Accepted Manuscript

Association of genetic variants with macronutrient intake in Circassian and Chechan populations in relation to diabetes


PII: S2214-5400(18)30030-6
DOI: doi:10.1016/j.mgene.2018.03.003
Reference: MGENE 415
To appear in: Meta Gene
Received date: 4 December 2017
Revised date: 21 February 2018
Accepted date: 4 March 2018

Please cite this article as: R. Dajani, S.J. Shbailat, A. Arafat, S. Sobar, H. Bitar, R. Tayyem, Z. Wei, H. Hakonarson, Association of genetic variants with macronutrient intake in Circassian and Chechan populations in relation to diabetes. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Mgene(2017), doi:10.1016/j.mgene.2018.03.003

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Title: Association of genetic variants with macronutrient intake in Circassian and Chechan populations in relation to diabetes.

Authors:
Dajani R†, Shbailat SJa, Arafat Ab, Sobar Sc, Bitar Hd, Tayyem Re, Wei Zf, Hakonarson Hg,h

aDepartment of Biology and Biotechnology, The Hashemite University, Zarqa, Jordan
bNational Center for Diabetes, Endocrinology and Genetics, Amman, Jordan
cCell Therapy Center, University of Jordan, Amman, Jordan
dFaculty of Medicine, University of Jordan, Amman, Jordan
eFaculty of Agriculture, University of Jordan, Amman, Jordan
fNew Jersey Institute of Technology, New Jersey, USA
gCenter for Applied Genomics, The Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA.
hDepartment of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA.

†Correspondence should be addressed to RD
Department of Biology and Biotechnology, The Hashemite University, Zarqa, Jordan
email: rdajani@hu.edu.
Telephone: 00962798859335

Running title: Macronutrient intake in relation to diabetes in ethnic populations
Abstract

Background and Aims: Type two Diabetes (T2D) is a complex disease including environmental and genetic factors. Macronutrient intake variation can be partly related to genetics. We propose in this study to analyse GWAS datasets for the Circassian and Chechan populations in order to determine associations between the GWAS data and nutrition intake in relation to diabetes.

Methods and Results: We analyzed a total of 36 traits including the macro/micronutrient intake in association with Genome wide association studies data for 34 persons with diabetes Chechans and Circassians. Within the Circassian population, there was a statistically significant association between carbohydrate and calorie intake and T2D associated SNPs on Histone deacetylase 9 gene- and between Vitamin B₂ intake and SNPs on a second locus of potential interest on chromosome 11 near LOC101928989 and teneurin transmembrane protein 4. Caffeine intake was also associated with significant SNPs unrelated to T2D. On the other hand, only calorie intake correlated with the occurrence of significant SNPs in Chechans, none of them were related to T2D.

Conclusion: we have identified a genetic association between intake of carbohydrate, calorie, vitamin B₂ and caffeine in the Circassian population and calorie intake in the Chechan population. Three of these intake traits (carbohydrate, calorie and vitamin B₂) were correlated with T2D development in Circassians. The association between macronutrient intake and diabetes development can shed light on causative variants for the pathogenesis of T2D.
Keywords
Circassian; Chechan; SNPs; Diabetes; ethnic

Introduction
Diabetes is among the most common non-communicable diseases globally. The International Diabetes Federation estimates that there are currently 194 million people aged 20 to 79 with diabetes worldwide and that this will increase to 333 million by 2025 (1). Diabetes is the fifth main cause of death in Jordan. According to a 2007 study by the Heart and Capillary Disease Prevention directorate (HCDP) of the Ministry of Health in Jordan, diabetes afflicts 16% of Jordanians over the age of 18. Another 23.8% of Jordanians over 18 years old are on the brink of becoming persons with diabetes, while the diabetes prevalence in Jordan is 30.5% among both children and adults (2). Despite extensive research efforts, the genetic basis of T2D remains largely unknown (3).

Over the past decade, Genome Wide Association Studies (GWAS) have been used to link regions of the human genome with complex diseases (4, 5). Researchers have linked haplotypes to type 2 diabetes, coronary heart disease and macular degeneration (3, 4, 6). Although, it has been difficult to uncover the functional significance of discovered SNPs, GWAS remain a way to discover genetic risk factors
for complex diseases. Most GWAS have been performed on European populations. There is a need to conduct GWAS on different ethnic populations to discover novel risk factors, especially if we are to apply these discoveries on other populations.

Previously, we performed GWAS on two ethnic populations of ancient descent, the Circassians and the Chechans. Both the Circassians and the Chechans, are the largest indigenous nationalities of North Caucasus (7, 8, 9). The Circassian and Chechan populations are descendants of a single ancient origin with later divisions along linguistic and geographic borders (8, 10). Previous analysis of classical genetic markers such as blood groups and serum proteins have shown statistical significant genetic diversity in the Caucasus (7, 11). The genetic diversity has been confirmed by mitochondrial DNA and Y chromosome analysis (8, 11). Under military pressure groups of Circassians and Chechans immigrated to Jordan 140 years ago. Circassians and Chechans in Jordan are endogamous and have managed to keep their separate sense of identity and ethnicity (12). The Circassian and Chechan communities represent unique populations that are genetically distinct from the Arab population.

A recent study has shown that the prevalence of impaired fasting glycemia and diabetes was 18.5% and 9.6%, respectively, for Circassians, and 14.6% and 10.1%, respectively, for Chechans. In both ethnicities, the prevalence of impaired fasting glycemia and diabetes was significantly higher in men, older age groups, married, subjects of lower educational level, past smokers, and subjects with obesity. Moreover, low high density lipoprotein (HDL) cholesterol was the most common abnormality observed in the Circassian and Chechan populations in Jordan (13).

A following study, using the same Circassian and Chechan persons with diabetes compared them to normal participants from both ethnicities. They found that
intake of all macronutrients and most of micronutrients (vitamin B\textsubscript{12}, folic acid, vitamin C, iron, selenium, and zinc) did not differ between participants with normal blood glucose and those with impaired fasting glucose or diabetes in both Circassian and Chechan populations (13,14).

In view of high incidence of T2D, we performed GWAS to search for T2D disease genes. Our GWAS studies demonstrated that multiple variants underlie T2D in the Chechen and Circassian populations in Jordan, some of which are shared with other ethnic groups. No single SNP met genome-wide significance criteria at $5 \times 10^{-8}$ level, in either populations (in publication). We propose in this study to further analyse the GWAS datasets for the Circassian and Chechan populations in order to determine associations between the GWAS data and nutrition intake of these populations in relation to diabetes. Macronutrients intake and dietary behavior, which is related to many factors such as culture, economics, social and psychologic and health beliefs, varies among individuals. This variance can be partly related to genetic components. The heritability is reported to range from 8% to 70% (1). The genetics behind macronutrient intake can shed light on the biology of dietary behavior. There have been GWAS that have identified chromosomal regions associates with macronutrient intake (2-5).

The present study sheds more light on multiple genetic factors that underlie T2D in association with macro/micronutrient intake in Chechen and Circassian populations in Jordan. Some of these factors may be shared with other ethnic groups.

Materials and Method
Subjects: The sample consisted of 34 Circassians and 34 Chechans. Inclusion criteria included: diagnosis with T2D and receiving medical treatment for it, pure ethnicity (Circassian, Chechen) in the last 3 generations, living in Jordan, age >30 years old, age of diagnosis >18 years old, and time since diagnosis and commencement of treatment >6 months.

GWAS Data: GWAS data produced from two previous research projects funded by King Hussein institute for biotechnology and cancer (KHCIBC) (Chechan population) and the Higher council for science and technology (HCST) (Circassian population). High-throughput, genome-wide SNP genotyping was conducted using the Infinium II OMNI-Express BeadChip technology (Illumina), at the Center for Applied Genomics (CAG) at the Children’s Hospital of Philadelphia (CHOP), USA, according to manufacturer’s standard protocol. Both SNP genotype data and intensity data including information of Log R Ratio (LRR) and B Allele Frequency (BAF) was extracted from GenomeStudio project files.

Biochemical data: Biochemical data produced from two previous research projects funded by KHCIBC (Chechan population) and HCST (Circassian-population) included Lipid parameters (Total Cholesterol, HDL, low density lipoprotein (LDL) and triacylglycerol (TG)), and glucose were analyzed for all samples using Enzymatic assays. The following definitions and cutoff points were used in the study according to the (ADA) guidelines (2014): Glycemic control was considered satisfactory (good) if HbA1c levels were <7.0%, and unsatisfactory (poor) if HbA1c levels were ≥7.0%. An elevated LDL level was defined as ≥100 mg/dL, a low HDL level was defined as <40 mg/dL in males and <50 mg/dL in females, an elevated cholesterol level was defined as ≥200 mg/dL, and an elevated triglycerides level was defined as ≥150 mg/dL. Patients
were considered dyslipidemic if they had at least one of the previously mentioned lipid abnormalities.

**Macro/micronutrient intake data:** Macro/micronutrient data produced from two previous research projects funded by KHCIBC (Chechan population) and HCST (Circassian population). 24-hour recall was collected from participants to estimate mean nutrient intake. The individual was asked to recall information on portion sizes, cooking methods, recipe of dishes, condiments, and beverages for the past 24-hour. All details about estimating portion sizes were explained by trained research assistants. The intakes of total calories, macro- and micronutrients as well as other dietary substances were calculated using ESHA Food Processor (Version 7.9, 1987–2002: Salem, OR). Foods not listed in ESHA were entered item by item after getting its components and recipe from the participant.

**SNPs analysis:** We also performed a manual search of the NCBI database for all genes related to the list of the annotated nominally significant SNPs (P<0.05). We did not find databased-evidence discussing the exact SNP. Therefore, we used a multi-step search technique, which starts by identifying the SNP, determining its location and the related gene. Then, we searched NIH database for studies that discussed the related gene and if it is correlated with T2D. After that, we described the full name of the gene, its function (if known) and the major disease associations based on the NIH database.

**Statistical Analysis**

All statistical tests for association were carried out using the software package plink (http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml). Principal components analysis was carried out using Eigenstrat 3.0 (http://genepath.med.harvard.edu/~reich/Software.htm ). The single-marker analysis for
the genome-wide data was carried out for different traits using linear regression on allele counts with the first 10 principle components as covariates. P values and effect sizes (log odds ratios) with the corresponding 95% confidence intervals were calculated for the association analysis.

A Manhattan plot is a scatter plot used to display a large number of data-points such as genome-wide association studies (GWAS). In GWAS Manhattan plots, genomic coordinates are displayed along the X-axis, with the negative logarithm of the association P-value for each single nucleotide polymorphism (SNP) displayed on the Y-axis. Each dot on the Manhattan plot signifies a SNP. Because the strongest associations have the smallest P-values, their negative logarithms will be the greatest.

A Q–Q plot is a quantile-quantile probability plot where graphical method is used for comparing two probability distributions by plotting their quantiles against each other. First, the set of intervals for the quantiles is chosen. A point (x, y) on the plot corresponds to one of the quantiles of the second distribution (y-coordinate) plotted against the same quantile of the first distribution (x-coordinate). Thus the line is a parametric curve with the parameter which is the number of the interval for the quantile. If the two distributions being compared are similar, the points in the Q–Q plot will approximately lie on the line y = x. In a case-control comparison of genotype markers, if the cases are enriched for specific markers, they will deviate away from the control data at a certain significance level. If there is no difference, then y = x.

For calculation of genomic inflation factor also known as lambda(λ) in genome-wide association studies (GWAS), λ is defined as the median of the resulting chi-squared test statistics divided by the expected median of the chi-squared distribution. The median of a chi-squared distribution with one degree of freedom is 0.4549364. A λ value can be
calculated from z-scores, chi-square statistics, or p-values, depending on the output from the association analysis. The algorithm we used is automated in the software PLINK, which provides accurate measures on the population inflation. In our instance GIF was 1.00, indicating there was no evidence of population stratification causing inflation.

Quality control
After filtering out SNPs with minor allele frequency <0.05, missing rate > 0.05, Hardy-Weinberg Equilibrium Test P value < 0.0001, we ended up with ~588K SNPs for association analysis. The top hit SNP rs16871944 has reached genome wide significance level (p=3.35E-08).

An association being detected means that its effect size should be large enough at the given sample size. For example, as we show in our result section, we found the estimated effect size for SNP rs16871944 is beta=1071, with 95% confidence interval of [177, 1419]. Given that N=103, minor allele frequency = 0.27 (rs16871944), and response CALs SD=500, we can obtain its power as 0.85 when beta=1071. We now make a power vs effect size plot with effect size beta ranging from 177 to 1419. The tool website for GPC used is R package “GeneticsDesign” (Figure 6).

Results
We analyzed a total of 36 traits including the macro/micronutrient intake and biochemical data with valid values in association with the GWAS data for 34 persons with diabetes Chechans and 34 persons with diabetes Circassians (Table 1 -list of traits). For each trait that showed SNPs with P<0.001 we plotted a Manhattan plot and a QQ plot as described in the methods section.
Only one trait, Calorie intake (CALs) (genomic inflation factor 1.00), showed SNPs with P<0.001 in the Chechan population (Figure 1.a -QQ-plot; Figure 1.b -Manhattan plot). However, none of these SNPs has been reported in the literature to be related to diabetes. We further analyzed these SNPs and found a number of them associated with fatty liver disease. In Table 2 for each trait we show all the SNPS that were statistically significant and their related genes and diseases.

Four traits showed potential significant signals in the Circassian population. The first trait is carbohydrate intake (CARB) (genomic inflation factor 1) and the second trait is calorie intake (CAL) (genomic inflation factor 1.06736) (Figure 2.a -QQ-plot and b Manhattan plot, and Figure 3.a -QQ-plot and b Manhattan plot, respectively). Both showed SNPs with P<0.001 on Histone deacetylase 9 (HDAC9) gene which is reported to be relevant to diabetes. Other SNPs on other genomic regions that are related to CAL but not to T2D were associated with fatty liver disease Table 2.

The third trait that showed SNPs that were statistically significant was Vitamin B\textsubscript{2} (genomic inflation factor 1.09848) (Figure 4.a -QQ-plot and b. Manhattan plot). The SNPs were located on a second locus of potential interest on chromosome 11. This locus is not on any gene (intergenic) and the nearby genes are LOC101928989 and teneurin transmembrane protein 4 (TENM4). Other SNPs in other genomic regions that are associated with vitamin B2 trait but not with T2D were correlated with the occurrence of other diseases such as ovarian and cervical cancers Table 2.

The fourth trait that showed SNPs that were statistically significant was caffeine intake. Although caffeine intake (CAFF) (genomic inflation factor 1.03214) (Figure 5.a -QQ-plot and b. Manhattan plot) showed significant SNPs with P<0.001, none of them
were related to T2D. These SNPs were of significant interest, because they were associated with Huntington disease and lateral temporal epilepsy Table 2.

Discussion
In the present study, we analyzed a total of 36 traits including macro/micronutrient intake and biochemical data in Chechen and Circassian populations. Among these traits, only CALs intake was associated with significant SNPs in Chechans, however, none of these SNPs was related to T2D. On the other hand, CALs, CARB, Vitamin B₂ and CAFF intake were associated with significant SNPs in Circassians, and all these traits except CAFF were related to T2D. In addition, our research revealed 7 SNPs that were associated with macro/micronutrient intake but unrelated to diabetes, and found that they were associated with other diseases.

We found that SNPs correlated with CALs and CARB intake in Circassians were carried on *HDAC9* gene. This gene is reported to be relevant to T2D. It is highly expressed in insulin producing β-cells of pancreas (15, 16), and its product acts epigenetically to prevent the differentiation of these cells during embryogenesis and after birth (16). A previous study showed that one SNP carried on this gene was significantly associated with high BMI, T2D and hyperlipidemia, and as a result it significantly modified the susceptibility to coronary artery disease and increased the severity to coronary atherosclerosis (17). In our study, SNPs identified on *HDAC9* gene in T2D patients might contribute to the occurrence/and or severity of the disease by changing the expression of the gene, and consequently changing the epigenetic regulation of β-cells differentiation during development.
We also showed that Vitamin B₂ intake was associated with SNPs located on an intergenic region carried on chromosome 11. This region exists near LOC101928989 and TENM4 genes. The former is a hypothetical gene with still uncharacterized function, while the later encodes for Teneurin-4, a member of type II transmembrane glycoprotein family which includes Teneurin 1-4 (18). TENM4 is expressed in the central nervous system and in the mesenchymal tissues including cartilage and muscles (19). Despite its association with several diseases, no relevance was reported in literature between TENM4 and T2D. Nevertheless, chromosome 11, where the intergenic region is located, was shown to be related to T2D.

Chromosome 11 carries six genes that are implicated in insulin secretion [ATP binding cassette subfamily C member 8 (ABCC8, also known as SUR1), potassium voltage-gated channel subfamily J member 11 (KCNJ11, also known as Kir6.2), uncoupling protein 2 (UCP2) and melatonin receptor 1B (MTNR1B)], signaling [exostosin glycosyltransferase 2 (EXT2)] and resistance [uncoupling protein 3 (UCP3)] (20). SNPs on genes related to secretion (21, 22, 23, 24) or signaling (25, 26) of insulin were associated with T2D. Moreover, variants on UCP3 gene were correlated with insulin resistance causing T2D (27). The ethnic background is also associated with certain SNPs that might have a role in the occurrence of T2D (20). Some of them are carried on chromosome 11. For example, certain SNPs on UCP2-UCP3 gene cluster were correlated with T2D in Caucasian Americans (28). Other variants near Fanconi anemia complementation group F (FANCF) gene were associated with young-onset of T2D in Pima Indian Americans (29). Furthermore, variants in an intergenic region of chromosome 11p12 might contribute to the occurrence of T2D in Finnish population (30). Although chromosome 11 carries SNPs specific for certain ethnicity, it also carries
certain variants on genes such as potassium voltage-gated channel subfamily Q member 1 (KCNQ1) (31; 32), KCNJ11 (22; 33), ABCC8 (22), UCP2 (34; 35; 36) and ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1 (ARAP1, also known as CENTD2) (37; 33) that were reported to be related to T2D in wide range ethnic groups. Taken together, the SNPs we identified on chromosome 11 might have a role in down-regulating the insulin pathway or increasing the insulin resistance. Further studies are required to uncover the underlying mechanisms of different polymorphic alleles and to explore whether these SNPs are specific for Circassians or prevalent in other ethnicities.

We explored the significant variants that were associated with macro/micronutrient intake but unrelated to diabetes, and found that they were associated with other diseases. One SNP (rs2143571) was related to CALs intake. It was carried on the sorting and assembly machinery component 50 homolog (SAMM50) gene and associated with susceptibility to nonalcoholic fatty liver disease (NAFLD) (38). Four SNPs were related to Vitamin B2 intake. Two of them were carried on DNA methyltransferase 3 alpha (DNMT3A) (rs11887120) and deoxyuridine triphosphate (DUT) (rs3784621) genes (39; 40). The former was correlated with the occurrence of ovarian cancer (39), while the later was correlated with cervical cancer associated with human papillomavirus persistence (40). The third SNP (rs1393350) was located on tyrosinase (TYR) gene and used as a predictor of the eye color (41,42). The last SNP (rs12904216) was carried on solute carrier family 12 member 1 (SLC12A1) gene and was associated with developing thiazolidine-related edema in patients with T2D (43). Although this SNP was unrelated to T2D in our study, the difference in ethnicity between our study (Circassians) and the previous one (Taiwanese) may explain this inconsistency. Finally, two SNPs were related to CAFF intake. One of them...
(rs2276881) is carried on *huntingtin (HTT)* gene and associated with Huntington’s disease (44), while the other (rs7099034) was located at *glioma inactivated 1 (LGI1)* gene and might have a role in the occurrence of lateral temporal epilepsy (45).

Our study has its limitations since it is based on a select cohort of T2D individuals, and as a result, they may not be generalizable. Moreover, the small size of the cohort may have prevented us from detecting more modest effects. Although we observed consistent patterns of association, many of these associations were of nominal significance. Nevertheless, there have been GWAS studies that have been successful with subject numbers that are comparable to ours (46, 47). Therefore, replication in an independent cohort should be conducted to support or refute the obtained results. However, obtaining a second cohort from these populations is extremely difficult as these are the first ever samples collected. In addition, nutrients intake estimation was based on 24-hour recall which depends on memory, and as a consequence recall bias is expected.

In summary, our study showed new SNPs are associated with T2D, and that they play a role in nutrient-specific choice and dietary preference. Although the associations reported here are significant, the effects are small in magnitude. Given the complexity of the investigated traits, it is likely that other as yet unidentified SNPs and copy number variants, contribute to macro/micronutrient intake. Future studies should focus on conducting the same experiments in non-European cohorts to determine whether or not the identified SNPs are associated with macro/micronutrient intake in other ethnicities.
Funding:

This research was funded by the Hashemite University.

Conflict of interest: None

Author contribution: RD designed research project and wrote the manuscript; HH and ZW contributed to the design of the research project and performed the data analysis and contributed to the writing; SS, RT, SS, HB and AA contributed to data analysis and writing of the manuscript

References


The expression of teneurin-4 in the avian embryo: potential roles in patterning of 
19. Suzuki N, Mizuniwa C, Ishii K, Nakagawa Y, Tsuji K, Muneta T, Sekiya I, 
Akazawa C (2014a) Teneurin-4, a transmembrane protein, is a novel regulator 
20. Kaul N, Ali S (2016) Genes, Genetics, and Environment in Type 2 Diabetes: 
21. Crispim D, Fagundes NJ, dos Santos KG, Rheinheimer J, Bouças AP, de Souza 
UCP2 gene are associated with proliferative diabetic retinopathy in patients with 
variants in the KCNJ11-ABCC8 region with type 2 diabetes. Genet Mol Res 
12(3):2990-3002.
(2014) ABCC8 genetic variants and risk of diabetes mellitus. Gene 545(2):198- 
204.
24. Ren J, Xiang AH, Trigo E, Takayanagi M, Beale E, Lawrence JM, Hartiala J, 
Richey JM, Allayee H, Buchanan TA, Watanabe RM (2014) Genetic variation 
in MTNR1B is associated with gestational diabetes mellitus and contributes only 
to the absolute level of beta cell compensation in Mexican Americans. 
Diabetologia 57(7):1391-1399.


Figure 1 a QQ-plot of Chechan for train CALs (calorie intake)

Figure 1 b Manhattan Plot of Chechan for trait CALs (calorie intake). SNPs are sorted by chromosomal location along X axis against their –log10(P-value) shown on Y axis. The horizontal line is for the suggested significance level p=10-5.
Figure 2 a QQ plot of Circassians for trait CARB (Carbohydrate intake).

Figure 2 b Manhattan Plot of Circassian for trait CARB (Carbohydrate intake). SNPs are sorted by chromosomal location along X axis against their –log10(P-value) shown on Y axis. The horizontal line is for the suggested significance level p=10-5.
Figure 3 a QQ plot for trait CALs (Calorie intake) in Circassian population

Figure 3 b Manhattan Plot of Circassian for trait CALs (Calorie intake). SNPs are sorted by chromosomal location along X axis against their $-\log_{10}(P\text{-value})$ shown on Y axis.

The horizontal line is for the suggested significance level $p=10^{-5}$.
Figure 4 a QQ Plot of Circassian for trait Vitamin B2

Figure 4 b Manhattan Plot of Circassian for trait Vitamin B2. SNPs are sorted by chromosomal location along X axis against their $-\log_{10}(P\text{-value})$ shown on Y axis. The horizontal line is for the suggested significance level $p=10^{-5}$
Figure 5 a QQ Plot of Circassian for trait CAFF

Figure 5 b Manhattan Plot of Circassian for trait CAFF. SNPs are sorted by chromosomal location along X axis against their $-\log_{10}(P\text{-value})$ shown on Y axis. The horizontal line is for the suggested significance level $p=10^{-5}$
Figure 6 A power vs effect size plot with effect size beta ranging from 177 to 1419. The tool website for GPC used is R package “GeneticsDesign”
<table>
<thead>
<tr>
<th>Number</th>
<th>Trait</th>
<th>ABBREVIATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Protein</td>
<td>PROT</td>
</tr>
<tr>
<td>2</td>
<td>Carbohydrate</td>
<td>CARB</td>
</tr>
<tr>
<td>3</td>
<td>Fiber</td>
<td>FIBER</td>
</tr>
<tr>
<td>4</td>
<td>Lactose</td>
<td>LACT</td>
</tr>
<tr>
<td>5</td>
<td>Fat</td>
<td>FAT</td>
</tr>
<tr>
<td>6</td>
<td>Monounsaturated fatty acid</td>
<td>MONO</td>
</tr>
<tr>
<td>7</td>
<td>Saturated fatty acid</td>
<td>SAT</td>
</tr>
<tr>
<td>8</td>
<td>Polyunsaturated fatty acid</td>
<td>POLY</td>
</tr>
<tr>
<td>9</td>
<td>Trans fatty acid</td>
<td>TFA</td>
</tr>
<tr>
<td>10</td>
<td>Cholesterol</td>
<td>CHOL</td>
</tr>
<tr>
<td>11</td>
<td>Vitamin A (international unit)</td>
<td>AIU</td>
</tr>
<tr>
<td>12</td>
<td>Vitamin B1</td>
<td>B1</td>
</tr>
<tr>
<td>13</td>
<td>Vitamin B2</td>
<td>Vitamin B2</td>
</tr>
<tr>
<td>14</td>
<td>Vitamin B3</td>
<td>B3</td>
</tr>
<tr>
<td>15</td>
<td>Vitamin B12</td>
<td>B12</td>
</tr>
<tr>
<td>16</td>
<td>Vitamin B6</td>
<td>B6</td>
</tr>
<tr>
<td>17</td>
<td>Biotin</td>
<td>BIOT</td>
</tr>
<tr>
<td>18</td>
<td>Vitamin C</td>
<td>VITC</td>
</tr>
<tr>
<td>19</td>
<td>Vitamin D</td>
<td>DIU</td>
</tr>
<tr>
<td>20</td>
<td>Vitamin E</td>
<td>EIU</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>21</td>
<td>Folic acid</td>
<td>FOLAC</td>
</tr>
<tr>
<td>22</td>
<td>Vitamin K</td>
<td>VITK</td>
</tr>
<tr>
<td>23</td>
<td>Pantothenic acid</td>
<td>PANTHOTHIE</td>
</tr>
<tr>
<td>24</td>
<td>Calorie</td>
<td>CALC</td>
</tr>
<tr>
<td>25</td>
<td>Iron</td>
<td>IRON</td>
</tr>
<tr>
<td>26</td>
<td>Selenium</td>
<td>SELNIUM</td>
</tr>
<tr>
<td>27</td>
<td>Caffeine</td>
<td>CAFF</td>
</tr>
<tr>
<td>28</td>
<td>Omega-3</td>
<td>OMEGA3</td>
</tr>
<tr>
<td>29</td>
<td>Omega-6</td>
<td>OMEGA6</td>
</tr>
<tr>
<td>30</td>
<td>Zinc</td>
<td>ZINC</td>
</tr>
<tr>
<td>31</td>
<td>Glucose</td>
<td>GLUCOSE</td>
</tr>
<tr>
<td>32</td>
<td>Glycosylated hemoglobin</td>
<td>HBAIC</td>
</tr>
<tr>
<td>33</td>
<td>Cholesterol</td>
<td>CHOL</td>
</tr>
<tr>
<td>34</td>
<td>High-Density lipoprotein</td>
<td>HDL</td>
</tr>
<tr>
<td>35</td>
<td>Low-Density lipoprotein</td>
<td>LDL</td>
</tr>
<tr>
<td>36</td>
<td>Triglyceride</td>
<td>TRIG</td>
</tr>
</tbody>
</table>

Table 1 Macro/micronutrient intake and biochemical traits for Chechan and Circassian populations
<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene ID</th>
<th>Locus ID</th>
<th>Chromosome #</th>
<th>Biochemical factor</th>
<th>Disease</th>
<th>Functional class</th>
<th>P-value</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs21435</td>
<td>SAMM50</td>
<td>25813</td>
<td>22</td>
<td>CALS</td>
<td>Genetic Variants in the SAMM50 Gene Create Susceptibility to Nonalcoholic Fatty Liver Disease in a Chinese Han Population</td>
<td>Intron-variant</td>
<td>0.000006646</td>
<td>(Chen et al. 2015)</td>
</tr>
<tr>
<td>rs11887</td>
<td>DNM3A</td>
<td>1788</td>
<td>2</td>
<td>B2MG</td>
<td>Ovarian CA</td>
<td>intron-variant</td>
<td>0.000002139</td>
<td>(Kelemen et al. 2008)</td>
</tr>
<tr>
<td>rs13 933 50</td>
<td>TYR</td>
<td>7299</td>
<td>11</td>
<td>B2MG</td>
<td>Prediction of eye color in the Slovenia population using the IrisPlex SNPs.</td>
<td>Intronic variant</td>
<td>0.0000 4833</td>
<td>(Kastelic and Drobnic 2012; Kastelic et al. 2013)</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>------</td>
<td>----</td>
<td>------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------</td>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td>rs37 846 21</td>
<td>DUT</td>
<td>1854</td>
<td>15</td>
<td>B2MG</td>
<td>risk for HPV persistence and progression to cervical cancer.</td>
<td>Intronic variant</td>
<td>0.0000 3019</td>
<td>(Wang et al. 2010)</td>
</tr>
<tr>
<td>rs12 904 2A1</td>
<td>SLC1 2A1</td>
<td>10537 08066</td>
<td>15</td>
<td>B2MG</td>
<td>Genetic predisposition</td>
<td>intronic variant,upstr</td>
<td>0.0000 6383</td>
<td>(Chang et al. 2011)</td>
</tr>
<tr>
<td>rs22</td>
<td>HTT</td>
<td>3064</td>
<td>4</td>
<td>CAFF</td>
<td>Huntingtonton’s disease</td>
<td>reference synonymous codon</td>
<td>0.0000</td>
<td>1753</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>------</td>
<td>---</td>
<td>------</td>
<td>------------------------</td>
<td>-----------------------------</td>
<td>--------</td>
<td>-----</td>
</tr>
<tr>
<td>768</td>
<td></td>
<td>81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs70</td>
<td>LGI1</td>
<td>10537</td>
<td>10</td>
<td>CAFF</td>
<td>Lateral temporal epilepsy</td>
<td>intron-variant</td>
<td>0.0000</td>
<td>001995</td>
</tr>
<tr>
<td>990</td>
<td></td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>211</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 List of macro/micronutrient associated loci associated with other diseases from both Circassian and Chechan populations
Abbreviation list

CAG: Center for Applied Genomics

GWAS: Genome Wide Association Studies

HDL: high density lipoprotein

HSDP: the Heart and Capillary Disease Prevention directorate

KHCIBC: King Hussein Cancer Institute for Biotechnology and Cancer

LDL: low density lipoprotein

SNPs: Single Nucleotide Polymorphisms

T2D: Type two diabetes
Dajani

Highlights
Title: Association of genetic variants with macro/micronutrient intake in Circassian and Chechan populations in relation to diabetes"

- Type two Diabetes (T2D) is a complex disease including environmental and genetic factors. Food intake plays a role in the manifestation of T2D. Macronutrient and micronutrient intake varies among individuals. This variance can be partly related to genetic components.
- This is the first study to study the association between GWAS and food intake in diabetic patients in the Circassian and Chechan population in Jordan.
- We were able to show that within the Circassian population, traits including carbohydrate, calorie and vitamin B₂ intake that were found to correlate with the occurrence of T2D were also shown to be associated with single nucleotide polymorphisms (SNPs) located on certain genomic regions.
- The association between macronutrient intake and diabetes development can shed light on causative variants for the pathogenesis of T2D.