Genome-wide association study identifies novel type II diabetes risk loci in Jordan subpopulations

Rana Dajani[1,*], Jin Li[2,3,*], Zhi Wei[4], Michael E. March[2], Qianghua Xia[3,5], Yousef Khader[6], Nancy Hakooz[7], Raja Fatahallah[8], Mohammed El-Khateeb[8], Ala Arafat[8], Tareq Saleh[1], Abdel Rahman Dajani[1], Zaid Al-Abbadi[1], Mohamed Abdul Qader[1], Abdel Halim Shiyab[9], Anwar Bateiha[6], Kamel Ajlouni[8] and Hakon Hakonarson[2,5,10]

[1] Department of Biology and Biotechnology, Hashemite University, Zarqa, Jordan
[2] Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, United States of America
[3] Department of Cell Biology, Tianjin Medical University, Tianjin, China
[4] Department of Computer Science, New Jersey Institute of Technology, Newark, NJ, United States of America
[5] Divisions of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA, United States of America
[6] Department of Community Medicine, Public Health and Family Medicine, Faculty of Medicine, Jordan University for Science and Technology, Irbid, Jordan
[7] Department of Biopharmaceutics and Clinical Pharmacy, School of Pharmacy, University of Jordan, Amman, Jordan
[8] National Center for Diabetes, Endocrinology and Genetics, Amman, Jordan
[9] Department of Anthropology, Yarmouk University, Irbid, Jordan
[10] The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States of America
[*] These authors contributed equally to this work.

ABSTRACT

The prevalence of Type II Diabetes (T2D) has been increasing and has become a disease of significant public health burden in Jordan. None of the previous genome-wide association studies (GWAS) have specifically investigated the Middle East populations. The Circassian and Chechen communities in Jordan represent unique populations that are genetically distinct from the Arab population and other populations in the Caucasus. Prevalence of T2D is very high in both the Circassian and Chechen communities in Jordan despite low obesity prevalence. We conducted GWAS on T2D in these two populations and further performed meta-analysis of the results. We identified a novel T2D locus at chr20p12.2 at genome-wide significance (rs6134031, $P = 1.12 \times 10^{-8}$) and we replicated the results in the Wellcome Trust Case Control Consortium (WTCCC) dataset. Another locus at chr12q24.31 is associated with T2D at suggestive significance level (top SNP rs4758690, $P = 4.20 \times 10^{-5}$) and it is a robust eQTL for the gene, *MLXIP* ($P = 1.10 \times 10^{-14}$), and is significantly associated with methylation level in *MLXIP*, the functions of which involves cellular glucose response. Therefore, in this first GWAS of T2D in Jordan subpopulations, we identified novel and unique susceptibility loci which may help inform the genetic underpinnings of T2D in other populations.

Subjects Genetics, Genomics, Diabetes and Endocrinology, Medical Genetics
Keywords eQTL, Methylation, Meta-analysis, Type 2 diabetes, Genome-wide association study

Submitted 3 April 2017
Accepted 6 July 2017
Published 17 August 2017

Corresponding authors
Rana Dajani, rdajani@hu.edu.jo
Hakon Hakonarson, hakonarson@email.chop.edu

Academic editor
Daniela Foti

Additional Information and Declarations can be found on page 13

DOI **10.7717/peerj.3618**

INTRODUCTION

Diabetes is among the most common non-communicable diseases globally. It has been estimated that there are currently about 194 million people at the age of 20 to 79 with diabetes worldwide and that this number will further increase to 333 million by 2025 (*Wild et al., 2004*). Diabetes is the fifth main cause of death in Jordan, afflicting 16 percent of Jordanian adult citizens; another 23.8 percent of adults in Jordan are also on the brink of becoming diabetics according to a study from 2007 by the Heart and Capillary Disease Prevention directorate (HCDP) of the Ministry of Health in Jordan; and the rate of diabetes prevalence in Jordan is 30.5 percent among both children and adults (*Ajlouni et al., 2008*). Thus, diabetes presents a significant public health burden to the Jordan community. Type II Diabetes (T2D) is the major type of diabetes, which accounts for 95% percent of all diabetes cases worldwide.

Despite extensive research efforts for more than a decade and some notable successes, much of the genetic basis of common human diseases remains unresolved (*Hirschhorn & Daly, 2005*). The genome-wide association study (GWAS) has been a powerful approach for identifying novel susceptibility loci for complex diseases (*Barrett & Cardon, 2006*; *Pe'er et al., 2006*), such as T2D. To date, more than 80 T2D susceptibility loci have been uncovered by GWAS. However, the heritability attributed to these loci remains as low as just 10% (*Imamura et al., 2016*). In addition, these studies have mostly focused on populations of European ancestry and East Asians, with a few studies on South Asians and Mexicans. The genetic determinants of T2D in Middle East populations have not been extensively studied by GWAS and limited evidence suggested that at least some of the reported T2D loci showed differential associations in different populations in the Middle East (*Mtiraoui et al., 2012*). It has also been reported that the presentation of T2D is different between Middle East immigrants and European patients (*Glans et al., 2008*), implying some different genetic basis between populations. Given the prevalence of the disease in the region, more research is warranted to understand the genetic basis of T2D specific to given Middle Eastern populations.

The Circassians and the Chechens are two ethnic populations of ancient descent in Jordan, both of which are the largest indigenous nationalities of the North Caucasus (*Barbujani, Nasidze & Whitehead, 1994*; *Bulayeva, 2006*; *Nasidze et al., 2001*). These two populations are descendants of a single ancient origin with later divisions along linguistic and geographic borders (*Nasidze et al., 2004*; *Nasidze et al., 2001*). After immigrating to Jordan 140 years ago, Circassians and Chechens in Jordan are endogamous and have managed to keep their separate sense of identity and ethnicity during the last one hundred years in Jordan (*Kailani, 2002*). Previous analysis of classical genetic markers such as blood groups and serum proteins have also shown statistical significant genetic diversity in the Caucasus (*Barbujani, Nasidze & Whitehead, 1994*; *Barbujani et al., 1994*), which has been further confirmed by mitochondrial DNA and Y chromosome analysis (*Nasidze et al., 2004*; *Nasidze et al., 2001*). While a T2D GWAS has been conducted in the Lebanese population (*Ghassibe-Sabbagh et al., 2014*), the Lebanese are Arab in origin; Circassians and Chechans are a separate, non-Arab ethnic group. These are clearly different populations, with different

ancestries. The Circassian and Chechen communities may provide us an opportunity to study a genetically unique population and compare genetic basis for complex human diseases between different populations.

T2D has become an alarming public health issue in Jordan. Epidemiology studies showed that the prevalence of impaired fasting glycemia is 18.5% and 14.6% and prevalence of diabetes is 9.6% and 10.1% for Circassians and Chechens, respectively (*Dajani et al., 2012*). In view of the very high incidence of T2D in Jordan and the genetic distinctness of Circassian and Chechan populations, we performed a GWAS to search for genetic factors contributing to T2D in these two populations and compared the results with European population.

MATERIALS & METHODS

Ethics statement

The study has been approved by the institutional review board committee at the National Center for Diabetes, Endocrinology and Genetics of Jordan (approval number: 457/9.MS). The written informed consent was given by all participants.

Study subjects and sample collection

A random sample of $N = 144$ from the Chechen population in Jordan and a random sample of $N = 140$ from the Circassian population in Jordan were recruited to participate in the study. Each participant in the study filled out a survey that included pedigree information. The identities of parents, grandparents, and great-grandparents (both maternally and paternally) were reported in the survey and any individual with non-Chechen heritage for even one person in his/her pedigree was excluded for the Chechen subpopulation; the same identity confirmation was conducted for the Circassian subpopulation.

A subject was defined as affected by diabetes mellitus if this diagnosis is known to the patient or, according to the ADA definitions, if fasting serum glucose is 7 mmol/L (126 mg/dl) or more. Impaired fasting glucose was defined as a fasting serum glucose level of $\geq$6.1 mmol/L (100 mg/dl) but <7 mmol/L. The glycemic control was assessed using HbA_{1c}. Patients with previously diagnosed diabetes who had HbA_{1c} >7% were defined as having 'unsatisfactory' glycemic control.

Sample collection

A total of 9 ml of whole blood was drawn in EDTA tubes from the subjects by vacutainer system. Genomic DNA was isolated from whole blood sample using the phenol-chloroform protocol.

Genotyping and quality control

We performed high-throughput, genome-wide SNP genotyping, using the InfiniumII OMNI-Express BeadChip technology (Illumina), at the Center for Applied Genomics (CAG) at the Children's Hospital of Philadelphia (CHOP), USA. Sample quality control (QC) was performed based on the following measures: sample call rate, overall heterozygosity, relatedness testing and other metrics. Samples were excluded from analysis for SNP call rate <95%, heterozygosity beyond five standard deviation of the mean. One sample from each pair of duplicated or cryptic related samples was removed. For each pair

of duplicate or related samples the sample with the highest SNP call rate was kept in the dataset. In the SNP-based QC, SNPs with a call rate <95%, minor allele frequency <1% or showing significant deviation from Hardy-Weinberg-Equilibrium (HWE test P-value $<10^{-4}$) in the controls were removed. All QC steps were carried out using the software package PLINK (*Purcell et al., 2007*).

Principal component analysis (PCA)

PCA was conducted to confirm ethnic identity and to generate covariates to control for population stratification in the association analysis. LD-pruning was performed using PLINK, and only independent ($r^2 < 0.2$), autosomal non-GC/AT SNPs were included in the PCA, which was conducted using EIGENSTRAT (*Price et al., 2006*) version 3.0.

Association analysis and meta-analysis

The single-marker analysis for the genome-wide data was carried out using logistic regression on allele counts with the first 10 principle components as covariates. P values and odds ratios with the corresponding 95% confidence intervals were calculated for the association analysis in Chechen and Circassian subpopulations separately. Both association and meta-analysis were performed using PLINK.

The WTCCC cohort

The cohort of European population was from WTCCC, which has been reported before (*Wellcome Trust Case Control Consortium, 2007*). All the samples were genotyped on Affymetrix Genome-Wide Human SNP Array 5.0. We similarly performed sample and SNP based QC steps and excluded non-European subjects based on PCA. Logistic regression was performed including the first three principal components as covariates.

Imputation analysis

The regional imputation at the locus of chr12q24.31 was conducted in two steps. First, the genotype data were prephased with SHAPEIT (*Delaneau, Marchini & Zagury, 2012*; *Delaneau, Zagury & Marchini, 2013*) version 2, and then genotype imputation was performed using IMPUTE 2 (*Howie, Donnelly & Marchini, 2009*; *Marchini et al., 2007*) with the 1000 Genome Phase 3 (https://mathgen.stats.ox.ac.uk/impute/1000GP%20Phase%203%20haplotypes%206%20October%202014.html) as the reference panel. Missing data likelihood score test was conducted to assess the association of each imputed SNP genotype with T2D using software SNPTEST (*Marchini et al., 2007*) V2, including the first three principal components as covariates. SNPs with info score <0.8 or with HWE-test p-value $<1 \times 10^{-06}$ were excluded from association testing.

Analysis of methylation data

Genomic DNA of a subset of samples in the biorepository of CAG was isolated from peripheral blood mononuclear cells. Genome-wide methylation profiling was conducted on the Infinium HumanMethylation450 BeadChip Kit at CAG according to the manufacturers' protocols. Methylation data were exported from the Illumina GenomeStudio and loaded into the R statistical package (r-project.org) using the lumi package (*Du et al., 2010*; *Lin et al., 2008*). After adjusting for quantile color balance and background level and simple

Table 1 The number of samples after quality control filtering.

Ethnicity	Cases		Controls		Total
	N	Male %	*N*	Male %	*N*
Chechen	34	47%	109	40%	143
Circassian	33	39%	105	45%	138
Total	67		214		281

Notes.
N, Number.

scaling normalization, *M*-value density and CpG-site intensity were plotted and aberrant chips were removed. These samples have also been genotyped at CAG and their genetic ethnicity was checked by PCA. We extracted the *M*-values (the log2 ratio between the methylated and unmethylated probe intensities) and the genotype information of the 425 subjects of European ancestry. We removed subjects of missing genotype at SNP rs4758690 and extreme outlier values of methylation *M*-values ($\geq$median *M*-value of the genotype group $\pm$ 3SD) and then assessed the association between the additive genotype at rs4758690 and methylation *M*-value in gene *MLXIP* using linear regression including sex, age, and 10 genotype-derived principle components. Box-plots were generated using R package.

RESULTS

Identification of novel T2D signals in Jordan subpopulations

To understand the genetic basis for T2D in Jordan populations, we conducted GWAS in Chechen and Circassian subpopulations of Jordan. The sample information after QC is summarized in Table 1. Specifically, for the Chechen subpopulation, we have 34 cases and 109 controls; for the Circassian subpopulation, we have 33 cases and 105 controls (Table 1). Approximately 645,000 SNPs in each subpopulation passed QC. We conducted logistic regression analyses separately in each population, including ten genotype-derived principal components as covariates. There was no signal that reached genome-wide significance, however there are several SNPs at suggestive level of significance ($P < 1 \times 10^{-4}$) in each subpopulation (Tables S1–S2). Then we performed meta-analysis of the association results from the two subpopulations. In the meta-analysis, we observed a signal at genome-wide significant level (SNP rs6134031, P-value $= 1.12 \times 10^{-8}$) under both fixed effect model and random effect model (Fig. S1, Fig. 1, Table 2). This SNP is located at the 5′ of the *JAG1* gene (Fig. 1). In addition, there is another signal with multiple SNPs showing suggestive evidence of association (P-value $< 1 \times 10^{-4}$), with SNP rs4758690 having the lowest P-value at 4.20×10^{-5} (Fig. S1, Table 2, Fig. 1). SNP rs4758690 is located in the intron of *MLXIP*, a gene involved in transcriptional regulation of genes in glucose metabolism. Taken together, these results demonstrate significant GWAS associations to novel T2D susceptibility loci in Jordan subpopulations.

Test the association signals in European population

We then investigated whether these association signals exist in populations of other ethnicities. We examined the association of these SNPs in the T2D dataset of the Wellcome Trust Case Control Consortium (WTCCC) (*Wellcome Trust Case Control Consortium,*

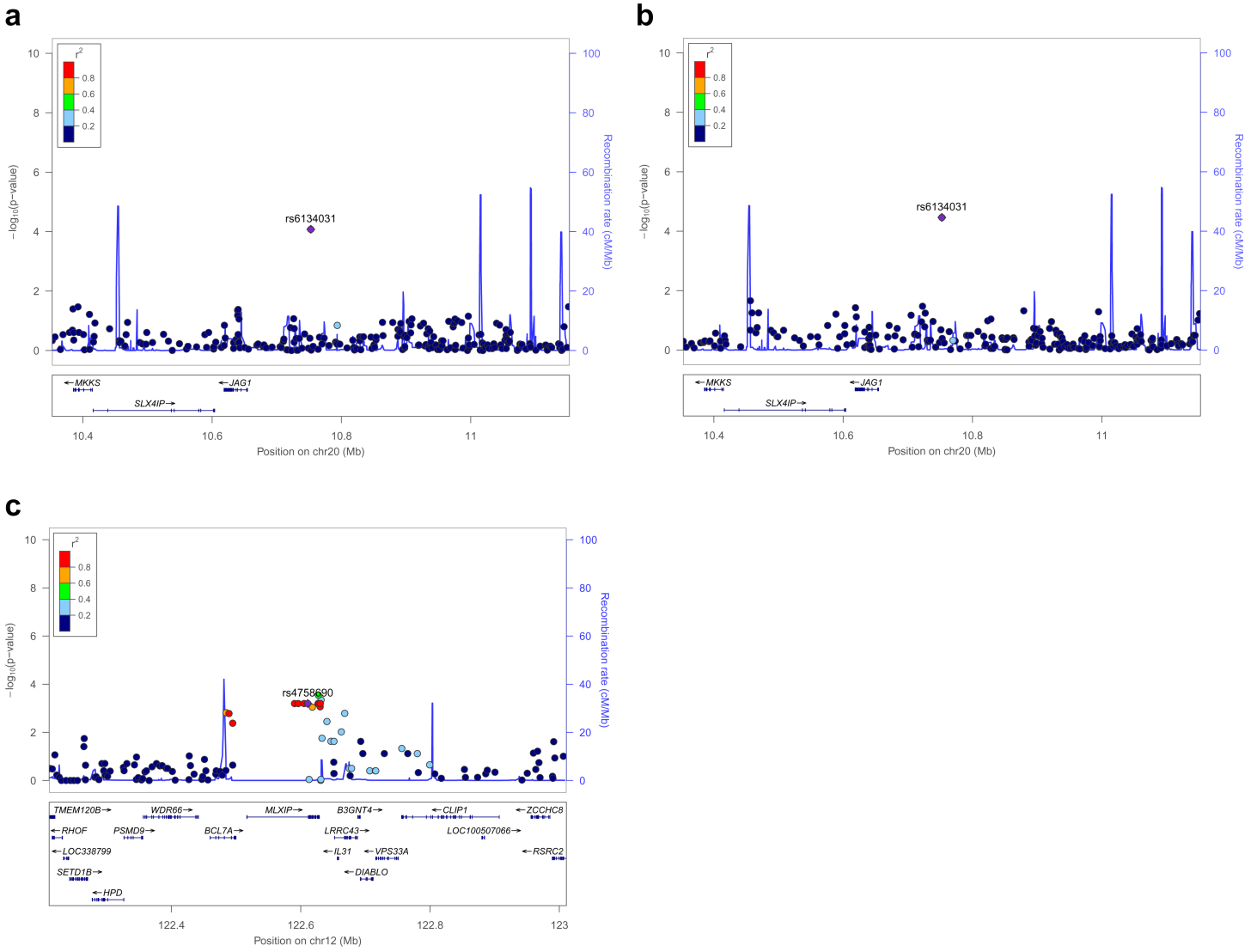

Figure 1 The regional association plots for the top associated loci. (A) chr20p12.2 locus in Circassian population; (B) chr20p12.2 locus in Chechen population; (C) chr12q24.31 in Chechen population. The top associated SNP at each locus is shown in purple and the LD between the remaining SNPs and the index SNP are indicated by their colors. The r^2 values were calculated from the each population using software PLINK (*Purcell et al., 2007*). The recombination rates are shown by the light blue lines and the genomic positions are on human genome build hg19. The plots were made using software LocusZoom (*Pruim et al., 2010*).

2007) which is composed of 1,999 cases and 3,004 controls, genotyped on the Affymetrix Genome-Wide Human SNP Array 5.0. After QC, 1,952 cases and 2,960 controls of European ancestry remained for association analysis by logistic regression. The top SNP in the Jordan analysis, rs6134031 demonstrated nominally significant association with T2D in the WTCCC cohort ($P = 0.012$) and the same direction of effect (Table 2). The SNP rs4758690 is not genotyped on the Affymetrix GW5.0 Array, so we conducted imputation over this region in the replication cohort. Based on the imputed genotype data, we did not observe a significant association to rs4758690 (OR = 1.01, $P = 0.61$).

Table 2 Top associations ($P < 5 \times 10^{-5}$) found in meta-analysis of Circassian and Chechen subpopulations.

SNP	Chr	Pos (hg19)	Gene	A1/A2	Ethnicity	MAF cases/controls	OR (95% CI)	*P*-value
rs6134031	20	10752610	*JAG1*	T/C	Circassian	0.50/0.25	9.48 (3.09,29.07)	8.36×10^{-5}
					Chechen	0.51/0.23	9.84 (3.33,29.02)	3.45×10^{-5}
					Meta		9.66	1.12×10^{-8}
					European	0.28/0.26	1.12 (1.03,1.23)	0.012
rs4758690	12	122610909	*MLXIP*	G/A	Circassian	0.59/0.41	2.41 (1.19,4.91)	0.015
					Chechen	0.60/0.38	3.89 (1.78,8.47)	6.36×10^{-4}
					Meta		3.00	4.20×10^{-5}
					European	0.53/0.52	1.01 (0.93,1.09)	0.61

Notes.
SNP, single nucleotide polymorphism; Chr, chromosome; Pos, Position; A1, minor allele; A2, major allele; MAF, minor allele frequency; OR, odds ratio; CI, confidence interval.

Correlation of T2D variants with *MLXIP* gene expression and methylation

Interrogating these T2D variants in the GTEx dataset (*GTEx Consortium, 2015*), we uncovered a nominally significant association between SNP rs6134031 and *JAG1* expression, in Esophagus–Muscularis (Beta $= -0.15$, $P = 0.0073$, Fig. S2) and a marginal correlation in pancreatic tissue which is of potential biological relevance to T2D (Beta $= -0.13$, $P = 0.071$, Fig. S2). Though it is not significant, we did observe a trend of association between the doses of minor allele T and a lower expression of *JAG1*.

On the other hand, we found a genome-wide significant eQTL effect of SNP rs4758690 for gene *MLXIP* expression in transverse colon (Beta $= 0.46$, $P = 1.10 \times 10^{-14}$) and small intestine terminal ileum (Beta $= 0.50$, $P = 4.20 \times 10^{-7}$) tissue specimens (Fig. 2). A similar significant eQTL effect was reported for *MLXIP* expression in normal pre-pouch ileum in another study examining eQTLs in human intestine tissues (*Kabakchiev & Silverberg, 2013*).

Further, we found that SNP rs4758690 is significantly associated with the methylation probe cg22729539 ($P = 3.07 \times 10^{-5}$) residing within an intron of the longest isoform of *MLXIP* (Fig. 3). This site is absent in other short isoforms. We observed a positive correlation between the eQTL and the methylation data at this locus. As methylation is one of the important mechanisms regulating gene expression, these results are of potential interest. The minor allele G confers a lower expression of *MLXIP* compared to the major allele A, as well as a reduced methylation level at probe cg22729539, consistent with previous reports that gene body methylation was found to be positively correlated with gene expression (*Yang et al., 2014*). In addition, cg22729539 resides in a region with multiple histone modifications and transcription factor binding in pancreatic islets and liver cells which are central to T2D (Fig. S3) and additional T2D relevant cell lines (Table S3) (*Bhandare et al., 2010*; *Encode Project Consortium, 2012*; *Parker et al., 2013*; *Pasquali et al., 2014*; *Roadmap Epigenomics et al., 2015*). The bound transcription factors include CEBPB which is known to function in adipogenesis (*Darlington, Ross & MacDougald, 1998*), ER stress and pancreatic β cell failure (*Matsuda et al., 2010*) (Table S3), therefore this region may function as active cis-regulatory element, regulating *MLXIP* expression.

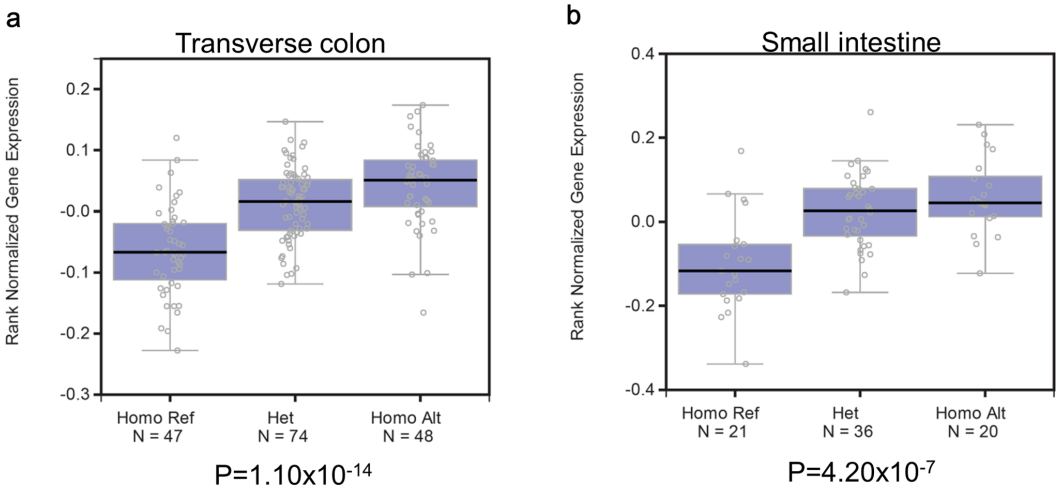

Figure 2 Box plots showing the association between SNP rs4758690 genotype and gene *MLXIP* expression level. (A) in tissue transverse colon, beta = 0.46, $P = 1.10 \times 10^{-14}$; (B) in tissue small intestine, beta = 0.50, $P = 4.20 \times 10^{-7}$. The *in silico* analyses were conducted at GTEx Protal (*GTEx Consortium, 2015*). The sample groups of different rs4758690 genotype were indicated on the *X*-axis; and the relative expression level of *MLXIP* is shown on the *Y*-axis. The median value of *MLXIP* expression level in each genotype group is represented by the dark black horizontal line in the box plot. In the both figures, the reference allele is G and the alternative allele is A.

The expression of *JAG1* and *MLXIP*

The biological relevance of these two genes to T2D was further strengthened by their expression pattern. For *JAG1*, it is reported to be highly expressed in arteries and in bronchial epithelial cells and lung tissue, with a particularly high level of expression in the gastrointestinal tract tissues, such as small intestine and colon (Figs. S4 and S5). For the gene *MLXIP*, high levels of expression have been consistently noticed in colon tissue as reported in different studies (Figs. S6 and S7). Both of these genes demonstrated medium level of expression in certain tissues highly relevant to T2D, including *JAG1* in adipose, pancreas, and smooth muscle (Figs. S4 and S5), and *MLXIP* in muscle, pancreas and pancreatic islet cells (Figs. S6 and S7).

The overall expression pattern of *JAG1* is similar to that of the gene Coagulation Factor III (*F3*) (correlation > 0.7), genetic polymorphisms of which have been shown to be associated with T2D in different ethnicity groups (*Palmer et al., 2012*; *Yamada et al., 2006*; *Yamaguchi et al., 2007*) and the expression of which is significantly higher in monocytes and neutrophils of diabetes and prediabetic subjects (*Ichikawa et al., 1998*).

Consistent with the expression pattern, knockout of *JAG1* in a mouse model resulted in defects in endocrine/exocrine glands, homeostasis/metabolism, and the liver/biliary system (Fig. S8) (*Blake et al., 2017*; *Finger et al., 2017*). *MLXIP*-deficient mice displayed distinct metabolic features including increased serum lactate and alanine levels, consumption of fatty acids for energy production during exercise, and increased glycolytic capacity in skeletal muscles. These features are associated with T2D in humans (*Crawford et al., 2010*; *Imamura et al., 2014*; *Karpe, Dickmann & Frayn, 2011*).

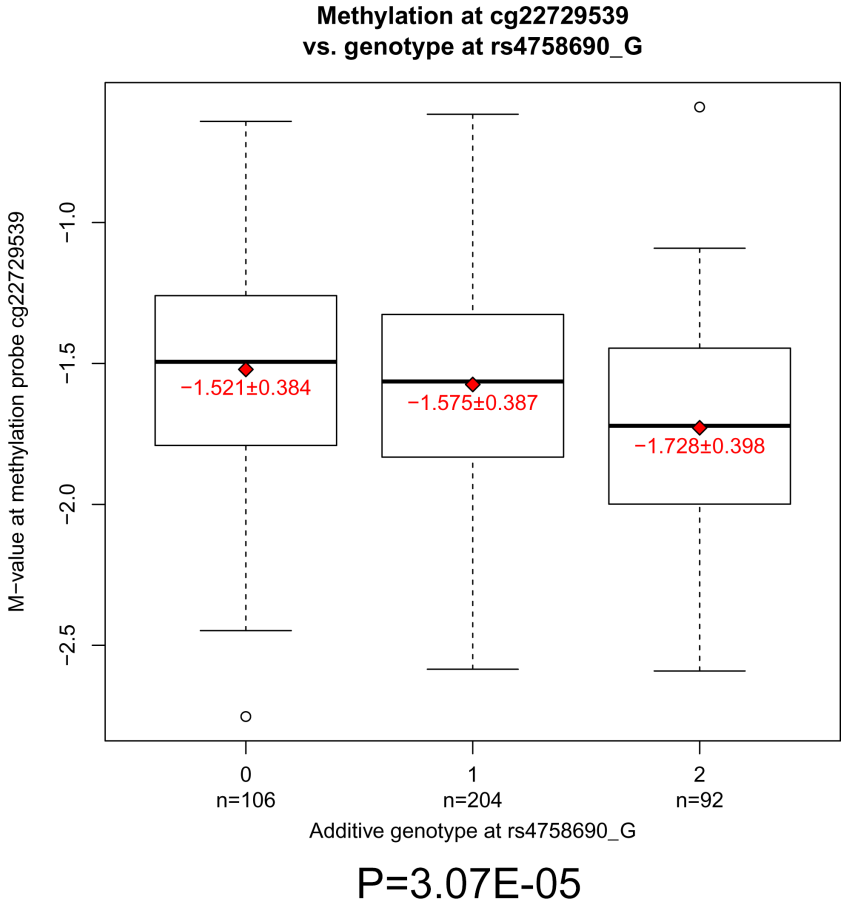

Figure 3 The association between SNP rs4758690 genotype and methylation status in gene *MLXIP*. *M*-values for methylation probe cg22729539 are plotted against the additive genotype at SNP rs4758690. Dark horizontal lines in the boxplots indicate the median *M*-value of each genotype group, the boxes represent the first to third quartiles, and the ends of whiskers of the boxplot show 1.5 times the interquartile range (IQR). Open circles represent data points outside of the range of 1.5 IQR. Red diamonds indicate the means of each genotype group, with the values of the mean ± standard deviation shown in red text. The number of individuals in each group with additive genotype of minor allele G is shown below the *X*-axis.

Replication of previously reported T2D loci

Previous genetic and genomic studies of T2D have yielded fruitful results. Based on literature review and a search of the NHGRI-EBI GWAS catalog (*Welter et al., 2014*), we generated a list of 182 genes which have been reported to be associated with T2D. Among them, 86 have intragenic SNPs or nearby SNPs that are nominally significant in our meta-analysis of Jordan subpopulations (Table S4), demonstrating the validity of our study even with a small sample size and support for common genetic basis of T2D in different ethnicities.

DISCUSSION

In this first GWAS of T2D in Jordan subpopulations, we identified a novel genome-wide significant locus at chr20p12.2 close to gene *JAG1* and replicated the association in the

samples of European ancestry of the WTCCC dataset. JAG1 is expressed in T2D relevant tissues and knockout of *JAG1* resulted in T2D related phenotypes in mice. We also found an interesting locus of suggestive significance at 12q24.31 in the intron of *MLXIP*. We further showed there is strong eQTL effect of the top associated SNP at this locus with correlation between its genotype and methylation of *MLXIP*, suggesting this locus may confer a cis-regulatory effect on *MLXIP* expression and this effect is at least in part mediated through methylation.

JAG1 encodes a ligand for receptor Notch 1, functioning in the Notch signaling pathway which is important for multiple cellular functions, especially during normal development and pathogenesis of cancer (*Bray, 2016*). Accumulative evidence demonstrate a critical role of the Notch signaling pathway in the regulation of metabolism and that perturbations in Notch signaling may lead to the development of obesity and T2D. It has been shown that overactivation of Notch signaling results in stimulation of glycogenolysis and gluconeogenesis in the liver, counteracting insulin effects (*Bi & Kuang, 2015*; *Pajvani et al., 2013*; *Pajvani et al., 2011*). Another role of Notch signaling in diabetes mellitus is to increase lipogenesis via mechanistic target of rapamycin complex 1, resulting in the development of hyperglycemia and fatty liver (*Bi & Kuang, 2015*; *Pajvani et al., 2013*), dysfunctions associated with T2D. Positive correlation of Notch signaling with insulin resistance and fatty liver has been reported in humans (*Valenti et al., 2013*). Key roles of Notch signaling also include regulation of adipocyte homeostasis and skeletal muscle homeostasis (*Bi & Kuang, 2015*). One upstream regulator of JAG1, HMGA1 is also involved in the molecular mechanism of T2D (*Bianco et al., 2015*). It has been reported that the expression of *JAG1* is down-regulated upon HMGA1 depletion by siRNA (*Pegoraro et al., 2013*). *HMGA1* encodes a non-histone chromatin associated protein, involved in multiple important cellular functions underlying pathogenesis of T2D, such as insulin production (*Arcidiacono et al., 2014*), in insulin action (*Iiritano et al., 2012*).

MLXIP encodes MondoA which interacts with MLX. Together they activate transcription of genes involved in glucose metabolism (*Sloan & Ayer, 2010*). Recent studies demonstrate that in addition to regulation of glucose-sensing transcription, *MLXIP* plays an important role in Myc activation and subsequent metabolic pathway reprogramming (*Carroll et al., 2015*). It is well known that Myc has important functions in the pathogenesis of diabetes, through both regulating cell cycle entry and maintaining expansion, regeneration and normal function of beta-cells (*Tiwari et al., 2016*). It has been shown that abnormal activation of Myc resulted in decreased beta-cell differentiation, proliferation and reduced insulin secretion (*Cheung et al., 2010*). On the other hand, insufficient Myc expression leads to hyperglycemia and beta-cell inactivity (*Guo et al., 2013*).

The pathological events that can lead to the development of T2D are diverse, such as deficiency and malfunction of beta-cells together with insulin resistance in multiple tissues, including liver and adipose tissues (*Tiwari et al., 2016*). The likely underlying genes for the novel T2D signals that we identified through GWAS are key players of signaling pathways that could lead to the development of T2D.

It is interesting that in our study, we observed a positive correlation between methylation and *MLXIP* expression that was associated with the rs4758690 SNP. While methylation

at promoter sites usually results in gene silencing, methylation at other gene sites often enhances gene expression (*Yang et al., 2014*) or affects splicing (*Jones, 2012*). The presence of histone modification marks and transcription factor binding in the vicinity of methylation probe cg22729539 suggests that this region contains cis-regulatory elements that actively regulate transcription. These epigenetic factors, like DNA methylation and histone modification, may interact with each other to influence gene expression in either the same or opposite directions (*Banovich et al., 2014*; *Cedar & Bergman, 2009*). DNA methylation could also affect nearby transcription factor binding, such as transcription factor CEBPB, which plays an important role in adipogenesis (*Darlington, Ross & MacDougald, 1998*), ER stress and pancreatic β cell failure (*Matsuda et al., 2010*). The coordination between a variety of genetic and epigenetic factors may regulate the expression of *MLXIP*, and further the development of T2D.

The two SNPs, rs6134031 and rs4758690 have been reported to be associated with other human traits, though genome-wide significance was not reached in those studies. In the NHGRI-EBI GWAS catalog (*Welter et al., 2014*), SNP rs6134031 has been reported to be associated with Plasma omega-6 polyunsaturated fatty acid levels (linoleic acid, n-6 PUFAs) (rs6134031-T, beta = 0.0372, P-value $= 4 \times 10^{-6}$) (*Dorajoo et al., 2015*). The relationship between n-6 PUFAs and T2D is debatable. Generally, n-6 PUFAs are considered to be proinflammatory and n-3 PUFAs to be anti-inflammatory. Thus, high dietary intake of n-6 PUFAs and elevated (n-6) to (n-3) ratio are associated with chronic inflammatory diseases including T2D (*Patterson et al., 2012*; *Simopoulos, 2016*). However, a recent study by *Forouhi et al. (2016)* in a large number of European subjects found that different types of n-6 PAFUs are differentially associated with risk of T2D. Linoleic acid (LA) and eicosadienoic acid (EDA) were inversely associated with T2D (OR < 1), arachidonic acid (AA) was not significantly associated, and γ-linolenic acid (GLA), dihomo-GLA, docosatetraenoic acid (DTA), docosapentaenoic acid (n6-DPA) are positively associated (OR $>$ 1). Thus the relationship between n-6 PUFAs (and its subtypes) and T2D needs to be further evaluated in more studies. SNP rs4758690 is also associated with height (P-value $= 2.396 \times 10^{-5}$), however the effect size and direction of effect are not available (*Lango Allen et al., 2010*). A systematic review and meta-analysis of 18 studies revealed that significant inverse association between height and T2D risk was only observed in women, not men (*Janghorbani, Momeni & Dehghani, 2012*). Thus the genotype of these 2 SNPs are important for inter-related human traits, suggesting these traits share common molecular underpinnings.

Our study has started to reveal the similarities and differences of the genetic basis of T2D between Jordan subpopulations and other ethnicities. Despite the small sample size, we were able to replicate almost half of the loci that were reported to be associated with T2D in genetic and genomic studies in other populations. The replication of these associations suggests some common genetic basis underlying the development of T2D among different ethnicities. For complex traits and diseases, there are many GWAS loci which could not be replicated across different ethnicities, such as the SNP rs7756992 in the *CDKAL1* gene which strongly associates with T2D in subjects of European ancestry, but displayed no association in a population of West Africa (*Steinthorsdottir et al., 2007*). Among the 37

SNPs associated with T2D in European or Asian populations, only two were replicated in a Qatari population (*O'Beirne et al., 2016*). In the Jordan subpopulations examined, we observed a significant association of rs6134031 and T2D, with a very large effect size. In the WTCCC, including only subjects of European ancestry, the LD structure for this region is different and the association of rs6134031 with T2D is less strong. The association at SNP rs4758690 is nominally significant in both Jordan subpopulations, however it is not significant in WTCCC subjects of European ancestry. The identification of these two loci suggested unique genetic determinants for T2D in the Jordan subpopulations. The separate GWAS performed in Chechen and Circassian subpopulations also suggest distinct genetic factors for T2D in each of these two ethnicities. As reviewed by *Rosenberg et al. (2010)*, such ethnic population differences may arise from variations in disease allele frequency, effect direction, effect size, distinct LD patterns, and trait/disease phenotype prevalence. Therefore, it is important to carry out genetic studies in different ethnic groups.

A major limitation of our study is the small size, which reduces the statistical power to detect a true effect of the genetic variants. The small sample size may lead to *p*-values of true associations failing to reach stringent significance thresholds, like the genome-wide significance threshold of 5×10^{-8}, resulting in false negatives (type II error). Therefore, we also considered other biological evidence when interpreting our results and we were encouraged by the replication of the *JAG1* locus and the strong eQTL signal observed for *MLXIP*, due to their strong biological relevance to T2D. As reported and discussed by other studies, true association may not always reach the conventionally corrected conservative threshold of 5×10^{-8} for declaring a genome-wide significance (*Nishizawa et al., 2014*). In our case, future studies with larger sample sizes of Jordan populations are needed to replicate the findings from our study and to further identify other genetic loci.

CONCLUSION

Taken together, our results from the first GWAS of T2D conducted in two subpopulations in Jordan have identified novel genetic factors underlying T2D; we additionally demonstrate there is common genetic basis among the different ethnicities as well as certain unique genetic factors that underlie T2D in the Jordan subpopulations. Identification of these novel genetic risk factors will offer the potential to gain further insight into the development of T2D and may help with the development of novel treatments precisely for the Jordan populations, which will reduce disease burden and promote health.

ACKNOWLEDGEMENTS

We would like to thank the Circassian and Chechen communities for their cooperation in conducting this study.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the King Hussein Center for Cancer and Biotechnology and the Higher Council for Science and Technology, Amman, Jordan, and by an Institutional Development Award from The Children's Hospital of Philadelphia to the Center for Applied Genomics. There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
King Hussein Center for Cancer and Biotechnology.
Higher Council for Science and Technology, Amman, Jordan.
The Children's Hospital of Philadelphia to the Center for Applied Genomics.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Rana Dajani conceived and designed the experiments, contributed reagents/materials/-analysis tools, wrote the paper, reviewed drafts of the paper.
- Jin Li, Zhi Wei and Michael E. March performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Qianghua Xia analyzed the data, prepared figures and/or tables, reviewed drafts of the paper.
- Yousef Khader, Nancy Hakooz, Raja Fatahallah, Mohammed El-Khateeb, Ala Arafat, Tareq Saleh, Abdel Rahman Dajani, Zaid Al-Abbadi, Mohamed Abdul Qader, Abdel Halim Shiyab, Anwar Bateiha and Kamel Ajlouni contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Hakon Hakonarson conceived and designed the experiments, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper.

Human Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The study has been approved by the Institutional Review Board committee at the National Center for Diabetes, Endocrinology and Genetics of Jordan. The written informed consent was given by all participants.

Data Availability

The following information was supplied regarding data availability:

The raw data has been uploaded as a Supplementary File.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.3618#supplemental-information.

REFERENCES

Ajlouni K, Khader YS, Batieha A, Ajlouni H, El-Khateeb M. 2008. An increase in prevalence of diabetes mellitus in Jordan over 10 years. *Journal of Diabetes and its Complications* **22**:317–324 DOI 10.1016/j.jdiacomp.2007.01.004.

Arcidiacono B, Iiritano S, Chiefari E, Brunetti FS, Gu G, Foti DP, Brunetti A. 2014. Cooperation between HMGA1, PDX-1, and MafA is Essential for Glucose-Induced Insulin Transcription in Pancreatic Beta Cells. *Frontiers in Endocrinology* **5**:Article 237 DOI 10.3389/fendo.2014.00237.

Banovich NE, Lan X, McVicker G, Van de Geijn B, Degner JF, Blischak JD, Roux J, Pritchard JK, Gilad Y. 2014. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLOS Genetics* **10**:e1004663 DOI 10.1371/journal.pgen.1004663.

Barbujani G, Nasidze IS, Whitehead GN. 1994. Genetic diversity in the Caucasus. *Human Biology* **66**:639–668.

Barbujani G, Whitehead GN, Bertorelle G, Nasidze IS. 1994. Testing hypotheses on processes of genetic and linguistic change in the Caucasus. *Human Biology* **66**:843–864.

Barrett JC, Cardon LR. 2006. Evaluating coverage of genome-wide association studies. *Nature Genetics* **38**:659–662 DOI 10.1038/ng1801.

Bhandare R, Schug J, Le Lay J, Fox A, Smirnova O, Liu C, Naji A, Kaestner KH. 2010. Genome-wide analysis of histone modifications in human pancreatic islets. *Genome Research* **20**:428–433 DOI 10.1101/gr.102038.109.

Bi P, Kuang S. 2015. Notch signaling as a novel regulator of metabolism. *Trends in Endocrinology and Metabolism* **26**:248–255 DOI 10.1016/j.tem.2015.02.006.

Bianco A, Chiefari E, Nobile CG, Foti D, Pavia M, Brunetti A. 2015. The association between HMGA1 rs146052672 aariant and type 2 diabetes: a transethnic meta-analysis. *PLOS ONE* **10**:e0136077 DOI 10.1371/journal.pone.0136077.

Blake JA, Eppig JT, Kadin JA, Richardson JE, Smith CL, Bult CJ, Mouse Genome Database Group. 2017. Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. *Nucleic Acids Research* **45**:D723–D729 DOI 10.1093/nar/gkw1040.

Bray SJ. 2016. Notch signalling in context. *Nature Reviews. Molecular Cell Biology* **17**:722–735 DOI 10.1038/nrm.2016.94.

Bulayeva KB. 2006. Overview of genetic-epidemiological studies in ethnically and demographically diverse isolates of Dagestan, Northern Caucasus, Russia. *Croatian Medical Journal* **47**:641–648.

Carroll PA, Diolaiti D, McFerrin L, Gu H, Djukovic D, Du J, Cheng PF, Anderson S, Ulrich M, Hurley JB, Raftery D, Ayer DE, Eisenman RN. 2015. Deregulated Myc

requires MondoA/Mlx for metabolic reprogramming and tumorigenesis. *Cancer Cell* **27**:271–285 DOI 10.1016/j.ccell.2014.11.024.

Cedar H, Bergman Y. 2009. Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews Genetics* **10**:295–304 DOI 10.1038/nrg2540.

Cheung L, Zervou S, Mattsson G, Abouna S, Zhou L, Ifandi V, Pelengaris S, Khan M. 2010. c-Myc directly induces both impaired insulin secretion and loss of beta-cell mass, independently of hyperglycemia *in vivo*. *Islets* **2**:37–45 DOI 10.4161/isl.2.1.10196.

Crawford SO, Hoogeveen RC, Brancati FL, Astor BC, Ballantyne CM, Schmidt MI, Young JH. 2010. Association of blood lactate with type 2 diabetes: the Atherosclerosis Risk in Communities Carotid MRI Study. *International Journal of Epidemiology* **39**:1647–1655 DOI 10.1093/ije/dyq126.

Dajani R, Khader YS, Fatahallah R, El-Khateeb M, Shiyab AH, Hakooz N. 2012. Diabetes mellitus in genetically isolated populations in Jordan: prevalence, awareness, glycemic control, and associated factors. *Journal of Diabetes and its Complications* **26**:175–180 DOI 10.1016/j.jdiacomp.2012.03.009.

Darlington GJ, Ross SE, MacDougald OA. 1998. The role of C/EBP genes in adipocyte differentiation. *Journal of Biological Chemistry* **273**:30057–30060 DOI 10.1074/jbc.273.46.30057.

Delaneau O, Marchini J, Zagury JF. 2012. A linear complexity phasing method for thousands of genomes. *Nature Methods* **9**:179–181 DOI 10.1038/nmeth.1785.

Delaneau O, Zagury JF, Marchini J. 2013. Improved whole-chromosome phasing for disease and population genetic studies. *Nature Methods* **10**:5–6 DOI 10.1038/nmeth.2307.

Dorajoo R, Sun Y, Han Y, Ke T, Burger A, Chang X, Low HQ, Guan W, Lemaitre RN, Khor CC, Yuan JM, Koh WP, Ong CN, Tai ES, Liu J, Van Dam RM, Heng CK, Friedlander Y. 2015. A genome-wide association study of n-3 and n-6 plasma fatty acids in a Singaporean Chinese population. *Genes & Nutrition* **10**:Article 53 DOI 10.1007/s12263-015-0502-2.

Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM. 2010. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**:587 DOI 10.1186/1471-2105-11-587.

Encode Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**:57–74 DOI 10.1038/nature11247.

Finger JH, Smith CM, Hayamizu TF, McCright IJ, Xu J, Law M, Shaw DR, Baldarelli RM, Beal JS, Blodgett O, Campbell JW, Corbani LE, Lewis JR, Forthofer KL, Frost PJ, Giannatto SC, Hutchins LN, Miers DB, Motenko H, Stone KR, Eppig JT, Kadin JA, Richardson JE, Ringwald M. 2017. The mouse Gene Expression Database (GXD): 2017 update. *Nucleic Acids Research* **45**:D730–D736 DOI 10.1093/nar/gkw1073.

Forouhi NG, Imamura F, Sharp SJ, Koulman A, Schulze MB, Zheng J, Ye Z, Sluijs I, Guevara M, Huerta JM, Kroger J, Wang LY, Summerhill K, Griffin JL, Feskens EJ, Affret A, Amiano P, Boeing H, Dow C, Fagherazzi G, Franks PW, Gonzalez C,

Kaaks R, Key TJ, Khaw KT, Kuhn T, Mortensen LM, Nilsson PM, Overvad K, Pala V, Palli D, Panico S, Quiros JR, Rodriguez-Barranco M, Rolandsson O, Sacerdote C, Scalbert A, Slimani N, Spijkerman AM, Tjonneland A, Tormo MJ, Tumino R, Van der AD, Van der Schouw YT, Langenberg C, Riboli E, Wareham NJ. 2016. Association of Plasma Phospholipid n-3 and n-6 polyunsaturated fatty acids with type 2 diabetes: the EPIC-interact case-cohort study. *PLOS Medicine* **13**:e1002094 DOI 10.1371/journal.pmed.1002094.

Ghassibe-Sabbagh M, Haber M, Salloum AK, Al-Sarraj Y, Akle Y, Hirbli K, Romanos J, Mouzaya F, Gauguier D, Platt DE, El-Shanti H, Zalloua PA. 2014. T2DM GWAS in the Lebanese population confirms the role of TCF7L2 and CDKAL1 in disease susceptibility. *Scientific Reports* **4**:7351 DOI 10.1038/srep07351.

Glans F, Elgzyri T, Shaat N, Lindholm E, Apelqvist J, Groop L. 2008. Immigrants from the Middle-East have a different form of Type 2 diabetes compared with Swedish patients. *Diabetic Medicine* **25**:303–307 DOI 10.1111/j.1464-5491.2007.02366.x.

GTEx Consortium. 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**:648–660 DOI 10.1126/science.1262110.

Guo S, Dai C, Guo M, Taylor B, Harmon JS, Sander M, Robertson RP, Powers AC, Stein R. 2013. Inactivation of specific beta cell transcription factors in type 2 diabetes. *Journal of Clinical Investigation* **123**:3305–3316 DOI 10.1172/JCI65390.

Hirschhorn JN, Daly MJ. 2005. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* **6**:95–108.

Howie BN, Donnelly P, Marchini J. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLOS Genetics* **5**:e1000529 DOI 10.1371/journal.pgen.1000529.

Ichikawa K, Yoshinari M, Iwase M, Wakisaka M, Doi Y, Iino K, Yamamoto M, Fujishima M. 1998. Advanced glycosylation end products induced tissue factor expression in human monocyte-like U937 cells and increased tissue factor expression in monocytes from diabetic patients. *Atherosclerosis* **136**:281–287 DOI 10.1016/S0021-9150(97)00221-9.

Iiritano S, Chiefari E, Ventura V, Arcidiacono B, Possidente K, Nocera A, Nevolo MT, Fedele M, Greco A, Greco M, Brunetti G, Fusco A, Foti D, Brunetti A. 2012. The HMGA1-IGF-I/IGFBP system: a novel pathway for modulating glucose uptake. *Molecular Endocrinology* **26**:1578–1589 DOI 10.1210/me.2011-1379.

Imamura M, Chang BH, Kohjima M, Li M, Hwang B, Taegtmeyer H, Harris RA, Chan L. 2014. MondoA deficiency enhances sprint performance in mice. *Biochemical Journal* **464**:35–48 DOI 10.1042/BJ20140530.

Imamura M, Takahashi A, Yamauchi T, Hara K, Yasuda K, Grarup N, Zhao W, Wang X, Huerta-Chagoya A, Hu C, Moon S, Long J, Kwak SH, Rasheed A, Saxena R, Ma RC, Okada Y, Iwata M, Hosoe J, Shojima N, Iwasaki M, Fujita H, Suzuki K, Danesh J, Jorgensen T, Jorgensen ME, Witte DR, Brandslund I, Christensen C, Hansen T, Mercader JM, Flannick J, Moreno-Macias H, Burtt NP, Zhang R, Kim YJ, Zheng W, Singh JR, Tam CH, Hirose H, Maegawa H, Ito C, Kaku K, Watada H, Tanaka

Y, Tobe K, Kawamori R, Kubo M, Cho YS, Chan JC, Sanghera D, Frossard P, Park KS, Shu XO, Kim BJ, Florez JC, Tusie-Luna T, Jia W, Tai ES, Pedersen O, Saleheen D, Maeda S, Kadowaki T. 2016. Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. *Nature Communications* **7**:Article 10531 DOI 10.1038/ncomms10531.

Janghorbani M, Momeni F, Dehghani M. 2012. Hip circumference, height and risk of type 2 diabetes: systematic review and meta-analysis. *Obesity Reviews* **13**:1172–1181 DOI 10.1111/j.1467-789X.2012.01030.x.

Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* **13**:484–492 DOI 10.1038/nrg3230.

Kabakchiev B, Silverberg MS. 2013. Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. *Gastroenterology* **144**:1488–1496 DOI 10.1053/j.gastro.2013.03.001.

Kailani W. 2002. Chechens in the Middle East: between original and host cultures. Caspian studies program. *Available at http://belfercenter.ksg.harvard.edu/publication/12785/chechens_in_the_middle_east.html* (accessed on 3 August 2017).

Karpe F, Dickmann JR, Frayn KN. 2011. Fatty acids, obesity, and insulin resistance: time for a reevaluation. *Diabetes* **60**:2441–2449 DOI 10.2337/db11-0425.

Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Ferreira T, Wood AR, Weyant RJ, Segre AV, Speliotes EK, Wheeler E, Soranzo N, Park JH, Yang J, Gudbjartsson D, Heard-Costa NL, Randall JC, Qi L, Vernon Smith A, Magi R, Pastinen T, Liang L, Heid IM, Luan J, Thorleifsson G, Winkler TW, Goddard ME, Sin Lo K, Palmer C, Workalemahu T, Aulchenko YS, Johansson A, Zillikens MC, Feitosa MF, Esko T, Johnson T, Ketkar S, Kraft P, Mangino M, Prokopenko I, Absher D, Albrecht E, Ernst F, Glazer NL, Hayward C, Hottenga JJ, Jacobs KB, Knowles JW, Kutalik Z, Monda KL, Polasek O, Preuss M, Rayner NW, Robertson NR, Steinthorsdottir V, Tyrer JP, Voight BF, Wiklund F, Xu J, Zhao JH, Nyholt DR, Pellikka N, Perola M, Perry JR, Surakka I, Tammesoo ML, Altmaier EL, Amin N, Aspelund T, Bhangale T, Boucher G, Chasman DI, Chen C, Coin L, Cooper MN, Dixon AL, Gibson Q, Grundberg E, Hao K, Juhani Junttila M, Kaplan LM, Kettunen J, Konig IR, Kwan T, Lawrence RW, Levinson DF, Lorentzon M, McKnight B, Morris AP, Muller M, Suh Ngwa J, Purcell S, Rafelt S, Salem RM, Salvi E, Sanna S, Shi J, Sovio U, Thompson JR, Turchin MC, Vandenput L, Verlaan DJ, Vitart V, White CC, Ziegler A, Almgren P, Balmforth AJ, Campbell H, Citterio L, De Grandi, A, Dominiczak A, Duan J, Elliott P, Elosua R, Eriksson JG, Freimer NB, Geus EJ, Glorioso N, Haiqing S, Hartikainen AL, Havulinna AS, Hicks AA, Hui J, Igl W, Illig T, Jula A, Kajantie E, Kilpelainen TO, Koiranen M, Kolcic I, Koskinen S, Kovacs P, Laitinen J, Liu J, Lokki ML, Marusic A, Maschio A, Meitinger T, Mulas A, Pare G, Parker AN, Peden JF, Petersmann A, Pichler I, Pietilainen KH, Pouta A, Ridderstrale M, Rotter JI, Sambrook JG, Sanders AR, Schmidt CO, Sinisalo J, Smit JH, Stringham HM, Bragi Walters G, Widen E, Wild SH, Willemsen G, Zagato L, Zgaga L, Zitting P, Alavere H, Farrall M, McArdle WL, Nelis M, Peters MJ, Ripatti S, Van Meurs JB, Aben KK,

Ardlie KG, Beckmann JS, Beilby JP, Bergman RN, Bergmann S, Collins FS, Cusi D, Den Heijer M, Eiriksdottir G, Gejman PV, Hall AS, Hamsten A, Huikuri HV, Iribarren C, Kahonen M, Kaprio J, Kathiresan S, Kiemeney L, Kocher T, Launer LJ, Lehtimaki T, Melander O, Mosley Jr TH, Musk AW, Nieminen MS, O'Donnell CJ, et al. 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**:832–838 DOI 10.1038/nature09410.

Lin SM, Du P, Huber W, Kibbe WA. 2008. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Research* **36**:e11 DOI 10.1093/nar/gkm1075.

Marchini J, Howie B, Myers S, McVean G, Donnelly P. 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genetics* **39**:906–913 DOI 10.1038/ng2088.

Matsuda T, Kido Y, Asahara S, Kaisho T, Tanaka T, Hashimoto N, Shigeyama Y, Takeda A, Inoue T, Shibutani Y, Koyanagi M, Hosooka T, Matsumoto M, Inoue H, Uchida T, Koike M, Uchiyama Y, Akira S, Kasuga M. 2010. Ablation of C/EBPbeta alleviates ER stress and pancreatic beta cell failure through the GRP78 chaperone in mice. *Journal of Clinical Investigation* **120**:115–126 DOI 10.1172/JCI39721.

Mtiraoui N, Turki A, Nemr R, Echtay A, Izzidi I, Al-Zaben GS, Irani-Hakime N, Keleshian SH, Mahjoub T, Almawi WY. 2012. Contribution of common variants of ENPP1, IGF2BP2, KCNJ11, MLXIPL, PPARgamma, SLC30A8 and TCF7L2 to the risk of type 2 diabetes in Lebanese and Tunisian Arabs. *Diabetes et Metabolisme* **38**:444–449 DOI 10.1016/j.diabet.2012.05.002.

Nasidze I, Ling EY, Quinque D, Dupanloup I, Cordaux R, Rychkov S, Naumova O, Zhukova O, Sarraf-Zadegan N, Naderi GA, Asgary S, Sardas S, Farhud DD, Sarkisian T, Asadov C, Kerimov A, Stoneking M. 2004. Mitochondrial DNA and Y-chromosome variation in the caucasus. *Annals of Human Genetics* **68**:205–221 DOI 10.1046/j.1529-8817.2004.00092.x.

Nasidze I, Risch GM, Robichaux M, Sherry ST, Batzer MA, Stoneking M. 2001. Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. *European Journal of Human Genetics* **9**:267–272 DOI 10.1038/sj.ejhg.5200615.

Nishizawa D, Fukuda K, Kasai S, Hasegawa J, Aoki Y, Nishi A, Saita N, Koukita Y, Nagashima M, Katoh R, Satoh Y, Tagami M, Higuchi S, Ujike H, Ozaki N, Inada T, Iwata N, Sora I, Iyo M, Kondo N, Won MJ, Naruse N, Uehara-Aoyama K, Itokawa M, Koga M, Arinami T, Kaneko Y, Hayashida M, Ikeda K. 2014. Genome-wide association study identifies a potent locus associated with human opioid sensitivity. *Molecular Psychiatry* **19**:55–62 DOI 10.1038/mp.2012.164.

O'Beirne SL, Salit J, Rodriguez-Flores JL, Staudt MR, Abi Khalil C, Fakhro KA, Robay A, Ramstetter MD, Al-Azwani IK, Malek JA, Zirie M, Jayyousi A, Badii R, Al-Nabet Al-Marri A, Chiuchiolo MJ, Al-Shakaki A, Chidiac O, Gharbiah M, Bener A, Stadler D, Hackett NR, Mezey JG, Crystal RG. 2016. Type 2 diabetes risk allele loci in the qatari population. *PLOS ONE* **11**:e0156834 DOI 10.1371/journal.pone.0156834.

Pajvani UB, Qiang L, Kangsamaksin T, Kitajewski J, Ginsberg HN, Accili D. 2013. Inhibition of Notch uncouples Akt activation from hepatic lipid accumulation by decreasing mTorc1 stability. *Nature Medicine* **19**:1054–1060 DOI 10.1038/nm.3259.

Pajvani UB, Shawber CJ, Samuel VT, Birkenfeld AL, Shulman GI, Kitajewski J, Accili D. 2011. Inhibition of Notch signaling ameliorates insulin resistance in a FoxO1-dependent manner. *Nature Medicine* **17**:961–967 DOI 10.1038/nm.2378.

Palmer ND, McDonough CW, Hicks PJ, Roh BH, Wing MR, An SS, Hester JM, Cooke JN, Bostrom MA, Rudock ME, Talbert ME, Lewis JP, DIAGRAM Consortium, MAGIC Investigators, Ferrara A, Lu L, Ziegler JT, Sale MM, Divers J, Shriner D, Adeyemo A, Rotimi CN, Ng MC, Langefeld CD, Freedman BI, Bowden DW. 2012. A genome-wide association search for type 2 diabetes genes in African Americans. *PLOS ONE* **7**:e29202 DOI 10.1371/journal.pone.0029202.

Parker SC, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, Van Bueren KL, Chines PS, Narisu N, NISC Comparative Sequencing Program, Black BL, Visel A, Pennacchio LA, Collins FS. 2013. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proceedings of the National Academy of Sciences of the United States of America* **110**:17921–17926 DOI 10.1073/pnas.1317023110.

Pasquali L, Gaulton KJ, Rodriguez-Segui SA, Mularoni L, Miguel-Escalada I, Akerman I, Tena JJ, Moran I, Gomez-Marin C, Van de Bunt M, Ponsa-Cobas J, Castro N, Nammo T, Cebola I, Garcia-Hurtado J, Maestro MA, Pattou F, Piemonti L, Berney T, Gloyn AL, Ravassard P, Gomez-Skarmeta JL, Muller F, McCarthy MI, Ferrer J. 2014. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nature Genetics* **46**:136–143 DOI 10.1038/ng.2870.

Patterson E, Wall R, Fitzgerald GF, Ross RP, Stanton C. 2012. Health implications of high dietary omega-6 polyunsaturated Fatty acids. *Journal of Nutrition and Metabolism* **2012**:539426 DOI 10.1155/2012/539426.

Pe'er I, De Bakker PI, Maller J, Yelensky R, Altshuler D, Daly MJ. 2006. Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nature Genetics* **38**:663–667 DOI 10.1038/ng1816.

Pegoraro S, Ros G, Piazza S, Sommaggio R, Ciani Y, Rosato A, Sgarra R, Del Sal G, Manfioletti G. 2013. HMGA1 promotes metastatic processes in basal-like breast cancer regulating EMT and stemness. *Oncotarget* **4**:1293–1308 DOI 10.18632/oncotarget.1136.

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**:904–909 DOI 10.1038/ng1847.

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ. 2010. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**:2336–2337 DOI 10.1093/bioinformatics/btq419.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome

association and population-based linkage analyses. *American Journal of Human Genetics* **81**:559–575 DOI 10.1086/519795.

Roadmap Epigenomics C, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, Ward LD, Sarkar A, Quon G, Sandstrom RS, Eaton ML, Wu YC, Pfenning AR, Wang X, Claussnitzer M, Liu Y, Coarfa C, Harris RA, Shoresh N, Epstein CB, Gjoneska E, Leung D, Xie W, Hawkins RD, Lister R, Hong C, Gascard P, Mungall AJ, Moore R, Chuah E, Tam A, Canfield TK, Hansen RS, Kaul R, Sabo PJ, Bansal MS, Carles A, Dixon JR, Farh KH, Feizi S, Karlic R, Kim AR, Kulkarni A, Li D, Lowdon R, Elliott G, Mercer TR, Neph SJ, Onuchic V, Polak P, Rajagopal N, Ray P, Sallari RC, Siebenthall KT, Sinnott-Armstrong NA, Stevens M, Thurman RE, Wu J, Zhang B, Zhou X, Beaudet AE, Boyer LA, De Jager PL, Farnham PJ, Fisher SJ, Haussler D, Jones SJ, Li W, Marra MA, McManus MT, Sunyaev S, Thomson JA, Tlsty TD, Tsai LH, Wang W, Waterland RA, Zhang MQ, Chadwick LH, Bernstein BE, Costello JF, Ecker JR, Hirst M, Meissner A, Milosavljevic A, Ren B, Stamatoyannopoulos JA, Wang T, Kellis M. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**:317–330 DOI 10.1038/nature14248.

Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. 2010. Genome-wide association studies in diverse populations. *Nature Reviews Genetics* **11**:356–366 DOI 10.1038/nrg2760.

Simopoulos AP. 2016. An increase in the Omega-6/Omega-3 fatty acid ratio increases the risk for obesity. *Nutrients* **8**:Article 128 DOI 10.3390/nu8030128.

Sloan EJ, Ayer DE. 2010. Myc, mondo, and metabolism. *Genes Cancer* **1**:587–596 DOI 10.1177/1947601910377489.

Steinthorsdottir V, Thorleifsson G, Reynisdottir I, Benediktsson R, Jonsdottir T, Walters GB, Styrkarsdottir U, Gretarsdottir S, Emilsson V, Ghosh S, Baker A, Snorradottir S, Bjarnason H, Ng MC, Hansen T, Bagger Y, Wilensky RL, Reilly MP, Adeyemo A, Chen Y, Zhou J, Gudnason V, Chen G, Huang H, Lashley K, Doumatey A, So WY, Ma RC, Andersen G, Borch-Johnsen K, Jorgensen T, Van Vliet-Ostaptchouk JV, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Rotimi C, Gurney M, Chan JC, Pedersen O, Sigurdsson G, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. 2007. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nature Genetics* **39**:770–775 DOI 10.1038/ng2043.

Tiwari S, Roel C, Tanwir M, Wills R, Perianayagam N, Wang P, Fiaschi-Taesch NM. 2016. Definition of a Skp2-c-Myc pathway to expand human beta-cells. *Scientific Reports* **6**:28461 DOI 10.1038/srep28461.

Valenti L, Mendoza RM, Rametta R, Maggioni M, Kitajewski C, Shawber CJ, Pajvani UB. 2013. Hepatic notch signaling correlates with insulin resistance and nonalcoholic fatty liver disease. *Diabetes* **62**:4052–4062 DOI 10.2337/db13-0769.

Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**:661–678 DOI 10.1038/nature05911.

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, Parkinson H. 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research* **42**:D1001–D1006 DOI 10.1093/nar/gkt1229.

Wild S, Roglic G, Green A, Sicree R, King H. 2004. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* **27**:1047–1053 DOI 10.2337/diacare.27.5.1047.

Yamada Y, Matsuo H, Segawa T, Watanabe S, Kato K, Kameyama T, Yokoi K, Ichihara S, Metoki N, Yoshida H, Satoh K, Nozawa Y. 2006. Assessment of genetic factors for type 2 diabetes mellitus. *International Journal of Molecular Medicine* **18**:299–308.

Yamaguchi S, Yamada Y, Matsuo H, Segawa T, Watanabe S, Kato K, Yokoi K, Ichihara S, Metoki N, Yoshida H, Satoh K, Nozawa Y. 2007. Gender differences in the association of gene polymorphisms with type 2 diabetes mellitus. *International Journal of Molecular Medicine* **19**:631–637.

Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. 2014. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* **26**:577–590 DOI 10.1016/j.ccr.2014.07.028.