论文

基于环境的编码方法在预测HLA-A*0201结合多肽中的应用

赵璞, 李通化

同济大学化学系, 上海 200092

摘要：

T淋巴细胞对抗原的识别是产生与调节有效免疫应答的关键, T细胞只识别主要组织相容性复合物(MHC)呈递上来的抗原, 因此MHC与抗原多肽的结合就成为一系列免疫应答过程中基础的一环. 为了辅助疫苗设计, 多种机器学习技术已被普遍应用于MHC结合多肽的预报领域中. 本文以支持向量机(SVM)为手段, 以HLA-A*0201的实验数据集为对象, 对多种肽段编码方法形成的模型进行评价, 得到的AUC值的范围在0.932~0.936之间. 提出一种新的利用抗原多肽结合环境的编码方法, 使预报的AUC值提高到0.953. 对独立数据集进行建模预报, 同样证明环境编码模型的预报准确率高于传统编码方法的准确率.

关键词： HLA-A*0201; 结合多肽预测; 支持向量机(SVM); 编码; 受试者工作曲线(ROC)

Prediction of HLA-A*0201 Binding Peptides Using Binding-environment-based Peptide Representation

ZHAO Pu, LI Tong-Hua*

Department of Chemistry, Tongji University, Shanghai 200092, China

Abstract:

In all vertebrates, there is a large genomic region or gene family that has a major influence on graft survival referred to as the Major Histocompatibility Complex(MHC). T cells only recognize antigens as a complex with MHC molecules. Therefore MHC binding peptides prediction is an important step in T cells epitopes discovery. To facilitate vaccine design, computational methods have been developed for predicting MHC-binding peptides. A large variety of machine-learning techniques are commonly used in this field. This work explored Support Vector Machine(SVM) as such a method for developing prediction systems of HLA-A*0201 by using experiment dataset. Data representations play a key role in SVM models, so we examined different types of inputs variables for predicting HLA-binding peptides. The AUCs of these SVM models were 0.932—0.936. Then this work proposed a new way to encode peptides, which uses the information of peptides' binding environment, and achieved an impressive AUC of 0.953. The results of independent dataset prediction showed that the overall performance of our novel environmental encoding based SVM models is improved in comparison to other traditional encodings.

参考文献：

[1]Brusic V., Bajic V. B., Petrovsky N.. Methods[J], 2004, 34: 436—443
[2]Hagmann M.. Science[J], 2000, 290(5489): 80—82
[3]Sette A.. Science[J], 2000, 290(5499): 2074—2075
[4]Mallios R. R.. Bioinformatics[J], 2001, 17(10): 942—948

扩展功能

本文信息

Supporting info

PDF(219KB)

[HTML全文]
(${article.html_WenJianDaXiao}KB)

参考文献[PDF]

参考文献

服务与反馈

把本文推荐给朋友

加入我的书架

加入引用管理器

引用本文

Email Alert

文章反馈

浏览反馈信息

本文关键词相关文章

HLA-A*0201; 结合多肽预测; 支持向量机(SVM); 编码; 受试者工作曲线(ROC)

本文作者相关文章

PubMed

[5]Nielsen M., Lundegaard C., Worning P., *et al.*. Bioinformatics[J], 2004, 20(9): 1388—1397

[6]Reche P. A., Glutting J. P., Zhang H., *et al.*. Immunogenetics[J], 2004, 56: 405—419

[7]Doytchinova I. A., Walshe V. A., Jones N. A., *et al.*. J. Immunol.[J], 2004, 172: 7495—7502

[8]Altuvia Y., Sette A., Sidney J., *et al.*. Hum. Immunol.[J], 1997, 58: 1—11

[9]LIN Zhi-Hua(林治华), WU Yu-Zhang(吴玉章), WANG Ying(王瑛), *et al.*. Chem. J. Chinese Universities (高等学校化学学报)[J], 2003, 24(5): 826—830

[10]Bhasin M., Raghava G. P. S.. Vaccine[J], 2004, 22: 3195—3204

[11]Honeyman M. C., Brusic V., Stone N. L., *et al.*. Nat. Biotechnol.[J], 1998, 16: 966—969

[12]Zhao Y. D., Pinilla C., Valmori D., *et al.*. Bioinformatics[J], 2003, 19: 1978—1984

[13]Peters B., Bui H. H., Frankild S., *et al.*. PLoS Comput. Biol.[J], 2006, 2(6): 574—584

[14]CHEN Xiao-Mei(陈晓梅), RAO Han-Bing(饶含兵), HUANG Wen-Li(黄文丽), *et al.*. Chem. J. Chinese Universities(高等学校化学学报)[J], 2007, 28(11): 2171—2178

[15]ZHOU Xiang-Dong(周向东), LI Tong-Hua(李通化), BIAN Fang(边防), *et al.*. Chem. J. Chinese Universities(高等学校化学学报)[J], 2000, 21(2): 216—218

[16]XU Jia-Xi(许家喜). Chem. J. Chinese Universities(高等学校化学学报)[J], 2000, 21(11): 1688—1693

[17]HUANG Jin-Yan(黄金艳). Knowledge Based Study on Bioinformatics(基于知识的生物信息学研究)[D], Shanghai: Tongji University, 2007: 24—25

[18]Kidera A., Konishi Y., Oka M., *et al.*. J. Prot. Chem.[J], 1985, 4(1): 23—55

[19]ZHOU Peng(周鹏), LI Zhi-Liang(李志良), TIAN Fei-Fei(田菲菲), *et al.*. Chin. Sci. Bull.(科学通报)[J], 2006, 11: 1259—1263

[20]Nielsen M., Lundegaard C., Worning P., *et al.*. Protein Sci.[J], 2003, 12: 1007—1017

[21]Chelvanayagam G.. Immunogenetics[J], 1996, 45: 15—26

[22]Zhang G. L., Bozic I., Kwoh C. K., *et al.*. J. Immunol. Methods[J], 2007, 320: 143—154

[23]Kawashima S., Pokarowski P., Pokarowska M., *et al.*. Nucleic Acids Res.[J], 2008, 36: D202—D205

[24]Kenn M., Qian B., Goldstein R. A.. Prot. Struct. Func. Genet.[J], 2000, 41: 498—503

[25]Swets J. A.. Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers[M], Mahwah: Lawrence Erlbaum Associates, 1996: 417—428

[26]Fawcett T.. Pat. Recog. Lett.[J], 2006, 27: 861—874

[27]Swets J. A.. Science[J], 1988, 240: 1285—1293

本刊中的类似文章

文章评论

序号	时间	反馈人	邮箱	标题
				UGG Sale
				UGGS Disc
				Cheap U
				boots onl
				sale ugg b
				sale che