

# 通过生物信息学验证*RN1H4*基因预测尤文肉瘤患者预后的作用

孙 楷<sup>1</sup>, 王春燕<sup>2</sup>, 于 壮<sup>1\*</sup>

<sup>1</sup>青岛大学附属医院, 山东 青岛

<sup>2</sup>潍坊市坊子区人民医院, 山东 潍坊

收稿日期: 2022年2月11日; 录用日期: 2022年3月4日; 发布日期: 2022年3月15日

## 摘 要

尤文肉瘤是恶性程度极高的骨原发肿瘤, 好发于儿童和青少年。尤文肉瘤预后极差, 具有高度的侵袭性, 在青少年儿童肿瘤发病率中排在第二位。我们基于GEO数据库(Gene Expression Omnibus)中尤文肉瘤相关的数据, 构建了一个由7个基因构成的风险分数计算模型。该模型联合年龄性别构建的预测模型可用于预测患者预后。在该风险分数计算模型中, 筛选出了*RN1H4*基因。该基因在高风险组中显著表达。*RN1H4*高表达患者预后较差, 辅助T细胞、单核细胞、中性粒细胞和树突细胞的浸润和*RN1H4*基因的表达量具有显著相关性。这些细胞浸润丰度也显著影响了病人预后。我们又检测了*RN1H4*基因高低表达两组差异基因。对差异基因做了GO和KEGG分析, 预测了*RN1H4*在尤文肉瘤中发挥的作用。揭示了*RN1H4*在尤文肉瘤中的作用机制。

## 关键词

生物信息学, 尤文肉瘤, 预后, *RN1H4*, GEO数据库

# Verifying the Function of *RN1H4* Gene Predicting Prognosis of Ewing Sarcoma by Bioinformatics

Kai Sun<sup>1</sup>, Chunyan Wang<sup>2</sup>, Zhuang Yu<sup>1\*</sup>

<sup>1</sup>The Affiliated Hospital of Qingdao University, Qingdao Shandong

<sup>2</sup>The People's Hospital of Fangzi District Weifang City, Weifang Shandong

Received: Feb. 11<sup>th</sup>, 2022; accepted: Mar. 4<sup>th</sup>, 2022; published: Mar. 15<sup>th</sup>, 2022

\*通讯作者 Email: yuzhuang2002@163.com

文章引用: 孙楷, 王春燕, 于壮. 通过生物信息学验证 *RN1H4* 基因预测尤文肉瘤患者预后的作用[J]. 临床医学进展, 2022, 12(3): 1758-1768. DOI: 10.12677/acm.2022.123253

## Abstract

Ewing sarcoma is a highly malignant primary tumor of bone, usually occurring in children and adolescents. Ewing sarcoma has a very poor prognosis and is highly aggressive, ranking second in the incidence of tumors in children and adolescents. We constructed a risk score calculation model consisting of 7 genes based on Ewing sarcoma data from the GEO Database (Gene Expression Omnibus). The prediction model combined with age and sex can be used to predict the prognosis of patients. The *RN1H4* gene was screened from the risk score model. The gene was significantly expressed in the high-risk group. The prognosis of patients with high expression of *RN1H4* was poor, and the infiltration of helper T cells, monocytes, neutrophils and dendritic cells was significantly correlated with the expression of *RN1H4*. The abundance of these infiltrates also significantly affected patient outcomes. We also detected the high and low expression of *RN1H4* gene. GO and KEGG enrichment analyses were performed to predict the role of *RN1H4* in Ewing's sarcoma. The mechanism of *RN1H4* in Ewing's sarcoma was revealed.

## Keywords

Bioinformatics, Ewing Sarcoma, Prognosis, *RN1H4*, GEO Database

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

尤文肉瘤(Ewing Sarcoma)是一种极度恶性的原发骨肿瘤,易复发,具有极高的侵袭性,在青少年肿瘤发病率中排在第二位[1] [2] [3] [4]。尤文肉瘤目前的治疗方案主要是手术切除、化疗和放疗。目前未发生转移的尤文肉瘤的五年生存率为 15%~75% [5]。但是还有约 25%的病人在诊断初期便发生了转移[5] [6]。目前靶向药物在肿瘤治疗中展现了广阔的前景,例如纳武利尤单抗治疗 PD-1 高表达的非小细胞肺癌[7]具有显著的疗效。靶向治疗在肉瘤中,尤其是尤文肉瘤中发展缓慢。所以,继续筛选一个新型的靶向分子,为尤文肉瘤靶向治疗发展提供潜在靶点。

核受体超家族 1H 组 4 (*NR1H4*)在肝脏和小肠中胆汁酸生成和分泌起到作用[8]。在正常情况下,该基因在调节胆汁分泌、稳定肝肠循环有重要作用[9]。它也可以通过调节 *MYC* 通路影响到结肠癌肿瘤细胞的存活[10]。同时也可以影响 *Ras* 通路和 DNA 的甲基化[11]。该基因但是该基因在尤文肉瘤中的作用尚未探索。该基因与肿瘤免疫微环境的关系尚未阐明。

## 2. 方法

### 2.1. 数据下载

从 GEO 数据库(<https://www.ncbi.nlm.nih.gov/geo/>)中下载 2 个数据集(GSE17679、GSE63155) [12] [13],将两个数据集合并后,数据进行清洗和标准化。

### 2.2. 构建风险分数计算公式

依据获得的合并表达矩阵和病人预后,将病人分为练习组和检验组,数量比为 7:3。在训练组中通过

多重 COX 回归分析筛选和构建风险分数计算公式，同时在检验组中进行检验。依据风险分数的中位值将病人分为高低风险两组。风险分数联合年龄、性别构建列线图。通过 ROC 曲线和检验风险分数计算公式预测生存的准确性。通过 Calibration 曲线验证列线图预测能力。

### 2.3. 表达差异基因筛选和富集分析

依据 *NR1H4* 基因表达量中位值，将病人分成 *NR1H4* 高低表达两组。利用 R 语言软件和“limma 包” [14] 计算差异基因。差异基因筛选标准为： $\text{adjust P value} < 0.05$  且  $|\text{LogFoldChang}| > 0.5$ 。获得的差异表达基因进行 KO 和 KEGG 的富集分析。GO 和 KEGG 富集分析通过“clusterProfiler 包” [13]。

### 2.4. 免疫细胞浸润相关分析

通过 CIBERSORT [15] 算法对所有病人 22 种免疫细胞进入情况进行计算。免疫细胞分布和 *NR1H4* 基因表达量进行相关性分析，统计方法是 Spearman 法。

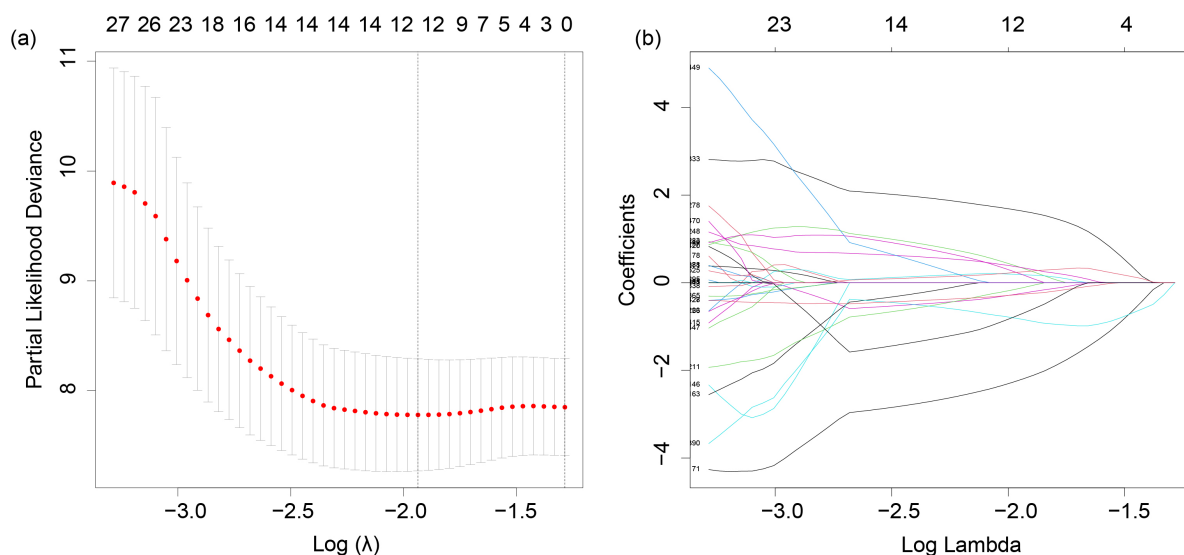
### 2.5. 统计学方法

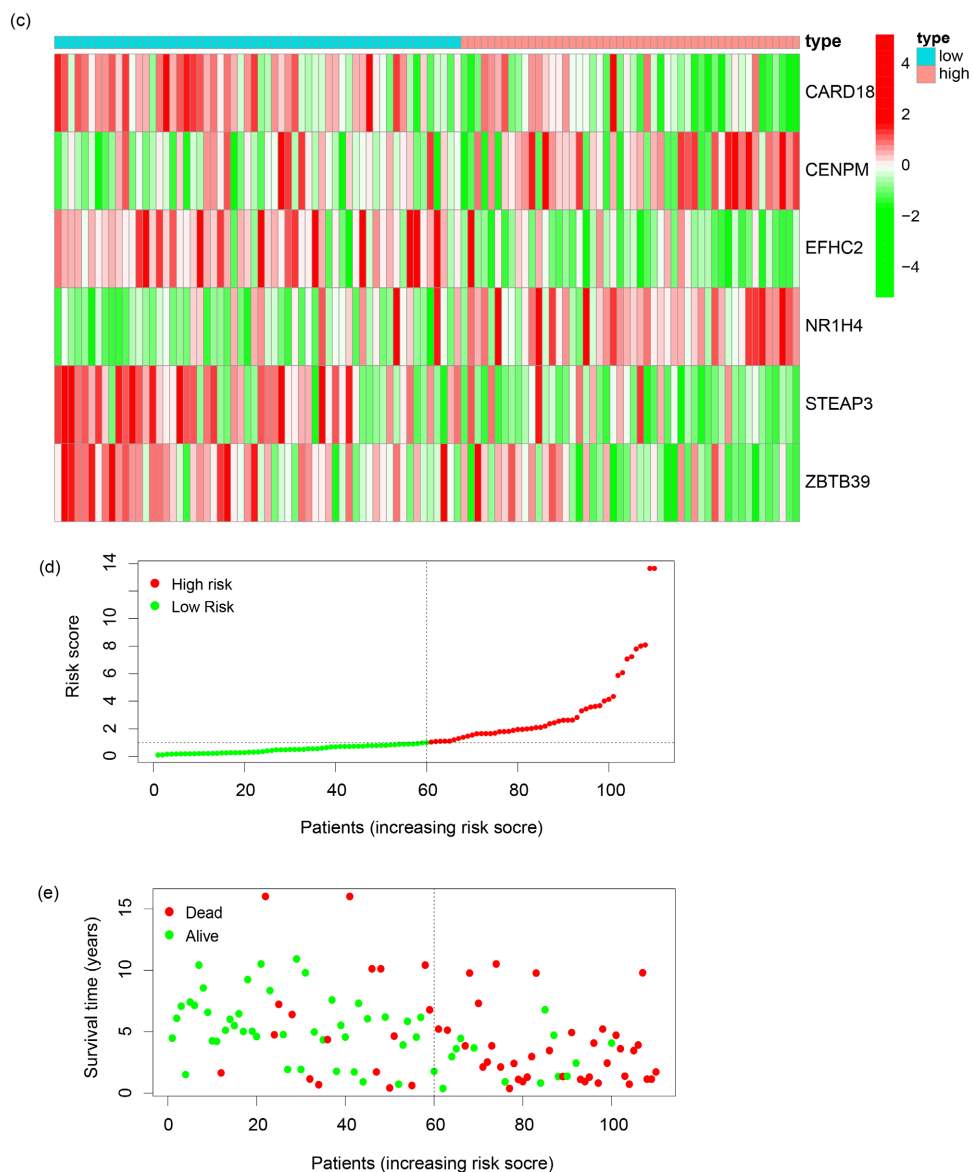
该研究中差异性分析使用 T 检验进行分析，设定  $P < 0.05$  为具有统计学差异。

## 3. 结果

### 3.1. 构建风险分数计算

在训练组中，通过 Lasso 回归和多重 COX 模型(图 1(a)，图 1(b))，构建出了基于 6 个基因组成的风险分数计算公式。基因分别是 *CARD18*、*CENPM*、*EFHC2*、*NR1H4*、*STEAP3*、*ZBTB39*。通过热图展现在各个病人中，六种基因的表达情况(图 1(c))。其对应的系数由表 1 所示。风险分数是每个病人 6 种基因的表达量乘以基因各自的系数的总和。通过图 1(d)和图 1(e)来展现，高风险组病人生存期较低风险组病人生存期短，高风险组病人死亡比例较低风险组病人高。ROC 曲线显示，该风险分数计算公式预测病人预后具有良好的准确性(图 2(a)~(c))。风险分数计算公式为  $\text{Riskscore} = \text{CARD18} * -3.624607368 + \text{CENPM} * 2.565920347 + \text{NR1H4} * 2.57112032 + \text{EFHC2} * -1.648318634 + \text{STEAP3} * -1.494970509 + \text{ZBTB39} * -4.248323426$ 。





**Figure 1.** (a), (b): Six genes were screened out by Lasso regression and multiple Cox regression; (c): The expression of six genes in all patients were showed by heat map; (d): Dot plot shows the risk scores distribution in high-/low-risk group. (e): Dot plot shows the survival state and overall survival in high-/low-risk group.

**图 1.** (a), (b): Lasso 回归多重 Cox 回归筛选 6 个基因; (c): 热图显示所有病人 6 个基因的表达量; (d): 点图表示高低风险组病人风险分数分布情况; (e): 高低风险组病人生存状态差异

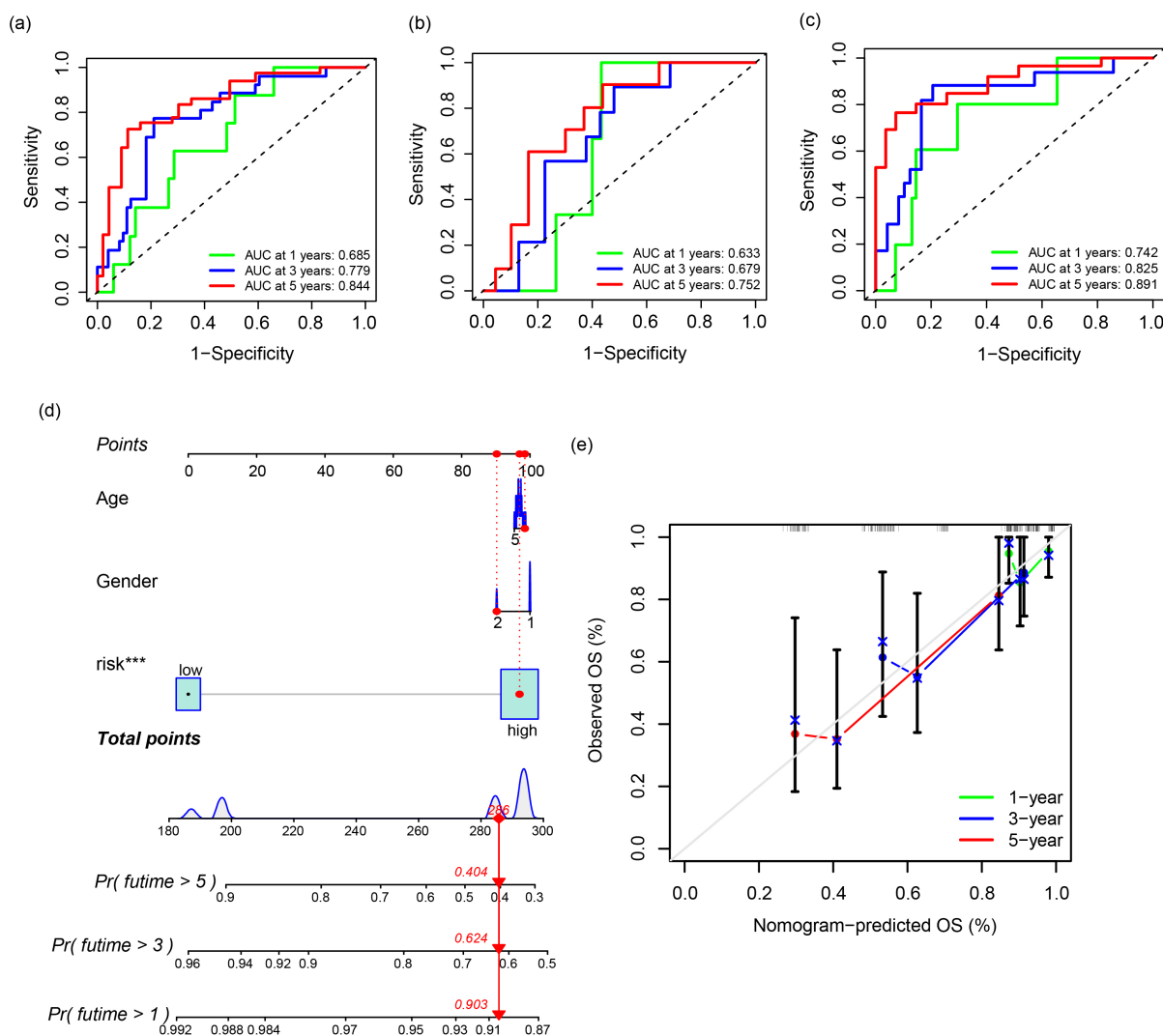
**Table 1.** The six genes and their coefficient

**表 1.** 六种基因及系数

Gene	Coefficient
CARD18	-3.624607368
CENPM	2.565920347
EFHC2	-1.648318634
NR1H4	2.57112032
STEAP3	-1.494970509
ZBTB39	-4.248323426

### 3.2. 构建列线图

风险分数计算公式联合性别、年龄，构建了列线图(图 2(d))。该列线图可用于临床预测病人预后。针对列线图，Calibration 曲线鉴定了它预测的准确度(图 2(e))。该预测模型具有良好的预测效果。



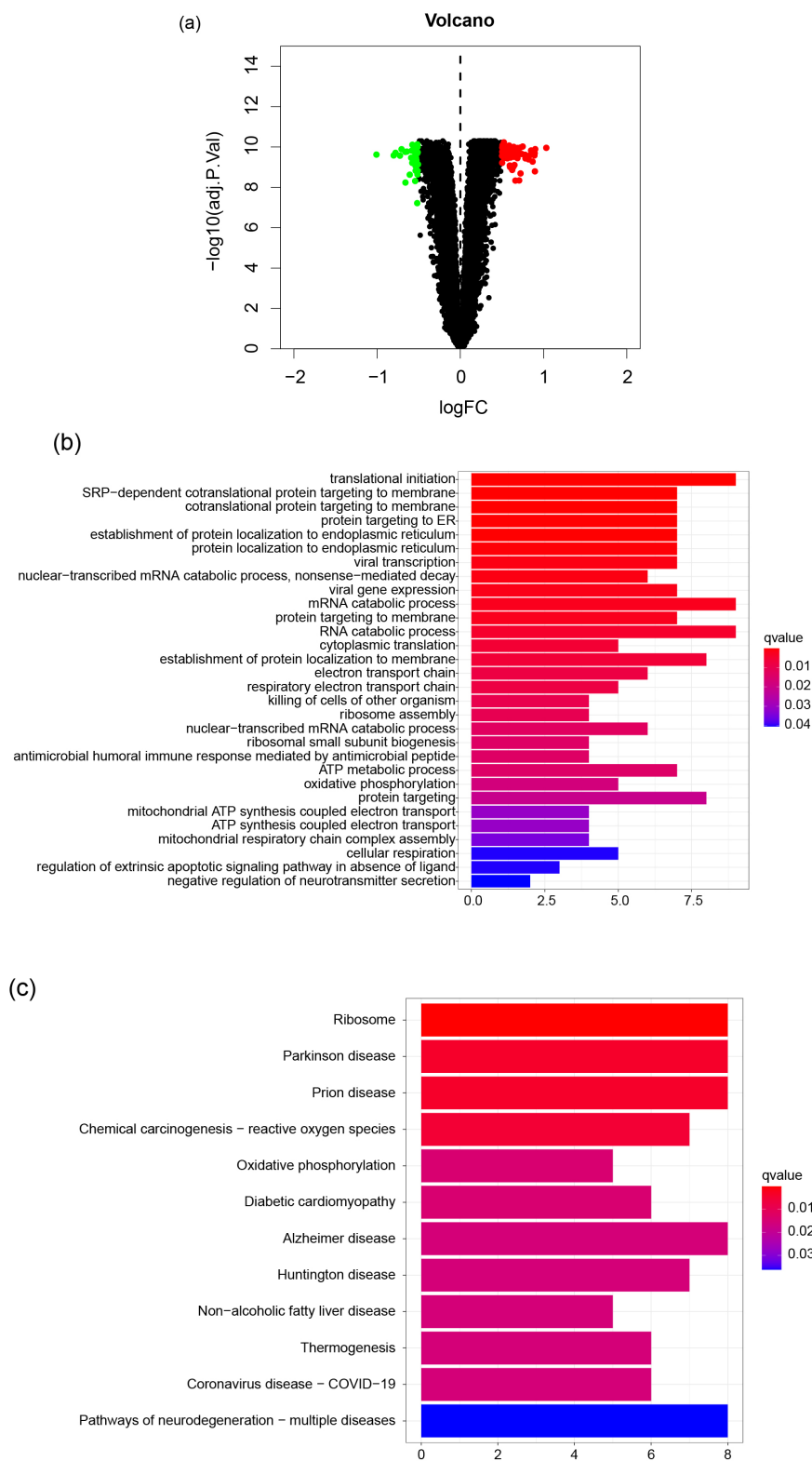
**Figure 2.** (a): The accuracy of risk score predicting all patents' prognosis; (b): The accuracy of risk score predicting train group patents' prognosis; (c): The accuracy of risk score predicting test group patents' prognosis; (d): Nomogram that consisted by risk score, age and gender were constructed. 1 stands for male and 2 stands for female; (e): The accuracy of nomogram predicting patients' prognosis was detected by Calibration

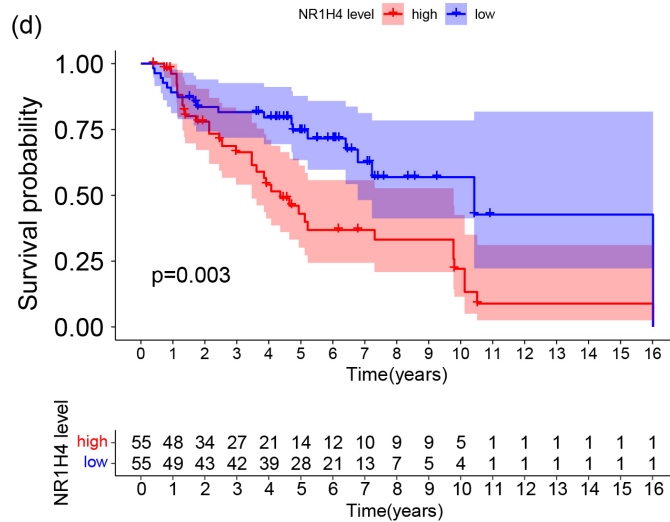
**图 2.** (a): 风险分数在整体病人中，预测预后的准确性；(b): 风险分数在检验集中，预测预后的准确性；(c): 风险分数在训练集中，预测预后的准确性；(d): 风险分数联合年龄和性别，构建列线图。在性别选项中：1 代表男性，2 代表女性；(e): Calibration 曲线检验列线图预测预后的准确性

### 3.3. 差异表达基因的鉴定和富集分析

*NRIH4* 基因在高风险组中显著富集，根据 *NRIH4* 基因的表达量中位值，病人分为高低表达组。两组病人生存预后显著差异(图 3(d))。两组共筛选了 96 个基因。相对于低表达组，高表达组有 31 个低表达基因和 65 个高表达基因(图 3(a))，我们对这些基因进行 GO 和 KEGG 分析(图 3(b)，图 3(c))。GO 分

析结果显示差异表达基因主要集中在蛋白定位和内质网功能。而 KEGG 富集分析结果是蛋白质溶酶体。GO 和 KEGG 结果共同证明，高低表达组病人在蛋白质代谢方面有显著差异。



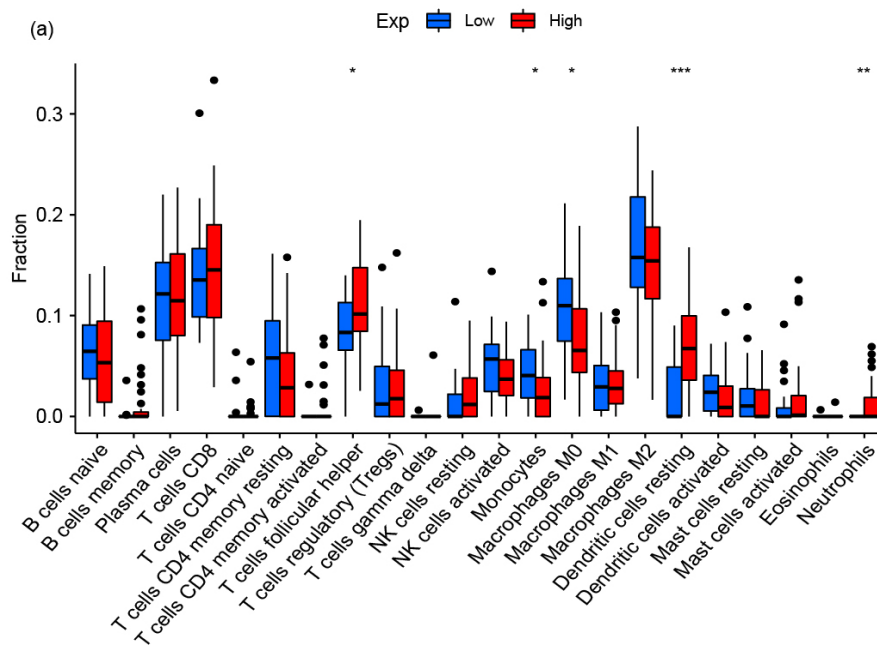


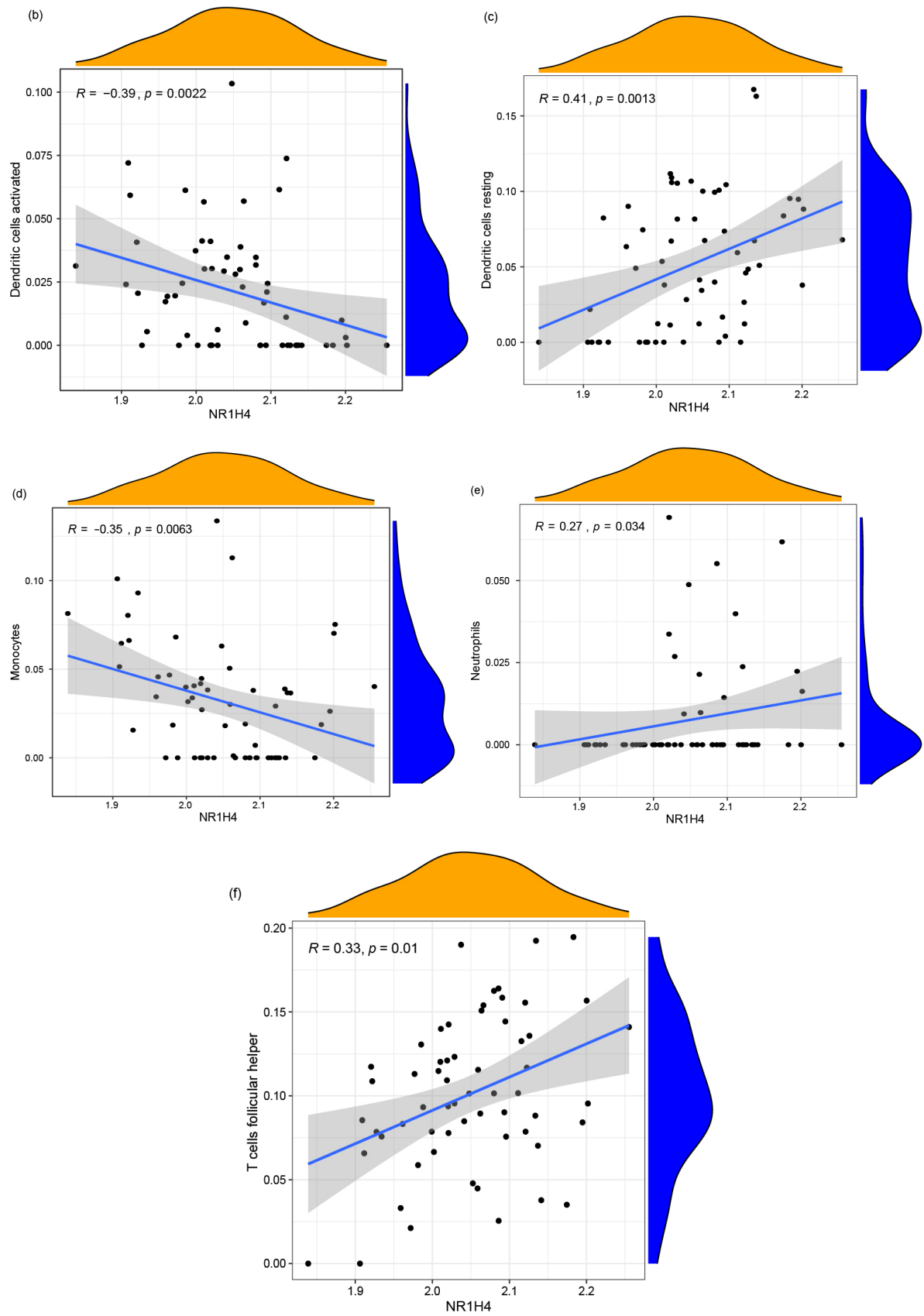
**Figure 3.** (a): The different expression genes between *NR1H4* high expression group and *NR1H4* low expression group; (b): GO enrichment analysis about the different expression genes; (c): KEGG enrichment analysis with different expression genes; (d): The overall survival difference between high/low *NR1H4* expression groups

**图 3.** (a): *NR1H4* 高低表达组之间的差异基因分布; (b): 差异基因在 GO 富集分析中的结果; (c): 差异基因在 KEGG 富集分析中的结果; (d): *NR1H4* 高低表达组之间生存预后差异

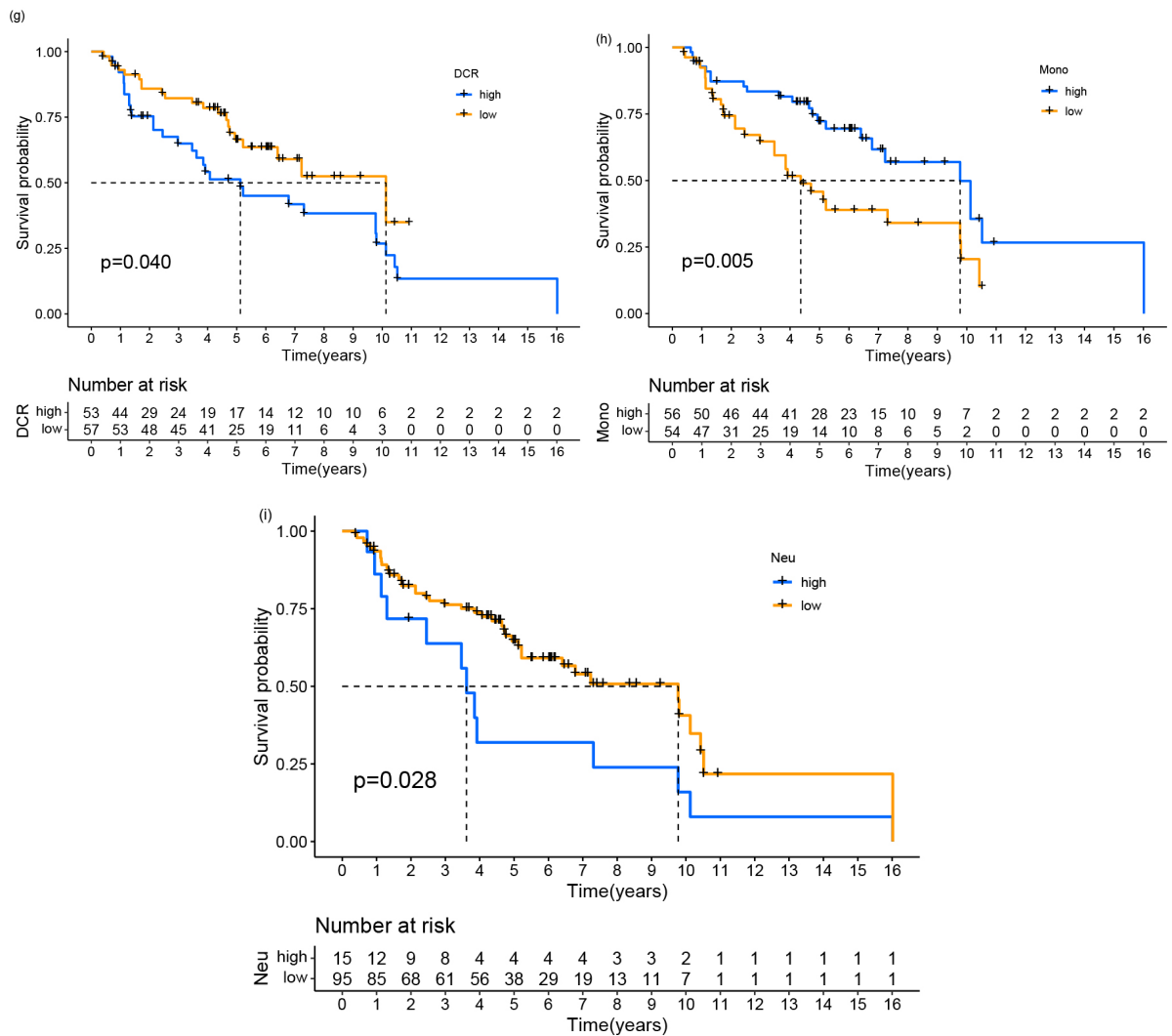
### 3.4. 免疫细胞浸润相关性和差异

高低表达病人在免疫细胞浸润具有显著的不同(图 4(a))。伴随 *NR1H4* 基因表达量上升, 单核细胞和激活态树突细胞浸润丰度逐渐降低, 中性粒细胞、滤泡 T 细胞和休眠态树突细胞浸润丰度显著升高(图 4(b)-(f))。通过 KM 曲线比较了以上五种免疫细胞高低浸润的预后差异, 其中休眠态树突细胞和中性粒细胞高浸润组生存预后较差, 单核细胞高浸润组生存预后较好(图 4(g)-(i))。









**Figure 4.** (a): The infiltration difference of immune cells between high/low *NRIH4* expression groups; (b): The correlation between dendritic cells activated and *NRIH4* expression level is negative; (c): The correlation between dendritic cells resting and *NRIH4* expression level is positive; (d): The correlation between monocytes and *NRIH4* expression level is negative; (e): The correlation between neutrophils and *NRIH4* expression level is positive; (f): The correlation between T cells follicular helper and *NRIH4* expression level is positive; (g): The survival difference between high/low dendritic cells resting infiltration groups; (h): The patients in high monocyte infiltration groups has better prognosis; (i): The high neutrophil infiltration patients has worse prognosis

**图 4.** (a): *NRIH4* 高低表达组之间, 免疫细胞浸润差异; (b): 激活态树突状细胞分布与 *NRIH4* 表达成负相关; (c): 休眠态树突状细胞浸润与 *NRIH4* 表达成正相关; (d): 单核细胞浸润程度与 *NRIH4* 表达成负相关; (e): 中性粒细胞浸润分布和 *NRIH4* 表达成正相关; (f): 滤泡 T 细胞分布与 *NRIH4* 表达成正相关; (g): 休止态树突状细胞高低浸润组生存差异; (h): 单核细胞高浸润组病人预后较好; (i): 中性粒细胞高浸润组病人预后较差

#### 4. 讨论

尤文肉瘤容易发生早期转移。肉瘤出现血运转移较为常见, 而尤文肉瘤往往发生在具有造血功能的骨骼中或附近, 这导致尤文肉瘤极易发生早期转移[16] [17]。该特点使得尤文肉瘤病人往往没有接受根治手术的机会。即便接受根治切除术后, 也极易容易发生复发和远处转移[18] [19] [20]。尤文肉瘤因为其常常生长在具有造血功能的骨骼中或附近, 这使得放射疗法在治疗尤文肉瘤的治疗中的作用有限[21]。而目前化疗对尤文肉瘤的效果不佳, 病人预后往往十分不理想, 寻找新靶点、开发新的靶向药物在尤文肉瘤

治疗中极为重要。

在本文研究中, 我们通过 Lasso 回归和多重 Cox 回归分析, 构建了风险分数计算公式。该公式通过训练集的筛选和验证集的验证, 具有很好的代表性。通过 ROC 曲线验证也证明了该公式具有良好的预后预测功能。我们又联合临床特征, 如性别、年龄, 构建了列线图, 用于预测临床病人的生存预后。我们在构建风险分数计算公式中, 发现 *NR1H4* 基因在高风险人群中表达显著富集。我们以 *NR1H4* 表达量中位值为标准, 将病人分为高低表达两组。高表达组病人预后显著差于低表达组。这说明 *NR1H4* 高表达是尤文肉瘤差预后的标志物。我们又分析了高低表达组两组病人差异表达基因, 共筛选出 96 个差异表达基因。对这 96 个差异表达基因进行 GO 和 KEGG 附近分析。结果显示, 高低表达组病人之间, 在蛋白定位、蛋白降解和内质网功能方面有显著差异。说明这些功能的差异可能会显著影响尤文肉瘤病人的预后。我们还检测高低表达组病人在免疫细胞浸润方面的差异。我们发现六种细胞的分布和 *NR1H4* 基因的表达量具有相关性。其中伴随 *NR1H4* 基因表达量上升, 单核细胞和激活态树突细胞浸润丰度逐渐降低, 中性粒细胞、滤泡 T 细胞和休眠态树突细胞浸润丰度显著升高。我们又比较了以上 6 个免疫细胞高低浸润组之间生存差异。其中休眠态树突细胞和中性粒细胞高浸润容易出现较差预后, 而单核细胞的高浸润容易出现较好预后。这说明, *NR1H4* 高表达和休眠态树突细胞、中性粒细胞高浸润相关, 与单核细胞浸润程度降低相关。而这些因素相互作用, 使患者预后变差。

## 5. 结论

*NR1H4* 是尤文肉瘤的预后标志物, 它的高表达与单核细胞和激活态树突细胞浸润丰度逐渐降低、中性粒细胞、滤泡 T 细胞和休眠态树突细胞浸润丰度显著升高相关。这些因素相互作用给病人带来较差预后。

## 参考文献

- [1] Ewing J. (1972) Classics in Oncology. Diffuse Endothelioma of Bone. *CA: A Cancer Journal for Clinicians*, **22**, 95-98. <https://doi.org/10.3322/canjclin.22.2.95>
- [2] Grünewald, T.G.P., Cidre-Aranaz, F., Surdez, D., Tomazou, E.M., de Álava, E., Kovar, H., Sorensen, P.H., Delattre, O. and Dirksen, U. (2018) Ewing Sarcoma. *Nature Reviews Disease Primers*, **4**, Article No. 5.
- [3] Gaspar, N., Hawkins, D.S., Dirksen, U., Lewis, I.J., Ferrari, S., Le Deley, M.C., et al. (2015) Ewing Sarcoma: Current Management and Future Approaches through Collaboration. *Journal of Clinical Oncology*, **33**, 3036-3046. <https://doi.org/10.1200/JCO.2014.59.5256>
- [4] Balamuth, N.J and Womer, R.B. (2010) Ewing's Sarcoma. *The Lancet Oncology*, **11**, 184-192. [https://doi.org/10.1016/S1470-2045\(09\)70286-4](https://doi.org/10.1016/S1470-2045(09)70286-4)
- [5] (2018) Ewing Sarcoma. *Nature Reviews Disease Primers*, **4**, Article No. 6. <https://doi.org/10.1038/s41572-018-0007-6>
- [6] 刘杰, 赵田, 李钦传, 陈国涵. 肺癌免疫治疗进展[J/OL]. 中国胸心血管外科临床杂志: 1-9. <http://kns.cnki.net/kcms/detail/51.1492.r.20220126.1110.024.html>, 2022-02-09.
- [7] de Aguiar Vallim, T.Q., Tarling, E.J. and Edwards, P.A. (2013) Pleiotropic Roles of Bile Acids in Metabolism. *Cell Metabolism*, **17**, 657-669. <https://doi.org/10.1016/j.cmet.2013.03.013>
- [8] Lee, Y.J., Lee, E.Y., Choi, B.H., Jang, H., Myung, J.K. and You, H.J. (2020) The Role of Nuclear Receptor Subfamily 1 Group H Member 4 (NR1H4) in Colon Cancer Cell Survival through the Regulation of c-Myc Stability. *Molecules and Cells*, **43**, 459-468.
- [9] Savola, S., Klami, A., Myllykangas, S., Manara, C., Manara, C., Scotlandi, K., et al. (2011) High Expression of Complement Component 5 (C5) at Tumor Site Associates with Superior Survival in Ewing's Sarcoma Family of Tumour Patients. *International Scholarly Research Notices*, **2011**, Article ID: 168712. <https://doi.org/10.5402/2011/168712>
- [10] Bailey, A.M., Zhan, L., Maru, D., Shureiqi, I., Pickering, C.R., Kiriakova, G., Izzo, J., He, N., Wei, C., Baladandayuthapani, V., Liang, H., Kopetz, S., Powis, G. and Guo, G.L. (2014) FXR Silencing in Human Colon Cancer by DNA Methylation and KRAS Signaling. *American Journal of Physiology-Gastrointestinal and Liver*, **306**, G48-G58. <https://doi.org/10.1152/ajpgi.00234.2013>

- 
- [11] Jiang, L., Zhang, H., Xiao, D., Wei, H. and Chen, Y. (2021) Farnesoid X Receptor (FXR): Structures and Ligands. *Computational and Structural Biotechnology Journal*, **19**, 2148-2159. <https://doi.org/10.1016/j.csbj.2021.04.029>
- [12] Volchenboum, S.L., Andrade, J., Huang, L., Barkauskas, D.A., Krailo, M., Womer, R.B., *et al.* (2015) Gene Expression Profiling of Ewing Sarcoma Tumors Reveals the Prognostic Importance of Tumor-Stromal Interactions: A Report from the Children's Oncology Group. *The Journal of Pathology: Clinical Research*, **1**, 83-94. <https://doi.org/10.1002/cjp2.9>
- [13] Yu, G., Wang, L.-G. Han, Y. and He, Q.-Y. (2012) clusterProfiler: An R Package for Comparing biological Themes among Gene Clusters. *OMICS: A Journal of Integrative Biology*, **16**, 284-287. <https://doi.org/10.1089/omi.2011.0118>
- [14] Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015). *Limma* Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Research*, **43**, e47. <https://doi.org/10.1093/nar/gkv007>
- [15] Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M. and Alizadeh, A.A. (2015) Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nature Method*, **12**, 453-457. <https://doi.org/10.1038/nmeth.3337>
- [16] Riggi, N., Suvà, M.L. and Stamenkovic, I. (2021) Ewing's Sarcoma. *New England Journal of Medicine*, **384**, 154-164. <https://doi.org/10.1056/NEJMra2028910>
- [17] Eaton, B.R., Claude, L., Indelicato, D.J., Vatner, R., Yeh, B., Schwarz, R. and Laack, N. (2021) Ewing Sarcoma. *Pediatric Blood & Cancer*, **68**, e28355. <https://doi.org/10.1002/pbc.28355>
- [18] Zöllner, S.K., Amatruda, J.F., Bauer, S., Collaud, S., de Álava, E., DuBois, S.G., Hards, J., Hartmann, W., Kovar, H., Metzler, M., Shulman, D.S., Streitbürger, A., Timmermann, B., Toretsky, J.A., Uhlenbruch, Y., Vieth, V., Grünwald T.G.P. and Dirksen, U. (2021) Ewing Sarcoma-Diagnosis, Treatment, Clinical Challenges and Future Perspectives. *Journal of Clinical Medicine*, **10**, Article No. 1685. <https://doi.org/10.3390/jcm10081685>
- [19] Pappo, A.S. and Dirksen, U. (2018) Rhabdomyosarcoma, Ewing Sarcoma, and Other Round Cell Sarcomas. *Journal of Clinical Oncology*, **36**, 168-179. <https://doi.org/10.1200/JCO.2017.74.7402>
- [20] Morales, E., Olson, M., Iglesias, F., Dahiya, S., Luetkens, T. and Atanackovic, D. (2020) Role of Immunotherapy in Ewing Sarcoma. *Journal for Immunotherapy of Cancer*, **8**, e000653. <https://doi.org/10.1136/jitc-2020-000653>
- [21] Luo, C., Wei, J. and Han, W. (2016) Spotlight on Chimeric Antigen Receptor Engineered T Cell Research and Clinical Trials in China. *Science China Life Sciences*, **59**, 349-359. <https://doi.org/10.1007/s11427-016-5034-5>