresses Lipolysis and Facilitates Benign Adipose Tissue Expansion. *Diabetes* (2014).

5. Jobsis, F. F. Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science* **198**, 1264-1267 (1977).

6. Sevick, E. M., Chance, B., Leigh, J., Nioka, S. & Maris, M. Quantitation of time- and frequency-resolved optical spectra for the determination of tissue oxygenation. *Anal. Biochem.* **195**, 330-351 (1991).

7. van der Harst, P. *et al*. Seventy-five genetic loci influencing the human red blood cell. *Nature* **492**, 369-375 (2012).

8. Lobigs, L. M., Knight, E. J., Schumacher, Y. O. & Gore, C. J. Within-subject haemoglobin variation in elite athletes: a longitudinal investigation of 13 887 haemoglobin concentration readings. *Drug Test. Anal.* **8**, 228-234 (2016).

9. Kaelin, W. G.,Jr & Ratcliffe, P. J. Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Mol. Cell* **30**, 393-402 (2008).

10. Rahtu-Korpela, L. *et al*. Hypoxia-Inducible Factor Prolyl 4-Hydroxylase-2 Inhibition Protects Against Development of Atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **36**, 608-617 (2016).

11. Marsch, E. *et al*. Deficiency of the oxygen sensor prolyl hydroxylase 1 attenuates hypercholesterolaemia, atherosclerosis, and hyperglycaemia. *Eur. Heart J.* (2016).

12. Rantakallio, P. Groups at risk in low birth weight infants and perinatal mortality. *Acta Paediatr. Scand.* **193**, Suppl 193:1+ (1969).

13. Jarvelin, M. R. *et al*. Early life factors and blood pressure at age 31 years in the 1966 northern Finland birth cohort. *Hypertension* **44**, 838-846 (2004).

14. Raitakari, O. T. *et al*. Cohort profile: the cardiovascular risk in Young Finns Study. *Int. J. Epidemiol.* **37**, 1220-1226 (2008).

15. Matsuda, M. Measuring and estimating insulin resistance in clinical and research settings. *Nutr. Metab. Cardiovasc. Dis.* **20**, 79-86 (2010).

16. Wurtz, P. *et al*. Metabolic signatures of insulin resistance in 7,098 young adults. *Diabetes* **61**, 1372-1380 (2012).

17. Wurtz, P. *et al*. Metabolic signatures of adiposity in young adults: Mendelian randomization analysis and effects of weight change. *PLoS Med.* **11**, e1001765 (2014).

18. Wurtz, P. *et al*. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation* **131**, 774-785 (2015).

19. Tammelin, T., Nayha, S. & Rintamaki, H. Cardiorespiratory fitness of males and females of northern Finland birth cohort of 1966 at age 31. *Int. J. Sports Med.* **25**, 547-552 (2004).

20. Hirsila, M., Koivunen, P., Gunzler, V., Kivirikko, K. I. & Myllyharju, J. Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *J. Biol. Chem.* **278**, 30772-30780 (2003).

21. Locke, A. E. *et al*. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).

22. Wannamethee, S. G., Perry, I. J. & Shaper, A. G. Hematocrit and risk of NIDDM. *Diabetes* **45**, 576-579 (1996).

23. Facchini, F. S., Carantoni, M., Jeppesen, J. & Reaven, G. M. Hematocrit and hemoglobin are independently related to insulin resistance and compensatory hyperinsulinemia in healthy, non-obese men and women. *Metabolism* **47**, 831-835 (1998).

24. Cirillo, M., Laurenzi, M., Trevisan, M. & Stamler, J. Hematocrit, blood pressure, and hypertension. The Gubbio Population Study. *Hypertension* **20**, 319-326 (1992).

25. Bottiger, L. E. & Carlson, L. A. Relation between serum cholesterol and triglyceride concentration and haemoglobin values in non-anaemic healthy persons. *Br. Med. J.* **3**, 731-733 (1972).

26. Koivunen, P. & Kietzmann, T. Hypoxia-Inducible Factor Prolyl 4-Hydroxylases and Metabolism. *Trends Mol. Med.* (2018).

27. Voss, J. D., Masuoka, P., Webber, B. J., Scher, A. I. & Atkinson, R. L. Association of elevation, urbanization and ambient temperature with obesity prevalence in the United States. *Int. J. Obes. (Lond)* **37**, 1407-1412 (2013).

28. Woolcott, O. O., Ader, M. & Bergman, R. N. Glucose homeostasis during short-term and prolonged exposure to high altitudes. *Endocr. Rev.* **36**, 149-173 (2015).

29. Bakris, G. L. *et al*.
Effects of a novel anemia treatment, FG-4592 - an oral hypoxia-inducible prolyl hydroxylase inhibitor (HIF-PHI) on blood pressure and cholesterol in patients with chronic kidney disease *J. Clin. Hypertension* **14**, 487-489 (2013).

30. Olson, E. *et al*. Short-term treatment with a novel HIF-prolyl hydroxylase inhibitor (GSK1278863) failed to improve measures of performance in subjects with claudication-limited peripheral artery disease. *Vasc. Med.* **19**, 473-482 (2014).

Acknowledgements This study was supported by the Academy of Finland Grants 266719 and 308009, the S. Jusélius Foundation, the Emil Aaltonen Foundation and the Jane and Aatos Erkko Foundation to P.K. NFBC1966 received financial support from University of Oulu Grant no. 65354 and 24000692, Oulu University Hospital Grant no. 2/97, 8/97 and 24301140, Ministry of Health and Social Affairs Grant no. 23/251/97, 160/97, 190/97, National Institute for Health and Welfare, Helsinki Grant no. 54121, Regional Institute of Occupational Health, Oulu, Finland Grant no. 50621, 54231, ERDF European Regional Development Fund Grant no. 539/2010 A31592, the European Union's Horizon 2020 research and innovation programme grant agreement no. 633595 (DynaHEALTH), no. 733206 (LIFECYCLE) and (iHEALTH), Academy of Finland, University Hospital Oulu and NHLBI grant 5R01HL087679-02 through the STAMPEED program.

We thank the late professor Paula Rantakallio (launch of NFBC1966), the participants in the 46-year study and the NFBC project center. The Young Finns Study has been financially supported by the Academy of Finland: grants 286284 (T.L) , 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; and Diabetes Research Foundation of Finnish Diabetes Association.

Author contributions P.K., J.A., J.K., J.T., R.S. and J.M. designed the analyses. V.K. performed the statistical analyses. J.A., J.T., V.K., J.K., J.M., S.K.-K. and P.K. analyzed the data. M.-R.J., J.A., S.K.-K., T.T., M.A.-K. and O.T.R. designed individual studies. J.A., J.T., R.S., E.Y.D., J.K., T.T., M.A.-K., J.A. K.P., O.T.R., M.-R.J. and P.S. collected and generated data in individual studies. P.K., M.-R.J., J.A., J.T., V.K., J.K., J.M. and S.K.-K. wrote the manuscript.

Competing financial interests The authors have no competing financial interests.

Materials and Correspondence Correspondence should be addressed to P.K. (peppi.koivunen@oulu.fi) or M.-R.J. (m.jarvelin@imperial.ac.uk) and request for materials to M.-R.J.

Figure legends and Figures

Figure 1 Association of Hb levels with anthropometric and metabolic measures in mice and human. a, Association of Hb levels with body weight in 1-year-old C57Bl/6 male mice (n = 31). **b,** Association of the highest quartile (High Hb) and the lowest quartile (Low Hb) of Hb levels with glucose tolerance test (GTT) area under the curve (AUC) (**b**) and HOMA-IR scores, respectively, in the 1-year-old C57Bl/6 male mice (**c**). **d,** Association of Hb levels with BMI in NFBC1966 at age of 46. Unadjusted regression line (blue) with 95% confidence intervals (gray) of blood Hb levels (g/l) with BMI (kg/m^2). **e,** Forest plot representing of the effect size of association of Hb levels with log(BMI), waist and hip circumference and waist-hip ratio in NFBC1966 (red), YFS (blue) and meta-analysis (black), respectively, and Hb levels with body fat percentage, visceral fat area and body muscle mass ratio in NFBC1966, respectively. Effects sizes are reported in SD units. Data in **a** is average ± SEM. ***P <0.001. The data in **b** was adjusted sex, smoking and physical activity.

Figure 2 Hb levels associate with key metabolic parameters. Forest plots representing (**a**) the effect sizes of the association in SD units of Hb levels with fasting glucose and log(fasting insulin) levels, log(HOMA-IR) and log(HOMA-B) indexes, log(area under the curve (AUC) of glucose in oral glucose tolerance test (OGTT))*, log(AUC of insulin in OGTT)* and 2 h glucose in levels in a 2 h OGTT*, log(Matsuda index)*, systolic and diastolic blood pressure, fasting serum cholesterol, LDL cholesterol, HDL cholesterol and log(triglyceride) and log(high-sensitivity C-reactive protein (CRP)) levels in meta-analysis of NFBC1966 at age of 46 and YFS at mean age of 42 years, respectively. Asterisk indicates the associations only analyzed in NFBC1966. The data were adjusted sex, smoking and physical activity (blue), and additionally

for BMI (red). **b,** Effect sizes of association in SD units of Hb levels with log(BMI), waist and hip circumference and waist-hip ratio in NFBC1966 at age of 31 (red) and 46 (blue) years, respectively. The *P* values represent the difference in effect size between 31 and 46 years. **c**, Effect sizes of the association in SD units of Hb levels with fasting glucose and log(fasting insulin) levels, HOMA-IR and HOMA-B indexes, systolic and diastolic blood pressure, fasting serum cholesterol, LDL cholesterol, HDL cholesterol and triglyceride levels and log(high-sensitivity CRP) in NFBC1966 at age of 31 (red) and 46 (blue) years, respectively. The *P* values for the difference in effect size between age 31 and 46 years are indicated. The data in **b** and **c** were adjusted for sex, smoking and physical activity, and in **c** for BMI.

Figure 3 Higher Hb levels associate with metabolic signatures of adiposity, insulin resistance and cardiovascular risk. Effect sizes in SD units of association of Hb levels with systemic metabolite levels in random effects meta-analysis of NFBC1966 at age of 46 and YFS at mean age of 42 years. The data were adjusted for BMI, sex, smoking and physical activity and corrected for multiple testing.

Figure 4 Lower Hb levels associate with decreased oxygen consumption at rest and upregulation of hypoxia-induced genes. a, Associations of Hb levels with oxygen consumption at rest (Effect size (Beta) = 0.26, *P* = 0.02 for meta-analysis of both genders), oxygen consumption at rest with body fat percentage (Effect size (Beta) = 0.19, *P* = 0.001 for meta-analysis of both genders) and Hb levels with body fat percentage (Effect size (Beta) = 0.14, *P* = 0.057 for meta-analysis of both genders) in a subpopulation (n = 123) of NFBC1966 at age of 31. **b**, Gene-set enrichment analysis of the YFS lowest Hb quartile (Hb < 132 g/l, n = 392) vs. YFS highest Hb quartile (Hb > 152 g/l, n = 371) and hypoxia-induced genes. Cohort n = 1,636. The

analysis was adjusted for sex, age, numbers of thrombocytes and leukocytes, the first five principal components of the transcriptomics data, BMI, research center and three technical microchip variables. **c,** Hb levels, body weight, fasting blood glucose, fasting serum insulin, HOMA-IR, 2h blood glucose levels in GTT, total cholesterol and triglyceride levels of 3-month-old C57Bl/6 male mice (n = 21) before (baseline) and 2 weeks after venesection (post venesection). **d**, Schematic summary of the discovered associations. Hypothesized pathways are depicted by dashed lines. Data in **c** is average ± SEM. *P <0.05, **P <0.01, ***P <0.001.

Figure 1

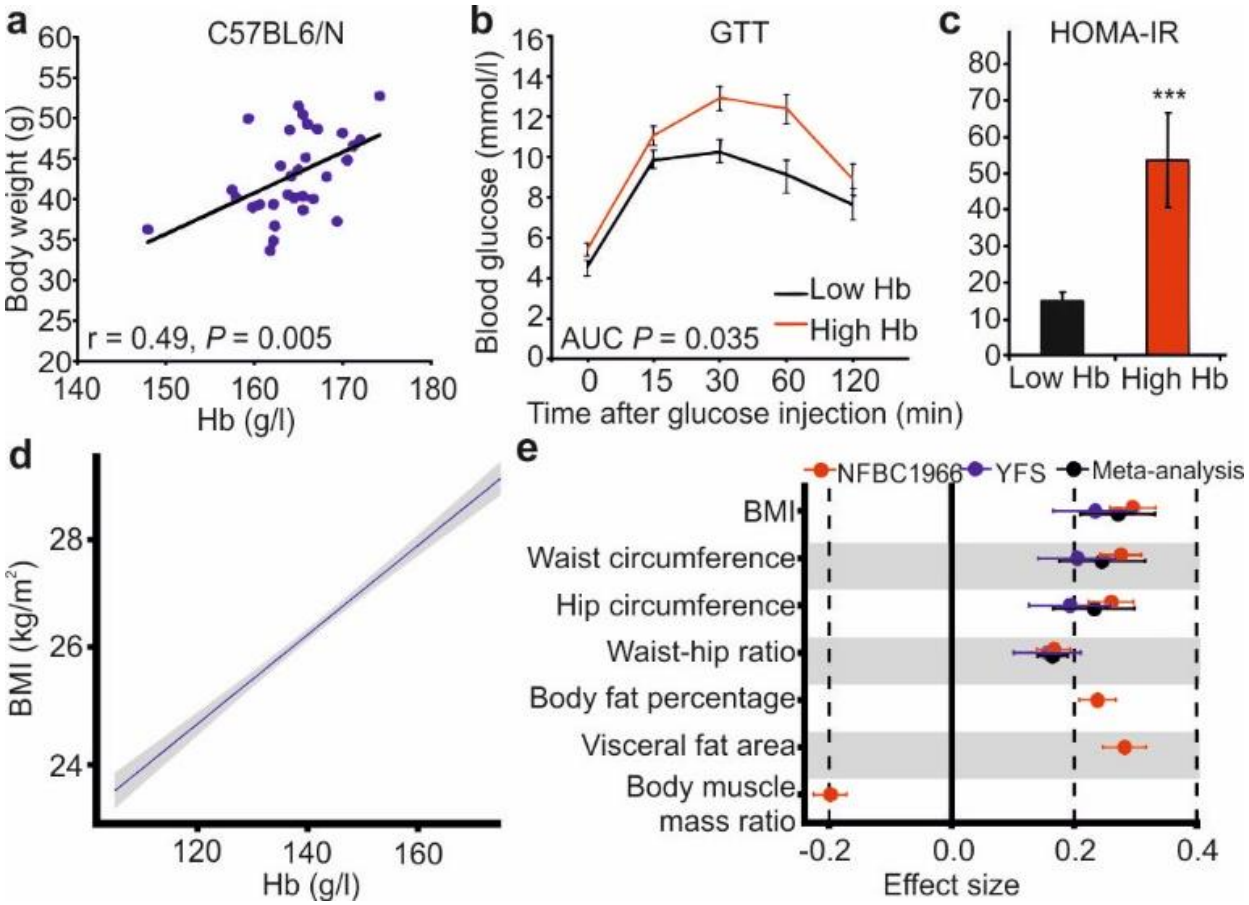

a
C57BL6/N
Body weight (g)
r = 0.49, P = 0.005
Hb (g/l)
b
GTT
Blood glucose (mmol/l)
Low Hb
High Hb
AUC P = 0.035
Time after glucose injection (min)
c
HOMA-IR
***
Low Hb
High Hb
d
BMI (kg/m²)
Hb (g/l)
e
NFBC1966
YFS
Meta-analysis
BMI
Waist circumference
Hip circumference
Waist-hip ratio
Body fat percentage
Visceral fat area
Body muscle mass ratio
Effect size

Figure 2

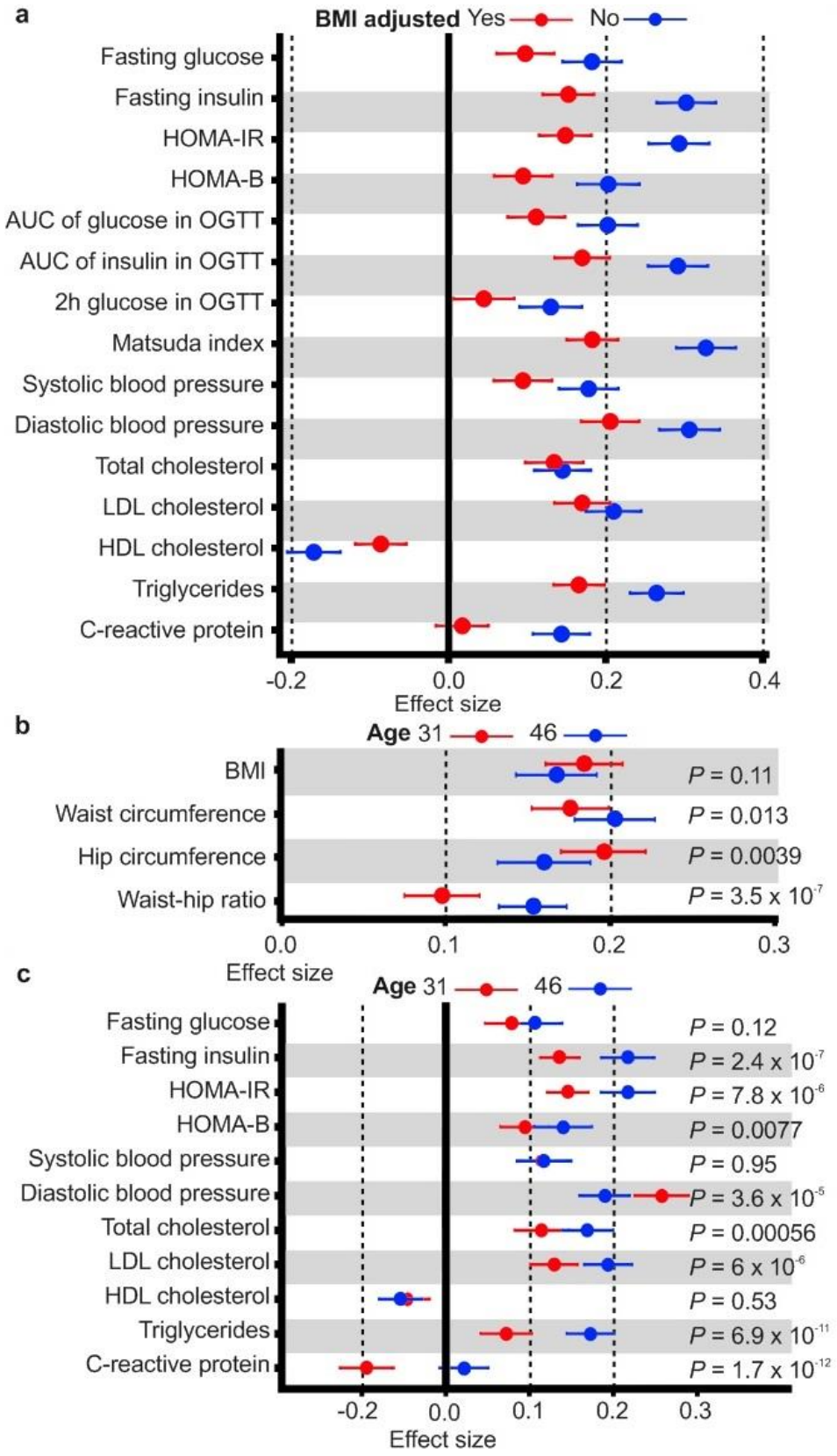

a
BMI adjusted Yes No
Fasting glucose
Fasting insulin
HOMA-IR
HOMA-B
AUC of glucose in OGTT
AUC of insulin in OGTT
2h glucose in OGTT
Matsuda index
Systolic blood pressure
Diastolic blood pressure
Total cholesterol
LDL cholesterol
HDL cholesterol
Triglycerides
C-reactive protein
-0.2
0.0
0.2
0.4
Effect size
b
Age 31 46
BMI
Waist circumference
Hip circumference
Waist-hip ratio
P = 0.11
P = 0.013
P = 0.0039
P = 3.5 x 10-7
0.0
0.1
0.2
0.3
Effect size
c
Age 31 46
Fasting glucose
Fasting insulin
HOMA-IR
HOMA-B
Systolic blood pressure
Diastolic blood pressure
Total cholesterol
LDL cholesterol
HDL cholesterol
Triglycerides
C-reactive protein
P = 0.12
P = 2.4 x 10-7
P = 7.8 x 10-6
P = 0.0077
P = 0.95
P = 3.6 x 10-5
P = 0.00056
P = 6 x 10-6
P = 0.53
P = 6.9 x 10-11
P = 1.7 x 10-12
-0.2
0.0
0.1
0.2
0.3
Effect size

Figure 3

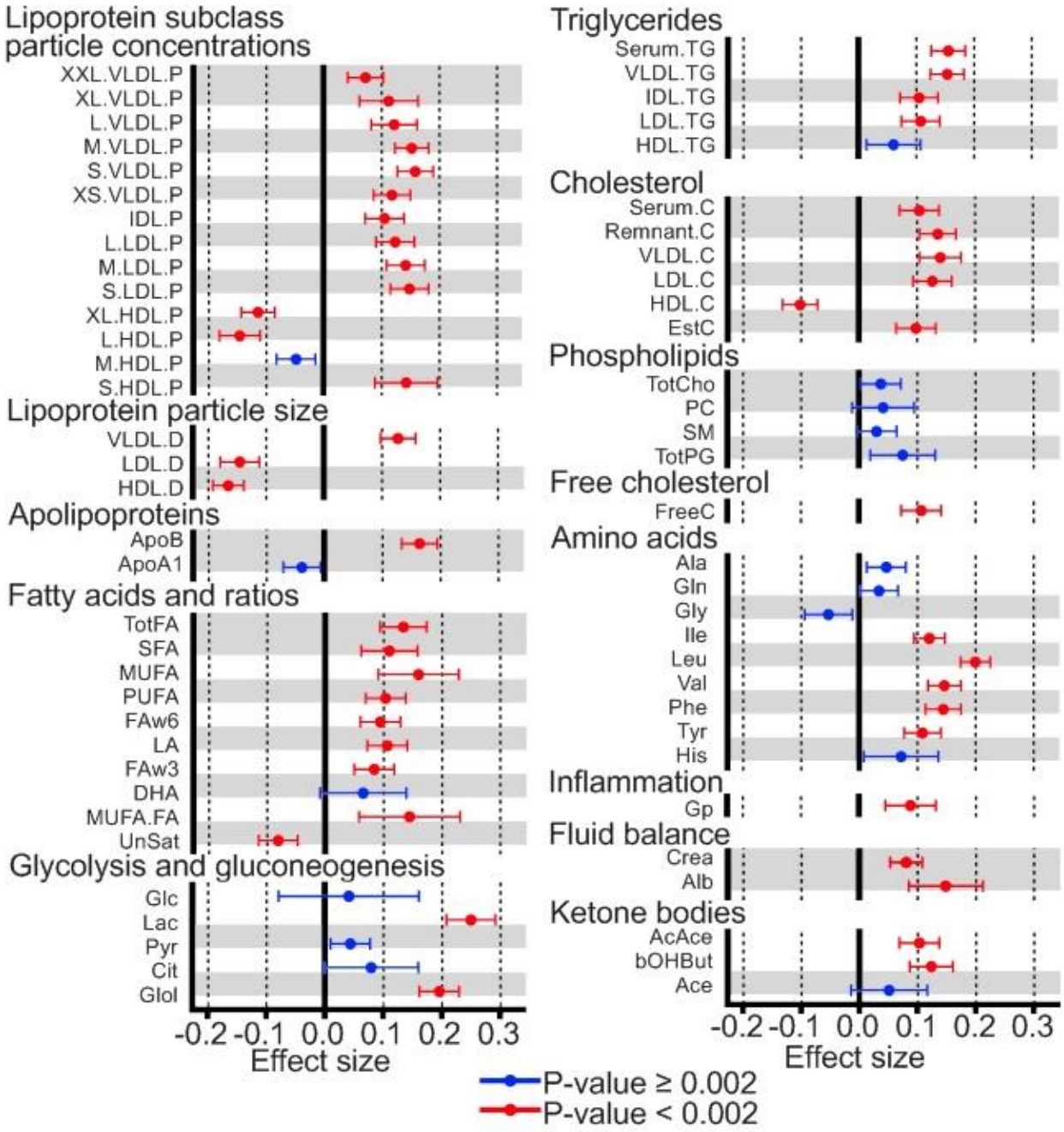
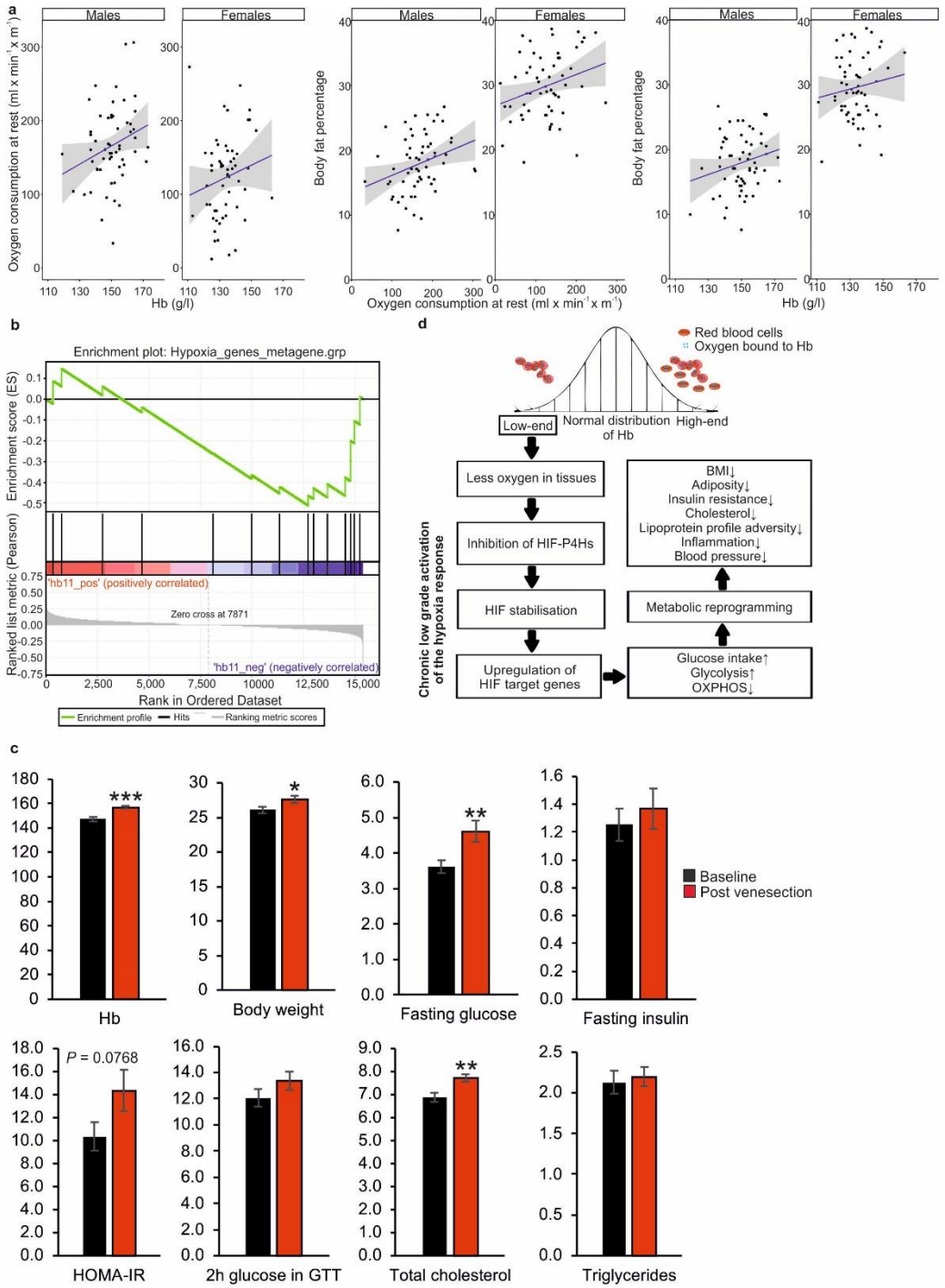

Lipoprotein subclass particle concentrations
XXL.VLDL.P
XL.VLDL.P
L.VLDL.P
M.VLDL.P
S.VLDL.P
XS.VLDL.P
IDL.P
L.LDL.P
M.LDL.P
S.LDL.P
XL.HDL.P
L.HDL.P
M.HDL.P
S.HDL.P
Lipoprotein particle size
VLDL.D
LDL.D
HDL.D
Apolipoproteins
ApoB
ApoA1
Fatty acids and ratios
TotFA
SFA
MUFA
PUFA
FAw6
LA
FAw3
DHA
MUFA.FA
UnSat
Glycolysis and gluconeogenesis
Glc
Lac
Pyr
Cit
Glol
-0.2 -0.1 0.0 0.1 0.2 0.3
Effect size
Triglycerides
Serum.TG
VLDL.TG
IDL.TG
LDL.TG
HDL.TG
Cholesterol
Serum.C
Remnant.C
VLDL.C
LDL.C
HDL.C
EstC
Phospholipids
TotCho
PC
SM
TotPG
Free cholesterol
FreeC
Amino acids
Ala
Gln
Gly
Ile
Leu
Val
Phe
Tyr
His
Inflammation
Gp
Fluid balance
Crea
Alb
Ketone bodies
AcAce
bOHBut
Ace
-0.2 -0.1 0.0 0.1 0.2 0.3
Effect size
P-value ≥ 0.002
P-value < 0.002

Figure 4


a
Males
Females
Oxygen consumption at rest (ml x min⁻¹ x m⁻²)
Hb (g/l)
Body fat percentage
Oxygen consumption at rest (ml x min⁻¹ x m⁻²)
b
Enrichment plot: Hypoxia_genes_metagene.grp
Enrichment score (ES)
Ranked list metric (Pearson)
'hb11_pos' (positively correlated)
Zero cross at 7871
'hb11_neg' (negatively correlated)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
d
Red blood cells
Oxygen bound to Hb
Low-end
Normal distribution of Hb
High-end
Less oxygen in tissues
Inhibition of HIF-P4Hs
HIF stabilisation
Upregulation of HIF target genes
Glucose intake↑
Glycolysis↑
OXPHOS↓
Metabolic reprogramming
BMI↓
Adiposity↓
Insulin resistance↓
Cholesterol↓
Lipoprotein profile adversity↓
Inflammation↓
Blood pressure↓
Chronic low grade activation of the hypoxia response
c
Baseline
Post venesection
Hb
Body weight
Fasting glucose
Fasting insulin
P = 0.0768
HOMA-IR
2h glucose in GTT
Total cholesterol
Triglycerides

Methods

Study populations

We tested our hypothesis in two Finnish populations, Northern Finland Birth Cohort 1966 (NFBC1966)[1, 2] and Cardiovascular Risk in Young Finns Study (YFS)[3]. The NFBC1966 originally included 12,068 mothers who gave birth to 12,231 live born children (96.3% of all estimated births in the two northernmost provinces of Finland, Oulu and Lapland, during the year 1966). The data collection started in 1965 when the mothers were pregnant and so far, data has been collected at ages 1, 14, 31 and 46 years. At 31- and 46-years, the data collection included postal questionnaires and clinical examinations with anthropometric and blood pressure measurements and blood sampling (n = 3624 at 31-years and n = 5351 at 46-years, Extended data Fig. 1). Maximal exercise test was performed for a subpopulation at 31-years (n = 123).

Blood analyses

Blood samples of NFBC1966 were taken after an overnight fasting period, centrifuged immediately and before analysing samples of 31-years were stored firstly at −20 °C and later at −80 °C, whereas samples of 46-years were analysed immediately without storing. NFBC1966 blood samples were analysed in NordLab Oulu (former name Oulu University Hospital, Laboratory), a testing laboratory (T113) accredited by Finnish Accreditation Service (FINAS) (EN ISO 15189). For YFS blood analyses, see[3].

Blood haemoglobin (Hb) level and red blood cell parameters

Blood Hb and red cell parameters of NFBC1966 were determined using a Coulter STKR analyser (Beckman Coulter, Fullerton, CA, USA) at age 31-years and a Sysmex XE-2100 analyser (Sysmex Corporation, Kobe, Japan) at age 46-year. The whole blood Hb levels of NFBC1966 were determined

using spectrophotometric methods, red blood cells (RBC) were measured using electric resistance detecting methods (impedance technology) with hydrodynamic focusing. At age 31-years hematocrit (HCT) was measured by the sum of RBC counted in a specified volume of diluted blood and at age 46-years HCT were determined applying the RBC pulse-height detection.

Serum lipids

Serum cholesterol (total, HDL and LDL) and triglycerides of NFBC1966 were determined using a Hitachi 911 automatic analyser and commercial reagents (Roche, Boehringer Mannheim, Germany) at 31-years[4, 5] and an enzymatic assay method (Advia 1800; Siemens Healthcare Diagnostics Inc., Tarrytown, NY, USA) at 46-years.

High sensitivity C-reactive protein

High sensitivity C-reactive protein (CRP) in NFBC1966 was analysed by an immunoenzymometric assay (Medix Biochemica, Espoo, Finland) at 31-years and by an immune nefelometric assay at 46-years (BN ProSpec, Siemens Healthcare Diagnostics Inc., Newark, DE, USA).

Glucose and insulin levels and fasting indices

At age 31-years fasting plasma glucose and serum insulin levels of NFBC1966 were determined by glucose dehydrogenase method (Granutest 250, Diagnostica Merck, Darmstadt, Germany) and by radioimmunoassay (Pharmacia Diagnostics, Uppsala, Sweden), respectively. At 46-years fasting plasma glucose and serum insulin levels of NFBC1966 were analysed by an enzymatic dehydrogenase method (Advia 1800, Siemens Healthcare Diagnostics, Tarrytown, NY, USA) and by a chemiluminometric immunoassay (Advia Centaur XP, Siemens Healthcare Diagnostics, Tarrytown, NY, USA), respectively.

To evaluate the insulin resistance and β-cell function we calculated fasting indices HOMA-IR (fasting plasma glucose x fasting serum insulin / 22.5) and HOMA2-β ((20 x fasting serum insulin) / (fasting plasma glucose – 3.5) x 100).

Oral glucose tolerance test and diabetes

A two hour oral glucose tolerance test (OGTT) was performed after overnight (12 hour) fasting period for NFBC1966 at 46-years (n = 4446). Medication for diabetes or just before test measured capillary blood glucose level >8.0 mmol/l were used as exclusion criteria. Both serum insulin and plasma glucose were measured at baseline and 30, 60 and 120 minutes after 75g glucose intake. OGTT glucose and insulin values were used to calculate insulin and glucose area under curve (glucose-AUC and insulin-AUC), Matsuda index for insulin sensitivity (10 000 x ((fasting plasma glucose x fasting serum insulin) x ((fasting plasma glucose + 30 min plasma glucose + 60 min plasma glucose + 120 min plasma glucose) / 4) x ((fasting serum insulin + 30 min serum insulin + 60 min serum insulin + 120 min serum insulin) / 4)))[6].

Serum metabolomics

Metabolic measures were obtained using the same high-throughput serum NMR metabolomics platform in the same analysis laboratory as described previously[7]. This methodology provided information on 228 serum measures, including lipoprotein subclass distribution and lipoprotein particle concentration, low molecular weight metabolites, such as amino acids, 3-hydroxybutyrate and creatinine, and detailed molecular information on serum lipids. The metabolite data provided by this platform has been utilized in various studies in epidemiology and genetics that have been recently reviewed[8]. The method provides metabolic measures in concentration units facilitating interpretation and replication of findings.

Anthropometric measurements

Body weight and height and waist and hip circumferences of NFBC1966 participants were measured at 31- and 46-years. Body weight was measured with digital scale, which was calibrated regularly. Height was measured twice (mean of the two measurements was used) by using standard and calibrated stadiometer. Finally, body mass index (BMI) was calculated as the ratio of weight and height squared. Waist and hip circumferences were measured twice and mean of the two measurements was used. The waist-hip ratio was assessed as the ratio between circumferences of the waist (at the level midway between lowest rib margin and the iliac crest) and the hip (at the widest trochanters). In addition, body fat mass, fat percentage, muscle mass and visceral fat area of NFBC1966 participants at 46-years were measured by InBody 720 bioelectrical impedance analyser (Biospace Co., Ltd., Seoul, Korea). All anthropometric measurements were done after overnight (12 h) fasting period. For YFS, see[3].

Brachial blood pressure measurement

In NFBC1966 brachial systolic and diastolic blood pressure was measured two times at 31-years and three times at 46-years with 1 min interval after 15 min of rest on the right arm of the seated participants using an automated oscillometric blood pressure device and appropriately sized cuff (Mercury sphygmomanometer at 31-years and Omron Digital Automatic Blood Pressure Monitor Model M10-IT at 46-years). Finally, the mean of two lowest systolic values and their diastolic values was used in the analyses.

Confounders and exclusions

Smoking and physical activity were used as confounders. NFBC1966 participants reported their smoking habits at 31- and 46-years and participants we categorised them as never, former and current smokers at both time points. Physical activity was inquired and categorised as inactive, moderate or

high. Individuals using combined oral contraceptives, hormone replacement therapy or serum lipids lowering drugs were omitted from the metabolomics analyses, as these have been shown to greatly influence these measures (total n = 957). Individuals with known diabetes before clinical examinations were omitted from analyses. Diabetes was defined according to self-reported diagnoses and medications, hospital outpatient and inpatient registers and medication registers from Social Insurance Institution of Finland. For YFS, see[3].

Analyses in mice

All animal experiments were performed according to protocols approved by the Provincial State Office of Southern Finland (license number ESAVI/6154/04.10.07/2014). The investigations conform to the Guiding Principles for research involving animals. Male C57Bl/6N mice (Charles River, Germany) were maintained in plastic cages in a constant 21°C temperature with a 12 h light-dark cycle and had free access to food. After weaning, the mice were fed Teklad Global 18% Protein Rodent Diet (18.6% protein, 6.2% fat, 44.2% carbohydrates; Envigo) and at 5-weeks of age the mice were transferred to Teklad Global 19% Protein Extruded Rodent Diet (19.0% protein, 9.0% fat, 44.9% carbohydrates; Envigo) to slightly support the increase in adipocity upon aging. At 11-months of age the mice were subjected to glucose tolerance test (GTT). GTT was performed on mice fasted for 12 h and anesthetized with fentanyl/fluanisone and midazolam. Mice were injected intraperitoneally with 1 mg/g glucose, and blood glucose concentrations were monitored with a glucometer. Serum insulin values were determined with Rat/Mouse Insulin ELISA kit (EZRMI-13K; Millipore), and HOMA-IR scores were calculated from the glucose and insulin values. The mice were weighed, sampled for hemoglobin determined (by HemoCue Hb 201 hemoglobin meter) and sacrificed at the age of one year. For the venesection experiment 3-month-old C57Bl/6N male mice fed Teklad Global 18% Protein Rodent Diet were fasted for 12 h and anesthetized as above. Mice were weighted and their fasting blood glucose values and Hb levels were determined. Samples were collected for

determination of serum fasting insulin, total cholesterol and triglyceride levels. Mice were then injected intraperitoneally with 2 mg/g glucose and blood glucose concentration after 2 h was monitored. The mice were then subjected to venesection of 0.2 ml. After 14 days the same analyses (body weight, Hb, fasting glucose, fasting insulin, 2-h glucose in GTT, total cholesterol and triglycerides) were repeated and the mice were sacrificed. Statistical analyses for comparison between two groups were done with two-tailed Student's *t test.* Pearson's correlation coefficient was calculated to compare linear dependences between two variables. Areas under the curve were calculated by the summary measures method.

Statistical analyses

Regression models were used to assess the associations between Hb (main explanatory variable) and anthropometric and metabolic variables (outcome). To adjust for potential confounding factors, smoking, physical activity and sex were included as explanatory variables in the regression models. For the analyses using YFS data, age was also included as an explanatory variable. The analyses were conducted including or excluding observations with an absolute value more than 3 standard deviations off from the mean of Hb or the (potentially transformed) outcome variable. Linear associations were assumed for all continuous variables. To ease the comparison of magnitudes of association across outcomes, the effect size estimates are reported in standard deviation units. To account for multiple comparisons in the anthropometric and metabolome-wide analyses, principal component analysis was used to estimate the number of independent tests for a Bonferroni-style multiple testing correction. As 23 principal components explained 95% of the variation in the anthropometric and metabolic outcomes, a *P* value of 0.002 (0.05/23) was used as a nominal threshold for significance in these analyses. R 3.3.2 software was used for all statistical analyses (R Core Team (2013): R: a language and environment for statistical computing, Vienna, Austria, http://www.R-project.org/).

Anthropometric measures

BMI was log-transformed to eliminate skewness and heteroscedasticity in model residuals. Height was included as an additional explanatory variable in the regression models, excluding the model for BMI. The analyses were first conducted separately for the 46-age year follow-up in NFBC1966 and for YFS, and then meta-analyzed for those outcomes available in both cohorts using an inverse-variance weighted random-effects model. The meta-analysis effect sizes were calculated and reported in the units in which they were measured. For the NFBC1966 dataset, similar regression models were conducted for red blood cell counts (Eryt) and hematocrit (Hct) instead of Hb, to examine the association between these measures and anthropometric variables.

Metabolic parameters

Fasting insulin levels, Matsuda and HOMA indices, AUCs and trigyceride and hsCRP levels were log-transformed to ensure symmetry and homoscedasticity of model residuals. In the hsCRP variable, including zeros in the values, the transformation used was $log(x + c/2)$, where $c = min(x_{x \neq 0})$, *i.e.* the smallest non-zero value in the variable. The analyses were conducted for the 46-year follow-up data in NFBC1966 with both including and excluding BMI as an additional explanatory variable.

NMR metabolomics

In total, 150 metabolites from the NMR metabolomics panel were analyzed (Extended data Table 8). Of these, 144 were measured in concentrations, and cube-root transformation was applied to these variables[9]. The rest of the variables were left untransformed, except the degree of unsaturation, which had a notably skewed distribution and thus log-transformed to ensure homoscedasticity in model residuals. The users of combined oral contraceptives, hormone replacement therapy and lipid lowering drugs (total n = 957) were omitted because their influence on these measures[10,11]. The analyses were conducted first separately in the NFBC1966 (at 46 years) and YFS datasets, and then

meta-analyzed using inverse-variance weighted random-effects models. BMI was included as an additional explanatory variable.

Longitudinal models in NFBC1966

Extended linear models using generalized least squares were used to examine the longitudinal change in association between Hb levels and the outcomes available at two time points in the NFBC1966. Time of measurement (ages of 31 years or 46 years) was included in the model as a categorical variable, and interaction terms with time and all other variables (excluding sex and height where applicable) were included in the model to allow varying effects between cross-sectional measurements. The error terms were allowed to be correlated within each individual and heteroskedastic between the time points. These analyses were conducted using the *gls()* function available in the *nlme* package[12]. Due to the different methods used to measure hsCRP at 31 and 46 years, an inverse normal rank transformation was applied to set the values on a comparable scale. As the glucose concentrations at the NFBC1966 31-year follow-up were measured from whole blood these values were transformed to corresponding concentrations in plasma[13] for compatibility with the data at 46-year follow-up.

Cross-sectional and longitudinal analysis of Hb levels on metabolites

To analyse both cross-sectional and longitudinal effects of Hb levels on the metabolites, Hb levels were decomposed to mean Hb ($\overline{Hb_j}$) for each individual j and variation of Hb from the individual mean $\left(Hb_{jk} - \overline{Hb_j}\right)$ for each individual j at time point k (1 = 31 years, 2 = 46 years). The following model was fit for each metabolite separately:

$$metabolite_{jk} = \beta_0 + \beta_C\overline{Hb_j} + \beta_L\left(Hb_{jk} - \overline{Hb_j}\right) + \beta_T Time_k + \beta_S Sex_j + \beta_Z Z_{jk} + \beta_{TZ} Time_k Z_{jk} + \epsilon_{jk}.$$

This model allows simultaneous estimation of the cross-sectional (between-subject) effect β_C and the longitudinal (within-subject) effect β_L[14], while controlling for potentially time-varying covariates Z_{jk}, in this case BMI, smoking and physical activity. The error terms ϵ_{jk} were allowed to be correlated within each individual and heteroskedastic between the time points. These analyses were conducted using the *gls()* function available in the *nlme* package[2].

Maximal exercise test subsample in NFBC1966

Regression models of the effects of Hb levels on oxygen consumption at rest divided by height (VO_2rest/h), VO_2rest/h on body fat percentage and Hb levels on body fat percentage were conducted for the subset of NFBC1966 participants at 31 years who underwent a maximal exercise test with direct measurement of oxygen consumption (M901, ergospirometer, Medikro, Finland)[15]. These analyses included sex as an additional explanatory variable.

Mendelian randomization

Bidirectional Mendelian randomization (MR) was conducted to examine a potential causality of Hb levels on BMI and *vice versa*. In the MR approach, genetic variants predicting an observational exposure are used as proxies for the exposure. If these variants fulfil instrumental variable assumptions[16], then association between the variants and the outcome can give evidence for causation, eliminating problems of confounding or reverse causation. Using a P value threshold of 1×10^{-8}, 56 and 17 single nucleotide polymorphisms (SNPs) from independent genetic loci associated with BMI and Hb, respectively, were identified from European ancestry genome-wide meta-analyses for the two traits[17,18]. All BMI- and Hb-associated SNPs were available in the Hb and BMI meta-analysis results, respectively. After harmonizing the effect alleles for each SNP, the effect estimates and their standard errors of BMI-related SNPs were extracted from the summary statistics available for Hb and *vice versa*. Causal effects were estimated from these summary data using both the inverse-variance

weighted (IVW) method[19] and MR-Egger regression[20]. The IVW method is a weighted regression of SNP-outcome on SNP-exposure coefficients, weighted by the inverse of the squared SNP-outcome standard errors. The slope of this model is an estimate for the causal effect. The intercept is constrained at zero, implying the assumption that the genetic data predicts the outcome only through the exposure. MR-Egger regression relaxes this assumption by including an intercept term, which is an estimate for the average pleiotropic effect of a SNP on the outcome. Testing whether this intercept equals zero serves as a test for directional pleiotropy. Assuming that magnitude of the SNP-exposure associations across all variants are independent of their pleiotropic effects, the slope in MR Egger regression provides a consistent estimate of the causal effect. Sources of bias to these causal effect estimates include deviations of the SNP-exposure estimates from the true associations, use of weak instruments, and participant overlap in the meta-analysis from where the effect estimates were calculated. In order to quantify the potential bias, relative bias estimates (the reciprocals of the F parameters[21]) were calculated for the IVW method estimates and I_{GX}^2 statistics[22] for the MR Egger estimates.

Polygenic risk score analyses

We conducted polygenic risk score (PRS) analyses to further examine the association between Hb, BMI and lipids. We generated PRSs for all exposure-outcome combinations using exposure GWAS results[23-26] for SNP weights, optimised for each outcome in NFBC1966 at 46 years. We used PRSice software[27] for PRS calculation, using a clumping distance of 250kb and an r^2 threshold of 0.1. Ten first genetic principal components were added as covariates to account for population stratification. PRS was calculated for a range of p-value cut-offs (from 5×10^{-8} to 1 by 0.05 increments) and we report the PRS using a p-value cut-off giving the highest R^2 with the target outcome. As NFBC1966 is part of all base GWASs used for PRS calculation, we might expect some inflation in the PRS variance explained. As a sensitivity test, we tested optimising PRS using BMI GWAS results from

UKBiobank (http://www.nealelab.is/blog/2017/9/11/details-and-considerations-of-the-uk-biobank-gwas) on Hb in NFBC1966 at 46 years. The R^2 for the PRS using the external GWAS results was 0.35 % (positive association, p = 4e-4), compared to 0.17 % (positive association, p = 0.01) with GWAS results with NFBC1966 included, showing no evidence for inflation in the variance explained.

Gene Set Enrichment Analysis

To examine the hypothesized activation of hypoxia response pathway, a Gene Set Enrichment Analysis (GSEA)[28] was conducted for the subset of YFS participants at 34-49 years with Hb levels and whole-blood genome-wide expression profiling data (n = 1636)[29]. The tested hypoxia response pathway included 22 genes induced at least four-fold by hypoxia in human monocytes (Extended data Table 11)[30]. The GSEA was performed between YFS participants in lowest (Hb <132, n = 392) and highest (Hb > 152, n = 371) quartile of Hb distribution. The analysis was done adjusted for age, sex, BMI, YFS study center, leukocyte and thrombocyte counts, five first principal components of transcriptomics data and three technical variables from microarray hybridizations (chip, plate and well). The analysis was done with residual gene expression profiles calculated by linear regression of each gene expression probe and above mentioned variables. A nominal P value < 0.05 was considered significant.

References

1. Rantakallio, P. Groups at risk in low birth weight infants and perinatal mortality. *Acta Paediatr. Scand.* **193**, Suppl 193:1+ (1969).

2. Jarvelin, M. R. *et al*. Early life factors and blood pressure at age 31 years in the 1966 northern Finland birth cohort. *Hypertension* **44**, 838-846 (2004).

3. Raitakari, O. T. *et al*. Cohort profile: the cardiovascular risk in Young Finns Study. *Int. J. Epidemiol.* **37**, 1220-1226 (2008).

4. Wieland, H. & Seidel, D. A simple specific method for precipitation of low density lipoproteins. *J. Lipid Res.* **24**, 904-909 (1983).

5. Sugiuchi, H. *et al*. Direct measurement of high-density lipoprotein cholesterol in serum with polyethylene glycol-modified enzymes and sulfated alpha-cyclodextrin. *Clin. Chem.* **41**, 717-723 (1995).

6. Matsuda, M. & DeFronzo, R. A. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes Care* **22**, 1462-1470 (1999).

7. Soininen, P. *et al.* High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst* **134**, 1781–1785 (2009).

8. Würtz, P. *et al.* Quantitative Serum Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic Technologies. *Am. J. Epidemiol.* **18**6, 1084-1096 (2017).

9. Cox, N. J. Speaking Stata: Graphing model diagnostics. *Stata J.* **4**, 449–475 (2004).

10. Wang, Q. *et al*. Effects of hormonal contraception on systemic metabolism: cross-sectional and longitudinal evidence. *Int. J. Epidemiol.* **45**, 1445-1457 (2016).

11. Wurtz, P. *et al*. Metabolomic Profiling of Statin Use and Genetic Inhibition of HMG-CoA Reductase. *J. Am. Coll. Cardiol.* **67**, 1200-1210 (2016).

12. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-128, http://CRAN.R-project.org/package=nlme (2016).

13. Holtkamp, H. C., Verhoef, N. J. & Leijnse, B. The difference between the glucose concentrations in plasma and whole blood. *Clin Chim Acta* **59**, 41-49 (1975).

14. Fitzmaurice, G. M., Laird, N. M. & Ware, J. H. *Applied Longitudinal Analysis*, 2nd Edition. John Wiley and Sons, New Jersey (2011).

15. Tammelin, T., Nayha, S. & Rintamaki, H. Cardiorespiratory fitness of males and females of northern Finland birth cohort of 1966 at age 31. *Int. J. Sports Med.* **25**, 547-552 (2004).

16. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, (R1):R89-98 (2014).

17. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).

18. van der Harst P. *et al.* Seventy-five genetic loci influencing the human red blood cell. *Nature* **492**, 369-375 (2012).

19. Burgess, S., Butterworth, A. & Thompson S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 656-665 (2013).

20. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512-525 (2015).

21. Burgess, S., Davies, N. M., Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet. Epidemiol.* **40**, 597-608 (2016).

22. Bowden, J., Del Greco M. F., Minelli, C., Davey Smith, G., Sheehan, N. A. & Thompson J. R. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int. J. Epidemiol.* **45**, 1961-1974 (2016).

23. Dupuis J., *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105-116 (2010).

24. van der Harst P., *et al.* Seventy-five genetic loci influencing the human red blood cell. *Nature* **492**, 369-375 (2012).

25. Locke A. E., *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).

26. Global Lipids Genetics C. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274-1283 (2013).

27. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466-1468 (2015).

28. Subramanian, A. *et al*. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545-15550 (2005).

29. Raitoharju E. *et al.* Blood microRNA profile associates with the levels of serum lipids and metabolites associated with glucose metabolism and insulin resistance and pinpoints pathways underlying metabolic syndrome: the cardiovascular risk in Young Finns Study. *Mol. Cell. Endocrinol.* **391**, 41-49 (2014).

30. Bosco, M. C. *et al.* Hypoxia modifies the transcriptome of primary human monocytes: modulation of novel immune-related genes and identification of CC-chemokine ligand 20 as a new hypoxia-inducible gene. *J. Immunol.* **177**, 1941-1955 (2006).