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

Rana Dajani Corresp., [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 R Dajani** [1] , **Zaid Al-Abbadi** [1] , **Mohamed Abdul Qader** [1] , **Abdel H Shiyab** [9] , **Anwar Bateiha** [6] , **Kamel Ajlouni** [8] , **Hakon Hakonarson** Corresp. [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, Pennsylvania, United States

[3] Department of Cell Biology, Tianjin Medical University, Tianjin, China

[4] Department of Computer Science, New Jersey Institute of Technology, Newark, New Jersey, United States

[5] Divisions of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States

[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 Faculty 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

Corresponding Authors: Rana Dajani, Hakon Hakonarson
Email address: rdajani@hu.edu.jo, hakonarson@email.chop.edu

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, $P=1.12x10^{-8}$) 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, $P= 4.20x10^{-5}$) and it is a robust eQTL for the gene, *MLXIP* ($P=1.10x10^{-14}$), 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.

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[5,3], Yousef Khader[6], Nancy Hakooz[7], Raja Fatahallah[8], Mohammad El-Khateeb[8], Ala Arafat[8], Tarek Saleh[1], Abdel Rahman Dajani[1], Zaid Al-Abbadi[1], Mohamed Abdul Qader[1], Abdel Haleem Shiyab[9], Anwar Bateiha[6], Kamel Ajlouni[8], 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, Pennsylvania 19104, USA

[3] Department of Cell Biology, Tianjin Medical University, Tianjin, China

[4] Department of Computer Science, New Jersey Institute of Technology, New Jersey, USA

[5] Divisions of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, USA

[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 Faculty 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, Pennsylvania, 19104, USA.

* RD and JL contributed equally to this work.

Correspondence should be addressed to
Dr. Hakon Hakonarson
The Children's Hospital of Philadelphia and University of Pennsylvania
34th street and Civic Center Blvd., Philadelphia, PA, 19104
Email: hakonarson@email.chop.edu
267-426-0088 (phone); 267-426-0363 (fax)
and
Rana Dajani
Associate Professor Molecular Cell Biology
Biology Department
Hashemite University
P. O. Box 150459
Zarqa 13115 Jordan
Tel : +962 (5) 3903333
Fax : +962 (5) 3826613
email : rdajani@hu.edu.jo
Cell phone: 00962798859335

Abstract

The prevalence of Type II Diabetes (T2D) has been increasing and has become a disease of significant public health burden in Jordan. The genetic determinants of T2D in Middle-East populations have not been well studied by genome-wide association studies (GWAS). 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, $P=1.12x10^{-8}$) and we replicated the results in the Wellcome Trust Case Control Consortium (WTCCC) dataset. Another locus at chr12q24.31 is associated with T2D at a suggestive significance level (top SNP rs4758690, $P= 4.20x10^{-5}$). This SNP is a robust eQTL for the gene, *MLXIP* ($P=1.10x10^{-14}$), 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.

73

74

75 **Introduction**

76 Diabetes is among the most common non-communicable diseases globally. It has been estimated
77 that there are currently about 194 million people at the age of 20 to 79 with diabetes worldwide
78 and that this number will further increase to 333 million by 2025 (Wild et al. 2004). Diabetes is
79 the fifth main cause of death in Jordan, afflicting 16 percent of Jordanian adult citizens; another
80 23.8 percent of adults in Jordan are also on the brink of becoming diabetics according to a study
81 from 2007 by the Heart and Capillary Disease Prevention directorate (HCDP) of the Ministry of
82 Health in Jordan; and the rate of diabetes prevalence in Jordan is 30.5 percent among both
83 children and adults (Ajlouni et al. 2008). Thus diabetes presents a significant public health
84 burden to the Jordan community. Type II Diabetes (T2D) is the major type of diabetes, which
85 accounts for 95% percent of all diabetes cases worldwide.

86

87 Despite extensive research efforts for more than a decade and some notable successes, much of
88 the genetic basis of common human diseases remains unresolved (Hirschhorn & Daly 2005). The
89 genome-wide association study (GWAS) has been a powerful approach for identifying novel
90 susceptibility loci for complex diseases (Barrett & Cardon 2006; Pe'er et al. 2006), such as T2D.
91 To date, more than 80 T2D susceptibility loci have been uncovered by GWAS. However, the
92 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 et al. 1994a; 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 et al. 1994a; Barbujani et al. 1994b), which has been further confirmed by mitochondrial DNA and Y chromosome analysis (Nasidze et al. 2004; Nasidze et

113 al. 2001). While a T2D GWAS has been conducted in the Lebanese population (Ghassibe-
114 Sabbagh et al. 2014), the Lebanese are Arab in origin; Circassians and Chechans are a separate,
115 non-Arab ethnic group. These are clearly different populations, with different ancestries. The
116 Circassian and Chechen communities may provide us an opportunity to study a genetically
117 unique population and compare genetic basis for complex human diseases between different
118 populations.
119
120 T2D has become an alarming public health issue in Jordan. Epidemiology studies showed that
121 the prevalence of impaired fasting glycemia is 18.5% and 14.6% and prevalence of diabetes is
122 9.6% and 10.1% for Circassians and Chechens respectively(Dajani et al. 2012). In view of the
123 very high incidence of T2D in Jordan and the genetic distinctness of Circassian and Chechan
124 populations, we performed a GWAS to search for genetic factors contributing to T2D in these
125 two populations and compared the results with European population.
126
127 **Materials & Methods**
128 *Ethics Statement*
129 The study has been approved by the institutional review board committee at the
130 National Center for Diabetes, Endocrinology and Genetics of Jordan (approval number:
131 457/9.MS). The written informed consent was given by all participants.
132

133 *Study subjects and sample collection*

134 A random sample of N = 144 from the Chechen population in Jordan and a random sample of N
135 = 140 from the Circassian population in Jordan were recruited to participate in the study. Each
136 participant in the study filled out a survey that included pedigree information. The identities of
137 parents, grandparents, and great-grandparents (both maternally and paternally) were reported in
138 the survey and any individual with non-Chechen heritage for even one person in his/her pedigree
139 was excluded for the Chechen subpopulation; the same identity confirmation was conducted for
140 the Circassian subpopulation.
141 A subject was defined as affected by diabetes mellitus if this diagnosis is known to the patient or,
142 according to the ADA definitions, if fasting serum glucose is 7mmol/L (126 mg/dl) or more.
143 Impaired fasting glucose was defined as a fasting serum glucose level of $\geq$6.1mmol/L (100
144 mg/dl) but < 7mmol/L. The glycemic control was assessed using HbA_{1c}. Patients with previously
145 diagnosed diabetes who had HbA_{1c} >7% were defined as having 'unsatisfactory' glycemic
146 control.
147
148 *Sample collection*
149 Nine ml of whole blood was drawn in EDTA tubes from the subjects by vacutainer system.
150 Genomic DNA was isolated from whole blood sample using the phenol-chloroform protocol.
151
152 *Genotyping and quality control*

153 We performed high-throughput, genome-wide SNP genotyping, using the InfiniumII OMNI-
154 Express BeadChip technology (Illumina), at the Center for Applied Genomics (CAG) at the
155 Children's Hospital of Philadelphia (CHOP), USA. Sample quality control (QC) was performed
156 based on the following measures: sample call rate, overall heterozygosity, relatedness testing and
157 other metrics. Samples were excluded from analysis for SNP call rate < 95%, heterozygosity
158 beyond five standard deviation of the mean. One sample from each pair of duplicated or cryptic
159 related samples was removed. For each pair of duplicate or related samples the sample with the
160 highest SNP call rate was kept in the dataset. In the SNP-based QC, SNPs with a call rate <95%,
161 minor allele frequency < 1% or showing significant deviation from Hardy-Weinberg-Equilibrium
162 (HWE test P-value < 10^{-4}) in the controls were removed. All QC steps were carried out using the
163 software package PLINK (Purcell et al. 2007)
164
165 *Principal component analysis (PCA)*
166 PCA was conducted to confirm ethnic identity and to generate covariates to control for
167 population stratification in the association analysis. LD-pruning was performed using PLINK,
168 and only independent (r^2<0.2), autosomal non-GC/AT SNPs were included in the PCA, which
169 was conducted using EIGENSTRAT (Price et al. 2006) version 3.0.
170
171 *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 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 et al. 2012; Delaneau et al. 2013) version 2, and then genotype imputation was performed using IMPUTE 2 (Howie et al. 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

192 assess the association of each imputed SNP genotype with T2D using software SNPTEST
193 (Marchini et al. 2007) V2, including the first three principal components as covariates. SNPs
194 with info score <0.8 or with HWE-test p-value < $1x10^{-06}$ were excluded from association testing.
195
196 *Analysis of methylation data*
197 Genomic DNA of a subset of samples in the biorepository of CAG was isolated from peripheral
198 blood mononuclear cells. Genome-wide methylation profiling was conducted on the Infinium
199 HumanMethylation450 BeadChip Kit at CAG according to the manufacturers' protocols.
200 Methylation data were exported from the Illumina GenomeStudio and loaded into the R
201 statistical package (r-project.org) using the lumi package (Du et al. 2010; Lin et al. 2008). After
202 adjusting for quantile color balance and background level and simple scaling normalization, M-
203 value density and CpG-site intensity were plotted and aberrant chips were removed. These
204 samples have also been genotyped at CAG and their genetic ethnicity was checked by PCA. We
205 extracted the M-values (the log2 ratio between the methylated and unmethylated probe
206 intensities) and the genotype information of the 425 subjects of European ancestry. We removed
207 subjects of missing genotype at SNP rs4785690 and extreme outlier values of methylation M-
208 values (≥median M-value of the genotype group±3SD) and then assessed the association between
209 the additive genotype at rs4785690 and methylation M-value in gene *MLXIP* using linear
210 regression including sex, age, and 10 genotype-derived principle components. Box-plots were
211 generated using R package.

212

213 **Results**

214 **Identification of novel T2D signals in Jordan subpopulations**. To understand the genetic basis
215 for T2D in Jordan populations, we conducted GWAS in Chechen and Circassian subpopulations
216 of Jordan. The sample information after QC is summarized in Table 1. Specifically, for the
217 Chechen subpopulation, we have 34 cases and 109 controls; for the Circassian subpopulation, we
218 have 33 cases and 105 controls (Table 1). Approximately 645,000 SNPs in each subpopulation
219 passed QC. We conducted logistic regression analyses separately in each population, including
220 ten genotype-derived principal components as covariates. There was no signal that reached
221 genome-wide significance, however there are several SNPs at suggestive level of significance (P
222 $<1x10^{-4}$) in each subpopulation (Supplementary Tables 1-2). Then we performed meta-analysis
223 of the association results from the two subpopulations. In the meta-analysis, we observed a signal
224 at genome-wide significant level (SNP rs6134031, P-value= $1.12x10^{-8}$) under both fixed effect
225 model and random effect model (Supplementary Fig. 1, Fig. 1, Table 2). This SNP is located at
226 the 5' of the *JAG1* gene (Fig. 1). In addition, there is another signal with multiple SNPs showing
227 suggestive evidence of association (P-value $<1x10^{-4}$), with SNP rs4758690 having the lowest P-
228 value at $4.20x10^{-5}$ (Supplementary Fig. 1, Table 2, Fig. 1). SNP rs4758690 is located in the intron
229 of *MLXIP*, a gene involved in transcriptional regulation of genes in glucose metabolism. Taken
230 together, these results demonstrate significant GWAS associations to novel T2D susceptibility
231 loci in Jordan subpopulations.

232

233 **Test the association signals in European population.** We then investigated whether these
234 association signals exist in populations of other ethnicities. We examined the association of these
235 SNPs in the T2D dataset of the Wellcome Trust Case Control Consortium (WTCCC)(Wellcome
236 Trust Case Control 2007) which is composed of 1999 cases and 3004 controls, genotyped on the
237 Affymetrix Genome-Wide Human SNP Array 5.0. After QC, 1952 cases and 2960 controls of
238 European ancestry remained for association analysis by logistic regression. The top SNP in the
239 Jordan analysis, rs6134031 demonstrated nominally significant association with T2D in the
240 WTCCC cohort (P=0.012) and the same direction of effect (Table 2). The SNP rs4758690 is not
241 genotyped on the Affymetrix GW5.0 Array, so we conducted imputation over this region in the
242 replication cohort. Based on the imputed genotype data, we did not observe a significant
243 association to rs4758690 (OR= 1.01, P=0.61).

244

245 **Correlation of T2D variants with *MLXIP* gene expression and methylation**. Interrogating
246 these T2D variants in the GTEx dataset (GTEx Consortium 2015), we uncovered a nominally
247 significant association between SNP rs6134031 and *JAG1* expression, in Esophagus –
248 Muscularis (Beta=-0.15, P=0.0073, Supplementary Fig. 2) and a marginal correlation in
249 pancreatic tissue which is of potential biological relevance to T2D (Beta=-0.13, P=0.071,
250 Supplementary Fig. 2). Though it is not significant, we did observe a trend of association
251 between the doses of minor allele T and a lower expression of *JAG1*.

252 On the other hand, we found a genome-wide significant eQTL effect of SNP rs4758690 for gene
253 *MLXIP* expression in transverse colon (Beta=0.46, P= 1.10×10^{-14}) and small intestine terminal
254 ileum (Beta=0.42, P=4.20×10^{-7}) tissue specimens (Fig. 2). A similar significant eQTL effect was
255 reported for *MLXIP* expression in normal pre-pouch ileum in another study examining eQTLs in
256 human intestine tissues (Kabakchiev & Silverberg 2013).
257 Further, we found that SNP rs4758690 is significantly associated with the methylation probe
258 cg22729539 (P=3.07×10^{-5}) residing within an intron of the longest isoform of *MLXIP* (Fig. 3).
259 This site is absent in other short isoforms. We observed a positive correlation between the eQTL
260 and the methylation data at this locus. As methylation is one of the important mechanisms
261 regulating gene expression, these results are of potential interest. The minor allele G confers a
262 lower expression of *MLXIP* compared to the major allele A, as well as a reduced methylation
263 level at probe cg22729539, consistent with previous reports that gene body methylation was
264 found to be positively correlated with gene expression (Yang et al. 2014). In addition,
265 cg22729539 resides in a region with multiple histone modifications and transcription factor
266 binding in pancreatic islets and liver cells which are central to T2D (Supplementary Fig. 3) and
267 additional T2D relevant cell lines (Supplementary Table 3) (Bhandare et al. 2010; Encode
268 Project Consortium 2012; Parker et al. 2013; Pasquali et al. 2014; Roadmap Epigenomics et al.
269 2015). The bound transcription factors include CEBPB which is known to function in
270 adipogenesis (Darlington et al. 1998), ER stress and pancreatic β cell failure (Matsuda et al.

271 2010) (Supplementary Table 3), therefore this region may function as active cis-regulatory
272 element, regulating *MLXIP* expression.
273
274 **The expression of *JAG1* and *MLXIP***. The biological relevance of these two genes to T2D was
275 further strengthened by their expression pattern. For *JAG1*, it is reported to be highly expressed
276 in arteries and in bronchial epithelial cells and lung tissue, with a particularly high level of
277 expression in the gastrointestinal tract tissues, such as small intestine and colon (Supplementary
278 Figs. 4 and 5). For the gene *MLXIP*, high levels of expression have been consistently noticed in
279 colon tissue as reported in different studies (Supplementary Figs. 6 and 7). Both of these genes
280 demonstrated medium level of expression in certain tissues highly relevant to T2D, including
281 *JAG1* in adipose, pancreas, and smooth muscle (Supplementary Figs. 4 and 5), and *MLXIP* in
282 muscle, pancreas and pancreatic islet cells (Supplementary Figs. 6 and 7).
283 The overall expression pattern of *JAG1* is similar to that of the gene Coagulation Factor III (*F3*)
284 (correlation >0.7), genetic polymorphisms of which have been shown to be associated with T2D
285 in different ethnicity groups (Palmer et al. 2012; Yamada et al. 2006; Yamaguchi et al. 2007) and
286 the expression of which is significantly higher in monocytes and neutrophils of diabetes and
287 prediabetic subjects (Ichikawa et al. 1998).
288 Consistent with the expression pattern, knockout of *JAG1* in a mouse model resulted in defects in
289 endocrine/exocrine glands, homeostasis/metabolism, and the liver/biliary system (Supplementary
290 Fig. 8) (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 et al. 2011).

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 (Supplementary Table 4), 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 over activation 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.

350 It is interesting that in our study, we observed a positive correlation between methylation and
351 *MLXIP* expression that was associated with the rs4758690 SNP. While methylation at promoter
352 sites usually results in gene silencing, methylation at other gene sites often enhances gene
353 expression (Yang et al. 2014) or affects splicing (Jones 2012). The presence of histone
354 modification marks and transcription factor binding in the vicinity of methylation probe
355 cg22729539 suggests that this region contains cis-regulatory elements that actively regulate
356 transcription. These epigenetic factors, like DNA methylation and histone modification, may
357 interact with each other to influence gene expression in either the same or opposite directions
358 (Banovich et al. 2014; Cedar & Bergman 2009). DNA methylation could also affect nearby
359 transcription factor binding, such as transcription factor CEBPB, which plays an important role
360 in adipogenesis (Darlington et al. 1998), ER stress and pancreatic β cell failure (Matsuda et al.
361 2010). The coordination between a variety of genetic and epigenetic factors may regulate the
362 expression of *MLXIP*, and further the development of T2D.
363
364 The two SNPs, rs6134031 and rs4758690 have been reported to be associated with other human
365 traits, though genome-wide significance was not reached in those studies. In the NHGRI-EBI
366 GWAS catalog (Welter et al. 2014), SNP rs6134031 has been reported to be associated with
367 Plasma omega-6 polyunsaturated fatty acid levels (linoleic acid, n-6 PUFAs) (rs6134031-T,
368 beta= 0.0372, P-value= $4x10^{-6}$) (Dorajoo et al. 2015). The relationship between n-6 PUFAs and
369 T2D is debatable. Generally, n-6 PUFAs are considered to be proinflammatory and n-3 PUFAs

370 to be anti-inflammatory. Thus, high dietary intake of n-6 PUFAs and elevated (n-6) to (n-3) ratio
371 are associated with chronic inflammatory diseases including T2D (Patterson et al. 2012;
372 Simopoulos 2016). However, a recent study by Forouhi N.G. et al in a large number of European
373 subjects found that different types of n-6 PAFUs are differentially associated with risk of T2D
374 (Forouhi et al. 2016). Linoleic acid (LA) and eicosadienoic acid (EDA) were inversely
375 associated with T2D (OR<1), arachidonic acid (AA) was not significantly associated, and γ-
376 linolenic acid (GLA), dihomo-GLA, docosatetraenoic acid (DTA), docosapentaenoic acid (n6-
377 DPA) are positively associated (OR>1). Thus the relationship between n-6 PUFAs (and its
378 subtypes) and T2D needs to be further evaluated in more studies. SNP rs4758690 is also
379 associated with height (P-value=2.396×10^{-5}), however the effect size and direction of effect are
380 not available (Lango Allen et al. 2010). A systematic review and meta-analysis of 18 studies
381 revealed that significant inverse association between height and T2D risk was only observed in
382 women, not men (Janghorbani et al. 2012). Thus the genotype of these 2 SNPs are important for
383 inter-related human traits, suggesting these traits share common molecular underpinnings.
384
385 Our study has started to reveal the similarities and differences of the genetic basis of T2D
386 between Jordan subpopulations and other ethnicities. Despite the small sample size, we were
387 able to replicate almost half of the loci that were reported to be associated with T2D in genetic
388 and genomic studies in other populations. The replication of these associations suggests some
389 common genetic basis underlying the development of T2D among different ethnicities. For

390 complex traits and diseases, there are many GWAS loci which could not be replicated across
391 different ethnicities, such as the SNP rs7756992 in the *CDKAL1* gene which strongly associates
392 with T2D in subjects of European ancestry, but displayed no association in a population of West
393 Africa (Steinthorsdottir et al. 2007). Among the 37 SNPs associated with T2D in European or
394 Asian populations, only 2 were replicated in a Qatari population (O'Beirne et al. 2016). In the
395 Jordan subpopulations examined, we observed a significant association of rs6134031 and T2D,
396 with a very large effect size. In the WTCCC, including only subjects of European ancestry, the
397 LD structure for this region is different and the association of rs6134031 with T2D is less strong.
398 The association at SNP rs4758690 is nominally significant in both Jordan subpopulations,
399 however it is not significant in WTCCC subjects of European ancestry. The identification of
400 these two loci suggested unique genetic determinants for T2D in the Jordan subpopulations. The
401 separate GWAS performed in Chechen and Circassian subpopulations also suggest distinct
402 genetic factors for T2D in each of these two ethnicities. As reviewed by Rosenberg et al
403 (Rosenberg et al. 2010), such ethnic population differences may arise from variations in disease
404 allele frequency, effect direction, effect size, distinct LD patterns, and trait/disease phenotype
405 prevalence. Therefore, it is important to carry out genetic studies in different ethnic groups.
406
407 A major limitation of our study is the small size, which reduces the statistical power to detect a
408 true effect of the genetic variants. The small sample size may lead to p-values of true
409 associations failing to reach stringent significance thresholds, like the genome-wide significance

410 threshold of $5x10^{-8}$, resulting in false negatives (type II error). Therefore, we also considered
411 other biological evidence when interpreting our results and we were encouraged by the
412 replication of the *JAG1* locus and the strong eQTL signal observed for *MLXIP*, due to their
413 strong biological relevance to T2D. As reported and discussed by other studies, true association
414 may not always reach the conventionally corrected conservative threshold of $5x10^{-8}$ for declaring
415 a genome-wide significance (Nishizawa et al. 2014). In our case, future studies with larger
416 sample sizes of Jordan populations are needed to replicate the findings from our study and to
417 further identify other genetic loci.
418

419 **Conclusion**

420 Taken together, our results from the first GWAS of T2D conducted in two subpopulations in
421 Jordan have identified novel genetic factors underlying T2D; we additionally demonstrate there
422 is common genetic basis among the different ethnicities as well as certain unique genetic factors
423 that underlie T2D in the Jordan subpopulations. Identification of these novel genetic risk factors
424 will offer the potential to gain further insight into the development of T2D and may help with the
425 development of novel treatments precisely for the Jordan populations, which will reduce disease
426 burden and promote health.
427

428 **Acknowledgments**

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

References

Ajlouni K, Khader YS, Batieha A, Ajlouni H, and El-Khateeb M. 2008. An increase in prevalence of diabetes mellitus in Jordan over 10 years. *J Diabetes Complications* 22:317-324.

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

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

Barbujani G, Nasidze IS, and Whitehead GN. 1994a. Genetic diversity in the Caucasus. *Hum Biol* 66:639-668.

Barbujani G, Whitehead GN, Bertorelle G, and Nasidze IS. 1994b. Testing hypotheses on processes of genetic and linguistic change in the Caucasus. *Hum Biol* 66:843-864.

Barrett JC, and Cardon LR. 2006. Evaluating coverage of genome-wide association studies. *Nat Genet* 38:659-662. 10.1038/ng1801

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

Bi P, and Kuang S. 2015. Notch signaling as a novel regulator of metabolism. *Trends Endocrinol Metab* 26:248-255. 10.1016/j.tem.2015.02.006

Bianco A, Chiefari E, Nobile CG, Foti D, Pavia M, and Brunetti A. 2015. The Association between HMGA1 rs146052672 Variant and Type 2 Diabetes: A Transethnic Meta-Analysis. *PLoS One* 10:e0136077. 10.1371/journal.pone.0136077

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

463 Bray SJ. 2016. Notch signalling in context. *Nat Rev Mol Cell Biol* 17:722-735.
464 10.1038/nrm.2016.94
465 Bulayeva KB. 2006. Overview of genetic-epidemiological studies in ethnically and
466 demographically diverse isolates of Dagestan, Northern Caucasus, Russia. *Croat Med*
467 *J* 47:641-648.
468 Carroll PA, Diolaiti D, McFerrin L, Gu H, Djukovic D, Du J, Cheng PF, Anderson S, Ulrich M,
469 Hurley JB, Raftery D, Ayer DE, and Eisenman RN. 2015. Deregulated Myc requires
470 MondoA/Mlx for metabolic reprogramming and tumorigenesis. *Cancer Cell* 27:271-
471 285. 10.1016/j.ccell.2014.11.024
472 Cedar H, and Bergman Y. 2009. Linking DNA methylation and histone modification:
473 patterns and paradigms. *Nat Rev Genet* 10:295-304. 10.1038/nrg2540
474 Cheung L, Zervou S, Mattsson G, Abouna S, Zhou L, Ifandi V, Pelengaris S, and Khan M. 2010.
475 c-Myc directly induces both impaired insulin secretion and loss of beta-cell mass,
476 independently of hyperglycemia in vivo. *Islets* 2:37-45. 10.4161/isl.2.1.10196
477 Crawford SO, Hoogeveen RC, Brancati FL, Astor BC, Ballantyne CM, Schmidt MI, and Young
478 JH. 2010. Association of blood lactate with type 2 diabetes: the Atherosclerosis Risk
479 in Communities Carotid MRI Study. *Int J Epidemiol* 39:1647-1655.
480 10.1093/ije/dyq126
481 Dajani R, Khader YS, Fatahallah R, El-Khateeb M, Shiyab AH, and Hakooz N. 2012. Diabetes
482 mellitus in genetically isolated populations in Jordan: prevalence, awareness,
483 glycemic control, and associated factors. *J Diabetes Complications* 26:175-180.
484 10.1016/j.jdiacomp.2012.03.009
485 Darlington GJ, Ross SE, and MacDougald OA. 1998. The role of C/EBP genes in adipocyte
486 differentiation. *J Biol Chem* 273:30057-30060.
487 Delaneau O, Marchini J, and Zagury JF. 2012. A linear complexity phasing method for
488 thousands of genomes. *Nat Methods* 9:179-181. 10.1038/nmeth.1785
489 Delaneau O, Zagury JF, and Marchini J. 2013. Improved whole-chromosome phasing for
490 disease and population genetic studies. *Nat Methods* 10:5-6. 10.1038/nmeth.2307
491 Dorajoo R, Sun Y, Han Y, Ke T, Burger A, Chang X, Low HQ, Guan W, Lemaitre RN, Khor CC,
492 Yuan JM, Koh WP, Ong CN, Tai ES, Liu J, van Dam RM, Heng CK, and Friedlander Y.
493 2015. A genome-wide association study of n-3 and n-6 plasma fatty acids in a
494 Singaporean Chinese population. *Genes Nutr* 10:53. 10.1007/s12263-015-0502-2
495 Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, and Lin SM. 2010. Comparison of Beta-
496 value and M-value methods for quantifying methylation levels by microarray
497 analysis. *BMC Bioinformatics* 11:587. 1471-2105-11-587 [pii]
498 10.1186/1471-2105-11-587
499 Encode Project Consortium. 2012. An integrated encyclopedia of DNA elements in the
500 human genome. *Nature* 489:57-74. 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, and Ringwald M. 2017. The mouse Gene Expression Database (GXD): 2017 update. *Nucleic Acids Res* 45:D730-D736. 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, and 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 Med* 13:e1002094. 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, and Zalloua PA. 2014. T2DM GWAS in the Lebanese population confirms the role of TCF7L2 and CDKAL1 in disease susceptibility. *Sci Rep* 4:7351. 10.1038/srep07351

Glans F, Elgzyri T, Shaat N, Lindholm E, Apelqvist J, and Groop L. 2008. Immigrants from the Middle-East have a different form of Type 2 diabetes compared with Swedish patients. *Diabet Med* 25:303-307. 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. 10.1126/science.1262110

Guo S, Dai C, Guo M, Taylor B, Harmon JS, Sander M, Robertson RP, Powers AC, and Stein R. 2013. Inactivation of specific beta cell transcription factors in type 2 diabetes. *J Clin Invest* 123:3305-3316. 10.1172/JCI65390

Hirschhorn JN, and Daly MJ. 2005. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 6:95-108.

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

Ichikawa K, Yoshinari M, Iwase M, Wakisaka M, Doi Y, Iino K, Yamamoto M, and 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.

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, and Brunetti A. 2012. The HMGA1-

IGF-I/IGFBP system: a novel pathway for modulating glucose uptake. *Mol Endocrinol* 26:1578-1589. 10.1210/me.2011-1379

Imamura M, Chang BH, Kohjima M, Li M, Hwang B, Taegtmeyer H, Harris RA, and Chan L. 2014. MondoA deficiency enhances sprint performance in mice. *Biochem J* 464:35-48. 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, and Kadowaki T. 2016. Genome-wide association studies in the Japanese population identify seven novel loci for type 2 diabetes. *Nat Commun* 7:10531. 10.1038/ncomms10531

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

Jones PA. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 13:484-492. 10.1038/nrg3230

Kabakchiev B, and Silverberg MS. 2013. Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. *Gastroenterology* 144:1488-1496, 1496 e1481-1483. 10.1053/j.gastro.2013.03.001

Kailani W. 2002. Chechens in the Middle East: Between Original and Host Cultures. *Caspian Studies Program*.

Karpe F, Dickmann JR, and Frayn KN. 2011. Fatty acids, obesity, and insulin resistance: time for a reevaluation. *Diabetes* 60:2441-2449. 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 TH, Jr., Musk AW, Nieminen MS, O'Donnell CJ, Ohlsson C, Oostra B, Palmer LJ, Raitakari O, Ridker PM, Rioux JD, Rissanen A, Rivolta C, Schunkert H, Shuldiner AR, Siscovick DS, Stumvoll M, Tonjes A, Tuomilehto J, van Ommen GJ, Viikari J, Heath AC, Martin NG, Montgomery GW, Province MA, Kayser M, Arnold AM, Atwood LD, Boerwinkle E, Chanock SJ, Deloukas P, Gieger C, Gronberg H, Hall P, Hattersley AT, Hengstenberg C, Hoffman W, Lathrop GM, Salomaa V, Schreiber S, Uda M, Waterworth D, Wright AF, Assimes TL, Barroso I, Hofman A, Mohlke KL, Boomsma DI, Caulfield MJ, Cupples LA, Erdmann J, Fox CS, Gudnason V, Gyllensten U, Harris TB, Hayes RB, Jarvelin MR, Mooser V, Munroe PB, Ouwehand WH, Penninx BW, Pramstaller PP, Quertermous T, Rudan I, Samani NJ, Spector TD, Volzke H, Watkins H, Wilson JF, Groop LC, Haritunians T, Hu FB, Kaplan RC, Metspalu A, North KE, Schlessinger D, Wareham NJ, Hunter DJ, O'Connell JR, Strachan DP, Wichmann HE, Borecki IB, van Duijn CM, Schadt EE, Thorsteinsdottir U, Peltonen L, Uitterlinden AG, Visscher PM, Chatterjee N, Loos RJ, Boehnke M, McCarthy MI, Ingelsson E, Lindgren CM, Abecasis GR, Stefansson K, Frayling TM, and Hirschhorn JN. 2010. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467:832-838. 10.1038/nature09410

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

615 Marchini J, Howie B, Myers S, McVean G, and Donnelly P. 2007. A new multipoint method
616 for genome-wide association studies by imputation of genotypes. *Nat Genet* 39:906-
617 913. 10.1038/ng2088
618 Matsuda T, Kido Y, Asahara S, Kaisho T, Tanaka T, Hashimoto N, Shigeyama Y, Takeda A,
619 Inoue T, Shibutani Y, Koyanagi M, Hosooka T, Matsumoto M, Inoue H, Uchida T,
620 Koike M, Uchiyama Y, Akira S, and Kasuga M. 2010. Ablation of C/EBPbeta alleviates
621 ER stress and pancreatic beta cell failure through the GRP78 chaperone in mice. *J*
622 *Clin Invest* 120:115-126. 10.1172/JCI39721
623 Mtiraoui N, Turki A, Nemr R, Echtay A, Izzidi I, Al-Zaben GS, Irani-Hakime N, Keleshian SH,
624 Mahjoub T, and Almawi WY. 2012. Contribution of common variants of ENPP1,
625 IGF2BP2, KCNJ11, MLXIPL, PPARgamma, SLC30A8 and TCF7L2 to the risk of type 2
626 diabetes in Lebanese and Tunisian Arabs. *Diabetes Metab* 38:444-449.
627 10.1016/j.diabet.2012.05.002
628 Nasidze I, Ling EY, Quinque D, Dupanloup I, Cordaux R, Rychkov S, Naumova O, Zhukova O,
629 Sarraf-Zadegan N, Naderi GA, Asgary S, Sardas S, Farhud DD, Sarkisian T, Asadov C,
630 Kerimov A, and Stoneking M. 2004. Mitochondrial DNA and Y-chromosome variation
631 in the caucasus. *Ann Hum Genet* 68:205-221.
632 Nasidze I, Risch GM, Robichaux M, Sherry ST, Batzer MA, and Stoneking M. 2001. Alu
633 insertion polymorphisms and the genetic structure of human populations from the
634 Caucasus. *Eur J Hum Genet* 9:267-272.
635 Nishizawa D, Fukuda K, Kasai S, Hasegawa J, Aoki Y, Nishi A, Saita N, Koukita Y, Nagashima
636 M, Katoh R, Satoh Y, Tagami M, Higuchi S, Ujike H, Ozaki N, Inada T, Iwata N, Sora I,
637 Iyo M, Kondo N, Won MJ, Naruse N, Uehara-Aoyama K, Itokawa M, Koga M, Arinami
638 T, Kaneko Y, Hayashida M, and Ikeda K. 2014. Genome-wide association study
639 identifies a potent locus associated with human opioid sensitivity. *Mol Psychiatry*
640 19:55-62. 10.1038/mp.2012.164
641 O'Beirne SL, Salit J, Rodriguez-Flores JL, Staudt MR, Abi Khalil C, Fakhro KA, Robay A,
642 Ramstetter MD, Al-Azwani IK, Malek JA, Zirie M, Jayyousi A, Badii R, Al-Nabet Al-
643 Marri A, Chiuchiolo MJ, Al-Shakaki A, Chidiac O, Gharbiah M, Bener A, Stadler D,
644 Hackett NR, Mezey JG, and Crystal RG. 2016. Type 2 Diabetes Risk Allele Loci in the
645 Qatari Population. *PLoS One* 11:e0156834. 10.1371/journal.pone.0156834
646 Pajvani UB, Qiang L, Kangsamaksin T, Kitajewski J, Ginsberg HN, and Accili D. 2013.
647 Inhibition of Notch uncouples Akt activation from hepatic lipid accumulation by
648 decreasing mTorc1 stability. *Nat Med* 19:1054-1060. 10.1038/nm.3259
649 Pajvani UB, Shawber CJ, Samuel VT, Birkenfeld AL, Shulman GI, Kitajewski J, and Accili D.
650 2011. Inhibition of Notch signaling ameliorates insulin resistance in a FoxO1-
651 dependent manner. *Nat Med* 17:961-967. 10.1038/nm.2378
652 Palmer ND, McDonough CW, Hicks PJ, Roh BH, Wing MR, An SS, Hester JM, Cooke JN,
653 Bostrom MA, Rudock ME, Talbert ME, Lewis JP, Consortium D, Investigators M,

Ferrara A, Lu L, Ziegler JT, Sale MM, Divers J, Shriner D, Adeyemo A, Rotimi CN, Ng MC, Langefeld CD, Freedman BI, Bowden DW, Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G, McCulloch LJ, Ferreira T, Grallert H, Amin N, Wu G, Willer CJ, Raychaudhuri S, McCarroll SA, Langenberg C, Hofmann OM, Dupuis J, Qi L, Segre AV, van Hoek M, Navarro P, Ardlie K, Balkau B, Benediktsson R, Bennett AJ, Blagieva R, Boerwinkle E, Bonnycastle LL, Bostrom KB, Bravenboer B, Bumpstead S, Burtt NP, Charpentier G, Chines PS, Cornelis M, Couper DJ, Crawford G, Doney AS, Elliott KS, Elliott AL, Erdos MR, Fox CS, Franklin CS, Ganser M, Gieger C, Grarup N, Green T, Griffin S, Groves CJ, Guiducci C, Hadjadj S, Hassanali N, Herder C, Isomaa B, Jackson AU, Johnson PR, Jorgensen T, Kao WH, Klopp N, Kong A, Kraft P, Kuusisto J, Lauritzen T, Li M, Lieverse A, Lindgren CM, Lyssenko V, Marre M, Meitinger T, Midthjell K, Morken MA, Narisu N, Nilsson P, Owen KR, Payne F, Perry JR, Petersen AK, Platou C, Proenca C, Prokopenko I, Rathmann W, Rayner NW, Robertson NR, Rocheleau G, Roden M, Sampson MJ, Saxena R, Shields BM, Shrader P, Sigurdsson G, Sparso T, Strassburger K, Stringham HM, Sun Q, Swift AJ, Thorand B, Tichet J, Tuomi T, van Dam RM, van Haeften TW, van Herpt T, van Vliet-Ostaptchouk JV, Walters GB, Weedon MN, Wijmenga C, Witteman J, Bergman RN, Cauchi S, Collins FS, Gloyn AL, Gyllensten U, Hansen T, Hide WA, Hitman GA, Hofman A, Hunter DJ, Hveem K, Laakso M, Mohlke KL, Morris AD, Palmer CN, Pramstaller PP, Rudan I, Sijbrands E, Stein LD, Tuomilehto J, Uitterlinden A, Walker M, Wareham NJ, Watanabe RM, Abecasis GR, Boehm BO, Campbell H, Daly MJ, Hattersley AT, Hu FB, Meigs JB, Pankow JS, Pedersen O, Wichmann HE, Barroso I, Florez JC, Frayling TM, Groop L, Sladek R, Thorsteinsdottir U, Wilson JF, Illig T, Froguel P, van Duijn CM, Stefansson K, Altshuler D, Boehnke M, McCarthy MI, Soranzo N, Wheeler E, Glazer NL, Bouatia-Naji N, Magi R, Randall J, Johnson T, Elliott P, Rybin D, Henneman P, Dehghan A, Hottenga JJ, Song K, Goel A, Egan JM, Lajunen T, Doney A, Kanoni S, Cavalcanti-Proenca C, Kumari M, Timpson NJ, Zabena C, Ingelsson E, An P, O'Connell J, Luan J, Elliott A, McCarroll SA, Roccasecca RM, Pattou F, Sethupathy P, Ariyurek Y, Barter P, Beilby JP, Ben-Shlomo Y, Bergmann S, Bochud M, Bonnefond A, Borch-Johnsen K, Bottcher Y, Brunner E, Bumpstead SJ, Chen YD, Chines P, Clarke R, Coin LJ, Cooper MN, Crisponi L, Day IN, de Geus EJ, Delplanque J, Fedson AC, Fischer-Rosinsky A, Forouhi NG, Frants R, Franzosi MG, Galan P, Goodarzi MO, Graessler J, Grundy S, Gwilliam R, Hallmans G, Hammond N, Han X, Hartikainen AL, Hayward C, Heath SC, Hercberg S, Hicks AA, Hillman DR, Hingorani AD, Hui J, Hung J, Jula A, Kaakinen M, Kaprio J, Kesaniemi YA, Kivimaki M, Knight B, Koskinen S, Kovacs P, Kyvik KO, Lathrop GM, Lawlor DA, Le Bacquer O, Lecoeur C, Li Y, Mahley R, Mangino M, Manning AK, Martinez-Larrad MT, McAteer JB, McPherson R, Meisinger C, Melzer D, Meyre D, Mitchell BD, Mukherjee S, Naitza S, Neville MJ, Oostra BA, Orru M, Pakyz R, Paolisso

693 G, Pattaro C, Pearson D, Peden JF, Pedersen NL, Perola M, Pfeiffer AF, Pichler I,
694 Polasek O, Posthuma D, Potter SC, Pouta A, Province MA, Psaty BM, Rayner NW, Rice
695 K, Ripatti S, Rivadeneira F, Rolandsson O, Sandbaek A, Sandhu M, Sanna S, Sayer AA,
696 Scheet P, Seedorf U, Sharp SJ, Shields B, Sijbrands EJ, Silveira A, Simpson L, Singleton
697 A, Smith NL, Sovio U, Swift A, Syddall H, Syvanen AC, Tanaka T, Tonjes A,
698 Uitterlinden AG, van Dijk KW, Varma D, Visvikis-Siest S, Vitart V, Vogelzangs N,
699 Waeber G, Wagner PJ, Walley A, Ward KL, Watkins H, Wild SH, Willemsen G,
700 Witteman JC, Yarnell JW, Zelenika D, Zethelius B, Zhai G, Zhao JH, Zillikens MC,
701 Borecki IB, Loos RJ, Meneton P, Magnusson PK, Nathan DM, Williams GH, Silander K,
702 Salomaa V, Smith GD, Bornstein SR, Schwarz P, Spranger J, Karpe F, Shuldiner AR,
703 Cooper C, Dedoussis GV, Serrano-Rios M, Lind L, Palmer LJ, Franks PW, Ebrahim S,
704 Marmot M, Kao WH, Pramstaller PP, Wright AF, Stumvoll M, Hamsten A, Buchanan
705 TA, Valle TT, Rotter JI, Siscovick DS, Penninx BW, Boomsma DI, Deloukas P, Spector
706 TD, Ferrucci L, Cao A, Scuteri A, Schlessinger D, Uda M, Ruokonen A, Jarvelin MR,
707 Waterworth DM, Vollenweider P, Peltonen L, Mooser V, and Sladek R. 2012. A
708 genome-wide association search for type 2 diabetes genes in African Americans.
709 *PLoS One* 7:e29202. 10.1371/journal.pone.0029202
710 Parker SC, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA, van Bueren KL, Chines
711 PS, Narisu N, Program NCS, Black BL, Visel A, Pennacchio LA, Collins FS, National
712 Institutes of Health Intramural Sequencing Center Comparative Sequencing
713 Program A, and Authors NCSP. 2013. Chromatin stretch enhancer states drive cell-
714 specific gene regulation and harbor human disease risk variants. *Proc Natl Acad Sci*
715 *U S A* 110:17921-17926. 10.1073/pnas.1317023110
716 Pasquali L, Gaulton KJ, Rodriguez-Segui SA, Mularoni L, Miguel-Escalada I, Akerman I, Tena
717 JJ, Moran I, Gomez-Marin C, van de Bunt M, Ponsa-Cobas J, Castro N, Nammo T,
718 Cebola I, Garcia-Hurtado J, Maestro MA, Pattou F, Piemonti L, Berney T, Gloyn AL,
719 Ravassard P, Gomez-Skarmeta JL, Muller F, McCarthy MI, and Ferrer J. 2014.
720 Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated
721 variants. *Nat Genet* 46:136-143. 10.1038/ng.2870
722 Patterson E, Wall R, Fitzgerald GF, Ross RP, and Stanton C. 2012. Health implications of
723 high dietary omega-6 polyunsaturated Fatty acids. *J Nutr Metab* 2012:539426.
724 10.1155/2012/539426
725 Pe'er I, de Bakker PI, Maller J, Yelensky R, Altshuler D, and Daly MJ. 2006. Evaluating and
726 improving power in whole-genome association studies using fixed marker sets. *Nat*
727 *Genet* 38:663-667. 10.1038/ng1816
728 Pegoraro S, Ros G, Piazza S, Sommaggio R, Ciani Y, Rosato A, Sgarra R, Del Sal G, and
729 Manfioletti G. 2013. HMGA1 promotes metastatic processes in basal-like breast
730 cancer regulating EMT and stemness. *Oncotarget* 4:1293-1308.
731 10.18632/oncotarget.1136

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

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, and Willer CJ. 2010. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26:2336-2337. 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, and Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559-575. 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, and Kellis M. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* 518:317-330. 10.1038/nature14248

Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, and Boehnke M. 2010. Genome-wide association studies in diverse populations. *Nat Rev Genet* 11:356-366. 10.1038/nrg2760

Simopoulos AP. 2016. An Increase in the Omega-6/Omega-3 Fatty Acid Ratio Increases the Risk for Obesity. *Nutrients* 8:128. 10.3390/nu8030128

Sloan EJ, and Ayer DE. 2010. Myc, mondo, and metabolism. *Genes Cancer* 1:587-596. 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, and Stefansson K. 2007. A

771 variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet*
772 39:770-775. 10.1038/ng2043
773 Tiwari S, Roel C, Tanwir M, Wills R, Perianayagam N, Wang P, and Fiaschi-Taesch NM. 2016.
774 Definition of a Skp2-c-Myc Pathway to Expand Human Beta-cells. *Sci Rep* 6:28461.
775 10.1038/srep28461
776 Valenti L, Mendoza RM, Rametta R, Maggioni M, Kitajewski C, Shawber CJ, and Pajvani UB.
777 2013. Hepatic notch signaling correlates with insulin resistance and nonalcoholic
778 fatty liver disease. *Diabetes* 62:4052-4062. 10.2337/db13-0769
779 Wellcome Trust Case Control C. 2007. Genome-wide association study of 14,000 cases of
780 seven common diseases and 3,000 shared controls. *Nature* 447:661-678.
781 10.1038/nature05911
782 Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T,
783 Hindorff L, and Parkinson H. 2014. The NHGRI GWAS Catalog, a curated resource of
784 SNP-trait associations. *Nucleic Acids Res* 42:D1001-1006. 10.1093/nar/gkt1229
785 Wild S, Roglic G, Green A, Sicree R, and King H. 2004. Global prevalence of diabetes:
786 estimates for the year 2000 and projections for 2030. *Diabetes Care* 27:1047-1053.
787 Yamada Y, Matsuo H, Segawa T, Watanabe S, Kato K, Kameyama T, Yokoi K, Ichihara S,
788 Metoki N, Yoshida H, Satoh K, and Nozawa Y. 2006. Assessment of genetic factors for
789 type 2 diabetes mellitus. *Int J Mol Med* 18:299-308.
790 Yamaguchi S, Yamada Y, Matsuo H, Segawa T, Watanabe S, Kato K, Yokoi K, Ichihara S,
791 Metoki N, Yoshida H, Satoh K, and Nozawa Y. 2007. Gender differences in the
792 association of gene polymorphisms with type 2 diabetes mellitus. *Int J Mol Med*
793 19:631-637.
794 Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, and Liang G. 2014. Gene body
795 methylation can alter gene expression and is a therapeutic target in cancer. *Cancer*
796 *Cell* 26:577-590. 10.1016/j.ccr.2014.07.028

798 **Tables:**

799 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

800 N=Number

801

802 Table 2. Top associations ($P<5x10^{-5}$) found in meta-analysis of Circassian and Chechen

803 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)	$8.36x10^{-5}$
					Chechen	0.51/0.23	9.84 (3.33,29.02)	$3.45x10^{-5}$
					Meta		9.66	**$1.12x10^{-8}$**
					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)	$6.36x10^{-4}$
					Meta		3.00	**$4.20x10^{-5}$**
					European	0.53/0.52	1.01 (0.93,1.09）	0.61

804 SNP – single nucleotide polymorphism; Chr – chromosome; Pos – Position; A1 – minor allele; A2 – major allele; MAF – minor allele frequency;

805 OR – odds ratio; CI – confidence interval

806

807

Figures:

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).

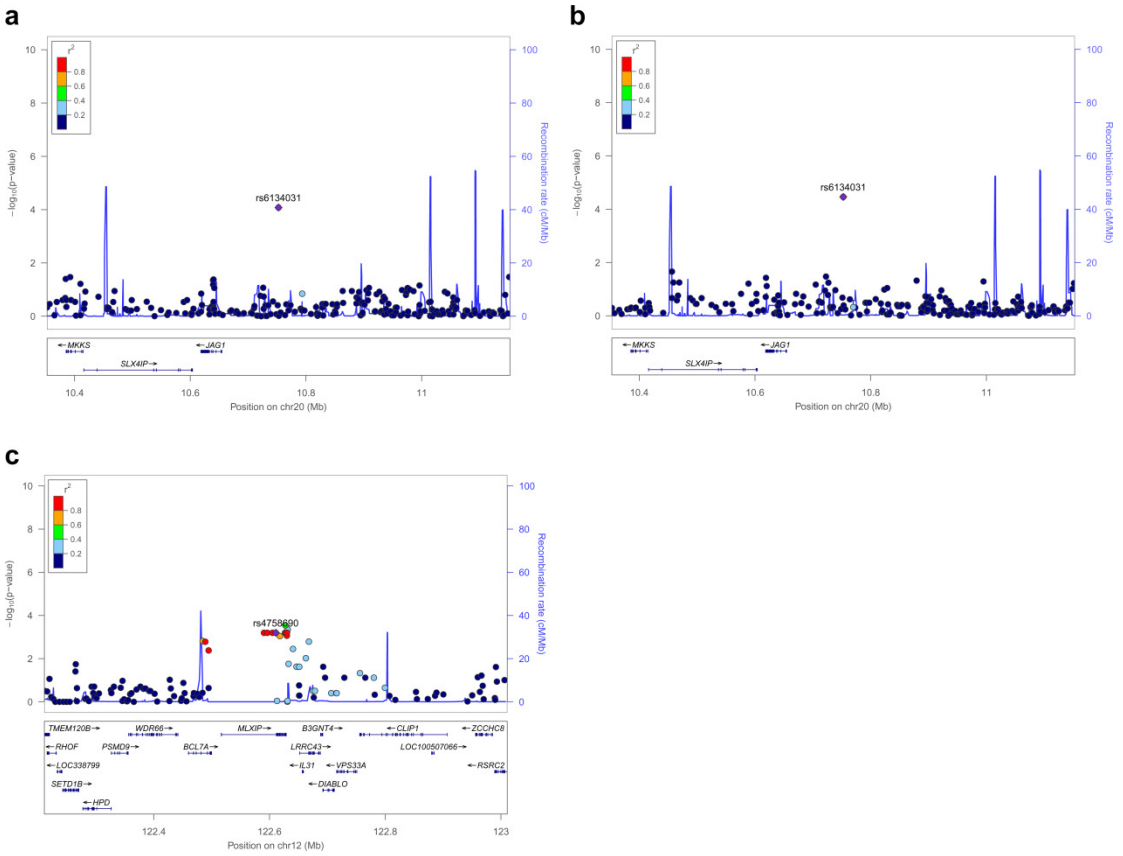
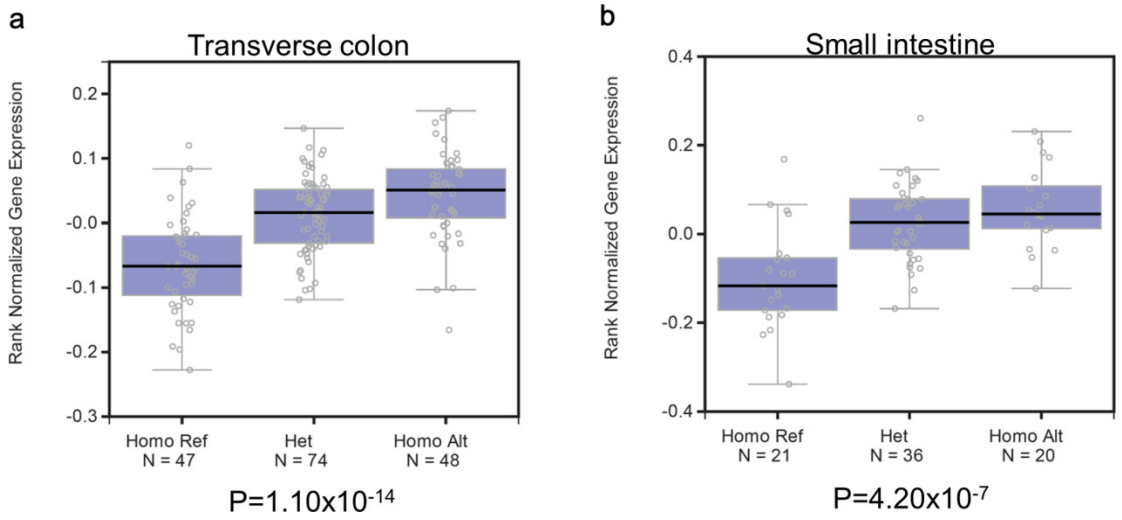


Figure 2: Box plots showing the association between SNP rs4758690 genotype and gene *MLXIP* expression level. **(a)** in tissue transverse colon, beta=0.46, P=1.10 x10^{-14}; **(b)** in tissue small intestine, beta=0.50, P=4.20x10^{-7}. The *in silico* analyses were conducted at GTEx Protal (GTEx Consortium 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.

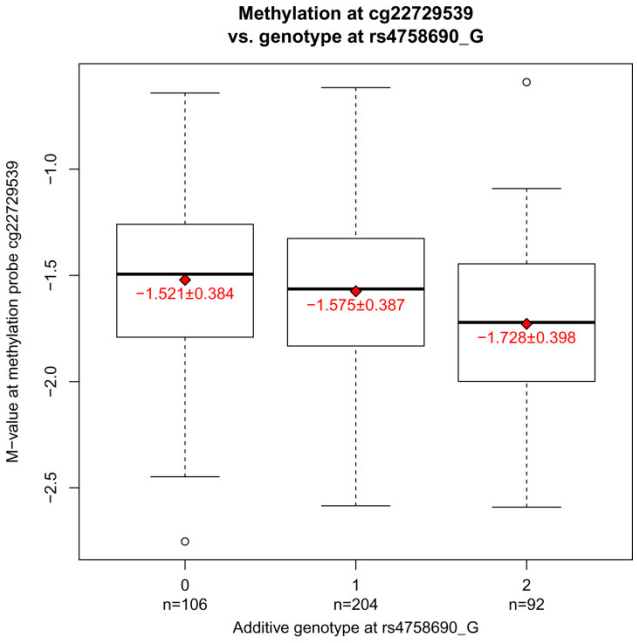


841

842

843

844 **Figure 3.** The association between SNP rs4758690 genotype and methylation status in gene
845 *MLXIP*. M-values for methylation probe cg22729539 are plotted against the additive genotype at
846 SNP rs4758690. Dark horizontal lines in the boxplots represent the median of the group, the
847 boxes show the 25%-75% quantiles, and the whiskers of the boxplot extend beyond those
848 quartiles to 1.5 times the interquartile range (IQR). Open circles indicate data outside the 1.5
849 IQR. Red diamonds indicate the means of each group, and the red text is the mean ± standard
850 deviation of each group. The number of individuals with each additive genotype of minor allele
851 G is shown below the X-axis.



852 P=3.07E-05

853