Open access • Posted Content • DOI:10.1101/2021.04.17.21255471

Random glucose GWAS in 493,036 individuals provides insights into diabetes pathophysiology, complications and treatment stratification — Source link

Lagou, Lagou, Lagou, Longda Jiang ...+113 more authors

Institutions: Katholieke Universiteit Leuven, Wellcome Trust Sanger Institute, University of Oxford, Imperial College London ...+41 more institutions

Published on: 20 Apr 2021 - medRxiv (Cold Spring Harbor Laboratory Press)

Topics: Blood sugar regulation, Type 2 diabetes, Diabetes mellitus, Glucose Measurement and Genome-wide association study

Random glucose GWAS in 493,036 individuals provides insights into diabetes pathophysiology, complications and treatment stratification

Authorship:

Vasiliki Lagou[1-3*], Longda Jiang[4-5*], Anna Ulrich[5*], Liudmila Zudina[5], Karla Sofia Gutiérrez González[6-7], Zhanna Balkhiyarova[5,8-10], Alessia Faggian[5,8], Shiqian Chen[11], Petar Todorov[12], Sodbo Sharapov[13], Alessia David[14], Letizia Marullo[15], Reedik Mägi[16], Roxana-Maria Rujan[17], Emma Ahlqvist[18], Gudmar Thorleifsson[19], He Gao[20], Evangelos Evangelou [20-21], Beben Benyamin[22-24], Robert Scott[25], Aaron Isaacs[6,26-27], Jing Hua Zhao[28], Sara M Willems[6], Toby Johnson[29], Christian Gieger[30-32], Harald Grallert[30,32], Christa Meisinger[33], Martina Müller-Nurasyid[34-37], Rona J Strawbridge[38-41], Anuj Goel[42-43], Denis Rybin[44], Eva Albrecht[36], Anne U Jackson[45], Heather M Stringham[45], Ivan R Corrêa, Jr. [46], Farber-Eber Eric[47], Valgerdur Steinthorsdottir[19], André G Uitterlinden[48-49], Patricia B Munroe[29,50], Morris J Brown[29], Schmidberger Julian[51], Oddgeir Holmen[52], Barbara Thorand[31-32], Kristian Hveem[52], Tom Wilsgaard[53-54], Karen L Mohlke[55], Wolfgang Kratzer[51], Haenle Mark[51], Wolfgang Koenig[56-58], Bernhard O Boehm[59], Tricia M Tan[11], Alejandra Tomas[60], Victoria Salem[61], Inês Barroso[62], Jaakko Tuomilehto[63-65], Michael Boehnke[45], Jose C Florez[66-68], Anders Hamsten[38-39], Hugh Watkins[42-43], Inger Njølstad[53-54], H-Erich Wichmann[31], Mark J Caulfield[29,50], Kay-Tee Khaw[69], Cornelia van Duijn[6,70-71], Albert Hofman[48,72], Nicholas J Wareham[25], Claudia Langenberg[25], John B Whitfield[73], Nicholas G Martin[73], Grant Montgomery[73-74], Chiara Scapoli[75], Ioanna Tzoulaki[20-21], Paul Elliott[20,76-77], Unnur Thorsteinsdottir[19,78], Kari Stefansson[19,78], Evan L Brittain[79], Mark I McCarthy[1,80-81^], Philippe Froguel[5,82], Patrick M Sexton[83-84], Denise Wootten[83-84], Leif Groop[18,85], Josée Dupuis[44], James B Meigs[67-68,86], Giuseppe Deganutti[17], Ayse Demirkan[8,87], Tune H Pers[12], Christopher A Reynolds[17,88], Yurii S Aulchenko[6,89-90], Marika A Kaakinen[5,8#], Ben Jones[11#], Inga Prokopenko[8,10,82#], on behalf of the Meta-Analysis of Glucose and Insulin-related Traits Consortium (MAGIC).

* These authors contributed equally to this research.

^ Current address: Genentech, 1 DNA Way, South San Francisco, CA 94080.

# These authors jointly directed this research.

Affiliations:

1)Wellcome Centre for Human genetics, University of Oxford, Oxford, United Kingdom; 2)Human Genetics, Wellcome Sanger Institute, Hinxton, Cambridgeshire, UK; 3)VIB-KU Leuven Center for Brain and Disease Research, Leuven, Belgium; 4)Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia; 5)Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK; 6)Genetic Epidemiology Unit, Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands; 7)Department of Molecular Diagnostic, Clinical Laboratory, Clinica Biblica Hospital, San José, Costa Rica; 8)Department of Clinical and Experimental Medicine, School of Biosciences and Medicine, University of Surrey, Guildford, UK; 9)Department of Endocrinology, Bashkir State Medical University, Ufa, Russian Federation; 10)Institute of Biochemistry and Genetics, Ufa Federal Research Centre Russian Academy of Sciences, Ufa, Russian Federation; 11)Section of Endocrinology and Investigative Medicine, Imperial College London, London, UK; 12)Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Denmark; 13)Laboratory of Glycogenomics, Institute of Cytology and Genetics, Prospekt Akademika Lavrent'yeva 10, Novosibirsk, Russia; 14)Centre for Bioinformatics and System Biology, Department of Life Sciences, Imperial College London, London, UK; 15)Department of Evolutionary Biology, Genetic Section, University of Ferrara, Ferrara, Italy; 16)Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia; 17)Centre for Sport, Exercise and Life Sciences, Faculty of Health and Life Sciences, Coventry University, Alison Gingell Building, UK; 18)Lund University Diabetes Centre, Department of Clinical Sciences Malmö, Lund University, Malmö, Sweden; 19)deCODE genetics/Amgen, Inc., Reykjavik, Iceland; 20)Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK; 21)Department of Hygiene and Epidemiology, University of Ioannina Medical School,

Ioannina, Greece; 22)Australian Centre for Precision Health, University of South Australia, Adelaide, Australia; 23)Allied Health and Human Performance, University of South Australia, Adelaide, Australia; 24)South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia; 25)MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge, Cambridge, UK; 26)CARIM School for Cardiovascular Diseases and Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, the Netherlands; 27)Department of Physiology, Maastricht University, Maastricht, the Netherlands; 28)Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 29)Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 30)Research Unit of Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum München Research Center for Environmental Health, Neuherberg, Germany; 31)Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 32)German Center for Diabetes Research (DZD), München-Neuherberg, Germany; 33)Independent Research Group Clinical Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 34)Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology and Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany; 35)Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-University, Munich, Germany; 36)Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 37)Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University, Mainz, Germany; 38)Cardiovascular Medicine Unit, Department of Medicine, Solna, Karolinska Institutet, Stockholm, Sweden; 39)Center for Molecular Medicine, Karolinska University Hospital Solna, Stockholm, Sweden; 40)Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK; 41)Health Data Research UK; 42)Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK; 43)Wellcome Centre

for Human Genetics, University of Oxford, Oxford, UK; 44)Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA; 45)Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA; 46)New England Biolabs, 240 County Road, Ipswich, Massachusetts, USA; 47)Vanderbilt Translational and Clinical Cardiovascular Research Center, Nashville, Tennessee, USA; 48)Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands; 49)Department of Internal Medicine, Erasmus Medical Center, Rotterdam, the Netherlands; 50)NIHR Barts Cardiovascular Biomedical Research Centre, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 51)Ulm University Medical Centre, Department of Internal Medicine I, Ulm, Germany; 52)Department of Public Health and General Practice, Norwegian University of Science and Technology, Trondheim, Norway; 53)Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, Tromsø, Norway; 54)Department of Clinical Medicine, Faculty of Health Sciences, University of Tromsø, Tromsø, Norway; 55)Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA; 56)Deutsches Herzzentrum München, Technische Universität München, Munich, Germany; 57)DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich Germany; 58)Institute of Epidemiology and Medical Biometry, University of Ulm, Ulm, Germany; 59)Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore and Department of Endocrinology, Tan Tock Seng Hospital, Singapore; 60)Section of Cell Biology and Functional Genomics, Imperial College London, London, UK; 61)Department Bioengineering, Imperial College London, Bessemer Building, South Kensington Campus, London, UK; 62)EXCEED (Exeter Centre of Excellence for Diabetes Research), University of Exeter Medical School, Exeter, UK; 63)Public Health Promotion Unit, Finnish Institute for Health and Welfare, Helsinki, Finland; 64)Department of Public Health, University of Helsinki, Helsinki, Finland; 65)Diabetes Research Unit, King Abdulaziz University, Jeddah, Saudi Arabia; 66)Center for Genomic Medicine and Diabetes Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 67)Programs in Metabolism and Medical and Population Genetics, Broad

Institute, Massachusetts, USA; 68)Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA; 69)Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 70)Centre for Medical Systems Biology, Leiden, the Netherlands; 71)Nuffield Department of Population Health, University of Oxford, Oxford, UK; 72)Netherlands Consortium for healthy ageing, the Hague, the Netherlands; 73)QIMR Berghofer Medical Research Institute, Brisbane Australia; 74)Institute for Molecualr Bioscience, The University of Queensland, St Lucia, Australia; 75)Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy; 76)MRC Centre for Environment and Health, Imperial College, London, UK; 77)National Institute for Health Research Imperial College London Biomedical Research Centre, Imperial College London, UK; 78)Faculty of Medicine, University of Iceland, Reykjavík, Iceland; 79)Vanderbilt University Medical Center and the Vanderbilt Translational and Clinical Cardiovascular Research Center, Nashville, Tennessee, USA; 80)Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom; 81)Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom; 82)UMR 8199 - EGID, Institut Pasteur de Lille, CNRS, University of Lille, F-59000 Lille, France; 83)Drug Discovery Biology, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia ; 84)ARC Centre for Cryo-electron Microscopy of Membrane Proteins, Monash Institute of Pharmaceutical Sciences, Monash University, Parkville, Victoria, Australia; 85)Finnish Institute for Molecular Medicine (FIMM), Helsinki University, Helsinki, Finland; 86)General Medicine Division, Massachusetts General Hospital, Boston, Massachusetts, USA; 87)Department of Genetics, University Medical Center Groningen, the Netherlands; 88)School of Life Sciences, University of Essex, Colchester, UK; 89)Laboratory of Glycogenomics, Institute of Cytology and Genetics SD RAS, Prospekt Akademika Lavrent'yeva 10, Novosibirsk, Russia; 90)PolyOmica, ‘s-Hertogenbosch, The Netherlands.

Correspondence should be addressed to:

Prof Inga Prokopenko

Section of Statistical Multi-Omics,

Department of Clinical and Experimental Medicine, University of Surrey

Leggett Building, Daphne Jackson Road, Manor Campus,

Guildford, Surrey, UK, GU2 7WG

Phone: +441483684900;

Email: i.prokopenko@surrey.ac.uk

Dr Ben Jones

Section of Endocrinology and Investigative Medicine

Imperial College London

Du Cane Road

London, UK, W12 0NN

Phone: +442033130348

Email: ben.jones@imperial.ac.uk

Dr Marika Kaakinen

Section of Statistical Multi-Omics,

Department of Clinical and Experimental Medicine, University of Surrey

Leggett Building, Daphne Jackson Road, Manor Campus,

Guildford, Surrey, UK, GU2 7WG

Phone: +441483683365;

Email: m.kaakinen@surrey.ac.uk

Abstract

Homeostatic control of blood glucose requires different physiological responses in the fasting and post-prandial states. We reasoned that glucose measurements under non-standardised conditions (random glucose; RG) may capture diverse glucoregulatory processes more effectively than previous genome-wide association studies (GWAS) of fasting glycaemia or after standardised glucose loads. Through GWAS meta-analysis of RG in 493,036 individuals without diabetes of diverse ethnicities we identified 128 associated loci represented by 162 distinct signals, including 14 with sex-dimorphic effects, 9 discovered through trans-ethnic analysis, and 70 novel signals for glycaemic traits. Novel RG loci were particularly enriched in expression in the ileum and colon, indicating a prominent role for the gastrointestinal tract in the control of blood glucose. Functional studies and molecular dynamics simulations of coding variants of *GLP1R*, a well-established type 2 diabetes treatment target, provided a genetic framework for optimal selection of GLP-1R agonist therapy. We also provided new evidence from Mendelian randomisation that lung function is modulated by blood glucose and that pulmonary dysfunction is a diabetes complication. Thus, our approach based on RG GWAS provided wide-ranging insights into the biology of glucose regulation, diabetes complications and the potential for treatment stratification.

Main text

Genetic factors are important determinants of glucose homeostasis and type 2 diabetes (T2D) susceptibility. Heritability of both fasting glucose (FG) and T2D is high, at 35-40%[1] and 30-60%[2], respectively. To date, more than 400 genetic loci have been described for T2D[3,4].

Genome-wide association studies (GWAS) for glycaemic traits in individuals without diabetes have identified genetic predictors of blood glucose, insulin and other metabolic responses during fasting or after oral or intravenous glucose challenge tests[5-8]. However, physiological glucose regulation involves responses to diverse nutritional and other stimuli that were, by design, omitted from such studies. Blood glucose is frequently measured at different times throughout the day in clinical practice and research studies (random glucose; RG). Whilst RG is inherently more variable than standardised measures, we reasoned that, across a very large number of individuals, it may more comprehensively represent complex glucoregulatory processes occurring in different organ systems. Therefore, to identify and functionally validate genetic effects influencing RG, explore its relationships with other traits and diseases, and utilise these data to inform approaches to T2D treatment stratification, we performed the first large-scale trans-ethnic GWAS meta-analysis for RG in individuals without diabetes.

RG GWAS significantly expands the catalogue of glycaemia-related genetic associations

We undertook RG GWAS in 493,036 individuals without diabetes of European (n=479,482) and other ethnic (n=16,554) descent with adjustment for age, sex and time since last meal (where available), along with exclusion of extreme hyperglycaemia (RG>20 mmol/L) and individuals with diabetes (**Supplementary Table 1)**. The covariate selection was done upon extensive phenotype modelling **(Methods, Supplementary Table 2, Supplementary Figure 1a).** We identified 162 distinct signals ($P<10^{-5}$) within 128 genetic loci reaching genome-wide significance ($P<5x10^{-8}$) (**Figure 1a**, **Supplementary Table 3**). Seventy RG signals had not previously been reported for glycaemic traits (**Table 1, Supplementary Table 3**). In Europeans, while the UK Biobank (UKBB) study provided 83.8% of the total study size, 128 detected

signals out of the 143 were directionally consistent in UKBB and other contributing studies grouped together (**Supplementary Table 3**). Adjustment for last meal timing (**Supplementary Figure 1b**) reduced effect sizes for several loci, including *ITPR3*, *RREB1*, *RGS17*, *RFX6*/*VGLL2* and *SYNGAP1*, suggesting that these may be more related to the post-prandial state. *RREB1*, *RFX6* are transcription factors implicated in the development and function of pancreatic beta cells[9,10], and *ITPR3* is a calcium channel involved in islet calcium dynamics in response to glucose and G protein-coupled receptor (GPCR) activation[11]. Neither adjustment for body-mass index (BMI), nor a more stringent hyperglycaemia cut-off (RG>11.1 mmol/L or HbA1c≥6.5%) (**Supplementary Figure 1c-e**) materially changed the magnitude and significance of the RG effect estimates, although when all covariate models were individually applied, nine additional signals at genome-wide significance were identified in UKBB (**Table 1, Supplementary Table 4**).

Several of the 162 signals identified in Europeans showed nominal significance (P<0.05) in specific UKBB ethnic groups, with *GCK* (rs2908286, $r^2_{1000GenomesAllEthnicities}$=0.83 with rs2971670 lead in Europeans) reaching genome-wide significance in the African descent individuals alone (**Supplementary Table 3**). Among the novel RG signals, *USP47* was nominally significant in the individuals of African, *FAM46* and *ACVR1C* in the Indian and *TRIM59/KPNA4* and *ZC3H13* in Chinese UKBB ancestry. Trans-ethnic meta-analyses combining Europeans and the other four UKBB ancestral groups revealed seven novel RG signals, including those at *FOXN3, EPS8* and *ISG20L2* (**Table 1**). Overall, while being only 16,554 individuals larger in sample size than the European meta-analysis, the trans-ethnic analysis expanded the novel locus discovery for RG by one tenth (**Supplementary Table 5**).

Among established glycaemic trait signals, the well-known FG loci *G6PC2* ($P<5.86\times10^{-754}$) and *GCK* ($P<6.93\times10^{-301}$), with key roles in gluconeogenesis[12] and glucose sensing[13], respectively, showed the strongest associations with RG (**Supplementary Table 3**). We also observed two thirds of RG signals overlapping with T2D-risk loci (**Supplementary Figure 1e**), including *SLC30A8, DGKB, TCF7L2, GRB10* and *THADA*. The direction of effects at these loci between RG, T2D and homeostasis model assessment of beta-cell function/insulin resistance (HOMA-B/-IR)[6] (**Supplementary Figures 1e-f and 2**, **Supplementary Table 6**) were consistent with their epidemiological correlation. Notably, 14 established[14,15], such as *DGKB*, *THADA*, *RSPO3*, *G6PC2,* and novel, including *TRIM59*, *POP7*, *SLC43A2*, and *SGIP1*, loci showed sex-dimorphic effects (**Methods, Table 1**, **Figure 1a**, **Supplementary Table 3**). Fine-mapping the associations at RG loci through conditional analysis (**Table 1**) we found three independent coding nonsynonymous rare (minor allele frequency, MAF<1%) variants at *G6PC2* with predicted (rs2232326) and established (rs138726309, rs2232323)[16] deleterious effects (**Supplementary Table 7**). Within *GCK*, we observed five rare independent ($r^2_{1000GenomesAllEthnicities}<0.001$) non-deleterious variants associated with RG at genome-wide significance, including a novel 3'UTR rs2908276 for T2D, glycaemic traits or obesity (**Supplementary Table 7**).

Next, we sought to pinpoint the most plausible set of causal variants by calculating 99% credible sets for each of RG loci. In the Europeans only analysis, 19 RG signals were explained by one variant with posterior probability of ≥99% of being causal. For another 20 signals, a lead variant had a posterior probability >80% (**Figure 1b, Supplementary Table 8**). The credible sets were narrowed down in trans-ethnic RG meta-analysis (median credible set size 12.5 in the Europeans only, and 11.0 in the trans-ethnic analysis) (**Supplementary Tables 9 and 10**). This analysis helped to prioritise *GLP1R* for functional studies, in addition to the

already deeply characterised *G6PC2* and *CCND2*[17], all three with lead SNPs of low frequency (1%≤MAF<5%) and posterior probability >99% of being causal.

The lead RG-associated SNPs at *GLP1R, NEUROD1,* and *EDEM3* loci in our analysis were low-frequency coding variants (**Supplementary Figure 3**). *NEUROD1* (Neuronal Differentiation 1) and *EDEM3* (ER Degradation Enhancing Alpha-Mannosidase Like Protein 3) are plausible candidates for glucose homeostasis with the former reported for glucosuria[18] and the latter linked to renal function[19,20]. Additionally, lead variants at three previously reported for FG (*GCKR, TET2* and *RREB1*) and two novel RG (*NMT1*, *WIPI1*) loci were all common (MAF≥5%) coding variants (**Supplementary Figure 3**).

Functional and structural characterisation of RG-associated *GLP1R* coding variants provides a possible framework for T2D treatment stratification

The *GLP1R* gene, identified in our analysis and in previous T2D[21] and glycaemic trait[22] GWAS, encodes a class B G protein-coupled receptor (glucagon-like peptide-1 receptor; GLP-1R) that is an established target for glucose-lowering and weight loss in T2D using drugs such as exenatide (exendin-4) and semaglutide[23]. Within *GLP1R,* the lead missense variant at rs10305492 (A316T) had a strong (0.058 mmol/l per allele) RG-lowering effect, second by size only to *G6PC2* locus variants. Previous attempts to functionally characterise A316T and further *GLP1R* variants experimentally have been inconclusive[24], so we adopted a strategy based on measuring ligand-induced coupling to mini-$G\alpha_s$[25], representing the most proximal part of the $G\alpha_s$-adenylate cyclase-cyclic adenosine monophosphate (cAMP) pathway that links GLP-1R activation to insulin secretion. Mini-$G\alpha_s$ coupling efficiency was predictive of RG

effect for 16 *GLP1R* coding variants detected in the UKBB dataset with effect allele frequency ranging from common (G168S, rs6923761, P=4.40x10^{-5}) to rare (R421W, rs146868158, P=0.054) (**Figure 2a, Supplementary Table 11**), thereby linking differences in experimentally measured GLP-1R function to blood glucose homeostasis.

To probe whether *GLP1R* coding variation could be therapeutically as well as physiologically relevant, we also measured responses to several endogenous and pharmacological GLP-1R agonists. Focussing on the two directly genotyped *GLP1R* missense variants in UKBB, we observed that A316T (rs10305492-A) showed increased responses, and R421W (rs146868158-T) showed reduced responses, to all ligands except exendin-4 (both variants) and semaglutide (A316T only), in line with their RG effects (**Figure 2b**). Agonist-induced GLP-1R endocytosis with R421W was normal despite its signalling deficit, suggestive of biased agonism[26]. The imputed common G168S variant, with relatively small RG-lowering effect (β=-0.0013 [SE=3.14x10^{-4}]), also showed subtle increases in function.

To gain structural insights into *GLP1R* variant effects we performed molecular dynamics simulations of the human GLP-1R bound to oxyntomodulin[27] (**Extended Data Tables 1-6**). A316T has a single amino acid substitution in the core of the receptor transmembrane domain (**Figure 2c**) that leads to an alteration of the hydrogen bond network in close proximity (**Video S1**). In A316T, residue T316$^{5.46}$ replaced Y242$^{3.45}$ in a persistent hydrogen bond with the backbone of P312$^{5.42}$ one turn of the helix above T316$^{5.46}$ (**Figures 2d-e, Video S1**). This triggers a local structural rearrangement that could transmit to the intracellular G protein binding site through transmembrane helix 3 (TM3) and TM5. A structural water molecule was found close to position 5.46 in both A316T and WT (water cluster α5, **Figure 2f**). The same

298 water bridged the backbone of Y241$^{3.44}$ and A316$^{5.46}$ in WT, or the backbone of Y241$^{3.44}$ and
299 the side chain of T316$^{5.46}$ in A316T. Given the importance of conserved water networks in the
300 process of activation of class A GPCRs[28,29], the presence of a stable hydrated spot close to
301 position 5.46[30] corroborates this site as important for tuning the intracellular conformational
302 landscape of GLP-1R. Also, a stabilising role for the water molecules at the binding site of the
303 G protein (water cluster apha5, **Figure 2f**) cannot be ruled out. Note that our results differ
304 from a previous analysis of A316T dynamics[22], which used an early model that does not fully
305 capture the full structural features of the current active GLP-1R models.
306
307 In analogy with A316T, molecular simulations with the G168S variant indicate the formation
308 of a stable new hydrogen bond between the side chain of residue S168$^{1.63}$ and A164$^{1.59}$,
309 located one turn above on the same helix (**Video S2, Figure 2g**). This moved the C-terminal
310 end of TM1 closer to TM2 and reduced the overall flexibility of ICL1 (**Figure 2h**), which could
311 potentially alter the role of ICL1 in G protein activation. In contrast to A316T and G168S, the
312 site of mutation R421W is consistent with persistent interactions with the G protein.
313 Simulations predicted a propensity of R421W to interact with a different region of the G
314 protein β-subunit to that engaged by WT (**Figure 2i**).
315
316 For a broader view of the impact of *GLP1R* coding variation, we screened an additional 178
317 missense variants identified from exome sequencing[31] for exendin-4-induced mini-G_s
318 coupling and endocytosis (**Figures 2j-k, Supplementary Table 12**). 110 variants showed a
319 reduced response in either or both pathways ("LoF1"), and 67 displayed a specific response
320 deficit that was not fully explained by differences in GLP-1R surface expression ("LoF2"), with
321 many of these defects being larger than in the analysis in **Figure 2a**.

Overall, these data suggest *GLP1R* variation influences blood glucose levels in health and is likely to be a direct modifier of responses to drug treatment[32]. As some patients fail to respond adequately to GLP-1R agonist treatment, and others are particularly sensitive to side effects[33], this approach may feed into optimised treatment selection in T2D.

Functional annotation of RG associations and intestinal health

Previous T2D and glycaemic trait GWAS have primarily implicated pancreatic, adipose and liver tissues[3]. To leverage our RG GWA results to identify additional cell and tissue types with aetiological roles in glucose metabolism, we performed a range of complementary functional annotation analyses in relation to RG GWAS. DEPICT[34], which predicts enriched tissue types from prioritised gene sets (**Methods**), highlighted intestinal tissues including ileum and colon, as well as pancreas, adrenal glands, adrenal cortex and cartilage (False Discovery Rate<0.20) (**Figures 3a-b, Supplementary Tables 13a-c**). Similarly, CELLECT[35], which facilitates cell-type prioritisation based on single cell RNAseq datasets (**Methods**), identified large intestinal tissue as the second ranked only to pancreatic cell types (**Figure 3c, Supplementary Table 14)**; interestingly, RG variants were related particularly to enriched expression in pancreatic polypeptide (PP) cells, exceeding even the more conventionally implicated insulin-secreting beta cells. Supporting evidence was obtained from transcriptome-wide association study (TWAS) analysis (**Methods**), where we identified a total of 216 (119 unique) significant genetically driven associations across the ten tested tissues; 52 (26 unique) of highlighted genes are located at genome-wide significant RG loci (**Supplementary Tables 15a)**. TWAS signals in skeletal muscle showed the largest overlap with RG signals, such as *GPSM1*[36] and

346 *WARS*; with combined results from ileum and colon also highly enriched, including the novel
347 *NMT1* and the established *FADS1/3* and *MADD* genes (**Figure 1a, Supplementary Tables 15a-**
348 **b**). Moreover, epigenetic annotations using the GARFIELD tool highlighted significant
349 ($P<2.5\times10^{-5}$, **Methods**) enrichment of RG-associated variants in foetal large intestine, as well
350 as blood, liver and other tissues (**Supplementary Figure 4, Supplementary Table 16**). Adult
351 intestinal tissues are not available in GARFIELD except for colon. Prompted by multiple
352 analyses highlighting a potential role for the digestive tract in glucose regulation, we assessed
353 the overlap between our signals and those from the latest microbiome GWAS[37] (**Methods**)
354 and identified three genera sharing signals with RG at two loci: *Collinsella* and
355 *LachnospiraceaeFCS020* at *ABO-FUT2* and *Slackia* at *G6PC2* (**Figure 1a, Supplementary Table**
356 **17**). The *ABO-FUT2* locus effects on RG could be mediated by abundance of bacteria
357 producing glucose from lactose and galactose[38].
358
359 eQTL colocalization analyses, using eQTLgen blood expression data from 31,684 individuals[39]
360 and the COLOC2 approach (**Methods**), identified 14 loci with strong links (posterior
361 probability >50%) to gene expression data, including *SMC4*, *TRIM59*, *EIF5A2*, *TET2*, *COG5*,
362 *CHMP5*, *NFX1*, *FNBP4*, *MADD*, *RAPSN*, *WARS1*, *HBM*, *NUFIP2*, and *PPDPF* (**Supplementary**
363 **Table 18***).* This further supported elucidation of biological candidates at novel and established
364 glycaemic loci.
365
366 Finally, we observed associations at two RG loci (*GCKR, HNF1A*) with nine total plasma N-
367 glycome traits[40] at a Bonferroni corrected threshold (**Methods, Figure 1a, Supplementary**
368 **Table 19**). These traits represent highly branched galactosylated sialylated glycans (attached
369 to alpha1-acid protein - an acute-phase protein[41]), known to lead to chronic low-grade

inflammation[42,43] and an increased risk of T2D[44-46] that might be explained by the role of N-glycan branching of the glucagon receptor in the glucose homeostasis[47]. In addition, ten glycans showed association with five RG loci (*GCKR, HNF1A*, *BAG1*, *PLUT*, *ACVR1C*) loci at a suggestive level of significance (**Figure 1a**). Among them, three are attached to immunoglobulin G molecules[41] and their increased relative abundances are associated with a lower risk of T2D[48] and diminished inflammation status[49].

Analysis of genetic relationships between RG and other metabolic or non-metabolic traits

To quantify the shared genetic contribution between RG and other phenotypes, we estimated their genetic correlations using linkage-disequilibrium score regression analyses. We detected positive genetic correlations between RG, squamous cell lung cancer (r_g=0.28, *P*=0.0015), and lung cancer (r_g=0.12, *P*=0.037, **Figure 4, Supplementary Table 20**); as well as inverse genetic correlations with lung function related traits, such as forced vital capacity (FVC, r_g=-0.090, *P*=0.0059) and forced expiratory volume in 1 second (FEV1, r_g=-0.054, *P*=0.017) (**Figures 3a and 4, Supplementary Table 20**). To investigate this further, we conducted a bi-directional Mendelian Randomisation (MR) analysis, which suggested a causal effect of RG and T2D on lung function, including FEV1 ($\beta_{MR\text{-}RG}$=-0.60, *P*=0.0015; $\beta_{MR\text{-}T2D}$=-0.049, *P*=1.27x10^{-13}) and FVC ($\beta_{MR\text{-}RG}$=-0.61, *P*=3.5x10^{-4}; $\beta_{MR\text{-}T2D}$=-0.062, *P*=1.42x10^{-21}), but not *vice versa* (**Methods, Supplementary Table 21**). Previous observational studies have highlighted worsening lung function, as defined by FVC, in T2D patients[50,51]. More recently, it was shown that patients with diabetes are at an increased risk of death from the viral infection COVID-19[52], with pulmonary dysfunction contributing to mortality[53]. Our data therefore support the

393 causal effect of glycaemic dysregulation on a decline in lung function as a novel complication
394 of diabetes.
395
396 Genome-wide genetic correlation analyses also showed strong positive genetic correlation of
397 RG with FG (r_g=0.88, P=6.93×10^{-61}, **Figure 4, Supplementary Table 20**). We meta-analysed RG
398 studies other than UKBB with FG GWAS summary statistics[54], observing 77 signals reaching
399 nominal significance that were directionally consistent in both UKBB and RG+FG
400 (**Supplementary Table 3**), providing an additional support to our RG findings. Given the large
401 genetic overlap between RG, other glycaemic traits and T2D, we evaluated the ability of a
402 trait-specific polygenic risk score (PRS) to predict RG, T2D and glycated haemoglobin (HbA1c)
403 levels using UKBB effect estimates and the Vanderbilt cohort (**Methods**). The RG PRS
404 explained 0.58% of the variance in RG levels when individuals with T2D were included,
405 (**Supplementary Table 22**) and 0.71% of the variance after excluding those who developed
406 T2D within one year of their last RG measurement. The RG PRS performance was comparable
407 to that of the FG loci PRS (0.38% vs. 0.42% for T2D; 0.40% vs. 0.44% for HbA1c) indicating
408 wide similarities with the latter.
409
410 We previously highlighted diverse effects of FG and T2D loci on pathophysiological processes
411 related to T2D development by grouping associated loci in relation to their effects on multiple
412 phenotypes[6]. Cluster analysis of the RG signals with 45 related phenotypes identified three
413 separate clusters that give insights into the aetiology of glucose regulation and associated
414 disease states (**Methods, Figure 1a, Supplementary Table 23, Supplementary Figures 5a-d**).
415 Cluster 1 ("metabolic syndrome" cluster) clearly separated 33 loci with effects on higher
416 waist-to-hip ratio, blood pressure, plasma triglycerides, insulin resistance (HOMA-IR) and

417 coronary artery disease risk, as well as lower testosterone and sex hormone binding globulin
418 levels in men. Cluster 3 was characterised in particular by insulin secretory defects[6]. Cluster 2
419 was less clearly defined by a primary effect on insulin release *versus* insulin action[3], but
420 interestingly included a sub-cluster of 21 loci which exert protective effects on inflammatory
421 bowel disease. Moreover, cluster 2 was notable for generally reduced impact on T2D risk in
422 comparison to clusters 1 and 3, underscoring the partial overlap between genetic
423 determinants of glycaemia and T2D that is known to exist[55].
424
425 **Discussion**
426
427 Taking advantage of data from 493,036 individuals, we have expanded by 58 the number of
428 loci associated with glycaemic traits. By using RG, our analysis integrates genetic contributions
429 to a wider range of physiological stages than possible with FG or other standardised
430 measures. Moreover, the greater statistical power obtained from large trans-ethnic meta-
431 analysis improves confidence in identification of potentially causal variants, thereby helping
432 to prioritise loci for more detailed functional analyses in the future. Our observation of ligand-
433 specific responses to the A316T, G168S and R421W *GLP1R* variants provides a mechanism
434 that can explain why some individuals respond better or worse to particular GLP-1R-targeting
435 drugs. We note that other class B GPCRs identified in our current analysis and other glycaemic
436 or T2D GWAS include *GIPR*, *GLP2R*[3] and *SCTR*[21], all of which are investigational targets for T2D
437 treatment. Our functional annotation analyses point to underexplored tissue mediators of
438 glycaemic regulation, with several sources of evidence highlighting a likely role of the
439 intestine. This observation is compatible with the well-described and profound effects of
440 gastric bypass surgery on T2D resolution[56], as well as links between the intestinal microbiome

441 and responses to several diabetes drugs[57]. Finally, through Mendelian randomisation we
442 were able to identify a causal effect of glucose levels and T2D on lung function, demonstrating
443 the utility of this approach for the corroboration of findings from observational studies and
444 elevating lung dysfunction as a new complication of diabetes.

Methods

Phenotype definition and model selection for RG GWAS

We used RG (mmol/l) measured in plasma or in whole blood (corrected to plasma level using the correction factor of 1.13). Individuals were excluded from the analysis, if they had a diagnosis of T2D or were on diabetes treatment (oral or insulin). Individual studies applied further sample exclusions, including pregnancy, fasting plasma glucose equal to or greater than 7 mmol/l in a separate visit, when available, and having type 1 diabetes. Detailed descriptions of study-specific RG measurements are given in **Supplementary Table 1**. All studies were approved by local ethics committees and all participants gave informed consent. We examined the distributions of untransformed and natural logarithmic transformed RG in the first set of six available cohorts. We observed that RG was approximately normally distributed after natural log transformation. We then determined the variables that could have a significant effect on RG by fitting several regression models using naturally log-transformed RG as the outcome with age, sex, BMI and time since last meal as predictors. Modelling of RG revealed significant effects (P<0.05) of age, sex, BMI and time since last meal (accounted for as T, T^2 and T^3) in these cohorts (**Supplementary Table 2**). Compared to RG models without T, inclusion of T, T^2 and T^3 increased the proportion of variance explained in the range of 1-6%. Thus, inclusion of this covariate is potentially equivalent to 1-6% increase in study sample size. For the GWAS, we included individuals based on two RG cut-offs: <20 mmol/l (20) to account for the effect of extreme RG values and <11.1 mmol/l (11), which is an established threshold for T2D diagnosis. We then evaluated six different models in GWAS according to covariates included and cut-offs used: 1) age (A) and sex (S), RG<20 mmol/L (**AS20**), 2) age, sex and BMI (B), RG<20 mmol/L (**ASB20**), 3) age and sex, RG<11.1 mmol/L

(**AS11**), 4) age, sex and BMI, RG<11.1 mmol/L (**ASB11**), 5) age, sex, T, T^2 and T^3, RG<20 mmol/L (**AST20**) and 6) age, sex, T, T^2 and T^3 and BMI, RG<20 mmol/L (**ASTB20**). Apart from above, additional adjustments for study site and geographical covariates were also applied.

Genotyping and quality control

Commercial genome-wide arrays and the Metabochip[58] were used by individual studies for genotyping. Studies with genome-wide arrays undertook imputation of missing genotypes using at least the HapMap II CEU reference panel via MACH[59], IMPUTE[60] or MINIMAC[61] software (**Supplementary Table 1**). For each study, samples reflecting duplicates, low call rate, gender mismatch, or population outliers were removed. Low-quality SNPs were excluded by the following criteria: call rate <0.95, minor allele frequency (MAF) <0.01, minor allele count <10, Hardy-Weinberg *P*-value $<10^{-4}$. GWAS were performed with PLINK, SNPTEST, EMMAX, R package LMEKIN, Merlin, STATA, and ProbABEL (**Supplementary Table 1**).

GWAS in the UKBB

For the GWAS of the UKBB data we excluded non-white non-European individuals and those with discrepancies in genotyped and reported sex. For the RG definition, we used the same criteria as in the other studies described above. To control for population structure, we adjusted the analyses for six first principal components. The GWAS was performed using the BOLT-LMM v2.3 software[62,63] restricting the analyses to variants with MAF>1% and imputation quality>0.4.

RG meta-analyses

The GWAS meta-analysis of RG consisted of four components: (i) 37,239 individuals from 10 European GWAS imputed up to the HapMap 2 reference panel, (ii) 3,156 individuals from three European GWAS with Metabochip coverage, (iii) 21,083 individuals from two European GWAS imputed up to 1000 genomes reference panel and iv) 401,810 individuals of white European origin from the UKBB and (iv) 16,983 individuals from the Vanderbilt cohort imputed to the HRC panel. We imputed the GWAS meta-analysis summary statistics of each component to all-ancestries 1000 Genomes reference panel[64] using summary statistics imputation method implemented in the SS-Imp v0.5.5 software[65]. SNPs with imputation quality score <0.7 were excluded. We then conducted inverse variance meta-analyses to combine the association summary statistics from all components using METAL (version from 2011-03-25)[66]. We focused our meta-analyses on models AS20 (17 cohorts, N_{max}=481,150) and AST20 (when time from last meal was available in the cohort) (12 cohorts, N_{max}=438,678). For FHS cohort, where no information was available for individuals with RG>11.1 (an established threshold for 2hGlu concentration, which is a criterion for T2D diagnosis), AS11 model results were used. In order to maximise the association power while taking into account T, we also performed meta-analysis using AST20 (when time from last meal was available in the cohort) combined with AS20 (otherwise) and we termed this analysis as AS20+AST20 in the following text (17 cohorts, N_{max}=480,250).

A signal was considered to be associated with RG if it had reached genome-wide significance ($P<5x10^{-8}$) in the meta-analysis of UKBB and other cohorts in either of our two models of interest (AS20) or (AST20) or in their combination (AS20+AST20). We report the *P*-value from the combined model, unless otherwise stated. Full results from all models are provided in the **Supplementary Table 3**. All the follow-up analyses were conducted using the combined AS20+AST20 model. We checked for nominal significance ($P<0.05$) and directional consistency

of the effect sizes for the selected leads in the combined model in UKBB results vs other cohort results. We further extended the check between UKBB results and meta-analysis of other cohorts including FG GWAS meta-analysis[54] excluding overlapping cohorts. This meta-analysis conducted in METAL was sample size and *P*-value based due to the measures being at different scale (natural logarithm transformed RG and untransformed FG).

Trans-ethnic analyses and meta-analysis

We performed GWAS in those non-European populations within UKBB that had a sample size of at least 1,500 individuals. These were Black (N=7,644), Indian (N=5,660), Pakistani (N=1,747) and Chinese (N=1,503). We further meta-analysed our European cohorts with the trans-ethnic UKBB cohorts. The analyses were performed with BOLT-LMM and METAL.

Sex-dimorphic analysis

To evaluate sex-dimorphism in our results, we meta-analysed the UKBB and the Vanderbilt cohort with the GMAMA software[67], which provides a 2 degrees of freedom (df) test of association assuming different effect sizes between the sexes. We considered a signal to show evidence of sex-dimorphism if the 2 df test *P*-value was $<5x10^{-8}$ and if the sex heterogeneity *P*-value (1 df) was <0.05.

Clumping and GCTA analysis

We performed a standard clumping analysis [PLINK 1.9 (v1.90b6.4)[68] criteria: $P \leq 5 \times 10^{-8}$, $r^2=0.01$, window-size=1Mb, 1000 Genomes Phase 3 data as linkage disequilibrium (LD) reference panel] to select a list of near-independent signals. We then performed a stepwise model selection analysis (GCTA conditional analysis) to replicate the analysis using GCTA

v1.93.0[69] with the following parameters: $P \leq 5 \times 10^{-8}$ and window-size=1Mb. We further checked for additional distinct signals by using a region-wide threshold of $P \leq 1 \times 10^{-5}$ for statistical significance.

GLP-1R pharmacological and structural analysis

Reagents

Custom peptides were purchased from Wuxi Apptec and were at least 95% pure. SNAP-Surface probes were purchased from New England Biolabs. BG-S-S-649[70] was provided by New England Biolabs on a collaborative basis. Furimazine was obtained from Promega.

Plasmids and cell line generation

Wild-type and variant GLP-1R expression plasmids, termed pcDNA5-SNAP$_f$-GLP-1R-SmBiT, were generated by Genewiz, as previously described[71], to the following design: a fast-labelling SNAP$_f$ tag and upstream signal peptide based on that of the 5-HT$_{3A}$ receptor (MDSYLLMWGLLTFIMVPGCQA), plus C-terminal SmBiT tag, were appended to the codon-optimised wild-type or variant human GLP-1R sequence (without the endogenous N-terminal signal peptide, which would lead to cleavage of the N-terminal SNAP-tag; accordingly, known missense variants in the signal peptide region were not included), and inserted into the pcDNA5/FRT/TO expression vector. These constructs allow bio-orthogonal labelling of expressed GLP-1R using SNAP-labelling probes and monitoring of cytosolic protein interactions made to GLP-1R. Constructs were used either for transient transfection or to generate stable cell lines. To obtain cell populations with inducible expression of SNAP-GLP-1R-SmBiT from a single genomic locus, Flp-In™ T-REx™ 293 cells[72] (Thermo Fisher) were co-transfected with pOG44 (Thermo Fisher) and wild-type or variant pcDNA5-SNAP$_f$-GLP-1R-

564 SmBiT in a 9:1 ratio, followed by selection with 100 µg/ml hygromycin. The resulting cell lines
565 were maintained in DMEM supplemented with 10% foetal bovine serum (FBS) and 1%
566 penicillin/streptomycin.

567

568 **Mini-G_s recruitment assay**

569 Assays were performed as previously described[71]. Where stable cell lines were used (i.e.
570 **Figures 2a and 2b**), wild-type or variant T-REx-SNAP-GLP-1R-SmBiT cells were seeded in 12-
571 well plates and transfected with 1 µg/well LgBiT-mini-G_s[25] (a gift from Prof Nevin Lambert,
572 Medical College of Georgia). The following day GLP-1R expression was induced by addition of
573 tetracycline (0.2 µg/ml) to the culture medium for 24 hours. For transient transfection assays
574 (i.e. **Figure 2j**), HEK293T cells in poly-D-lysine-coated white 96-well plates were transfected
575 using Lipofectamine 2000 with 0.05 µg/well wild-type or variant SNAP-GLP-1R-SmBiT plus
576 0.05 µg/well LgBiT-mini-G_s and the assay performed 24 hours later. Cells were then
577 resuspended in Hank's balanced salt solution (HBSS) + furimazine (Promega) diluted 1:50 and
578 seeded in 96-well half area white plates, or the same reagent added to adherent cells for
579 transient transfection assays. Baseline luminescence was measured over 5 min using a
580 Flexstation 3 plate reader at 37°C before addition of ligand or vehicle. Agonists were applied
581 at a series of concentrations spanning the response range. After agonist addition, luminescent
582 signal was serially recorded over 30 min, and ligand-induced effects were quantified by
583 subtracting individual well baseline. Signals were corrected for differences in cell number as
584 determined by BCA assay.

585

586 **High content imaging-based GLP-1R internalisation assay**

587 The assay was performed as previously described[71]. Where stable cell lines were used (i.e.
588 **Figures 2a and 2b**), wild-type or variant T-REx-SNAP-GLP-1R-SmBiT cells were seeded
589 (10,000/well) in poly-D-lysine-coated black, clear-bottom 96-well plates, in complete medium
590 supplemented with tetracycline (0.2 µg/ml) for 24 hours before the assay. Medium was
591 removed and cells labelled with 0.5 µM BG-S-S-649 (a gift from New England Biolabs) in
592 complete medium for 20 min at 37°C. Agonists were then applied in serum-free medium at
593 the indicated dose for a 30-min stimulation period to induce GLP-1R internalisation. A series
594 of concentrations spanning the response range were used. Cells were then washed with HBSS,
595 followed by a 5-min treatment ± 100 mM sodium 2-mercaptoethanesulfonate (Mesna) in
596 alkaline TNE buffer (pH 8.6) to cleave residual surface BG-S-S-649 without affecting that
597 internalised whilst bound to SNAP-GLP-1R. After re-washing, the plate was imaged using a
598 0.75 numerical aperture 20x phase contrast objective, with 9 fields-of-view (FOVs) per well
599 acquired for both transmitted phase contrast and epifluorescence. Flat-field correction of
600 epifluorescence images was performed using BaSiC[73] and cell segmentation was performed
601 using PHANTAST[74] for the phase contrast image. To determine specific GLP-1R labelling, cell-
602 free background per image was determined from the segmented epifluorescence image and
603 subtracted from the mean fluorescence intensity from the cell-containing regions. Ligand
604 induced effects were determined by subtracting the signal from vehicle-treated cells exposed
605 to Mesna. Responses were normalised to signal from labelled, untreated cells (i.e. total
606 surface labelling) within the same assay. GLP-1R surface expression levels were also obtained
607 from these assays from wells not treated with GLP-1RA or Mesna. For transient transfection
608 assays (i.e. **Figure 2j**), the assay was performed similarly but with the following changes: 1)
609 HEK293T cells in poly-D-lysine-coated black clear-bottom 96-well plates were transfected
610 using Lipofectamine 2000 with 0.1 µg/well wild-type or variant SNAP-GLP-1R-SmBiT and the

assay performed 24 hours later; 2) the plate was imaged as above both prior to and after ligand treatment (+subsequent Mesna cleavage); 3) surface labelling quantification was obtained from the pre-treatment read, and total internalised receptor was obtained from the post-treatment read.

Analysis of pharmacological data

Technical replicates within the same assay were averaged to give one biological replicate. For concentration-response assays (**Figures 2a and 2b**), ligand-induced responses were analysed by 3-parameter fitting in Prism 8.0 (GraphPad Software). As a composite measure of agonism[75], log_{10}-transformed E_{max}/EC_{50} values were obtained for each ligand/variant response. The wild-type response was subtracted from the variant response to give $\Delta log(max/EC_{50})$, a measure of gain- or loss-of-function for the variant relative to wild-type. Log_{10}-transformed surface expression levels were obtained for each variant relative to wild-type; these were then used to correct mini-G_s $\Delta log(max/EC_{50})$ values for differences in variant GLP-1R surface expression levels, by subtraction with error propagation. GLP-1R internalisation responses were already normalised to surface expression within each assay. Statistical significance between wild-type and variant responses was inferred if the 95% confidence intervals for $\Delta log(max/EC_{50})$ did not cross zero[75]. Changes to the profile of receptor response between mini-G_s recruitment and GLP-1R internalisation were inferred if $p<0.05$ with unpaired t-test analysis, with Holm-Sidak correction for multiple comparisons. For transient transfection assays (**Figure 2j**), responses were normalised to wild-type response and log_{10} transformed to give Log Δ response. Additionally, the impact of differences in surface expression on functional responses was determined by subtracting log-transformed normalised expression level from log-transformed normalised response.

635

636 **Variance explained in RG effects by mini-Gs recruitment at coding *GLP1R* variants**

637 RG (AST20) effects estimated in the UKBB study at 18 independent (r^2<0.02) coding *GLP1R*
638 variants (**Supplementary Table 10**) were regressed on mini-Gs coupling in response to GLP-1
639 stimulation (corrected for surface expression) giving more weight to variants with higher
640 minor allele frequency. Adjusted R^2 is reported as variance explained in RG effects by mini-
641 G_s coupling.

642

643 **Computational methods including molecular dynamics simulations**

644 The active state structure of GLP-1R in complex with OXM[27] and Gs protein was modelled as
645 previously described[30] and used to simulate the WT GLP-1R and G168S, A316T and R421W.
646 The systems were prepared for molecular dynamics (MD) simulations and equilibrated as
647 reported in[30]. AceMD3[76] was employed for production runs (four MD replicas of 500 ns each).
648 AquaMMapS analysis[77] was performed as previously described[30].

649

650 **Credible set analysis**

651 After selecting the signals with each region based on different M-A results from AS20, AST20
652 and AS20+AST20 models, we further performed a credible set analysis to obtain a list of
653 potential causal variants for each of the 143 selected signals. Based on the method adopted
654 from[78] under the assumption that there is one causal variant within each region, we created
655 99% credible sets. We also calculated credible sets for the trans-ethnic meta-analysis and
656 compared the results between the European only and trans-ethnic meta-analyses.

657

658 **DEPICT analysis**

DEPICT uses GWAS summary statistics and computes a prioritization of genes in associated loci, which are used to prioritise tissues via enrichment analysis. DEPICT v1 (rel 194) was used with default settings and RG GWAS summary statistics as input against a genetic background of SNPsnap data[79] derived from the 1000 Genomes Project Phase 3[80] in order to prioritise genes. Tissue and cell types enriched for prioritised genes were computed on normalised expression data comprised of 209 tissues and cell types from 37,427 Affymetrix U133 Plus 2.0 Array, as previously described[34]. We used 500 permutations for bias adjustment and 50 replications for false discovery rate estimation in our analysis in order to calculate empirical *P*-values and false discovery rate cutoffs for prioritised tissues.

CELLECT analysis

CELL type Expression-specific integration for Complex Traits (CELLECT)[35] v1.0.0 and Cell type EXpression-specificity (CELLEX)[35] v1.0.0 are two toolkits for genetic identification of likely etiologic cell types using GWAS summary statistics and single-cell RNA-sequencing (scRNA-seq) data. Tabula muris gene expression data[81], a scRNA-seq dataset derived from 20 organs from adult male and female mice, was pre-processed as described previously[82]. Briefly, expression values were normalised by using a scaling factor of 10k transcripts. The normalised values were transformed by taking log(x+1), followed by filtering out infrequently expressed genes, and keeping only those mouse transcripts with 1-1 mapping to human genes in Ensembl v.91. This data was supplied to CELLEX to compute a cumulative expression specificity metric (ESμ) of every gene for each Tabula muris cell type by combining four different expression specificity measures[82]. ESμ values were converted to stratified LD-score regression (S-LDSC) annotations using the 1000 Genomes Project SNPs and mapping each SNP

682 to the strongest ESμ value within 100kb. Cell types were prioritised by S-LDSC on the basis of
683 ESμ-derived annotations and GWAS summary statistics from the current RG meta-analysis.
684
685 **Genetically regulated gene expression analysis**
686 We used MetaXcan (S-PrediXcan) v0.6.10[83] to identify genes whose genetically predicted
687 gene expression levels are associated with RG in a number of tissues. The tested tissues were
688 chosen based on their involvement in glucose metabolism. Those were adipose visceral
689 omentum, adipose subcutaneous, skeletal muscle, liver, pancreas and whole blood.
690 Additionally, we tested ileum, transverse colon, sigmoid colon and adrenal gland, because
691 they were highlighted by DEPICT analysis. The models for the tissues of interest were trained
692 with GTEx Version 7 transcriptome data from European individuals[84]. The tissue
693 transcriptome models and 1000 Genomes[85] based covariance matrices of the SNPs used
694 within each model were downloaded from PredictDB Data Repository. The association
695 statistics between predicted gene expression and RG were estimated from the effects and
696 their standard errors coming from the AS20+AST20 model. Only statistically significant
697 associations after Bonferroni correction for the number of genes tested across all tissues ($P \leq$
698 8.996×10^{-7}) were included into the table. Genes, where less than 80% of the SNPs used in the
699 model were found in the GWAS summary statistics, were excluded due to low reliability of
700 association result.
701
702 **GARFIELD analysis**
703 We applied the GARFIELD tool v2[86] on the RG AS20+AST20 meta-analysis results to assess
704 enrichment of the RG-associated variants within functional and regulatory features.
705 GARFIELD integrates various types of data from a number of publicly available cell lines. Those

include genetic annotations, chromatin states, DNaseI hypersensitive sites, transcription factor binding sites, FAIRE-seq elements and histone modifications. We considered enrichment to be statistically significant if the RG GWAS *P*-value reached $P=1\times10^{-8}$ and the enrichment analysis *P*-value was $<2.5\times10^{-5}$ (Bonferroni corrected for 2040 annotations).

Genetic association with gut microbiome

We assessed the genetic overlap between RG GWAS results and those for gut microbiome. GWAS of microbiome profiles were publicly available and downloaded from the https://mibiogen.gcc.rug.nl/ [mibiogen.gcc.rug.nl]. For each of the 211 taxa, the corresponding *P*-values for the 143 RG GWAS SNPs and their proxies were extracted.

Genetic association with GLP-1 and GIP

We assessed the genetic overlap between RG GWAS results and those for glucagon-like peptide-1 (GLP-1) and gastric inhibitory polypeptide (GIP) measured at 0 and 120 minutes. We extracted the results for the 143 RG signals from the GWAS summary statistics for GLP-1 and GIP[87].

eQTL co-localization analysis

We further performed co-localization analysis using whole blood gene expression-QTL (eQTL) data provided by eQTLGen[39] and AS20+AST20 meta-analysis results. Only cis-eQTL data from eQTLGen was incorporated to reduce the computational burden. The COLOC2 Bayesian-based method[88] was used to interrogate the potential co-localization between RG GWAS signals and the genetic control of gene expression. We first extracted the RG GWAS test statistics of all the SNPs within +/-1Mb region around the 143 RG signals. Then, for each RG

730 signal, we matched the eQTLGen results with the RG results and performed COLOC2 analysis
731 evaluating the posterior probability (PP) of five hypotheses for each region: H_0, no
732 association; H_1, GWAS association only; H_2, eQTL association only; H_3, both GWAS and eQTL
733 association, but not co-localised; and H_4, both GWAS and eQTL association and co-localised.
734 Only GWAS signals with at least one nearby gene/probe reaching PP (H_4) ≥ 0.5 were reported.
735
736 **Genetic association with human blood plasma N-glycosylation**
737 We assessed the genetic association between 143 RG signals and 113 human blood plasma
738 N-glycome traits using previously published genome-wide summary association statistics[89].
739 The description of the analysed traits and details of the association analysis can be found
740 elsewhere[40]. We considered associations to be significant when *P*-
741 value<0.05/113/143=3.09e-6 (after Bonferroni correction). Association was considered as
742 suggestive when *P*-value<10^{-4}.
743
744 **Genetic correlation analysis**
745 We investigate the shared genetic component between RG and other traits, including
746 glycaemic ones, by performing genetic correlation analysis using the bivariate LD score
747 regression method (LDSC v1.0.0)[90]. To reduce multiple testing burden, only the GWAS results
748 of the UKBB model AS20 were used. We used GWAS summary statistics available in LDhub[91]
749 and the Meta-Analysis of Glucose and Insulin-related Traits Consortium (MAGIC) website
750 (https://www.magicinvestigators.org) for several traits including FG/FI[54], HOMA-B/HOMA-
751 IR[92]. In total, 228 different traits were included in the genetic correlation analysis with RG.
752 We considered *P*≤0.05 as the nominal significant level.
753

MR analysis

We applied a bidirectional two-sample MR strategy to investigate causality between RG and lung function, as well as T2D and lung function using independent genetic variants as instruments. MR can provide estimates of the effect of modifiable exposures on an outcome (e.g. disease) unaffected by classical confounding or reverse causation, whenever randomised clinical trials are not feasible. We looked for evidence for the presence of a causal effect of RG and T2D on two lung function phenotypes; FVC and FEV1 in a two-sample MR setting. Genome-wide summary statistics for the lung function phenotypes were available[93], involving cohorts from the SpiroMeta consortium and the UKBB study. T2D susceptibility variants and their effects were obtained from the largest-to-date T2D GWAS[4].

To avoid confounding due to sample overlap, lung function summary statistics used as outcome data were those estimated in the SpiroMeta consortium alone. Similarly, when testing the effect of lung function on RG, RG genetic effects used as outcome data were estimated in all cohorts except UK Biobank. There was no sample overlap between the lung function- and the T2D GWAS, thus allowing the use of T2D effects estimated in all contributing European studies. Genome-wide T2D summary statistics were available from a previous study[3] to test for the causal effect of lung function on T2D. All analyses were conducted using the R software package TwoSampleMR v0.5.4[94].

Instrument selection: Independent (established by conditional analyses for both RG and the lung function phenotypes) genome-wide significant (P<5x10^{-8}) variants were selected as genetic instruments. In total, 143 independent variants were defined for RG by the current study, 424 T2D signals were reported for Europeans by Vujkovic *et al.* and 130/162 independent signals were reported by Shrine *et al.* for FVC and FEV1, respectively. We looked for proxy variants with a minimum r^2 of 0.8 where the instrumental variant was not present

778 in the outcome data. Palindromic variants with minor allele frequency larger than 45% were
779 excluded to avoid uncertainty when harmonizing effects to the exposure-increasing allele.
780 After filtering, 136 variants were used to instrument RG and 413 variants were available as
781 T2D instruments. For FVC, 125 and 115 variants could be used as instruments in the RG and
782 T2D MR analyses, respectively. For FEV1, 157 and 140 variants served as instruments in the
783 RG and T2D MR analyses, respectively.
784 Causal effects were estimated using the inverse-variance weighted method, which combines
785 the causal estimates of individual instrumental variants (Wald ratios) in a random-effects
786 meta-analysis[95]. As a sensitivity analysis, we employed MR-Egger regression to obtain causal
787 estimates that are more robust to the inclusion of invalid instruments[96].
788
789 **PRS analysis**
790 We tested the ability of the RG genetic effects to predict RG, T2D and HbA1c. We compared
791 that to the predictive power of T2D and FG genetic instruments by computing PRS for RG, T2D
792 and FG and assessing their performance in predicting RG, T2D and HbA1c. PRS analyses
793 require base- and target data from independent populations. The base datasets in our
794 analyses were UKBB-only estimates from the present RG GWAS, meta-analysis estimates of
795 32 studies for T2D[97] and meta-analysis estimates from the MAGIC for FG[54]. We used the
796 second largest cohort, the Vanderbilt University Medical Centre (VUMC), as our target
797 dataset. PRS construction and model evaluation were done using the software PRSice
798 (v2.2.3)[98]. The PRS for an individual is the summation of the effect (trait-increasing) alleles
799 weighted by the effect size of the SNP taken from the base data. The SNPs in the base data
800 are clumped so that they are largely independent of each other and thus their effects can be
801 summed. To assess predictive power, PRS for RG, T2D and FG were regressed onto the

802 phenotypes of interest (i.e. RG, T2D and HbA1c) providing the coefficient of determination
803 (R^2) as an estimate for the correlation between the phenotype and the PRS in the VUMC
804 cohort. All models were adjusted for age, four principal components, sex and the cohort-
805 specific batch effect. Since the optimal *P*-value threshold for including SNPs in the PRS is
806 unknown a priori, PRS are calculated over a range of thresholds and regressed onto the
807 phenotype of interest, optimising prediction accordingly. The R^2 estimates for each trait were
808 derived by subtracting the R^2 from the null model (*Phenotype ~ sex + age + 4 principal*
809 *components + batch)* from the R^2 from the full model (*Phenotype ~ PRS + sex + age + 4*
810 *principal components + batch*) which contains the PRS at the best predicting P-value
811 threshold.
812
813 **Clustering of the RG signals with results for 45 other phenotypes**
814 We looked up the Z-scores (regression coefficient beta divided by the standard error) of the
815 distinct 143 RG signals in publicly available summary statistics of 45 relevant phenotypes. All
816 variant effects were aligned to the RG risk allele. HapMap2 based summary statistics were
817 imputed using SS-Imp v0.5.5[65] to minimise missingness. Missing summary statistics values
818 were imputed via mean imputation. The resulting variant-trait association matrix was scaled
819 by the square root of the study's mean sample size. We used agglomerative hierarchical
820 clustering with Ward's method to partition the variants into groups by their effects on the
821 considered outcomes. The clustering analysis was performed in R using function hclust() from
822 in-built stats package.
823
824 **Data availability**

825 GWAS summary statistics for RG analyses presented in this manuscript will be deposited on
826 https://www.magicinvestigators.org/downloads/ and will be also be available through the
827 NHGRI-EBI GWAS Catalog https://www.ebi.ac.uk/gwas/downloads/summary-statistics.

828

829 **References**

830 1. Santos, R.L. *et al.* Heritability of fasting glucose levels in a young genetically isolated
831 population. *Diabetologia* **49**, 667-72 (2006).
832 2. Almgren, P. *et al.* Heritability and familiality of type 2 diabetes and related
833 quantitative traits in the Botnia Study. *Diabetologia* **54**, 2811-9 (2011).
834 3. Scott, R.A. *et al.* An Expanded Genome-Wide Association Study of Type 2 Diabetes in
835 Europeans. *Diabetes* **66**, 2888-2902 (2017).
836 4. Vujkovic, M. *et al.* Discovery of 318 new risk loci for type 2 diabetes and related
837 vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis.
838 *Nat Genet* **52**, 680-691 (2020).
839 5. Chen, J. *et al.* The Trans-Ancestral Genomic Architecture of Glycaemic Traits. *bioRxiv*,
840 2020.07.23.217646 (2020).
841 6. Dimas, A.S. *et al.* Impact of type 2 diabetes susceptibility variants on quantitative
842 glycemic traits reveals mechanistic heterogeneity. *Diabetes* **63**, 2158-71 (2014).
843 7. Ingelsson, E. *et al.* Detailed physiologic characterization reveals diverse mechanisms
844 for novel genetic Loci regulating glucose and insulin metabolism in humans. *Diabetes*
845 **59**, 1266-75 (2010).
846 8. Scott, R.A. *et al.* Large-scale association analyses identify new loci influencing
847 glycemic traits and provide insight into the underlying biological pathways. *Nat*
848 *Genet* **44**, 991-1005 (2012).
849 9. Deng, Y.N., Xia, Z., Zhang, P., Ejaz, S. & Liang, S. Transcription Factor RREB1: from
850 Target Genes towards Biological Functions. *Int J Biol Sci* **16**, 1463-1473 (2020).
851 10. Piccand, J. *et al.* Rfx6 maintains the functional identity of adult pancreatic beta cells.
852 *Cell Rep* **9**, 2219-32 (2014).
853 11. Tsuboi, T. *et al.* Glucagon-like peptide-1 mobilizes intracellular Ca2+ and stimulates
854 mitochondrial ATP synthesis in pancreatic MIN6 beta-cells. *Biochem J* **369**, 287-99
855 (2003).
856 12. Bosma, K.J. *et al.* Pancreatic islet beta cell-specific deletion of G6pc2 reduces fasting
857 blood glucose. *J Mol Endocrinol* **64**, 235-248 (2020).
858 13. Rutter, G.A., Georgiadou, E., Martinez-Sanchez, A. & Pullen, T.J. Metabolic and
859 functional specialisations of the pancreatic beta cell: gene disallowance,
860 mitochondrial metabolism and intercellular connectivity. *Diabetologia* **63**, 1990-1998
861 (2020).
862 14. Hara, K. *et al.* Genome-wide association study identifies three novel loci for type 2
863 diabetes. *Hum Mol Genet* **23**, 239-46 (2014).
864 15. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic
865 architecture and pathophysiology of type 2 diabetes. *Nat Genet* **44**, 981-90 (2012).

16. Mahajan, A. *et al.* Identification and functional characterization of G6PC2 coding variants influencing glycemic traits define an effector transcript at the G6PC2-ABCB11 locus. *PLoS Genet* **11**, e1004876 (2015).
17. Pullen, T.J. & Rutter, G.A. Roles of lncRNAs in pancreatic beta cell identity and diabetes susceptibility. *Front Genet* **5**, 193 (2014).
18. Benonisdottir, S. *et al.* Sequence variants associating with urinary biomarkers. *Hum Mol Genet* **28**, 1199-1211 (2019).
19. Teumer, A. *et al.* Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat Commun* **10**, 4130 (2019).
20. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet* **51**, 957-972 (2019).
21. Spracklen, C.N. *et al.* Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* **582**, 240-245 (2020).
22. Wessel, J. *et al.* Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nat Commun* **6**, 5897 (2015).
23. Tomkin, G.H. Treatment of type 2 diabetes, lifestyle, GLP1 agonists and DPP4 inhibitors. *World J Diabetes* **5**, 636-50 (2014).
24. Koole, C. *et al.* Polymorphism and ligand dependent changes in human glucagon-like peptide-1 receptor (GLP-1R) function: allosteric rescue of loss of function mutation. *Mol Pharmacol* **80**, 486-97 (2011).
25. Wan, Q. *et al.* Mini G protein probes for active G protein-coupled receptors (GPCRs) in live cells. *J Biol Chem* **293**, 7466-7473 (2018).
26. Jones, B. *et al.* Targeting GLP-1 receptor trafficking to improve agonist efficacy. *Nat Commun* **9**, 1602 (2018).
27. Deganutti, G. *et al.* Dynamics of GLP-1R peptide agonist engagement are correlated with kinetics of G protein activation. *bioRxiv*, 2021.03.10.434902 (2021).
28. Venkatakrishnan, A.J. *et al.* Diverse GPCRs exhibit conserved water networks for stabilization and activation. *Proc Natl Acad Sci U S A* **116**, 3288-3293 (2019).
29. Yuan, S., Filipek, S., Palczewski, K. & Vogel, H. Activation of G-protein-coupled receptors correlates with the formation of a continuous internal water pathway. *Nat Commun* **5**, 4733 (2014).
30. Zhao, P. *et al.* Activation of the GLP-1 receptor by a non-peptidic agonist. *Nature* **577**, 432-436 (2020).
31. Karczewski, K.J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
32. Hauser, A.S. *et al.* Pharmacogenomics of GPCR Drug Targets. *Cell* **172**, 41-54 e19 (2018).
33. Sorli, C. *et al.* Efficacy and safety of once-weekly semaglutide monotherapy versus placebo in patients with type 2 diabetes (SUSTAIN 1): a double-blind, randomised, placebo-controlled, parallel-group, multinational, multicentre phase 3a trial. *Lancet Diabetes Endocrinol* **5**, 251-260 (2017).
34. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
35. Timshel, P.N., Thompson, J.J. & Pers, T.H. Genetic mapping of etiologic brain cell types for obesity. *Elife* **9**(2020).
36. Ding, Q. *et al.* Genome-wide meta-analysis associates GPSM1 with type 2 diabetes, a plausible gene involved in skeletal muscle function. *J Hum Genet* **65**, 411-420 (2020).

37. Kurilshikov, A. *et al.* Genetics of human gut microbiome composition. *bioRxiv*, 2020.06.26.173724 (2020).
38. Lopera-Maya, E.A. *et al.* Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. *bioRxiv*, 2020.12.09.417642 (2020).
39. Võsa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv*, 447367 (2018).
40. Sharapov, S.Z. *et al.* Defining the genetic control of human blood plasma N-glycome using genome-wide association study. *Hum Mol Genet* **28**, 2062-2077 (2019).
41. Clerc, F. *et al.* Human plasma protein N-glycosylation. *Glycoconj J* **33**, 309-43 (2016).
42. Novokmet, M. *et al.* Changes in IgG and total plasma protein glycomes in acute systemic inflammation. *Sci Rep* **4**, 4347 (2014).
43. Schmidt, M.I. *et al.* Markers of inflammation and prediction of diabetes mellitus in adults (Atherosclerosis Risk in Communities study): a cohort study. *Lancet* **353**, 1649-52 (1999).
44. Dotz, V. *et al.* Plasma protein N-glycan signatures of type 2 diabetes. *Biochim Biophys Acta Gen Subj* **1862**, 2613-2622 (2018).
45. Keser, T. *et al.* Increased plasma N-glycome complexity is associated with higher risk of type 2 diabetes. *Diabetologia* **60**, 2352-2360 (2017).
46. Wittenbecher, C. *et al.* Plasma N-Glycans as Emerging Biomarkers of Cardiometabolic Risk: A Prospective Investigation in the EPIC-Potsdam Cohort Study. *Diabetes Care* **43**, 661-668 (2020).
47. Johswich, A. *et al.* N-glycan remodeling on glucagon receptor is an effector of nutrient sensing by the hexosamine biosynthesis pathway. *J Biol Chem* **289**, 15927-41 (2014).
48. Lemmers, R.F.H. *et al.* IgG glycan patterns are associated with type 2 diabetes in independent European populations. *Biochim Biophys Acta Gen Subj* **1861**, 2240-2249 (2017).
49. Liu, D. *et al.* Ischemic stroke is associated with the pro-inflammatory potential of N-glycosylated immunoglobulin G. *J Neuroinflammation* **15**, 123 (2018).
50. Kopf, S. *et al.* Breathlessness and Restrictive Lung Disease: An Important Diabetes-Related Feature in Patients with Type 2 Diabetes. *Respiration* **96**, 29-40 (2018).
51. Sonoda, N. *et al.* A prospective study of the impact of diabetes mellitus on restrictive and obstructive lung function impairment: The Saku study. *Metabolism* **82**, 58-64 (2018).
52. Abdi, A., Jalilian, M., Sarbarzeh, P.A. & Vlaisavljevic, Z. Diabetes and COVID-19: A systematic review on the current evidences. *Diabetes Res Clin Pract* **166**, 108347 (2020).
53. Zhu, L. *et al.* Association of Blood Glucose Control and Outcomes in Patients with COVID-19 and Pre-existing Type 2 Diabetes. *Cell Metab* **31**, 1068-1077 e3 (2020).
54. Lagou, V. *et al.* Sex-dimorphic genetic effects and novel loci for fasting glucose and insulin variability. *Nat Commun* **12**, 24 (2021).
55. Marullo, L., El-Sayed Moustafa, J.S. & Prokopenko, I. Insights into the genetic susceptibility to type 2 diabetes from genome-wide association studies of glycaemic traits. *Curr Diab Rep* **14**, 551 (2014).
56. Mingrone, G. *et al.* Metabolic surgery versus conventional medical therapy in patients with type 2 diabetes: 10-year follow-up of an open-label, single-centre, randomised controlled trial. *Lancet* **397**, 293-304 (2021).

57. Whang, A., Nagpal, R. & Yadav, H. Bi-directional drug-microbiome interactions of anti-diabetics. *EBioMedicine* **39**, 591-602 (2019).
58. Voight, B.F. *et al.* The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet* **8**, e1002793 (2012).
59. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10**, 387-406 (2009).
60. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).
61. Fuchsberger, C., Abecasis, G.R. & Hinds, D.A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782-4 (2015).
62. Loh, P.R., Kichaev, G., Gazal, S., Schoech, A.P. & Price, A.L. Mixed-model association for biobank-scale datasets. *Nat Genet* **50**, 906-908 (2018).
63. Loh, P.R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-90 (2015).
64. Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
65. Rueger, S., McDaid, A. & Kutalik, Z. Evaluation and application of summary statistic imputation to discover new height-associated loci. *PLoS Genet* **14**, e1007371 (2018).
66. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
67. Magi, R. & Morris, A.P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinformatics* **11**, 288 (2010).
68. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
69. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**, 369-75, S1-3 (2012).
70. Fang, Z. *et al.* The Influence of Peptide Context on Signaling and Trafficking of Glucagon-like Peptide-1 Receptor Biased Agonists. *ACS Pharmacol Transl Sci* **3**, 345-360 (2020).
71. Fang, Z. *et al.* Ligand-Specific Factors Influencing GLP-1 Receptor Post-Endocytic Trafficking and Degradation in Pancreatic Beta Cells. *Int J Mol Sci* **21**(2020).
72. Ward, R.J., Alvarez-Curto, E. & Milligan, G. Using the Flp-In T-Rex system to regulate GPCR expression. *Methods Mol Biol* **746**, 21-37 (2011).
73. Peng, T. *et al.* A BaSiC tool for background and shading correction of optical microscopy images. *Nat Commun* **8**, 14836 (2017).
74. Jaccard, N. *et al.* Automated method for the rapid and precise estimation of adherent cell culture characteristics from phase contrast microscopy images. *Biotechnol Bioeng* **111**, 504-17 (2014).
75. Kenakin, T. A Scale of Agonism and Allosteric Modulation for Assessment of Selectivity, Bias, and Receptor Mutation. *Mol Pharmacol* **92**, 414-424 (2017).
76. Harvey, M.J., Giupponi, G. & Fabritiis, G.D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J Chem Theory Comput* **5**, 1632-9 (2009).

77. Cuzzolin, A., Deganutti, G., Salmaso, V., Sturlese, M. & Moro, S. AquaMMapS: An Alternative Tool to Monitor the Role of Water Molecules During Protein-Ligand Association. *ChemMedChem* **13**, 522-531 (2018).
78. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* **81**, 208-27 (2007).
79. Pers, T.H., Timshel, P. & Hirschhorn, J.N. SNPsnap: a Web-based tool for identification and annotation of matched SNPs. *Bioinformatics* **31**, 418-20 (2015).
80. Genomes Project, C. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).
81. Tabula Muris, C. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367-372 (2018).
82. Timshel, P.N., Thompson, J.J. & Pers, T.H. Mapping heritability of obesity by brain cell types. *bioRxiv*, 2020.01.27.920033 (2020).
83. Barbeira, A.N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* **9**, 1825 (2018).
84. Gamazon, E.R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, 1091-8 (2015).
85. Delaneau, O., Marchini, J., Genomes Project, C. & Genomes Project, C. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun* **5**, 3934 (2014).
86. Iotchkova, V. *et al.* Discovery and refinement of genetic loci associated with cardiometabolic risk using dense imputation maps. *Nat Genet* **48**, 1303-1312 (2016).
87. Almgren, P. *et al.* Genetic determinants of circulating GIP and GLP-1 concentrations. *JCI Insight* **2**(2017).
88. Dobbyn, A. *et al.* Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. *Am J Hum Genet* **102**, 1169-1184 (2018).
89. Sharapov, S. *et al.* Genome-wide association summary statistics for human blood plasma glycome. (Zenodo, 2018).
90. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
91. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279 (2017).
92. Fedko, I.O. *et al.* Genetics of fasting indices of glucose homeostasis using GWIS unravels tight relationships with inflammatory markers. *bioRxiv*, 496802 (2018).
93. Shrine, N. *et al.* New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet* **51**, 481-493 (2019).
94. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**(2018).
95. Burgess, S. *et al.* Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res* **4**, 186 (2019).
96. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* **44**, 512-25 (2015).

97. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* **50**, 1505-1513 (2018).
98. Choi, S.W. & O'Reilly, P.F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **8**(2019).

Acknowledgements

Airwave

The Airwave Health Monitoring Study was funded by the UK Home Office (780- TETRA, 2003-2018) and is currently funded by the MRC and ESRC (MR/R023484/1) with additional funding from the NIHR Imperial College Biomedical Research Centre (BRC) in collaboration with Imperial College NHS Healthcare Trust. We thank all Airwave participants for their contribution to the study.

Personal support: Paul Elliott acknowledges support from the MRC and PHE (MR/L01341X/1, 812 2014-2019) and currently from the MRC for the MRC Centre for Environment and Health 813 (MR/S019669/1). Paul Elliott and Ioanna Tzoulaki are supported by the UK Dementia Research Institute which 814 receives funding from UK DRI Ltd funded by the UK Medical Research Council, Alzheimer's 815 Society and Alzheimer's Research UK. Paul Elliott is associate director of the Health Data 816 Research UK London funded by a consortium led by the UK Medical Research Council.

BRIGHT

This work was funded by the Medical Research Council of Great Britain (grant number: G9521010D). The BRIGHT study is extremely grateful to all the patients who participated in the study and the BRIGHT nursing team. This work formed part of the research themes

contributing to the translational research portfolio for the NIHR Barts Cardiovascular Biomedical Research Centre. The funders had no role in study design, data collection and analysis.

deCODE

We thank participants in deCODE genetic studies whose contribution made this work possible.

EMIL

EMIL-Cohort, a population-based cohort; the study was approved by the ethical committee of the Chamber of Physicians Baden-Württemberg in the year 2002 (Registration Number 133-02; dates 05.09.2002 and 24.09.2002). We thank Silke Rosinger, Simone Claudi-Boehm, Rosina Sing, Sabine Schilling and Angelika Kurkhaus for technical support.

EPIC-Norfolk

The EPIC-Norfolk study (https://doi.org/10.22025/2019.10.105.00004) has received funding from the Medical Research Council (MR/N003284/1 and MC-UU_12015/1) and Cancer Research UK (C864/A14136). The genetics work in the EPIC-Norfolk study was funded by the Medical Research Council (MC_PC_13048). We are grateful to all the participants who have been part of the project and to the many members of the study teams at the University of Cambridge who have enabled this research.

FINRISK87

Support for FUSION was provided by NIH grants R01-DK062370 (to M.B.), R01-DK072193 (to K.L.M.), and intramural project number 1Z01-HG000024 (to F.S.C.). Genome-wide genotyping was conducted by the Johns Hopkins University Genetic Resources Core Facility SNP Center at the Center for Inherited Disease Research (CIDR), with support from CIDR NIH contract no. N01-HG-65403.

Framingham Heart Study

Also supported by National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) U01/UM1 DK078616 to Dr. Meigs

HUNT2/Tromson

The Nord-Trøndelag Health Study (the HUNT study) is a collaboration between HUNT Research Centre (Faculty of Medicine, Norwegian University of Science and Technology NTNU), Nord-Trøndelag County Council, Central Norway Health Authority, and the Norwegian Institute of Public Health. University of Tromsø, Norwegian Research Council (project number 185764).

InterAct

We thank all EPIC participants and staff and the InterAct Consortium members for their contributions to the study. The InterAct project received funding from the European Union (Integrated Project LSHM-CT-2006-037197 in the Framework Programme 6 of the European Community). We thank staff from the technical, field epidemiology and data teams of the Medical Research Council Epidemiology Unit in Cambridge, UK, for carrying out sample preparation, DNA provision and quality control, genotyping and data handling work.

KORA F3

The KORA research platform (KORA, Cooperative Research in the Region of Augsburg) was initiated and financed by the Helmholtz Zentrum München–German Research Center for Environment and Health, which is funded by the German Federal Ministry of Education and Research and by the state of Bavaria. Furthermore, part of this work was supported by the German National Genome Research Network (NGFN) and the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ.

Brisbane Adolescent Twin Study / SSAGA-NAG adult cohort

Genotyping and phenotyping were supported by the Australian National Health and Medical Research Council (389891, 389892, 496739), the EU 5th Framework Programme GenomEUtwin Project (QLG2-CT-2002-01254) and the U.S. National Institutes of Health (AA07535, AA13320, AA13321, AA13326, AA14041, DA12854). B.B. and G.W.M. are supported by National Health and Medical Research Council (NHMRC) Fellowship Schemes. Participants gave informed consent and the studies were approved by appropriate institutional review boards.

PPP-Botnia/Malmö Diet Cancer

Personal support: Emma Ahlqvist was funded by grants from the Swedish Research Council (2017-02688, 2020-02191).

PROCARDIS

PROCARDIS was supported by the European Community Sixth Framework Program (LSHM-CT- 2007-037273), AstraZeneca, the British Heart Foundation, the Wellcome Trust (Contract No. 075491/Z/04), the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the Swedish Heart-Lung Foundation, the Torsten and Ragnar Söderberg Foundation, the Strategic Cardiovascular and Diabetes Programs of Karolinska Institutet and Stockholm County Council, the Foundation for Strategic Research and the Stockholm County Council. Ethical permission was granted by the local ethical board for each centre.

Personal support: Anuj Goel has received support from the BHF, European Commission [LSHM-CT- 2007-037273, HEALTH-F2-2013-601456] and TriPartite Immunometabolism Consortium [TrIC]- NovoNordisk Foundation [NNF15CC0018486]. Hugh Watkins has received support from Wellcome Trust core awards (090532/Z/09/Z, 203141/Z/16/Z) and is member of the Oxford BHF Centre for Research Excellence (RE/13/1/30181). Rona J Strawbridge is supported by a UKRI Innovation-HDR-UK Fellowship (MR/S003061/1).

Rotterdam Study

The generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters for their help in creating the GWAS database, and Karol Estrada and Maksim V. Struchalin for their support in creation and analysis of imputed data. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and

Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

The Section of Endocrinology and Investigative Medicine is funded by grants from the MRC, BBSRC, NIHR, and is supported by the NIHR Biomedical Research Centre Funding Scheme. The views expressed are those of the author(s) and not necessarily those of the any of the funders, the NHS, the NIHR or the Department of Health.

UK Biobank

We thank UK Biobank for data availability, project number 37685.

Personal support: Marika A Kaakinen, Anna Ulrich, Zhanna Balkhiyarova and Inga Prokopenko are in part funded by the European Union's Horizon 2020 research and innovation programme LONGITOOLS, H2020-SC1-2019-874739. Marika A Kaakinen is also funded by the European Foundation for the Study of Diabetes (EFSD) Albert Renold Travel Fellowship.

Vanderbilt

The samples and/or dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by numerous sources: institutional funding, private agencies, and federal grants. These include the NIH funded Shared Instrumentation Grant S10OD017985 and S10RR025141; and CTSA grants UL1TR002243, UL1TR000445, and UL1RR024975. Genomic data are also supported by investigator-led

projects that include U01HG004798, R01NS032830, RC2GM092618, P50GM115305, U01HG006378, U19HL065962, R01HD074711; and additional funding sources listed at https://victr.vumc.org/biovu-funding/".

Personal support: NIH Grants R01 HL146588 and R01 HL146588-01S1.

Individual acknowledgements

Alejandra Tomas was supported by MRC project grant MR/R010676/1 and grants from Diabetes UK and the European Federation for the Study of Diabetes.

Alessia David is funded by Wellcome Trust grant 104955/Z/14/Z.

Ben Jones is supported by an Imperial Post-CCT Post-Doctoral Fellowship, MRC project grant MR/R010676/1, and grants from the European Federation for the Study of Diabetes, Academy of Medical Sciences, Imperial NIHR Biomedical Research Centre, Engineering and Physical Sciences Research Council, Society for Endocrinology and British Society for Neuroendocrinology.

Christopher Reynolds is a Royal Society Industry Fellow.

Inês Barroso was funded by an "Expanding excellence in England" award from Research England.

Inga Prokopenko is funded by the World Cancer Research Fund (WCRF UK) and World Cancer Research Fund International (2017/1641), the Wellcome Trust (WT205915), the European Union's Horizon 2020 research and innovation programme (DYNAhealth, project number 633595) the LONGITOOLS, H2020-SC1-2019-874739 and the Royal Society (IEC\R2\181075).

Mark McCarthy was a Wellcome Investigator and an NIHR Senior Investigator (Niddk. U01-DK105535, Wellcome: 090532, 098381, 106130, 203141, 212259).

Patrick M Sexton and Denise Wootten: The work was supported by NHMRC Ideas #1184726 and NHMRC program grant #1150083. Patrick M Sexton is a NHMRC Senior Principal Research Fellow (#1154434) and Denise Wootten is an NHMRC Senior Research Fellow (#1155302).

Vasiliki Lagou was supported by a fellowship from the Research Foundation-Flanders (FWO).

Victoria Salem is the recipient of a Diabetes UK Harry Keen Clinician Scientist Fellowship.

Sharapov Sodbo was supported by a grant from the Russian Science Foundation (RSF) No. 19-15-00115.

Tricia M Tan acknowledges support from the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC) Funding Scheme.

Author contributions

First author: V.L., L.J., A.U., **Central analysis and writing group**: V.L., L.J., A.U., L.Z., K.G., Z.B., A.F., L.M., **Additional analyses junior lead**: S.C., P.T., S.S., A.David, R.M., R.R., E.Ahlqvist, T.M.T., A.T., V.Salem, **GWAS cohort analyst**: G.T., H.G., E.E., B.B., R.S., A.I., J.Z., S.M.W., T.J., C.G., H.G., C.M., M.M., R.J.S., A.G., D.R., J.D., Y.S.A., M.A.K., **Metabochip cohort analyst**: E.Albrecht, A.U.J., H.M.S., **Cohort sample collection, genotyping, phenotyping or additional analysis**: I.R.C., F.E., V.Steinthorsdottir, A.G.U., P.B.M., M.J.B., S.J., O.H., B.T., K.H., T.W., K.L.M., **Metabochip cohort PI**: W.Kratzer, H.M., W.Koenig, B.O.B., **GWAS cohort PI**: J.T., M.B., J.C.F., A.Hamsten, H.Watkins, I.N., H.Wichmann, M.J.C., K.K., C.v.D., A.Hofman, N.J.W., C.L., J.B.W., N.G.M., G.M., I.T., P.E., U.T., K.S., E.L.B., J.B.M., **Additional analyses senior**: P.M.S., D.W., L.G., G.D., A.Demirkan, T.H.P., C.A.R., **Other senior author (analysis and writing group)**: Y.S.A., **Other senior author (writing)**: I.B., C.S., M.I.M., P.F., J.D., J.B.M., **Senior author**: M.A.K., B.J., I.P.

Competing interests

Alejandra Tomas has received grant funding from Sun Pharmaceuticals.

Ivan R Corrêa, Jr is an employee of New England Biolabs, Inc., a manufacturer and vendor of reagents for life science research.

Mark J Caulfield is Chief Scientist for Genomics England, a UK Government company. The views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. **Mark McCarthy** has served on advisory panels for Pfizer, NovoNordisk and Zoe Global, has received honoraria from Merck, Pfizer, Novo Nordisk and Eli Lilly, and research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier, and Takeda. As of June 2019, MMcC is an employee of Genentech, and a holder of Roche stock.

Patrick M Sexton receives grant funding from Laboratoires Servier.

Toby Johnson is now a GSK employee.

Yurii S Aulchenko is owner of Maatschap PolyOmica and PolyKnomics BV, private organizations providing services, research and development in the field of computational and statistical, quantitative and computational (gen)omics.

Tables

Table 1. Novel loci for glycaemic traits discovered through i) a GWAS meta-analysis of RG levels in up to 479,482 Europeans without diabetes, and ii) a trans-ethnic meta-analysis of up to 496,036 Europeans and individuals of other ancestries (Black, Indian, Pakistani, Chinese) in UKBB. Loci showing sex-dimorphic effects on glycaemic trait levels for the first time are also shown.

Signal	Nearest gene(s)	Lead variant	Chr	Position	Type	Alleles (effect/ other)	EAF	Effect (SE)	*P*-value	*P* het	N
European	*KDM4A*	rs3791033	1	44,134,077	primary	T/C	0.68	-0.0018 (0.00030)	4.5×10^{-9}	0.59	476,655
European	*FAM46C*	rs1966228	1	118,144,332	primary	A/G	0.74	0.0032 (0.00030)	5.8×10^{-22}	0.13	472,798
European, nonsyn	*EDEM3*	rs78444298	1	184,672,098	primary	A/G	0.019	0.0073 (0.0011)	1.5×10^{-11}	0.61	439,856
European	*ACVR1C*	rs146418816	2	158,432,811	primary	A/G	0.057	-0.0034 (0.00060)	5.6×10^{-8}	0.055	475,174
European	*ACVR1C*	rs2509903	2	158,514,510	secondary	T/C	0.14	0.0022 (0.00040)	6.7×10^{-8}	0.14	478,582
European	*RBMS1*	rs12692596	2	161,265,910	primary	T/C	0.37	0.0018 (0.00030)	1.3×10^{-9}	0.84	478,570
European	*G6PC2*	rs143869345	2	169,708,322	secondary	A/G	0.98	-0.0152 (0.0012)	1.7×10^{-37}	1.00	401,810
European, nonsyn	*NEUROD1*	rs8192556	2	182,542,998	primary	T/G	0.024	0.0053 (0.00090)	2.8×10^{-8}	0.50	439,856
European	*CACNA2D3*	rs34222465	3	55,123,055	primary	A/G	0.56	-0.0019 (0.00030)	5.5×10^{-10}	0.052	439,856
European	*MBNL1*	rs4679997	3	152,396,466	secondary	C/G	0.33	0.0017 (0.00030)	9.3×10^{-8}	0.28	473,926
European	*MBNL1*	rs78482374	3	152,492,522	secondary	A/T	0.037	-0.0042 (0.00080)	6.7×10^{-8}	0.22	455,510
European	*TRIM59,KPNA4*	rs56394279	3	160,171,092	primary	T/C	0.52	-0.0018 (0.00030)	1.2×10^{-9}	0.068	474,089
European	*MECOM*	rs73174306	3	169,194,244	primary	A/T	0.96	-0.0057 (0.00070)	1.4×10^{-14}	0.095	432,212
European	*LCORL*	rs75631642	4	18,049,216	secondary	T/C	0.78	-0.0017 (0.00040)	2.1×10^{-6}	0.13	466,061
European	*LCORL*	rs6840504	4	18,205,102	primary	T/C	0.45	0.0018 (0.00030)	1.2×10^{-9}	0.15	475,423
European	*ADRB2*	rs71584073	5	148,149,418	primary	T/C	0.93	0.0035 (0.00060)	3.3×10^{-10}	0.44	439,856
European	*SYNGAP1*	rs9461856	6	33,395,199	primary	A/G	0.48	-0.00030 (0.00030)	0.33	0.091	457,070
European	*ITPR3*	rs1830873	6	33,620,397	primary	C/G	0.57	0.00070 (0.00030)	0.021	0.87	452,301
European	*ARMC2,SESN1*	rs118126621	6	109,304,170	primary	A/G	0.025	0.0039 (0.0010)	5.0×10^{-5}	0.049	432,212
European	*POP7,EPO*	rs534043	7	100,312,724	primary	A/G	0.11	-0.0031 (0.00050)	1.7×10^{-11}	0.37	475,631
European	*PRKAR2B*	rs3801969	7	106,711,492	primary	T/G	0.43	0.0016 (0.00030)	2.2×10^{-8}	0.22	478,580
European	*A1CF*	rs61856594	10	52,637,925	primary	A/G	0.71	0.0022 (0.00030)	1.6×10^{-11}	0.58	473,354
European	*PRKG1*	rs4415704	10	53,561,613	primary	T/C	0.42	-0.0016 (0.00030)	5.6×10^{-8}	0.82	474,069
European	*LMO1*	rs9667977	11	8,541,291	secondary	T/C	0.46	-0.0013 (0.00030)	4.5×10^{-6}	0.71	457,903
European	*USP47*	rs34718245	11	11,863,080	primary	A/G	0.15	-0.0023 (0.00040)	4.3×10^{-8}	0.63	470,144
European	*PDE3B*	rs141521721	11	14,763,828	primary	A/C	0.023	0.0050 (0.0010)	1.8×10^{-7}	0.0059	439,856

European	*PDHX*	rs75479466	11	34,961,066	primary	A/G	0.083	0.0029 (0.00050)	$2.1x10^{-8}$	0.29	472,109
European	*OR4A5*	rs72913090	11	50,653,357	primary	A/C	0.92	0.0031 (0.00050)	$1.0x10^{-8}$	0.13	418,793
European	*TRIM48*	rs150587121	11	55,036,391	primary	T/C	0.91	0.0029 (0.00050)	$7.7x10^{-8}$	0.14	435,903
European	*OR8K3,OR8K1*	rs2170441	11	56,095,739	primary	A/G	0.076	-0.0031 (0.00060)	$1.7x10^{-8}$	0.28	422,873
European	*SOX5*	rs12581677	12	24,060,732	primary	A/G	0.91	0.0031 (0.00050)	$1.2x10^{-9}$	0.036	477,019
European	*MANSC4,KLHL42*	rs11049144	12	27,931,511	primary	A/C	0.22	-0.0021 (0.00040)	$1.5x10^{-9}$	0.010	455,032
European	*MANSC4,KLHL42*	rs10492373	12	27,959,998	primary	A/G	0.19	-0.0022 (0.00040)	$2.3x10^{-9}$	0.0095	479,267
European	*RNF6*	rs12874929	13	26,781,607	primary	A/G	0.77	-0.0027 (0.00030)	$1.1x10^{-14}$	0.97	476,730
European	*KL*	rs488166	13	33,554,352	primary	C/G	0.18	0.0042 (0.00040)	$3.0x10^{-27}$	0.064	478,334
European	*ZC3H13*	rs12429980	13	46,550,138	primary	A/C	0.30	-0.0018 (0.00030)	$7.3x10^{-9}$	0.40	474,764
European	*SPRY2*	rs1359790	13	80,717,156	primary	A/G	0.28	-0.0019 (0.00030)	$4.2x10^{-9}$	0.0010	477,640
European	*HECTD1,HEATR5A*	rs727675	14	31,733,642	primary	A/G	0.57	0.0017 (0.00030)	$7.5x10^{-9}$	0.86	477,060
European	*WARS*	rs45617834	14	101,295,801	secondary	C/G	0.97	0.0043 (0.00090)	$1.2x10^{-6}$	0.79	432,212
European	*HERC1*	rs67507374	15	64,038,340	primary	A/T	0.30	-0.0023 (0.00030)	$4.3x10^{-13}$	0.20	475,691
European	*ITFG3,RAB11FIP3*	rs111811257	16	541,818	secondary	T/C	0.040	-0.0046 (0.00070)	$5.5x10^{-10}$	0.76	432,212
European	*TAOK1,ABHD15*	rs9894551	17	27,880,124	primary	A/T	0.17	-0.0027 (0.00040)	$5.2x10^{-11}$	0.71	415,229
European	*HNF1B*	rs10908278	17	36,099,952	primary	A/T	0.52	-0.0017 (0.00030)	$5.2x10^{-9}$	0.0041	439,856
European, syn	*NMT1*	rs2239923	17	43,176,804	primary	T/C	0.29	0.0019 (0.00030)	$4.1x10^{-9}$	0.62	478,582
European, nonsyn	*WIPI1*	rs883541	17	66,449,122	primary	A/G	0.77	-0.0024 (0.00030)	$4.4x10^{-12}$	0.24	477,006
European	*SKA1,MAPK4*	rs2957989	18	48,075,733	primary	A/G	0.82	0.0021 (0.00040)	$2.2x10^{-8}$	0.72	458,445
European	*RALY*	rs6059497	20	32,446,960	primary	C/G	0.54	-0.0017 (0.00030)	$9.2x10^{-9}$	0.85	464,409
European	*HNF4A*	rs2267850	20	43,524,963	primary	T/C	0.27	-0.0019 (0.0003)	$3.8x10^{-9}$	0.92	458,445
European	*TSHZ2*	rs2255805	20	51,627,634	primary	T/C	0.57	-0.0018 (0.00030)	$5.5x10^{-10}$	0.99	457,514
European	*STX16-NPEPL1*	rs2296529	20	57,282,381	primary	T/C	0.77	0.0020 (0.0003)	$5.5x10^{-9}$	0.12	455,859
European	*STX16-NPEPL1*	rs73129529	20	57,404,701	secondary	C/G	0.11	0.0025 (0.00050)	$1.0x10^{-7}$	0.47	439,856
European	*EEF1A2,PPDPF*	rs6122466	20	62,139,177	primary	A/G	0.85	-0.0027 (0.00040)	$1.6x10^{-10}$	0.75	443,482
European	*MTMR3,HORMAD2*	rs5763882	22	30,597,426	primary	A/G	0.092	-0.0028 (0.00050)	$2.9x10^{-8}$	0.39	451,947
European	*MTMR3,HORMAD2*	rs6006399	22	30,598,516	primary	T/G	0.88	0.0025 (0.00040)	$2.4x10^{-8}$	0.79	478,119
European, UKBB only	*PEX7*	rs7756291	6	137235325	primary	T/C	0.55	-0.00080 (0.00030)	0.0084	0.64	456,157
European, UKBB only	*SLC38A4*	rs74832478	12	47193148	primary	T/G	0.07	0.0031 (0.00060)	$7.0x10^{-8}$	0.03	476,132
European, UKBB only	*INAFM2,C15orf52*	rs4143838	15	40622374	primary	T/C	0.95	-0.0036 (0.00070)	$2.3x10^{-7}$	0.12	418,793
European, UKBB only	*ADCY9,SRL*	rs2018506	16	4227922	primary	C/G	0.85	-0.0021 (0.00040)	$1.5x10^{-7}$	0.45	461,733
European, UKBB only	*ERN1*	rs57676627	17	62203128	primary	T/C	0.15	0.0022 (0.00040)	$4.0x10^{-7}$	0.03	432,212
European, UKBB only	*CELF5,NFIC*	rs55740449	19	3334232	primary	T/C	0.17	0.0020 (0.00040)	$5.1x10^{-7}$	0.79	439,856

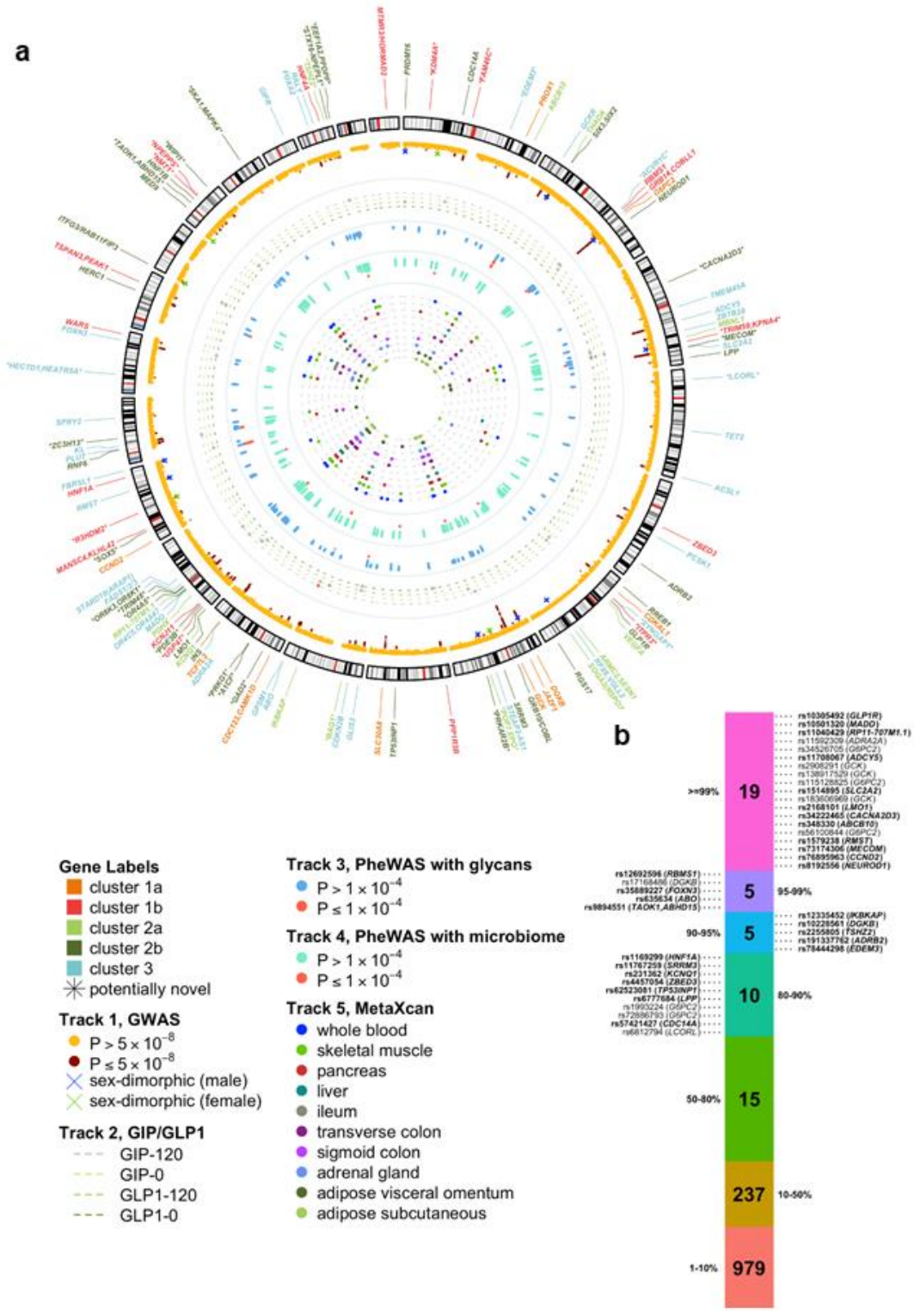
European, UKBB only	*RFX1*	rs2305780	19	14083761	primary	T/C	0.54	0.0015 (0.00030)	2.5×10^{-7}	0.24	439,856
Trans-ethnic M-A	*GATAD2B*	rs10908526	1	153,883,169	primary	C/T	0.51	0.0017 (0.00030)	1.2×10^{-8}	0.35	479,064
Trans-ethnic M-A	*RRNAD1*	rs3806415	1	156,698,265	primary	C/T	0.68	-0.0017 (0.00030)	4.8×10^{-8}	0.30	480,890
Trans-ethnic M-A	*PPP1CB,SPDYA*	rs111502507	2	29,009,180	primary	A/G	0.99	0.0064 (0.0011)	2.6×10^{-8}	0.97	439,427
Trans-ethnic M-A	*MINPP1,PAPSS2*	rs11202473	10	89,378,838	primary	G/A	0.63	-0.0017 (0.00030)	1.2×10^{-8}	0.53	490,575
Trans-ethnic M-A	*EPS8*	rs6488794	12	15,816,675	primary	A/G	0.030	0.0048 (0.00080)	1.4×10^{-8}	0.54	482,276
Trans-ethnic M-A	*SLC38A4*	rs74832478	12	47,193,148	primary	G/T	0.93	0.0033 (0.00060)	7.7×10^{-9}	0.026	492,686
Trans-ethnic M-A	*FOXN3*	rs12892260	14	89,580,986	secondary	T/C	0.94	-0.0033 (0.00060)	1.5×10^{-8}	0.23	448,766
Sex-dim: men	*PRDM16*	rs60330317	1	3,107,547	primary	G/A	0.82	0.0011 (0.00050)	0.021	0.0026	233,066
							0.82	0.0035 (0.00060)	6.1×10^{-9}		194,008
Sex-dim: women	*SGIP1*	rs7532598	1	66,998,624	primary	C/A	0.84	-0.0029 (0.00051)	1.2×10^{-8}	0.030	233,066
							0.84	-0.0012 (0.00062)	0.053		194,008
Sex-dim: men	*THADA*	rs149290349	2	43,451,957	in LD with primary	G/A	0.92	0.0035 (0.00074)	2.2×10^{-6}	3.6×10^{-4}	233,066
							0.92	0.0076 (0.00089)	1.0×10^{-17}		194,008
Sex-dim: men	*G6PC2*	rs13431652	2	169,753,415	in LD with primary	T/C	0.7	0.016 (0.00042)	1.0×10^{-1374}	5.6×10^{-4}	233,066
							0.7	0.018 (0.00050)	3.4E-286		194,008
Sex-dim: men	*TRIM59,KPNA4*	rs56394279	3	160,171,092	primary	C/T	0.49	0.0012 (0.00038)	0.0015	0.0075	233,066
							0.49	0.0028 (0.00046)	8.7×10^{-10}		194,008
Sex-dim: men	*SOGA3,RSPO3*	rs2800734	6	127,417,035	primary	G/A	0.71	0.0011 (0.00042)	0.0077	0.0032	233,066
							0.71	0.0031 (0.00051)	1.5×10^{-9}		194,008
Sex-dim: men	*DGKB,AGMO*	rs1974619	7	15,065,300	primary	C/T	0.45	-0.0037 (0.00038)	8.2×10^{-22}	1.2×10^{-5}	233,066
							0.45	-0.0063 (0.00046)	1.0×10^{-42}		194,008
Sex-dim: women	*SRRM3*	rs11773850	7	75,824,961	in LD with primary	G/A	0.98	-0.0087 (0.0013)	4.2×10^{-11}	0.0014	233,066
							0.98	-0.0021 (0.0016)	0.18		194,008
Sex-dim: men	*POP7,EPO*	rs534043	7	100,312,724	primary	A/G	0.11	-0.0019 (0.00060)	0.0014	0.0016	233,066
							0.11	-0.0048 (0.00072)	1.6×10^{-11}		194,008
Sex-dim: women	*R3HDM2*	rs7484541	12	57,714,803	in LD with primary	A/T	0.78	0.0032 (0.00046)	5.3×10^{-12}	0.030	233,066
							0.78	0.0016 (0.00056)	0.0037		194,008
Sex-dim: men	*RMST*	rs6538804	12	97,848,910	primary	C/G	0.60	0.0016 (0.00040)	7.3×10^{-5}	6.0×10^{-4}	233,066
							0.60	0.0037 (0.00048)	7.8×10^{-15}		194,008
Sex-dim: men	*FBRSL1*	rs11146926	12	133,125,450	primary	G/A	0.78	-0.0014 (0.00046)	0.0035	0.016	233,066
							0.78	-0.0031 (0.00056)	2.3×10^{-8}		194,008
Sex-dim: women	*FAM234A*	rs9929922	16	294,749	in LD with primary	A/G	0.82	0.0042 (0.00049)	4.8×10^{-18}	0.033	233,066
							0.82	0.0026 (0.00059)	9.6×10^{-6}		194,008

Sex-dim: women	*SLC43A2*	rs56405641	17	1,528,464	primary	C/T	0.91	-0.004 (0.00067)	$2.2x10^{-9}$	$2.7x10^{-4}$	233,066
							0.91	-0.00020 (0.00080)	0.81		194,008

Chr: chromosome; Pos: Position GRCh37; nonsyn: non-synonymoys; sex-dim: sex-dimorphic; EAF: allele frequency of the random glucose (RG) raising allele. A signal was annotated as "European" if it had reached genome-wide significance ($P<5x10^{-8}$) in the meta-analysis of European cohorts in either of our two models of interest with adjustment for age, sex with or without time since last meal (where available) along with exclusion of extreme hyperglycaemia (RG>20 mmol/L) or in their combination. A signal was annotated as "European, UKBB only" if it had reached genome-wide significance ($P<5x10^{-8}$) in UKBB in any of the six RG models (Methods). The EAF and *P*-values reported here are from the combined RG model. Heterogeneity among studies was assessed using the I^2 index. The Cochran's Q-test (for sex heterogeneity representing the differences in allelic effects between sexes) *P*-value is also shown. Sex-dimorphic effects and *P*-values are presented first for women.

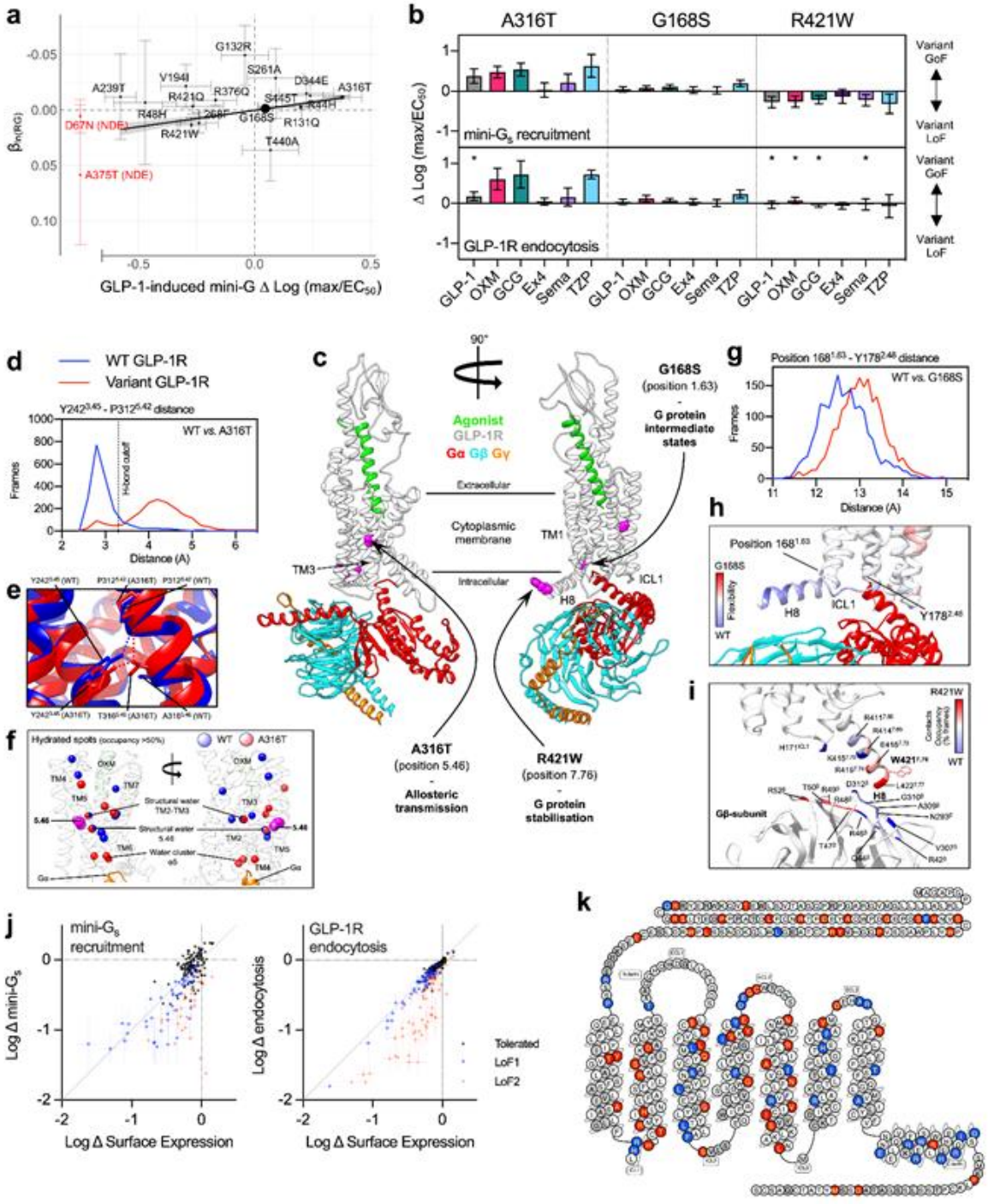
Figures

Figure 1. Summary of all RG loci identified in this study.



(**a**) **Circular Manhattan plot summarising findings from the present study. Outermost layer:** Gene names of the 162 distinct RG signals are labelled with different colours indicating three clusters defined in cluster analysis: 1a/b=metabolic syndrome, 2a/b=insulin release *versus* insulin action (with additional effects on inflammatory bowel disease for cluster 2a), 3=defects of insulin secretion (**Methods**). Asterisks annotate novel for glycaemic traits RG signals. **Track 1:** RG Manhattan plot reporting -log10(*P*-value) for RG-GWAS meta-analysis, signals reaching genome-wide significance (*P*-value<5×10^{-8}) are coloured in red. Crosses annotate genome-wide significant loci that show evidence of sex heterogeneity (**Methods**): blue crosses indicate signals with larger effects in men, green crosses – signals with larger effects in women. **Track 2:** Effects of RG genome-wide significant on four GIP/GLP-1-related traits GWAS. The colours of the dotted lines indicate four GIP/GLP-1-related traits, grey dot - signals reaching *P*-value<0.01 for a GIP/GLP-1-related trait, red dot – lead SNP has significant effect on GIP/GLP-1-related trait (Bonferroni-corrected *P*-value<1×10^{-4}). **Track 3:** Effects [-log10(*P*-value)] of lead RG variants in 113 glycan PheWAS. Blue dots - RG lead SNPs, red dots - lead SNPs reaching *P*-value<10^{-4}. **Track 4:** Effects [-log10(*P*-value)] of lead RG variants in 204 gut-microbiome PheWAS. Light green dots - RG lead SNPs, red dots - variants with significant effects at *P*-value<10^{-4}. **Track 5:** MetaXcan results for 10 selected tissues for RG GWAS meta-analysis (**Methods**), signals colocalising with genes (*P*-value<5×10^{-6}) are plotted for each tissue. (**b**) **Credible set analysis of RG associations in the European meta-analysis.** Variants from each of the RG signal credible sets are grouped based on their posterior probability (the percentiles labelled on the sides of the bar). SNP variants with posterior probability >80%, along with their locus names are provided. All variants from the credible set of the primary signals are highlighted in bold.

1324 **Figure 2. Functional and structural analysis of coding *GLP1R* variants.**



1325

1326 (**a**) Weighted regression of AST20 β_{RG} estimated in the UKBB study on *GLP1R* variant mini-G_s

1327 response to GLP-1 stimulation, with correction for variant surface expression, *n*=5-13. Size of

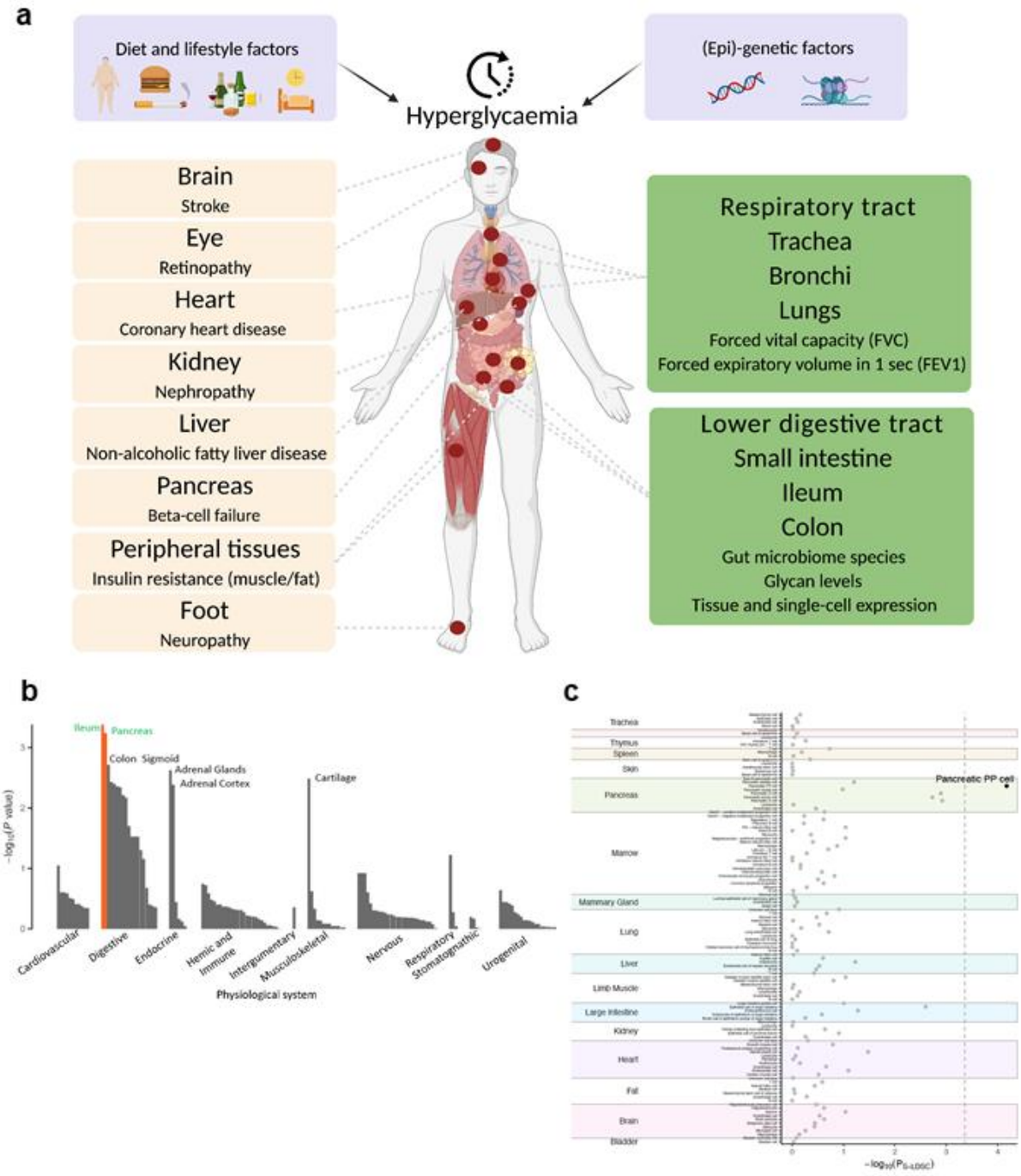
1328 dots is proportional to the weight (minor allele frequency) in the regression model (**Methods**).

1329 Error bars represent standard errors for β_{RG} and mini-G_s coupling in response to GLP-1

stimulation. The grey shaded area corresponds to the 95% confidence interval of the slope of the regression analysis (β=-0.027, 95%CI[-0.036 –{-0.016}], *P*-value=0.0001), which explained 65% of the variance in these associations. Variants in red showed no detectable surface expression (NDE) and are not included in regression analysis. (**b**) *GLP1R* variant mini-G_s coupling and receptor endocytosis, with surface expression correction, in response to GLP-1, oxyntomodulin (OXM), glucagon (GCG), exendin-4 (Ex4), semaglutide (Sema) and tirzepatide (TZP), *n*=6. Positive deviation indicates variant gain-of-function, with statistical significance inferred when the 95% confidence intervals shown do not cross zero. Responses are also compared between pathways by unpaired t-test, with * indicating statistically significant differences. (**c**) Architecture of the complex formed between the agonist-bound GLP-1R and Gs; the likely effect triggered by residues involved in GLP-1R isoforms A316T, G168S, and R421W (in magenta) are reported. (**d**) Distributions of the distance between $Y242^{3.45}$ side chain and $P312^{5.42}$ backbone computed during MD simulations of GLP-1R WT and A316T; the cut-off distance for hydrogen bond is shown. (**e**) Difference in the hydrogen bond network between GLP1-R WT and A316T. (**f**) Analysis of water molecules within the TMD of GLP1-R WT and A316T suggests minor changes in the local hydration of position 5.46 (unperturbed structural water molecule). (**g**) Distributions of the distance between position $168^{1.63}$ and $Y178^{2.48}$ during molecular dynamics simulations of GLP-1R WT and G168S. (**h**) During MD simulations the GLP-1R isoform S168G showed increased flexibility of ICL1 and H8 compared to WT, suggesting a different influence on G protein intermediate states. (**i**) Contact differences between Gs and GLP-1R WT or W421R; the C terminal of W421R H8 made more interactions with N terminal segment of Gs β subunit. (**j**) Mini-G_s and GLP-1R endocytosis responses to 20 nM exendin-4, plotted against surface GLP-1R expression, from 196 missense *GLP1R* variants transiently transfected in HEK293T cells (*n*=5 repeats per assay), with data

represented as mean ± standard error after normalization to wild-type response and log_{10}-transformation. Variants are categorised as "LoF1" when the response 95% confidence interval falls below zero or "LoF2" where expression-normalised 95% confidence interval falls below zero. (**k**) GLP-1R snake plot created using gpcr.com summarizing the functional impact of missense variants; for residues with >1 variant, classification is applied as LoF2>LoF1>tolerated.

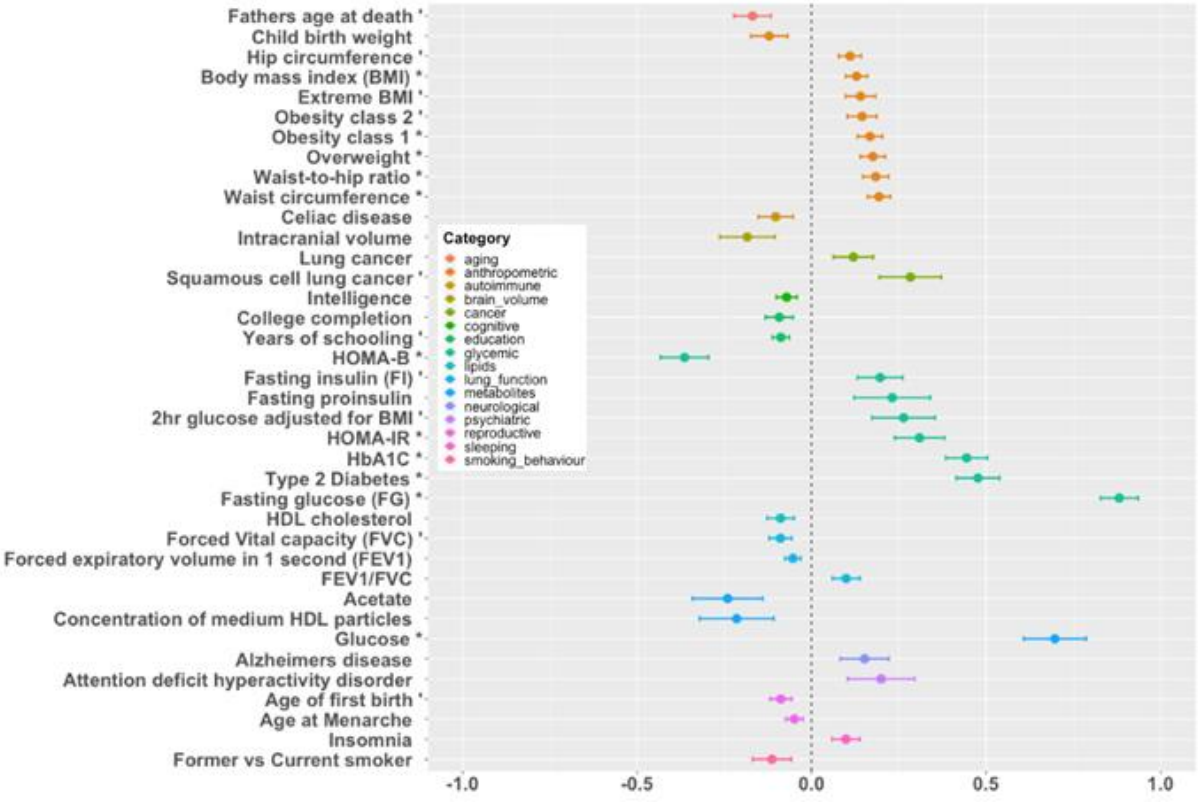
Figure 3. Deterioration of glucose homeostasis progressing into type 2 diabetes (T2D) and leading to complications in multiple organs and tissues - established (left, in peach colour) and new (right, in green).



1365

1366 (**a**) A human figure illustrating the main causes of hyperglycemia (a combination of lifestyle
1367 and genetic factors), and how hyperglycemia affects many organs and tissues. Complications
1368 on the left panel are well established for T2D. Those on the right panel are emerging ones and
1369 are supported by our current analyses. (**b**) Functional annotation of the RG GWAS results with
1370 DEPICT (**Methods**). (**c**) Functional annotation of the RG GWAS results with CELLECT
1371 (**Methods**).

1372 **Figure 4. Genome-wide genetic correlation between RG and a range of traits and diseases**



1373

1374 X axis provides the r_g genetic correlation values for traits or diseases (Y axis) reaching at

1375 least nominal significance. Correlations reaching a *P*-value<0.01 are labelled with " ' ", and

1376 those *P*-value<0.05/239 are labelled with " * ".