nature genetics

Effect of sequence variants on variance in glucose levels predicts type 2 diabetes risk and accounts for heritability

Erna V Ivarsdottir[1,2], Valgerdur Steinthorsdottir[1], Maryam S Daneshpour[3], Gudmar Thorleifsson[1], Patrick Sulem[1], Hilma Holm[1], Snaevar Sigurdsson[1], Astradur B Hreidarsson[4], Gunnar Sigurdsson[4], Ragnar Bjarnason[5,6], Arni V Thorsson[5], Rafn Benediktsson[4,6], Gudmundur Eyjolfsson[7], Olof Sigurdardottir[8], Isleifur Olafsson[9], Sirous Zeinali[10], Fereidoun Azizi[11], Unnur Thorsteinsdottir[1,6], Daniel F Gudbjartsson[1,2] & Kari Stefansson[1,6]

Sequence variants that affect mean fasting glucose levels do not necessarily affect risk for type 2 diabetes (T2D). We assessed the effects of 36 reported glucose-associated sequence variants[1] on between- and within-subject variance in fasting glucose levels in 69,142 Icelanders. The variant in *TCF7L2* that increases fasting glucose levels increases between-subject variance (5.7% per allele, $P = 4.2 \times 10^{-10}$), whereas variants in *GCK* and *G6PC2* that increase fasting glucose levels decrease between-subject variance (7.5% per allele, $P = 4.9 \times 10^{-11}$ and 7.3% per allele, $P = 7.5 \times 10^{-18}$, respectively). Variants that increase mean and between-subject variance in fasting glucose levels tend to increase T2D risk, whereas those that increase the mean but reduce variance do not ($r^2 = 0.61$). The variants that increase between-subject variance increase fasting glucose heritability estimates. Intuitively, our results show that increasing the mean and variance of glucose levels is more likely to cause pathologically high glucose levels than increase in the mean offset by a decrease in variance.

Despite recent advances in the genetics of T2D, understanding of the pathophysiology of the disease is still limited. Genome-wide association studies have yielded over 80 variants that associate with T2D, fasting glucose levels and other glycemic traits[2–6]. Although there is overlap between loci that affect fasting glucose and those that affect T2D, the effects of variants on mean fasting glucose do not predict their effects on T2D[1]. Further, none of the eight variants that associate with hemoglobin A1c (HbA1c), but not fasting glucose, associate with T2D, although HbA1c values above 6.5% are used as a diagnostic criterion for T2D[1].

Most reports on analysis of loci associated with quantitative traits have been confined to the effects of variants on the means of traits. However, variants can also affect the variability of traits (variance heterogeneity)[7]. Such loci have been reported for some human traits, including the major histocompatibility complex (MHC) region for rheumatoid arthritis[8], *FTO* for body mass index (BMI)[9], *SLC2A9* for serum urate[10], *LEPR* for C-reactive protein and *ICAM1* for soluble ICAM1 (ref. 11), as well as for traits in other species like rats[12], flies[13] and plants[14]. Further, variants can also affect the variability in measurements taken from the same individual. We refer to these two types of variability as between-subject and within-subject variance. Here we estimate the variance effects of variants that have been associated with fasting glucose levels[1] and examine how their effects on variance correlate with their effects on T2D risk. We also estimate how the effects of these variants on variance affect heritability estimates.

We chip genotyped 117,548 Icelanders with glucose measurements performed at three laboratories (**Fig. 1**, **Table 1**, **Supplementary Fig. 1** and **Supplementary Tables 1–4**). Of the subjects, 8,797 (7.5%) had T2D or were on diabetes medication[15]. Furthermore, 366 individuals had type 1 diabetes (T1D). The primary glucose variance association analysis was performed on individuals with fasting glucose levels (set I). Additionally, we generated three data sets for secondary analysis; one comprising individuals with fasting and/or non-fasting glucose levels (set II) and the previously listed data sets I and II after excluding T2D and T1D cases and individuals on diabetes medication.

Of the 36 known variants associated with glucose levels[1], 3 associated with between-subject variance consistently in all four analyses ($P < 0.05/36 = 0.0014$) (**Fig. 1a** and **Supplementary Tables 3–5**). One variant, rs7903146 in *TCF7L2*, is the strongest common T2D-associated variant[2,16]. The allele at this SNP associating with higher glucose levels and increased T2D risk was associated with greater between-subject glucose variance. In contrast, the alleles of rs560887 in *G6PC2* and rs2908289 in *GCK* that are associated with increased

[1]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [2]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. [3]Cellular and Molecular Endocrine Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. [4]Department of Endocrinology and Metabolic Medicine, Landspitali, National University Hospital of Iceland, Reykjavik, Iceland. [5]Children's Medical Center, Landspitali, National University Hospital of Iceland, Reykjavik, Iceland. [6]Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. [7]Laboratory in Mjodd, RAM, Reykjavik, Iceland. [8]Department of Clinical Biochemistry, Akureyri Hospital, Akureyri, Iceland. [9]Department of Clinical Biochemistry, Landspitali, National University Hospital of Iceland, Reykjavik, Iceland. [10]Iranian Molecular Medicine Network, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran. [11]Endocrine Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran. Correspondence should be addressed to D.F.G. (daniel.gudbjartsson@decode.is) or K.S. (kari.stefansson@decode.is).

Received 30 March; accepted 11 July; published online 7 August 2017; doi:10.1038/ng.3928

Table 1 Summary of the data

						n measurements		T2D		T1D		Age		YOB	
	n	Mean	Q1	Median	Q3	Mean	Range	*n*	%	*n*	%	Mean	s.d.	Mean	s.d.
Fasting glucose levels (set I)															
Male	28,981	5.8	5.0	5.5	6.2	3.0	1–55	3,296	11.4	76	0.3	61.6	15.2	1947.4	16.1
Female	40,161	5.4	4.8	5.2	5.7	3.0	1–94	3,059	7.6	78	0.2	59.0	17.1	1950.5	18.1
All glucose levels (set II)															
Male	51,911	6	5	5.6	6.5	8.2	1–234	4,676	9.0	185	0.4	62.9	16.2	1943.9	16.7
Female	65,637	5.6	4.8	5.3	6.0	7.8	1–280	4,121	7.3	181	0.3	60.3	18.2	1946.6	18.7
HbA1c (first measurement)															
Male	18,107	5.8	5.2	5.5	5.9	–	–	3,041	16.8	56	0.3	60.1	15.1	1948.9	15.8
Female	22,945	5.6	5.2	5.5	5.8	–	–	2,676	11.7	47	0.2	56.9	17.2	1952.1	17.7

T2D, type 2 diabetes; T1D, type 1 diabetes; YOB, year of birth; Q1, first quartile; Q3, third quartile.

glucose[1] associated with less between-subject variance. The variant in *G6PC2* does not associate significantly with T2D whereas the variant in *GCK* slightly increases T2D risk in the DIAGRAM Consortium (odds ratio (OR) = 1.04, $P = 0.018$; **Supplementary Table 2**).

We also estimated the effects of the 36 variants on the within-subject variance in glucose levels (**Fig. 1b**). The glucose-increasing alleles of three variants—rs560887 in *G6PC2*, rs6943153 in *GRB10* and rs2908289 in *GCK*—associated consistently with less within-subject variance in all four analyses (**Supplementary Tables 3–5**).

On the basis of a T2D meta-analysis (12,171 cases and 56,862 controls of European ancestry)[2], 22 of the 36 variants with an effect on mean fasting glucose levels also associate with T2D. However, their effects on fasting glucose levels and T2D risk were weakly correlated ($r^2 = 0.02$ between the effect on the mean (β) and log(OR), $P = 0.21$; an F test was performed in all regression analysis) (**Fig. 2a**). Interestingly, the effect of a variant on between-subject variance in fasting glucose combined with its effect on mean fasting glucose predicted the effect of this variant on T2D much better than the effect on the mean alone ($r^2 = 0.61$, P value for adding effect on between-subject variance = 5.7×10^{-8}). Even on its own, the effect on between-subject variance predicted the T2D effect reasonably well ($r^2 = 0.38$, $P = 3.3 \times 10^{-5}$) (**Fig. 2b**). Therefore, variants that increase both the mean and between-subject variance of glucose levels increase the risk of T2D more than variants that increase the mean but reduce the between-subject variance.

The effect on within-subject glucose variance was a worse predictor of T2D risk than the effect on between-subject variance ($r^2 = 0.24$) (**Supplementary Table 6**), and it did not improve prediction of T2D beyond the mean and between-subject effects ($P = 0.091$).

Interaction between sequence variants and environmental factors such as nutrition is a possible source of between-subject variance[11]. It has previously been reported that heterogeneity in T2D associations is introduced by BMI[17,18]. We estimated the interaction effects between the 36 glucose-associated variants and BMI on fasting glucose ($n = 39{,}986$). The interaction effects were correlated with the between-subject variance effects ($r^2 = 0.12$, $P = 0.020$) (**Supplementary Fig. 2** and **Supplementary Table 7**). These results show that the effects of variants are affected by environment, although only a small fraction of the effects on between-subject variance are mitigated through interaction with BMI.

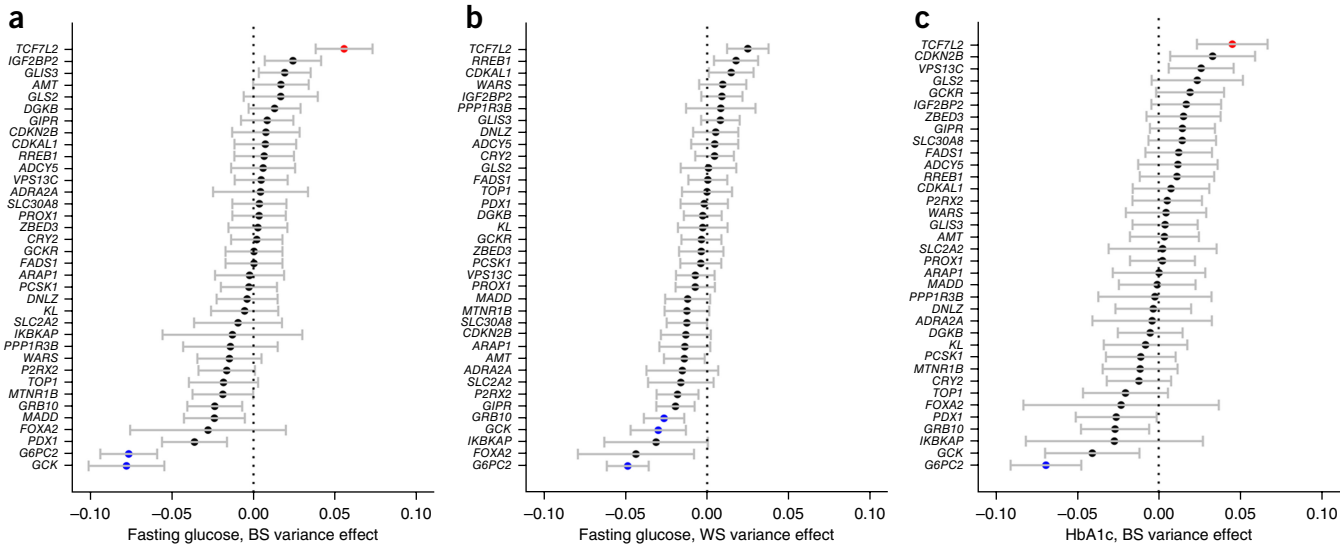

Figure 1 Effects of 36 published fasting-glucose-associated variants on between-subject and within-subject variance in fasting glucose levels and between-subject variance in HbA1c levels. Effects on variance are given for the allele that increases fasting glucose levels (**Supplementary Table 3**). Variants are colored blue if they significantly decrease the variance and red if they significantly increase it (likelihood-ratio test, $P < 0.05/36 = 0.0014$). (**a**) Effects on between-subject variance in fasting glucose ($\log(\alpha_{BS})$) and 95% confidence intervals for the estimated effects. (**b**) Effects on within-subject variance in fasting glucose levels ($\log(\alpha_{WS})$) and 95% confidence intervals for the estimated effects. (**c**) Effects on between-subject variance in HbA1c ($\log(\alpha_{BS})$) and 95% confidence intervals for the estimated effects.

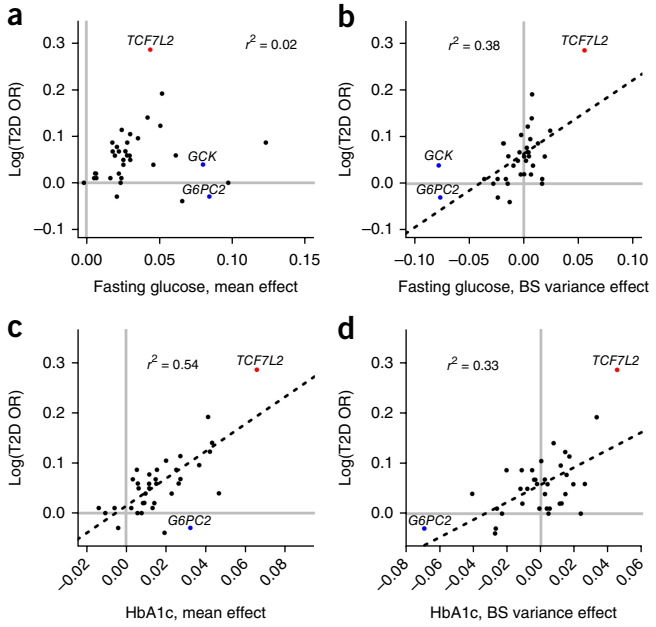

Figure 2 Effects of 36 published fasting-glucose-associated variants on fasting glucose and HbA1c, and between-subject variance in fasting glucose and HbA1c versus their effects on type 2 diabetes risk. Effects on fasting glucose were estimated in the Icelandic data, while effects on T2D risk were obtained from a T2D meta-analysis[2] (T2D-GENES Consortium, GoT2D Consortium, DIAGRAM Consortium; see URLs) (**Supplementary Tables 2**, **3** and **11**). Effects are given for the allele that increases fasting glucose levels. Variants are colored blue if they significantly decrease variance and red if they significantly increase it ($P < 0.05/36 = 0.0014$). (**a**) Fasting glucose mean effect (β) against log(T2D OR). (**b**) Fasting glucose between-subject variance effect ($\log(\alpha_{BS})$) against log(T2D OR). (**c**) HbA1c mean effect (β) against log(T2D OR). (**d**) HbA1c between-subject variance effect ($\log(\alpha_{BS})$) against log(T2D OR).

An undetected secondary variant can create a variance effect for the primary variant. However, secondary signals at the loci associated with between-subject variance had no impact on variance effects (**Supplementary Table 8**). Another possible source of effects on between-subject variance is interaction between loci. For the three variants associated with between-subject variance, we found no interaction (**Supplementary Table 9**).

To validate these variance effects, we analyzed a sample of 10,437 Iranians with 44,470 fasting glucose measurements from the prospective Tehran Lipid and Glucose Study[19]. We replicated the association of the variants in *TCF7L2*, *GCK* and *G6PC2* with between-subject variance and the association of the *G6PC2* and *GCK* variants with within-subject variance (**Supplementary Tables 5** and **10**).

HbA1c reflects the average plasma glucose concentration over 3 months, and an HbA1c value above 6.5% is used as a diagnostic criterion for T2D[15]. HbA1c measurements were available for 41,052 Icelanders with genotype information (**Table 1** and **Supplementary Table 1**). The number of measurements per subject was correlated with HbA1c. Therefore, we only used the first measurement for each subject in our analysis.

The pattern of effect for the 36 markers on between-subject HbA1c variability is consistent with the results for fasting glucose (**Fig. 1**, **Supplementary Fig. 3** and **Supplementary Table 11**). Of the 36 variants, the variants in *TCF7L2* and *G6PC2* were associated with between-subject variance (4.5% increase per allele, $P = 4.5 \times 10^{-5}$ and 6.9% decrease per allele, $P = 4.0 \times 10^{-10}$, respectively; a likelihood-ratio test was performed in all genome-wide associations). As for fasting glucose, the effect on between-subject variance in HbA1c increased the prediction accuracy of the effect on T2D ($r^2 = 0.54$ for the mean only, $r^2 = 0.77$ for the mean and between-subject variance effect, P value for adding the between-subject variance effect = 1.4×10^{-6}) (**Fig. 2c**,**d** and **Supplementary Table 6e**).

Eight variants have been reported to affect HbA1c without affecting fasting glucose, none of which have an effect on T2D[1,20]. These variants associate with red blood cell homeostasis and iron metabolism (**Supplementary Table 12**). Interestingly, the HbA1c-increasing allele for all eight markers lowered between-subject variance (**Supplementary Figs. 4** and **5**, and **Supplementary Table 13**), of which two were significantly associated with lower between-subject variance ($P < 0.05/8 = 0.0063$): rs10159477[G] in *HK1* was associated with 5.1% lower variance per allele ($P = 0.0024$) and rs6474359[T] in *ANK1* was associated with 8.0% lower variance per allele ($P = 0.0044$). The increase in the mean was offset by lower variance for carriers of these variants, and these individuals are therefore less likely to have high HbA1c measures. This may explain why carriers of these HbA1c-increasing variants are not likely to be misclassified as diabetic[20].

We constructed genetic risk scores (GRSs), based on the 36 variants, for both mean and between-subject variance of fasting glucose levels. Both GRSs were associated with T2D ($P < 3.1 \times 10^{-39}$; **Fig. 3** and **Supplementary Table 14**). Adding the GRS for between-subject variance to the GRS for the mean increased residual Nagelkerke's pseudo-r^2 from 0.4% to 1.0% ($P = 5.4 \times 10^{-67}$; **Supplementary Table 14**). Similarly, GRSs based on the 36 variants for glucose levels and the 8 variants for HbA1c measures were associated with T2D ($P < 3.4 \times 10^{-28}$; **Fig. 3** and **Supplementary Table 14**). This shows that the effects of variants on between-subject variance have an impact on genetic T2D risk prediction that is comparable to that from their effects on the mean.

The heritability of a trait is the fraction of variance attributable to genetics. Classical estimates of heritability ignore the impact of variants on phenotypic variance. Most heritability estimates are based on relating the correlation between relative pairs to the genetic sharing between relatives[21]. Correlation between relatives corresponds to the ratio of their covariance and the geometric mean of their phenotypic variances. Variants that affect variance will have a substantial impact on the denominator. However, their effect on covariance is unpredictable. In our data, we had fasting glucose measures and genotypic information for 35,965 sibling pairs and 38,527 parent–offspring pairs. To investigate the effect of variants on the covariance between relatives, we calculated the covariance for genotype-concordant relative pairs and estimated the relationship between genotype and covariance. For the 36 variants associated with glucose levels, the mean covariance trend in siblings and parent–offspring pairs correlated positively with the between-subject variance effect ($r^2 = 0.22$, $P = 2.1 \times 10^{-3}$) (**Fig. 4a** and **Supplementary Table 15**). If the increase in covariance per allele was higher than the variance effect, the correlation was also increased and the variants therefore also increased the estimated narrow-sense heritability. The variant in *TCF7L2* had the strongest trend of 17.6% increased covariance ($P = 4.1 \times 10^{-4}$) (**Fig. 4b**). The between-subject variance effect of *TCF7L2* was 5.7% per allele, and the correlation was therefore increased by 11.3% per allele.

We have shown that variants in *TCF7L2*, *GCK*, *G6PC2* and *GRB10* that affect mean fasting glucose levels also associate with variance in glucose. The variance effects remain after the removal of diabetic cases and individuals on diabetes medication. The two variants that lower between-subject variance do not associate with T2D risk, and their variance effect is thus not driven by a diabetes medication.

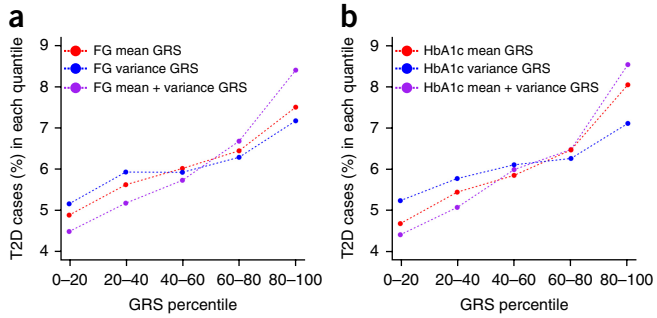

Figure 3 The percentage of type 2 diabetes cases in each quantile of the genetic risk scores. Combination of the GRSs for the mean and between-subject variance was weighted with the coefficients from logistic regression between T2D and the GRSs (**Supplementary Table 15**). (**a**) Fasting glucose (FG) GRS based on the 36 fasting-glucose-associated variants. (**b**) HbA1c GRS based on the 36 fasting-glucose-associated variants and 8 HbA1c-associated variants.

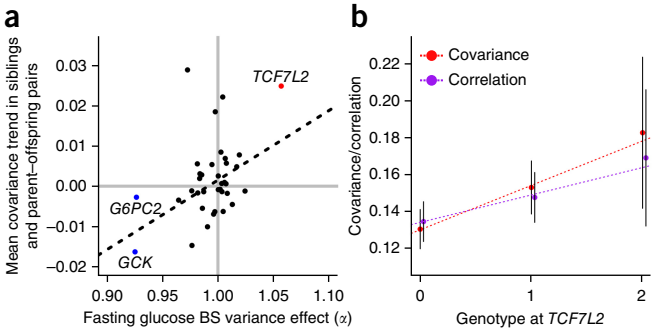

Figure 4 Covariance between genotype-concordant relative pairs. (**a**) Effects of 36 published fasting-glucose-associated variants on between-subject variance in fasting glucose levels and their glucose level covariance trends in pairs of relatives (**Supplementary Table 15**). Effects are given for the allele that increases fasting glucose levels. Variants are colored blue if they significantly decrease the variance and red if they significantly increase it ($P < 0.05/36 = 0.0014$). (**b**) Estimated covariance and correlation of fasting glucose measurements among pairs of relatives with the same genotype at *TCF7L2* and the 95% confidence intervals for the covariance and correlation estimates.

Conversely, removal of diabetic cases could create a variance effect in the presence of an effect on the mean, although we do not observe this phenomenon in our data. It is, however, likely that variants' effects on variance are at least partly due to their interaction with other variants and/or with environmental factors. This hypothesis is supported by the correlation between the variants' between-subject variance effects and their interaction with BMI.

We have also shown that variants that increase both mean fasting glucose levels and between-subject glucose variance increase T2D risk more than variants that increase fasting glucose but reduce the between-subject variance. These results largely account for the apparent discrepancy between the effects of variants on fasting glucose and their effects on T2D risk. This result is intuitively appealing, as T2D is primarily a disease of too high glucose; variants that increase both the mean and variance for glucose are more likely to be associated with pathologically high glucose levels than variants that only increase the mean or even have an increase in the mean offset by lower variance.

The variants in *GCK*, *G6PC2* and *TCF7L2* all affect fasting glucose levels, but their effects on T2D risk are not proportionate to their effects on glucose[22]. This may reflect different roles in glucose regulation. *GCK* and *G6PC2* encode enzymes that regulate glucose homeostasis, effectively establishing the glucose set point. Variants that increase mean glucose through these proteins will be countered by pressure to keep the glucose level within the physiological range, leading to reduced variance associated with these variants both within and between subjects. Similarly, variants that associate with increased HbA1c but not fasting glucose or T2D all associate with erythrocyte physiology and iron homeostasis and, where significant, lower HbA1c variance. Overall, this indicates low tolerance for variability in homeostatic regulation. In contrast, the variant associated with the highest variance in glucose levels is located in *TCF7L2*, which encodes a transcription factor that is thought to affect glucose levels through complex regulation of beta cell mass and function[23]. This variant affects beta cell response to glucose, leading to greater sensitivity to the environment and, thus, greater variability in glucose levels among carriers.

Only 2% of the heritability of fasting glucose levels is attributable to the effect of the 36 glucose-associated variants on mean levels. We have shown that variants that increase between-subject variance create positive covariance between individuals beyond their effects on the mean, increasing heritability estimates based on correlation between relative pairs. The effect of these markers on heritability is substantial and so is their contribution to the missing heritability of fasting glucose levels. Further, the effects of variants on the variability between individuals in glucose and HbA1c levels are as important for genetic risk prediction as the effects of variants on the mean.

URLs. T2D-GENES Consortium, GoT2D Consortium, DIAGRAM Consortium (2016-09-12), http://www.type2diabetesgenetics.org/.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors thank the subjects of the Icelandic deCODE study and the Iranian study for their participation. We also thank the staff at deCODE Genetics core facilities and all our colleagues for their contributions to this work. This research project has been supported by grant no. 121 NRCI Research Project and with the support of the National Research Council of the Islamic Republic of Iran.

AUTHOR CONTRIBUTIONS

E.V.I., V.S., P.S., H.H., U.T., D.F.G. and K.S. designed the study and interpreted the results. E.V.I. and D.F.G. performed statistical analysis. M.S.D., G.T., S.S., A.B.H., G.S., R. Bjarnason, A.V.T., R. Benediktsson, G.E., O.S., I.O., S.Z. and F.A. performed recruitment and phenotyping. The manuscript was drafted by E.V.I., V.S., D.F.G. and K.S. All authors contributed to the final version of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Scott, R.A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* **44**, 991–1005 (2012).
2. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
3. Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–244 (2014).
4. Manning, A.K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669 (2012).

5. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
6. Gaulton, K.J. *et al.* Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* **47**, 1415–1425 (2015).
7. Rönnegård, L. & Valdar, W. Recent developments in statistical methods for detecting genetic loci affecting phenotypic variability. *BMC Genet.* **13**, 63 (2012).
8. Wei, W.-H. *et al.* Major histocompatibility complex harbors widespread genotypic variability of non-additive risk of rheumatoid arthritis including epistasis. *Sci. Rep.* **6**, 25014 (2016).
9. Yang, J. *et al. FTO* genotype is associated with phenotypic variability of body mass index. *Nature* **490**, 267–272 (2012).
10. Topless, R.K. *et al.* Association of *SLC2A9* genotype with phenotypic variability of serum urate in pre-menopausal women. *Front. Genet.* **6**, 313 (2015).
11. Paré, G., Cook, N.R., Ridker, P.M. & Chasman, D.I. On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the Women's Genome Health Study. *PLoS Genet.* **6**, e1000981 (2010).
12. Perry, G.M.L. *et al.* Sex modifies genetic effects on residual variance in urinary calcium excretion in rat (*Rattus norvegicus*). *Genetics* **191**, 1003–1013 (2012).
13. Mackay, T.F. & Lyman, R.F. *Drosophila* bristles and the nature of quantitative genetic variation. *Phil. Trans. R. Soc. Lond. B* **360**, 1513–1527 (2005).
14. Shen, X., Pettersson, M., Rönnegård, L. & Carlborg, Ö. Inheritance beyond plain heritability: variance-controlling genes in *Arabidopsisthaliana*. *PLoS Genet.* **8**, e1002839 (2012).
15. Steinthorsdottir, V. *et al.* Identification of low-frequency and rare sequence variants associated with elevated or reduced risk of type 2 diabetes. *Nat. Genet.* **46**, 294–298 (2014).
16. Grant, S.F. *et al.* Variant of transcription factor 7-like 2 (*TCF7L2*) gene confers risk of type 2 diabetes. *Nat. Genet.* **38**, 320–323 (2006).
17. Perry, J.R. *et al.* Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in *LAMA1* and enrichment for risk variants in lean compared to obese cases. *PLoS Genet.* **8**, e1002741 (2012).
18. Cauchi, S. *et al.* The genetic susceptibility to type 2 diabetes may be modulated by obesity status: implications for association studies. *BMC Med. Genet.* **9**, 45 (2008).
19. Azizi, F. *et al.* Prevention of non-communicable disease in a population in nutrition transition: Tehran Lipid and Glucose Study phase II. *Trials* **10**, 5 (2009).
20. Soranzo, N. *et al.* Common variants at 10 genomic loci influence hemoglobin A_{1C} levels via glycemic and nonglycemic pathways. *Diabetes* **59**, 3229–3239 (2010).
21. Falconer, D.S. & Mackay, T.F.C. in *Introduction to Quantitative Genetics* 4th edn, Ch. 10 (Pearson, 1996).
22. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
23. Mitchell, R.K. *et al.* Selective disruption of *Tcf7l2* in the pancreatic β cell impairs secretory function and lowers β cell mass. *Hum. Mol. Genet.* **24**, 1390–1399 (2015).

ONLINE METHODS

Study subjects. *Iceland.* Measurements of glucose levels were available for a total of 117,548 Icelanders genotyped using Illumina chips. All study participants provided informed consent, and the study was approved by the Data Protection Commission of Iceland and the Icelandic National Bioethics Committee.

Iran. The Iranian subjects are part of the ongoing Tehran Lipid and Glucose Study[19], including 10,437 Iranians with 44,470 fasting glucose measurements genotyped using Illumina chips. All study participants provided informed consent. The study has been approved by the National Research Council of the Islamic Republic of Iran (no. 121) and has been performed with the approval of the Human Research Review Committee of the Endocrine Research Center, Shahid Beheshti University (M.C.).

SNP selection. The 36 fasting-glucose-associated variants were identified in a genome-wide association meta-analysis of up to 133,010 individuals of European ancestry without diabetes, including individuals genotyped using the Metabochip[1].

Whole-genome sequencing. The process used for whole-genome sequencing of the 8,453 Icelanders and the subsequent imputation have been described in a recent publication[24].

Association testing. *Mean effect.* Both fasting and non-fasting glucose measurements were transformed to a standard normal distribution using a rank-based inverse-normal transformation within each sex and each source separately and adjusted for age at measurement using a generalized additive model[25]. For each SNP, a classical linear regression, using the genotype as an additive covariate and mean glucose levels per subject as a response, was fit to test for association.

Between-subject variance effect. For each SNP, we fit a normal model where the mean glucose level per subject was regressed against the genotype and the between-subject variance was assumed to change multiplicatively with the genotype so that for non-carriers, heterozygotes and homozygotes the between-subject variance was assumed to be σ^2, $\alpha_{BS}\sigma^2$ and $\alpha_{BS}^2\sigma^2$, respectively (**Supplementary Note**).

Within-subject variance effect. For each SNP, we fit a normal model where glucose level measurements were regressed against the genotype and the within-subject variance was assumed to change multiplicatively with the genotype so that for non-carriers, heterozygotes and homozygotes the within-subject variance was assumed to be σ^2, $\alpha_{WS}\sigma^2$ and $\alpha_{WS}^2\sigma^2$, respectively (**Supplementary Note**).

Subjects in the data sets were related, causing the χ^2 test statistic to have mean >1 and median >0.675. We used a subset of 640,250 common SNPs to estimate the inflation factor λ and computed all P values by dividing the corresponding χ^2 values by λ to adjust for both relatedness and potential population stratification[26]. For the fasting glucose data set (I), $\lambda = 1.14$, and $\lambda = 1.21$ when estimating between-subject and within-subject variance effects, respectively.

BMI interaction effect. For each SNP, we fit an interaction regression model, using the genotype, BMI and the interaction term between the genotype and BMI as covariates and mean fasting glucose levels as the response. Both glucose levels and BMI measurements were transformed to a standard normal distribution using a rank-based inverse-normal transformation within each sex and each source separately and adjusted for age at measurement using a generalized additive model[25].

Thresholds for significance. In the set of 36 variants, significance thresholds for between-subject and within-subject variance effect were set to control the false discovery rate at 5% using standard Bonferroni correction ($P < 0.05/36 = 0.0014$).

Trend analysis. We assessed the relationship between the effects of sequence variants on mean and variance effects on glucose levels and their effect on T2D (log(OR)) using the following models:

A. T2D effect versus glucose mean effect: $\log(OR) = y_1\beta + \varepsilon$;
B. T2D effect versus glucose between-subject variance effect: $\log(OR) = y_2\log(\alpha_{BS}) + \varepsilon$;
C. T2D effect versus glucose mean and between-subject variance effect: $\log(OR) = y_1\beta + y_2\log(\alpha_{BS}) + \varepsilon$;
D. T2D effect versus glucose mean effect, between-subject variance effect and the interaction between glucose mean and between-subject variance effect: $\log(OR) = y_1\beta + y_2\log(\alpha_{BS}) + y_3(\beta \times \log(\alpha_{BS})) + \varepsilon$;
E. T2D effect versus glucose within-subject variance effect: $\log(OR) = y_4\log(\alpha_{WS}) + \varepsilon$;
F. T2D effect versus glucose mean and within-subject variance effect: $\log(OR) = y_1\beta + y_4\log(\alpha_{WS}) + \varepsilon$;
G. T2D effect versus glucose mean, between-subject variance and within-subject variance effect: $\log(OR) = y_1\beta + y_2\log(\alpha_{BS}) + y_4\log(\alpha_{WS}) + \varepsilon$

where β is the glucose mean effect, α_{BS} is the between-subject variance effect and α_{WS} is the within-subject variance effect. All models were fitted with a simple weighted linear regression where each variant was weighted by $f(1 - f)$, where f is the minor allele frequency of the variant, such that rare variants have less weight in the computation than common variants. The estimates and measures of goodness of fit are given in **Supplementary Table 6**.

Genetic risk scores. GRSs were constructed for both fasting glucose and HbA1c levels by combining the effect allele counts for the selected variants weighted by either the estimated mean effect or the between-subject variance effect of each allele on the trait.

Heritability. The correlation between close relative pairs is usually used to estimate heritability[21]. To assess how much variants effecting between-subject variance can contribute to heritability estimates, for each SNP, we estimated the covariance between siblings having the same genotype. Then, we performed a weighted linear regression between the estimated covariance and the genotype to assess the covariance trend. We weighted by the number of siblings having the genotype divided by the squared phenotypic variance given the genotype (**Supplementary Note**). This was repeated for parent–offspring pairs. The correlation between relatives is the ratio of their covariances and the geometric mean of their phenotypic variances. The correlation trend was therefore computed as the ratio of the covariance trend and variance trend (**Supplementary Note**).

A **Life Sciences Reporting Summary** for this paper is available.

Code availability. The code used to detect between-subject and within-subject variance effects is available as **Supplementary Code**.

Data availability. The authors declare that the data supporting the findings of this study are available within the article, its supplementary information files and upon request.

24. Gudbjartsson, D.F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
25. Hastie, T. & Tibshirani, R. Generalized additive models. *Stat. Sci.* **1**, 297–310 (1986).
26. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).

doi:10.1038/ng.3928

nature research

Corresponding Author: Gudbjartsson

Date: 05.07.2017

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   The sample size was determined by how many glucose measurements was available to us from the three laboratories in Iceland.

2. Data exclusions

   Describe any data exclusions.

   Data was excluded for subjects under 18 years old.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   We replicated our findings in an Iranian dataset.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   NA

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   NA

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☒	☐	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☒	☐	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *p* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

   Describe the software used to analyze the data in this study.

   We used R to analyze the data and produce figures along with custom code.

Nature Genetics: doi:10.1038/ng.3928

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	No eukaryotic cell line was used.
b. Describe the method of cell line authentication used.	No eukaryotic cell line was used.
c. Report whether the cell lines were tested for mycoplasma contamination.	No eukaryotic cell line was used.
d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell lines were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	See Study subjects chapter in Methods

Nature Genetics: doi:10.1038/ng.3928