Gene 539 (2014) 68–74

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Constructing regulatory networks to identify biomarkers for insulin resistance

Yuxin Yang [a], Yi Wang [b], Kai Zhou [c], An Hong [a,*]

[a] Jinan University, College of Life Science and Technology, 601 Huangpu Blvd, Guangzhou, Guangdong 510632, China
[b] United States Department of Agriculture-Agricultural Research Service, Western Regional Research Center, Albany, CA 94710, USA
[c] The Affiliated Hospital of Luzhou Medical College, 25 Taiping Ave., Luzhou, Sichuan 646000, China

ARTICLE INFO

Article history:
Received 10 July 2013
Received in revised form 21 January 2014
Accepted 25 January 2014
Available online 8 February 2014

Keywords:
Regulatory impact factor
Co-expression
Transcription factor
Microarray

ABSTRACT

Insulin resistance (IR) is a physiological condition in which cells fail to respond to the insulin hormone. Despite advances in the diagnosis and treatment of IR, novel molecular targets are still needed to improve the accuracy of diagnosis and the outcomes of therapy. Here, we present a systems approach to identify molecular biomarkers for IR. We downloaded the gene expression profile of IR from the Gene Expression Omnibus (GEO), generated a regulatory network by mapping co-expressed genes to transcription factors (TFs) and calculated the regulatory impact factor of each transcription factor. Finally, we selected a list of potential molecular targets that could be used as therapeutic targets or diagnostic biomarkers, including ETS1, AR, ESR1 and Myc. Our studies identified multiple TFs that could play an important role in the pathogenesis of IR and provided a systems understanding of the potential relationships among these genes. Our study has the potential to aid in the understanding of IR and provides a basis for IR biomarker discovery.

1. Introduction

Insulin regulates glucose uptake and circulating free fatty acid (FFA) concentrations. Insulin resistance (IR) is defined as an impaired sensitivity to insulin in its main target organs: adipose tissue, the liver and muscle. IR is primarily manifested in the skeletal muscle and adipose tissue, which fail to take up adequate glucose. Elevated glucose production in the liver also occurs, and IR increases circulating FFA concentrations and causes ectopic fat accumulation. In recent years, increasing evidence has emerged that obesity is associated with inflammation that is causally involved in the development of IR. Most importantly, IR has been established as a precursor of type 2 diabetes (T2D) (Harris, 1995; Lyssenko et al., 2008) and cardiovascular disease (Ferrannini et al., 2007; Lima et al., 2009; Reaven, 2008). Along with hyperinsulinemia and β-cell dysfunction, IR is a major pathophysiological determinant of dysglycemia (impaired fasting glycemia (IFG) and impaired glucose tolerance (IGT)) and T2D (Abdul-Ghani and DeFronzo, 2009; Defronzo, 2009). Conditions of high cardiovascular disease (CVD) risk, such as hypertension, dyslipidemia and atherosclerosis, have also been associated with IR (Biddinger et al., 2008; Mulvihill et al., 2011).

Although a variety of methods for measuring IR have been developed, there is a need for new biological markers that can be used to detect IR. Current methods include the gold-standard hyperinsulinemic euglycemic clamp (HI clamp), the insulin tolerance test, steady state plasma glucose (SSPG) following fixed somatostatin/glucose/insulin infusions and modeling based on the oral glucose tolerance test (OGTT) or the frequently sampled intravenous glucose tolerance test (FSIVGTT) (Bergman, 2007). However, these procedures are mostly confined to clinical research settings due to cost and time constraints. Fasting insulin and derived indices (HOMA, QUICKI) have been widely used (Muniyappa et al., 2008), but a lack of insulin measurement standardization strongly limits their accuracy and has prevented their adoption in routine clinical practice. In fact, in recent years, many insulin resistance-related biomarkers have been confirmed, such as a-hydroxybutyrate (a-HB) (Gall et al., 2010), growth and differentiation factor-15 (GDF-15) (Vila et al., 2011), soluble CD163 (Parkner et al., 2012), adiponectin (Lu et al., 2008), circulating fatty acid synthase (FASN) (Fernandez-Real et al., 2010) and YKL-40 (Rathcke et al.,

Abbreviations: TFs, transcription factors; GEO, Gene Expression Omnibus; FFA, free fatty acid; T2D, type 2 diabetes; IFG, impaired fasting glycemia; IGT, impaired glucose tolerance; CVD, cardiovascular diseases; HI clamp, hyperinsulinemic euglycemic clamp; SSPG, steady state plasma glucose; OGTT, oral glucose tolerance test; FSIVGTT, frequently sampled intravenous glucose tolerance test; a-HB, a-hydroxybutyrate; GDF-15, growth and differentiation factor-15; FASN, fatty acid synthase; IS, insulin sensitive; FDR, false discovery rates; NCBI, National Center for Biotechnology Information; DAVID, The Database for Annotation, Visualization and Integrated Discovery; KEGG, Kyoto Encyclopedia of Genes and Genomes; ETS1, v-ets erythroblastosis virus E26 oncogene homolog 1; AR, androgen receptor; ESR1, estrogen receptor 1; MYC, v-myc myelocytomatosis viral oncogene homolog; NAFLD, nonalcoholic fatty liver disease; EPCs, endothelial progenitor cells.

* Corresponding author.
E-mail address: tha@jnu.edu.cn (A. Hong).

http://dx.doi.org/10.1016/j.gene.2014.01.061

2006). These biomarkers not only help researchers to understand the mechanism of insulin resistance but also serve as signs for identifying individuals at risk of progression to insulin resistance. These biomarkers have the potential to provide simple and effective strategies for disease prevention and patient monitoring.

In this paper, we introduce a systems approach that can be used for the discovery of new molecular biomarkers for IR. Based on gene expression data, we reconstructed a regulatory network and selected a list of potential molecular targets that could be used for the discovery of therapeutic targets and diagnostic biomarkers. These targets may also provide novel insights into IR pathogenesis.

2. Methods

2.1. Affymetrix microarray analysis

We obtained gene expression datasets (Table 1) from the NCBI (National Center for Biotechnology Information) GEO (Gene Expression Omnibus) (http://www.ncbi.nlm.nih.gov/geo/) database. By using "insulin resistance" as a search term, we found 57 human expression datasets in the database through 2013. We chose three datasets for our analysis based on three criteria. First, these datasets contain both insulin-resistant and insulin-sensitive samples. Second, the three datasets examine different tissues. Finally, each of the three datasets has a sufficient number of insulin-resistant samples. In these datasets, we considered only the samples that are insulin sensitive (IS) or insulin resistant (IR) and come from different human organs. The original expression datasets from all of the chips were processed into expression estimates using the RMA method (robust multi-array average) with the default settings as implemented in Bioconductor. Next, a linear model was constructed. The differentially expressed genes (DEG) were extracted by using a t-test with a p-value cutoff of 0.05.

2.2. Differential co-expression analysis

To determine the correlations of gene pairs in the datasets, we used the CoExpress tool (http://www.bioinformatics.lu/CoExpress/), which is a user-friendly software tool for the interactive comparison of expression profiles. CoExpress can be used to build pairwise gene co-expression matrices.

For each co-expression gene pair, we also used the DiffCorr package in R (Fukushima, 2012) to export a list of significantly differential correlations as a text file. This package can calculate the differences in the correlations, the corresponding p-values and the results of Fisher's z-test while controlling false discovery rates (FDR).

2.3. Construction of the IR regulatory network

To reconstruct an IR regulatory network from the co-expressed gene pairs, we first downloaded the set of human transcriptional regulation interactions from HTRIdb (Bovolenta et al., 2012). HTRIdb is a repository of experimentally verified interactions between human TFs and their respective target genes. We constructed an IR regulatory network model based on the regulatory interactions between the co-expressed gene pairs.

Table 1
Gene expression datasets used in this study.

Sample origin	Sample type and numbers	Raw data source ID	Reference
Skeletal muscle	20 IS, 20 IR	GSE22309	Wu et al. (2007)
Adipose tissue	20 IS, 19 IR	GSE20950	Hardy et al. (2011)
Hepatokines	7 IS, 10 IR	GSE23343	Misu et al. (2010)

2.4. Computation of the regulatory impact factor

The regulatory impact factor (RIF) appears to be a robust and valuable methodology for identifying the regulators that show the highest evidence of contributing to differential expression under two different biological conditions. The RIF is a metric assigned to each TF that combines the change in co-expression of the TF and its potential targets. The RIF is computed as described by Reverter et al. (2010).

2.5. Pathway enrichment analysis

To facilitate the functional annotation and analysis of large lists of genes, we used DAVID (The Database for Annotation, Visualization and Integrated Discovery) for KEGG (Kyoto Encyclopedia of Genes and Genomes) term enrichment analysis. DAVID identifies the canonical pathways that are associated with a given list of genes by calculating the hypergeometric test p-value for the probability of an association between this set of genes and a canonical pathway (Huang da et al., 2009). A p-value of less than 0.05 and a count greater than 2 were chosen as the cut-off criteria. We calculated the p-values and adjusted the raw p-values to obtain FDR by using the Benjamini–Hochberg method for multiple-testing correction.

3. Results

3.1. Identification of differentially expressed genes in IR

To identify co-expressed gene pairs involved in IR in different tissues, we downloaded the publicly available microarray datasets GSE22309, GSE20950 and GSE23343 from the GEO database and used the CoExpress tool to identify co-expressed gene pairs between IR and normal samples. We found 1775 genes with 173,206 co-expressed gene pairs in skeletal muscle, 3774 genes with 596,396 co-expressed gene pairs in adipose tissue and 10,089 genes with 411,194 co-expressed gene pairs in hepatokines. A total of 355 genes were identified in all three tissues (Fig. 1). We calculated the statistical significance of gene overlap between each pair of groups using the hypergeometric distribution. The p-values were far less than 0.05, indicating that the overlap between these sets does not occur by chance.

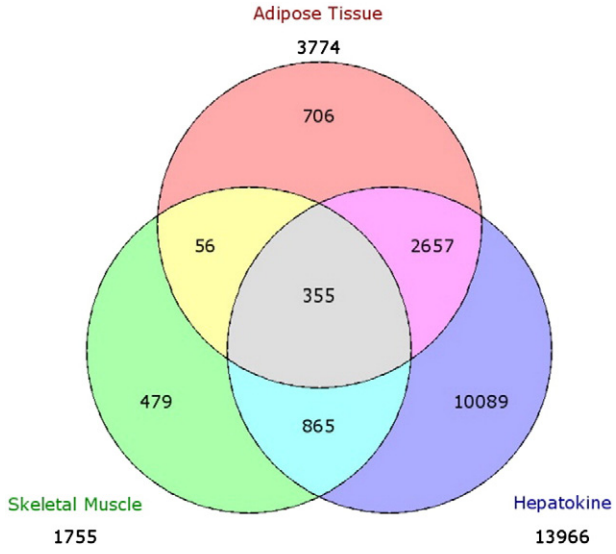

Fig. 1. Venn diagram that shows the co-expression genes in the three datasets.