Prediction of Fasting Blood Glucose Level within a Korean Population

Hyoung Doo Shin[1]*, Yeon Su Kim[1], In Ki Baek[1] and Hyun Sub Cheong[2]

[1]Department of Life Science, Sogang University, Seoul, Republic of Korea
[2]Research Institute for Life Science, GW Vitek, Inc. Seoul, Republic of Korea

Abstract

Fasting blood glucose (FBG) level is prevalent trait to predict several diseases. In the present study, We identified markers for prediction of blood glucose level. The genotype data of a total of 5,013 samples (2,373 men and 2,640 women) were obtained from the Korea Association Resource (KARE) project. We collected markers from a genome-wide association study (GWAS) catalog which included additional markers from nearby regions of GWAS catalog markers. In order to establish a FBG prediction model, we selected significant single nucleotide polymorphisms (SNPs) using a 10-fold cross-validation. In addition, we validated our prediction model using the final validation set. We selected a total of 7 SNP comprised of 2 SNPs (rs7754840 and rs12699673) for the men and 5 SNPs (rs102275, rs1574285, rs2908289, rs6494307 and rs917793) for the women from the 10-fold cross-validation process. The results of the 10-fold across-validation process in the men and the women indicated upward trends of FBG levels. Also, we validated our prediction model using the final validation set. In the final validation set, increased trends were observed across all of the sets. Our prediction model for FBG may be helpful to further FBG related studies.

Keywords: Genome-wide association study • SNP • Fasting blood glucose • weighted genetic risk score • SNP marker

Abbreviations: FBG: Fasting blood glucose; SNP: Single nucleotide polymorphism; GWAS: Genome-wide association study; KARE: Korea Association Resource; wGRS: Weighted genetic risk score; DM: Diabetes mellitus; T2DM: Diabetes mellitus type 2; GDM: Gestational diabetes mellitus; MAF:Minor allele frequency; LD: Linkage disequilibrium; CDKAL1: CDK5 regulatory subunit associated protein 1 like 1; DGKB: Diacylglycerol kinase beta; TMEM258: Transmembrane protein 258; GLIS3: GLIS family zinc finger 3; GCK: Glucokinase; C2CD4B: C2 calcium dependent domain containing 4B

Introduction

Diabetes mellitus (DM) is one of the most serious health problems in the world. Of the various diagnosis factors for DM, fasting blood glucose (FBG) level is an important factor for diagnosis of prediabetes as it has a high association with the development rate of diabetes. In addition, more recent studies have observed that prediabetes status is associated with risk of coronary heart disease in both diabetic and nondiabetic subjects. The comparison between subjects with prediabetes and normal subjects has also found that prediabetic status increases the risk of cardiovascular disease and other related symptoms such as stroke and coronary heart disease.

Due to the importance of FBG, large-scale genome-wide association studies have been conducted and have found numerous markers significantly associated with FBG level, such as G6PC2, FOXA2, MTNR1B and LEPR. Following replication studies also have identified the associations between these genes and FBG level. For example, one previous study reported significant association between the rs10830963 in MTNR1B and FBG level in Bosnian and Herzegovinian population. The rs560887 in G6PC2 also showed significant associations with reduced FBG levels in normal subject of Hispanic populations.

Several function studies have also provided evidences of a relationship between the significant SNPs and FBG level. The previous study demonstrated that two SNPs (rs560887 and rs2232316) in G6PC2 are responsible for increasing of FBG level by enhancing G6PC2 pre-mRNA splicing and regulating transcription factor binding affinity. Another study showed that rs1137100 in the LEPR gene is associated with glucose homeostasis.

Based on these previous observations, it is clear that there is a significant association between FBG levels and SNPs. And also, according to the relation of FBG with disease occurrence, we anticipate that the prediction of FBG level using SNP may suggest possible role in public health.

In the present study, we aimed to establish a prediction model of FBG for Korean men and women using the weighted genetic risk score (wGRS) method. We used the genotype data obtained from the Korea Association Resource (KARE) project. Only the SNPs which were reported in previous studies were used to construct prediction model. Also, we applied 10-fold cross-validation to our analysis procedure to improve the validity of our analysis.

Materials and Methods

Study subjects

A total 8,840 subjects (4,182 men and 4,658 women) initially used for this study. The genotype data was derived from KARE project that approved by Public Institutional Bioethics Committee designated by the Ministry of Health and Ware (P01-201502-31-002). For quality of data, we removed inadequate samples and SNPs that have the data of call rate with under 0.98. The data of SNPs were also deleted with minor allele frequency (MAF) under 0.05. Finally, the 5,013 subjects (2,373 men and 2,640 women) were selected for further

analyses. Among the 5,013 samples, we used 90% (2,135 men and 2,376 women) of samples as a SNP selection set for 10-fold cross validation process. The rest of samples (238 men and 264 women) were used as a final validation set [1,2].

Statistical analysis

The SNP were collected from GWAS catalog markers which were associated with blood glucose studies. In order to obtain genotype data, we used data of the collected markers of GWAS and other data derived from KARE data (2,581 SNPs) of nearby regions which are ±50kb from the GWAS markers. Among the SNPs, we performed calculating the coefficients of linkage disequilibrium (LD) to avoid high LD values using the Haploview software. Consequently, we obtained 207 SNP markers, which have shown related blood glucose studies previously. The regression analysis was conducted for the p-value that was used for identification SNPs in the training sets. We conducted the regression analysis using the GoldenHelixSVS8 software (Bozeman, MT, USA) [3-5].

SNP selection for fasting blood glucose

In order to prediction model, men subjects (2,135 samples) and women subjects (2,376 samples) were used for testing SNP selection sets separately. Using the training sets (1,921 men and 2,138 women) and the test sets (214 men and 238 women) of genotype data, 10-fold cross validation was conducted for identify SNPs. The SNP markers were selected for less than 0.05 of p-value across all sets from the top of 20 SNPs that were identified LD. Each p-value of 20 SNP markers is shown in Table 2 and Table 3. These SNPs were used for the wGRS [6,7]. The wGRS was calculated by using the following formula: Weighted genetic risk score $(wGRS) = \sum_{i=1}^{n} (w_i \times [SNP]_i)$ where w_i is weight that means the regression slope of selected SNPs, $[SNP]_i$ is the number of allele of FBG. We conducted analysis separately by gender, using these score to divide into 4 blocks after ascending sort of score. Then we calculated average of FBG values of each blocks and obtained the graphs. For validation these SNPs, the final validation sets (238 in men subjects and 264 in women subjects) were performed statistical analysis as same manner, resulting in trend line [8].

Results

Clinical profile of the study subjects

In the present study, we used the genotype data obtained from a total of 5,013 individuals comprised of 2,373 men and 2,640 women subjects. Of the total samples, 90% (n=4,511) were used in the SNP selection set to obtain the significant SNPs for constructing the prediction model [9]. The remaining 10% (n=502) were used in the Final validation set. The average age of men population was slightly lower than that of the women population in all analysis groups (51.57 vs. 52.62; 51.61 vs. 52.57; and 51.26 vs. 53.06 in the total subjects, SNP selection set and final validation set, respectively). The FBG level was generally higher in men than women (90.46 vs. 85.74; 90.32 vs. 85.53; 91.70 vs. 86.44 in total subjects, SNP selection set and final validation set). Detailed clinical profile of the samples is described in (Table 1).

SNP selection to construct prediction model

To select SNPs for FBG prediction model, we performed 10-fold cross validation based on regression analyses using a total of 207 SNPs which collected from a GWAS catalog and nearby SNPs. After performing the 10-fold cross-validation process, we found that two SNPs (rs7754840 in CDKAL1 and rs12699673 located near DGKB) had constantly significant association with FBG level for men subject in all 10 training sets (Table 2). Similarly, we selected five SNPs (rs102275 in TMEM258, rs1574285 in GLIS3, rs2908289 in GCK, rs6494307 located near C2CD4B and rs917793 in YKT6) which had consistent significant association with FBG level for women subject (Table 3). Detailed information of selected SNPs was listed in (Table 4) with their allele information, location and genotype counts.

Results of prediction model for Fasting blood glucose level

Using the selected SNPs, we constructed a prediction model for FBG using wGRS method. We observed upward trends of blood glucose levels with increasing of wGRS in both men and women population (Figure 1). Finally, we applied wGRS of both men and women population in the final validation set. As we expected, fasting blood glucose levels increased with wGRS (Figure 2).

The black line and red line represents the average FBG of the training sets and corresponding test sets, respectively. The standard deviations were used as error bars. (A) The analysis results of 10-fold cross validation using the men population. (B) The analysis results of 10-fold cross validation using the women population (Figure 1).

The black line represent result from the SNP selection set and the red line was obtained from Final validation set. (A) The FBG trends in the men population using 2 SNPs. (B) The FBG trends in the women population using 5 SNPs (Figure 2).

Discussion

As diabetes mellitus type 2 (T2DM) is important to public health problem, numerous studies have been focused on identifying significant markers for T2DM. However, several previous reports have indicated that FBG might play a role in the prevention and treatment for various glucose-related diseases. One previous study showed that an increase in FBG level was significantly associated with upward diagnosis factors such as BMI and HDL-c. Another recent study supported the correlation between FBG level and various clinical factors such as HbA1c, insulin, C-peptide and triglycerides. These observations indicated that a prediction model for FBG level might contribute to public health not only with respect to diseases related to blood glucose but various factors of diseases [10].

In the present study, we used a total of 5,013 individuals to establish a prediction model for FBG level. Due to the low average FBG level of women compared to men, we decided to construct gender-specific FBG models rather than a single comprehensive model using total subjects. Based on our results, we selected two SNPs (rs7754840 and rs12699673) for men and five SNPs (rs6494307, rs1574285, rs102275, rs2908289 and rs917793) for women population. Numerous previous studies have demonstrated the relationship between the SNPs and glucose-related phenotypes [11].

Table 1. The clinical profile of the samples.

	Sex	Number of Subjects	Age (Min-Max)	FBG (Min-Max) (mg/dL)
Total subjects	Male	2,373	51.57 (40-69)	90.46 (53-488)
	Female	2,640	52.62 (40-69)	85.74 (45-461)
SNP selection set	Male	2,135	51.61 (40-69)	90.32 (53-488)
	Female	2,376	52.57 (40-69)	85.53 (45-311)
Final validation set	Male	238	51.26 (40-69)	91.70 (66-291)
	Female	264	53.06 (40-69)	86.44 (65-461)

Average clinical characteristic of analysis groups including total, SNP selection set and final validation set;FBG, fasting blood glucose

Table 2. P-values of the top 20 SNPs in the SNP selection set and 10-fold across-validation using men population.

Markers	SNP selection set	Training sets for men during 10-fold cross-validation (n=1,921)									
	(n=2,135)	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Set 10
rs7754840	1.85×10^{-5}	3.05×10^{-5}	0.0001	0.0002	0.0003	1.25×10^{-5}	3.37×10^{-5}	3.10×10^{-5}	0.0002	7.15×10^{-5}	2.19×10^{-5}
rs12699673	0.002	0.003	0.01	0.01	0.01	0.01	0.004	0.004	0.003	0.004	0.005
rs6494307	0.009	0.01	0.006	0.01	0.002	0.02	0.03	0.02	0.07	0.006	0.01
rs4246215	0.02	0.02	0.03	0.02	0.02	0.04	0.09	0.04	0.03	0.05	0.01
rs895636	0.04	0.03	0.05	0.10	0.05	0.06	0.21	0.10	0.005	0.06	0.05
rs11558471	0.05	0.10	0.14	0.06	0.10	0.10	0.06	0.05	0.11	0.13	0.06
rs11857366	0.05	0.08	0.03	0.02	0.13	0.04	0.02	0.02	0.01	0.05	0.02
rs1337919	0.08	0.16	0.31	0.21	0.18	0.29	0.11	0.15	0.16	0.24	0.36
rs3847554	0.09	0.08	0.14	0.19	0.12	0.17	0.1	0.11	0.12	0.15	0.13
rs780093	0.10	0.13	0.10	0.12	0.14	0.14	0.15	0.15	0.03	0.12	0.12
rs10830962	0.11	0.08	0.13	0.08	0.05	0.32	0.11	0.13	0.15	0.15	0.17
rs11619319	0.11	0.29	0.20	0.15	0.10	0.38	0.29	0.31	0.11	0.24	0.33
rs1209523	0.12	0.20	0.38	0.29	0.23	0.39	0.17	0.21	0.23	0.3	0.44
rs10811661	0.14	0.08	0.08	0.05	0.08	0.12	0.07	0.08	0.33	0.07	0.19
rs4402960	0.20	0.16	0.15	0.18	0.45	0.19	0.24	0.07	0.06	0.23	0.22
rs12915227	0.24	0.20	0.20	0.17	0.30	0.13	0.11	0.07	0.02	0.10	0.12
rs13179048	0.28	0.33	0.61	0.61	0.44	0.49	0.71	0.33	0.17	0.69	0.56
rs4237150	0.48	0.78	0.45	0.66	0.48	0.46	0.44	0.57	0.49	0.51	0.15
rs4972516	0.58	0.95	0.58	0.63	0.34	0.66	0.51	0.70	0.61	0.80	0.44
rs340835	0.78	0.80	0.94	0.72	0.91	0.86	0.89	0.81	0.88	0.93	0.72

Table 3. P-values of the top 20 SNPs in the SNP selection set and 10-fold across-validation using women population.

Markers	SNP selection set (n=2,376)	Training sets for women during 10-fold cross-validation (n=2,138)									
		Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	Set 7	Set 8	Set 9	Set 10
rs6494307	0.002	0.003	0.003	0.003	0.002	0.003	0.003	0.001	0.001	0.002	0.001
rs1574285	0.002	0.002	0.004	0.002	0.01	0.004	0.005	0.002	0.001	0.003	0.006
rs102275	0.002	0.005	0.003	0.002	0.002	0.0006	0.002	0.002	0.001	0.001	0.001
rs2908289	0.01	0.009	0.02	0.01	0.03	0.01	0.004	0.02	0.02	0.01	0.04
rs917793	0.02	0.02	0.01	0.01	0.03	0.02	0.004	0.03	0.02	0.02	0.05
rs11558471	0.02	0.02	0.04	0.03	0.08	0.05	0.02	0.04	0.03	0.05	0.08
rs10244051	0.03	0.08	0.06	0.05	0.10	0.16	0.09	0.02	0.02	0.03	0.04
rs2206734	0.03	0.04	0.01	0.03	0.06	0.03	0.02	0.005	0.03	0.01	0.06
rs895636	0.07	0.31	0.18	0.18	0.15	0.15	0.11	0.07	0.07	0.07	0.05
rs10811661	0.08	0.11	0.19	0.17	0.22	0.08	0.16	0.21	0.19	0.18	0.25
rs10830962	0.08	0.07	0.05	0.05	0.02	0.05	0.16	0.05	0.05	0.06	0.02
rs2018860	0.11	0.21	0.09	0.15	0.10	0.14	0.41	0.06	0.11	0.11	0.07
rs16860794	0.19	0.32	0.27	0.43	0.24	0.47	0.30	0.24	0.40	0.38	0.22
rs10838524	0.23	0.15	0.21	0.24	0.24	0.31	0.40	0.26	0.29	0.28	0.29
rs6048205	0.24	0.18	0.48	0.39	0.31	0.57	0.39	0.52	0.42	0.36	0.33
rs340835	0.24	0.27	0.26	0.26	0.26	0.67	0.56	0.32	0.31	0.33	0.31
rs10209020	0.29	0.20	0.44	0.37	0.17	0.23	0.21	0.43	0.36	0.20	0.17
rs11041816	0.31	0.22	0.66	0.38	0.39	0.49	0.62	0.29	0.14	0.20	0.14
rs1209523	0.35	0.26	0.61	0.46	0.43	0.49	0.53	0.79	0.61	0.55	0.58
rs3847554	0.37	0.31	0.28	0.25	0.13	0.15	0.54	0.52	0.47	0.49	0.28

Of the two polymorphisms for men population, rs7754840 (linked to rs9356744, r2=0.902) which is located in CDKAL1 was significantly associated with the gestational diabetes mellitus (GDM). And it was used in the prediction model for T2DM in a Japanese population. This significance was also found in other populations such as Iranian population and in large-scale meta-analysis using various populations. Another study suggested an association between DPP-4 inhibitors which was used for T2DM treatment with rs7754840. The other selected SNP, rs2191349 (linked to rs12699673, r2=0.893) located near DGKB was responsible for various phenotypes of glucose metabolism. One previous study using Korean population proved an effect of rs2191349 in beta-cell function. Other several studies have also support the association between the reduction of insulin and fasting glucose concentration.

As with the men population, numerous evidences has been found for the polymorphisms in the women subjects, with the exception two SNPs (rs6494307 located near C2CD4B and rs1574285 in GLIS3). The two SNPs, rs174550 (correlated with rs102275 in TMEM258, r2=1.000) and rs4607517 (correlated with rs917793 in YKT6, r2=0.886), have been found to affect beta-cell function and were associated with GDM or prediabetes. In addition, rs730497 (absolute LD with rs2908289 in GCK, r2=1.000) showed an effect

Table 4. Information of seven SNPs for FBG prediction using Korean population.

Population	Markers	Gene	Location	Allele information			Genotype Count (Average fasting blood glucose)			Linked GWAS catalog SNP	Reference (PMID)
				Minor	Major	MAF	C/C	C/R	R/R		
Man	rs7754840	CDKAL1	6:20661019	C	G	0.472	685	1,135	553	rs9356744	Hwang, et al.
							-88.59	-89.3	-95.13	(r²=0.902)	-25187374
	rs12699673	Intergenic	7:15009979	C	T	0.303	1,145	1,016	212	rs2191349	Manning, et al.
							-92.14	-89.55	-85.69	(r²=0.893)	-22581228
Woman	rs102275	TMEM258	11:61790331	C	T	0.312	1,275	1,084	281	rs174550	Dupuis, et al.
							-86.85	-84.89	-82.83	(r²=1.000)	-20081858
	rs1574285	GLIS3	9:4283137	G	T	0.429	854	1,307	479	rs4237150	Hwang, et al.
							-83.85	-85.61	-88.78	(r²=0.888)	-25187374
	rs2908289	GCK	7:44184343	A	G	0.189	1,732	820	88	rs730497	Hwang, et al.
							-84.74	-87.5	-85.34	(r²=1.000)	-25187374
	rs6494307	Intergenic	15:62102491	G	C	0.438	839	1,287	514	rs7173964	Manning, et al.
							-86.42	-86.03	-83.27	(r²=0.981)	-22581228
	rs917793	YKT6	7:44206254	T	A	0.219	1,614	894	132	rs4607517	Dupuis, et al.
							-84.69	-87.2	-86.32	(r²=0.886)	-20081858

C/C, the homozygote of the major allele; C/R, the heterozygote; R/R, the homozygote of the minor allele. MAF, Minor allele frequency.
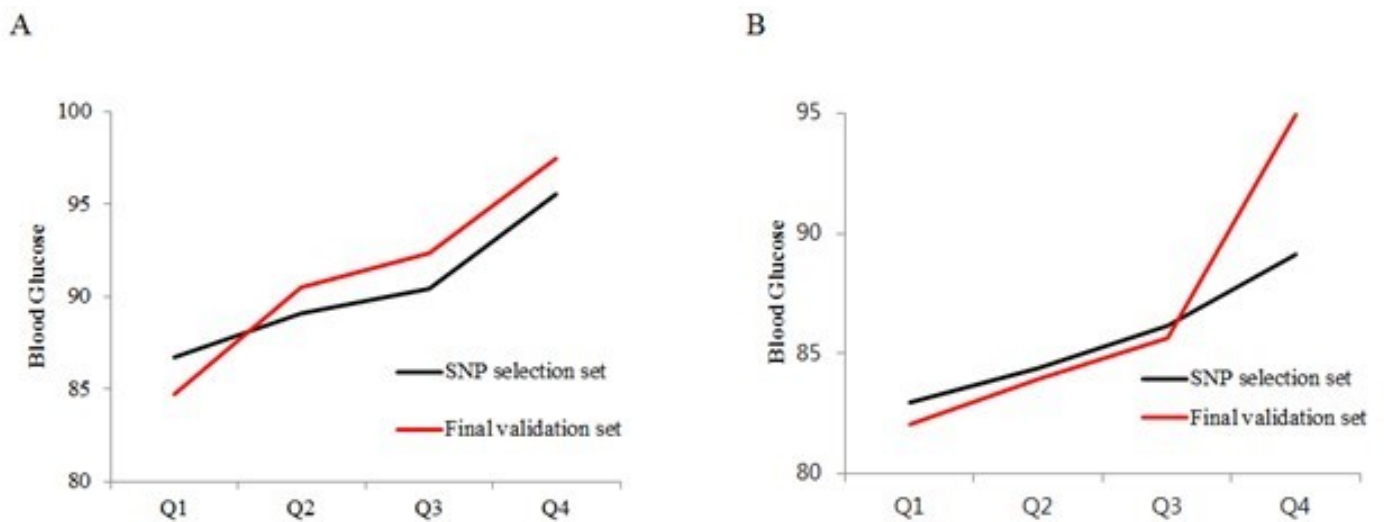

Figure 1. The average FBG level of men and women populations of the 10-fold cross-validation.
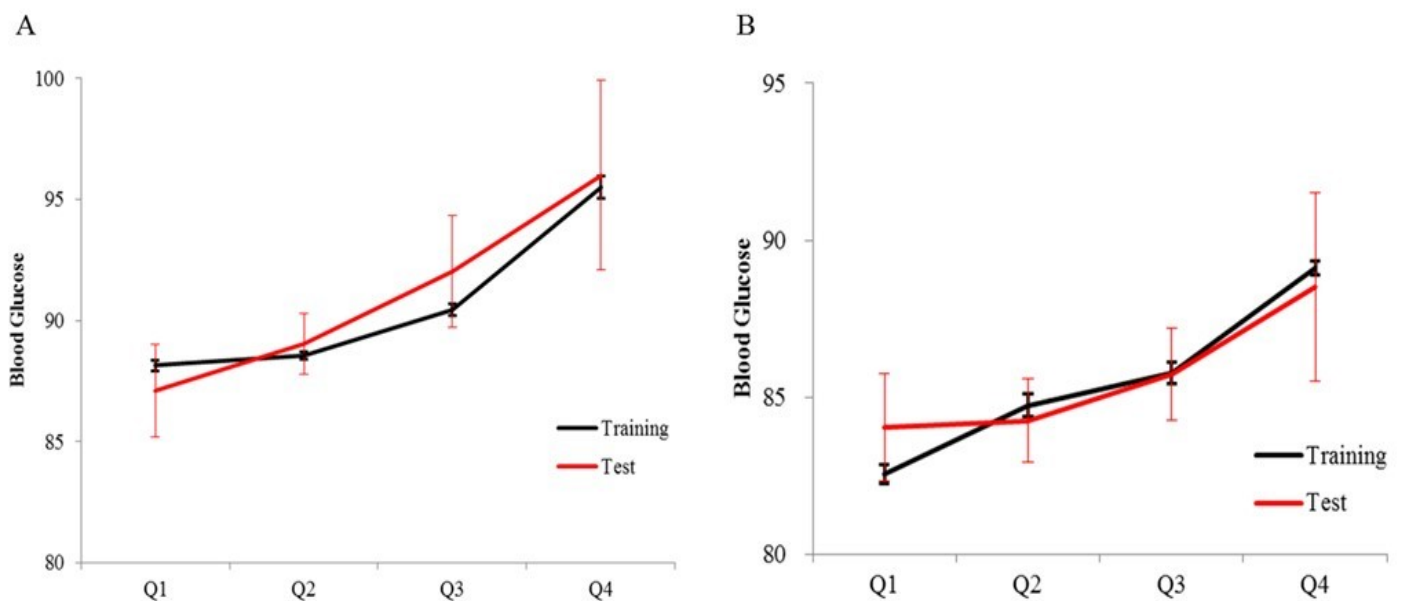

Figure 2. The FBG level of the final validation.

to HbA1c level which was an important diagnosis factor for chronic glycemic and hyperglycemia. However, there were no evidences of rs6494307 and rs1574285 impact on fasting glucose level or glucose related diseases. Further studies may be needed to identify the role of the SNPs in glucose metabolism.

The lack of confirmation using other independent cohorts was also one of limitation of the present study. Due to the problem, we selected only significant SNPs which had been reported on in previously studies. In this study, we set aside 10% of the total samples as a final validation set to overcome this limitation. Future studies might consider our limitation to build a more precise prediction model for FBG level using population-specific markers.

Conclusion

In summary, we constructed gender-specific FBG prediction models using genotype data from a Korean population. For the models, we selected two SNPs for men and five SNPs for women. Both of our models showed constantly upward trends of FBG level with increasing of wGRS. Prediction models of the present study might be useful for further glucose concentration studies and glucose-related diseases.

Acknowledgement

Conflict of Interest

All authors declare that they have no conflict of interest.

References

1. An, Ping, Iva Miljkovic, Bharat Thyagarajan and Aldi T. Kraja, et al. "Genome-wide association study identifies common loci influencing circulating glycated hemoglobin (HbA1c) levels in non-diabetic subjects: The Long Life Family Study (LLFS)." *Metab* 63 (2014): 461-468.

2. Baerenwald, D. A., A. Bonnefond, N. Bouatia-Naji and B. P. Flemming, et al. "Multiple functional polymorphisms in the G6PC2 gene contribute to the association with higher fasting plasma glucose levels." *Diabetologia* 56 (2013): 1306-1316.

3. Barrett, Jeffrey C., B. Fry, J. D. M. J. Maller and Mark J. Daly. "Haploview: Analysis and visualization of LD and haplotype maps." *J Bioinform* 21 (2005): 263-265.

4. Bouatia-Naji, Nabila, Ghislain Rocheleau, Leentje Van Lommel and Katleen Lemaire, et al. "A polymorphism within the G6PC2 gene is associated with fasting plasma glucose levels." *Sci* 320 (2008): 1085-1088.

5. Cho, Yoon Shin, Min Jin Go, Young Jin Kim and Jee Yeon Heo, et al."A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits." *Nat Genet* 41 (2009): 527-534.

6. Danaei, Goodarz, Saman Fahimi, Yuan Lu and Bin Zhou, et al. "Effects of diabetes definition on global surveillance of diabetes prevalence and diagnosis: A pooled analysis of 96 population-based studies with 331 288 participants." *Lancet Diabetes Endocrinol* 3 (2015): 624-637.

7. DECODE Study Group and European Diabetes Epidemiology Group. "Is the current definition for diabetes relevant to mortality risk from all causes and cardiovascular and noncardiovascular diseases?." *Diabetes Care* 26 (2003): 688-696.

8. Demirci, F.Y., A.S. Dressen, R.F. Hamman and C.H. Bunker, et al. "Association of a common G6PC2 variant with fasting plasma glucose levels in non-diabetic individuals." *Ann Nutr Metab* 56 (2010): 59-64.

9. Saxena, Richa, Benjamin F. Voight, Valeriya Lyssenko and Noël P. Burtt, et al. "Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels." *Science* 316 (2007): 1331-1336.

10. Emerging Risk Factors Collaboration. "Diabetes mellitus, fasting blood glucose concentration and risk of vascular disease: A collaborative meta-analysis of 102 prospective studies." *Lancet* 375 (2010): 2215-2222.

11. Lu, Y., K. Hajifathalian, M. Ezzati and M. Woodward, et al. "Global Burden of Metabolic Risk Factors for Chronic Diseases Collaboration (BMI Mediated Effects). Metabolic mediators of the effects of body-mass index, overweight and obesity on coronary heart disease and stroke: A pooled analysis of 97 prospective cohorts with 1.8 million participants." *Lancet* 383 (2014): 970-983.