

RESEARCH ARTICLE

Exome sequencing-based identification of novel type 2 diabetes risk allele loci in the Qatari population

Sarah L. O’Beirne^{1,2}, Jacqueline Salit¹, Juan L. Rodriguez-Flores¹, Michelle R. Staudt¹, Charbel Abi Khalil³, Khalid A. Fakhro^{3,4}, Amal Robay³, Monica D. Ramstetter^{1,5}, Joel A. Malek³, Mahmoud Zirie⁶, Amin Jayyousi⁶, Ramin Badii⁷, Ajayeb Al-Nabet Al-Marri⁷, Abdulbari Bener⁶, Mai Mahmoud⁸, Maria J. Chiuchiolo¹, Alya Al-Shakaki³, Omar Chidiac³, Dora Stadler⁸, Jason G. Mezey^{1,5}, Ronald G. Crystal^{1,2*}

1 Department of Genetic Medicine, Weill Cornell Medical College, New York, New York, United States of America, **2** Division of Pulmonary and Critical Care Medicine, Department of Medicine, Weill Cornell Medical College, New York, New York, United States of America, **3** Department of Genetic Medicine, Weill Cornell Medical College-Qatar, Doha, Qatar, **4** Division of Translational Medicine, Sidra Medical Research Centre, Doha, Qatar, **5** Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY, United States of America, **6** Department of Medicine, Hamad Medical Corporation, Doha, Qatar, **7** Laboratory Medicine and Pathology, Hamad Medical Corporation, Doha, Qatar, **8** Department of Medicine, Weill Cornell Medical College-Qatar, Doha, Qatar

☞ These authors contributed equally to this work.

* geneticmedicine@med.cornell.edu



OPEN ACCESS

Citation: O’Beirne SL, Salit J, Rodriguez-Flores JL, Staudt MR, Abi Khalil C, Fakhro KA, et al. (2018) Exome sequencing-based identification of novel type 2 diabetes risk allele loci in the Qatari population. PLoS ONE 13(9): e0199837. <https://doi.org/10.1371/journal.pone.0199837>

Editor: Nicholette D. Palmer, Wake Forest School of Medicine, UNITED STATES

Received: October 5, 2017

Accepted: June 14, 2018

Published: September 13, 2018

Copyright: © 2018 O’Beirne et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All sequence read data and individual and population VCF files were submitted to the Sequence Read Archive (SRA) section of the NCBI SRA database (SRA accession #SRP061463, BioProject 288292). In addition, phenotype and covariate data and the population VCF has been made available through our website <http://geneticmedicine.weill.cornell.edu/genome.html>.

Funding: This grant was supported by QNRF NPRP 09-741-3-193 and general funds of the Department

Abstract

Background

Type 2 diabetes (T2D) susceptibility is influenced by genetic and lifestyle factors. To date, the majority of genetic studies of T2D have been in populations of European and Asian descent. The focus of this study is on genetic variations underlying T2D in Qataris, a population with one of the highest incidences of T2D worldwide.

Results

Illumina HiSeq exome sequencing was performed on 864 Qatari subjects (574 T2D cases, 290 controls). Sequence kernel association test (SKAT) gene-based analysis identified an association for low frequency potentially deleterious variants in 6 genes. However, these findings were not replicated by SKAT analysis in an independent cohort of 12,699 exomes, primarily due to the absence of low frequency potentially deleterious variants in 5 of the 6 genes. Interestingly one of the genes identified, catenin beta 1 (*CTNNB1*, β -catenin), is the key effector of the Wnt pathway and interacts with the nuclear receptor transcription factor 7-like 2 (*TCF7L2*), variants which are the most strongly associated with risk of developing T2D worldwide. Single variant analysis did not identify any associated variants, suggesting the SKAT association signal was not driven by individual variants. None of the 6 associated genes were among 634 previously described T2D genes.

Conclusions

The observation that genes not previously linked to T2D in prior studies of European and Asian populations are associated with T2D in Qatar provides new insights into the

of Genetic Medicine, Weill Cornell Medical College. Hamad Medical Corporation (HMC), is not a commercial company, but the premier not-for-profit health care provider for Doha, Qatar. The authors who are employees of HMC have the same no conflict as the other authors on this paper. The studies reported in this paper were grant supported (Qatar Foundation). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

complexity of T2D pathogenesis and emphasizes the importance of understudied populations when assessing genetic variation in the pathogenesis of common disorders.

Introduction

The prevalence of type 2 diabetes (T2D) in Qatar is one of the highest in the world [1]. T2D is a complex disease with both inheritance and superimposed environmental factors such as diet and lifestyle playing a role in susceptibility to the condition [2,3]. Genetic studies in T2D including genome-wide association studies (GWAS) and exome sequencing have helped to explain the inherited basis and pathogenesis of the condition and identified genes that influence pancreatic beta-cell function/insulin secretion and insulin resistance [4–6]. However, the prevalence of T2D varies widely among populations and despite the growing epidemic of T2D in the Middle East, the majority of these studies have been carried out in populations of European or Asian descent, populations where the prevalence of T2D is much lower than in Qatar [1,6]. Given genetic diversity among populations, the information from GWAS findings in Europeans or Asians may not be transferable to other populations including the Qataris.

We have previously demonstrated that the common T2D risk alleles identified in the European and Asian populations do not replicate in the Qatari population [7]. In the present study, exome sequencing of 864 Qataris (approximately 0.25% of the entire Qatari population) was applied for comparison of 574 Qataris with T2D to 290 Qatari non-diabetic controls. Using sequence kernel association analysis, low frequency (0.01 to 0.1 minor allele frequency (MAF)) potentially deleterious variants associated with T2D were identified in 6 genes, including catenin beta 1 (*CTNNB1*, β -catenin) the key effector of the Wnt pathway, which interacts with the nuclear receptor transcription factor 7-like 2 (*TCF7L2*), the gene most strongly associated with risk of developing T2D worldwide [8]. Additionally, *CTNNB1* interacts with the Wnt pathway member dishevelled segment polarity protein 1 (*DVL1*), also identified in this analysis. However, it was not possible to replicate the study findings in an independent cohort of 12,699 exomes made up of individuals from various populations in the T2D-GENES cohort [9], downloaded with permission from the European Genome Archive (<https://www.ebi.ac.uk/ega/studies/EGAS00001001460>) and analysed using the identical pipeline.

Subjects and methods

Study population

Under protocols approved by the Institutional Review Boards of Hamad Medical Corporation (HMC) and Weill Cornell Medical College Qatar (WCMC-Q), subjects were recruited from HMC clinics and written informed consent obtained. A total of 864 subjects (574 cases with T2D and 290 controls) were included in the study (S1 Table). Most of the subjects were participants in our previous study [7]. T2D was diagnosed based on the American Diabetes Association (ADA) criteria including fasting blood glucose ≥ 126 mg/dL and/or 2 hour plasma glucose ≥ 200 mg/dL during an oral glucose tolerance test and/or HbA1C $\geq 6.5\%$ [10]. All subjects were over the age of 30 with a family history of a minimum three generations of ancestry in Qatar. Detailed subject assessment is described in this article's S1 File.

Sequencing and variant detection

Exome sequencing was performed at the New York Genome Center (NYGC) as previously described [11]. Reads were mapped to the GRCh37 human reference genome and prepared

for variant calling using GATK best practices [12], and a call set for all 864 exomes was produced by simultaneous genotyping using the GATK UnifiedGenotyper algorithm [12]. Data quality was optimized and verified based on multiple metrics described in S1 File. Variants ($n = 295,515$) were categorized based on MAF into singleton (MAF < 0.001 , $n = 46,874$), rare (MAF 0.001 to 0.01, $n = 190,771$, including $n = 64,258$ doubletons), low-frequency (MAF 0.01 to 0.1, $n = 58,671$) and common (MAF > 0.1 , $n = 46,073$). Singleton and rare variants were excluded from the SKAT analysis due to the enrichment of false positive variants in this dataset, as demonstrated in our prior study of the same exome dataset *versus* whole genome sequence data [11]. Additionally, common variants were excluded as our prior study demonstrated that they are not associated with T2D in Qatar [7]; leaving $n = 58,671$ variants.

Variants in protein coding genes were identified using SnpEff v.4.2 [13] which uses ENSEMBL v.75 gene models to assign variants to genes, and to determine variant functional region and impact on the assigned gene. SnpEff classified variants in protein coding genes into 4 impact categories (modifier, low, moderate and high) based on their potential for altering protein structure and function. The focus of this study was on variants of moderate and high potential for deleteriousness, which included missense and loss of function (LoF) single nucleotide variants (SNV). In addition, quantitative scores of deleteriousness were calculated using CADD v.1.3 [14]. Variants were also annotated with respect to allele frequency in the cohort, cases, controls, 1000 Genomes Phase 3 v.5 [15] and ExAC v.3.1 [16], calculated using VCFTools v.0.1.14 [17].

Statistical analysis

To identify genes linked to diabetes most effectively, the analysis was limited to low frequency (MAF 0.01 to 0.1) variants with moderate or high potential for altering protein structure or function ($n = 20,492$). Four distinct association analyses were conducted on the whole-exome sequence data genotypes, including gene-based analysis (sequence kernel association test, SKAT) [18] and variant-based single variant analysis (SVA) [19], with Bonferroni multiple testing correction for all genes and for known T2D genes [20]. The genetic models tested included an association test for each variant after applying prior filters (SVA), and an association test for each gene (SKAT) containing at least 1 variant after filtering. The SVA was conducted using EMMAX [19], with age, gender, body mass index (BMI) and kinship matrix (calculated by EMMAX-KIN) as covariates. The SKAT test was conducted using the SKAT v.1.2.2 (<https://cran.r-project.org/web/packages/SKAT/>) library in R v.3.3.2. (<https://cran.r-project.org/>). Two gene set filters were considered- all protein coding genes, and a subset of tested genes overlapping with 634 genes previously linked to T2D (Supplemental Table 20 of Fuchsberg *et al* [20]). These gene sets were used for multiple testing corrections by the Bonferroni method [21] with $\alpha = 0.05$. Population structure was accounted for in the SKAT and SVA analysis using the kinship matrix, with kinship analysis conducted using EMMAX-KIN v.10Mar2010 [19]. While no relatives were knowingly included in the analysis, the Qatari population has a high prevalence of consanguineous marriage [19,22], increasing the likelihood that relatives may have been included inadvertently.

The Qatari population can be divided into 3 major ancestry clusters (Arab/Bedouin (Q1), Persian/South Asian (Q2) and Sub-Saharan African (Q3)) [11,22,23]. A majority of individuals in this study belonged to the Q1 ($n = 605$) and Q2 ($n = 210$) clusters, with a smaller number ($n = 49$) from the Q3 subpopulation, proportions similar to those observed in the general Qatari population. Consanguinity is high and comparable to other Middle Eastern populations in the Q1 and Q2 subpopulations, while consanguinity is lower and more comparable to African populations in the Q3 subpopulation [24]. The relatively higher consanguinity among the

Q1 and Q2 individuals included in this study was confirmed in another study which reconstructed large pedigrees of 1st, 2nd, and 3rd cousins among $n = 1376$ Qataris including the $n = 864$ Qataris from this study [11]. A kinship matrix was utilized to control for population structure, accounting for both near and distant relationships (including consanguineous sub-populations). Adequate population structure correction was verified by QQ plots of the variant distributions. Although no “genomic control” lambda adjustment was conducted, the lambda value was calculated for both the SKAT and SVA distributions [25], to assess inflation due to residual population structure or polygenic disease risk [26].

To exclude the possibility of gaps in coverage depth for significant genes, the coverage depth mean and 95% confidence interval was calculated for each exon, in each individual, in each significant gene using Samtools with exons defined by the Agilent target list bed file, and summary of the results was calculated in R.

Replication

To replicate any associations identified in this study in an independent cohort, SKAT test p values were calculated for 12,699 exomes from the T2D-GENES cohort [9,20], obtained with permission from the European Genome Archive (<https://www.ebi.ac.uk/ega/studies/EGAS00001001460>). SKAT was calculated for each gene that was associated with T2D in the analysis using the identical code and parameters used for analysis of Qatari subjects.

Investigation of known T2D loci

Prior studies have identified at least $n = 81$ variants linked to T2D in one or more human populations [20]. To determine whether any of these variants are linked to T2D in Qatar, confidently genotyped variants that overlapped with the set of $n = 81$ were extracted from the $n = 295,515$ genotyped in Qataris by exome sequencing. The case and control allele frequency for known alleles was compared between published reports related to the identified variants in the Qatari data. In addition, the power to detect an association in Qatar in the current study was calculated for each of the identified variants (S6 Table) using Purcell’s Genetic Power Calculator [27], assuming the same odds ratio as in published studies [28–32], as previously described [7].

Data sharing

All sequence read data and individual and population VCF files were submitted to the Sequence Read Archive (SRA) section of the NCBI SRA database (SRA accession #SRP061463, BioProject 288292). In addition, phenotype and covariate data and the population VCF was made available through our website <http://geneticmedicine.weill.cornell.edu/genome.html>. Code for replicating the analysis was also made available (https://github.com/juansearch/_WAS).

Results

Demographics

A total of 864 subjects passed quality control, all were over the age of 30 and self-reported to have three generations of native Qatari ancestry. Of these, 574 were classified as T2D cases based on ADA criteria for fasting plasma glucose, oral glucose tolerance test and/or HbA1C level, as detailed in S1 File. The controls ($n = 290$) had normal fasting glucose and HbA1C $\leq 6.5\%$. The majority of the subjects were female both for cases with T2D (344/574, 60%) and controls (169/290, 58%; $p > 0.5$). T2D cases were significantly older with a mean age of 56 ± 10

yr vs 46 ± 9 yr in controls ($p < 10^{-10}$). BMI was significantly higher in T2D cases (33 ± 7 kg/m²) compared to controls (31 ± 7 kg/m²; $p < 10^{-4}$). HbA1C was significantly higher in T2D cases ($8.4 \pm 1.9\%$) compared to controls ($5.7 \pm 0.4\%$, $p < 10^{-10}$; S1 Table).

Association tests

Both gene-based sequence kernel association test (SKAT) and variant-based SVA association tests were conducted on potentially deleterious low frequency variants. After variant filtering based on allele frequency (1% to 10%) and function (missense or LoF), exome sequencing identified a total of $n = 20,492$ potentially deleterious SNVs in $n = 9,378$ protein coding genes (S4 Table). These variants were tested for association with T2D using the SKAT with age, gender, BMI and population structure as covariates. Rather than exclude relatives, the kinship matrix of all 864 Qataris was used to control for population structure (S1 Fig). Population structure was also examined using Principal Components Analysis of Qataris (S2A Fig) and Qataris with 1000 Genomes Populations (S2B Fig), and confirmed to show the range of diversity previously described in this population [11,23]. Two multiple testing corrections were applied, limiting the analysis to either all genes or known T2D genes [20], combined with application of Bonferroni [21] multiple testing correction.

Six genes had Bonferroni-significant p values in the SKAT analysis (S2 Table). None of the associated genes were among 332 known T2D genes tested, and Bonferroni multiple testing correction limited to known T2D genes did not identify any significant genes. In addition, SVA did not identify any individual variants associated with T2D by either multiple testing correction. The lack of significant SVA results could be attributed to low power given the sample size, as well as prior studies showing the lack of replication for T2D SVA hits in Qatar [7].

The 6 genes that were significant under Bonferroni multiple testing correction in the SKAT analysis of potentially deleterious low frequency variants included *CTNNB1*, *DLL1*, *DTNB*, *DVLI*, *EPB41L3*, and *KIF12* (Table 1, Fig 1A.). The QQ plot for the analysis showed no evidence of major inflation in SKAT analysis ($\lambda = 1.15$), with no need for λ adjustment (Fig 1B), and similar findings for SVA of the same variants ($\lambda = 1.03$) (Fig 1C). Of 9 potentially deleterious low frequency variants in these 6 genes, 8 had combined annotation dependent depletion (CADD) scores > 9 (Table 2). Coverage depth mean and 95% confidence interval across exons in these 6 genes was 60.8 ± 0.27 , excluding gaps in coverage depth that

Table 1. Qatari type 2 diabetes (T2D) at-risk genes identified by sequence kernel association test (SKAT) of low-frequency potentially deleterious protein coding single nucleotide polymorphisms (SNP)¹.

Gene	Gene name	Chr	Start	End	Potentially deleterious coding variants ²	SKAT p ³	Significance threshold
<i>KIF12</i>	Kinesin family member 12	9	114,091,623	114,100,099	1	2.37×10^{-9}	Bonferroni
<i>DVLI</i>	Dishevelled segment polarity protein 1	1	1,335,278	1,349,142	1	3.30×10^{-7}	Bonferroni
<i>EPB41L3</i>	Erythrocyte membrane protein band 4.1 like 3	18	5,392,381	5,630,666	3	9.91×10^{-7}	Bonferroni
<i>DTNB</i>	Dystrobrevin beta	2	25,377,220	25,673,647	2	1.20×10^{-6}	Bonferroni
<i>DLL1</i>	Delta like canonical Notch ligand 1	6	170,282,200	170,291,075	1	3.34×10^{-6}	Bonferroni
<i>CTNNB1</i>	Catenin beta 1	3	41,199,451	41,240,448	1	3.35×10^{-6}	Bonferroni

¹ The SKAT analysis presented was conducted on all 864 Qataris, limited to 20,492 potentially deleterious (missense or loss of function, LoF) low-frequency (minor allele frequency 1% to 10%) variants in 9,378 protein coding genes. None of these genes were previously linked to T2D, and the known T2D genes [20] with the lowest p value was CCAAT/enhancer binding protein alpha (*CEBPA*), with a p value of 2.93×10^{-4} . A total of 54 genes had p-values lower than *CEBPA*, including the 6 shown.

² Potentially deleterious defined as missense or LoF.

³ SKAT p: Sequence kernel association test p value for the gene.

<https://doi.org/10.1371/journal.pone.0199837.t001>

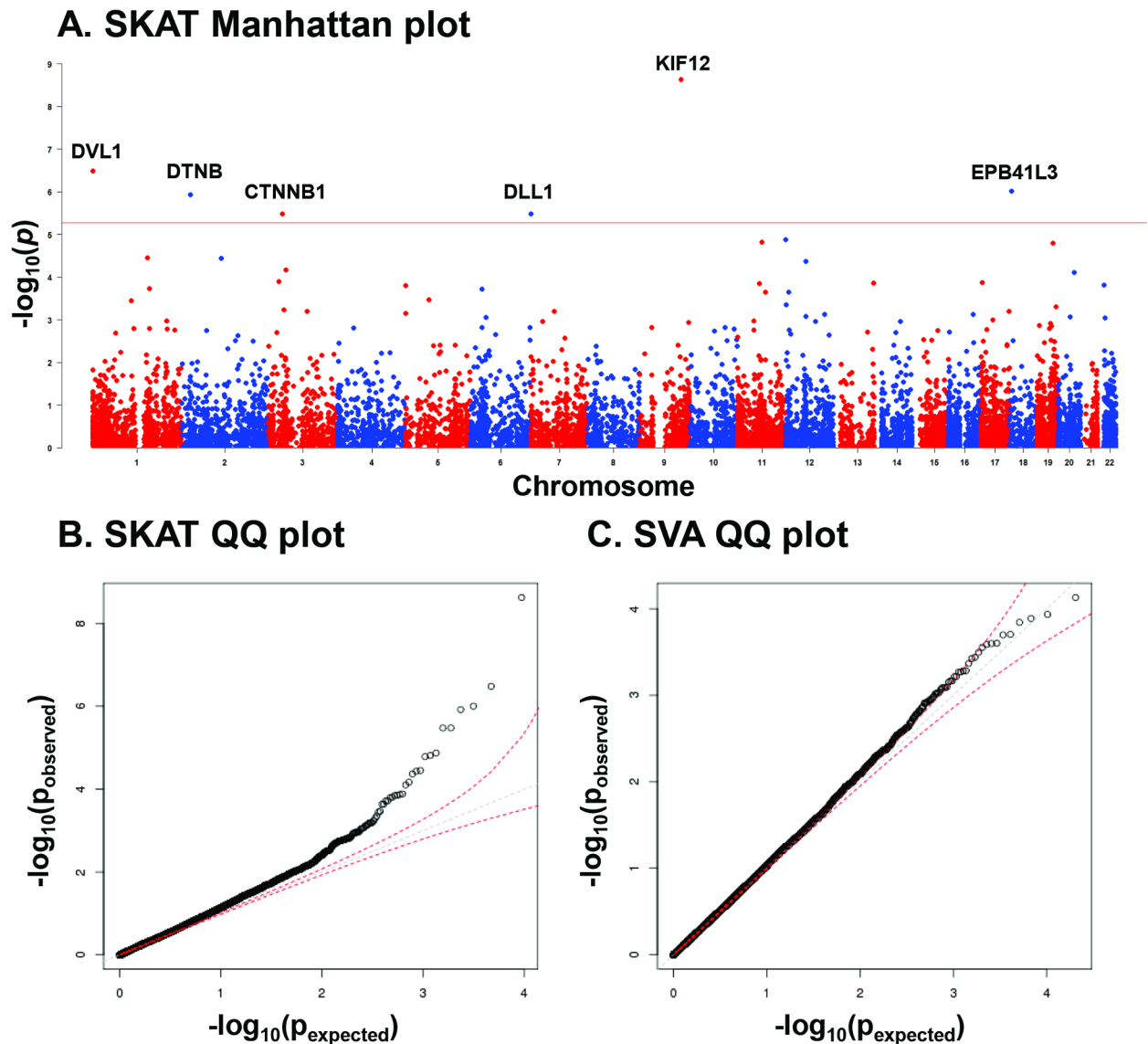


Fig 1. Manhattan and quantile-quantile (QQ) plots of genome-wide sequence kernel association analysis (SKAT) and QQ plot of single variant analysis (SVA) of 20,492 low-frequency potentially deleterious variants in 9,378 genes in all 864 Qataris (574 cases with type 2 diabetes (T2D) and 290 controls), with age, gender, BMI and kinship as covariates. Six genes were significant in the analysis, with p values below Bonferroni multiple testing threshold ($\alpha = 0.05$). A. Manhattan plot of genome wide association analysis. The plots show $-\log_{10}(p)$ on the y-axis and the chromosomal position of each variant on the x-axis. Genes are ranked by uncorrected p values. Genes with $p <$ the Bonferroni correction p value threshold are labeled in the plot [above the red horizontal line, $\log_{10}(0.05/22003) = 5.64$]. B. QQ plot between observed and expected p values for SKAT. The red broken lines represent the confidence band indicating the range of results consistent with a 95% interval around the null (grey broken line). No lambda correction was applied. The observed lambda value was $\lambda = 1.15$ for SKAT. C. QQ plot between observed and expected p values for SVA of potentially deleterious variants using EMMAX [19], with age, gender, BMI and kinship as covariates. The red broken lines represent the confidence band indicating the range of results consistent with a 95% interval around the null (grey broken line). No lambda correction was applied. The observed lambda value was $\lambda = 1.03$ for SVA.

<https://doi.org/10.1371/journal.pone.0199837.g001>

could produce false positive or false negative variants. None of the genes replicated in the T2D-GENES cohort (S3 Table), primarily due to lack of potentially deleterious low frequency variants in this cohort for 5 of 6 genes. The only gene where a replication test was possible was *EPB41L3*, which contained 4 potentially deleterious low frequency variants ($p > 0.81$).

Table 2. Single nucleotide polymorphisms (SNP) in qatari type 2 diabetes (T2D) at-risk genes^{1,2}.

Gene	Minor allele frequency ³																		
	Cases						Controls												
	Chr	Pos	rsID	SVA p value	CADD ⁴	Transcript change	Protein change	Function	Minor/Major alleles	Hom ⁵ Minor (cases)	Het ⁶ (cases)	Hom Major (cases)	Hom Minor (controls)	Het (controls)	Hom Major (controls)	Cases	Controls	ExAC	1000 Genomes
<i>KIF12</i>	9	116,859,679	.	0.8899	28	c.134T>C	p.Leu45Pro	Missense	G/A	0	3	571	0	25	265	0.0026	0.0431	0.0000	0.0000
<i>DVLI</i>	1	1,271,676	.	0.4723	9	c.1934A>C	p.His645Pro	Missense	G/T	0	16	558	0	36	254	0.0139	0.0621	0.0000	0.0000
<i>EPB41L3</i>	18	5,397,367	.	0.0145	10	c.2531T>C	p.Leu844Pro	Missense	G/A	0	14	560	0	34	256	0.0122	0.0586	0.0000	0.0000
<i>EPB41L3</i>	18	5,416,160	rs8082898	0.6001	9	c.1724A>G	p.Tyr575Cys	Missense	C/T	0	16	558	0	7	283	0.0139	0.0121	0.0340	0.0565
<i>EPB41L3</i>	18	5,478,295	rs117900256	0.6378	23	c.326G>T	p.Ser109Ile	Missense	A/C	1	37	536	0	16	274	0.0340	0.0276	0.0130	0.0122
<i>DTNB</i>	2	25,611,134	.	0.5382	25	c.1672A>C	p.Thr558Pro	Missense	G/T	0	11	563	0	31	259	0.0096	0.0535	0.0000	0.0000
<i>DTNB</i>	2	25,611,140	rs562264712	0.5472	23	c.1666A>C	p.Thr556Pro	Missense	G/T	0	37	537	0	40	250	0.0322	0.0690	0.0000	0.0116
<i>DLL1</i>	6	170,592,620	rs200861263	0.0460	0	c.1747T>C	p.Cys583Arg	Missense	G/A	0	61	513	0	69	221	0.0531	0.1190	0.0002	0.0000
<i>CTNNB1</i>	3	41,278,119	rs77750814	0.5528	23	c.1995C>A	p.Asp665Glu	Missense	A/C	0	75	499	0	6	284	0.0653	0.0103	0.1090	0.0000

¹ Single variant analysis (SVA) was conducted to identify associations between low frequency potentially deleterious variants and T2D using EMMAX v.10Mar2010 on all 864 Qataris, using age, gender, BMI and a kinship matrix calculated using EMMAX-KIN as covariates. To determine if single variants in the 6 associated genes (Table 1) were driving the SKAT association signal, the SVA p values are presented for potentially deleterious variants in these genes. Variants were functionally annotated using SnpEff v.4.2 using ENSEMBL v.75 gene models, and potentially deleterious variants were either missense or loss of function variants.

² Shown (from left-to-right) is the gene symbol, chromosome (Chr) and position (Pos) of the variant, DbsNP v.147 rsID for the variant (or “.” if novel), the SVA p value, combined annotation dependent depletion (CADD) score, transcript change (in reference-alternate allele order), protein change (in reference-alternate allele order), variant function, minor and major alleles, genotype counts for cases and controls, minor allele frequency in cases, controls, ExAC and 1000G.

³ The Qatari minor allele frequency was quantified in ExAC v.0.3.1 [16] and in 1000 Genomes Phase 3 v.5 [15].

⁴ CADD scores were calculated for each variant to further assess the potential for deleteriousness [14].

⁵ Hom: Homozygous

⁶ Het: Heterozygous

<https://doi.org/10.1371/journal.pone.0199837.t002>

Integration of SKAT and SVA results

Though SVA did not identify variants significantly associated with T2D in Qatar, it provided information on the directionality of variant effect for the 9 variants identified within the 6 SKAT-significant T2D associated genes, based on variant frequency in cases and controls. In the case of 6 of the 9 variants, the minor allele in Qatar had a higher frequency in controls and appeared to be protective (*KIF12* p.Leu45Pro, *DVL1* p.His645Pro, *EPB41L3* p.Leu844Pro, *DTNB* p.Thr558Pro, *DTNB* p.Thr556Pro, and *DLL1* p.Cys583Arg), while the other 3 variants (*EPB41L3* p.Tyr575Cys, *EPB41L3* p.Ser109Ile and *CTNNB1* p.Asp665Glu.) appeared to be risk increasing with a higher frequency in cases.

Investigation of known T2D loci

While no single variant was significant after multiple testing correction, the allele frequency in cases and controls was compared to prior studies for $n = 6$ variants previously linked to T2D [20,29–32] located within protein coding exons genotyped in this study. These variants included well-known T2D associated variants, such as the *PPARG* p.Pro12Ala variant (rs1801282) that was previously shown not replicate in Qataris [33] (S5 Table). For four of the six variants (67%) the known risk allele from prior reports was the major allele and, in five of the six variants (83%), the risk allele had a higher allele frequency in cases not surprising given the high prevalence of T2D in Qatar. The lowest p value among the 6 variants was for rs515071 ($p < 0.036$), which is located near an intron/exon junction of ankyrin 1 (*ANK1*). While the expected odds ratio in the Qatari population is unknown, assuming the same odds ratio as published studies and given the sample size, this study had insufficient power to detect an association for these 6 variants in single variant analysis (S6 Table), confirmed by the lack of significant SVA associations. However, the advantage of SKAT in this study is that it combines data across variants in the same gene, which leads to additional power to detect an association and a less stringent multiple testing correction.

Discussion

The prevalence of T2D in Qatar is one of the highest in the world [1]. In the present study, exome sequencing, a strategy not previously applied in the study of T2D in Middle Eastern populations, identified 6 genes (*KIF12*, *DVL1*, *EPB41L3*, *DTNB*, *DLL1*, *CTNNB1*) with potentially deleterious low frequency variants significantly associated with T2D. Interestingly, none of these genes replicated in an independent cohort of 12,699 exomes from T2D-GENES. This observation provides new insights into the complexity of the pathogenesis of T2D and emphasizes the importance of understudied populations. Out of 6 genes associated with T2D, 5 contained potentially deleterious variants present only in Qataris, highlighting the unique genetics of T2D in this population.

Prior studies of T2D risk alleles in middle eastern populations

To date, GWAS performed mostly in European and Asian populations, have identified over 80 loci associated with T2D that include 634 genes [4–6,34,35] and exome sequencing has identified over 10 T2D risk genes [4–6,20,36]. However, these associations may not replicate in other populations due to variability in risk allele frequency in the control and/or the specific phenotype being assessed [37]. There have been no prior GWAS or exome studies of T2D risk alleles in Middle Eastern populations. We previously reported that from a panel of 37 single nucleotide polymorphisms (SNP) commonly associated in GWAS of T2D in Europeans and Asians, only two SNPs in *TCF7L2*, rs7903146 and rs4506565 were associated with T2D in

Qataris, suggesting that the genetic risks for T2D are different in Qataris compared to Europeans and Asians [7]. Further, candidate gene studies of T2D in Middle Eastern populations yielded inconsistent results. While it was reported that many of the known European risk alleles were associated with T2D in Lebanese Arabs [38], the Saudi Arabian population [39], and in Tunisians and Moroccans [40], others did not demonstrate an association between the European T2D risk SNPs in these populations. For example, though a link between SNPs in *TCF7L2* and T2D was reported in Moroccans [40], only a marginal association was found in Arabs [41] and several European SNPs were not associated with T2D in Tunisians [42]. Finally, only 2 of 23 loci associated with BMI in other populations were linked to obesity in Qataris [43].

Exome analysis in the Qatari population

Exome sequencing identifies low frequency population specific coding variants that would not otherwise be detected by array genotyping, with the potential to explain some of the heritability not identified by GWAS [44,45]. Out of $n = 81$ SNPs associated with T2D by GWAS [20], very few have led to the discovery of causal variants [44,46]. In contrast, exome sequencing assesses only the coding portion of the genome, and when the analysis is limited to potentially deleterious (missense and LoF) variants that alter protein function, the associated variants have a plausible mechanism of functional importance [47]. However, unless sample or effect sizes are very large, most statistical tests are underpowered to identify rare variants. To circumvent this challenge, SKAT an association test, which uses a multiple regression model to test for association between multiple variants in a sequenced region and a phenotype, was developed [18]. Recent updates to SKAT allowed adjustment for covariates and kinship in the statistical model. Furthermore, in contrast to burden tests, SKAT allows different variants in this study to have positive, negative or no effect on the phenotype [18]. Using SKAT, this study identified 6 genes with potentially deleterious low frequency variants associated with T2D in the Qatari population. These genes have not previously been linked to T2D risk in GWAS or exome sequencing analysis of European or Asian populations, confirming prior studies failing to replicate known T2D risk alleles in the Qatari population [7]. For 5 of the genes potentially deleterious variants were identified in Qataris only, with no potentially deleterious variants observed in the large T2D-GENES cohort used for replication testing [9].

The gene most strongly associated with T2D in Qataris and replicated in other populations was *DVLI*. This gene encodes a member of a family of intracellular scaffolding proteins that act downstream of transmembrane Wnt receptors playing an important role in the signalling pathway including interacting with and stabilizing *CTNNB1*. It was previously linked to autosomal dominant Robinow syndrome 2, [48] and exhibited an accelerated rate of evolution in primates as compared to rodents, most prominent in the lineage from primates to humans [49]. With regards to diabetes, methylation at CpGs in the body of *DVLI* has been associated with the development of T2D [50], and expression of *DVLI* is reduced in adipose tissue of non-diabetic subjects with insulin resistance compared to controls [51]. In addition, *DVLI* was previously identified within a quantitative trait locus region for T2D [52].

CTNNB1 (previously known as β -catenin) mediates canonical Wnt signalling, a pathway critically involved in many processes including embryogenesis, cell growth and motility [53]. Wnt signalling is tightly regulated during growth and development of the pancreas and islet cells, and the Wnt pathway is involved in pancreatic beta-cell proliferation, glucose homeostasis and lipid metabolism [53]. Interestingly *TCF7L2* is a nuclear receptor for *CTNNB1*, and *TCF7L2* variants are the most strongly associated with risk of developing T2D, and have been reproduced in many populations [8,53], including Qataris [7]. Additionally, prior studies have

demonstrated that during WNT signalling, *DVL1* a component of the pathway acts to stabilize *CTNNB1* leading to accumulation of *CTNNB1* in the nucleus and subsequent transcription of Wnt-responsive genes [54]. Activation of the Wnt/ β -catenin pathway has been demonstrated to improve pancreatic beta-cell regeneration in diabetic rats [55], a discovery with potential therapeutic implications in T2D.

As with the Wnt pathway, Notch signalling has been shown to be critical for normal pancreatic development, controlling differentiation of pancreatic progenitor cells into endocrine and exocrine cells [56], and has also been implicated in the development of T2D. *DLL1* is a delta-like Notch ligand. Notch signalling has been demonstrated to control development of the murine endocrine pancreas, with *DLL1* knockout mice displaying premature endocrine cell development and subsequent pancreatic hypoplasia [57]. *In vitro* induction of Delta-Notch pathway ligand expression including *DLL1* in adult human pancreatic cells leads to the development of insulin expressing cells [58]. Additionally, a SNP in an intra-genic region spanning several genes including *DLL1* has been associated with type 1 diabetes (T1D) [59]. Also in T1D, it was demonstrated that *DLL1* protein expression is elevated in smooth skeletal muscle in cases compared to controls [60].

Kinesin family proteins (KIF) are a group of proteins with motor and microtubule cytoskeleton functions. Members of this superfamily are recognized to play an important role in the control of blood sugar both through pancreatic beta-cell insulin secretion and insulin stimulated glucose uptake in target tissues [61]. The *KIF12* gene codes for a cytoplasmic scaffold protein involved in microtubule motor functions. *KIF12* is expressed in fetal liver, adult brain and pancreatic islets, as well as renal tumors, and pancreatic cancer. Lipotoxicity due to chronic exposure to excess fatty acids, as may occur in obesity, is believed to damage pancreatic beta-cell *via* increased oxidative stress leading to beta-cell dysfunction and diabetes [62]. In the pancreas and kidney, *KIF12* expression is induced by the hepatocyte nuclear factor (*HNF1*) α / β [63], which in turn is inactivated by fatty acids [64]. *HNF1* α , is an activator encoded by the most frequently mutated gene in human monogenic diabetes (MODY3) [65]. *KIF12* has been demonstrated to mediate an antioxidant cascade in beta-cells in mice, and *KIF12* knockout mice have impaired glucose-induced insulin secretion and increased beta-cell oxidative stress [66]. This gene potentially links dietary fatty acids and obesity to increased oxidative stress in beta-cell leading to beta-cell dysfunction and T2D.

EPB41L3 (DAL-1, protein 4.1B) is a member of the membrane/cytoskeleton-associated protein 4.1 superfamily, and has not previously been linked to T2D. In addition to acting as a membrane cytoskeleton component, it is believed to function as a tumor suppressor, and is downregulated in several cancers including non-small cell lung cancer and breast cancer [67]. It has been linked to pancreatic beta-cell carcinoma in mice [68], however its role in the condition in humans remains unknown. *EPB41L3* inhibits cell proliferation and promotes apoptosis by modulating the activity of proteins including protein arginine methyltransferase 3 (*PRMT3*) [67,69]. Interestingly, *PRMT3* has been demonstrated to play a role in mycophenolic acid induced pancreatic beta-cell apoptosis *in vitro* [70].

DTNB is a component of the dystrophin-associated protein complex, which acts as a scaffold for signalling proteins, with abnormalities in this complex leading to muscular dystrophy [71]. It is expressed predominantly in the brain and has not previously been linked with T2D.

Study limitations

The main limitations of this study can be grouped into three categories: population sampling, genetic assay selection, and analytical model selection. With respect to population sampling, the cohort sampled in this study is small relative to recently published exome cohorts where

T2D-associated genes were not identified. Given the small size of the cohort, it is challenging to argue that this study could find an association where larger studies did not. However, a recent population-specific study finding potentially deleterious variants linked to T2D in Greenland [36] suggests that there remain novel and population-specific undiscovered T2D-associated loci. By focusing on low frequency potentially deleterious variants in the exome of Qataris, this study provides additional evidence that population-specific variants are linked to T2D.

With respect to genetic assay selection, exome sequencing was utilized to identify low frequency and population specific variants, based on the knowledge from targeted genotyping that known T2D loci do not replicate in the Qatari population [7], and that Qataris variants are poorly represented on conventional genotyping arrays [11]. While rare variants are also of interest, a larger cohort would have been required to sample these variants with high confidence, as singletons in exomes were enriched with false positives [11]. The focus in this study was on potentially deleterious variants in protein coding exons, as these variants have a clear impact on protein function. While off-target and non-coding (intronic, regulatory) variants represent the majority of common variants linked to T2D [20], and can also be sampled by exome sequencing, variant detection false positive rates increase outside target regions and require stringent filters that can produce false negatives in regions of interest [72]. Additionally, this study excluded common variants, contrary to the common-disease-common-variant hypothesis that the same variants and the same genes are associated with T2D in every human population [73]. Given the evolutionary history of human populations, including the isolated evolution of the Arab/Bedouin (Q1) and Persian/South Asian (Q2) Qatari sub-populations (which represent the majority of Qataris in this study and in the general population) for tens of thousands of years [24], the accumulation of recent variants in the Qatari population that influence metabolism is plausible [74].

The associated genes in this study did not replicate in the T2D-GENES cohort [9], primarily due to the lack of potentially deleterious variants in 5 of the 6 associated genes. However, from a broader perspective there are 3 different ways to consider replication- variant replication, gene replication and pathway replication. The original form of replication is variant replication, where a variant associated with T2D in for example a European population replicates in the Qatar population, a scenario that this prior study failed to observe [7]. The second form of replication is gene replication, where different variants in the same gene lead to the same phenotype- loss of function of a protein. This also was not observed in the current study, due to the great genetic distance between the populations compared, and accumulation of low-frequency variants over time. However, our study did observe pathway replication, whereby various genes in the same pathway were linked to T2D in different studies. One of the genes identified, catenin beta 1 (*CTNNB1*, β -catenin), interacts with the nuclear receptor transcription factor 7-like 2 (*TCF7L2*), variants in which are the most strongly associated with risk of developing T2D worldwide. Thus it appears that Qataris have variants which alter the function pathways associated with T2D, however these variants are located different genes from other populations.

Finally, with respect to analytical model limitations, the current study chose the SKAT method, correcting for population structure using a kinship matrix, and including the covariates of age, gender and BMI. Based on the QQ plots observed, population stratification correction was adequate, and appears to account for both near and distant relationships. The diversity of the Qatari population is a known issue, as the Arab/Bedouin (Q1) and Persian/South Asian (Q2) clusters are characterized by high inbreeding, while the Sub-Saharan African (Q3) cluster exhibits low consanguinity and reflects more recent migration to Qatar [24]. As a further argument supporting the use of kinship matrices for population stratification, our

prior study comparing the Qatari population in comparison to 1000 Genomes [24] reconstructed the demographic history of human migration out-of-Africa using a neighbour-joining tree of the kinship matrix, comparable to the degree of resolution obtained using principal components analysis [75]. In addition, the kinship matrix of exome data for 1376 Qataris enabled reconstruction of extended families among Arab/Bedouin (Q1), Persian/South Asian (Q2), and Sub Saharan African Qataris (Q3), identifying 1st, 2nd, and 3rd degree relationships not known prior to the study [11]. The SKAT model used here was limited with respect to computation, which scaled on the order of the cube of the sample size. Thus, analysis of only the 6 genes significantly associated in Qatar was conducted on the T2D-GENES cohort. Faster versions of SKAT are in development and could be applied in future studies.

Conclusions

The genetic architecture of T2D is heterogeneous, and can be explained by variants across the spectrum of allele frequency and function. Based on prior studies by this group and others, and confirmed in this study, known T2D loci did not replicate in Qatar. The most likely scenario to explain this lack of replication is the unique genetics of the Qatari population, the majority of whom (Q1 and Q2 subpopulations) have lived in genetic isolation for thousands of years practicing within-tribe consanguineous marriage. As a result, risk alleles for T2D were not found to be shared between Qataris and other populations. This study confirms the hypothesis of population-specific T2D risk loci by applying exome sequencing and SKAT analysis, identifying 6 novel genes linked to T2D only in Qatar. Replication in the T2D-GENES cohort failed, primarily due to the lack of potentially deleterious variants in the same genes in this cohort. Future studies in larger cohorts of closely related Arab populations will hopefully confirm the unique architecture of T2D in a region with high disease prevalence.

Supporting information

S1 Table. Demographics of Type 2 diabetes cases and controls.

(PDF)

S2 Table. Results of Type 2 diabetes association analyses conducted on 864 Qataris.

(PDF)

S3 Table. Replication of sequence kernel association test (SKAT) significant genes by burden test in an independent cohort.

(PDF)

S4 Table. Potentially deleterious low frequency snps in Qataris.

(PDF)

S5 Table. Known protein coding loci linked to Type 2 diabetes.

(PDF)

S6 Table. Power calculation for known protein coding loci linked to type 2 diabetes.

(PDF)

S1 Fig. Kinship distribution. In order to account for population structure and relatedness in the analysis, kinship was calculated for each pair of Qataris ($n = 864$ individuals) using KING v.2 [27]. Shown is the frequency distribution of kinship scores, where higher numbers on the x -axis indicate closer relationships between pairs of Qataris. The y -axis shows the number of pairs with each score.

(PDF)

S2 Fig. Principal components analysis of Qatari subjects alone and in combination with 1000 Genomes Phase 3 populations. Principal components (PC) analysis was conducted for **A.** $n = 864$ Qataris [Arab (Q1) in red, Bedouin (Q1) in pink, Persian (Q2) in blue, South Asian (Q2) in green, Sub-Saharan African (Q3) in orange]; and for **B.** Qataris in combination with 1000 Genomes Phase 3 populations [5] using PLINK2 [10]. Shown is a plot of PC1 (x -axis) and PC2 (y -axis); 1000 Genomes in squares: Europeans in red, South Asians in blue, East Asians (Q2) in green, Americans in grey, Africans in orange, and Qataris in black circles. (PDF)

S1 File. Supplemental methods.

(PDF)

S2 File. Supplemental references.

(PDF)

Acknowledgments

We would like to thank L. Abu Raddad, Weill Cornell Medical College-Qatar, for help in the management of NPRP 09-741-3-193; and N. Mohamed for editorial assistance.

Author Contributions

Conceptualization: Sarah L. O’Beirne, Juan L. Rodriguez-Flores, Ronald G. Crystal.

Data curation: Sarah L. O’Beirne, Jacqueline Salit, Michelle R. Staudt, Charbel Abi Khalil, Khalid A. Fakhro, Joel A. Malek, Mahmoud Zirie, Amin Jayyousi, Alya Al-Shakaki, Omar Chidiac.

Formal analysis: Sarah L. O’Beirne, Jacqueline Salit, Juan L. Rodriguez-Flores, Monica D. Ramstetter, Ramin Badii, Ajayeb Al-Nabet Al-Marri, Maria J. Chiuchiolo, Jason G. Mezey.

Funding acquisition: Sarah L. O’Beirne, Mai Mahmoud, Dora Stadler, Ronald G. Crystal.

Investigation: Sarah L. O’Beirne, Juan L. Rodriguez-Flores.

Methodology: Sarah L. O’Beirne, Jacqueline Salit, Juan L. Rodriguez-Flores, Joel A. Malek, Jason G. Mezey.

Project administration: Michelle R. Staudt, Charbel Abi Khalil, Khalid A. Fakhro, Amal Robay, Joel A. Malek, Ramin Badii, Abdulbari Bener, Mai Mahmoud, Maria J. Chiuchiolo, Dora Stadler, Ronald G. Crystal.

Resources: Ronald G. Crystal.

Supervision: Charbel Abi Khalil, Khalid A. Fakhro, Amal Robay, Mahmoud Zirie, Amin Jayyousi, Ajayeb Al-Nabet Al-Marri, Ronald G. Crystal.

Validation: Jacqueline Salit, Juan L. Rodriguez-Flores, Monica D. Ramstetter, Ronald G. Crystal.

Writing – original draft: Sarah L. O’Beirne, Juan L. Rodriguez-Flores, Ronald G. Crystal.

Writing – review & editing: Sarah L. O’Beirne, Jacqueline Salit, Juan L. Rodriguez-Flores, Michelle R. Staudt, Ronald G. Crystal.

References

1. International Diabetes Federation (2014) IDF Diabetes Atlas 6th edition. http://www.idf.org/sites/default/files/DA-regional-factsheets-2014_FINAL.pdf.
2. Stumvoll M, Goldstein BJ, van Haefen TW (2005) Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* 365: 1333–1346. [https://doi.org/10.1016/S0140-6736\(05\)61032-X](https://doi.org/10.1016/S0140-6736(05)61032-X) PMID: 15823385
3. McCarthy MI (2010) Genomics, type 2 diabetes, and obesity. *N Engl J Med* 363: 2339–2350. <https://doi.org/10.1056/NEJMra0906948> PMID: 21142536
4. Grarup N, Sandholt CH, Hansen T, Pedersen O (2014) Genetic susceptibility to type 2 diabetes and obesity: from genome-wide association studies to rare variants and beyond. *Diabetologia* 57: 1528–1541. <https://doi.org/10.1007/s00125-014-3270-4> PMID: 24859358
5. Bonnefond A, Froguel P (2015) Rare and common genetic events in type 2 diabetes: what should biologists know? *Cell Metab* 21: 357–368. <https://doi.org/10.1016/j.cmet.2014.12.020> PMID: 25640731
6. Sanghera DK, Blackett PR (2012) Type 2 Diabetes Genetics: Beyond GWAS. *J Diabetes Metab* 3.
7. O'Beirne SL, Salit J, Rodriguez-Flores JL, Staudt MR, Abi Khalil C, Fakhro KA et al (2016) Type 2 diabetes risk allele loci in the Qatari population. *PLOS ONE* (in press).
8. Cauchi S, El AY, Choquet H, Dina C, Kremler F, Weitgasser R et al. (2007) TCF7L2 is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis. *J Mol Med (Berl)* 85: 777–782. <https://doi.org/10.1007/s00109-007-0203-4> PMID: 17476472
9. Flannick J, Fuchsberger C, Mahajan A, Teslovich TM, Agarwala V, Gaulton KJ et al. (2017) Sequence data and association statistics from 12,940 type 2 diabetes cases and controls. *Sci Data* 4: 170179. <https://doi.org/10.1038/sdata.2017.179> PMID: 29257133
10. Standards of medical care in diabetes—2013. *Diabetes Care* 36 Suppl 1: S11–S66.
11. Fakhro KA, Staudt MR, Ramstetter MD, Robay A, Malek JA, Badii R et al. (2016) The Qatar genome: a population-specific tool for precision medicine in the Middle East. *Hum Genome Var* 3: 16016. <https://doi.org/10.1038/hgv.2016.16> PMID: 27408750
12. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43: 491–498. <https://doi.org/10.1038/ng.806> PMID: 21478889
13. Cingolani P, Platts A, Wang IL, Coon M, Nguyen T, Wang L et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6: 80–92. 19695 [pii]; <https://doi.org/10.4161/fly.19695> PMID: 22728672
14. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 46: 310–315. <https://doi.org/10.1038/ng.2892> PMID: 24487276
15. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO et al. (2015) A global reference for human genetic variation. *Nature* 526: 68–74. [nature15393 \[pii\]; https://doi.org/10.1038/nature15393](https://doi.org/10.1038/nature15393) PMID: 26432245
16. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D et al. (2017) The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res* 45: D840–D845. [gkw971 \[pii\]; https://doi.org/10.1093/nar/gkw971](https://doi.org/10.1093/nar/gkw971) PMID: 27899611
17. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158. [btr330 \[pii\]; https://doi.org/10.1093/bioinformatics/btr330](https://doi.org/10.1093/bioinformatics/btr330) PMID: 21653522
18. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA et al. (2012) Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet* 91: 224–237. <https://doi.org/10.1016/j.ajhg.2012.06.007> PMID: 22863193
19. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB et al. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42: 348–354. [ng.548 \[pii\]; https://doi.org/10.1038/ng.548](https://doi.org/10.1038/ng.548) PMID: 20208533
20. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ et al. (2016) The genetic architecture of type 2 diabetes. *Nature*. [nature18642 \[pii\]; https://doi.org/10.1038/nature18642](https://doi.org/10.1038/nature18642) PMID: 27398621
21. Bland JM, Altman DG (1995) Multiple significance tests: the Bonferroni method. *BMJ* 310: 170. PMID: 7833759

22. Rodriguez-Flores JL, Fuller J, Hackett NR, Salit J, Malek JA, Al-Dous E et al. (2012) Exome sequencing of only seven Qataris identifies potentially deleterious variants in the Qatari population. *PLOS ONE* 7: e47614. <https://doi.org/10.1371/journal.pone.0047614> PONE-D-12-10481 [pii]. PMID: 23139751
23. Hunter-Zinck H, Musharoff S, Salit J, Al-Ali KA, Chouchane L, Gohar A et al. (2010) Population genetic structure of the people of Qatar. *Am J Hum Genet* 87: 17–25. S0002-9297(10)00266-1 [pii]; <https://doi.org/10.1016/j.ajhg.2010.05.018> PMID: 20579625
24. Rodriguez-Flores JL, Fakhro K, Agosto-Perez F, Ramstetter MD, Arbiza L, Vincent TL et al. (2016) Indigenous Arabs are descendants of the earliest split from ancient Eurasian populations. *Genome Res* 26: 151–162. gr.191478.115 [pii]; <https://doi.org/10.1101/gr.191478.115> PMID: 26728717
25. Hao K, Li C, Rosenow C, Wong WH (2004) Detect and adjust for population stratification in population-based association study using genomic control markers: an application of Affymetrix Genechip Human Mapping 10K array. *Eur J Hum Genet* 12: 1001–1006. <https://doi.org/10.1038/sj.ejhg.5201273> PMID: 15367915
26. Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, Willer CJ et al. (2011) Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* 19: 807–812. [ejhg.2011.39](https://doi.org/10.1038/ejhg.2011.39) PMID: 21407268
27. Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19: 149–150. PMID: 12499305
28. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J et al. (2000) The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 26: 76–80. <https://doi.org/10.1038/79216> PMID: 10973253
29. Dayeh TA, Olsson AH, Volkov P, Almgren P, Ronn T, Ling C (2013) Identification of CpG-SNPs associated with type 2 diabetes and differential DNA methylation in human pancreatic islets. *Diabetologia* 56: 1036–1046. <https://doi.org/10.1007/s00125-012-2815-7> PMID: 23462794
30. Imamura M, Maeda S, Yamauchi T, Hara K, Yasuda K, Morizono T et al. (2012) A single-nucleotide polymorphism in ANK1 is associated with susceptibility to type 2 diabetes in Japanese populations. *Hum Mol Genet* 21: 3042–3049. dds113 [pii]; <https://doi.org/10.1093/hmg/dds113> PMID: 22456796
31. Sakai K, Imamura M, Tanaka Y, Iwata M, Hirose H, Kaku K et al. (2013) Replication study for the association of 9 East Asian GWAS-derived loci with susceptibility to type 2 diabetes in a Japanese population. *PLOS ONE* 8: e76317. <https://doi.org/10.1371/journal.pone.0076317> PONE-D-13-17352 [pii]. PMID: 24086726
32. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H et al. (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 316: 1336–1341. 1142364 [pii]; <https://doi.org/10.1126/science.1142364> PMID: 17463249
33. Badii R, Bener A, Zirie M, Al-Rikabi A, Simsek M, Al-Hamaq AO et al. (2008) Lack of association between the Pro12Ala polymorphism of the PPAR-gamma 2 gene and type 2 diabetes mellitus in the Qatari consanguineous population. *Acta Diabetol* 45: 15–21. <https://doi.org/10.1007/s00592-007-0013-8> PMID: 17805473
34. Billings LK, Florez JC (2010) The genetics of type 2 diabetes: what have we learned from GWAS? *Ann N Y Acad Sci* 1212: 59–77. <https://doi.org/10.1111/j.1749-6632.2010.05838.x> PMID: 21091714
35. Prasad RB, Groop L (2015) Genetics of type 2 diabetes-pitfalls and possibilities. *Genes (Basel)* 6: 87–123. genes6010087 [pii]; <https://doi.org/10.3390/genes6010087> PMID: 25774817
36. Moltke I, Grarup N, Jorgensen ME, Bjerregaard P, Treebak JT, Fumagalli M et al. (2014) A common Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2 diabetes. *Nature* 512: 190–193. nature13425 [pii]; <https://doi.org/10.1038/nature13425> PMID: 25043022
37. Hara K, Shojima N, Hosoe J, Kadowaki T (2014) Genetic architecture of type 2 diabetes. *Biochem Biophys Res Commun* 452: 213–220. <https://doi.org/10.1016/j.bbrc.2014.08.012> PMID: 25111817
38. Almawi WY, Nemr R, Keleshian SH, Echtay A, Saldanha FL, AlDoseri FA et al. (2013) A replication study of 19 GWAS-validated type 2 diabetes at-risk variants in the Lebanese population. *Diabetes Res Clin Pract* 102: 117–122. <https://doi.org/10.1016/j.diabres.2013.09.001> PMID: 24145053
39. Al-Daghri NM, Alkharfy KM, Alokail MS, Alenad AM, Al-Attas OS, Mohammed AK et al. (2014) Assessing the contribution of 38 genetic loci to the risk of type 2 diabetes in the Saudi Arabian Population. *Clin Endocrinol (Oxf)* 80: 532–537.
40. Cauchi S, Ezzidi I, El AY, Mtiraoui N, Chaieb L, Salah D et al. (2012) European genetic variants associated with type 2 diabetes in North African Arabs. *Diabetes Metab* 38: 316–323. <https://doi.org/10.1016/j.diabet.2012.02.003> PMID: 22463974
41. Alsmadi O, Al-Rubeaan K, Mohamed G, Alkayal F, Al-Saud H, Al-Saud NA et al. (2008) Weak or no association of TCF7L2 variants with Type 2 diabetes risk in an Arab population. *BMC Med Genet* 9: 72. <https://doi.org/10.1186/1471-2350-9-72> PMID: 18655717

42. Ezzidi I, Mtiraoui N, Cauchi S, Vaillant E, Dechaume A, Chaieb M et al. (2009) Contribution of type 2 diabetes associated loci in the Arabic population from Tunisia: a case-control study. *BMC Med Genet* 10: 33. <https://doi.org/10.1186/1471-2350-10-33> PMID: 19368707
43. Tomei S, Mamtani R, Al AR, Elkum N, Abdulmalik M, Ismail A et al. (2015) Obesity susceptibility loci in Qataris, a highly consanguineous Arabian population. *J Transl Med* 13: 119. <https://doi.org/10.1186/s12967-015-0459-3> PMID: 25890290
44. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH et al. (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 11: 446–450. <https://doi.org/10.1038/nrg2809> PMID: 20479774
45. Li B, Leal SM (2008) Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 83: 311–321. <https://doi.org/10.1016/j.ajhg.2008.06.024> PMID: 18691683
46. McCormack SE, Grant SF (2015) Allelic expression imbalance: tipping the scales to elucidate the function of type 2 diabetes-associated loci. *Diabetes* 64: 1102–1104. <https://doi.org/10.2337/db14-1836> PMID: 25805763
47. Huyghe JR, Jackson AU, Fogarty MP, Buchkovich ML, Stancakova A, Stringham HM et al. (2013) Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat Genet* 45: 197–201. <https://doi.org/10.1038/ng.2507> PMID: 23263489
48. White J, Mazzeu JF, Hoischen A, Jhangiani SN, Gambin T, Alcino MC et al. (2015) DVL1 frameshift mutations clustering in the penultimate exon cause autosomal-dominant Robinow syndrome. *Am J Hum Genet* 96: 612–622. S0002-9297(15)00068-3 [pii]; <https://doi.org/10.1016/j.ajhg.2015.02.015> PMID: 25817016
49. Dorus S, Vallender EJ, Evans PD, Anderson JR, Gilbert SL, Mahowald M et al. (2004) Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. *Cell* 119: 1027–1040. S0092867404011432 [pii]; <https://doi.org/10.1016/j.cell.2004.11.040> PMID: 15620360
50. Hedman AK, Zilmer M, Sundstrom J, Lind L, Ingelsson E (2016) DNA methylation patterns associated with oxidative stress in an ageing population. *BMC Med Genomics* 9: 72. <https://doi.org/10.1186/s12920-016-0235-0> [pii]. PMID: 27884142
51. Yang X, Jansson PA, Nagaev I, Jack MM, Carvalho E, Sunnerhagen KS et al. (2004) Evidence of impaired adipogenesis in insulin resistance. *Biochem Biophys Res Commun* 317: 1045–1051. <https://doi.org/10.1016/j.bbrc.2004.03.152> S0006291X04006527 [pii]. PMID: 15094374
52. Brown AC, Olver WI, Donnelly CJ, May ME, Naggert JK, Shaffer DJ et al. (2005) Searching QTL by gene expression: analysis of diabetes. *BMC Genet* 6: 12. 1471-2156-6-12 [pii]; <https://doi.org/10.1186/1471-2156-6-12> PMID: 15760467
53. Smith U (2007) TCF7L2 and type 2 diabetes—we WNT to know. *Diabetologia* 50: 5–7. <https://doi.org/10.1007/s00125-006-0521-z> PMID: 17102948
54. MacDonald BT, Tamai K, He X (2009) Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev Cell* 17: 9–26. S1534-5807(09)00257-3 [pii]; <https://doi.org/10.1016/j.devcel.2009.06.016> PMID: 19619488
55. Figeac F, Uzan B, Faro M, Chelali N, Portha B, Movassat J (2010) Neonatal growth and regeneration of beta-cells are regulated by the Wnt/beta-catenin signaling in normal and diabetic rats. *Am J Physiol Endocrinol Metab* 298: E245–E256. ajpendo.00538.2009 [pii]; <https://doi.org/10.1152/ajpendo.00538.2009> PMID: 19920216
56. Kim W, Shin YK, Kim BJ, Egan JM (2010) Notch signaling in pancreatic endocrine cell and diabetes. *Biochem Biophys Res Commun* 392: 247–251. S0006-291X(09)02488-7 [pii]; <https://doi.org/10.1016/j.bbrc.2009.12.115> PMID: 20035712
57. Apelqvist A, Li H, Sommer L, Beatus P, Anderson DJ, Honjo T et al. (1999) Notch signalling controls pancreatic cell differentiation. *Nature* 400: 877–881. <https://doi.org/10.1038/23716> PMID: 10476967
58. Heremans Y, Van De Casteele M, in't VP, Gradwohl G, Serup P, Madsen O et al. (2002) Recapitulation of embryonic neuroendocrine differentiation in adult human pancreatic duct cells expressing neurogenin 3. *J Cell Biol* 159: 303–312. <https://doi.org/10.1083/jcb.200203074> jcb.200203074 [pii]. PMID: 12403815
59. Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, Kim CE et al. (2011) A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet* 7: e1002293. <https://doi.org/10.1371/journal.pgen.1002293> PGENETICS-D-10-00576 [pii]. PMID: 21980299
60. D'Souza DM, Zhou S, Rebalka IA, MacDonald B, Moradi J, Krause MP et al. (2016) Decreased Satellite Cell Number and Function in Humans and Mice With Type 1 Diabetes Is the Result of Altered Notch Signaling. *Diabetes* 65: 3053–3061. db15-1577 [pii]; <https://doi.org/10.2337/db15-1577> PMID: 27335233

61. Seog DH, Lee DH, Lee SK (2004) Molecular motor proteins of the kinesin superfamily proteins (KIFs): structure, cargo and disease. *J Korean Med Sci* 19: 1–7. <https://doi.org/10.3346/jkms.2004.19.1.1> PMID: 14966333
62. Lee Y, Hirose H, Ohneda M, Johnson JH, McGarry JD, Unger RH (1994) Beta-cell lipotoxicity in the pathogenesis of non-insulin-dependent diabetes mellitus of obese rats: impairment in adipocyte-beta-cell relationships. *Proc Natl Acad Sci U S A* 91: 10878–10882. PMID: 7971976
63. Gong Y, Ma Z, Patel V, Fischer E, Hiesberger T, Pontoglio M et al. (2009) HNF-1beta regulates transcription of the PKD modifier gene Kif12. *J Am Soc Nephrol* 20: 41–47. <https://doi.org/10.1681/ASN.2008020238> PMID: 19005009
64. Johnstone KA, Diakogiannaki E, Dhayal S, Morgan NG, Harries LW (2011) Dysregulation of Hnf1b gene expression in cultured beta-cells in response to cytotoxic fatty acid. *JOP* 12: 6–10. v12i01a02 [pii]. PMID: 21206094
65. Yamagata K, Oda N, Kaisaki PJ, Menzel S, Furuta H, Vaxillaire M et al. (1996) Mutations in the hepatocyte nuclear factor-1alpha gene in maturity-onset diabetes of the young (MODY3). *Nature* 384: 455–458. <https://doi.org/10.1038/384455a0> PMID: 8945470
66. Yang W, Tanaka Y, Bundo M, Hirokawa N (2014) Antioxidant signaling involving the microtubule motor KIF12 is an intracellular target of nutrition excess in beta cells. *Dev Cell* 31: 202–214. <https://doi.org/10.1016/j.devcel.2014.08.028> PMID: 25373778
67. Bernkopf DB, Williams ED (2008) Potential role of EPB41L3 (protein 4.1B/Dal-1) as a target for treatment of advanced prostate cancer. *Expert Opin Ther Targets* 12: 845–853. <https://doi.org/10.1517/14728222.12.7.845> PMID: 18554153
68. Terada N, Ohno N, Yamakawa H, Baba T, Fujii Y, Christofori G et al. (2003) Protein 4.1B in mouse islets of Langerhans and beta-cell tumorigenesis. *Histochem Cell Biol* 120: 277–283. <https://doi.org/10.1007/s00418-003-0573-9> PMID: 14574582
69. Singh V, Miranda TB, Jiang W, Frankel A, Roemer ME, Robb VA et al. (2004) DAL-1/4.1B tumor suppressor interacts with protein arginine N-methyltransferase 3 (PRMT3) and inhibits its ability to methylate substrates in vitro and in vivo. *Oncogene* 23: 7761–7771. <https://doi.org/10.1038/sj.onc.1208057> [pii]. PMID: 15334060
70. Huh KH, Cho Y, Kim BS, Joo DJ, Kim MS, Kim YS (2014) PRMT3: new binding molecule to RhoGDI-alpha during mycophenolic acid-induced beta-cell death. *Transplant Proc* 46: 1229–1232. S0041-1345(13)01412-7 [pii]; <https://doi.org/10.1016/j.transproceed.2013.12.016> PMID: 24815167
71. Blake DJ, Nawrotzki R, Loh NY, Gorecki DC, Davies KE (1998) beta-dystrobrevin, a member of the dystrophin-related protein family. *Proc Natl Acad Sci U S A* 95: 241–246. PMID: 9419360
72. Samuels DC, Han L, Li J, Quanguh S, Clark TA, Shyr Y et al. (2013) Finding the lost treasures in exome sequencing data. *Trends Genet* 29: 593–599. S0168-9525(13)00127-3 [pii]; <https://doi.org/10.1016/j.tig.2013.07.006> PMID: 23972387
73. Schork NJ, Murray SS, Frazer KA, Topol EJ (2009) Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev* 19: 212–219. S0959-437X(09)00088-4 [pii]; <https://doi.org/10.1016/j.gde.2009.04.010> PMID: 19481926
74. Pritchard JK (2001) Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 69: 124–137. S0002-9297(07)61452-9 [pii]; <https://doi.org/10.1086/321272> PMID: 11404818
75. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38: 904–909. <https://doi.org/10.1038/ng1847> PMID: 16862161