Dupuis, J; Langenberg, C; Prokopenko, I; Saxena, R; Soranzo, N; Jackson, AU; Wheeler, E; Glazer, NL; Bouatia-Naji, N; Gloyn, AL; Lindgren, CM; Mgi, R; Morris, AP; Randall, J; Johnson, T; Elliott, P; Rybin, D; Thorleifsson, G; Steinthorsdottir, V; Henneman, P; Grallert, H; Dehghan, A; Hottenga, JJ; Franklin, CS; Navarro, P; Song, K; Goel, A; Perry, JR; Egan, JM; Lajunen, T; Grarup, N; Spars, T; Doney, A; Voight, BF; Stringham, HM; Li, M; Kanoni, S; Shrader, P; Cavalcanti-Proena, C; Kumari, M; Qi, L; Timpson, NJ; Gieger, C; Zabena, C; Rocheleau, G; Ingelsson, E; An, P; O'Connell, J; Luan, J; Elliott, A; McCarroll, SA; Payne, F; Roccasecca, RM; Pattou, F; Sethupathy, P; Ardlie, K; Ariyurek, Y; Balkau, B; Barter, P; Beilby, JP; Ben-Shlomo, Y; Benediktsson, R; Bennett, AJ; Bergmann, S; Bochud, M; Boerwinkle, E; Bonnefond, A; Bonnycastle, LL; Borch-Johnsen, K; Bttcher, Y; Brunner, E; Bumpstead, SJ; Charpentier, G; Chen, YD; Chines, P; Clarke, R; Coin, LJ; Cooper, MN; Cornelis, M; Crawford, G; Crisponi, L; Day, IN; de Geus, EJ; Delplanque, J; Dina, C; Erdos, MR; Fedson, AC; Fischer-Rosinsky, A; Forouhi, NG; Fox, CS; Frants, R; Franzosi, MG; Galan, P; Goodarzi, MO; Graessler, J; Groves, CJ; Grundy, S; Gwilliam, R; Gyllensten, U; Hadjadj, S; Hallmans, G; Hammond, N; Han, X; Hartikainen, AL; Hassanali, N; Hayward, C; Heath, SC; Hercberg, S; Herder, C; Hicks, AA; Hillman, DR; Hingorani, AD; Hofman, A; Hui, J; Hung, J; Isomaa, B; Johnson, PR; Jrgensen, T; Jula, A; Kaakinen, M; Kaprio, J; Kesaniemi, YA; Kivimaki, M; Knight, B; Koskinen, S; Kovacs, P; Kyvik, KO; Lathrop, GM; Lawlor, DA; Le Bacquer, O; Lecoeur, C; Li, Y; Lyssenko, V; Mahley, R; Mangino, M; Manning, AK; Martnez-Larrad, MT; McAteer, JB; McCulloch, LJ; McPherson, R; Meisinger, C; Melzer, D; Meyre, D; Mitchell, BD; Morken, MA; Mukherjee, S; Naitza, S; Narisu, N; Neville, MJ; Oostra, BA; Orr, M; Pakyz, R; Palmer, CN; Paolisso, G; Pattaro, C; Pearson, D; Peden, JF; Pedersen, NL; Perola, M; Pfeiffer, AF; Pichler, I; Polasek, O; Posthuma, D; Potter, SC; Pouta, A; Province, MA; Psaty, BM; Rathmann, W; Rayner, NW; Rice, K; Ripatti, S; Rivadeneira, F; Roden, M; Rolandsson, O; Sandbaek, A; Sandhu, M; Sanna, S; Sayer, AA; Scheet, P; Scott, LJ; Seedorf, U; Sharp, SJ; Shields, B; Sigurethsson, G; Sijbrands, EJ; Silveira, A; Simpson, L; Singleton, A; Smith, NL; Sovio, U; Swift, A; Syddall, H; Syvnen, AC; Tanaka, T; Thorand, B; Tichet, J; Tnjes, A; Tuomi, T; Uitterlinden, AG; van Dijk, KW; van Hoek, M; Varma, D; Visvikis-Siest, S; Vitart, V; Vogelzangs, N; Waeber, G; Wagner, PJ; Walley, A; Walters, GB; Ward, KL; Watkins, H; Weedon, MN; Wild, SH; Willemsen, G; Witteman, JC; Yarnell, JW; Zeggini, E; Zelenika, D; Zethelius, B; Zhai, G; Zhao, JH; Zillikens, MC; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consor-

tium; Borecki, IB; Loos, RJ; Meneton, P; Magnusson, PK; Nathan, DM; Williams, GH; Hattersley, AT; Silander, K; Salomaa, V; Smith, GD; Bornstein, SR; Schwarz, P; Spranger, J; Karpe, F; Shuldiner, AR; Cooper, C; Dedoussis, GV; Serrano-Ros, M; Morris, AD; Lind, L; Palmer, LJ; Hu, FB; Franks, PW; Ebrahim, S; Marmot, M; Kao, WH; Pankow, JS; Sampson, MJ; Kuusisto, J; Laakso, M; Hansen, T; Pedersen, O; Pramstaller, PP; Wichmann, HE; Illig, T; Rudan, I; Wright, AF; Stumvoll, M; Campbell, H; Wilson, JF; Anders Hamsten on behalf of Procardis Consortium; MAGIC investigators; Bergman, RN; Buchanan, TA; Collins, FS; Mohlke, KL; Tuomilehto, J; Valle, TT; Altshuler, D; Rotter, JI; Siscovick, DS; Penninx, BW; Boomsma, DI; Deloukas, P; Spector, TD; Frayling, TM; Ferrucci, L; Kong, A; Thorsteinsdottir, U; Stefansson, K; van Duijn, CM; Aulchenko, YS; Cao, A; Scuteri, A; Schlessinger, D; Uda, M; Ruokonen, A; Jarvelin, MR; Waterworth, DM; Vollenweider, P; Peltonen, L; Mooser, V; Abecasis, GR; Wareham, NJ; Sladek, R; Froguel, P; Watanabe, RM; Meigs, JB; Groop, L; Boehnke, M; McCarthy, MI; Florez, JC; Barroso, I (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nature genetics, 42 (2). pp. 105-16. ISSN 1061-4036 DOI: 10.1038/ng.520

Downloaded from: http://researchonline.lshtm.ac.uk/4253/

DOI: 10.1038/ng.520

Usage Guidelines

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by-nc-nd/2.5/

ONLINE METHODS

Cohort description

The consortia participating in MAGIC contributed a maximum total of 122,743 individuals. The Stage 1 discovery set included 36,466-46,186 individuals (depending on trait), from 17 population-based cohort studies and four case-control studies. The Stage 2 replication set included up to 76,558 individuals from 33 sample collections, including 28 population-based and five case-control collections. Detailed information on all studies is provided in Supplementary Tables 1a (Stage 1 discovery) and 1b (Stage 2 replication). All participants were adults of white European ancestry from the United States or Europe. Individuals were excluded from the analysis if they had a physician diagnosis of diabetes, were on diabetes treatment (oral or insulin), or had a fasting plasma glucose equal to or greater than 7 mmol/L. Some individuals with FG <7 mmol/L but who would have tested abnormally after an oral glucose challenge could have been included; we estimated this number to be as low as <1% in Framingham and 1.6% in Inter99, two population cohorts in which all relevant data were available. Individual studies applied further sample exclusions, including pregnancy, non-fasting individuals, type 1 diabetes, outliers ±3 SD of distribution for either FG or FI, as detailed in Supplementary Table 1a and 1b. Individual Stage 1 discovery cohort sizes ranged between 458 and 6,479 samples; Stage 2 replication cohorts ranged between 554 and 8,010 samples. All studies were approved by local research ethic committees and all participants gave informed consent.

Type 2 diabetes association

The association analysis of lead SNPs with T2D as a dichotomous trait was carried out under the additive genetic model in 27 case-control cohorts totaling 40,655 cases and 87,022 controls of European descent. These included 8,130 cases and 38,987 controls from eight DIAGRAM+ Consortium studies and 32,525 additional T2D cases and 48,035 additional controls from 19 cohorts genotyped de novo, listed as cohort (N cases/N controls): FUSION_stage2 (1,203/1,261), METSIM_CC (854/3,469), Addition/Ely (892/1,612), Cambridgeshire Case Control Study (541/527), Norfolk Diabetes Case Control Study (6,056/6,428), deCODE (1,465/23,194), DGDG (690/730), DGI (1,022/1,075), ERGO (1,178/4,761), EUROSPAN (268/3,710), FUSION (1,161/1,174), KORA S3 (433/1,438), T2D Wellcome Trust Case Control Consortium (1,924/2,938), HPFS (1,146/1,241), Nurses' Health Study (1,532/1,754), Danish (3,652/4,992), KORA_replication consisting of cases from KORAS1-S4 and the Augsburg Diabetes Family Study (ADFS) and controls from KORA S4 (1,047/1,491), OxGN_58BC (UKRS2) (612/1,596), UKT2DGC (4,979/6,454), Framingham Heart Study_CC (674/7,664), NHANES (289/1,219), Partners/Roche (534/649), Umeå (1,327/1,424), French_CC (2,155/1,862), GCI Poland_DGI_Stage2 (969/969), GCI_US_DGI_Stage2 (1,191/1,171), and MDC_MDR_DGI_Stage2 (2,814/3,234). According to the best sample-specific model in some cohorts, age and BMI were used as covariates for adjustment of the case-control association. The meta-analysis of the cohort-specific summary statistics (odds ratios and 95% confidence intervals) was performed using a fixed effects inverse-variance approach with GWAMA (http://www.well.ox.ac.uk/gwama/index.shtml).

Quantitative trait measurements

Fasting glucose (FG, in mmol/L) was measured from fasting whole blood, plasma or serum or a combination of these. Whole blood FG levels were corrected to plasma FG using a correction factor of 1.13. Fasting insulin was measured as described in Supplementary Tables 1a and 1b for each of the cohorts. Indices of β-cell function (HOMA-B) and insulin resistance (HOMA-IR) were derived from paired fasting glucose and insulin measures using the homeostasis model assessment[18].

Genotyping, imputation and quality control

Genotyping of individual cohorts was carried out using commercial genome-wide arrays as detailed in Supplementary Tables 1a and 1b. For genome-wide SNP sets, different criteria were used to filter out poor-quality SNPs and samples prior to imputation. Criteria generally applied for exclusion of samples were: (i) call-rate <0.95, (ii) individuals with heterozygosity outside the population-specific bounds and (iii) ethnic outliers. Criteria generally applied for exclusion of SNPs were: (i) minor allele frequency (MAF) <0.01, (ii) Hardy-Weinberg equilibrium $P < 10^{-4}$ or 10^{-6} and (iii) call-rate <0.95. Imputation of additional autosomal SNPs from the HapMap CEU reference panel was performed using the software MACH[23], IMPUTE[22] or BIMBAM[68] with parameters and pre-imputation filters as specified in the Supplementary Tables 1a and 1b. SNPs were also excluded if the cohort-specific imputation quality as assessed by r2.hat<0.3 (MACH) or proper-info<0.4 (IMPUTE) or observed/expected dosage variance <0.3 (BIMBAM), or if their mapping and/or strand annotation was ambiguous. In total, up to 2.5

million genotyped or imputed autosomal SNPs were considered for meta-analysis. SNPs were considered for meta-analysis if they were available for at least 20% of maximum available sample size, or ≥10,000 individuals informative for each SNP.

Statistical analyses

We excluded from analysis people with diabetes (those on diabetes treatment or with FG ≥7 mmol/L), non-fasting participants and pregnant women. In each cohort we used log-transformed trait values for FI, HOMA-IR and HOMA-B and untransformed FG as the dependent variable in linear regression models that included terms for sex, age (except NFBC 1966, where all subjects were 31 years old), study site (if applicable), geographical covariates (if applicable) and age squared (Framingham only) to assess the association of additively coded genotypes with trait values. Association testing was performed using software that takes genotype and imputation uncertainty into account, using a missing data likelihood test implemented in SNPTEST[22] or by using allele dosages in the linear regression model in MACH2QTL[23], GenABEL[69] or lmekin from the R kinship package[70]. Regression estimates for the effect of the additively coded SNP were pooled across studies in a meta-analysis using a fixed effect inverse-variance approach[71]. The individual cohort results, but not the final meta-analysis results, were corrected for residual inflation of the test statistic using the genomic control method[72]. Final GC values were 1.05 for FG, 1.046 for HOMA-B, 1.04 for HOMA-IR and 1.041 for FI.

Replication SNP selection and analysis

Twenty-five lead SNPs from among the most significant association results in the Stage 1 discovery meta-analyses were selected for replication. In order to account for the correlation between traits and to ensure independent signals, highly significant associations detected in two or more traits were selected only once. All selected loci had an $r^2<0.5$ with the nearest other selected loci. From each unique locus, the SNP with the smallest *P* value was chosen. All SNPs had a minimum sample size of at least 80% of the overall discovery sample. Variants known to be associated with T2D (i.e. *SLC30A8*, *TCF7L2*) and reaching the genome-wide significance threshold ($P<5\times10^{-8}$) were not included in the replication list. SNPs were also selected on the basis of low heterogeneity between studies, although loci with biologic plausibility were selected even if there was some evidence of heterogeneity. Seventeen SNPs from the glucose and HOMA-B analyses and 8 SNPs from the insulin and HOMA-IR analyses were taken for Stage 2 replication. Although previously described, variants in *G6PC2*, *GCK*, *GCKR* and *MTNR1B* were selected for replication to serve as "positive controls" in all study samples. Up to 4 alternate proxy SNPs (maximizing LD with the index SNP) were selected for each locus to accommodate the capacities of different platforms. In the cases where index SNPs failed in the initial stage of genotyping, replication results were obtained for proxy SNPs in strong LD with the original index SNP whenever possible. SNPs with Hardy-Weinberg equilibrium P-values ≤0.001 were excluded. In cases where more than one proxy SNP was genotyped but the index SNP was unavailable, the proxy SNP's LD with the index SNP and its call rate was used to select the SNP with the best quality genotyping to be included in the meta-analysis.

Genotype data for 25 signals or proxies were obtained from 33 independent replication cohorts, including both *in silico* data from pre-existing GWAS (8) and *de novo* genotyping (25). Phenotype definition and association testing between fasting traits and these 25 SNPs was performed in the same manner in each cohort. The inverse variance method was then applied to derive pooled effect estimates from the Stage 2 replication samples using METAL (http://www.sph.umich.edu/csg/abecasis/Metal/index.html) and GWAMA software. We then carried out a pooled analysis of the Stage 1 discovery cohorts and Stage 2 replication samples to determine which SNPs reached genome-wide significance, as determined by a $P < 5\times10^{-8}$. Heterogeneity was assessed using the I^2 index[48].

Notes on replication genotyping

Amish

The Amish trait data is reported for the Heredity and Phenotype Interaction Heart Study (HAPI), Amish Family Longevity Study (LS), and Amish Family Diabetes Study (AFDS). All studies genotyped 15 SNPs (rs10830963, rs4607517, rs11605924, rs11708067, rs1416802, rs588262, rs4675095, rs6947696, rs4912494, rs11920090, rs174550, rs7034200, rs4243291, rs457420, rs1881413). Other SNPs were typed on different sample subsets: AFDS only (rs2191349, rs10493846); HAPI only (rs560887); HAPI, LS, and Pharmacogenomics of Anti-Platelet Interaction study (PAPI) (rs780094, rs6479526, rs340835, rs11167682); HAPI, LS, AFDS, PAPI (rs4918635, rs855228); HAPI, LS, Amish Family Calcification Study (rs11039149). The genotyping statistics in Supplementary Table 1b are reported for the AFDS + HAPI + LS cohorts.

FUSION stage 2

The FUSION stage 2 cohort includes some Health 2000 samples, none of which overlap with the Health 2000 cohort.

SNP score

For the 16 SNPs reaching genome-wide significance of association (in either the discovery stage alone or in the combined replication and discovery meta-analysis), we defined a risk score as the weighted sum of the number of expected risk alleles, where the sum of the weights was set to the number of SNPs (16) and the weights were proportional to the estimate of the effect size for each SNP. Mean fasting glucose levels according to the number of weighted risk alleles were computed in some of the largest cohorts (Framingham, ARIC, NFBC 1966) with all 16 SNPs available (genotyped or imputed).

Bioinformatic analysis and functional annotation

To perform a preliminary assessment of the underlying functionality at the associated loci, we first expanded the set of SNPs to include those in strong LD with the index SNP (defined as pairwise r^2 >0.8 according to HapMap Phase II CEU data). We then mapped the genomic locations of all the SNPs in this expanded set to several non-mutually-exclusive genomic annotation sets: non-synonymous sites, splice sites, intergenic regions, 5' UTR, 3' UTR and introns from dbSNP version 129 (http://www.ncbi.nlm.nih.gov/projects/SNP/); 1kb and 5kb regions upstream of transcription start sites from Ensembl version 49 (http://www.ensembl.org); intergenic predicted transcription factor binding sites, CpG islands,

ORegAnno elements, Encode region ancestral repeats, EvoFold elements, multi-species conserved sequences and positively selected gene regions from the University of California Santa Cruz human table browser (http://genome.ucsc.edu/cgi-bin/hgTables); predicted microRNA target sites from TargetScan 4.2 (http://www.targetscan.org); validated enhancers from the Vista Enhancer Browser (http://enhancer.lbl.gov); predicted *cis*-regulatory modules from the PreMod database (http://genomequebec.mcgill.ca/PreMod) and validated non-coding RNAs from RNAdb (http://research.imb.uq.edu.au/rnadb/). The potential functional effect of non-synonymous substitutions were evaluated using three prediction programs: SIFT ((http://blocks.fhcrc.org/sift/SIFT.html), PolyPhen (http://genetics.bwh.harvard.edu/pph/), and PANTHER (http://www.pantherdb.org/tools/csnpScoreForm.jsp).

GRAIL

We used GRAIL (Gene Relationships Across Implicated Loci, www.broad.mit.edu/mpg/grail) to examine the putative relationship between candidate genes at validated loci based on concomitant appearance in published scientific text. GRAIL is a bioinformatic annotation tool that, given several genomic regions or SNPs associated with a particular phenotype or disease, searches for similarities in the published scientific text among the associated genes (Raychaudhury et al, submitted). It scores regions for functional relatedness by defining associated regions based on the interval between recombination hotspots flanking furthest neighbouring SNPs with $r^2>0.5$ to the index SNP, and identifies overlapping genes in that region. Based on textual relationships between genes (as determined from a download of Pubmed abstracts on December 16, 2006), GRAIL assigns a *P*-value to each

region suggesting its degree of functional connectivity, and picks the best candidate gene after taking into account multiple comparisons.

We considered the following SNPs and candidate genes: rs10830963 (*MTNR1B*), rs2191349 (*DGKB*), rs4607517 (*GCK*), rs11920090 (*SLC2A2*), rs11708067 (*ADCY5*), rs560887 (*G6PC2*), rs780094 (*GCKR*), rs11605924 (*CRY2*), rs7034200 (*GLIS3*), rs340874 (*PROX1*), rs10885122 (*ADRA2A*), rs7944584 (*NR1H3*), rs174550 (*FEN1, FADS1, C11orf9, C11orf10, FADS2*), and rs11071657 (*FAM148B*). In addition, the following keywords describing functional connections were used: “glucose”, “diabetes”, “islet”, “diacylglycerol”, “circadian”, “insulin”, “drosophila”, “liver”, “clock”, “cyclase”, “pancreatic”, “adenylyl”, “memory”, “beta”, “mice”, “islets”, “phosphatase”, “camp”, “light”, “activity”. A total of 7 genes (*MTNR1B*, *DGKB*, *GCK*, *SLC2A2*, *ADCY5*, *G6PC2* and *GCKR*) out of 14 had a significant association with functional connectivity (at $P<0.1$) compared to 1.4 expected under the null, demonstrating that this gene set is enriched in relationships with each other.

eQTL analysis

The validated association signals were searched for previous evidence of expression quantitative trait loci (eQTLs) using several data sources. Liver eQTL association results were obtained from Schadt *et al*.[39]. Cortex eQTL association results were obtained from Myers *et al*.[40]. Epstein-Barr virus-transformed lymphoblastoid cell eQTLs from [41] were retrieved using the mRNA by SNP browser (http://www.sph.umich.edu/csg/liang/asthma/). For each region, we limited our analysis to *cis* eQTLs given the difficulty of reliably interpreting *trans* effects. Genes or SNPs within 1 MB from the lead SNP were considered. The r^2 values between the lead

SNPs and eQTL SNPs were retrieved from the HapMap Phase 2 data (CEU Panel), and only SNPs with r^2 >0.6 were considered.

Of the 12 SNPs showing association with liver and located at < 1 MB from the lead SNP, 5 had no r^2 data in HapMap and were located at large distances from the MAGIC lead SNP (mean 320 Kb, range 48-725 Kb). Of the remaining 7, rs174548 at the *FADS1* (fatty acid desaturase 1, $P_{eQTL} = 1.74 \times 10^{-5}$) locus was located 130 bp away from the lead SNP rs174550 and in strong LD (pairwise $r^2 = 0.8$). All the remaining SNPs did not fit our criteria for selection, although we note that a second lead SNP (rs780094 at *GCKR*) was also moderately associated (r^2=0.49, distance = 74 Kb) with a strong effect eQTL (rs4665969 at *IFT172*, $P_{eQTL} = 3.97 \times 10^{-23}$). For circulating lymphoblastoid cells, the only *cis* effect fitting our criteria was observed for the MAGIC SNP rs174550 (*FADS1*), which was located 24 Kb from a known eQTL centered on the *FADS2* gene ($P_{eQTL} = 3.1 \times 10^{-4}$). Finally, for cortex, the only eQTL was found at four SNPs within *LOC131076* (rs6769837, rs7648255, rs12636058, rs6438726), all located >870 Kb from the MAGIC lead SNP (rs11708067 at *ADCY5*, LD metrics not available).

Gene expression studies

Adult total RNA samples, except pancreatic islets and flow-sorted β cells, were purchased from Clontech (Clontech-Takara Bio Europe, Saint-Germain-en-Laye, France). Adult human islets (n=2) were available through existing collections at Oxford University and were obtained with full ethical consent. Flow-sorted β cells were obtained from two adult brain-dead donors (preparations >92% insulin-positive cells), in accordance with French legislation and the local ethical committee, as previously described[73].

Tissue panel (Oxford): Samples were treated with DNase 1 (Ambion) to ensure that residual genomic contamination was removed. For each tissue, 1 μg of total RNA was used to generate cDNA by random primed first strand synthesis (Applied Biosystems) according to the manufacturer's protocol. Reverse transcription was also performed on all samples in the absence of the enzyme reverse transcriptase, and these samples were used as negative controls. Primers were designed to cover all RefSeq transcripts. Resulting cDNA for each tissue was diluted 1:100 and 4 μl used in a 10 μl qRT-PCR reaction with 5.5 μl gene expression mastermix (Applied Biosystems) and 0.5 μl gene specific assay (Applied Biosystems). All samples were run in triplicate. A standard curve was generated by pooling 1 μl of each cDNA and serially diluting (1:50, 1:100, 1:200, 1:400, 1:800) and running as above. Expression levels were determined with respect to the mean of 4 endogenous controls (β-actin, B2M, HPRT, TOP1) and normalizing to the mean of the 1:100 standard for the assay of interest. For ease of presentation the maximum gene expression has been equalled to 1 and all other tissue expressions reported as a fraction of this.

Tissue panel (Cambridge): Adult human total RNA samples (cerebellum, cortex, spleen, pancreas, lung, kidney, liver, skeletal muscle, heart, testes, adipocyte and total brain) were obtained from Clontech. Random-primed first-strand cDNA synthesis was performed with 100 ng RNA using Super Script II (Invitrogen) according to manufacturer's instructions. Primers were design to cover the majority of protein coding transcripts. For the standard curve 200 ng of a pool of all RNA samples was amplified using the same protocol. The resulting cDNA for each tissue was diluted five-fold and 5 μl of each sample were used in a 12 μl SYBR Green PCR Master Mix (Applied Biosystem). The cDNA for the standard curve was diluted two-fold and

used as above. Primers (SIGMA) were designed to anneal to all annotated isoforms of any given gene. Quantitative PCR reactions were done in triplicate on an ABI 7900HT (Applied Biosystems). Expression levels were calculated from their average crossing points, expressed relative to the control gene *Top1* (Topoisomerase 1), and normalized to gene-specific expression in pancreas. For the purpose of presentation for each gene the maximal expression was equalled to one and the rest reported as fraction of this number. The results of these duplicate experiments, which largely confirm those reported in the text, are shown in Supplementary Figure 3.

Flow-sorted β cells (Lille): Samples were treated with DNase 1 (Ambion) to ensure that residual genomic contamination was removed. For each tissue, 1 μg of total RNA was used to generate cDNA by random primed first strand synthesis (Applied Biosystems) according to the manufacturer's protocol. Reverse transcription was also performed on all samples in the absence of the enzyme reverse transcriptase, and these samples were used as negative controls. Total RNA was extracted using Nucleospin RNA II kit (Macherey Nagel) according to the manufacturer's instructions. Resulting cDNA for each tissue was diluted 1:10 and 4 μl used in a 20 μl qRT-PCR reaction with 10 μl gene expression mastermix (Applied Biosystems) and 1 μl gene specific assay (Applied Biosystems). Data is presented as the most expressed gene (*GLIS3*) normalized to one and all other genes reported as a fraction of this number.