RSC Advances

1   **RSC Advances**

2   **Full Length Article**

3   **Urinary metabolomics study on an induced-stress rat model using UPLC-QTOF/MS**

4   Yuan-yuan Xie †[a], Li Li †[a, b], Qun Shao[c,*], Yi-ming Wang[a], Qiong-Lin Liang[a], Hui-Yun Zhang[e], Peng

5   Sun[e], Ming-qi Qiao[e], Guo-An Luo[a, d*]

6   [a]*Department of Chemistry, Tsinghua University, Beijing, PR China, 100084.*

7   [b]*The Second College of Clinical Medicine, Guangzhou University of Chinese Medicine, Guangzhou, PR*

8   *China, 510120.*

9   [c]*School of Life Science, University of Bradford, Bradford, West Yorkshire, UK, BD71DP.*

10 [d]*State Key Laboratory for Quality Research in Chinese Medicine, Macau University of Science and*

11 *Technology, Avenida Wai Long, Taipa, Macau, PR China*

12 [e]*Shandong University of Traditional Chinese Medicine, Jinan, PR China*

13

14

15

16 †These authors contributed equally to work.

17 Address correspondence to:

18 * Prof. Guo-an Luo

19 Department of Chemistry, Tsinghua University, Beijing 100084, P.R. China

20 Tel/Fax: +86-10-62781688, E-mail: luoga@mail.tsinghua.edu.cn.

21 * Doctor Qun Shao

22 School of Life Science, University of Bradford, Bradford, West Yorkshire, BD71DP, UK.

23 Tel: +86-44-(0)1274 236041, Fax: +86-44-(0)1274 236155 E-mail: q.shao@bradford.ac.uk

24 **Abstract**

25 A urinary metabolomics method based on ultra-performance liquid chromatography coupled with

26 quadrupole/time-of-flight mass spectrometry (UPLC-QTOF/MS) was employed to investigate the

27 pathogenesis and therapeutic effects of a *Baixiangdan* capsule on rats undergoing electric-induced stress

28 for five days. Multivariate analysis techniques, such as principal component analysis (PCA) and partial least

29 squares-discriminant analysis (PLS-DA), were applied to observe the temporal changes in the metabolic

30 states of the electric-stressed rats visually, as well as the recovering tendency of the rats treated with the

31 *Baixiangdan* capsule. Artificial intelligence technology (artificial neural networks and neurofuzzy logic)

32 was used to identify the potential biomarkers, and the results showed a high overlap with the PLS-DA

33 model. A total of 14 potential biomarkers representing major cause-effect relationships between the

34 variations in the endogenous metabolites and the dynamic pathological processes associated with the

35 stress induced by the electric stimulation were identified, including amino acid metabolites, such as

36 2-aminoadipic acid, hippuric acid, spermine, 4-hydroxyglutamate and L-phenylalanine, in addition to

37 prostaglandin F3a and melatonin. The results indicated that the pathways corresponding to

38 L-phenylalanine, tyrosine, tryptophan, arginine, proline metabolism, pantothenic acid, and coenzyme A

39 synthesis were disturbed in the electric-stressed rats, and the application of the *Baixiangdan* capsule may

40 regulate the aforementioned metabolic pathways back to their initial states. The application of artificial

41 intelligence technologies provided powerful and promising tools to model the complex metabolomic

42 data and to discover hidden knowledge regarding the potential biomarkers associated with the

43 development of disease, which is also suitable for other complex biological data sets.

44 **Keywords:** metabolomics, urine, premenstrual syndrome, UPLC-QTOF/MS, artificial intelligence,

45 biomarkers

46 **1. Introduction**

47 Metabolomics, one of the major platforms in systems biology, is used to study perturbations in response

48 to physiological challenges, toxic insults or disease processes by measuring low-molecular-weight

49 metabolites (<1 kDa) and their dynamic changes in complex biological samples [1-3]. Recently, an

50 increasing number of publications described the application of metabolomics approaches in traditional

51 Chinese medicine (TCM) research, which demonstrates that metabolomics is a powerful tool for

52 assessing the holistic efficacy of TCM formulae because the global metabolic state of an entire organism

53 can be represented *via* a single metabolic profile analysis [4-6]. Normally, information-rich metabolomics

54 data are acquired from high-field nuclear magnetic resonance (NMR) [7], gas chromatography mass

55 spectrometry (GC-MS) [8], or/and UPLC-QTOF/MS [9]. Furthermore, many multivariate analysis

56 techniques, such as principal component analysis (PCA), partial least squares discriminant analysis

57 (PLS-DA), orthogonal partial least squares discriminant analysis (OPLS-DA) and support vector

58 machine recursive feature elimination (SVM-RFE), are commonly used to find informative biomarkers

59 for subsequent studies [10-11].

60 Neurofuzzy logic, which combines the adaptive learning capabilities of artificial neural networks (ANNs)

61 with the generality of representation from fuzzy logic, is one of the artificial intelligence (AI)

62 technologies that had proven to be an effective tool for analysing complex biological data sets [12]. Fuzzy

63 logic modelling implementing the adaptive B-spline modelling of observation data (ASMOD) algorithm

64 can be applied to generate a number of training models and perform training tests to determine which

65 one best fits the data. The quality of the models is assessed using various statistical fitness criteria, e.g.,

66 Akaike's Information Theoretic Criterion (AIC), final prediction error (FPE), cross validation (CV),

67 generalised cross validation (GCV), minimum descriptor length (MDL) and structure risk minimization

68 (SRM). These aim at minimizing a criterion containing two terms—one involving the prediction errors

69   computed in the data set and the other involving the complexity of the structure of the trained models.

70   These training parameters are then investigated to obtain the ideal model with best predictions of the

71   validation data, as well as generating intelligible rules in an "if then" format that explicitly represents the

72   cause-effect relationships contained in the experimental data. During the training process, the

73   improvement of the model training is assessed via these fitness criteria, which is different from the cross

74   validation approach commonly applied in neural networks, where the test data are used. In this method,

75   neural networks are used to optimise certain parameters of the fuzzy systems and automatically extract

76   fuzzy rules from the numerical data. Five back propagation learning algorithms, including Standard

77   Incremental, Standard Batch, RPROP, Quickprop, and Angle Driven Learning, are used to adjust the

78   weights of the network connections during the training. A change in the weights will affect the

79   contribution of each input variable and therefore largely influence the way that a trained network gives

80   predictions [13]. Neurofuzzy logic had been successfully applied in tablet film coatings, pharmaceutical

81   formulations and processing [14]. However, the application of neurofuzzy logic in metabolomics data

82   analysis to discover hidden knowledge regarding potential biomarkers associated with the development

83   of disease remains relatively new.

84   Premenstrual syndrome (PMS), a typical stress-related emotional disease affecting 8 % of women of

85   child bearing age, is a collection of emotional symptoms, with or without physical symptoms, related to

86   a woman's menstrual cycle. Emotional symptoms, referred to as premenstrual dysphoric disorder

87   (PMDD), such as depression and anxiety, must be consistently present to diagnose PMS [15]. Although

88   the duration of PMS symptoms are shorter than depression due to other etiologies, such as severe

89   depression, post-traumatic stress disorder and anxiety disorders, their influences on the quality of life of

90   a patient in the luteal phase can be as great as or worse than other disorders. PMS has attracted much

91   attention in the international medical community, though the exact etiology remains unclear even after

92   more than 40 years of systematic research [16]. The theory of TCM believes that the pathogenesis of PMS

93   is closely related to liver dysfunction, in which liver-Qi invasion syndrome and liver-Qi depression

94   syndrome are the two principal subtypes [17]. Therefore, TCM formulae that function to smooth the liver

95   and regulate vital energy are normally used to relieve the symptoms of PMS [18]. The *Baixiangdan*

96   capsule is a novel modernised composite medicine prepared using Radix Paeoniae alba and Cortex

97   moutan radicis extracts, together with Rhizoma Cyperi volatile oil, which exhibits a favourable efficacy

98   for the treatment of PMS due to liver-Qi invasion syndrome [19].

99   The electrical stimulation of female Sprague-Dawley (SD) rats produces a series of abnormal

100   behavioural and physiological responses similar to the symptoms of liver-Qi invasion syndrome PMS

101   in humans, and it is often used as an animal model to study the pathogenesis of PMS [20]. The feasibility

102   of establishing a model of liver-Qi invasion syndrome PMS in SD rats using electrical stimulation has

103   been proven [21]. Previously, a serum metabolomics approach based on UPLC/QTOF-MS was developed

104   to evaluate the therapeutic effects of the *Baixiangdan* capsule on liver-Qi invasion syndrome PMS in

105   rats. The therapeutic mechanism of the *Baixiangdan* capsule is related to the regulation of metabolism

106   by corticosteroids (e.g., tetrahydrodeoxycorticosterone, 5α-tetrahydrocortisol, epinephrine), oestrogen

107   (e.g., pregnanediol, estrone) and excitatory/ inhibitory amino acid neurotransmitters (e.g., lysine,

108   5-hydroxylysine, acetylcysteine) [22]. In the present study, we applied a urinary metabolomics method to

109   investigate the time-related biochemical abnormalities in liver-Qi invasion syndrome PMS due to

110   electrical stimulation for 5 days and assessed the therapeutic effects of the *Baixiangdan* capsule.

111   Artificial intelligence technology (artificial neural networks and neurofuzzy logic) was used to identify

112   the metabolic pathways and the potential biomarkers related to PMS to achieve the most comprehensive

113   metabolome coverage and provide a more in-depth understanding of the pathophysiological processes of

114   PMS.

115

116   **2. Material and methods**

117   **Chemicals and reagents**

118   HPLC-grade acetonitrile was purchased from J. T. Baker (Phillipsburg, NJ, USA). The following

119   compounds were obtained from Sigma-Aldrich (Louis, Mo, USA): 2-aminoadipic acid, hippuric acid,

120   spermine, 4-hydroxyglutamate, L-phenylalanine, melatonin, L-methionine, proline, genistein and

121   leucine-enkephalin. Ultrapure water (18.2 MΩ) was prepared using a Milli-Q water purification system

122   (Millipore, France). All of the other chemicals that were used were analytical grade.

123   The *Baixiangdan* capsule was a TCM prescription prepared using Radix Paeoniae alba extract, Cortex

124   moutan radices extract and Rhizoma Cyperi volatile oil, which was provided by Shandong Traditional

125   Chinese Medicine University. The preparation process was strictly carried out according to the fixed

126   processing parameters. The *Baixiangdan* capsules used in this study were placed under a careful quality

127   control to ensure their identity throughout all of the experiments. Three representative components

128   (paeoniflorin, paeonol and α-cyperone) were used as quality indicators during the HPLC evaluation [23].

129

130   **Animal handling and sample collection**

131   Healthy and non-pregnant female Sprague-Dawley rats (190-200 g in weight) were supplied by the

132   Experimental Animal Center of Shandong Traditional Chinese Medicine University (serial number

133   SCXK (Lu) 20050015 on the certificate of conformity, Jinan, China). All of the animals were

134   maintained in an environmentally controlled room under a controlled temperature (22–25°C) and

135   relative humidity (50 ± 5 %) on a 12 h light/dark cycle (lights on from 08:00 to 20:00). The experiments

136   were conducted in a specific pathogen free (SPF) grade laboratory according to the guidelines provided

137   in the Guiding Principles for the Care and Use of Laboratory Animals approved by the Committee for

138 Animal Experiments at Shandong Traditional Chinese Medicine University (Jinan, China). The animals

139 were acclimated for 1 week before use. A standard diet and water were provided to the rats *ad libitum*.

140 A total of 30 animals in diestrus and metestrus were selected using vaginal smears together with the

141 behavioural assessment described previously [20]. The animals were randomly divided into the following

142 3 groups with 10 rats in each group: (1) control group (CG), (2) stress group (SG), (3) *Baixiangdan*

143 capsule-dosed group (BCDG). Each rat was maintained in an individual tailor-made cage. The PMS rat

144 model was produced using electrically induced stimuli with a digital pulse stimulator [21]. The SG and

145 BCDG rats were treated with the electrically induced stimuli (0.5 mA pulses at a voltage of 2700~3300

146 V and a pulse width of 0.3 s) continuously for 5 days. Each application of the electric stimuli lasted for

147 5 minutes twice during the day and for 10 minutes three times in the evening. The BCDG rats were

148 administered a water solution of the *Baixiangdan* capsule at a dose of 10 mL/ kg·w·d (1 mL of the

149 solution is equivalent to 1 g of the crude herbs) via an intra-gastric gavage once a day, amounting to

150 eight times the clinical dosage. Meanwhile, the CG and SG rats were administered the same volume of

151 water via oral gavage. The 24-h urine samples were collected over the 5-day electric-stimuli period.

152 The urine samples at the starting point (without electric stimuli, day 0) were collected 24-h prior to the

153 start of the experiment. The collected urine samples were stored at -80°C until the sample preparation.

154 Because of the individual differences between the rats, not all of the rats urinated regularly every 24 h.

155 At the end of the experiment, only 140 urine samples were collected for the UPLC-QTOF/MS analysis.

156

157 **Sample preparation**

158 Prior to the analyses, the samples were thawed at room temperature. The urine samples were

159 centrifuged at 13000 rpm for 20 min at 4°C, then the supernatant was analysed via UPLC-QTOF/MS.

160 Three parallel sample solutions were prepared and analysed for accuracy.

161

162 **UPLC-QTOF/MS Analysis**

163 The chromatographic separation was performed on an ACQUITY UPLC BEH C_{18} column (2.1×100

164 mm, 1.7 μm, Waters Corp, Milford, MA, USA) using a Waters ACQUITY UPLC^{TM} system equipped

165 with a binary solvent delivery system, an auto-sampler, and a PDA detector. The column was

166 maintained at 30°C and eluted at a flow rate of 0.4 mL/min, using a mobile phase of water with 0.2 %

167 (by volume) formic acid (A) and acetonitrile (B). The gradient program was optimised as follows: 0-18

168 min, 0 % B to 35 % B; 18-20 min, 35 % B to 95 % B; 20-22 min, 95 % B; 22-25 min, 95 % B to 0 % B;

169 25-28 min, equilibration with 0 % B. The column eluent was directed to the mass spectrometer without

170 a split.

171 The mass spectrometry was performed on a Waters Q-TOF Premier mass spectrometer (Waters Corp.,

172 Manchester, UK) with the electrospray ionization source (ESI) operation in the positive ion mode ("V"

173 mode of operation). The ESI-MS parameters for LC/TOF-MS were: capillary voltage 3200 V; cone

174 voltage 35 V; nitrogen was used as the drying gas, and the desolvation gas flow rate was set as 700 L/h

175 at a temperature of 350°C; cone gas rate was 50 L/h; source temperature 110°C; The scan time was 0.1s;

176 inter-scan delay was 0.02 s. All of the analyses were acquired using an independent reference

177 lock-mass ion via the LockSpray^{TM} interface to ensure accuracy and reproducibility.

178 Leucine-enkephalin was used as the reference compound (*m/z* 556.2771 for the negative-ion mode) at a

179 concentration of 50 pg/μL and flow rate of 10 μL/min. The data were collected in the centroid mode

180 from *m/z* 50 to *m/z* 1000 using a LockSpray frequency of 10 s, and the data were averaged over 10

181 scans for the correction.

182

183 **Data Processing**

8

184   The data were combined into a single matrix by aligning the peaks with the exact mass / retention time

185   pair (EMRT) from each data file along with their associated intensities using MarkerLynx Applications

186   Manager version 4.1 (Waters Corp., Manchester, UK). The parameters included a retention time (t_R)

187   range from 0 to 28 min, a mass range from 50 to 1000 Da, and the mass tolerance was 0.02 Da. The

188   minimum intensity was set at 15 % of the base peak intensity, the maximum mass per t_R was set at 6,

189   and the t_R tolerance was set at 0.02 min. An original data list was obtained using a database (peak

190   matrix) containing 421 data records (144 data records were obtained from the stress group, 127 data

191   records were obtained from the *Baixiangdan* capsule-dosed group and 150 data records were obtained

192   from the control group) and 4000 independent variables (biochemical substances). Prior to the

193   multivariate statistical analyses, the data from each chromatogram were normalised to a constant

194   integrated intensity relative to the number of peaks to partially compensate for the concentration bias of

195   each sample. The processing of the data normalization had little effect on the conclusion of the

196   trajectory analysis, which aimed to improve the clustering tightness in the PLS-DA model by

197   comparing the results of the area-normalised data model with that of the non-normalised data model

198   (data were not shown). The between-subject data X was then Pareto-scaled to facilitate the analyses of

199   the major effects in the data. Upon grouping the information, the processed original data list was then

200   divided into three datasets and exported and processed via PCA and PLS-DA analyses using the

201   software package SIMCA-P version 11.5 (Umetrics AB, Umeå, Sweden).

202   Two commercial AI software tools representing the two technologies were used in this study:

203   INForm4.3 for the neural networks and FormRules 3.0 for the neurofuzzy logic. Both software

204   packages were provided by Intelligensys Ltd., UK. The algorithm and data processing methods of these

205   two software programs are as follows, and have also been described previously [13-14, 23]. For the data

206   prepossessing of the AI analysis, PCA was initially applied to reduce the dimensions of the SG dataset

207 containing the 144 data records, from 4000 independent variables to 140, according to the significance

208 of the contribution to the PCA model. FormRules 3.0 implements the ASMOD algorithm to generate

209 the neurofuzzy logic model, which enables the discovery of differential features (potential biomarkers).

210 The reduced variable dataset containing only the discovered differential features as independent

211 variables was established during the neurofuzzy logic modelling, and the hidden relationships among

212 the these differential features were also discovered. Structure Risk Minimization (SRM) was used to

213 assess the quality of the models in this study. INForm 4.3, which is embedded with a multi-layer

214 perceptron neural network, was applied to validate the robustness of the discovered differential features

215 by comparing the quality of the models based upon the original dataset and the reduced variable dataset.

216 One of the back-propagation learning algorithms, such as Standard Incremental, Standard Batch,

217 RPROP, Quickprop and Angle Driven Learning, was selected to obtain the optimal prediction accuracy.

218 The work flow of the data analyses using the AI techniques is shown in Figure 1.

219 **Figure 1 The work flow of the data analyses using AI techniques, including reducing the**

220 **variables using a neurofuzzy logic model and the predictive ability assessment using**

221 **a neural network model.**

222

223 **3. Results**

224 **Establishment of the metabolic fingerprints**

225 To optimize the experimental conditions, a pre-investigation had been conducted before the full study.

226 The fingerprints of a small batch of test urinary samples were acquired in both the positive and negative

227 mode. Higher noise and matrix effect in ESI negative mode had been observed. The higher baseline in

228 ESI negative mode led to the neglect of some low abundance metabolites and the concomitance of

229 multiple adduction ions. After considering the maximization of the number of detectable metabolites and

230   the quality of the acquired data, the full-scan detection was eventually set in ESI positive mode. After a

231   careful optimisation of the flow rate and the column temperature for the chromatography and the

232   capillary voltage, flow, and the temperature of the desolvation gas for the mass spectrometry detector,

233   the optimal parameters were fixed as listed in *section 2.4*. As a result, a higher flow rate (0.4 mL/min)

234   was used to achieve higher analysis efficiency on the UPLC column and to reduce the run time.

235   Meanwhile, the tolerance in the backpressure elevation and the effect on the spray and desolvation were

236   also considered. The flow and temperature of the desolvation gas were set at 700 L/h and 350°C,

237   respectively, to remove any redundant solvent resulting from the high flow rate and to improve the

238   efficiency of the desolvation and ionization. Using the optimised conditions, the representative base

239   peak intensity chromatograms of the rat urine obtained in ESI positive mode for the different groups are

240   shown in Figure 2. After completing the processing described in Section 2.5, a list of 4000 compounds

241   was exported for each sample, and the standard quality control (QC) samples were pooled (small

242   aliquots of each biological sample to be studied were pooled and thoroughly mixed). Between each

243   analytical unit of 20 analytes, the QC sample was analysed to provide a robust quality assurance for

244   each metabolic feature that was detected. The precision and repeatability of the UPLC-MS method

245   were validated via a duplicate analysis of six injections of the same QC sample and six parallel samples

246   prepared using the same preparation protocol, respectively. The relative standard deviations of the

247   retention time and area were less than 5.0 %. The resulting data showed that the precision and

248   repeatability of the proposed method were satisfactory for metabolomics analysis.

249

250   Figure 2 Representative base peak intensity chromatogram of the rat urine obtained in ESI positive

251   mode based of UPLC-QTOF/MS. (A) control group (CG); (B) stress group; (C) *Baixiangdan* capsule

252   dosed group (BCDG).

253

254 **Urinary metabolic profiling data processing using PCA and PLS-DA**

255 The PCA and PLS-DA analyses of the dataset containing 144 data records obtained from the SG rats on

256 days 0 (prior to the electrically induced stress), 1, 2, 3, 4 and 5 were performed first. The PCA score

257 plot (Figure S1) showed clear differences between the urine samples collected on days 0, 1, 2, 3, 4, and

258 5, which visualised the general changes in the holistic metabolic profile of the endogenous metabolites

259 during the electric stimulations.

260 The supervised pattern recognition (PLS-DA) was more focused on the actual class discriminating

261 variations compared to the unsupervised approach (PCA). Figure S2 (A) shows the score plot of the

262 PLS-DA model using the dataset from the SG rat urine samples to discriminate between the different

263 days of induced stress, and it is similar to the PCA result. The parameters of this PLS-DA model were

264 $R^2X_{(cum)}$=0.427, $R^2Y_{(cum)}$=0.952, $Q^2Y_{(cum)}$=0.912, which means that 42.7 % of the independent variables

265 were applied to construct the model, 95.2 % of the samples (data records) fit the established discriminant

266 mathematic model, and the prediction accuracy of this model was 91.2 %. After being processed via

267 PLS-DA in SIMCA-P, the mean-centred PLS-DA score plots were generated to trace and compare the

268 dynamic changes in the metabolic events in the rats undergoing electric stimulation for 5 days. In the

269 PLS-DA graph, each spot represented a sample and each assembly of samples indicated a particular

270 metabolic pattern at a different time point. The loci marked by arrows represent the trend of the mean

271 metabolite pattern changes. As shown in Figure S2 (A), the metabolic state of each group on day 1 had

272 deviated from the initial position (day 0, prior to the electrically induced stress), and the greatest

273 difference was observed on day 2, which indicates that in response to the electric stimulation, the

274 metabolism of the endogenous substances and the metabolic profiles of the urine compared to the initial

275 state (day 0) were significantly altered. From day 3 to day 5, the trajectory direction gradually returned

276   to that observed on day 1, indicating the recovery of the disturbed metabolic state. The VIP (variable

277   importance in the projection) value of each variable in the model was ranked according to its

278   contribution to the classification. The VIP list of the retention time-exact mass pairs was obtained from

279   the PLS-DA using SIMCA-P. To select the potential biomarkers worthy of preferential study in the next

280   step, these differential metabolites were validated using Student's *t* test. The critical *p*-value was set to

281   0.05 for the significantly different variables in this study. Following the criteria listed above, 14

282   significantly different endogenous metabolites present in the urine of on the 5th day were selected for

283   further study. The identification of the potential biomarkers was then carried out as follows, and the

284   results are listed in Table 1. The possible elemental compositions of the selected compounds were

285   generated using the software program Masslynx according to the following procedure: the calculated

286   mass, mass deviation (mDa and ppm), double-bond equivalent, formula, and *i*-fit value (the isotopic

287   pattern of the selected ion) were calculated using the selected *m/z* ions. A lower *i*-fit value and smaller

288   mass deviation indicate a more accurate elemental composition. The structural information was

289   obtained by searching freely accessible databases (KEGG (http://www.genome.jp) and HMDB

290   (http://www.hmdb.ca)) using the detected molecular weights and elemental compositions.

291   As a result, 14 potential biomarkers were identified based on the accurate elemental compositions and

292   the retention time and 9 were confirmed using the available reference standards by matching their

293   retention time and accurate mass measurement. Among them, 2-aminoadipic acid (**1**), 5-oxoproline (**2**),

294   shikimate-5-phosphate (**4**), 4-hydroxyglutamate (**5**), hippuric acid (**10**), 5-(2-hydroxyethyl)-4-

295   methyliazole (**11**) and melatonin (**13**) were found to be increased in the urine samples from the

296   electric-stressed rats compared to their initial state. Conversely, prostaglandin F3α (**3**), biocytin (**6**),

297   genistein (**7**), deoxyadenosine (**8**), 6-keto-prostagladin F1α (**9**), 2,3-diaminopropionic acid (**12**) and

298   5-amino-valerate (**14**) were decreased [23].

299 Meanwhile, the MS spectra dataset of the CG rats during the five testing days were also analysed using

300 PLS-DA. Compared to the pathological variations observed in the SG rats, the trajectory of the CG rats

301 was irregular, as shown in supporting information Figure S3, which suggests that the electric stimuli on

302 female rats may lead to systemic metabolic variation. To determine the treatment-related metabolic

303 pattern alterations, another PLS-DA model ($R^2X_{(cum)}$=0.423, $R^2Y_{(cum)}$=0.973, $Q^2Y_{(cum)}$=0.877) was

304 constructed with a dataset containing 127 data records obtained from the BCDG rats. As shown in

305 Figure S2 (B), a classification between different treatment days was clearly achieved, and the trajectory

306 of the metabolic profiles illustrated the temporal metabolic variations in the urine metabolites and

307 exhibited a recovering tendency back to the initial state (day 0) following treatment with the

308 *Baixiangdan* capsule.

309

310 **Feature selection and identification of the significant metabolites using AI technology**

311 Due to the complexity and nonlinearity of metabolomics data, AI technologies provide a meaningful

312 method for the discovery of feature information hidden in data. Neural networks are computational

313 systems capable of mimicking the mechanisms of human learning. They enable the detection of

314 complex relationships between a set of inputs and outputs and estimate the magnitude of the

315 relationships without requiring a mathematical description of how the output is functionally dependent

316 on the input. They are useful for processing unstructured and nonlinear data for the recognition of

317 patterns in high-dimensional data. Neurofuzzy logic is a hybrid AI technology that combines the

318 learning capabilities of neural networks with the generality of fuzzy logic, and it is able to generate

319 knowledge regarding the patterns hidden in data in an interpretable format [13].

320 In this study, the top 140 independent variables were selected from the ranking order generated by the

321 PCA according to the significance of its contribution to the PCA model. A reduced variable dataset was

14

322   then formed, which included the 140 independent variables from the original dataset. Further data

323   mining activities were then conducted using this new dataset, and the ASMOD algorithm was applied

324   to generate the neurofuzzy logic model. A total of 14 independent variables were discovered to be

325   differential features. Therefore, two datasets that included the same data records but different

326   dimensions (number of independent variables), in which one contained the 140 variables and the other

327   contained the 14 selected differential features as independent variables, have been established. Next, a

328   further investigation using a multi-layer perceptron neural network was carried out to validate the

329   robustness of the discovered reduced variable dataset by comparing the quality of the models based on

330   the two established datasets. During the modelling process, the two datasets were both randomly

331   divided into a validation set (28 data records, 20 % of the 144 data records were selected using the

332   "Smart Selection" function in INForm4.3) and a training set (116 data records, the remaining 80 % of

333   the 144 data records). Two neural network models were generated using the two selected training set

334   datasets. Then, the predictabilities of these two neural network models were tested against the validation

335   datasets. The validation R^2 that was computed using the validation dataset was used to evaluate the

336   predictability of the neural network model. As shown in Figures 5, the validation R^2 of the validation

337   dataset containing the 140 independent variables is 0.9506 (Figures 3A), and it is 0.9539 (Figures 3 B)

338   for the 14 discovered differential features (independent variables) after reducing the dimensions using

339   neurofuzzy logic. The similarity between the validation R^2 values indicates that the predictability of the

340   neural network model did not deteriorated by reducing the dimension. The major knowledge of the

341   relationships between the independent variables and the dependent variables (Grouping Information)

342   still remains in the reduced variable dataset. The 14 discovered differential features are sufficient to

343   explain the variability associated with the relationship between the independent and dependent variables

344   (Grouping Information) and to represent the cause-effect relationships between the variations in the

345   endogenous metabolites and the dynamic pathological processes associated with the stress induced by

346   the electric stimulation. Therefore, the 14 differential feature metabolites discovered via AI analysis

347   were considered to be potential biomarkers related to the development of induced stress. They were

348   identified using the methods described in *section 3.3*. As shown in Table 1, eight of the potential

349   biomarkers, including 2-aminoadipic acid (**1**), prostaglandin F3α (**3**), shikimate-5-phosphate (**4**),

350   4-hydroxyglutamate (**5**), genistein (**7**), hippuric acid (**10**), 2,3-diaminopropionic acid (**12**) and

351   melatonin (**13**), were discovered by both the PLS-DA and the AI analysis. The six remaining

352   differential metabolites were only discovered by the AI, of which, spermine (**15**), L-phenylalanine (**16**),

353   pantothenol (**18**) and xanthosine (**20**) were significantly decreased in the electric-stressed rats, whereas

354   proline (**17**) and L-methionine (**19**) were significantly increased.

355   Figure 3 The predictions given by the ANN models generated using a dataset containing various
356   numbers of independent variables. (A) 140 independent variables, (B) 14 independent variables
357   (reduced dimensions).

358   Table 1 Identification of the Significantly Different Endogenous Metabolites in the Model Rats'
359   Urine.

360

361   **4. Discussion**

362   PMS is a typical stress-related emotional disease that affects 8 % of women of child-bearing age.

363   Emotional symptoms, such as anxiety, must be consistently present to diagnose PMS. The electrical

364   stimulation on female SD rats can produce a series of abnormal behavioural and physiological

365   responses that are similar to the emotional symptoms of PMS, including a reduction in exploratory

366   behaviour and plasma hormone level alterations (prolactin, estradiol and progesterone) [20]. Previously, a

367   serum metabolomics approach based on UPLC/QTOF-MS had been developed to evaluate the

368   therapeutic effects of the *Baixiangdan* capsule on liver-Qi invasion syndrome PMS rats [22]. The

369   therapeutic mechanism of the *Baixiangdan* capsule is related to the regulation of the metabolism of

370   corticosteroids, oestrogen and excitatory/inhibitory amino acid neurotransmitters.

371   The present study developed a urinary metabolomics method on the basis of UPLC-QTOF/MS to

372   investigate the temporal variations in the metabolic profiles of rats that underwent electric stimulation

373   in 5 days. AI techniques integrating neurofuzzy logic and neural networks were applied for the first

374   time here to find and understand the correlation of the selected potential biomarkers to the occurrence

375   and development of liver-Qi syndrome PMS induced by electric stimulation. The minimal dataset,

376   containing 14 differential features (metabolites) that are sufficient to explain the variability of the

377   endogenous metabolites associated with the dynamic pathological processes induced by electric

378   stimulation, was obtained using neurofuzzy logic modelling. Therefore, the 14 differential feature

379   metabolites were considered to be potential biomarkers for discriminating the different urine metabolic

380   profiles in different days. Seven sub-models, implying hidden interactions between 14 potential

381   biomarkers, were constructed according to the intelligible rules in an "if then" format explicitly

382   representing the cause-effect relationships contained in the experimental data during the neurofuzzy

383   logic modelling. However, these important correlations among variables are usually neglected in the

384   commonly used multivariate analysis with VIP values as the weight sum of the PLS loadings to

385   evaluate the variable contributions for distinguishing different metabolic states.

386   As shown in Table 1, nine of the fourteen differential endogenous metabolites discovered using the

387   neurofuzzy logic model were found to be monoamine neurotransmitter metabolites, including

388   2-aminoadipic acid (**1**), hippuric acid (**10**), 4-hydroxyglutamate (**5**), 2,3-diaminopropionic acid (**12**),

389   spermine (**15**), L-phenylalanine (**16**), proline (**17**), pantothenol (**18**) and L-methionine (**19**). These

390   results are consistent with previous reports regarding the pathogenesis of PMS, which is related to

391 amino acid metabolism and neural signal transmission.

392 Glutamate is the most abundant fast excitatory neurotransmitter in the mammalian nervous system,

393 which is responsible for mediating a broad range of nervous system functions *via* glutamate receptors.

394 It may be involved in the metabolism of proteins and glucose in brain, as well as promoting oxidation

395 and improving the function of the central nervous system[24]. The previously study have explored the

396 relationship between pathogenesis of PMS and glutamate by determining the concentrations of

397 glutamate in serum and different regions of brain (including hypothalamus, limbic lobe, frontal cortex

398 and hippocampus) using pre-column derivatization HPLC method. The glutamate levels in serum,

399 hypothalamus and limbic lobe decreased significantly in PMS model rats when compared with normal

400 ones, while those in frontal cortex and hippocampus were found to be increasing after model

401 establishment[25].

402 2-aminoadipic acid (**1**) is a primary metabolite in the lysine metabolic pathway, which antagonizes

403 neuro-excitatory activity modulated by the glutamate receptor, N-methyl-D-aspartate (NMDA).

404 Aminoadipic has also been shown to inhibit the production of kynurenic acid, a broad spectrum

405 excitatory amino acid receptor antagonist [26]. The disorder of 2-aminoadipic acid has been associated

406 with varying neurological symptoms[27]. Meanwhile, the metabolism of lysine also rely on the regulation

407 of glutamate receptor, implied that the level of 2-aminoadipic acid in urine should be related to that of

408 glutamate.

409 4-hydroxyglutamate (**5**), an intermediate in the metabolism of gamma-hydroxyglutamic acid.

410 Specifically 4-hydroxyglutamate combines with 2-oxoglutarate to produce 4-hydroxy-2-oxoglutarate

411 and glutamate [28]. Therefore, the level of 4-hydroxyglutamate should also be closely related to that of

412 glutamate. Hippuric acid (**10**) is an acyl glycine formed by the conjugation of benzoic aicd with glycine

413 on the basis of the action of glycine N-acyltransferase. And glycine combines with α-ketoglutaric acid

414   could produce glyoxylic acid and glutamic acid. The up-regulation of these metabolites, including

415   2-aminoadipic acid (**1**), 4-hydroxyglutamate (**5**) and hippuric acid (**10**), in the urine represent an

416   increase in the excitatory amino acid glutamate[29].

417   In addition, shikimate-5-phosphate (**4**), which is the precursor of chorismic acid and tryptophan, was

418   also discovered by the two data mining approaches. Shikimate-5-phosphate has been reported to

419   participate in the metabolism of phenylalanine. Both L-phenylalanine (**16**) and tryptophan are required

420   for the biosynthesis of monoamine neurotransmitters and play an important role in the pathogenesis of

421   emotional disorders. Decreased phenylalanine levels were detected in the urine of the electric-stressed

422   rats, which was in agreement with other reports [13]. However, L-phenylalanine was only discovered via

423   the AI analysis, in addition to spermine (**15**), proline (**17**), pantothenol (**18**), L-methionine (**19**) and

424   xanthosine (**20**). Proline is also a derivative of glutamate, which generates hydroxyproline and then

425   decomposes into 4-hydroxyglutamate, an excitatory amino acid neurotransmitter.

426   Melatonin (**13**), which is involved in the metabolic pathway of 5-HT, was also found using the

427   neurofuzzy logic model. It has been suggested that PMS is related to a systemic imbalance of the

428   neurotransmitter 5-hydroxy tryptamine (5-HT) [30-32]. The emergence of symptoms such as emotional

429   instability, irritability, and anxiety are related to a decrease in 5-HT levels [33]. The increased melatonin

430   in the urine of the stressed rat model indicates the down-regulation of 5-HT. Melatonin exhibits

431   extensive physiological activities, and its daily and seasonal rhythms are considered to be closely

432   related to the functional regulation of immunity and the neuroendocrine and reproductive systems. The

433   biosynthesis of melatonin is also rhythmic, in addition to melatonin precursors and its related synthesis

434   enzymes, e.g., *N*-acetyltransferase, HIOMT, 5-HT [34]. The rhythm of *N*-acetyltransferase and HIOMT

435   exhibit the same tendencies as melatonin, while 5-HT is the opposite [35]. Therefore, a rise in melatonin

436   may be attributed to the premenstrual moods of dysphoria and irritability. Other biomarkers, such as

437   11-epi-prostagladin F2α (**3**) and 6-keto-prostagladin F1α (**9**) have been found to be involved in signal

438   transduction and the regulation of physiological activities, such as the synthesis of lipoproteins and

439   carbohydrates, which are related to the development of stress/emotion-related diseases.
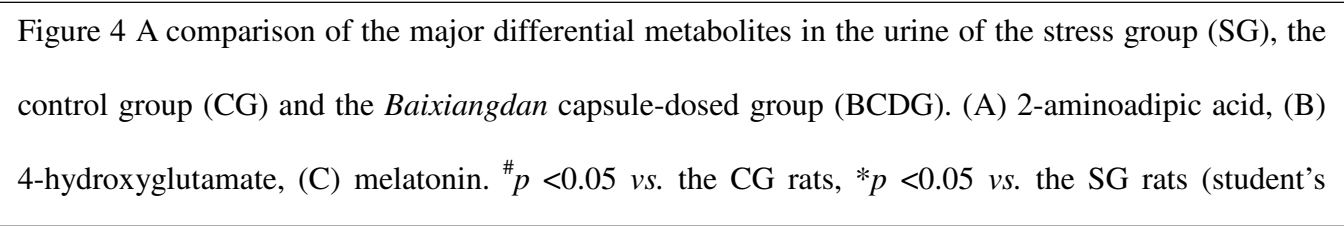
440   Genistein (**7**) was also identified as a differential component, and it has been reported to be related to

441   energy metabolism. Genistein is also a primary component of rat feed, which is made from soybeans.

442   The stress experiment may inevitably cause the loss of appetite; therefore, the reduction of genistein in

443   the urine between the experimental and initial states could be due to various factors.

444   The predictive abilities of these 14 potential biomarkers were then evaluated using a neural network

445   model. When using neural network algorithms, intelligible rules are generated on the basis of "unseen"

446   data to provide accurate predictions. This is different from the cross validation approach that is

447   commonly applied in multivariate analyses and neurofuzzy logic, where the test data are used. Therefore,

448   the evaluated results obtained using neural network modelling are more credible. The 14 potential

449   biomarkers discovered using the AI analysis closely related to the occurrence and development of

450   liver-Qi syndrome PMS, indicating that the AI appeared to be more effective than the PLS-DA analysis

451   for the data mining.

452   Upon analysing the dynamic trajectories of the holistic metabolic profiles for the 5 different days of

453   electric stimulation in the PLS-DA score plots, the greatest difference was observed on day 2, which

454   indicates that as a response to the electric stimulation, the metabolism of the endogenous substances

455   and the metabolic profiles in the urine were significantly altered compared to the initial state. From day

456   3 to day 5, the trajectory direction gradually moved back to that of day 1, meaning that the

457   experimental animals accommodated for the electric stimulation, and the stress states were relieved; the

458   same intensity of stimulation could not cause a similar response. Therefore, with the exception of the

459   fact that urine metabolites are the final products of all physiological and pathological processes, the

460   adaption to the stress state and the potential biomarkers discovered in this study showed significant

461   differences to the ones that were discovered in previous serum metabolomics studies.

462   The *Baixiangdan* capsule is a new TCM prescription that has been used for the treatment of PMS [33-34].

463   Different urine metabolite patterns were observed *via* PLS-DA, implicating the potential efficacy of the

464   *Baixiangdan* capsule on the electric-stress rat model. As shown in Figure 4, the average intensities of

465   2-aminoadipic acid, 4-hydroxyglutamate and melatonin in the urine of the different groups (CG, SG

466   and BCDG) were compared. During the first four days, compared with the levels observed in the CG

467   rats, the levels of the three metabolites in the SG rats gradually rose. However, in the treatment group

468   (BCDG), the levels decreased significantly compared to the SG rats. The results suggest that several

469   Figure 4 A comparison of the major differential metabolites in the urine of the stress group (SG), the

470   control group (CG) and the *Baixiangdan* capsule-dosed group (BCDG). (A) 2-aminoadipic acid, (B)

471   4-hydroxyglutamate, (C) melatonin. [#]p <0.05 *vs.* the CG rats, *p <0.05 *vs.* the SG rats (student's

472   patients suffering from PMS; therefore, relieving the typical psychological symptoms, such as

473   emotional instability, irritability and anxiety.

474

475

476

477

478   In the present study, AI analysis was applied for the discovery of potential biomarkers related to the

479   dynamic pathological processes of liver-Qi invasion syndrome PMS in an induced-stress rat model for

480   the first time. The explored potential biomarkers have been proved to be valuable according the

481   biochemical interpretations referred to the corresponding literatures. However, some remaining

482   questions would still be necessary and essential to regard the roles of obtained biomarkers in certain

483   metabolic pathways, and better understand the exact pathogenesis of PMS. For example, validation of

484   these discovered biomarkers on the basis of biological experimental evidences; illustration on

485   corrections of obtained potential biomarkers in serum/plasma and urine; precise quantitative

486   determinations of potential biomarkers to give rational thresholds for disease diagnosis and efficacy

487   evaluation, *etc.* should be solved in the following investigations. These problems would be solved in

488   the sequential investigations.

489   **5. Conclusions**

490   A urinary metabolomics method based on ultra-performance liquid chromatography coupled with

491   quadrupole/time-of-flight mass spectrometry (UPLC-QTOF/MS) was employed to investigate the

492   pathogenesis and therapeutic effect of the *Baixiangdan* capsule on electric-induced stress in rats for five

493   days. Artificial intelligence technology (artificial neural networks and neurofuzzy logic) was used for

494   the first time for the discovery of the differential metabolites in the data mining of this metabolomics

495   study. The ANN model exhibited a desirable fitness and predictive ability, and the metabolic signatures

496   discovered using neurofuzzy logic were helpful for understanding the hidden cause-effect relationships

497   between the experimental data. The potential mechanism of the electric stress was elucidated, and

498   excitatory amino acid neurotransmitters related to the typical psychological symptoms of PMS,

499   including anxiety and irritability, were found to be potential biomarkers for the diagnosis and

500   therapeutic evaluation of PMS. This research demonstrates that artificial intelligence technologies are

501   powerful and promising tools for modelling complex metabonomic data and discovering hidden

502   knowledge regarding potential biomarkers associated with the development of diseases, which is also

503   suitable for other complex biological data sets.

504   **Ethical conduct of research**

505   The authors state that they have obtained appropriate institutional review board approval or have

506   followed the principles outlined in the Declaration of Helsinki for all human or animal experimental

507   investigations.

508

509   **Acknowledgments**

513 **Supporting Information Available**

514 **Figure S1** The PCA score plot of the rat urine data on days 0 (prior to the electrically induced stress), 1,

515 2, 3, 4 and 5. The toleration ellipse curve in the PCA score plot was drawn using Hotelling's T2 with a

516 confidence value of 95 %.

517 **Figure S2** PLS-DA scores plots of normal rat urine data on days 0, 1, 2, 3, 4 and 5. Figure S2 The

518 PLS-DA score plots of the rat urine data on days 0 (prior to the electrically induced stress), 1, 2, 3, 4 and

519 5. (A) The dynamic mean-centred PLS-DA score plot of the rat urine data from the model group and the

520 control group $(Q^2Y_{(cum)}=0.912, R^2X_{(cum)}=0.427, R^2Y_{(cum)}=0.952)$. (B) The dynamic mean-centred

521 PLS-DA score plot of the rat urine data from the *Baixiangdan*-dosed group and the control group

522 $(Q^2Y_{(cum)}=0.877, R^2X_{(cum)}=0.423, R^2Y_{(cum)}=0.973)$.

523 **Figure S3** PLS-DA scores plots of normal rat urine data on days 0, 1, 2, 3, 4 and 5.

524 **References**

525 1. E. J. Want, I. D. Wilson, H. Gika, G. Theodoridis, R. S. Plumb, J. Shockcor, E. Holmes, and J. K.

526   Nicholson, *Nat. Protoc.*, 2010, **5**, 1005–1018.

527 2. J. K. Nicholson, *Mol. Syst. Biol.*, 2006, **2**, 52.

528 3. J. C. Lindon, J. K. Nicholson, E. Holmes, H. Antti, M. E. Bollard, H. Keun, O. Beckonert, T. M.

529   Ebbels, M. D. Reily and D. Robertson, *Toxicol. Appl. Pharm.*, 2003, **187**, 137–146.

530 4. X. P. Liang, X. Chen, Q. L. Liang, H. Y. Zhang, P. Hu, Y. M. Wang, and G. A. Luo, *J. Proteome. Res.*,

531   2011, **5**, 790–799.

532 5. P. Wang, H. Sun, H. Lv, W. Sun, Y. Yuan, Y. Han, D. Wang, A. Zhang and X. J. Wang, *J. Pharm.*

533   *Biomed. Anal.*, 2010, **53**, 631–645.

534 6. B. Sun, L. Li, S. M. Wu, Q. Zhang, H. J. Li, H. B. Chen, F. M. Li, F. T. Dong and X. Z. Yan, *Anal.*

535   *Biochem.*, 2009, **395**, 125–133.

536 7. H. M. Jia, Y. F. Feng, Y. T. Liu, X. Chang, L. Chen, H. W. Zhang, G. Ding and Z. M. Zou, *PLoS One*,

537   2013, **8**, e63624.

538 8. L. D. Han, J. F. Xia, Q. L. Liang, Y. Wang, Y. M. Wang, P. Hu, P. Li, and G. A. Luo, (2011), *Anal.*

539   *Chim. Acta*, 2011, **689**, 85–91.

540 9. Z. T. Jiang, J. B. Sun, Q. L. Liang, Y. F. Cai, S. S. Li, Y. Huang, Y. M. Wang and G. A. Luo, *Talanta*,

541   2011, **84**, 298–304.

542 10. X. Li, X. Lu, J. Tian and G. W. Xu, *Anal. Chem.*, 2009, **81**, 4468–4475.

543 11. J. Chen, Y. Zhang, X. Y. Zhang, R. Cao., S. L. Chen, Q. Huang, X. Lu, X. P. Wang, X. H. Wu, C. J.

544   Xu, G. W. Xu and X. H. Lin, *Metabolomics.*, 2011, **7**, 614–622.

545 12. J. S. R. Jang, C. T. Sun and E. Mizutani, Neuro-fuzzy and soft computing: a computational approach

546   to learning and machine intelligence. Prentice-Hall International (UK) Limited, London, 1997.

547   13. Q. Shao, R. C. Rowe and P. York, *Eur. J. Pharm. Sci.*, 2006, 28, 394–404.

548   14. Q. Shao, R. C. Rowe and P. York, *Eur. J. Pharm. Sci.*, 2007, 31, 129–136.

549   15. P. I. Sundstrom, S. Smith and M. Gulinello, *Arch. Womens Ment Health*, 2003, **6**, 23–41.

550   16. M. Steiner, *Lancet.*, 2000, **356**, 1126–1127.

551   17. H. Gao, (2009), Research progress on the correlation between anger pathogenesis and Premenstrual

552   syndrome. *Chin. J. Med. Guide*, 2009, **11**, 283–285.

553   18. Y. Zhao, H. Xu, L. M. Lei and H. L. Zhu, *Int. J. Lab. Med.*, 2013, **34**, 2627–2628.

554   19. S. G. Sun, Y. Lu, A. H. Wang, X. L. Sui, M. Huang, J. Liu, J. Zhu, Z. F. Li, H, Y. Zhang and M. Q.

555   Qiao, *China Pharmacy.* 2011, **22**, 209–211.

556   20. H. Y. Zhang, M. Q. Qiao, W. C. Zhu and J. Wang, *Chin. Trad. Pat. Med.*, 2002, **24**, 118–119.

557   21. L. Li, P. Sun, Q. L. Liang, H. Y. Zhang, Y. Wang, M. Q. Qiao and G. A. Luo, *Chin. Trad. Pat. Med.*,

558   2011, **33**, 762–767.

559   22. Y. M. Li, B. Zhang and X. H. Fan, *Chin. Trad. Pat. Med.*, 2009, **31**, 1690-1694.

560   23. G. A. Luo, Y. M. Wang, Q. L. Liang and Q. F. Liu, *Systems Biology for Traditional Chinese*

561   *Medicine*, Science Press, Beijing, China, 2010.

562   24. X. Zhou and X. Q. Wang, (2003), *Chin. J. Neurosci.*, 2003, **19**, 130–133 .

563   25. L. Sun, Analysis of progesterone and amino acid in serum and different brain regions of rats with

564   premenstrual syndrome liver -qi invasion . Master thesis, Shandong University of Traditional

565   Chinese Medicine, 2008.

566   26. H.Q. Wu, U. Ungerstedt, R. Schwarcz, *Eur J Pharmacol*. 1995, **25**, 55–61

567   27. K. Danhauser , S.W. Sauer, T.B. Haack, T. Wieland, C. Staufner, E. Graf, J. Zschocke, T.M. Strom,

568   T. Traub, J.G. Okun, T. Meitinger, G.F. Hoffmann, H. Prokisch, S. Kölker. *Am J Hum Genet*. 2012,

569   **91**, 1082–1087.

570   28. A. Goldstone , E. Adams, *J Biol Chem.* 1962, **237**, 3476–3485.

571   29. X. Y. Wang, C. Y. Zeng, J. C. Lin, T. L. Chen, T. Zhao, Z. Y. Jia, X. Xie, Y. P. Qiu, M. M. Su, T.

572   Jiang, M. W. Zhou, A. H. Zhao and W. Jia, *J. Proteome. Res.*, 2012, **11**, 6223−6230.

573   30. X. Wang, T. Zhao, Y. P. Qiu, T. Jiang, M. W. Zhou, A. H. Zhao and W. Jia, *J. Proteome. Res.*, 2009, **8**,

574   2511−2518.

575   31. S. Ramcharan, E. J. Love, G. H. Fick and A. Goldfien, *J. Clin. Epidemiol.*, 1992, **45**, 377– 392 .

576   32. J. E. Borenstein, B. B. Dean, J. Endicott, J. Wong, C. Brown, V. Dickerson and K. A. Yonkers, *J.*

577   *Reprod. Med.*, 2003, **48**, 515– 524.

578   33. A. Rapkin, *Psychoneuroendocrinol.*, 2003, **28**, 39 – 53.

579   34. J. Axelrod and H. Weissbach, *Science*, 1960, **131**, 1312 – 1313.

580   35. M. Abe and T. Masanori, *Exp. Eye Res.,* 1999, **68**, 255-262.

581   36. X. H. Liu, H. Y. Zhang and Q. T. Zhao, *J. Trad. Chin. Med.* 2007, **48**, 834–836.

582   37. Y.G. Hu and L.Xue, *Pharmacol. Clin. Chin. Mater. med.* 2007, **23**, 151–153.

583 **Table 1 Identification of the Significantly Different Endogenous Metabolites in the Model Rats' Urine**

No.	t_R (min)	m/z	Elemental composition	identification results	Data mining	Model[b]
1	5.75	162.0759	$C_{16}H_{12}NO_4$	2-aminoadipic acid[c]	PLS-DA, ANN	+(↑)[a]
2	6.42	130.0869	$C_5H_8NO_3$	5-oxoproline	PLS-DA	+(↑)
3	19.19	353.2466	$C_{20}H_{33}O_5$	11-epi-prostaglandin F2α	PLS-DA, ANN	+(↓)
4	10.54	255.0259	$C_7H_{12}O_8P$	shikimate-5-phosphate	PLS-DA, ANN	+(↑)
5	5.30	164.0805	$C_5H_{10}NO_5$	4-hydroxyglutamate[c]	PLS-DA, ANN	+(↑)
6	19.33	373.2744	$C_{16}H_{29}N_4O_4S$	biocytin	PLS-DA	+(↓)
7	12.47	271.0607	$C_{15}H_{11}O_5$	genistein[c]	PLS-DA, ANN	+(↓)
8	8.27	252.1595	$C_{10}H_{14}N_5O_3$	deoxyadenosine	PLS-DA	+(↓)
9	19.19	371.2589	$C_{20}H_{35}O_6$	6-keto-prostaglandin F1α	PLS-DA	+(↓)
10	4.60	180.0619	$C_9H_{10}NO_3$	hippuric acid[c]	PLS-DA, ANN	+(↑)
11	5.78	144.0601	$C_6H_{10}NOS$	5-(2-hydroxyethyl)-4-methyliazole	PLS-DA	+(↑)
12	4.69	105.0683	$C_3H_8N_2O_2$	2,3-diaminopropionic acid	PLS-DA, ANN	+(↓)
13	6.30	233.1219	$C_{13}H_{17}N_2O_2$	melatonin[c]	PLS-DA, ANN	+(↑)
14	4.60	118.0921	$C_5H_{12}NO_2$	5-amino-valerate	PLS-DA	+(↓)
15	16.31	203.1827	$C_{10}H_{27}N_4$	spermine[c]	ANN	+(↓)
16	1.03	166.0808	$C_9H_{12}NO_2$	L-phenylalanine[c]	ANN	+(↓)
17	5.78	116.0722	$C_5H_{10}NO_4$	proline[c]	ANN	+(↑)
18	8.26	206.1565	$C_9H_{20}NO_4$	pantothenol	ANN	+(↓)
19	1.06	150.0911	$C_5H_{12}NO_2S$	L-methionine[c]	ANN	+(↑)
20	11.34	285.0753	$C_{10}H_{12}N_4O_6$	xanthosine	ANN	+(↓)

584 [a] " ↑ " represents a higher level of metabolites, whereas " ↓ " represents a lower level of metabolites. All of the data represent the
585 intensity values of the metabolites on day 5. "+" represents a statistically significant difference ($p<0.05$).[b] Compared to the initial state.
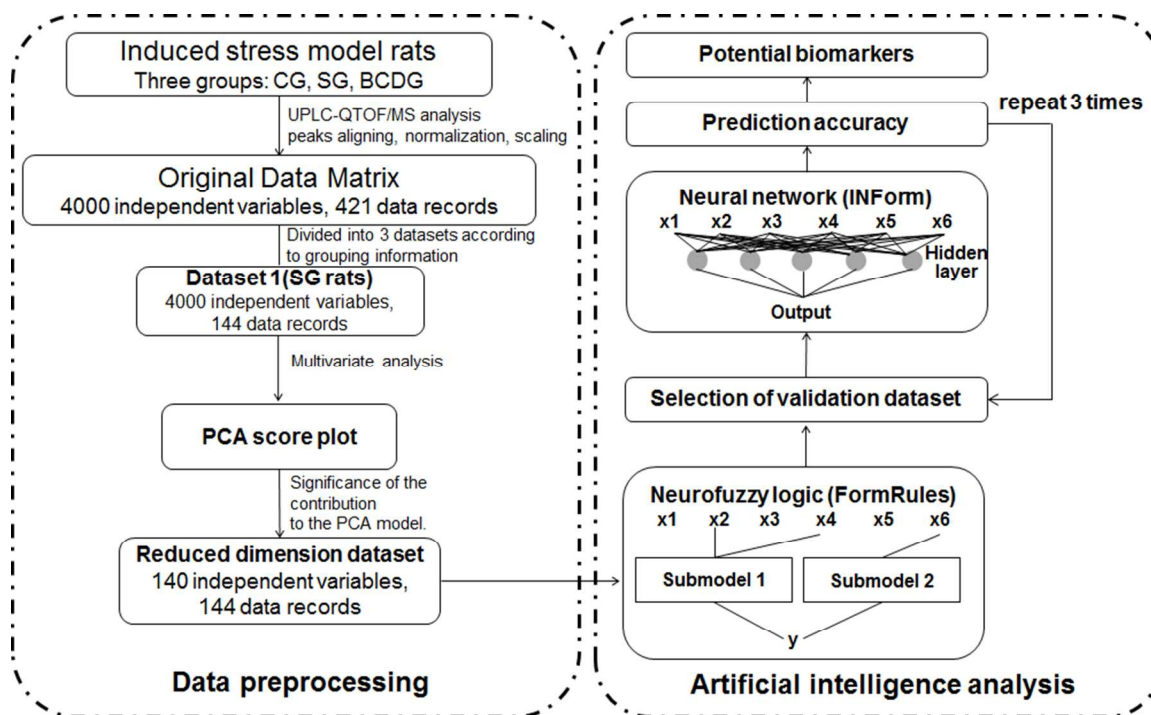586 [c] Confirmed using authentic standards.

Figure legends

Figure 1 The work flow of the data analyses using AI techniques, including reducing the variables using a neurofuzzy logic model and the predictive ability assessment using a neural network model.

Figure 2 The representative base peak intensity chromatograms of the rat urine obtained using the ESI positive mode of the UPLC-QTOF/MS. (A) normal group; (B) model group; (C) *Baixiangdan* capsule-dosed group.

Figure 3 The predictions given by the ANN models generated using a dataset containing various numbers of independent variables. (A) 140 independent variables, (B) 14 independent variables.

Figure 4 A comparison of the major differential metabolites in the urine of the model group (MG), the normal group (NG) and the *Baixiangdan* capsule-dosed group (BADG). (A) 2-aminoadipic acid, (B) 4-hydroxyglutamate, (C) melatonin. $^{\#}p < 0.05$ *vs.* the NG rats, $^{*}p < 0.05$ *vs.* the MG rats (student's t-test).

Figure 1 The work flow of the data analyses using AI techniques, including reducing the variables using

a neurofuzzy logic model and the predictive ability assessment using a neural network model.
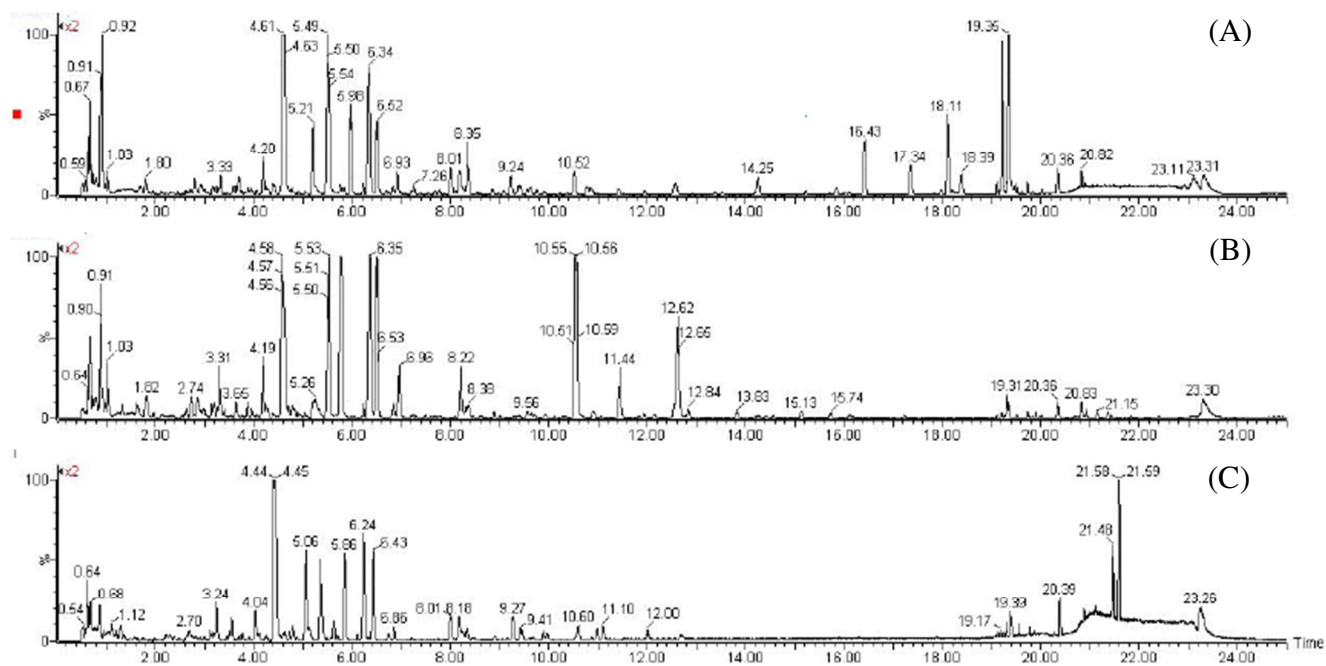
Figure 2. The representative base peak intensity chromatograms of the rat urine obtained using the ESI positive mode of the UPLC-QTOF/MS. (A) control group (CG); (B) stress group; (C) *Baixiangdan* capsule dosed group (BCDG).
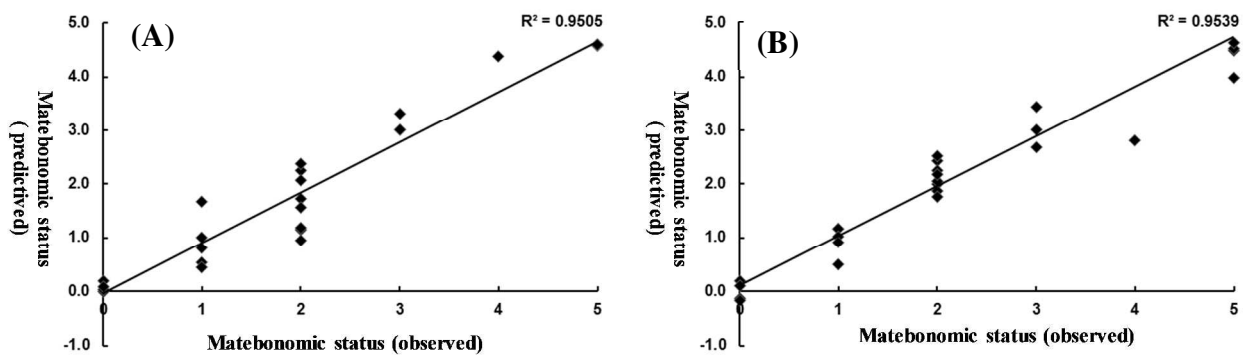
Figure 3 The predictions given by the ANN models generated using a dataset containing various numbers of independent variables. (A) 140 independent variables, (B) 14 independent variables (reduced dimensions).
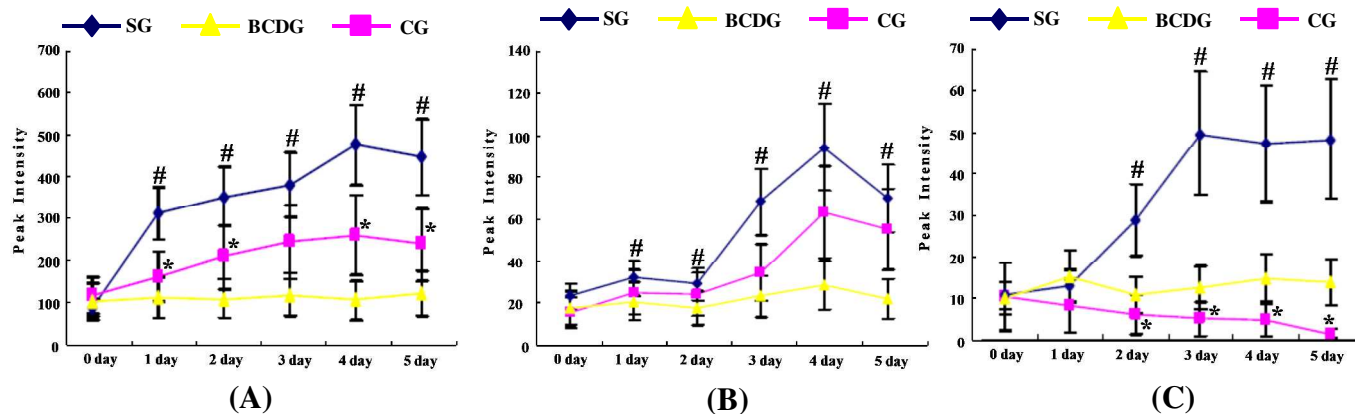
Figure 4 A comparison of the major differential metabolites in the urine of the stress group (SG), the control group (CG) and the *Baixiangdan* capsule-dosed group (BCDG). (A) 2-aminoadipic acid, (B) 4-hydroxyglutamate, (C) melatonin. #p <0.05 *vs.* the CG rats, *p <0.05 *vs.* the SG rats (student's t-test).

Graphical Abstract

An integrating application of multivariable analysis asn artificial intelligence technology (artificial neural networks and neurofuzzy logic) was firstly used to find out potential biomarkers related to the occurrence and development of liver-Qi syndrome PMS induced by electric stimulation in rats.