

Biomarker 筛选与验证

Boyun Tang

May 7, 2012

基于表达谱芯片的 Biomarker 筛选

评价准则

- 较好的预后预测 (Good prognostic value / prediction accuracy)
- 稳定的生物学解读 (Functional similarity / Semantic similarity)

基于表达谱芯片的 Biomarker 筛选

评价准则

- 较好的预后预测 (Good prognostic value / prediction accuracy)
- 稳定的生物学解读 (Functional similarity / Semantic similarity)

难点与挑战

- 属于高维拟合问题 $p \gg N$
- Response 值多为截尾数据

The NEW ENGLAND JOURNAL *of* MEDICINE

ESTABLISHED IN 1812

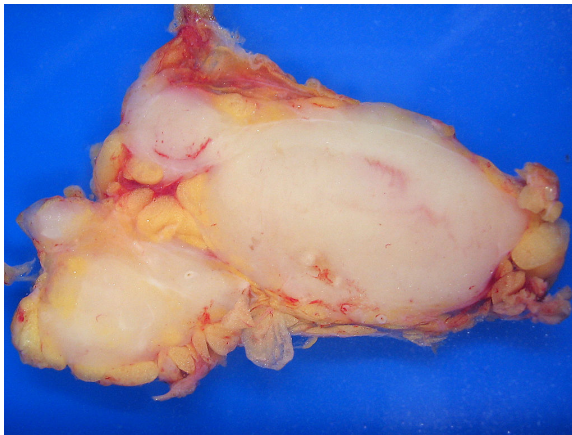
NOVEMBER 18, 2004

VOL. 351 NO. 21

Prediction of Survival in Follicular Lymphoma Based on Molecular Features of Tumor-Infiltrating Immune Cells

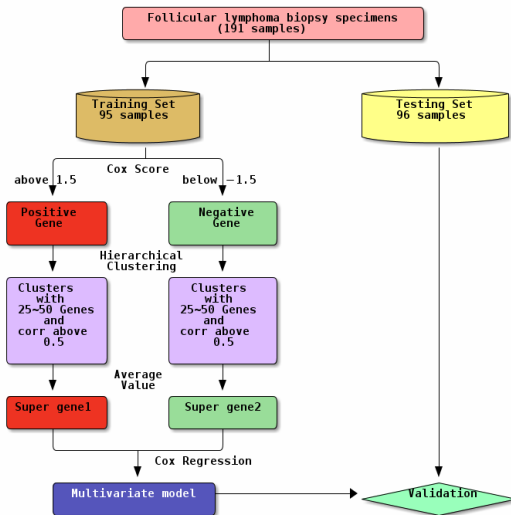
Sandeep S. Dave, M.D., George Wright, Ph.D., Bruce Tan, M.D., Andreas Rosenwald, M.D., Randy D. Gascoyne, M.D., Wing C. Chan, M.D., Richard I. Fisher, M.D., Rita M. Braziel, M.D., Lisa M. Rimsza, M.D., Thomas M. Grogan, M.D., Thomas P. Miller, M.D., Michael LeBlanc, Ph.D., Timothy C. Greiner, M.D., Dennis D. Weisenburger, M.D., James C. Lynch, Ph.D., Julie Vose, M.D., James O. Armitage, M.D., Erlend B. Smeland, M.D., Ph.D., Stein Kvaloy, M.D., Ph.D., Harald Holte, M.D., Ph.D., Jan Delabie, M.D., Ph.D., Joseph M. Connors, M.D., Peter M. Lansdorp, M.D., Ph.D., Qin Ouyang, Ph.D., T. Andrew Lister, M.D., Andrew J. Davies, M.D., Andrew J. Norton, M.D., H. Konrad Muller-Hermelink, M.D., German Ott, M.D., Elias Campo, M.D., Emilio Montserrat, M.D., Wyndham H. Wilson, M.D., Ph.D., Elaine S. Jaffe, M.D., Richard Simon, Ph.D., Liming Yang, Ph.D., John Powell, M.S., Hong Zhao, M.S., Neta Goldschmidt, M.D., Michael Chiorazzi, B.A., and Louis M. Staudt, M.D., Ph.D.

Follicular Lymphoma

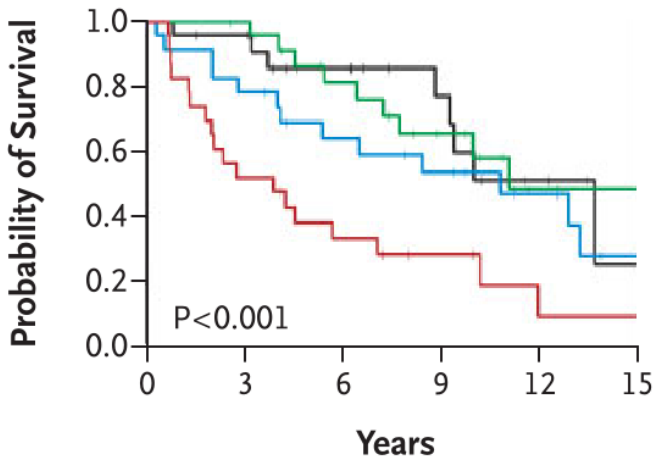


Median survival is around 10 years, but the range is wide, from less than one year, to more than 20 years. Some patients may never need treatment. The overall survival rate at 5 years is 72-77%.

文章思路

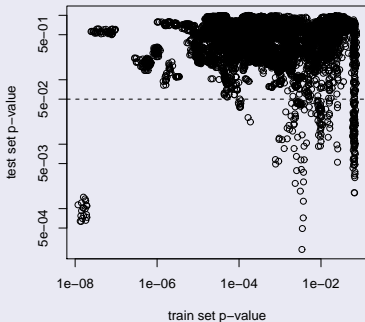


结果



深入分析

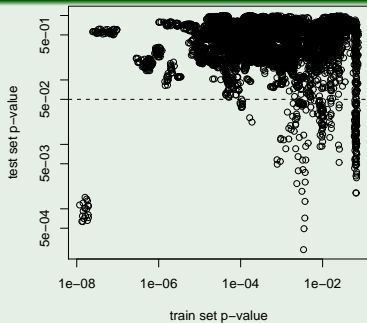
疑点



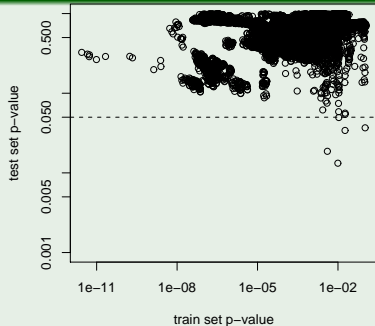
- Total points (Clustering pairs): 11628
- Points on testing set with $p < 0.05$: 239

深入分析

比较



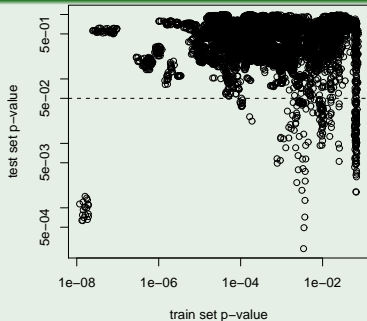
原始数据



互换训练集测试集

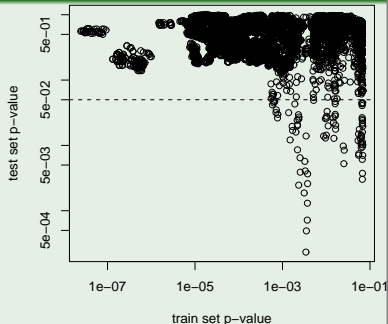
深入分析

比较



原始数据

聚类区间 (25,50)



聚类区间 (30,60)

深入分析

结论

- 看似完美的数据，未必**可靠**，很可能仅仅是由于**过拟合**
- 稳健的分析结果，必须经得起Permutation test的反复验证。

参考链接

- <http://www.nejm.org/doi/full/10.1056/NEJMoa041869>
- <http://www-stat.stanford.edu/~tibs/FL/report/>