

基于随机森林回归模型的中药复方金复康 对肺癌细胞增殖作用的组方优化

郑婷婷^{1,2}, 杨铭³, 阙祖俊², 金郁⁴, 焦丽静², 赵小东⁵, 许玲^{1*}

(1. 上海中医药大学附属岳阳中西医结合医院, 上海 200437; 2. 上海中医药大学附属龙华医院中医肿瘤研究所, 上海 200032; 3. 上海中医药大学附属龙华医院, 上海 200032; 4. 华东理工大学药学院, 上海 200237; 5. 上海交通大学生物医学工程学院, 上海 200240)

[摘要] 目的:以治疗肺癌疗效确切的中药复方金复康为研究对象,从抑制肺癌细胞增殖方面对其处方组成进行优化,并为中药复方的处方优化提供新的途径。方法:应用筛选试验设计(Plackett-Burman),然后应用随机森林算法,建立基于设计处方集的预测模型,61个组方设计矩阵为输入,以相应的61个组方的半数抑制浓度(half maximal inhibitory concentration, IC_{50})的自然对数为输出建立随机森林回归模型。并用5次10折交互验证优化模型参数 $mtry$,最后应用网格化搜索算法得到最优的配伍组,并进行肺癌细胞增殖实验验证,以金复康组成药物的不同组合作用于肺腺癌细胞株 A549 的 IC_{50} 为指标,从抑制肺癌细胞增殖方面对金复康的组方配伍进行优化。结果:通过随机森林模型结合网格化搜索算法得到的最优配伍组为黄芪、麦冬、重楼、女贞子和绞股蓝,且经实验验证,在抑制肺癌细胞增殖方面,金复康优化方相对于金复康原方具有更好的抑制细胞增殖效应。结论:基于随机森林模型结合网格化搜索算法能够为复杂性中药复方的组方优化提供方法学参考。

[关键词] 随机森林;金复康;细胞增殖;组方优化

[中图分类号] R285 **[文献标识码]** A **[文章编号]** 1005-9903(2017)04-0177-06

[doi] 10.13422/j.cnki.syfjx.2017040177

[网络出版地址] <http://www.cnki.net/kcms/detail/11.3495.R.20161129.1056.016.html>

[网络出版时间] 2016-11-29 10:56

Optimization of Effect of Traditional Chinese Medicine Compound Jinfukang on Lung Cancer Cell Proliferation Based on Random Forest Regression Model

ZHENG Ting-ting^{1,2}, YANG Ming³, QUE Zu-jun², JIN Yu⁴, JIAO Li-jing², ZHAO Xiao-dong⁵, XU Ling^{1*}

(1. Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine (TCM), Shanghai 200437, China; 2. Cancer Institute of Longhua Hospital, Shanghai University of TCM, Shanghai 200032, China; 3. Longhua Hospital, Shanghai University of TCM, Shanghai 200032, China; 4. School of Pharmacy, East China University of Science and Technology, Shanghai 200237, China; 5. School of Biomedical Engineering, Shanghai Jiaotong University, Shanghai 200240, China)

[Abstract] **Objective:** To study the accurate curative effect of traditional Chinese medicines (TCM) compound Jinfukang in treating lung cancer, and optimize the prescription composition to inhibit lung cancer cell proliferation, so as to provide a new approach for the optimization of TCM compound prescription. **Method:** Screening test design (Plackett-Burman) and random forest algorithm were applied in establishing the prediction model based on design prescription set. With 61 composition design matrixes as the input, and natural logarithm of

[收稿日期] 20160419(019)

[基金项目] 国家自然科学基金项目(81373623);上海市教育委员会科研创新项目(15ZZ066);浦东新区卫生系统重点学科群建设项目(PWZXq2014-12);上海市卫生和计划生育委员会科研项目(20134173)

[第一作者] 郑婷婷,在读硕士,从事中医药防治恶性肿瘤研究,Tel:15216679182,E-mail:1120437130@qq.com

[通讯作者] *许玲,教授,博士生导师,从事中医药防治恶性肿瘤研究,E-mail:xulq67@aliyun.com

their corresponding IC_{50} as the output, random forest regression model was established. The optimization model parameter entry was verified by 5-time and 10-fold interaction. Finally, a grid search algorithm was applied to get the optimal group, and verify the lung cancer cell proliferation experiment. With the IC_{50} obtained from different formula of Jinfukang on A549 lung adenocarcinoma cancer cells as an index, the composition of Jinfukang were optimized to inhibit lung cancer cell proliferation. **Result:** Through random forest model combined with grid search algorithm, we got the optimal compatible herbs, namely Astragali Radix, Ophiopogonis Radix, Parisidis Rhizoma, Ligustri Lucidi Fructus and Gynostemmatis Pentaphylli Herba seu Radix. Through the experiment, optimized Jinfukang showed a better effect in cell inhibiting and proliferation than original Jinfukang in the aspects of inhibition of lung cancer cell proliferation. **Conclusion:** The random forest model combined with the grid search algorithm complexity can provide methodology reference for the optimization of composition of complex TCM compound formula.

[**Key words**] random forest; Jinfukang; cell proliferation; composition optimization

金复康口服液(简称金复康)是由上海市名中医刘嘉湘教授研发的治疗肺癌的中成药,主要由黄芪、北沙参、麦冬、天冬、女贞子、石见穿等 12 味药组成,具有益气养阴、清热解毒的功效。金复康作为治疗肺癌的治疗药物广泛应用于临床,前期临床研究表明金复康作为铂类化疗药物的辅助药物可稳定非小细胞肺癌患者的病灶,有效抑制 III 期肺癌患者的远处转移,改善患者的临床症状,提高患者的生存质量,调节肺癌患者的免疫功能,明显减轻化疗患者的毒副反应^[1-2]。体内研究表明金复康可以抑制人肺癌细胞异种移植的裸小鼠的肿瘤生长并减小肿瘤体积。体外研究表明金复康可以抑制肺癌细胞的增殖,诱导凋亡并改变相关基因的表达^[3-4]。研究已经证实,肺癌的发生发展经历肺癌细胞在原发灶的增殖、突破基底黏附层,进入血管,随着血流在远处脏器停留,新生血管形成,最终肺癌细胞在远处脏器形成转移灶^[5]。那么,金复康主要是在哪一个环节发挥作用,或者是否存在金复康中某几味药的联合在某一环节发挥着更重要的作用,因此,本研究通过应用 R 语言程序包,利用随机森林回归,以不同的金复康组方药物作用于肺腺癌细胞株 A549 所得到的半抑制浓度为评价指标,以细胞增殖相关实验进行验证,建立金复康优化方,这是一次基于有效中成药进行靶点更明确的组方优化的探索研究。

1 材料

处方中饮片均购自上海华宇药业有限公司,经上海中医药大学附属龙华医院药剂科杨铭副主任药师鉴定,所用中药饮片均符合 2015 年版《中国药典》要求。

金复康的 12 味药(黄芪、北沙参、麦冬、女贞子、山茱萸、绞股蓝、淫羊藿、葫芦巴、石上柏、石见

穿、重楼和天冬)设计成 62 个组方,每个组方的药物采用醇提物,其制备方法为将各组方药物 100 g 使用粉碎机进行粉碎,每组药物加 5 倍体积的 70% 乙醇和 30% 的超纯水,超声提取 3 次,每次 60 min,取上清液,用旋转蒸发仪回收乙醇,后置于 $-80\text{ }^{\circ}\text{C}$ 冰箱过夜,次日使用冷冻干燥机使其冻干成粉^[6]。

F12K 培养基,胎牛血清(Gibco 公司,批号分别为 21127-022, 10099-141), CCK-8 试剂盒(日本 Dojindo 公司,批号 JR745)。细胞计数板(上海浩然生物技术有限公司,批号 15003471), 96 孔板(美国 costar 公司,批号 NY14831)。人肺腺癌 A549 细胞株(中科院上海细胞库,目录号为 TCHu150)。

MCO-18AIC 型二氧化碳恒温培养箱(日本 Sanyo 公司), BAS323S-CW 型电子分析天平(德国 Sartorius 公司), SpectraMax M5 型多功能酶标仪(美国 MD 公司), DK-600 型电热恒温水浴槽(上海精宏实验设备有限公司), 5418 型高速离心机(德国 Eppendorf 公司), Cedex XS 型细胞计数仪(瑞士 Roche 公司), Milli-Q Integra13 型超纯水机(美国 Millipore 公司)。

2 方法

2.1 主效应筛选设计 以 12 味药为因素,并分别以“用”与“不用”为两水平,应用 Plackett-Burman 筛选试验设计,产生 60 个设计处方,加上空白组(0 味中药组)及金复康全方组共计 62 个实验组,然后采用 CCK-8 法测出每个组方的半数抑制浓度(IC_{50})。

2.2 金复康组方药物对肺癌细胞增殖活性的检测 将 A549 细胞于含 10% 胎牛血清和 1% 双抗的 F12K 培养液中, $37\text{ }^{\circ}\text{C}$ 5% CO_2 的培养箱中培养, 2 ~ 3 d 传代 1 次。取对数生长期细胞,胰酶消化、离心、重悬、计数、制成细胞悬液,用培养液调整细胞密度

为 5×10^4 个/mL, 然后将细胞每孔 100 μL 接种于 96 孔板中, 静置 3 min, 然后置于培养箱中继续培养, 24 h 后依次加入 62 个不同组方的药物, 每个组方的质量浓度设为 1 000, 500, 250, 125, 62.5 $\text{mg} \cdot \text{L}^{-1}$ 组及空白组 (0 $\text{mg} \cdot \text{L}^{-1}$), 每个组方的每个浓度均设 5 个复孔, 混匀后置于培养箱中继续培养, 48 h 后将每孔加入 CCK-8 试剂 10 μL , 置于培养箱中, 1 h 后在波长为 490 nm 处测定其吸光度 A 。每个组方的 IC_{50} 均在 GraphPad Prime5 软件中计算得出^[7]。

$$\text{细胞存活率} = A_{\text{试验组}} / A_{\text{对照组}} \times 100\%$$

CCK-8 实验结果显示, 第 1 ~ 61 个配伍组的 IC_{50} 用均数加减标准差形式表示为 (746.49 \pm 657.93) $\text{mg} \cdot \text{L}^{-1}$, 其中金复康原方的 IC_{50} 为 574.2 $\text{mg} \cdot \text{L}^{-1}$ 。60 个组方的 IC_{50} 见表 1。

表 1 金复康 60 个组方的 IC_{50}

| No. | IC_{50} | No. | IC_{50} | No. | IC_{50} | No. | IC_{50} |
|-----|------------------|-----|------------------|-----|------------------|-----|------------------|
| 1 | 1 108 | 16 | 122.2 | 31 | 638.1 | 46 | 2 182 |
| 2 | 1 151 | 17 | 1 916 | 32 | 167.7 | 47 | 678.1 |
| 3 | 563.9 | 18 | 809.4 | 33 | 751.7 | 48 | 5 378 |
| 4 | 249.6 | 19 | 425.3 | 34 | 600.5 | 49 | 2 449 |
| 5 | 1 126 | 20 | 162.6 | 35 | 2 122 | 50 | 443.3 |
| 6 | 1 401 | 21 | 245.6 | 36 | 1 253 | 51 | 17 249 |
| 7 | 71.21 | 22 | 602.4 | 37 | 129.2 | 52 | 127.1 |
| 8 | 180.1 | 23 | 136.8 | 38 | 248.4 | 53 | 260.8 |
| 9 | 1 884 | 24 | 515 | 39 | 388.7 | 54 | 1 190 |
| 10 | 504.9 | 25 | 2 026 | 40 | 3 695 | 55 | 301.9 |
| 11 | 375.1 | 26 | 272.2 | 41 | 6 756 | 56 | 343.1 |
| 12 | 1 236 | 27 | 1 243 | 42 | 382.9 | 57 | 316.2 |
| 13 | 281 | 28 | 370.5 | 43 | 472.6 | 58 | 482.3 |
| 14 | 390.7 | 29 | 1 128 | 44 | 98 955 | 59 | 591.4 |
| 15 | 2 798 | 30 | 625.5 | 45 | 297.1 | 60 | 491.1 |

2.3 随机森林模型建立 随机森林算法是一类集合模型, 可用于回归与分类。它通过多次 bootstrap 抽样从原始样本中抽取多个随机样本, 并对每个 bootstrap 样本进行决策树建模, 然后将这些决策树组合在一起, 通过投票得到最终分类或者预测的结果。在组方优化问题中, 随机森林可解释若干个药物 (X_1, X_2, \dots, X_k) 对该配伍组药理效应的作用, 在建立分类/回归树的时候, 随机森林会应用 Bootstrap 重新抽样法随机的在原始设计处方数据集重新选择 n 个观测值, 其中有的配伍组会被选多次, 有的配伍组则没被选到。与此同时, 随机森林算法随机的从 k 个药物变量中选择部分药物变量进行分类/

回归树节点确定, k 值即模型 mtry 值可由交叉验证优化得到^[8-9]。

本研究使用 R 语言的随机森林包构建组方配伍与药效间的随机森林回归模型, 即以 61 个组方设计矩阵为输入, 以相应的 61 个组方的 IC_{50} 的自然对数为输出建立随机森林回归模型。在其他参数默认的情况下, 以模型拟合的均方根误差 (root-mean-square error, RMSE) 为指标, 采用 5 次 10 折交互验证优化模型参数 mtry, 结果见表 2, 当 mtry 赋值为 5 时, 其对应的 RMSE 最低为 0.857 2^[10]。

表 2 金复康 61 个组方 IC_{50} 的 5 次 10 折交互验证

Table 2 Results of 5-time and 10-fold interaction validation of IC_{50} of sixty-one formulas from Jinfukang

| mtry | RMSE |
|------|---------|
| 2 | 0.881 0 |
| 3 | 0.873 2 |
| 4 | 0.865 5 |
| 5 | 0.857 2 |
| 6 | 0.866 6 |
| 7 | 0.857 2 |
| 8 | 0.873 2 |
| 9 | 0.865 9 |
| 10 | 0.880 5 |
| 11 | 0.879 3 |
| 12 | 0.880 7 |

2.4 随机森林模型的检验 建模当中的一项重要内容就是模型的检验和评价, 通过模型的检验来评估模型的拟合效果。该研究模型的评价指标主要是模型的决定系数 (R^2) 和均方根误差 (RMSE)^[11]。通过模型的 R^2 和 RMSE 对模型的拟合效果进行评价, 在得到优化的模型参数后, 重新拟合模型得到的 $R^2 = 0.825 9$, $\text{RMSE} = 0.857 2$ 模型拟合效果较理想。且从预测值和实测值散点图 1, 可以看出预测值和实测值吻合较好。

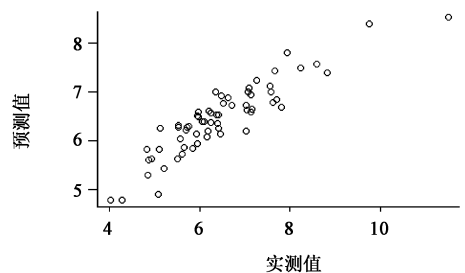


图 1 随机森林模型实测值和预测值的拟合度

Fig. 1 Fitting degree of measured values and predicted values based on random forest regression

3 结果

3.1 模型预测 在此模型的基础上运用网格化搜索算法优化金复康的组方,网格化搜索的基本思想是遍历药物的所有配伍组合,将各变量因子按照步长分成若干个水平,把各变量因子的不同水平组合起来,代入回归方程,筛选目标变量的最优解^[12-14]。结果得到最佳的组合是黄芪、麦冬、重楼、女贞子和

绞股蓝。这个组合的 IC_{50} 自然对数值的模型预测结果为 4.79。

3.2 模型验证 在得到最佳组合后,再从模型的预测结果中选择两个相对较好的组方,进行体外的细胞实验验证,结果见表 3,3 个验证组方的 IC_{50} 自然对数值的实验值与模型预测值的偏差较小,预测 RMSE 为 0.682 8。

表 3 优选组方的验证试验

Table 3 Results of experiment verification of optimization formula

| 处方 | 处方组成 | IC_{50} 自然对数值 | |
|------|----------------------|-----------------|------|
| | | 实验值 | 预测值 |
| 最佳处方 | 黄芪,麦冬,重楼,女贞子,绞股蓝 | 4.27 | 4.79 |
| 处方一 | 黄芪,麦冬,重楼,绞股蓝 | 5.07 | 4.91 |
| 处方二 | 黄芪,麦冬,重楼,女贞子,绞股蓝,淫羊藿 | 6.14 | 5.09 |

4 金复康优化方和金复康全方在抑制细胞增殖方面的比较

4.1 CCK-8 法检测 2 个组方对 A549 细胞活性的影响 取对数生长期细胞,胰酶消化、计数、制成细胞悬液,调整细胞密度为 5×10^4 个/mL,然后将细胞以每孔 100 μ L 接种于 96 孔板中,置于培养箱中培养 24 h 后加入 2 个组方不同浓度的药物,每组 5 复孔,混匀后置于培养箱中继续培养 48 h 后,加入 CCK-8 试剂 10 μ L,1 h 后于波长 490 nm 处测定其吸光度 A。 IC_{50} 均在 GraphPad Prime5 软件中计算得出。

研究表明,金复康全方和金复康优化方都可明显抑制 A549 细胞的活性,但是金复康优化方的效果更明显,金复康优化方的 IC_{50} 为 $71.21 \text{ mg} \cdot \text{L}^{-1}$,金复康全方的 IC_{50} 为 $574.2 \text{ mg} \cdot \text{L}^{-1}$,结果见图 2,其中 7 号方为金复康优化方,62 号方为金复康全方。

400 个/孔的细胞数接种于 6 孔板中,2 mL/孔,每组 3 复孔,用含 20% FBS 的 RPMI-1640 培养基进行培养。加药干预后于培养箱中连续培养 12 d 后,吸弃上清,用 4% 的多聚甲醛室温固定细胞 60 min,吸弃固定液,加入吉姆萨染液染色 20 min,吸弃染液,将 6 孔板竖直浸入水中,洗净板上染液,晾干,对克隆进行计数,并在扫描仪上进行扫描。

研究表明,金复康全方和金复康优化方都能抑制细胞克隆形成,但金复康优化方的效果更明显,且差异具有统计学意义 ($P < 0.05$),结果见图 3。

4.3 2 个组方对 A549 细胞周期的影响 将对数生长期的 A549 细胞以 1×10^5 个/孔接种于 6 孔板中,加药处理 48 h 后,收集细胞,用 PBS 洗 1 次,弹击 EP 管底,使细胞沉淀分散,加入预冷的 80% 乙醇重悬沉淀,4 $^{\circ}\text{C}$ 固定过夜。次日离心弃上清,用 PBS 洗 1 遍,加入 RnaseA 100 μ L,37 $^{\circ}\text{C}$ 孵育 30 min,在避光环境下加入 PI 染液 400 μ L 混匀,4 $^{\circ}\text{C}$ 避光孵育 30 min,流式细胞仪检测,结果用 Flowjo 软件多正态拟合程序进行曲线拟合分析,计算 DNA 含量,得出细胞 G_1 期,S 期, G_2/M 期的比率。

研究表明,金复康全方和金复康优化方都能将细胞周期阻滞在 G_2 期,但金复康优化方的效果更明显,且差异具有统计学意义 ($P < 0.05$),结果见图 4。

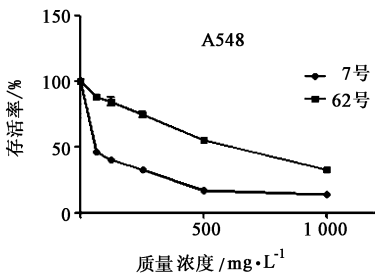


图 2 7 号方和 62 号方对 A549 细胞活性的影响 ($\bar{x} \pm s, n = 3$)

Fig. 2 Effect of No. 7 and No. 62 formulas on A549 cell activity ($\bar{x} \pm s, n = 3$)

4.2 检测 2 个组方对细胞克隆形成的影响 取对数生长期的细胞,消化后制成细胞悬液,计数,以

5 讨论

中药复方一般药味比较多,成分比较复杂,尽管临床疗效确切,但是对于作用环节的明确、客观适应证的把握等方面都存在着困难。在精准医学的发展要求下,中成药的作用靶点的明晰,适应证的具体化

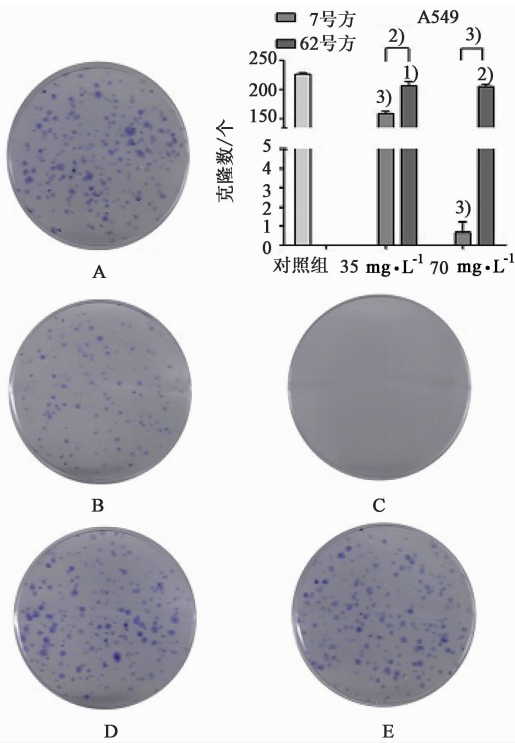


图 3 7 号方和 62 号方对 A549 细胞克隆形成的影响
A. 对照组 ($0 \text{ mg}\cdot\text{L}^{-1}$); B. 7 号方 $35 \text{ mg}\cdot\text{L}^{-1}$ 组; C. 7 号方 $70 \text{ mg}\cdot\text{L}^{-1}$ 组; D. 62 号方 $35 \text{ mg}\cdot\text{L}^{-1}$ 组; E. 62 号方 $70 \text{ mg}\cdot\text{L}^{-1}$ 组; 与对照组比较¹⁾ $P < 0.05$, ²⁾ $P < 0.01$, ³⁾ $P < 0.001$ (图 4 同)

图 3 7 号方和 62 号方对 A549 细胞克隆形成的影响
Fig. 3 Effect of No. 7 and No. 62 formulas on A549 cell colony formation

已显现出重要性,因此,开展基于有效中成药的优化研究,是一种加强合理用药的积极探索。

对于组方优化的建模优化研究,一般情况下均是“小样本”情形,通常的处理流程为对组方数据进行建模,对模型进行评价,再在可接受的模型上寻优找到效应较佳的组合。由于本文的响应变量 (IC_{50} 的自然对数) 是连续型的,所以建立的模型是回归模型,笔者在预实验时发现简单的线性模型并不能很好的刻画组方配伍与药效之间的关联性,可见组方配伍与药效间的关系呈现复杂的非线性,随机森林算法是处理复杂非线性关系的一种表现较好的模型,在本文的分析中亦能发现建立的模型能够满足组方配伍效应仿真的需要。但是,“群举法”并不能解决建模问题,其只能在已建立的模型上进行探讨,本文的寻优思路其实就是“群举法”,即在已建立的模型上,使用网格化搜索策略,“穷尽所有配伍组合”,并对其药效值进行仿真,然后得到最优的配伍组。而且,大量的研究都证明了随机森林算法有较高的预测准确率和较好的泛化能力和较少的调节参数等优点,且不易出现过度拟合,适用于各种数据

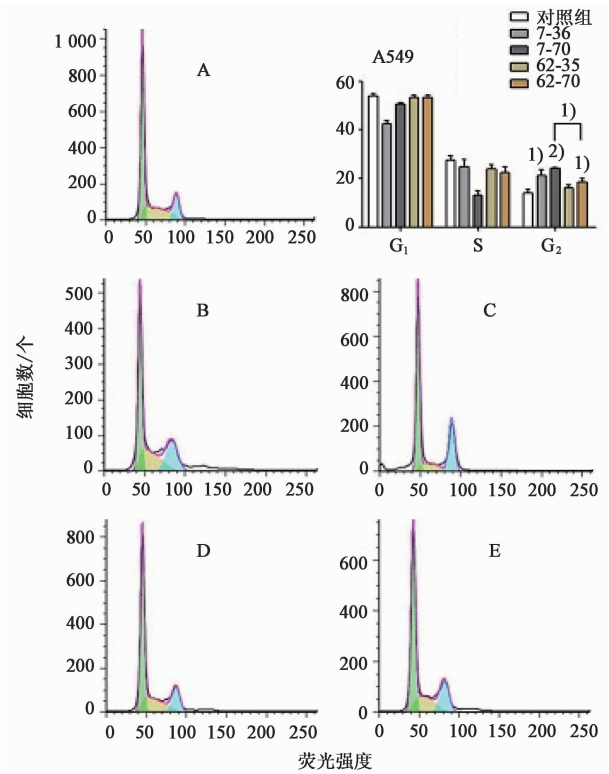


图 4 7 号方和 62 号方对 A549 细胞周期的影响
Fig. 4 Effect of No. 7 and No. 62 formulas on A549 cell cycle

集的运算,在医学领域得到了广泛的应用^[15]。因此,本研究利用该算法,结合肺癌增殖抑制研究方法,采用体外实验对金复康 62 个不同组方进行了筛选,获得了 5 味中药的组合方:黄芪补益脾肺之气,麦冬润肺生津,女贞子补益肝肾,绞股蓝益气健脾,上 4 味药以益气养阴扶正为主,而重楼清热解毒祛邪为主,绞股蓝也兼具清热解毒加强了祛邪的效应。金复康原方主要用于气阴两虚型的肺癌,本次优化获得的 5 味药仍保留了益气养阴为主的治则,同时明确了该 5 味药在抑制肿瘤增殖中起更重要的作用。

随机森林算法在本研究中的应用,使得多种药物组合的优化大大节省了时间和精力,提示数据挖掘的方法将有更广阔的空间应用于中医药的研究。一般中药复方的优化主要从两个方面入手:即组成的优化和剂量的优化^[16-17],本研究主要是对金复康组成进行优化。但是本次研究仅是通过体外细胞学的研究进行了初步尝试,动物体内的研究证实以及临床适应症的具体化及临床疗效评价等尚有许多工作需要。相信,随着科技发展及交叉学科的合作,必将有利于中医药的应用更加精准。

[参考文献]

[1] 孙玺媛,姜梅,张伟,等. 金复康口服液对非小细胞

- 人肺腺癌吉非替尼获得性耐药的影响[J]. 中药材, 2014, 37(7): 1254-1258.
- [2] 刘嘉湘, 潘敏求, 黎月恒, 等. 金复康口服液治疗原发性非小细胞肺癌临床研究[J]. 肿瘤, 2001, 21(6): 463-465.
- [3] 孙玺媛, 姜梅, 张忠太, 等. 金复康口服液对人肺腺癌耐吉非替尼 PC-9R 细胞凋亡的影响[J]. 辽宁中医杂志, 2014, 56(10): 2229-2232.
- [4] 孙建立, 刘嘉湘. 金复康口服液对裸鼠人肺腺癌细胞凋亡相关基因表达的影响[J]. 上海中医药杂志, 2007, 41(10): 69-71.
- [5] Wood S L, Pernemalm M, Crosbie P A, et al. The role of the tumor-microenvironment in lung cancer-metastasis and its relationship to potential therapeutic targets[J]. *Cancer Treat Rev*, 2014, 40(4): 558-566.
- [6] 蔡华丹. 茵陈蒿汤水提物和醇提物的成分及药效比较研究[D]. 北京:北京中医药大学, 2013.
- [7] XIA B, YANG S, LIU T, et al. miR-211 suppresses epithelial ovarian cancer proliferation and cell-cycle progression by targeting Cyclin D1 and CDK6[J]. *Mol Cancer*, 2015, 14(1):1-13.
- [8] Svetlichnyy D, Imrichova H, Fiers M, et al. Identification of high-impact cis-regulatory mutations using transcription factor specific random forest models [J]. *PLoS Comput Biol*, 2015, 11(11): e1004590.
- [9] Poona N, Van Niekerk A, Nadel R, et al. Random forest (RF) wrappers for waveband selection and classification of hyperspectral data[J]. *Appl Spectrosc*, 2016, 70(2): 322-333.
- [10] GUAN C, NIU X, SHI F, et al. Predicting a DNA-binding protein using random forest with multiple mathematical features[J]. *Biomed Mater Eng*, 2015, 26 Suppl 1: S1883-1889.
- [11] 雍凯. 随机森林的特征选择和模型优化算法研究[D]. 哈尔滨:哈尔滨工业大学, 2008.
- [12] 杨传华, 李运伦, 蒋海强, 等. 均匀设计和抗心肌缺血效应的人参健心胶囊组方优化研究[J]. 中国中西医结合杂志, 2015, 2015(9): 1105-1108.
- [13] 徐培平, 朱宇同, 张美义, 等. 多目标优化技术在中药复方药物筛选及组方优化中的应用[J]. 世界科学技术—中国药现代化, 2005, 7(2): 35-39.
- [14] Volken W, Frei D, Manser P, et al. An integral conservative gridding algorithm using *Hermitian curve* interpolation[J]. *Phys Med Biol*, 2008, 53(21): 6245-6263.
- [15] 曹正凤. 随机森林算法优化研究[D]. 北京:首都经济贸易大学, 2014.
- [16] 刘齐林, 李雪婷, 王沛, 等. 痛风宁微丸处方药物组方优化[J]. 中国实验方剂学杂志, 2010, 16(17): 137-139.
- [17] 杨伟华, 李运伦, 蒋海强, 等. 均匀设计和抗心肌缺血效应的人参健心胶囊组方优化研究[J]. 中国中西医结合杂志, 2015, 35(9): 1105-1108.

[责任编辑 邹晓翠]