Research paper

RNA-seq analysis of the transcriptome of the liver of cynomolgus monkeys with type 2 diabetes

Xinyu Li[a,b,1], Zijing Lin[a,1], Xiaorong Zhan[a,*], Jie Gao[a], Lijie Sun[a], Yan Cao[a], Hui Qiu[a]

[a] Department of Endocrinology, First Affiliated Hospital Harbin Medical University, Harbin, Heilongjiang, PR China
[b] Department of Pharmacology, The State-Province Key Laboratories of Biomedicine-Pharmaceutics of China, Key Laboratory of Cardiovascular Research, Ministry of Education, Harbin Medical University, Harbin, Heilongjiang, PR China

ARTICLE INFO

ABSTRACT

Genetic and environmental factors such as high-fat diet are involved in the development of type 2 diabetes mellitus (T2DM). Cynomolgus monkey shares similar genetic makeup, tissue structures, physiology and metabolic function to human. This study aimed to establish T2DM model in cynomolgus monkey and compare expression profiles of hepatic genes and their associated pathways in normal cynomolgus monkeys and those with T2DM. We employed RNA-seq technique and identified 1451 differentially expressed genes (DEGs) with a false discovery rate (FDR) of 0.1% between normal and T2DM animals. KEGG pathway analysis revealed that DEGs were associated with 12 KEGG pathways ($P < 0.05$). Two of these pathways were associated with metabolism and five were related to immunity. Unexpected, we found ECM-receptor interaction pathway. In conclusion, our data suggest that three major pathways may be implicated in the development of T2DM, including steroid biosynthesis, immune response and ECM. Further characterization of these pathways may provide new targets for the prevention and therapy of T2DM.

1. Introduction

Type 2 diabetes mellitus (T2DM) is a multifactorial disease caused by a complex interaction between environmental and genetic factors. A diet high in sucrose and fat is thought to be an environmental factor that leads to the development of obesity and T2DM by impairing the action of insulin on the liver (Esposito et al., 2010; Walker et al., 2007; Storlien et al., 1988). Rodents and humans have significant differences in characteristics and lifestyles, although the rodents are fed a diet similar to that of humans (Rees and Alcolado, 2005). Monkeys have similar biological characteristics and lifestyles as humans because of the evolutionary relationship between primates and humans (Yan et al., 2011). Cynomolgus monkeys are one of the most widely used non-human primate experimental animals, their genetic makeup indicates highly similar tissue structures, physiology and metabolic functions to those of humans, and they share high similarity in disease-associated genes and drug targets (Yan et al., 2011).

Cynomolgus monkey is an ideal model for T2DM research because the incidence of spontaneous diabetes in cynomolgus monkey is similar to that of human, and it presents insulin resistance, abnormal glucose metabolism and hyperinsulinemia (Yue et al., 2016; Zhou et al., 2014). In addition, cynomolgus monkey has a longer life cycle, allowing us to adequately investigate the abnormalities in glucose and lipid metabolism, pathological changes in the pancreas, liver, kidney, and various other complications caused by T2DM (Cefalu et al., 2004).

The liver plays a pivotal role in energy metabolism. Under the control of hormones, especially insulin, the liver stores or releases glucose as needed by the body. Assessing hepatic insulin resistance is usually synonymous with measuring hepatic glucose production (Kotronen et al., 2008). By focusing on the liver as an important target organ in T2DM, novel disease-associated genes can be evaluated. RNA-seq is a new technique to analyze changes in gene expression across the entire transcriptome and has been applied to a rapidly increasing number of organisms (Wang et al., 2009).

In this study we established T2DM model in cynomolgus monkey, and employed RNA-seq technique to identify differentially expressed genes (DEGs) in liver transcriptome of T2DM cynomolgus monkeys. Furthermore, we described Gene Ontology and pathway enrichment analysis of DEGs and their associated over-represented biological KEGG pathways.

2. Materials and methods

2.1. Animals

140 adult male cynomolgus monkeys, age 9–11 years old, were housed continuously at the Southern China Primate Research Center accredited by the Chinese Association for Accreditation of Laboratory Animal Care, and all procedures were approved by the Animal Care and Use Committee of the National Institute on Aging Intramural Program.

2.2. T2DM cynomolgus monkey model

During baseline assessments, all monkeys were maintained on standard cynomolgus monkey chow. After baseline assessment, the monkeys were randomized into two groups: a healthy standard diet (SD; 20 monkeys) or a high-fat diet (HFD; 120 monkeys). The healthy standard diet was a purified diet consisting of 13% kcal of fat. The high-fat diet was a specially formulated purified ingredient diet with 42% kcal of fat. Monkeys received two meals per day at estimated ad libitum levels for 15 months, and water was always available. The 120 monkeys fed with HFD were subsequently screened with an oral glucose tolerance test (OGTT), referring to the Human's Standard set by World Health Organization (WHO), and 13 of the monkeys were diagnosed with diabetes.

2.3. Groups

In the normal control group (NC group, n = 3), three monkeys were randomly selected from 20 healthy monkeys being fed SD. From the group of 13 monkeys with T2DM induced by HFD (T2DM group, n = 3), three monkeys were randomly selected.

2.4. Measurements of metabolic status

After 14–16 h of fasting, the animals were deeply anesthetized with pentobarbital (Sigma-Aldrich, MO, USA), and body weight was measured. Blood samples were collected to assess metabolic status, including the levels of blood glucose, plasma triglycerides, CHOL, HDL, LDL, and serum-free fatty acids, by using a Hitachi 7600 Biochemical Analyzer (Hitachi, Tokyo, Japan). The homeostatic model assessment of insulin resistance (HOMA-IR) was calculated according to the formula glucose (mmol/l) × insulin (μU/ml) / 22.5.

2.5. cDNA library preparation and Illumina sequencing for transcriptome analysis

A portion of the liver was removed and weighed. Approximately 100 mg liver samples were dissected, rinsed in RNase-free phosphate buffer, snap-frozen in liquid nitrogen, and stored at −80 °C. Total RNA was extracted using TRIzol reagent (Invitrogen, USA) following the manufacturer's protocol. The integrity of total RNA was examined by 1% agarose gel electrophoresis. In addition, the spectrophotometric absorbances of RNA samples at 230, 260 and 280 nm were used to determine the purity and concentration of total RNA. The samples for transcriptome analysis were prepared using Illumina TruSeq RNA Sample Prep Kit following the manufacturer's recommendations. Briefly, mRNA was at equimolar concentration (500 ng of the original total RNA for each liver) from each group (a mixture of RNA from three monkeys from T2DM group in equal proportions or a mixture of RNA from three monkeys from control group in equal proportions) using oligo (dT) magnetic beads. Using a fragmentation buffer, the mRNA was fragmented into short fragments of approximately 200 bp, and then the first strand cDNA was synthesized with random hexamer priming using the mRNA fragments as templates. The double-stranded cDNAs were purified using a QiaQuick PCR extraction kit (Qiagen) and eluted with elution buffer (EB) for end repair and poly(A) addition. Finally,

sequencing adapters were ligated onto the 5′ and 3′ ends of the fragments. The fragments were purified by agarose gel electrophoresis and enriched by PCR amplification to create a cDNA library.

The cDNA library was sequenced on the Illumina sequencing platform (HiSeq™ 2000). After removing the sequencing adapters and low-complexity reads, all of the RNA-seq reads were mapped to the reference genome using SOAPaligner/SOAP2. No > 5 mismatches were allowed in the alignment, which was specified by parameter –v in the program. The annotation database Ensembl Genes v67 was used as a reference. The gene expression level was calculated using the RPKM method (reads per kilobase transcriptome per million mapped reads) (Mortazavi et al., 2008) using the following formula:

$$RPKM = \frac{10^6 C}{NL/10^3}$$

Given the expression of gene A, C is the number of reads that are uniquely aligned to gene A, N is the total number of reads that are uniquely aligned to all genes, and L is the number of bases on gene A. The RPKM method is able to eliminate the influence of different gene lengths and sequencing discrepancies when calculating gene expression levels.

2.6. Differentially expressed gene (DEG) analysis

Differentially expressed genes between T2DM and control samples were identified by using the significance of digital gene expression profiles (Audic and Claverie, 1997). The false discovery rate (FDR) was controlled using the Benjamini and Hochberg algorithm (Benjamini and Hochberg, 1995). Genes were considered to be differentially expressed based on the following criteria: FDR ≤ 0.001.

2.7. GO functional enrichment and KEGG pathway analysis of DEGs

Functional enrichment was performed using the human genome as the reference set and the hypergeometric test to calculate the enrichment significance. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways or Gene Ontology (GO) terms were considered. If the whole genome had a total of N genes, of which K were involved in the functional category under investigation, and the set of DEGs for analysis had a total of M genes, of which x were involved in the same function category, then the P value could be calculated to evaluate the enrichment significance for that function category as follows:

$$P = 1 - F(x|\,N, K, M) = 1 - \sum_{t=0}^{x} \frac{\binom{K}{t}\binom{N-K}{M-t}}{\binom{N}{M}}$$

Then, the enriched significance P value was adjusted using the Benjamini and Hochberg algorithm (Benjamini and Hochberg, 1995). Finally, KEGG pathways or GO terms with adjusted P values < 0.05 and including at least two differentially expressed genes were considered.

2.8. Quantitative real-time PCR

cDNA was synthesized using 1.5 μg of extracted total RNA and a High Capacity RNA-to-cDNA Kit (Applied Biosystems, Foster City, CA, USA). Quantitative real-time polymerase chain reaction (qRT-PCR) was performed using an ABI 7300 Real-Time PCR system (Applied Biosystems) and DyNAmo HS SYBR Green qPCR kit (Finnzymes, Finland). Amplification was carried out for 1 cycle at 95 °C for 5 min followed by 40 cycles at 95 °C for 15 s, annealing temperature for 15 s, and 72 °C for 30 s, then 1 cycle of 72 °C for 10 min. The primer sequences and annealing temperature were listed in Table 1. qRT-PCR for all target genes was performed using two different reference genes, β-actin (ACTB) and glyceraldehyde-3-phosphate dehydrogenase