ABSTRACT

As the predominant group inhabiting the human gastrointestinal tract, bifidobacteria play a vital role in human nutrition, therapeutics, and health by shaping and maintaining the gut ecosystem, reducing blood cholesterol, and promoting the supply of nutrients. The interaction between bacterial cells and human intestinal epithelial cell lines has been studied for decades in an attempt to understand the mechanisms of action. These studies, however, have been limited by lack of genomic and proteomic database to aid in achieving comprehensive understanding of these mechanisms at molecular levels. Microarray data (GSE: 74119) coupled with isobaric tags for relative and absolute quantitation (iTRAQ) were performed to detect differentially expressed genes and proteins in HT-29 cells after incubation with *Bifidobacterium bifidum*. Real-time quantitative PCR, gene ontology, and Kyoto Encyclopedia of Genes and Genomes analyses were further conducted for mRNA validation, functional annotation, and pathway identification, respectively. According to the results of microarray, 1,717 differentially expressed genes, including 1,693 upregulated and 24 downregulated genes, were selected and classified by the gene ontology database. The iTRAQ analysis identified 43 differentially expressed proteins, where 29 proteins were upregulated and 14 proteins were downregulated. Eighty-two candidate genes showing consistent differences with microarray and iTRAQ were further validated in HT-29 and Caco-2 cells by real-time quantitative PCR. Nine of the top genes showing interesting results with high confidence were further investigated in vivo in mice intestine samples. Integration of genomic and proteomic data provides an approach to identify candidate genes that are more likely to function in ubiquitin-mediated proteolysis, positive regulation of apoptosis, membrane proteins, and transferase catalysis. These findings might contribute to our understanding of molecular mechanisms regulating the interaction between probiotics and intestinal epithelial cell lines.

Key words: HT-29 cells, *Bifidobacterium bifidum*, microarray, iTRAQ

INTRODUCTION

The microbial inhabitants of the gastrointestinal tract constitute the most complex ecosystem in the human body and form a closely integrated unit with host developmental processes (Hooper and Gordon, 2001; McFall-Ngai, 2002; Phillips, 2006; Candela et al., 2008; Yin, 2013). As the predominant group normally inhabiting the gastrointestinal tract (Modler, 1994; Guarner and Malagelada, 2003), *Bifidobacterium* are the most widely used probiotic bacteria (Maepherson and Harris, 2004), representing about 8 to 10% of the normal adult fecal flora and reaching concentrations of $10^{10}$ cfu/g in the intestine (Tannock, 1995; Gibson et al., 2004). During the past 2 decades, accumulating data have shown that *Bifidobacterium* play a vital role in human nutrition, as well as therapeutic and health benefits by shaping and maintaining the gut ecosystem (Chow and Lee, 2008), reducing blood cholesterol (Modler, 1994), preventing diarrheas (Salminen et al., 1996), increasing immunity (Gill et al., 2000; Turroni et al., 2014), and promoting the supply of nutrients. Consequently, *Bifidobacterium* have become an increasingly interesting bacteria for probiotic applications in pharmaceutical and dairy products (Candela et al., 2005).

During the last few years, the use of human intestinal epithelial cell line (IEC) HT-29 has increased dramatically in many research fields including the pharmaceutical sciences (Artursson et al., 2001). The interaction between *Bifidobacterium* and IEC have been studied for decades to understand the mechanisms of action, such as adhesion and colonization properties (Candela et al., 2005; Guglielmetti et al., 2008; Qin et al., 2010; Kainu-
The genome of *Bifidobacterium bifidum* has recently been sequenced and annotated (Turroni et al., 2010; Ferrario et al., 2015), revealing metabolic pathways for host-derived glycan foraging (Turroni et al., 2010); however, there are gaps in information about bifidobacterial physiology and ecology. The molecular mechanisms that allow bifidobacteria to interact with the host are not confined to the regulation of immune-related factors, the integrity of the intestinal mucosal barrier, and competition with binding sites of harmful bacteria. As only a few molecules have been described thus far, more investigations are necessary to elucidate the mechanisms between bifidobacteria and their host, such as the protein expressed by transaldolase gene (O’Connell Motherway et al., 2011; González-Rodríguez et al., 2013; Turroni et al., 2014), antimicrobial properties (Gagnon et al., 2004; Zhang et al., 2010), and effect on expression profiles (Riedel et al., 2006) after interaction of *Bifidobacterium* strains with human IEC. Candela et al. (2008) showed that *Bifidobacterium* can modulate IL-8 production (Candela et al., 2008), which is an acute inflammatory response; however, the mechanism behind this is still not clearly. Therefore, the elucidation of function represents a significant step toward understanding those molecular mechanisms.

Microarray technology offers a high-throughput approach to determine the expression of thousands of positional candidate genes in samples (Buratti et al., 2013; Algahtani et al., 2014), and provides expression levels as absolute values computed as the number of transcripts observed for individual genes. Integration of gene expression information and pathway information will help prioritize positional candidate genes (Hornshøj et al., 2009). However, the regulation at the mRNA level does not necessarily correlate with changes at the protein level (Pawar et al., 2013), and microarrays do not always estimate absolute expression levels accurately (Held et al., 2006; Fu et al., 2009). Hence, proteomics present another powerful tool for global investigation of a great multitude of differentially expressed proteins, and a complementary method to study steady-state gene expression and perturbation-induced changes. An isobaric tags for relative and absolute quantitation (iTRAQ)-based quantitative proteomics approach can achieve a greater depth of proteome coverage, including a broader range of protein classes, and has the potential to reveal underlying molecular mechanisms (Yun et al., 2011; Sun et al., 2012; Tse et al., 2013; Zhang et al., 2014).

To fully understand the function of probiotics, enhancing our comprehensive understanding of bifidobacteria and the role they play in the intestinal ecosystem, we analyzed genomic and proteomic data in enterocyte-like HT-29 cells after interaction with *B. bifidum* by coupling microarray with the iTRAQ-labeling technique. These technologies also provided information on the role of cells on *B. bifidum* influence in IEC, which is still poorly understood. Differentially expressed genes and proteins were picked out in accordance with the ratio of standard and real-time PCR for further validation. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were also conducted for mRNA validation as well as functional annotation and pathway identification, respectively.

**MATERIALS AND METHODS**

### Cells and Animals

Nontransformed human Caucasian colon adenocarcinoma cells HT-29 (ATCC: HTB-38) and Caco-2 (ATCC: HTB 37) were routinely cultured in 75-cm² flasks and maintained in Dulbecco’s modified Eagle’s medium (HyClone Laboratories Inc., Logan, UT) containing 25 mM glucose supplemented with inactivated 10% fetal bovine serum (Hangzhou Sijiqing Bioengineering Material Co. Ltd., Zhejiang, China), 1% sodium pyruvate, and antibiotics (penicillin 100 IU/mL and streptomycin 100 mg/mL) in a 5% CO₂ atmosphere with 90% relative humidity at 37°C. Cells were subcultured after 72 h to 80% confluence. Cells from passage number 37 to 70 were used for the experiments. BALB/c mice (6–8 wk old, 18 ± 2 g) were purchased from the Hebei Municipal Center for Disease Control and Prevention, Shijiazhuang, China.

### Preparation of Samples

*Bifidobacterium bifidum* was isolated from humans and identified as ATCC 29521 by sequencing of 16S rDNA. Before the experiment, this microorganism was recovered through 2 successive transfers in de Man, Rogosa, Sharpe (Qingdao Hope Bio-Technology Co. Ltd., Qingdao City, China) agar and broth medium supplemented with 0.5 g of l-cysteine/L at 37°C for 24 h in an anaerobic incubator. For cell culture assays, the overnight culture of *B. bifidum* was transferred to a 50-mL centrifuge tube and harvested by centrifugation (6,000 × g, 10 min, 4°C), and washed twice with deionized water. The supernatant was gently discarded, and the bacterial pellet was resuspended to 1 × 10⁸ cfu/mL of bacteria per milliliter with Dulbecco’s modified Eagle’s medium for later use.
The HT-29 and Caco-2 cells were seeded in 75-cm² (250-mL) cell culture dishes (Greiner Bio-One, Monroe, NC) at a density of 1 × 10⁶ cells/20 mL of the medium. After 72 h, when a confluence level of 95% was attained, the medium was refreshed. Subsequently, the cells were treated with B. bifidum bacterial suspension (T samples), whereas untreated cells (NT samples) were used as controls. After incubation for 3 h, the cell or intestine was detached with a cell scraper for RNA and total protein extraction.

**RNA Isolation and Microarray Analysis**

According to the manufacturer’s procedure, total RNA was isolated by a routine method which uses a Trizol total RNA reagent (catalog no. DP405-02, Tiangen Biotech, Beijing, China). The total RNA was isolated from monolayered HT-29 and Caco-2 cells; after incubation, the quality of RNA was assessed by absorbance (A) ratios of A₂₆₀ to A₂₈₀ close to 2.

The candidate genes from HT-29 cells after incubation with B. bifidum were analyzed by using the Illumina microarray platform (Illumina, San Diego, CA). Total RNA (300 ng) was amplified once using the MessageAmp II aRNA Kit according to the manual (version 0404; Ambion, Thermo Fisher Scientific, Waltham, MA). Part of the amplified RNA (10 μg) was reverse transcribed into cDNA (another part of the cDNA was saved for later real-time PCR validation), then in vitro transcribed and biotin-labeled to produce cRNA, which was subsequently hybridized to meet the standard of the Human HT-12 v4 Expression BeadChips (Illumina). After washing and scanning on the Illumina BeadArray Reader [Supplemental Figure S1 (http://dx.doi.org/10.3168/jds.2015-10577)], a flowchart of the microarray procedure, background correction and normalization were performed with 974 probes [including a negative control, gene intensity (housekeeping gene) control, hybridization controls, low stringency controls, biotin, and high stringency controls] for quality control. The 936 probes represented low expression of all points in the foreground signal value minus the average value of the background). Subsequent analysis was performed using the ratio method to identify differentially expressed genes. Each gene was plotted as a point on a scatter plot using a gene locus chip. Most of the scatter points located near the diagonal (ratio = 1) and others are away from the diagonal, which is in accord with the basic rule of statistics.

**Functional Annotation and the Analysis of Metabolic Pathways**

The differentially expressed genes were analyzed for membership in specific KEGG biological pathways, using the Web Gestalt resource (web-based gene set analysis toolkit, available at http://bioinfo.vanderbilt.edu/webgestalt; Yamada et al., 2004). The individual genes in that pathway, which are located in the candidate intervals, then become functional positional candidates. We also investigated the interrelatedness of the pathways, as the KEGG pathway database includes multiple references to other KEGG pathways. The entries were grouped into 23 categories.

**Total Protein Extraction**

The cells were detached with a cell scraper for total protein extraction after 3 h of incubation. The samples were homogenized in 500 μL of lysis buffer containing 7 M urea, 2 M thiourea, 1% (wt/vol) ASB-14 (Sigma-Aldrich, St. Louis, MO), 2 mM EDTA, and 1 mM phenylmethylsulfonyl fluoride. The suspension was mixed with 10 mM dithiothreitol after incubation in lysis buffer for 5 min. Thereafter, the Psupernatant was retained after centrifugation (16,018 × g, 10 min, 4°C) and incubation with 10 mM dithiothreitol for 1 h, followed by sonication (15 min). A final concentration of 55 mM iodoacetamide was added to the supernatant and resulting mixture was stored in dark room for 45 min to induce cysteine alkylation. Cold acetone was subsequently added at −20°C for 2 h. The sample pellets were dissolved in 0.5 M tetraethylammonium bromide followed by centrifugation (16,018 × g, 10 min, 4°C). The resulting supernatant was used for quantitative analysis after centrifugation.

**iTRAQ Quantification and Analysis**

Total protein was sent to Beijing Genomics for proteomic analysis and quantification. Up- and down-
regulation at the mRNA level do not necessarily correlate with changes at the protein level, suggesting that the relationship between mRNA and protein was not strictly linear and that they have different regulation mechanisms (de Sousa Abreu et al., 2009; Pawar et al., 2013).

The basic flow of iTRAQ in this experiment (a detailed experimental procedure and reagents can be found online: http://www.absciex.com/Downloads/Literature/mass-spectrometry-4370075B.pdf and http://www.sciex.com/products/standards-and-reagents/itraq-reagents) is shown in Figure 1. In the first step, total protein was extracted from the sample. Second, total protein was processed by the reductive alkylation, to reduce disulfide linkages, benefiting subsequent hydrolysis. Protein concentrations were then determined using the Bradford assay (Bradford, 1976) and qualitatively visualized through standard 1-dimensional gel separation techniques. Protein samples were then run by SDS-PAGE separation. Equimolar protein amounts were then digested by trypsin, and peptides labeled with the iTRAQ reagent. Labeled peptides were combined, and the mixed peptides were separated by strong cation exchange chromatography. The strong cation exchange chromatography fractions were then subjected to LC-MS/MS (Triple TOF5600, Sciex, Framingham, MA) analysis. If losses occurred during analysis, each sample experienced the same loss and the peptide ratios were preserved.

In this experiment, databases were selected using the following principles: for already-sequenced organisms, direct selection of the species database was used; for nonsequenced organisms, the most relevant categories of the sample proteome database were chosen. The resulting de novo peptide sequences were then searched against the [IPI: homo (91,464 sequences), V-3.87; B. bifidum NCBInr (22,631 sequences), txid1681] database manually to validate protein identification by Mascot software (ver. 2.3.02, Matrix Science, Boston, MA). Assessment of protein quality was done by iTRAQ and Triple TOF5600; and error of 0.05 Da was found when matching peptides. For the selection of differentially expressed proteins, we considered the following criteria: (1) the protein must have $P < 0.05$ across 2 independent iTRAQ experiments (Lu et al. 2008), and (2) the proteins must have at least a 1.2-fold change.

### In Vivo Bacterial Assays

After 1 wk of acclimatization to the laboratory environment, mice were randomly divided into 2 groups of 10 each. The mice in the control group were treated by gavage with normal saline at a dose of 20 mL/kg of BW once daily, and mice in the test group were treated

<table>
<thead>
<tr>
<th>KEGG identification</th>
<th>Pathway name</th>
<th>Abbreviation</th>
<th>Count</th>
<th>Fraction</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa05212</td>
<td>Pancreatic cancer</td>
<td>PC</td>
<td>11</td>
<td>0.392857</td>
<td>0.00305</td>
</tr>
<tr>
<td>hsa04912</td>
<td>GnRH signaling pathway</td>
<td>GnRH</td>
<td>13</td>
<td>0.464286</td>
<td>0.003582</td>
</tr>
<tr>
<td>hsa04010</td>
<td>MAPK signaling pathway</td>
<td>MAPK</td>
<td>25</td>
<td>0.892857</td>
<td>0.0039</td>
</tr>
<tr>
<td>hsa05200</td>
<td>Pathways in cancer</td>
<td>CANCER</td>
<td>28</td>
<td>1</td>
<td>0.007453</td>
</tr>
<tr>
<td>hsa05221</td>
<td>Acute myeloid leukemia</td>
<td>AML</td>
<td>9</td>
<td>0.321429</td>
<td>0.008127</td>
</tr>
<tr>
<td>hsa04310</td>
<td>Wnt signaling pathway</td>
<td>Wnt</td>
<td>16</td>
<td>0.571429</td>
<td>0.008823</td>
</tr>
<tr>
<td>hsa04621</td>
<td>NOD-like receptor signaling pathway</td>
<td>NOD</td>
<td>9</td>
<td>0.321429</td>
<td>0.012054</td>
</tr>
<tr>
<td>hsa04664</td>
<td>Fc epsilon RI signaling pathway</td>
<td>FcE</td>
<td>10</td>
<td>0.357143</td>
<td>0.01605</td>
</tr>
<tr>
<td>hsa04666</td>
<td>Fc gamma R-mediated phagocytosis</td>
<td>Fcγ R</td>
<td>11</td>
<td>0.392857</td>
<td>0.020829</td>
</tr>
<tr>
<td>hsa04620</td>
<td>Toll-like receptor signaling pathway</td>
<td>TLR</td>
<td>11</td>
<td>0.392857</td>
<td>0.030451</td>
</tr>
<tr>
<td>hsa04070</td>
<td>Phosphatidylinositol signaling system</td>
<td>PI</td>
<td>9</td>
<td>0.321429</td>
<td>0.032044</td>
</tr>
<tr>
<td>hsa04810</td>
<td>Regulation of actin cytoskeleton</td>
<td>AC</td>
<td>18</td>
<td>0.642857</td>
<td>0.038014</td>
</tr>
<tr>
<td>hsa04660</td>
<td>T-cell receptor signaling pathway</td>
<td>TCR</td>
<td>11</td>
<td>0.392857</td>
<td>0.035225</td>
</tr>
<tr>
<td>hsa04150</td>
<td>mTOR signaling pathway</td>
<td>mTOR</td>
<td>7</td>
<td>0.25</td>
<td>0.035627</td>
</tr>
<tr>
<td>hsa05222</td>
<td>Small cell lung cancer</td>
<td>SCLC</td>
<td>9</td>
<td>0.321429</td>
<td>0.035089</td>
</tr>
<tr>
<td>hsa05211</td>
<td>Renal cell carcinoma</td>
<td>RCC</td>
<td>8</td>
<td>0.285714</td>
<td>0.04187</td>
</tr>
<tr>
<td>hsa04060</td>
<td>Cytokine-cytokine receptor interaction</td>
<td>CCRI</td>
<td>20</td>
<td>0.714286</td>
<td>0.043083</td>
</tr>
<tr>
<td>hsa04210</td>
<td>Apoptosis</td>
<td>APOP</td>
<td>9</td>
<td>0.321429</td>
<td>0.045387</td>
</tr>
<tr>
<td>hsa04910</td>
<td>Insulin signaling pathway</td>
<td>INS</td>
<td>12</td>
<td>0.428571</td>
<td>0.046266</td>
</tr>
<tr>
<td>hsa04510</td>
<td>Focal adhesion</td>
<td>FA</td>
<td>16</td>
<td>0.571429</td>
<td>0.047241</td>
</tr>
<tr>
<td>hsa04370</td>
<td>VEGF signaling pathway</td>
<td>VEGF</td>
<td>8</td>
<td>0.285714</td>
<td>0.047132</td>
</tr>
<tr>
<td>hsa04930</td>
<td>Type II diabetes mellitus</td>
<td>T2DM</td>
<td>6</td>
<td>0.214286</td>
<td>0.048623</td>
</tr>
<tr>
<td>hsa04520</td>
<td>Adherens junction</td>
<td>AJ</td>
<td>8</td>
<td>0.285714</td>
<td>0.049268</td>
</tr>
</tbody>
</table>

1Pathway name abbreviation.
2Count of genes or loci in pathway.
3Fraction is the proportion with respect to the largest pathway (pathways in cancer, n = 28 genes). When groups contain biological replicates, we produced $P$-values using approach by Illumina Custom (normalization and differential analysis; Illumina Inc., San Diego, CA).
by gavage with *B. bifidum* cells at a dose of $5.0 \times 10^8$ cfu/d once daily. Twenty-four hours following the final administration, animals were euthanized with anesthetic, and large intestines were collected immediately and washed 3 times with sterile PBS, followed by RNA extraction. The protocol for the animal experiment...
was approved by Nanchang University Animal Ethical Committee and met all ethical requirements to conduct the experiment (approval number 0064257).

**Real-Time PCR**

The primers used in this study are listed in Supplemental Table S1 (http://dx.doi.org/10.3168/jds.2015-10577), were designed according to ratio of differentially expressed genes and key points in pathway, and were synthesized by Invitrogen China (Shanghai, China). The β-actin genes were used as housekeeping genes for the cDNA extraction from HT-29, Caco-2 cells, and large intestine. If PCR inhibition occurred (no amplification of the internal positive control of a cycle threshold >40), the cDNA samples were diluted 2- to 30-fold in PCR-grade water. If false-negative results were suspected, cDNA samples were purified again (Mérault et al. 2011). Quantitative real-time PCR was performed using a commercial kit (Real Master-Mix, SYBR Green, Tiangen Biotech). Amplification and detection were carried out in 96-well plates with a 7900 HT Fast Real-Time PCR System (Applied Biosystems, Foster City, CA). The PCR was carried out in a final volume of 20 μL. Thermal cycler conditions were 95°C for 3 min, followed by 40 cycles at 95°C for 5 s, and then 60°C for 30 s. Each test was performed in duplicate. Relative mRNA levels were determined by comparative critical threshold quantitative real-time PCR (in a separate tube; Schmittgen and Livak, 2008), as described by Applied Biosystems (2001).

**Bioinformatics Analysis**

Functional enrichment analysis was performed using GO, the pathway analysis was performed by KEGG, and the protein orthologs were identified by Clusters of Orthologous Groups of proteins database (COG; http://www.ncbi.nlm.nih.gov/COG/). Both analyses proved statistically significant, with P-values less than 0.01 and 0.05, respectively.

**RESULTS**

**Microarray Analysis**

Differential genes were selected in accordance with the ratio standard point (fold-change >2 or <0.5; Figure 2). A total of 1,717 differentially expressed genes, including 1,693 upregulated genes and 24 downregulated genes, were chosen and classified using the GO database [with classifications including Ontology: Molecular Function (MF), Cellular Component (CC), and Biological Process (BP)]. The results indicated that 192 genes were classified according to BP, 27 categorized according to CC, and 40 categorized according to MF. Their functions were mainly involved in cellular processes and regulation, cell signal transduction, cellular metabolic process and regulation, gene transcription and expression regulation, biological regulation and synthesis, response to biotic stimulus regulation, and immune system development (data not shown). It is noteworthy that the results contain more than one gene associated with the adhesion and membrane-bound categories of differentially expressed genes (with numbers representing gene location in the cell and database), such as GO:0005925~focal adhesion, GO:0005924~cell-substrate adherens junction, GO:0005604~basement membrane, GO:0005912~adherens junction, GO:0043231~intracellular membrane-bounded organelle, GO:0070161~anchoring junction, GO:0043227~membrane-bounded organelle, and GO:0031258~lamellipodium membrane. This discovery could provide a reference of bifidobacteria adhesion to intestinal epithelial cells and inhibit the adhesion of pathogens at the molecular level.

Differential genes were also classified by KEGG, and the entries were grouped into 23 categories (Table 1). These pathways regulate important processes, such as adherent junction and focal adhesion, a regulation of actin cytoskeleton that likely contribute to the adhesion process. We ranked the pathways by weighing the gene expression count. With the use of this approach, the pathways in cancer (28 genes) and MAPK signaling pathway (25 genes) had the highest rank (data not shown). The adherent junction pathway included 8 genes (PTPBR, TCF7, FYN, BAIAP2, CTNNDI1, S1X2IP, FER, and FARP2), all of which were identified as expressed in HT-29 cells. The KEGG pathways database includes multiple references to other KEGG pathways that are also highly interrelated. Many genes have membership in more than one biological function pathway, for example, PIK3CG involved 17 pathways (pancreatic cancer, pathways in cancer, acute myeloid leukemia), PLD1 (pancreatic cancer, GnrH signaling pathway, pathways in cancer, Fc-gamma R-mediated phagocytosis), and SOCS2 (insulin signaling pathway, type II diabetes mellitus).

**iTRAQ Analysis**

Protein concentrations for T and NT samples were 2.5 and 3.2 μg/μL per Bradford assay, respectively. Then we used the iTRAQ approach to identify differentially expressed proteins in the T samples as compared with NT samples, and the samples were also identified.
by basic information statistics. A total of 145,779 spectra and 7,079 unique spectra were used to assign 4,199 unique peptides, representing 1,758 proteins, as shown in Figure 3A.

Almost 80% of identified proteins had a mass distribution <100 kDa (Figure 3B), and most protein mass distributions ranged from 10 to 80 kDa. Identified proteins accounted for almost 80%, as protein mass and the distribution of peptide length ranged from 5 to 37 in AA, and the highest distribution was at 9 AA compared with all peptides (Figure 3C). Figure 3D shows a trend where most of the identified proteins contain <10 peptides, and the number of proteins decreased as the number of matched peptides increased. Distribution of protein sequence coverage showed that the number of 0 to 5% protein sequence accounted for 51.3% (901/1,758) among the total number of proteins (Figure 3E).

The iTRAQ quantification and LC-MS/MS analysis identified relative changes in total protein for T and NT samples. Using a 1.2-fold increase or decrease in the level of protein abundance as a benchmark for potential physiologically significant changes \( (P < 0.05) \) resulted in 43 downregulated proteins being defined, with 29 being upregulated and 14 being downregulated (Figure 4). Of the differentially expressed proteins, 18/43 (41.9%) were identified by more than 5 unique peptides, 4/43 (9.3%) by 4 unique peptides, 8/43 (18.6%) by 3 unique peptides, 12/43 (27.9%) by 2 unique peptides, and 1/43 (2.3%) by 1 peptide. We also observed and compared up- or downregulation of different proteins between different samples by clustering analysis. Observed data clustered to a similar extent by using the shortest branches, which were connected to the data by Euclidean distance. The 2 data sets have short Euclidean distance, meaning they have similar functions (Supplemental Figure S3; http://dx.doi.org/10.3168/jds.2015-10577). The differentially expressed proteins were then classified using GO, COG, and KEGG pathways.

Functional annotation and GO analysis of the identified differentially expressed proteins were carried out using the NCBI database (http://www.ncbi.nlm.nih.gov/) and Blast-2-GO (https://www.blast2go.com/). To visualize the annotation of gene sets, Web Gene Ontology Annotation Plot (WEGO; http://wego.genomics.org.cn/cgi-bin/wego/index.pl) was performed to plot the distribution of GO annotation (Zhong et al., 2013; Zhang et al., 2014; Figure 5 and Supplemental Figure S4; http://dx.doi.org/10.3168/jds.2015-10577).

In the MF category (Figure 5 and Supplemental Figure
Figure 3. (A) Basic information statistics: x-axis shows the identified category, and the y-axis shows the number. Total spectra is the total number of secondary spectrum; spectra is the matched number of spectra; unique spectra is the spectra number of unique match peptides; peptide is the number of identified peptides; unique peptide is the number of unique peptide sequence; and protein is the number of the identified protein. (B) Protein mass distribution: x-axis shows the identified protein molecular mass (in kDa), and the y-axis shows the distribution of identified proteins. (C) Peptide length distribution: x-axis shows AA residues number of peptides, and the y-axis shows the length of the peptide as a percentage of all peptides. (D) Protein number distribution: x-axis shows the peptides number of identified the protein range, and the y-axis shows the number of proteins. (E) Distribution of protein’s sequence coverage. The graph shows the protein proportion of different coverage, the different colors represent different sequence coverage range, and the percentage of pie chart shows the ratio of the number of proteins in a range of different coverage per the total number of proteins. Color version available online.
differently expressed proteins are mainly associated with binding (47.40%), catalytic activity (29.13%), structural molecule activity (6.04%), enzyme regulator activity (3.94%), and transporter activity (3.86%). The BP category (Figure 6 and Supplemental Figure S4B) indicated that the differential proteins are mainly related to cellular process (12.48%), metabolic process (11.07%), biological regulation (7.40%), single-organism process (7.12%), and response to stimulus (6.12%). The CC category (Figure 5 and Supplemental Figure S4C) showed that differential proteins are mainly involved in cell part (18.87%), cell (18.87%), organelle (16.25%), organelle part (12.54%), and macromolecular complex (8.97%).

The analysis identified a protein orthologs classification via the COG database, allowing us to predict the possible functions of these proteins and find further functional classification. Function class of R (general function prediction only) had the highest number (244/1,583, 15.42%) of proteins, J (translation, ribosomal structure, and biogenesis) had 214 (13.52%) proteins, and N (cell motility) had the lowest number (1) of proteins (Figure 6).

To give a general view of the transcriptome in terms of pathway, pathway-enrichment analysis was conducted by KEGG for differential proteins, and pathway mapping of data revealed that the changing proteins not only belonged to 1 pathway, but 37 differently

Figure 4. Protein ratio distribution for control versus test. A ratio >0 or <0 means up- or downregulated expression, respectively; if no significant change exists in the amount of the same protein from 2 samples when the relative quantification was applied, then the ratio is close to 1 in abundance of protein. The points were marked with dark gray (red; upregulated proteins) and light gray (green; downregulated proteins) when the ratio was >1.2-fold. The dark gray (red) and light gray (green) dots are potentially differentially expressed proteins, which need to be validated by PCR in a future study. Color version available online.

Figure 5. Distribution of the gene ontology (GO) terms for differentially expressed proteins. The GO classification map represents the number or percent of proteins and the distribution of ontology (molecular function, biological process, cellular component). The pie chart (Figure 3E) depicts the categories of regulated proteins identified in HT-29 cells by GO analysis. Color version available online.
expressed proteins resulting in 44 pathways (data not shown). We further identified 5 pathways significantly enriched with association signals at \( P < 0.05 \), including 16 (43.24%) differentially expressed proteins. The pathways encompassed Staphylococcus aureus infection (8.11%), aminoacyl-tRNA biosynthesis (13.51%), systemic lupus erythematosus (8.11%), complement and coagulation cascades (5.41%), and pathogenic Escherichia coli infection (8.11%).

### Analysis of Differentially Expressed Genes with Real-Time PCR in Different Cells

Microarray technology offers a high-throughput approach to determine the expression of thousands of positional candidate genes in T and NT samples. Integration of gene expression information by combining KEGG pathway information with GO analysis can help prioritize the locations of candidate genes. The genomic and proteomic analyses of HT-29 cells yielded 1,717 differentially expressed genes, of which 82 genes were further validated according to key points in pathways by low-density array real-time PCR in HT-29 and Caco-2 cells. All 82 genes were selected because of good expression in both samples, and 79.3% (65/82) of measurements were validated by real-time PCR as similar to the Illumina microarray. As illustrated in Figure 7A to D, the gene expression yields of the HT-29 and Caco-2 cells had consistent changes (up or downregulated) after incubation with \( B. \) bifidum and PBS for 3 h. However, to select the best candidates displaying the most consistent alterations and high specificity, fold change data were analyzed and only the changes validated in more than 50% of the cases were considered, leading to a selection of 30 differential genes with high confidence (data not shown).

We selected 9 interesting genes from the 30 high-confidence genes and investigated these in mice intestine samples, resulting in 8 upregulated genes and 1 downregulated gene (Figure 8; primers used in mice large intestine for 9 differential genes are listed in Supplemental Table S2; http://dx.doi.org/10.3168/jds.2015-10577). Nine gene expression levels were collected for \( USP27X, ALG8, CHN2, GLOD5, FAM13A, ARHGEF6, TNFRSF12A, \) and \( TGFB2 \), all of which were upregulated, and \( LOC440386 \), which was downregulated. The results matched those obtained by cDNA microarray results. As illustrated in Figure 7A to D and Figure 8,

![Figure 6. Clusters of Orthologous Groups (COG) database function classification of cell sequence. The x-axis shows the COG classification entry, and the y-axis shows the number of the corresponding proteins by functional classification. Color version available online.](image-url)
Figure 7. Performance (means ± SD) of selected 82 genes according to key points in the pathway to quantify fold changes by representative relative abundance in HT-29 cells (A, B) and Caco-2 cells (C, D). The housekeeping gene β-actin was used to normalize the input amounts of RNA, and the relative expression ratio for each gene was presented as a log2 value in the histogram compared with the control group by using comparative quantitative real-time PCR. A ratio greater than zero or below zero (<0) indicated upregulation or downregulation of the gene expression, respectively, and the whiskers correspond to 1 SD (A, B, C, D).
the 9 gene expression yields of the mice large intestine were consistent with changes (up or downregulated) in HT-29 and Caco-2 cells by incubation with \textit{B. bifidum} and PBS for 3 h. The results also indicated that the large intestine has a higher expression than HT-29 and Caco-2 cells by data normalization and fold change calculation. We therefore suggested that the large intestine reflects more comprehensive characteristics of intestinal cells.

**DISCUSSION**

Up- or downregulation at the mRNA level does not necessarily correlate with changes at the protein level (Pawar et al., 2013). Therefore, in our study, cDNA microarrays were used for transcript expression and protein abundance data for identical samples was generated using iTRAQ-based proteomics to analyze and detect differentially expressed genes and proteins changes across samples. We found that the degree of detectable overlapping genes was modest by microarray and iTRAQ. This might be in part because some genes without transcript-protein concordance were identified, which may arise from annotation errors or differential regulation of translation, turnover, or alternative splicing (Hornshøj et al., 2009). With microarray data, spot intensities are provided for all genes printed on the microarray slides, regardless of whether the genes are expressed or not (Hornshøj et al., 2009; Pawar et al., 2013). Typically, poor correlation has been reported between transcript and protein levels, with little or no correlation at all in some cases (Gygi et al., 1999; Bitton et al., 2008). Where differences do remain, these are more likely to represent biological rather than technical effects, thus enhancing our ability to identify and pursue genes for which protein expression is modulated post-transcription.

From the results of KEGG, functional annotation, and GO analysis using the 2 technologies, we found that \textit{B. bifidum} affected a wide variety of pathways, reflecting the complexity of their interaction with the host gut. We also found higher degrees of correlation involved in adhesion and biological regulation during the interaction, which agrees with observations from previous studies (Foroni et al., 2011; Turroni et al., 2013). The agreement pathways across these 2 technologies are focal adhesion, adhering junction, type II diabetes mellitus (insulin signaling pathway), and pancreatic cancer. Functional annotation and GO analysis of the 2 technologies identified biological adhesion, biological regulation, cell proliferation, and immune system process in BP; cell, cell part, cell-substrate junction, extracellular matrix part, and membrane-bounded organelle in the CC; and enzyme regulator activity and binding in the MF. This is in agreement with findings at both the transcript and protein level.

Our results demonstrate the efficacy of high-throughput sequencing for measuring gene expression levels. A total of 1,717 differential genes were detectable by transcript-based technologies and 82 genes were further validated with best expression in both samples; thus, up to 79.3% were coincident with Illumina microarray, a validation rate close to other studies (Schröder et al., 2011). Nine interesting genes were further researched by real-time PCR in the mouse large intestine, which involved ubiquitin-mediated proteolysis, positive regulation of apoptosis, membrane proteins, and transferase catalysis. Higher expression levels of genes in microarray represent upregulation in real-time PCR and their lower expression levels represent downregulation. The detailed description about the function of these genes is provided below. Ubiquitin-specific proteases work as a nucleophile to attack lysine-glycine isopeptide bonds within ubiquitinated proteins, and reversible modification of target proteases regulates an assortment of signaling pathways either through proteasomal degradation or by altering the activity or localization (Sowa et al., 2009). Asparagine-linked glycosylation 8 acts as a transmembrane protein (Remminghorst and Rehm, 2006) that catalyzes the transfer of a glycosyl group from one compound (donor) to another (acceptor; Withers and Lougheed, 2001), and has been implicated in the roles of cell surface glycosylation and intercellular recognition. Chimaerin 2 encodes a guanosine triphosphate-metabolizing protein, important in cell proliferation and migration (Leung et al., 1994), which
causes a change in the level or activity of a second messenger or other downstream target, ultimately leading to a change in the function of the cell as a result of a single trigger reaction. A specific guanine nucleotide exchange factor, ARHGEF6, codes for regulators or effectors of small GTPases (Nodé-Langlois et al., 2006) and a form of programmed cell death induced by external or internal signals that trigger the activity of proteolytic caspasases, whose actions dismantle the cell and result in cell death. Rho GTPases, critical regulators of the actin cytoskeleton (Hall, 1998) and associated focal adhesion complexes (Ridley and Hall, 1992), can mediate signal transduction pathway from the external environment and play roles in physiological mechanisms including motility, migration, cell division, invasion, or extension of processes (Burridge and Wennerberg, 2004). Transforming growth factor-beta 2 has multifunctional peptides that regulate proliferation, differentiation, adhesion, and migration (Stein et al., 2014), and other functions in many cell types by transducing their signal through combinations of transmembrane receptors (Rosa et al., 1988; Bradl and Linnington, 1996; Kimura et al., 2013). Tumor necrosis factor receptor superfamily, member 12A (FN14) is an immediate-early response gene located on chromosome 16, expressed in transfected NIH 3T3 fibroblasts, which decreases cellular adhesion to extracellular matrix proteins and inhibits growth and migration in vitro (Meighan-Mantha et al., 1999). The FAM13A gene, which is mapped to chromosome 4q22, has a putative role in signal transduction (Silverman, 2012) and both anti-inflammatory and tumor suppression functions (Young et al., 2011). Therefore, the study of the interaction between the host gut and bifidobacteria function is associated and has been correlated with alterations to intracellular signaling genes. These genes are themselves regulated by nested signaling events, and unravelling the sequence of events that occurs is complicated, but key to establishing the definitive underlying molecular mechanism of action.

Microarray profiling of transcriptomes with large-scale validation is easily carried out by sensitive techniques, such as real-time PCR, which we used in the current study. However, further experiments are needed to understand the biological basis of these differences.

**CONCLUSIONS**

According to microarray and iTRAQ technology, a large number of differential genes and proteins were found when IEC co-incubate with *Bifidobacterium*. After validation in HT-29 and Caco-2 cells and in mouse samples, 9 of the top genes showed interesting results with high confidence. These genes have different functions and belong to different metabolism pathways. Future studies should research the function of each gene in IEC.

**ACKNOWLEDGMENTS**

This project was sponsored by National Natural Science Foundation of China (NSF 31170091, 31260363, 31360377, 81160494, 31000048), and Ganpo Talent 555 Engineering Project of Jiangxi Province (Nanchang City, China).

**REFERENCES**


