Identification of candidate blood biomarkers for the diagnosis of septicaemic melioidosis based on WGCNA

Li Yin*, Yuanyuan Chen*, Tingting Fu, Lin Liu and Qianfeng Xia

Key Laboratory of Tropical Translational Medicine of Ministry of Education, NHC Key Laboratory of Tropical Disease Control, School of Tropical Medicine and The Second Affiliated Hospital, Hainan Medical University, Haikou, PR China

ABSTRACT
Melioidosis is an infectious disease caused by *Burkholderia pseudomallei* (Bp), a gram-negative bacillus. Sepsis is the most prevalent type of melioidosis. Due to factors such as lack of precision and slow presentation of bacterial culture tests, the misdiagnosis rate could exceed 100 per cent. Therefore, more reliable, and adaptable diagnostic methods are urgently needed. Weighted gene co-expression network analysis (WGCNA) was employed to screen the featured modules specially expressed in sepsis patients caused by *Bp*. Two representative co-expression modules were selected to perform gene ontology (GO) and KEGG analysis using ClusterProfiler package based on R language. We found that antigen processing and presentation of exogenous peptide antigen via MHC class I pathway, cytosol to ER transport and cell killing related pathways enriched in darkmagenta module which significantly correlated with the sepsis caused by *Bp*. Eventually, a diagnostic 6-mRNA signature consisting of ASPHD2, LAP3, SEPT4, FAM26F, WARS and LGALS3BP was identified, which could discern the sepsis caused by *Bp* compared with other organisms. This will provide a new insight in screening markers for early detection of sepsis caused by *Bp*, and the interaction between pathogens and hosts. This should shed light on the early detection of *Bp*-caused infectious diseases.

Introduction
Melioidosis is an infectious disease caused by *Burkholderia pseudomallei*, a kind of Gram-negative bacillus. Cases have been reported in different regions around the world, including Thailand, China, and Australia etc. In China, it is mainly distributed in provinces of Hainan, Guangdong and Guangxi. People infected with *B. pseudomallei* are those who have been exposed mainly through inhalation or skin wounds to contaminated soil and water. Sepsis is the most prevalent type of melioidosis and miRNAs are extensively explored in the diagnosis of sepsis as serum sepsis biomarkers [1]. The misdiagnosis rate could reach 80–90% due to factors such as lack of specificity and slow presentation of bacterial culture results [2]. The patients could be misdiagnosed as pulmonary infection, pulmonary tuberculosis, hepatitis, malaria, and general sepsis, etc. Compared with other infections, *B. pseudomallei*-induced sepsis has a high mortality rate. Therefore, better diagnostic tests are needed to improve the earlier diagnosis, initiation of appropriate therapy and increase the survival rate of sepsis caused by *B. pseudomallei*. The ideal biomarkers should be able to rapidly and specifically screen patients with sepsis caused by *B. pseudomallei* from those patients caused by other pathogens.

Microarray has been proved a useful biomedical tool in the identification of biomarkers in many aspects [3,4]. The immune system in the peripheral blood of people infected by pathogenic bacteria will change [5]. So the gene expression profile based on microarray could be applied to explore the dysregulation of genes caused by pathogen.

Weighted gene co-expression network (WGCNA) could be used to find the modules which include highly correlated genes, and the modules could be related to external traits [6]. WGCNA has already shown its advantage in screening tumour-related hub genes [7–9].

In order to explore the genes changed in the peripheral blood of the septicemia patients caused by *B. pseudomallei*, 138 samples divided into four groups were performed bioinformatics analysis.

Materials and methods
Data processing
The dataset (GSE:69528) used in this manuscript was downloaded from the NCBI Gene Expression Omnibus (GEO) with a total of 138 samples [10]. The labelled cRNA of all samples were hybridised to Illumina HumanHT-12 V4.0 expression
BeadChip array. All the 138 samples were divided into four groups including Uninfected type 2 diabetes mellitus (U_DM), Uninfected healthy(U_H), patients with sepsis caused by B. pseudomallei (SB) and patients with sepsis caused by other pathogens(OS). The processed data was downloaded for the next analysis.

**Construction of weighted gene co-expression network and identification of modules related to external traits**

The genes with similar expression profiles could be grouped in one module by WGCNA, and the genes commonly participated in one biological process or pathway. WGCNA is designed to be an unsupervised analysis method which could group genes based on their expression profiles [11,12]. WGCNA package is accessible for free which can be used to find the modules of highly correlated genes [6].

**Functional enrichment analysis**

To explore the enrichment results of the genes within each module identified by WGCNA, the functional enrichment analysis of the biological process (GO:BP) and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) analysis [13–15] was done using the ClusterProfiler package [15]. In our analysis, the cut-off value for p-value and q-value was set to 0.05 and 0.2, respectively. Because of the redundancy of enriched GO terms, the simRel method was used to do the redundancy analysis between GO terms [16]. The final result was visualised by R language based on REVIOG [16–19].

**Results**

**Data processing**

The series expression matrix data of 138 samples were downloaded from NCBI. The four groups of U_DM, U_H, SB and OS contain 27, 28, 29 and 54 samples, respectively. According to the request of input files for WGCNA, the expression file and corresponding traits file was generated respectively.

**Construction of weighted gene co-expression network and identification of modules related to external traits**

The adjacency matrix for a scale-free topology network was defined by choosing the soft threshold power 9 which is performed using the pickSoftThreshold function in WGCNA as shown in Figure 1. This value directly affected the construction of the module and the division of genes within the module. This was the lowest value closest to the scale-free network. In our study, the dynamic tree cut method was
employed to identify these kinds of genes with similar expression patterns as well as their relevant biological processes and pathways. Next, the modules were clustered according to their eigengenes which stands for their correlations. Here, the cut height of 0.2 is chosen to merge the similar modules (Figure 2(A)). The dendrogram was generated using hierarchical clustering, in which the short vertical line corresponded a gene and the branches corresponded the co-expressed genes (Figure 2(B)).

**Correlation between modules and clinic traits**

Genes sharing the similar expression pattern (co-expressed gene) were clustered into the same module. The modules which highly correlates with sepsis caused by *B. pseudomallei* mainly were picked out based on the correlation between MEs (module eigengenes) and external traits as shown in Figure 3. Darkmagenta and darkgrey modules were identified significantly associated with septicaemia caused by *B. pseudomallei*. Module darkmagenta and darkgrey were positively and negatively correlated with SB respectively.

**Functional enrichment analysis**

The genes in dark magenta and dark grey modules were selected to perform the GO:BP and KEGG enrichment analysis using ClusterProfiler package, which was shown in Figure 4. According to the similarity of the GO terms got from ClusterProfiler, the GO term ID and *p*-value was selected to perform the REVIGO analysis with the SimRel method. Then the results were visualised using R language as shown in Figure 5. The child items belonging to the same parent item
Figure 3. Correlation matrix of module eigengene values obtained from WGCNA. Twenty-one modules were identified, and each module eigengene was tested for correlation with trait. Within each cell, upper values are correlation coefficients between module eigengene and the traits; lower values are the corresponding p-value.

Figure 4. Functional enrichment analysis of the genes from dark magenta and dark grey module. (A) The enriched GO:BP terms for dark magenta module. (B) The enriched GO:BP terms for dark grey module. (C) The enriched KEGG terms for dark magenta module. (D) The enriched KEGG terms for dark grey module.
are merged and represented by the most significant item. For dark magenta module, the enriched GO terms included antigen processing and presentation of exogenous peptide antigen via MHC class I, cytosol to ER transport, regulation of amino acid metabolism and multi-organism process, canonical Wnt signalling pathway, cell killing and cytolysis (Figures 4(A) and 5(A)). And the enriched KEGG pathway in dark magenta module included NOD-like receptor signalling pathway, Epstein–Barr virus infection, herpes simplex infection, antigen processing and presentation, primary immunodeficiency, allograft rejection (Figure 4(C)). For dark grey module, the enriched GO:BP terms included gas transport, the regulation of protein catabolism and haemoglobin metabolism (Figures 4(B) and 5(B)). The enriched KEGG
pathways included mitophagy (Figure 4D). Response to interferon-gamma pathway was enriched from the redundancy analysis results (Figure 5A).

The selection of candidate biomarkers related to sepsis caused by B.pseudomallei

R. Pankla et al. identified 37 transcript candidate diagnostic biomarkers from sepsis caused by melioidosis compared with by other organisms [20]. In our study, WGCNA was used to find the genes related closely to sepsis caused by B. pseudomallei. Two modules were identified. Six genes including ASPHD2, LAP3, SEPT4, FAM26F, WARS and LGALS3BP which has been identified in both methods as the candidate biomarkers as shown in Figure 6.

Discussion

The burden of disease caused by melioidosis is seriously underestimated all over the world. There are approximately 165,000 cases of melioidosis in the world each year, of which 89,000 are fatal, resulting in an annual death toll that is comparable to measles and much higher than dengue fever [21]. Melioidosis is a leading cause of community-acquired sepsis [22]. The rate of positive cultures of B. pseudomallei isolated from clinical specimens was low and time-consuming, which could delay diagnosis and treatment [23–25]. So, fast and accurate diagnostic methods of B. pseudomallei are urgently needed. Microarray data have demonstrated its application value in the peripheral blood of many diseases [26–29].

In order to identify the early blood biomarkers differentiating melioidosis from sepsis caused by pathogens, the whole blood signature was used to perform WGCNA analysis in our study. Two modules were identified which closely correlated with sepsis caused by B. pseudomallei. Next, the genes in dark magenta and dark grey module were selected to perform GO:BP and KEGG enrichment analysis. The enrichment results were integrated based on semantic similarity due to the redundancy of GO terms.

As one of two classes of Major histocompatibility complex (MHC), MHC class I present the peptide site including 8–10 amino acids for recognition by the appropriate T-cells. Contrary to MHC class I, MHC class II present antigens derived from extracellular proteins. According to the previous study, elevated mRNA expression related to MHC class II could discriminate septicemic melioidosis from B. pseudomallei [30–32]. So the detecting of the expression of the molecules may be conducive for the diagnosis of melioidosis [20]. In the present study, antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent (GO:0002479) enriched in dark magenta module. This process indicates an exogenous peptide antigen which has been degraded within the cell on its cell surface. In this process, the peptide transports from the cytosol to ER relying on TAP. Exogenous recombinant influenza A virus nucleoprotein has been shown to be processed for presentation of MHC class I antigens to CD8+ cells [32]. Although several mechanisms have been proposed to elucidate the processing of foreign proteins and MHC class I antigen presentation, the details remain unclear [30,31, 33]. At the same time, cytosol to ER transport and ubiquitin-dependent protein catabolic process enriched in sepsis caused by B. pseudomallei too. All of the above showed the importance of MHC class I antigen presentation in the host response to melioidosis. Cell killing pathway enriched in sepsis caused by B. pseudomallei too. Different from cell-autonomous death, cell killing refers to the induction of death of one cell by another cell, which is very important to the human defense system. This process is mediated and performed by natural killer cells (NK cells). NK cells are prone to kill the vulnerable cell which loses their MHC I due to infection or being cancer cells. The role of
MHC class I in the host-pathogen interaction should be explored deeply which could promote vaccine development. In our study, the response to interferon-gamma (IFNγ) pathway enriched too. IFNγ plays a vital in innate and adaptive immunity against bacterial and viral infections by inhibiting viral replication directly. Among the 6 genes overlapping 37 candidate signatures screened by Damien Chaussabel, LAP3 and WARS participated in IFNγ pathway [20]. Recently, it was shown that the galectin 3 binding protein (LGALS3BP, also known as 90K) is a virus-induced protein. It leads to IFN and pro-inflammatory production by binding with TRAF6 and TRAF3 complex as a scaffold [34]. In addition, it could suppress virus replication. It remains unclear how LGALS3BP works during bacterial infections. As far as we know, there has been no reports of ASPHD2 gene. But it has been suggested that it is one of the components of NK cell membrane protein.

Early diagnosis of sepsis infected by B. pseudomallei could save treatment time. The host’s response to pathogens can be obtained by analysing the blood genomic transcripts. And the specific biomarkers could be identified.

Our research provides a new perspective for the screening of markers for early diagnosis of sepsis caused by B. pseudomallei, and the interaction between pathogens and hosts. It will provide a basis for the early diagnosis of diseases infected by B. pseudomallei.

Author contribution
The authors confirm contribution to the paper as follows: study conception and design: Qianfeng Xia and Li Yin; analysis and interpretation of the data: Li Yin, Yuanyuan Chen, Tingting Fu and Lin Liu; the drafting of the paper: Li Yin, Tingting Fu and Lin Liu. All authors reviewed the results and approved the final version of the manuscript.

Disclosure statement
No potential conflict of interest was reported by the author(s).

Funding
This study was supported by High-level Talent Project of Hainan Natural Science Foundation (820RC652), Major Science and Technology Program of Hainan Province (ZDKJ2020003, ZDKJ2021036), the National Natural Science Foundation of China. [81960002], Innovation and Entrepreneurship Training Program of Hainan Medical College [201911810018], Hainan Medical University Program for Talents Introduction [XRC190036].

Data availability statement
The data that support the findings of this study are openly available in GEO at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE13015, reference number (Rungnapa 20)

References
[1] Manetti AC, Maiese A, Paolo MD, et al. MicroRNAs and sepsis-induced cardiac dysfunction: a systematic review. IJMS. 2020; 22(1):321.

Zheng X, Xia Q, Xia L, et al. Endemic melioidosis in Southern China: past and present. Trop Med Infect Dis. 2019;4(1):1–15.

Marzancola MG, Sedighi A, Li PC. DNA microarray-based diagnostics. Methods Mol Biol. 2016;1368:161–178.

Yoo SM, Choi JH, Lee SY, et al. Applications of DNA microarray in disease diagnostics. J Microbiol Biotechnol. 2009;19(7):635–646.

Wonderlich ER, Caroline AL, McMillen CM, et al. Peripheral blood biomarkers of disease outcome in a monkey model of Rift Valley fever encephalitis. J Virol. 2018;92(3):e01662–17

Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

Loesgen DD, Hernandez-Vargas H, Voegele C, et al. Identification of novel long non-coding RNAs deregulated in hepatocellular carcinoma using RNA-sequencing. Oncotarget. 2016;7(22):31862–31877.

He Z, Sun M, Ke Y, et al. Identifying biomarkers of papillary renal cell carcinoma associated with pathological stage by weighted gene co-expression network analysis. Oncotarget. 2017;8(17): 27904–27914.

Liao X, Huang K, Huang R, et al. Genome-scale analysis to identify prognostic markers in patients with early-stage pancreatic ductal adenocarcinoma after pancreatieoduodenectomy. Onco Targets Ther. 2017;10:4493–4506.

Barrett T, Wilhite SE, Ledoux P, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013; 41(Database issue):D991–D995.

Costanzo M, VanderSluis B, Koch EN, et al. A global genetic interaction network maps a wiring diagram of cellular function. Science. 2016;353(6306):aaaf1420–aaaf1420.

Giulietti M, Occhipinti G, Principato G, et al. Weighted gene co-expression network analysis reveals key genes involved in pancreatic ductal adenocarcinoma development. Cell Oncol (Dordr). 2016;39(4):379–388.

Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015;43(Database issue):D1049–D1056.

Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.

Yu G, Wang LG, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16(5):284–287.

Wang JZ, Du Z, Payattakool R, et al. A new method to measure the semantic similarity of GO terms. Bioinformatics. 2007;23(10): 1274–1281.

Jiang JJ, Conrath DW. Semantic similarity based on corpus statistics and lexical taxonomy. 1997;10:1–15.

Resnik P. Semantic similarity in a taxonomy: an information based measure and its application to problems of ambiguity in natural language. 1999;11:95–130.

Supek F, Bosnjak M, Skunca N, et al. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011;6(7): e21800.

Pankla R, Buddhisa S, Berry M, et al. Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. Genome Biol. 2009;10(11):R27.

Limmuthurosatsuk D, Golding N, Dance DA, et al. Predicted global distribution of Burkholderia pseudomallei burden of melioidosis. LID - 15008 [pii]. (2058-5276 (Print))

Kelser EA. Melioidosis: a greater threat than previously suspected? Microbes Infect. 2016;18(11):661–668.

Gohler A, Trung TT, Hopf V, et al. Multitarget quantitative PCR improves detection and predicts cultivability of the pathogen Burkholderia pseudomallei. Appl Environ Microbiol. 2017;83(8): e03212–16.

Lau SK, Siridhar S, Ho CC, et al. Laboratory diagnosis of melioidosis: past, present and future. Exp Biol Med (Maywood). 2015; 240(6):742–751.

Lowe W, March JK, Bunnell AJ, et al. PCR-based methodologies used to detect and differentiate the Burkholderia pseudomallei complex: B. pseudomallei, B. mallei, and B. thailandensis. Curr Issues Mol Biol. 2014;16:23–54.
[26] Navi BB, Mathias R, Sherman CP, et al. Cancer-related ischemic stroke has a distinct blood mRNA expression profile. Stroke. 2019; 50(11):3259–3264.

[27] Sasayama D, Hori H, Nakamura S, et al. Identification of single nucleotide polymorphisms regulating peripheral blood mRNA expression with genome-wide significance: an eQTL study in the Japanese population. PLoS One. 2013;8(1):e54967.

[28] Song L, Du NC, Li FR, et al. Expression and clinical significance of peripheral blood T cell JAK2/STAT3 mRNA in chronic idiopathic thrombocytopenic purpura. Zhongguo Shi Yan Xue Ye Xue Za Zhi. 2018;26(6):1746–1751.

[29] Song Y, Liu Y, Wu P, et al. Genome-wide mRNA expression analysis of peripheral blood from patients with obsessive-compulsive disorder. Sci Rep. 2018;8(1):12583.

[30] Ke Y, McGraw CL, Hunter RL, et al. Nonionic triblock copolymers facilitate delivery of exogenous proteins into the MHC class I and class II processing pathways. Cell Immunol. 1997;176(2):113–121.

[31] Martinez-Kinader B, Lipford GB, Wagner H, et al. Sensitization of MHC class I-restricted T cells to exogenous proteins: evidence for an alternative class I-restricted antigen presentation pathway. Immunology. 1995;86(2):287–295.

[32] Voeten JT, Rimmelzwaan GF, Nieuwkoop NJ, et al. Antigen processing for MHC class I restricted presentation of exogenous influenza a virus nucleoprotein by B-lymphoblastoid cells. Clin Exp Immunol. 2001;125(3):423–431.

[33] Yewdell JW, Norbury CC, Bennink JR. Mechanisms of exogenous antigen presentation by MHC class I molecules in vitro and in vivo: implications for generating CD8+ T cell responses to infectious agents, tumors, transplants, and vaccines. Adv Immunol. 1999;73:1–77. Retrieved from doi: http://www.ncbi.nlm.nih.gov/pubmed/10399005.

[34] Gano X, Zhangchuan X, Feivan D, et al. Inducible LGALS3BP/90K activates antiviral innate immune responses by targeting TRAF6 and TRAF3 complex. PLoS Pathog. 2019;15(8):e1008002.