GWAS studies of SNPs related to type 2 diabetes in a Mongolian population sample in China

Haihua Bai, Haiping Liu, Suyalatu Suyalatu, Ming Chen, Qizhu Wu

Inner Mongolia University for the Nationalities, Tongliao, Inner Mongolia 028000, China.
hh_bai@163.com

Vladimir Babenko, Olga Posukh, Ludmila Osipova, Yuriy Orlov

Institute of Cytology and Genetics, Lavrentyeva 10, 630090, Novosibirsk, Russia,
{bob,posukh,ludos,orlov}@bionet.nsc.ru

Large scale genome wide association studies (GWAS) have identified approximately 70 single nucleotide polymorphisms (SNPs) conferring susceptibility to type 2 diabetes (T2D). T2D is a complex disease hall marked by insulin resistance and pancreatic beta-cell dysfunction. With the help of recent advances in genotyping and sequencing technology, GWAS studies have enormously contributed to the identification of susceptibility genes for T2D and many other complex disorders. Approximately 70 loci conferring susceptibility to T2D have been identified to date and substantial ethnic specific differences in genetic architecture underlying T2D has been observed among populations of different ethnic backgrounds [1]. Studying genetics of T2D in a multi-ethnic cohorts has been insightful in both fine-mapping casual variants and identifying new loci (DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, [2], indicating the use of investigating common variants in different ethnic samples [3]. However, most of these loci have not been replicated in a diverse populations and much heterogeneity has been observed across ethnic groups with different genetic backgrounds.

Different selection pressures of different climatic conditions dominate the directions of biological evolution. It will help us to reveal the genetic basis of the Buryats' environmental adaptability by studying the genetic diversity of the Buryats living in the cold region of Siberia permanently. Buryats is an important branch of the Mongolian, and the research of its incidence for type II diabetes and disease risk factors will contribute to perfecting the list of disease-causing genes of Mongolian type II diabetes, and is of important scientific significance in the molecular function and pathogenic mechanism of disease-causing genes.

Meanwhile it has high applied value on gene diagnosis, treatment, and prevention of type II diabetes. We analyzed GWAS identified SNPs in a Mongolian sample of northern China including Buryats representatives (497 diagnosed with T2D and 469 matched controls) for association with T2D and diabetes related quantitative traits.

We collected blood sample of 986 individuals of Mongolian ethnicity from the Inner Mongolia, China. Of them, 511 were diagnosed with T2D and 475 were matched healthy normal controls. T2D cases were registered based on the world health organization criteria and admitted to the affiliated hospital of the Inner Mongolia University for Nationalities. Non-diabetic healthy controls were selected based on matching age, sex and ethnic background from the same region. Aside from the diagnosis of T2D, we collected other diabetes related lipid traits, such as TC, HDL, LDL, and TG. We selected a list of SNPs previously found to be associated with T2D based on NHGRI GWAS catalog (www.genome.gov/gwastudies, November, 2012). We were able to genotype 34 SNPs located in or near 32 candidate genes. Following PCR amplifications, we purified PCR products through 1X Agencourt AMPure XP-Medium beads to finally get mixed Illumina pair-end libraries with insert sizes calculated by Aglient 2100 bio-analyzer (Agilent, USA) and concentrations estimated by Real Time PCR. Sequencing of 144 samples was performed on Illumina MiSeq and the other 643 samples were sequenced on Illumina Hiseq 2500. All sequencing steps were in strict accordance with Illumina recommended protocols.

The final sequencing depth reached > 200X, and the length of pair-end reads is 100bp. Reads with base quality of >=20 and 1 mismatch with adaptor sequences were kept for further analysis. BWA-0.5.9 (http://bio-bwa.sourceforge.net) was used to map all clean reads against the human reference genome of hg19 allowing 3bp mismatches across a single read. samtools's mpileup (SAMtools-0.1.18, http://samtools.sourceforge.net) command was used to obtain raw SNP genotypes [4]. These raw genotypes were further filtered according to the following criteria: SNPs with >=5% of missing call rat across the samples and samples with >=3% of missing genotypes (the latter corresponds to 10% of missing SNP call rate) were removed. At the end, we tested the genotypes of the SNPs for Hardy Weinberg disequilibrium, and dropped the ones that have P value $<1x10^{-6}$ in unaffected individuals. 28

SNPs passed these quality control filtering. Overall genotype call rate is 99.3% or higher across the sample.

We replicated a previous T2D association of rs2237897 near KNNQ1 (OR=1.48; P=0.001) in a Mongolian sample and replicated T2D association of ten other SNPs, namely, rs7578326 (IRS1), rs1531343 (HMGA2), rs8042680 (PRC1), rs7578597 (THADA), rs1333051 (CDKN2), rs6723108 (TMEM163), rs163182 (KCNQ1), rs1387153 (MTNR1B), rs243021 (BCL11A), and rs10229583 (PAX4) in the same sample. Further, we tested four lipids related to diabetes, namely, total cholesterol (TC), high density lipoprotein (HDL), low density lipoprotein (LDL), triglyceride (TG), and showed that risk allele of the strongest T2D associated SNP in our sample, rs757832 (IRS1) to be associated with increased level of TG. We observed significant level of T2D risk allele frequency difference between the Mongolian sample and "1000 genomes" Caucasian sample.

We tested association between candidate SNPs and status of T2D using logistic regression with age, sex and BMI as covariates. The study-wide significance was determined by applying Bonferroni correction using 28 tested SNPs (P value of 1.8×10^{-3}). Diabetes related quantitative trait associations (TC, HDL, LDL, and TG) were assessed similarly using regression analysis with age, sex, BMI, and T2D status as covariates. All quantitative trait measures were normalized by quantile normalization. Statistical tests, including 95% confidence intervals (CI), were performed using EPACTS (http://www.sph.umich.edu/csg/kang/epacts/). Difference of allele frequency between the Mongolian sample (healthy controls) and other "1000 genomes" (www.1000genomes.org) sample of interest was estimated by fst option of plink software (plink1.9).

After rigorous sample and marker level quality control filtering, genotypes of 28 SNPs on 966 individuals (including 497 with T2D cases and 469 non-diabetic ethnically matched controls) were kept for subsequent analyses. Overall, consistent with previous studies, T2D cases in current study have higher TC, TG, LDL-C values compared to controls, and have comparable HDL-C values with the controls, indicating TC, GT, and LDL-C are among risk factors. We replicated a previous T2D association of rs2237897 near KCNQ1 (OR=1.48; P=0.001) in our sample. In addition, we replicated associations of ten previously identified

SNPs in a different population, namely, rs7578326 (IRS1), rs1531343 (HMGA2), rs8042680 (PRC1), rs7578597 (THADA), rs1333051 (CDKN2), rs6723108 (TMEM163), rs163182 (KCNQ1), rs1387153 (MTNR1B), rs243021 (BCL11A), and rs10229583 (PAX4)

We chose to examine the association of 34 GWAS SNPs previously identified in European and East Asian populations, with susceptibility to T2D in a Mongolian population in China. The study confirmed the T2D association of rs2237897 in KCNQ1 reported in multiple populations, including samples of European, Mexican, Chinese, Japanese and Mongolian ethnic backgrounds [5]. We also replicated T2D association of two SNPs in another East Asian sample, namely, rs163182 (KCNQ1), rs10229583 (PAX4) in Chinese samples [6]. In addition, our study replicated T2D association of eight additional GWAS SNPs in the Mongolian sample.

This study was supported by the National Natural Science Foundation of China. (81160101,81060098) and RFBR (15-54-53091).

1. J. M. Keaton et al. (2014) A comparison of type 2 diabetes risk allele load between African Americans and European Americans, Hum Genet., **133**(12):1487-95.

2. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, A. Mahajan et al. (2014) Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility, *Nature Genetics*, **46**(3), 234–244.

3. H. Bai et al. (2014) The genome of a Mongolian individual reveals the genetic imprints of Mongolians on modern human populations, *Genome Biol Evol.*, **6**(12):3122-36.

4. H. Li & R. Durbin (2011) Inference of human population history from individual whole-genome sequences. *Nature*, **475**(7357), 493–496.

5. K. Hara et al. (2014) Genome-wide association study identifies three novel loci for type 2 diabetes, *Human Molecular Genetics*, **23**(1), 239–246.

6. R. Ma et al. (2013) Genome-wide association study in a Chinese population identifies a susceptibility locus for type 2 diabetes at 7q32 near PAX4, *Diabetologia*, **56**(6), 1291–1305.