## ONLINE METHODS

**Stage 1 samples, genotyping and analysis.** We combined results from eight genome-wide T2D-case control association studies involving European-descent samples: DGDG, deCODE, DGI, Rotterdam, EUROSPAN, FUSION, KORAgen and WTCCC. Sample characteristics are provided in **Supplementary Table 2**. Details of the SNP genotyping platforms, genotype and sample quality-control measures and autosomal T2D association analysis approach (additive model on ln (OR) scale) are provided in the **Supplementary Note**. Imputation on the X chromosome was performed using IMPUTE (see URLs) with males coded as homozygotes. For X chromosome SNPs, ORs were calculated under the assumption of X inactivation; in other words, the per-allele OR cited is based on the between-homozygote difference. We performed autosomal and X-chromosomal association analysis on directly genotyped SNPs where available, or on imputed data using the expected allele counts or genotype probabilities.

**Stage 1 meta-analysis.** The combined stage 1 sample included 8,130 cases and 38,987 controls with an effective sample size of 22,044 (that is, a sample size with power equivalent to 11,022 cases and 11,022 controls). We excluded <1,000 SNPs where the allele frequencies were <0.40 in at least one study and >0.60 in others to minimize possible difficulties due to mislabeled alleles. A total of 2,426,886 autosomal and 60,771 chromosome X chromosome SNPs were available for ≥3 studies; of these, 2,255,857 autosomal and 52,946 chromosome X SNPs were available for ≥17,000 effective total samples. We performed genomic control correction[64] of autosomal data from each individual study (separately for directly genotyped and for imputed data) by inflating the standard error of the estimated ln (OR) so the significance of the SNP matched that of the genomic control *P* value. For those autosomal SNPs with data on ≥17,000 effective total samples, we estimated the genomic control inflation factor to be 1.074 (1.069 after removing 9,939 SNPs from regions of known T2D association). We used inverse variance-weighted meta-analysis to combine association results for stage 1 and investigated evidence for heterogeneity of ORs using Cochran's *Q* statistic and $I^2$ (ref. 65). We present autosomal meta-analysis results based on individual study genomic control correction in the main paper. The **Supplementary Note** presents association results after a second round of genomic control adjustment based on the results of the meta-analysis. Individual loci were plotted using LocusZoom (see URLs).

**Selection of stage 1 SNPs for stage 2 genotyping.** We selected the most strongly associated SNP from each region containing ≥1 SNPs with a stage 1 fixed effects meta-analysis $P \leq 1 \times 10^{-5}$ based on data from ≥3 studies. We removed SNPs within 250 kb of the index SNPs for previously identified T2D associations at *TCF7L2*, *PPARG*, *KCNJ11*, *CDKAL1*, *CDKN2A*, *IGF2BP2*, *FTO*, *HHEX*, *SLC30A8*, *JAZF1*, *THADA*, *CDC123*, *TSPAN8*, *NOTCH2*, *ADAMTS9*, *HNF1B* and *WFS1*. We did not exclude *KCNQ1*, as we had evidence of a signal independent of the association previously identified in East Asian samples[8,9]. Of the 24 SNPs selected for follow-up, all had data from ≥3 studies; 22 had data from ≥17,000 effective total samples (**Supplementary Table 3**).

**Stage 2 samples, genotyping and analysis.** We followed up the 23 most strongly associated autosomal SNPs from stage 1 in 19 stage 2 studies by performing *in silico* replication (for 3 studies), *de novo* genotyping (for 15 studies) or a combination of both (for 1 study). The stage 2 sample for autosomal signal follow-up comprised up to 34,412 cases and 59,925 controls with an effective sample size of 79,246. For the single X-chromosome signal of interest (identified some time after the autosomal signals), we performed *de novo* genotyping in 4 studies including 8,535 cases and 12,326 controls. **Supplementary Tables 1** and **2** provide sample counts and summary characteristics. Details of SNP genotyping platforms, genotype quality control, imputation methods (where appropriate) and T2D association analysis approaches are summarized in the **Supplementary Note**.

**Stage 1 and 2 meta-analysis.** For autosomal SNPs, the maximum sample size for the combination of stages 1 and 2 was 42,542 cases and 98,912 controls, yielding a total effective sample size of 101,290. For the X-chromosome SNP of interest, the equivalent numbers were 16,665 cases and 51,313 controls. We performed a fixed effects meta-analysis to combine association results for the two stages.

**Analysis of CNV-tagging SNPs.** We combined autosomal CNV-tagging SNPs from a range of sources (see **Supplementary Note**). The union of these lists provided 5,219 unique CNV-tagging SNPs for which we had GWA data for ≥17,000 stage 1 samples.

**Stage 1 conditional analysis.** To identify additional signals after accounting for the effects of known T2D loci, we repeated the GWA analysis after conditioning simultaneously on 30 established and newly discovered T2D-associated SNPs (see **Supplementary Note**). We did not include SNPs representing the X chromosome *DUSP9* locus (as it had not yet been verified as genome-wide significant when the analysis was performed), nor the *KCNQ1* locus first identified in East Asian samples[8,9] (as rs2237895—genome-wide significant in European ancestry samples[9]—was not present in our tested SNPs or in HapMap); we also did not include the recent addition of six further susceptibility loci[11–13] (as these analyses preceded those reports). Analyses were performed as described for the initial stage 1 analyses, with inclusion of study-specific covariates in addition to covariates for the 30 SNPs. These conditional analyses were performed in studies that accounted for 97% of the original stage 1 effective sample size, the exceptions being the non-Erasmus Rucphen Family (ERF) study components of EUROSPAN. To maximize the number of samples analyzed, where genotypes were missing, we replaced them at the 30 conditioned SNPs with the expected allele count based on imputation. We used a $P < 10^{-4}$ to identify putative additional signals of interest in identified T2D loci and $P < 10^{-5}$ in other regions of the genome.

**Etiological heterogeneity (BMI and age-of-diagnosis–stratified analyses).** For established and newly discovered autosomal T2D-associated variants, we performed a T2D case-control analysis within two BMI strata: BMI ≤ 30 kg/m$^2$ (4,048 T2D cases and 25,096 controls from 7 studies; see **Supplementary Note** for details) and BMI > 30 kg/m$^2$ (2,877 T2D cases, 6,764 T2D controls from 6 studies). We performed a fixed effects meta-analysis for each BMI stratum and tested for heterogeneity of the meta-analysis BMI strata-specific OR using Cochran's *Q* statistic[65]. We performed two analyses to examine the relationship between known T2D-associated variants and AOD of T2D. First, we analyzed AOD as an untransformed continuous trait using linear regression under an additive genetic model with inclusion of study-specific covariates (see **Supplementary Note** for details of samples), performing a fixed-effects meta-analysis of study-specific AOD effects. In addition, for the five studies with ≥100 cases with AOD <45 years, we meta-analyzed data comparing genotype distributions between early onset (<45 years) and later-onset (≥45 years) cases.

**Calculation of sibling relative risk.** The contribution to the sibling relative risk of a SNP with risk allele frequency $p$ and corresponding allelic odds ratio $\beta$ is

$$\lambda_{\text{sib}} = \left[1 + \frac{p(1-p)(\beta-1)^2}{2\left[(1-p)+\beta p\right]^2}\right]^2$$

assuming a multiplicative genetic model[66]. Relative risk estimates were combined as products. We estimated sibling relative risks for the 32 established and newly discovered T2D-associated SNPs based on the stage 2 ORs. We also optionally included the five T2D-associated SNPs identified by MAGIC investigators[12,13]. To capture the parent-of-origin–specific effects at *KCNQ1*, *KLF14* and the additional locus on 11p15, we used previous data[11].

**Meta-analysis results for T2D SNPs for insulin and glucose-related traits.** To establish whether the known and newly discovered T2D susceptibility loci were associated with other continuous glycemic phenotypes, we obtained meta-analysis association data (regression coefficients and *P* values) from the MAGIC analysis of 21 studies of nondiabetic European-descent individuals[12,13]. The MAGIC meta-analysis comprised ~2.5 million genotyped or imputed autosomal SNPs and included up to 46,263 individuals for fasting glucose and up to 38,413 for fasting insulin. Surrogate estimates of beta-cell function (HOMA-B) and insulin resistance (HOMA-IR) derived from fasting variables by homeostasis model assessment[67] were also analyzed.




doi:10.1038/ng.609

**BMI meta-analysis results for T2D SNPs from the GIANT Consortium.** To establish whether the T2D susceptibility loci were associated with BMI, we examined BMI association data (using *z* score, direction of association and *P* values) for these variants from the GIANT Consortium. Because inclusion of datasets ascertained for T2D case or control status could, at T2D loci, lead to distortion of the relationship with BMI[68], we restricted these analyses to eight population-based studies (including 21,233 individuals[43]).

**eQTL data.** To identify T2D-associated SNPs that might influence gene expression, we used previously described mRNA expression data (23,720 mRNA transcripts measured in 603 subcutaneous adipose tissue and 745 peripheral blood samples typed on the Illumina 300K Beadchip[47]). For SNPs and RNA transcripts within a 2-Mb window centered on each T2D index SNP, we tested for association between the log of the average expression ratio of two fluorophores and the allele count (genotype data) or expected allele count (imputed data) using linear regression, with adjustment for age, sex and, in the case of blood samples, differential cell count. For each locus and transcript, we performed additional conditional analyses by including in the regression model either the most strongly associated eQTL SNP in the region or the T2D index SNP. These conditional analyses were designed to test whether a T2D signal and any detected *cis* eQTL association at the T2D index SNP are likely to reflect the same underlying association. *P* values were adjusted for relatedness of the individuals by simulating genotypes through the corresponding Icelandic genealogy[69].

**Tissue expression data.** Adult total RNA samples were purchased from Clontech. Adult human islets ($n = 2$) were available following extraction from pancreases obtained from cadaveric donors in accordance with national transplant regulations and under ethical approval from the Oxfordshire Research Ethics Committee B. Gene expression assays were purchased from Applied Biosystems. Genes at each locus were chosen on the basis of proximity to the index SNP and biological credibility; the probe chosen for each gene was designed to cover the widest range of known transcripts. Samples were treated with DNase1 (Ambion) to ensure 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 manufacturer's protocol. Resulting cDNA for each tissue was diluted 1:100 and 4 μl was used in a 10 μl quantitative RT-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, serially diluting (1:50, 1:100, 1:200, 1:400 and 1:800) and running as above. Expression levels were determined with respect to the mean of three endogenous controls (*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 set to one and all other tissue expressions are reported as a fraction of this. In addition, flow-sorted pancreatic beta-cells were available from two adult donors (see above); preparations contained >95% insulin-positive cells. cDNA was generated from 150 ng RNA and treated as above. Resulting cDNA was diluted 1:50 and 4 μl was used in a 10 μl RT-PCR reaction. Gene expression was measured as described above. Assays that failed to demonstrate expression in human islets were excluded from this experiment. Expression levels were calculated through normalization to two endogenous controls (*HPRT* and *B2M*) and with respect to the average 1:50 standard curve dilution.

**URLs.** IMPUTE, http://mathgen.stats.ox.ac.uk/impute/impute.html/; LocusZoom, http://csg.sph.umich.edu/locuszoom/.


64. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
65. Higgins, J.P., Thompson, S.G., Deeks, J.J. & Altman, D.G. Measuring inconsistency in meta-analyses. *Br. Med. J.* **327**, 557–560 (2003).
66. Lin, S., Chakravarti, A. & Cutler, D.J. Exhaustive allelic transmission disequilibrium tests as a new approach to genome-wide association studies. *Nat. Genet.* **36**, 1181–1188 (2004).
67. Matthews, D.R. *et al.* Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28**, 412–409 (1985).
68. Stolerman, E.S. *et al.* TCF7L2 variants are associated with increased proinsulin/insulin ratios but not obesity traits in the Framingham Heart Study. *Diabetologia* **52**, 614–620 (2009).
69. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nat. Genet.* **37**, 129–137 (2005).

# Corrigendum: Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis

**Benjamin F Voight, Laura J Scott, Valgerdur Steinthorsdottir, Andrew P Morris, Christian Dina, Ryan P Welch, Eleftheria Zeggini, Cornelia Huth, Yurii S Aulchenko, Gudmar Thorleifsson, Laura J McCulloch, Teresa Ferreira, Harald Grallert, Najaf Amin, Guanming Wu, Cristen J Willer, Soumya Raychaudhuri, Steve A McCarroll, Claudia Langenberg, Oliver M Hofmann, Josée Dupuis, Lu Qi, Ayellet V Segrè, Mandy van Hoek, Pau Navarro, Kristin Ardlie, Beverley Balkau, Rafn Benediktsson, Amanda J Bennett, Roza Blagieva, Eric Boerwinkle, Lori L Bonnycastle, Kristina Bengtsson Boström, Bert Bravenboer, Suzannah Bumpstead, Noisël P Burtt, Guillaume Charpentier, Peter S Chines, Marilyn Cornelis, David J Couper, Gabe Crawford, Alex S F Doney, Katherine S Elliott, Amanda L Elliott, Michael R Erdos, Caroline S Fox, Christopher S Franklin, Martha Ganser, Christian Gieger, Niels Grarup, Todd Green, Simon Griffin, Christopher J Groves, Candace Guiducci, Samy Hadjadj, Neelam Hassanali, Christian Herder, Bo Isomaa, Anne U Jackson, Paul R V Johnson, Torben Jørgensen, Wen H L Kao, Norman Klopp, Augustine Kong, Peter Kraft, Johanna Kuusisto, Torsten Lauritzen, Man Li, Aloysius Lieverse, Cecilia M Lindgren, Valeriya Lyssenko, Michel Marre, Thomas Meitinger, Kristian Midthjell, Mario A Morken, Narisu Narisu, Peter Nilsson, Katharine R Owen, Felicity Payne, John R B Perry, Ann-Kristin Petersen, Carl Platou, Christine Proença, Inga Prokopenko, Wolfgang Rathmann, N William Rayner, Neil R Robertson, Ghislain Rocheleau, Michael Roden, Michael J Sampson, Richa Saxena, Beverley M Shields, Peter Shrader, Gunnar Sigurdsson, Thomas Sparsø, Klaus Strassburger, Heather M Stringham, Qi Sun, Amy J Swift, Barbara Thorand, Jean Tichet, Tiinamaija Tuomi, Rob M van Dam, Timon W van Haeften, Thijs van Herpt, Jana V van Vliet-Ostaptchouk, G Bragi Walters, Michael N Weedon, Cisca Wijmenga, Jacqueline Witteman, The MAGIC investigators, The GIANT Consortium, Richard N Bergman, Stephane Cauchi, Francis S Collins, Anna L Gloyn, Ulf Gyllensten, Torben Hansen, Winston A Hide, Graham A Hitman, Albert Hofman, David J Hunter, Kristian Hveem, Markku Laakso, Karen L Mohlke, Andrew D Morris, Colin N A Palmer, Peter P Pramstaller, Igor Rudan, Eric Sijbrands, Lincoln D Stein, Jaakko Tuomilehto, Andre Uitterlinden, Mark Walker, Nicholas J Wareham, Richard M Watanabe, Gonçalo R Abecasis, Bernhard O Boehm, Harry Campbell, Mark J Daly, Andrew T Hattersley, Frank B Hu, James B Meigs, James S Pankow, Oluf Pedersen, H-Erich Wichmann, Inês Barroso, Jose C Florez, Timothy M Frayling, Leif Groop, Rob Sladek, Unnur Thorsteinsdottir, James F Wilson, Thomas Illig, Philippe Froguel, Cornelia M van Duijn, Kari Stefansson, David Altshuler, Michael Boehnke & Mark I McCarthy**

***Nat. Genet.* 42, 579–589 (2010); published online 27 June 2010; corrected after print 27 August 2010**

In the version of this article initially published, there was an error in Table 1. Specifically, for rs5945326, the risk and non-risk alleles were reversed. The correct risk allele at rs5945326 is A, the non-risk allele is G and the risk allele frequency in HapMap CEU is 0.79. These errors have been corrected in the HTML and PDF versions of the article.