

Research Article

An herbal drug combination identified by knowledge graph alleviates the clinical symptoms of plasma cell mastitis patients: a nonrandomized controlled trial

Caigang Liu^{1,*}, Hong Yu^{1,#}, Guanglei Chen^{1,#}, Qichao Yang², Zichu Wang², Nan Niu¹, Ling Han³, Dongyu Zhao⁴, Manji Wang^{5,*}, Yuanyuan Liu^{2,6}, Yongliang Yang^{2,*},

¹*Cancer Stem Cell and Translation Medicine Lab, Innovative Cancer Drug Research and Development Engineering Center of Liaoning Province, Department of Oncology, Shengjing Hospital of China Medical University, Shenyang, 110004, China;*

²*School of Bioengineering, Dalian University of Technology, Dalian, 116023, China;*

³*National Engineering Research Center of Pharmaceuticals of Traditional Chinese Medicine, China Resources Sanjiu Medical & Pharmaceutical Co., Ltd, Shenzhen, China;*

⁴*International Cancer Institute, Peking University Health Science Center, Peking University, Beijing, 100191, China;*

⁵*Shanghai BeautMed Corporation, Shanghai, 200030, China;*

⁶*Department of Biology, University of Copenhagen, Denmark.*

Word count of the manuscript text: 3,645

[#]These authors contributed equally to the work.

^{*}To whom correspondence may be addressed:

Caigang Liu, E-mail: angel-s205@163.com; **Address:** Department of Oncology, Shengjing Hospital of China Medical University, Shenyang, 110004, China (**lead contact**)

Manji Wang, E-mail: 147304368@qq.com

Yongliang Yang, E-mail: everbright99@foxmail.com

32 **Abstract**

33 ***Background***

34 Plasma cell mastitis (PCM) is a nonbacterial breast inflammation with severe and
35 intense clinical manifestation yet treatment methods for PCM are still rather limited.
36 Although the mechanism of PCM remains unclear, mounting evidences suggest that
37 the dysregulation of immune system is closely associated with the pathogenesis of
38 PCM. Drug combinations or combination therapy could exert improved efficacy and
39 reduced toxicity through hitting multiple discrete cellular targets.

40 ***Methods***

41 We have developed a knowledge graph architecture towards immunotherapy and
42 systematic immunity that consists of herbal drug-target interactions with a novel
43 scoring system to select drug combinations based on target-hitting rates and
44 phenotype relativeness. To this end, we employed this knowledge graph to identify an
45 herbal drug combination for PCM and we subsequently evaluated the efficacy of the
46 herbal drug combination in clinical trial.

47 ***Results***

48 Our clinical data suggests that the herbal drug combination could significantly reduce
49 the serum level of various inflammatory cytokines, downregulate serum IgA and IgG
50 level, reduce the recurrence rate and reverse the clinical symptoms of PCM patients
51 with improvements of general health status.

52 ***Conclusions***

53 In summary, we reported that an herbal drug combination identified by knowledge
54 graph can alleviate the clinical symptoms of plasma cell mastitis patients. We
55 demonstrated that the herbal drug combination holds great promise as an effective
56 remedy for PCM, acting through the regulation of immunoinflammatory pathways
57 and improvement of systematic immune level. In particular, the herbal drug
58 combination could significantly reduce the recurrence rate of PCM, a major obstacle
59 for PCM treatment. Our data suggests that the herbal drug combination is expected to
60 feature prominently in future PCM treatment.

61 ***Funding***

62 C. Liu's lab was supported by grants from the Public Health Science and Technology
63 Project of Shenyang (Grant: 22-321-32-18); Y. Yang's laboratory was supported by
64 the National Natural Science Foundation of China (Grant: 81874301), the
65 Fundamental Research Funds for Central University (Grant: DUT22YG122) and the
66 Key Research project of 'be Recruited and be in Command' in Liaoning Province
67 (2021JH1/10400050).

68 ***Clinical trial number***

69 ClinicalTrials.gov: NCT05530226

70 **Key Words:** Plasma cell mastitis; knowledge graph; traditional Chinese medicine;
71 drug combination; clinical trial

72 **Introduction**

73 Plasma cell mastitis (PCM) represents as a serious inflammatory condition of
74 breast that occurs in young and middle-aged females at non-pregnant and non-
75 lactating period¹. The main histopathological characteristics of PCM is the infiltration
76 of plasma cells and lymphocytes in breast tissue². Interestingly, PCM shares similarity
77 with breast cancer in the perspective of macroscopical or microscopical
78 characteristics³. On the other hand, mounting evidences suggest that the dysregulation
79 of immune system is closely associated with the pathogenesis of PCM. Recently, the
80 incidence rate of PCM is quickly rising yet the treatment methods for PCM are still
81 rather limited. In clinical practice, surgical resection and hormone therapy remains as
82 two major treatments for PCM. Unfortunately, neither surgical resection nor hormone
83 therapy could prevent recurrence of PCM. Moreover, the serious side effects for
84 hormone therapy are still problematic⁴. Therefore, the discovery of effective PCM
85 treatment or therapeutics with minimal side effects is clearly warranted.

86 Traditional Chinese Medicine (TCM) has a rather long history for the prevention
87 and treatment of complex diseases in eastern Asia⁵⁻⁷. Moreover, for decades, TCM
88 has been often used as alternative or complementary medicine in the west. Indeed,
89 Chinese herbal compounds has been successfully applied in the treatment of plasma
90 cell mastitis (PCM) in conjunction with western medicine⁸. In clinical practice, TCM
91 refers to herbal entity prescription or formulae (also called as ‘Fangji’) which may
92 exhibit coordinating or synergistic effects through the combination of multiple herb
93 drugs⁹. However, the design of formulae in TCM is solely based on the principle of

94 'syndrome differentiation' according to the medicinal properties of herbal entities.
95 Moreover, the molecular mechanisms for the 'formulae' in TCM remains rather
96 elusive.

97 Knowledge graph has emerged as an advanced technology in the field of artificial
98 intelligence which is able to connect entities in a graph based on their existing
99 intricate relationships¹⁰. In particular, knowledge graph can enable the rational design
100 and identification of combination therapies for a specific disease or phenotypes¹¹.
101 Recently, we developed and constructed a knowledge graph for the discovery of
102 herbal drug combination towards immunotherapy and systematic immunity.
103 Subsequently, we identified a synergistic combination of herbal drugs for PCM via a
104 scoring system based on target hitting rates and phenotype relativeness. To verify our
105 concept of design, we conducted a clinical trial experiment for the drug combination
106 of herbal compounds mentioned above (ClinicalTrials.gov Registration:
107 NCT05530226). Strikingly, our clinical results demonstrated that the herbal drug
108 combination identified by knowledge graph can markedly suppress various
109 inflammatory cytokines in serum, restore clinical symptoms and reduce the recurrence
110 rates of PCM patients with improved global health status.

111 **Materials and Methods**

112 *Construction of knowledge graph towards immunotherapy*

113 We employed data mining techniques to collect and compile 240 targets of
114 immunotherapy and systematic immunity from PubMed database. Next, we collected

115 and compiled 345 herbal drug entities officially released by the National Health
 116 Commission of China and National Administration of Traditional Chinese Medicine.
 117 The intricate relations between the herbal drug entities and the immunotherapy targets
 118 were extracted from the PubMed database. These intricate relations were subjected to
 119 further manual curation. We used thirteen ontology terms to describe the intricate
 120 relations (edges) in the knowledge graph. Moreover, 64 attributes of the medicinal
 121 properties for the herbal drug entities were collected and compiled from
 122 Pharmacopoeia of China. Finally, we built the knowledge graph via Neo4j and
 123 Py2Neo tools which consists of 895 nodes and 2197 edges.

124 *Scoring system of the knowledge graph*

125 To this end, we developed a scoring system to asses and predict synergistic drug
 126 combination of herbal drug entities (number of drugs, n) as below,

$$127 \quad Score = f(x) * g(x) * \sum_{i=1}^x \left\| \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} \right\| \quad [1]$$

128 Herein, $f(x)$ represents as the penalty function. $f(x)$ value will be set to 0 if the
 129 medicinal properties of the drug combination fall into the contraindication rules in
 130 Pharmacopoeia of China. Otherwise, $f(x)$ value will be set to 1. Herein, the penalty
 131 function is used to ensure that the herbal drug entities in the combination didn't
 132 violate the contraindication rules in Pharmacopoeia of China according to the
 133 medicinal attributes of herbal drugs.

$$134 \quad g(x) = 1 - p_i \quad [2]$$

135 $g(x)$ is the target diversity function as above and p_i is calculated as $p_i = w/t$; t
136 refers to the total number of targets within the knowledge graph, w refers to the total
137 number of overlapping targets that the drug combination may hit. Hence, the target
138 diversity function can be used as a measure to assess the diversity of the targets that
139 the drug combination may hit. In another word, if each drug entity in the combination
140 hits distinct targets, the $g(x)$ value will be set to 1. The last term of the scoring
141 system is used as a measure to assess the relativeness of each drug entity in the
142 combination and calculated as follows,

$$143 \quad \sum_{i=1}^x \left\| \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix} \right\| \quad [3]$$

144 In brief, h_1 represents the target hitting rates of each drug entity in the combination
145 and was calculated as follows, $h_1 = n_i/t$; n_i is the number of hitting targets for each
146 drug entity in the combination; Again, t refers to the total number of targets
147 constituting the knowledge graph for the disease; Noteworthy, the concept of hitting
148 rates towards discrete targets has been used in the scoring function for the selection of
149 synergistic drug combinations¹². h_2 represents the phenotype relativeness of each
150 drug entity in the combination and $h_2 = c_2 * 1/x$, where x is the number of drug
151 entities in the combination and c_2 is the parameter; Namely, if the drug entity is
152 related to the phenotype of the disease (co-occurrence with the disease phenotype in
153 the literature), then c_2 value is set to 1 otherwise c_2 value is set to 0; h_3 represents the

154 literature relativeness or confidence of each drug entity in the combination and
155 calculated as follows,

$$156 \quad h_3 = c_3 * \log \sqrt{\sum_{i=1}^x l \times (j + k)} \quad [4]$$

157 in which l is the number of studies/publications that validated the association of drug
158 entity with the specific disease (herein in the knowledge graph refers to cancer
159 immunotherapy), j and k refer to if the relations of the drug entity with the disease
160 have been validated in cell lines or patient (or animal) tissues, respectively. Namely, if
161 the drug entity was validated in cancer cell lines or patient tissues, the j or k value will
162 be set to 1, respectively. Otherwise, the j or k value will be set to 0; c_3 is the parameter
163 and set to 1 here. Therefore, herein, a high score of h_3 implicates that the drug
164 combination is more relevant to cancer immunotherapy with high confidence of
165 literature relativeness. Collectively, our scoring system can be used to select those
166 drug combinations that are most relevant with disease phenotypes and those drug
167 combinations that are able to hit most discrete targets related to immunotherapy.

168 ***Design of the clinical trial***

169 In brief, 160 female patients diagnosed as plasma cell mastitis (PCM) in Shengjing
170 Hospital Affiliated to China Medical University were recruited in the clinical trial
171 between January 2021 to February 2022. Patients were randomly 1:1 divided into
172 experimental group (EG) and control group (CG). Noteworthy, in order to
173 demonstrate the therapeutic effect of TCM drug combination, we selected patients

174 who were treated with western medicine in the real world during the same period.
175 Therefore, the two groups of patients were divided into TCM treatment group
176 (experimental group) and western medicine treatment group (control group). There
177 was no significant difference in baseline data such as age, body mass index, clinical
178 classification, marriage and child-bearing history between the two groups
179 (**Supplemental file 1A**). Patients in the CG group were orally treated with
180 methylprednisolone tablets, 20mg/ day once a day. The patients in the EG group were
181 orally treated with 20g/bag of herbal drug combination twice a day, once in the
182 morning and once in the evening for 2 months. The herbal drug combination was
183 prepared as granules in the following formulae: Taraxacum 15g, Fructus forsythiae
184 15g, Honeysuckle 10g, Uniflower swisscentaury root 8g, Herba violae 20g, Danshen
185 10g, Astragalus 20g, Liquorice 8g. The formula was determined by TCM experts on
186 the basis of ‘syndrome differentiation’ as described in Pharmacopoeia of China.
187 Furthermore, the herbal drug combination in the form of granules was provided and
188 prepared by Shengjing Hospital Affiliated to China Medical University according to
189 the standard requirement of clinical study by National Medical Products
190 Administration (NMPA).

191 ***Clinical trial protocol***

192 The clinical trial for the herbal drug combination was registered at ClinicalTrials.gov
193 and entitled as “A Single Arm Study of Traditional Chinese Medicine for Plasma Cell
194 Mastitis” with registration code of NCT05530226 (see

195 CONSORT_flowchart_diagram, **Figure 1 - Figure supplement 2**). The detailed
196 clinical trial protocol has been provided a separate document in the Supplemental
197 Files named as '**Supplemental file 2**'.

198 *Measurement of serum inflammatory cytokines by ELSIA assay*

199 Venous blood of the CG and EG groups were collected in sterile non-anticoagulant
200 test tube before and after treatment. The immune transmission turbidimetry was used
201 according to the procedure of CRP kit and automatic biochemical analyzer was used
202 to detect the level of CRP. The levels of serum cytokines were measured by ELISA
203 (Elabscience) according to the manufacturer's instructions.

204 *Measurement of serum immunoglobulin level*

205 The venous blood of PCM patients in the two groups were collected in sterile non-
206 anticoagulant tube before and after treatment. The serum IgG and IgA were measured
207 by rate scattering turbidimetry using Array 360 System automatic specific protein
208 analyzer (Beckman Company, USA).

209 *Assessment of clinical symptoms of PCM patients*

210 The clinical symptoms were evaluated by attending physician with board certification
211 in pathology. The patients were scored before and after treatment according to the
212 standard rating scale for PCM (**Supplemental file 1B**).

213 *Statistics*

214 All data were evaluated as mean \pm SEM. Statistical analysis of the quantitative
215 multiple group comparisons was performed using the one-way analysis of variance
216 (ANOVA) followed by Tukey's test; whereas pairwise comparisons were performed
217 using the t test by GraphPad Prism 8 (Graph Pad Software, La Jolla, CA, USA).
218 Results were considered to be statistically significant with $p < 0.05$.

219 **Results**

220 In previous study, we collected and compiled 240 targets for immunotherapy and
221 systematic immunity from literature data¹³. Recently, we collected 345 entities of
222 herbal drugs documented in TCM books and herbal drugs announced by National
223 Administration of Traditional Chinese Medicine through advanced text-mining
224 techniques. The existing intricate relationships between the herbal drugs and
225 immunotherapy targets were also extracted and compiled via advanced text-mining
226 techniques and manual curation for the construction of knowledge graph. We defined
227 an ontology list consisting of 13 ontology terms describing the relations (edges)
228 between herbal drug entities and the immunotherapy targets based on manual curation
229 of literature data. Moreover, we collected the attributes of the medicinal properties for
230 each herbal compound from Pharmacopoeia of China¹⁴. Totally, we compiled and
231 integrated 64 attributes of the medicinal properties for herbal drug entities into the
232 knowledge graph. These medicinal properties are useful throughout the design of
233 herbal drug combination. Finally, we built the knowledge graph via Neo4j and
234 Py2Neo tools which consists of 895 nodes and 2197 edges (**Figure 1** and **Figure**

235 **supplement 1**), which can be visited online (<http://www.ikgg.org/>). Subsequently, we
236 employed a scoring system (or so-called recommendation system) to assess and predict
237 synergistic herbal drug combination from the knowledge graph. Of note, the scoring
238 function is able to identify those herbal drug combinations that are most related with
239 specific phenotypes as well as herbal drug combinations that are able to hit most
240 discrete cellular targets, yet still following the principle of ‘syndrome differentiation’
241 as described in Pharmacopoeia of China (**Materials and Methods**). Here, syndrome
242 differentiation refers to the very basic principle of identifying and treating disease in
243 TCM. Syndrome (Zheng) is the presentation of the pathological changes during a
244 specific disease course including the location, cause and nature of a disease. Moreover,
245 ‘Jun-Chen-Zuo-Shi’ refers to the rules guided by syndrome differentiation to select
246 multiple herbal drug entities to treat a specific disease in TCM. To this end, we used
247 this scoring function to select herbal drug combinations consisting of eight herbal
248 entities. We chose to identify drug combinations with eight entities because ‘formulae’
249 consisting of eight drugs are regarded as ‘essence combination’ in TCM community.
250 In short, we employed a combination generator that is able to randomly generate drug
251 combinations with eight herbal drug entities for ten rounds, each of which consists of
252 1,000 random drug combinations. All the generated drug combinations from the ten
253 rounds were further ranked and evaluated. Noteworthy, the scoring results of the ten
254 rounds presented as normal distributions (**Figure 2 – Figure Supplement 1**). The top
255 twenty combinations from each round ranked by the scoring function were further

256 curated and inspected by experts in TCM. Remarkably, we identified a specific drug
257 combination which was ranked among top twenty choices in all ten rounds of
258 calculation. The drug combination was chosen for the further clinical study for two
259 reasons. First, this drug combination was among top twenty combinations in each
260 round of our calculations. Second, we asked experts in TCM to inspect the top twenty
261 combinations on the basis of ‘syndrome differentiation’ as described in
262 Pharmacopoeia of China and finally the combination consisting of eight herbal drug
263 entities including ‘Fructus forsythiae’, ‘Herba violae’, ‘Uniflower swisscentaury root’,
264 ‘Danshen’, ‘Astragalus’, ‘Taraxacum’, ‘Liquorice’ and ‘Honeysuckle’ was selected
265 for further clinical study.

266 Next, we extracted the subgraph for the herbal drug combination mentioned
267 above and created a network diagram for the drug combination via Cytoscape tools¹⁵
268 (**Figure 2**). In total, the eight herbal drug entities in the combination regulate 46
269 cellular targets related to immunotherapy and systematic immunity such as HIF-1¹⁶,
270 iNOS, IL-17, IL-6, IL-1 β , mTOR, NLPR3, PD-L1, STAT3¹⁷, TGF- β , TLR2 and
271 TLR4 etc. (**Figure 2**). Noteworthy, the medicinal properties of the eight drug entities
272 could be classified into three major categories of ‘Heat-clearing and detoxicating’,
273 ‘Qi-tonifying’ and ‘Blood-activating menstruation regulating’. Moreover, we
274 conducted pathway analysis for the herbal drug combination for plasma cell mastitis.
275 Interestingly, we revealed that the herbal drug combination may modulate a few
276 pathways related to systematic immunity including ‘Toll-like receptors cascades’,

277 'MAP kinase activation', 'Adaptive immune system', 'Growth hormone receptor
278 signaling', 'Cytokine signaling in immune system' and 'Innate immune system' via
279 reactome knowledgebase¹⁸ (**Figure 3**). We believe all these may account for the
280 therapeutic profiles of the herbal drug combination towards PCM. For instance, 'Toll-
281 like receptors cascades', 'Adaptive immune system', 'Cytokine signaling in immune
282 system' and 'Innate immune system' are critical cellular pathways for systematic
283 immunity which are directly associated with the pathogenesis of PCM. Moreover, the
284 'MAP kinase activation' pathway is associated with cellular defense and innate
285 immunity which are also crucial for the development and inflammatory conditions of
286 PCM.

287 Subsequently, we want to evaluate the efficacy of herbal drug combination in
288 clinical trial for PCM patients (Clinicaltrials.gov number: NCT05530226). The 'Jun-
289 Chen-Zuo-Shi' principle was examined for the herbal drug combination ('Formulae')
290 by TCM experts and the dosage for each drug entity from the drug combination was
291 adjusted by TCM experts. The clinical trial is a unrandomized, open label single arm
292 study investigating the efficacy and safety of the herbal drug combination. To reveal
293 the therapeutic effects of TCM drug combination, we selected patients who were
294 treated with western medicine in the real world as comparison. Therefore, the two
295 groups of patients were divided into TCM treatment group (experimental group) and
296 western medicine treatment group (control group). All patients were diagnosed with
297 PCM by biopsy of breast tissue before recruited into the clinical trial. Patients in the

298 CG group (the control group) were orally treated with methylprednisolone whereas
299 the patients in the EG group (the experimental or treatment group) were orally treated
300 with herbal drug combination (**Materials and Methods**). Of note,
301 methylprednisolone is a standard corticosteroid for the treatment of inflammatory
302 conditions in clinical practice¹⁹ and therefore methylprednisolone was used in the CG
303 group.

304 Efficacy was assessed every 2 cycles and the results were summarized after six
305 months of treatment. The baseline characteristics are shown in **Supplemental file 1A**.
306 Strikingly, our results demonstrated that a few inflammatory cytokines in the serum
307 including IL-2, IL-4, IL-6, IFN- γ , IL-1 β and TNF- α were significantly downregulated
308 in PCM patients after treatment of herbal drug combination as compared to the CG
309 group treated with methylprednisolone (**Figure 4**). We chose to measure these
310 cytokines in the experiments because they are often regarded as serum cytokine
311 markers during the pathogenic development of PCM²⁰. In addition, we found that
312 serum level of IgA and IgG level were markedly suppressed in the treatment group of
313 herbal drug combination as compared to the control group (**Figure 5**). Of note, both
314 IgA and IgG have been found to be crucial diagnostic serum markers for PCM
315 patients²¹. Moreover, IgA is regarded as a major component of mucosal immunity
316 which is closely related to the pathogenesis of PCM^{22,23}. Therefore, our data suggests
317 that the herbal drug combination may enable the regulation of mucosal immunity and
318 consequently downregulate IgA and IgG serum level. Furthermore, we conducted the

319 standard Quality of Life questionnaire studies for PCM patients in the clinical
320 experiment. Our results implicated that symptom score, pain score and global health
321 status of PCM patients are significantly improved after treatment of the herbal drug
322 combination as compared to the control group (**Figure 6**). Noteworthy, our results
323 demonstrated that the recurrence rate of PCM patients in the treatment group were
324 reduced to 3.75% as compared to the recurrence rate of 12.5% in the control group
325 (**Table 1**). Moreover, the incidence rate of adverse events of PCM patients in the
326 treatment group were reduced to 6.25% as compared to the recurrence rate of 11.25%
327 in the control group (**Table 1**). In addition, we observed that the clinical symptoms of
328 PCM patients in the EG group such as swelling, abscess and fistula were reversed
329 (**Figure 7, Supplemental file 3**) after treatment of herbal drug combinations. The
330 clinical symptom score in the EG group is ~4.68 as compared to the clinical symptom
331 score of ~5.98 in the CG group (**Table 2**). These results suggest that the herbal drug
332 combination may achieve better efficacy for the treatment of PCM as compared to
333 methylprednisolone.

334 **Discussion**

335 With the increasing amount of biomedical data, the traditional drug discovery
336 campaign has been revolutionized with the aid of artificial intelligence techniques to
337 accelerate the process and reduce the cost²⁴. In recent years, knowledge graph, a
338 technique that can provide structured relations among entities and unstructured
339 semantic relations associated with entities, has been introduced into the domain of

340 drug discovery¹¹. The advantage of employing knowledge graph for drug discovery
341 lies in the capabilities of revealing structured associations between drug entities,
342 cellular targets, biological pathways and phenotypes for human disorders. This is
343 useful for scientists to identify new indications or phenotypes for existing drugs, or
344 so-called drug repurposing. With the aid of scoring function or recommendation
345 system, knowledge graph can also be used to design and identify drug combinations
346 for a specific disease. Herein, for the first time, we introduced and employed the
347 concept of knowledge graph to identify herbal drug combinations for the severe
348 Plasma cell mastitis (PCM) with unmet medical needs.

349 Although the pathogenesis of PCM remains largely unclear, there have been
350 numerous reports implicating that the overactivation of immunoinflammatory
351 pathways play an important role in the development of PCM²⁵. The major advantage
352 of using Traditional Chinese Medicine is that herbal drug combination can hit
353 multiple discrete targets related to immunoinflammatory pathways with improved
354 efficacy and reduced toxicity. Herein, for the first time, we showcase an example that
355 identifies an herbal drug combination via knowledge graph towards PCM. In contrast
356 to using the conventional principle of ‘syndrome differentiation’, our knowledge
357 graph consisting of intricate relations between herbal drug entities and
358 immunotherapy targets coupled with scoring functions are able to automatically
359 identify novel herbal drug combinations which can hit most discrete targets, making
360 this strategy unique in the TCM community. This is because PCM is a rather complex

disease which pathogenesis may involve multiple targets and immunoinflammatory pathways. Hence, we made the hypothesis that drug combinations that can act on most discrete targets or pathways related to PCM might be more effective to alleviate the symptoms. Although we acknowledge that the inclusion of chemical ingredients from the herbal drugs may impact the outcome of our analysis and design, unfortunately, the inclusion of chemical ingredients in the knowledge graph is rather technically difficult due to the limited and incomplete datasets for the herbal drugs in the field of TCM. Nevertheless, our strategy captures the prominent feature of design for drug combinations towards a complex disease such as PCM. In the future, we plan to include multiple types of omics data such as genomic, transcriptomic, proteomic, metagenomic or metabolomics data into the knowledge graph to reveal novel targets and enable novel drug discovery.

We want to remind the reader that a third arm (a placebo group) added in the clinical study might be useful to fully reveal the therapeutic effects of the herbal drug combination. Unfortunately, we were unable to add a third arm due to some ethical concerns. This is because plasma cell mastitis (PCM) is a rather acute, severe and intense clinical manifestation of breast. Without any treatment, the serious inflammatory condition of PCM may quickly advance into breast cancer. Therefore, this is a limitation of our clinical study and we hope to design more delicate clinical trial in the future to demonstrate the therapeutic effects of herbal drug combination. Nevertheless, the detailed analysis of three indicators including symptom score, pain

382 score and score of life quality between the experiment group treated with herbal drug
383 combination and the control group treated with methylprednisolone in the clinical
384 study suggest that the herbal drug combination might be more effective to reverse the
385 clinical conditions of PCM as compared to the methylprednisolone treatment (see
386 Figure 6 - Figure supplement 1-2-3).

387 In the present study, our results revealed that the herbal drug combination
388 identified by knowledge graph could suppress a few key immunoinflammatory
389 cytokines, enhance the systematic immune levels and significantly reduce the
390 recurrence rates of PCM patients. Of note, recurrence has become one major obstacle
391 after surgical resection for PCM treatment in clinical practice. On the other hand,
392 hormone therapy may increase the risk of side effects for PCM patients. Therefore,
393 our approach of herbal drug combination may provide a new avenue for PCM
394 treatment with less recurrence rate and reduced incidence rate of adverse events.

395 **Conclusion**

396 In summary, we report the identification and clinical assessment of an herbal drug
397 combination towards Plasma cell mastitis (PCM). We demonstrated that the herbal
398 drug combination holds great promise as an effective remedy for PCM, acting through
399 the regulation of multiple cellular targets and immunoinflammatory pathways which
400 leads to the improvement of systematic immune level. In particular, the herbal drug
401 combination could significantly reduce the recurrence rate of PCM, a major obstacle
402 for PCM treatment. Our data suggests that the herbal drug combination is expected to

403 feature prominently in the future PCM treatment. Moreover, these promising results
404 underscore the potential of knowledge graph to identify drug combinations or other
405 novel therapeutics across various types of human disorders.

406 **Acknowledgments**

407 Y. Yang's laboratory was supported by the National Natural Science Foundation of
408 China (Grant: 81874301), the Fundamental Research Funds for Central University
409 (Grant: DUT22YG122) and the Key Research project of 'be Recruited and be in
410 Command' in Liaoning Province (2021JH1/10400050); C. Liu's lab was supported by
411 grants from the National Natural Science Foundation of China (No. 81572609), China
412 Medical University Major Construction Project (No. 2017ZDZX05) and Liaoning
413 Colleges Innovative Talent Support Program (Cancer Stem Cell Origin and Biological
414 Behavior).

415 **Conflict of interest statement**

416 Ling Han is an employee of China Resources Sanjiu Medical & Pharmaceutical;
417 Manji Wang is an employee of Shanghai BeautMed Corporation.

418 **Declaration of Ethics**

419 The protocol was approved by the Institutional Review Board (IRB) of the China
420 Medical University (approval number: 2021PS024T). This study was registered with
421 ClinicalTrials.gov: NCT05530226. All patients provided written informed consent.

422

423 **Legends for Supplementary files and Source code**

424 **Supplementary file 1A.** Baseline characteristics of patients in the clinical trial.

425 **Supplementary file 1B.** Clinical symptom rating scale for plasma cell mastitis
426 (PCM).

427 **Supplementary file 2.** Clinical protocol for the clinical trial study.

428 **Supplementary file 3.** Some detailed information for the six patients displayed in
429 Figure 7 of the main text.

430 **Source code 1.** Python source code for the scoring system of knowledge graph to
431 assess and select appropriate drug combinations.

432

433

434

435

436

437

438

439

440

441

442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459

Figures

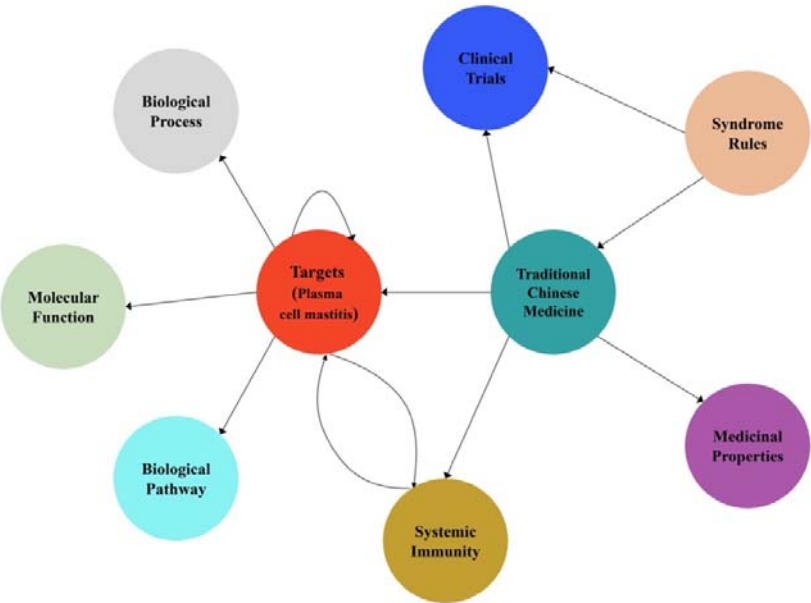


Figure 1. Schematic diagram of the knowledge graph for the drug discovery of plasma cell mastitis in the present work.

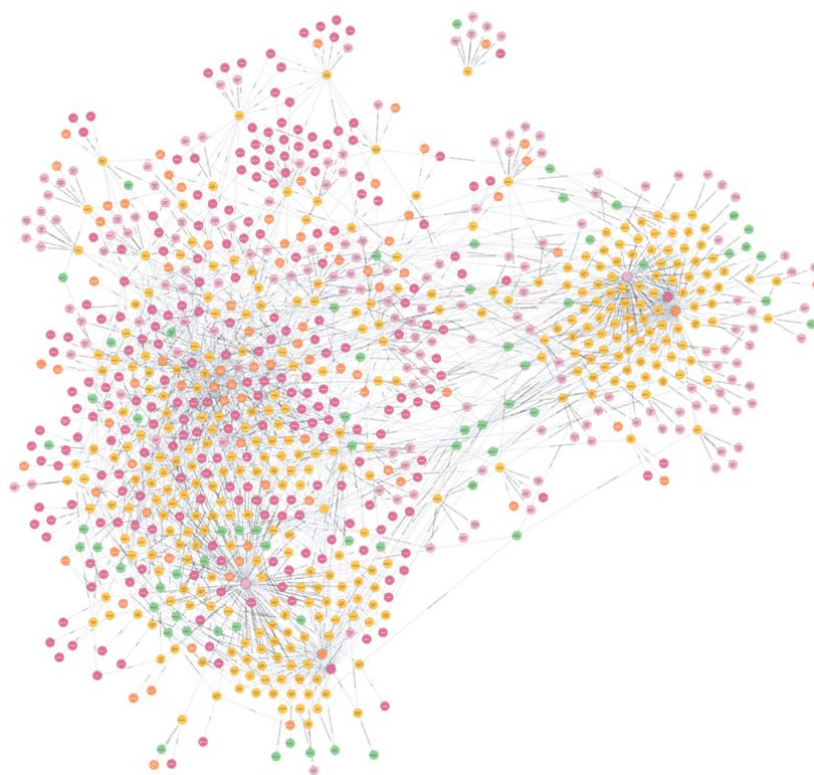
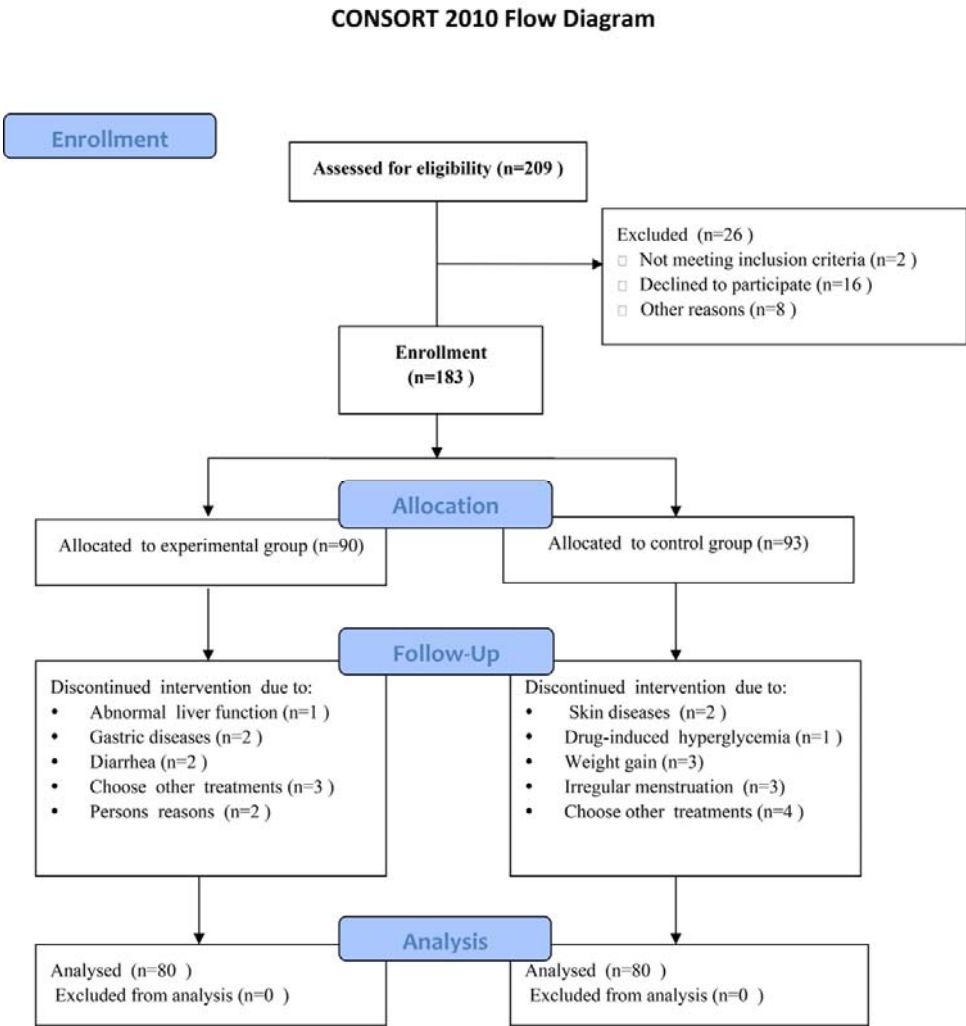


Figure 1 - Figure supplement 1. Snapshot of the medical knowledge graph towards immunotherapy. The knowledge graph was constructed by Neo4j and Python tools (<http://www.ikgg.org/kghti>). The yellow color represents the herbal drug entity, the pink color represents the cellular targets and the green color represents the attributes of herb drugs.



484

485 **Figure 1 - Figure supplement 2.** Consort flowchart diagram of the clinical trial for
486 the herbal drug combination.

487

488

489

490

491

492

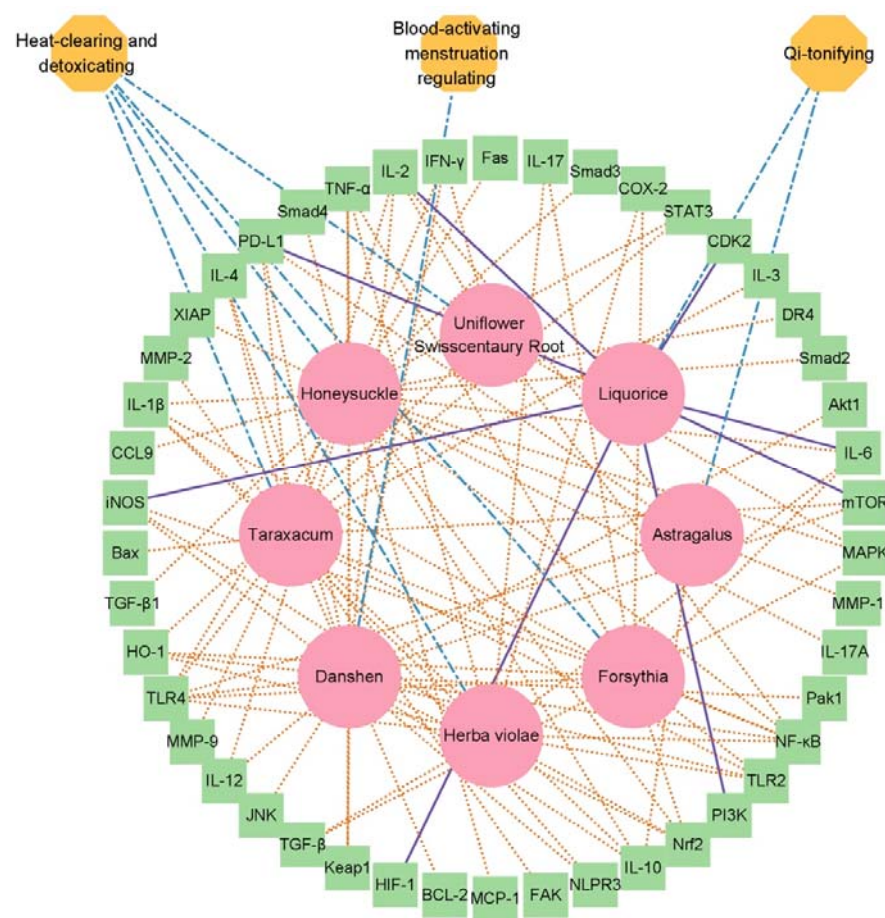


Figure 2. Network diagram of the herbal drug combination consisting of eight entities including ‘Honeysuckle’, ‘Taraxacum’, ‘Astragalus’, ‘Danshen’, ‘Forsythia’, ‘Herba violae’, ‘Liquorice’ and ‘Uniflower swisscentaury root’ displayed in red circles. In total, 46 cellular targets related to systemic immunity were hit by the drug combination such as NLPR3, IL-17, TLR4, STAT3, IL-6, iNOS and TLR2 et al which were displayed in green box model. Three major medicinal attributes (properties) were identified for these herbal drug entities including ‘Heat-clearing and detoxicating’, ‘Qi-tonifying’ and ‘Blood-activating menstruation regulation’.

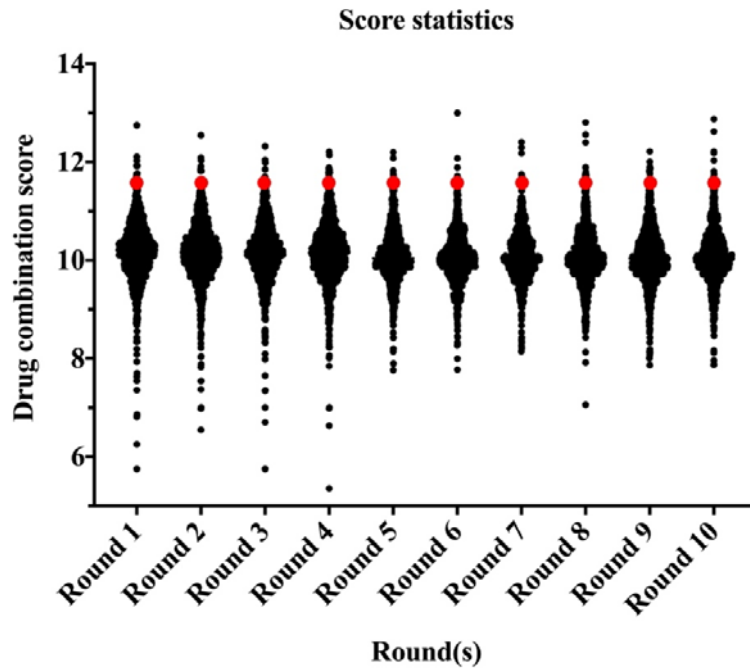


Figure 2 - Figure supplement 1. The score statistics of ten rounds of random herbal drug combinations (1,000 random combination from each round) for the treatment of plasma cell mastitis. Noteworthy, the scoring results of the ten rounds presented as normal distributions. The herbal drug combination identified for the clinical trial was among the top twenty choices in all the ten rounds and marked in red dots.

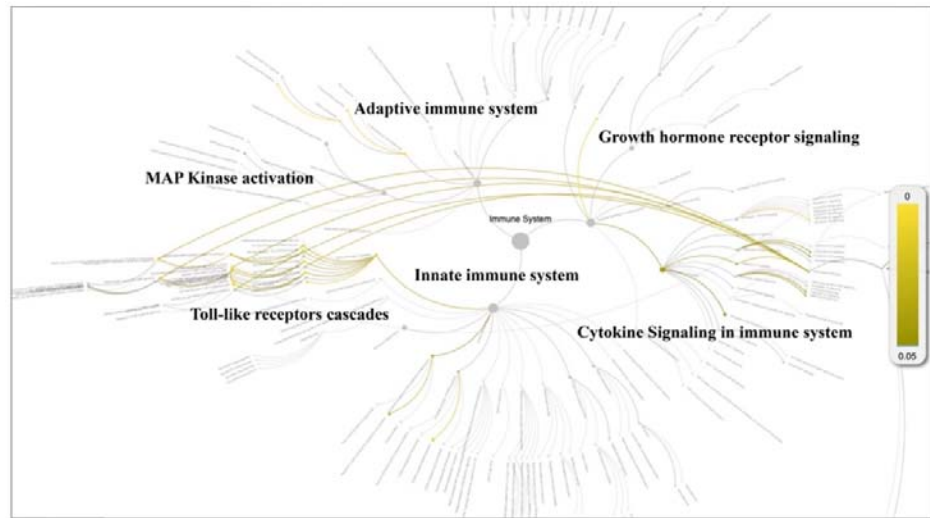


Figure 3. Pathway analysis for the potential cellular targets of the herbal drug combination via reactome knowledgebase. The statistically significant pathways were highlighted and displayed in yellow color ($P < 0.05$).

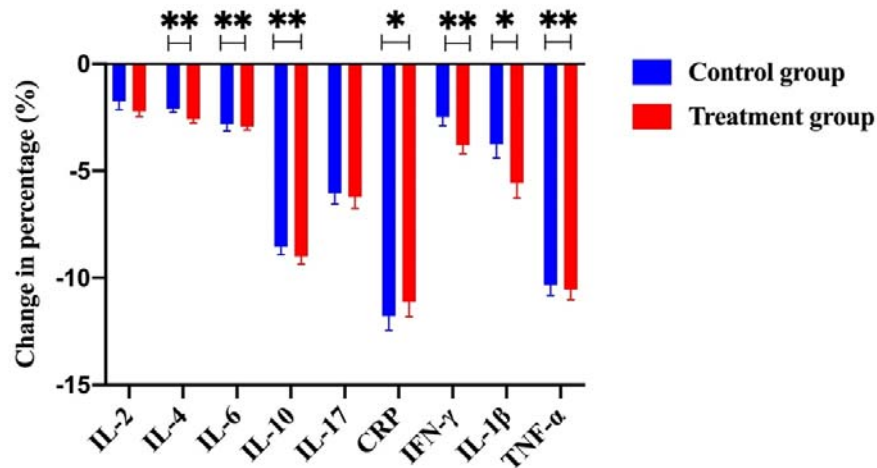


Figure 4. The change of percentage for numerous immunological cytokines including IL-2, IL-4, IL-6, IL-10, IL-17, CRP, IFN- γ , IL-1 β and TNF- α from the control group and the treatment group (with TCM treatment). Notably, a few key cytokines such as IL-2, IL-4, IFN- γ , IL-1 β and TNF- α were significantly downregulated in the treatment group as compared to the control group.

Figure 4 – source data. Patients' raw data of serum cytokines from the control group (without TCM treatment) and the treatment group (with TCM treatment) in the clinical trial (ClinicalTrials.gov: NCT05530226).

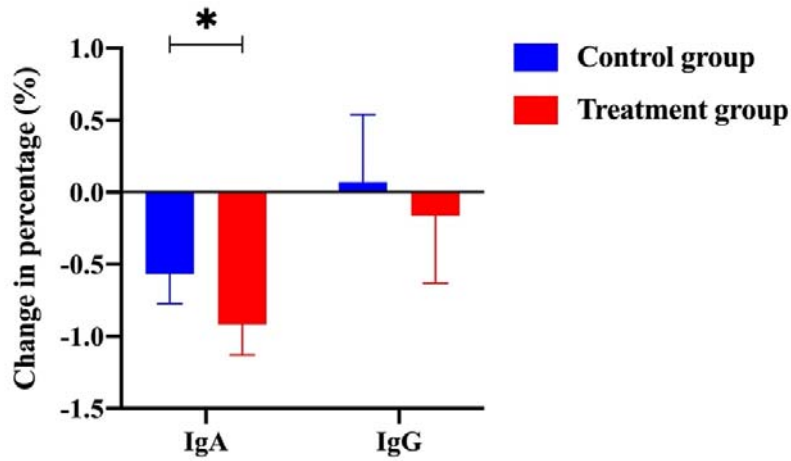


Figure 5. The change of percentage for IgA level and IgG level from the control group and the treatment group (with TCM treatment).

Figure 5 – source data. Patients’ raw data of serum IgA and IgG level from the control group (without TCM treatment) and the treatment group (with TCM treatment) in the clinical trial (ClinicalTrials.gov: NCT05530226).

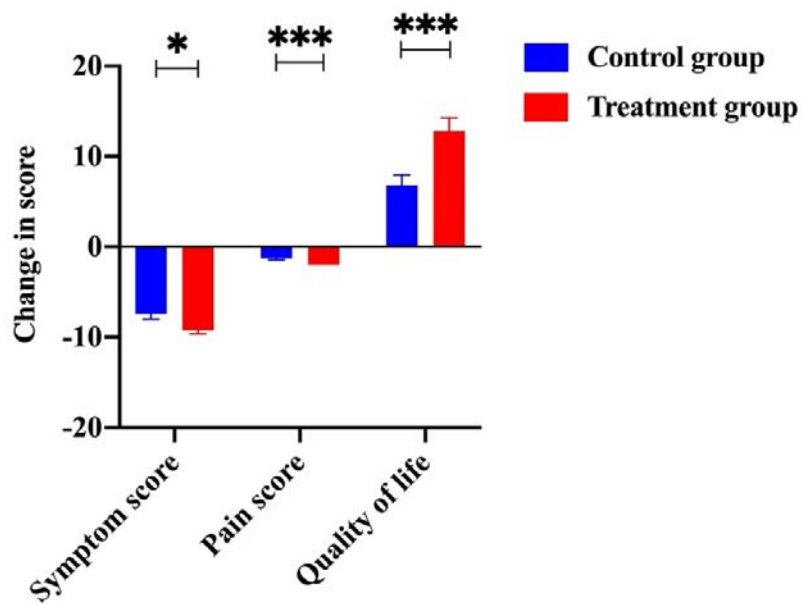


Figure 6. The change of scores from items including symptom, pain and living quality. The scores of symptoms, pain and living quality from the treatment group were significantly improved as compared to the control group.

Figure 6 – source data 1. Patients’ raw data of symptom scores from the control group (without TCM treatment) and the treatment group (with TCM treatment) in the clinical trial (ClinicalTrials.gov: NCT05530226).

Figure 6 – source data 2. Patients’ raw data of pain scores from the control group (without TCM treatment) and the treatment group (with TCM treatment) in the clinical trial (ClinicalTrials.gov: NCT05530226).

Figure 6 – source data 3. Patients’ raw data of quality of life from the control group (without TCM treatment) and the treatment group (with TCM treatment) in the clinical trial (ClinicalTrials.gov: NCT05530226).

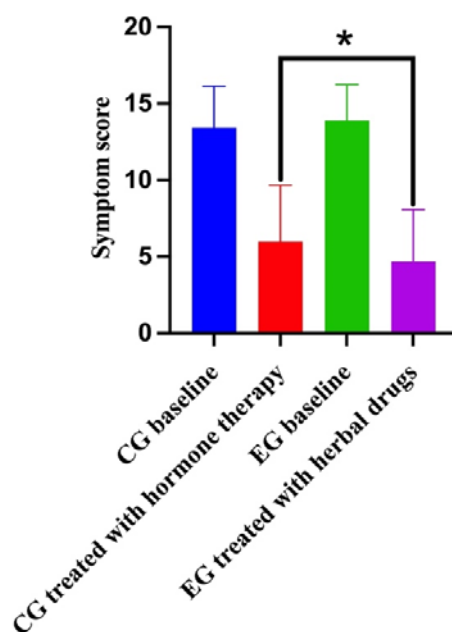


Figure 6 - Figure supplement 1. The comparison of symptom score between different groups. CG baseline (control group baseline), EG baseline (experimental group baseline); CG treated with methylprednisolone (control group treated with methylprednisolone), EG treated with herbal drugs (experimental group treated with herbal drug combinations).

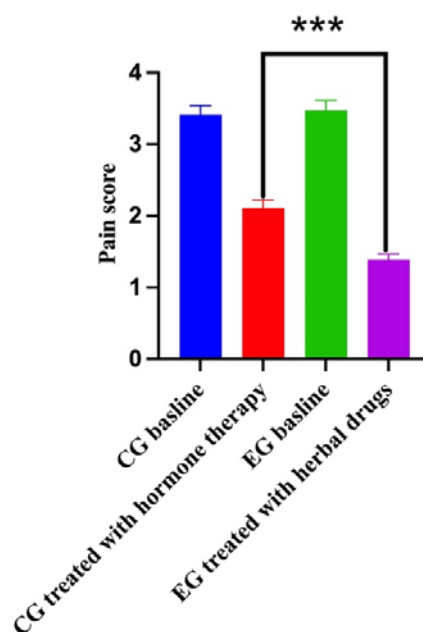


Figure 6 - Figure supplement 2. The comparison of pain score between different groups. CG baseline (control group baseline), EG baseline (experimental group baseline); CG treated with methylprednisolone (control group treated with methylprednisolone), EG treated with herbal drugs (experimental group treated with herbal drug combinations).

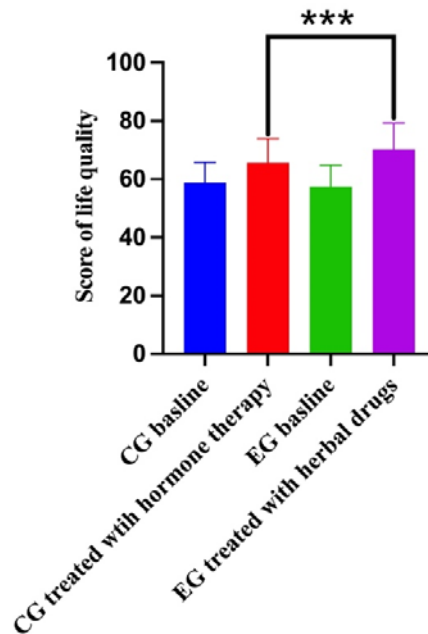


Figure 6 - Figure supplement 3. The comparison of score of life quality between different groups. CG baseline (control group baseline), EG baseline (experimental group baseline); CG treated with methylprednisolone (control group treated with methylprednisolone), EG treated with herbal drugs (experimental group treated with herbal drug combinations).

679

680

681

682

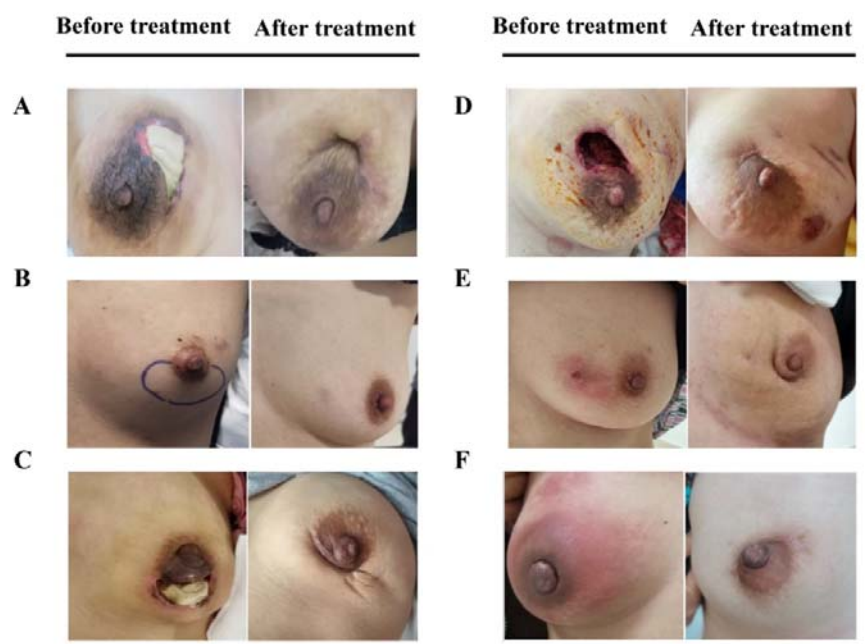
683

684

685

686

687 **Figure 7.** The comparison of whole breast for six representative PCM patients in the
688 EG group before treatment and after treatment (written informed consents have been
689 provided from all the patients; some detailed information was provided in the
690 Supplemental Files).
691



692 **Table 1.** Comparison of operation rate, recurrence rate and incidence of adverse
693 reactions between the two groups (EG: experimental group; CG: control group).

694

Groups	N (Patients)	Operation	Recurrence	Incidence of adverse events
EG	80	25 (31.25%)	3 (3.75%)	5 (6.25%)
CG	80	47 (58.75%)	10 (12.5%)	9 (11.25%)
X ²		12.22	4.103	1.252
p		<0.001	0.043	0.263

695

696

697 **Table 2.** Clinical symptom scores between the two groups in the clinical trial (EG:
698 experimental group; CG: control group; The clinical symptom rating scale for PCM
699 was displayed in Supplemental file 1B).

700

701

Groups	N (Patients)	Baseline	Treated	t	p
EG	80	13.90±2.37	4.68±3.40	22.19	<0.001
CG	80	13.423±2.70	5.98±3.68	14.282	<0.001
t		1.189	2.323		
p		0.236	0.021		

702

703

References

1. Cutler, M. Plasma-cell mastitis; report of a case with bilateral involvement. *Br Med J* **1**, 94-96 (1949).
2. Yu, J.J., *et al.* Mouse model of plasma cell mastitis. *J Transl Med* **10 Suppl 1**, S11 (2012).
3. Cheng, L., *et al.* Mastitis, a Radiographic, Clinical, and Histopathologic Review. *Breast J* **21**, 403-409 (2015).
4. Fleming, L., *et al.* The impact of medication side effects on adherence and persistence to hormone therapy in breast cancer survivors: A quantitative systematic review. *Breast* **64**, 63-84 (2022).
5. Tu, Y. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nat Med* **17**, 1217-1220 (2011).
6. Tu, Y. Artemisinin-A Gift from Traditional Chinese Medicine to the World (Nobel Lecture). *Angew Chem Int Ed Engl* **55**, 10210-10226 (2016).
7. Yin, X., *et al.* Effect of Electroacupuncture on Insomnia in Patients With Depression: A Randomized Clinical Trial. *JAMA Netw Open* **5**, e2220563 (2022).
8. Zhang, J., Xu, J., Zhang, J. & Ren, Y. Chinese herbal compound combined with western medicine therapy in the treatment of plasma cell mastitis: A protocol for systematic review and meta-analysis. *Medicine (Baltimore)* **99**, e22858 (2020).
9. Li, S., Zhang, B., Jiang, D., Wei, Y. & Zhang, N. Herb network construction and co-module analysis for uncovering the combination rule of traditional Chinese herbal formulae. *BMC Bioinformatics* **11 Suppl 11**, S6 (2010).
10. Ye, Q., *et al.* A unified drug-target interaction prediction framework based on knowledge graph and recommendation system. *Nat Commun* **12**, 6775 (2021).
11. Zeng, X., Tu, X., Liu, Y., Fu, X. & Su, Y. Toward better drug discovery with knowledge graph. *Curr Opin Struct Biol* **72**, 114-126 (2022).
12. Jin, W., *et al.* Deep learning identifies synergistic drug combinations for treating COVID-19. *Proc Natl Acad Sci U S A* **118**(2021).
13. Zhang, Y., *et al.* Checkpoint therapeutic target database (CKTTD): the first comprehensive database for checkpoint targets and their modulators in cancer immunotherapy. *J Immunother Cancer* **8**(2020).
14. Hao, Y.F. & Jiang, J.G. Origin and evolution of China Pharmacopoeia and its implication for traditional medicines. *Mini Rev Med Chem* **15**, 595-603 (2015).
15. Reimand, J., *et al.* Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc* **14**, 482-517 (2019).
16. Nan, Q., *et al.* Mechanism underlying efficacy of Shugan Sanjie decoction on plasma cell mastitis, based on network pharmacology and experimental verification. *J Tradit Chin Med* **42**, 400-407 (2022).

- 746 17. Liu, Y., *et al.* Sinomenine hydrochloride inhibits the progression of plasma
747 cell mastitis by regulating IL-6/JAK2/STAT3 pathway. *Int Immunopharmacol*
748 **81**, 106025 (2020).
- 749 18. Jassal, B., *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res* **48**,
750 D498-D503 (2020).
- 751 19. Lv, J., *et al.* Effect of Oral Methylprednisolone on Clinical Outcomes in
752 Patients With IgA Nephropathy: The TESTING Randomized Clinical Trial.
753 *JAMA* **318**, 432-442 (2017).
- 754 20. Liu, Y., *et al.* Activation of the IL-6/JAK2/STAT3 pathway induces plasma
755 cell mastitis in mice. *Cytokine* **110**, 150-158 (2018).
- 756 21. Xing, M., Zhang, S., Zha, X. & Zhang, J. Current Understanding and
757 Management of Plasma Cell Mastitis: Can We Benefit from What We Know?
758 *Breast Care (Basel)* **17**, 321-329 (2022).
- 759 22. Betts, C.B., *et al.* Mucosal Immunity in the Female Murine Mammary Gland.
760 *J Immunol* **201**, 734-746 (2018).
- 761 23. Bharathan, M. & Mullarky, I.K. Targeting mucosal immunity in the battle to
762 develop a mastitis vaccine. *J Mammary Gland Biol Neoplasia* **16**, 409-419
763 (2011).
- 764 24. Yang, Y., Adelstein, S.J. & Kassis, A.I. Target discovery from data mining
765 approaches. *Drug Discov Today* **14**, 147-154 (2009).
- 766 25. Liu, Y., *et al.* IL-6/STAT3 signaling pathway is activated in plasma cell
767 mastitis. *Int J Clin Exp Pathol* **8**, 12541-12548 (2015).
768
769