



生命科学与信息科学仿佛量子纠缠的恋人，缘起 20 世纪中叶。DNA 结构阐明、分子生物学兴起；系统论、控制论、信息论纷纷问世并引起广泛关注；计算机走进人类社会并产生重要影响。风起云涌，因缘际会，计算生物学、理论生物学等悄然问世，生命被视为一个信息系统。进入 20 世纪 80 年代，随着人类基因组计划的酝酿与实施，生物信息学从计算与理论的窠臼中脱颖而出，不断融合软件工程与信息技术的最新进展，日新月异。由此带来数据驱动的生物医学研究范式，与假说驱动的传统研究范式相辅相成，极大推进与改变着生物医学各学科领域。近年来，从阿尔法狗 (AlphaGo) 到阿尔法狐 (AlphaFold)，人工智能不断震惊科学界、产业界与普通公众。最新版本的阿尔法狐几乎实现了序列到结构的从头预测，这一任务背后的科学问题被视为生物信息学的圣杯。迄今，任务接近完成，但结构密码仍未阐明，提示从工程到科学，我们需要更多领域更多人的智慧。

从 DNA、RNA 到蛋白质，从基因组、转录组、蛋白质组到表观组，从分子、细胞、组织、器官到系统，从单个物种到生态系统，从序列、结构到网络，从数据、模型到理论，从生物、医学到人工智能，生物信息学海纳百川。这是一个武林。数学家、物理学家、化学家、生物学家、医学家、计算机与信息科学家群贤毕至；谷歌、微软、IBM、百度、华为、字节跳动，少长咸集。中国生物信息学研究的前辈们如郝柏林、李衍达、张春霆、罗辽复、陈润生、罗静初、孙之荣等老师，生动体现了生物信息学的交叉学科属性。在他们的示范引领与培养带动之下，各学科众多青年学子纷纷登上生物信息学舞台。近年来，生物信息学的发展在我国已成燎原之势。哈尔滨医科大学等一批高校开设了生物信息学本科专业，多种生物信息学专业期刊或专辑问世，各省生物信息学一级学会先后成立，包括 2019 年底成立的四川省生物信息学学会。2020 年 9 月，习近平主席在科学家座谈会上的讲话中指出：“坚持面向世界科技前沿、面向经济主战场、面向国家重大需求、面向人民生命健康，不断向科学技术广度和深度进军”。生物信息学充分体现了四个面向。

作为一所以电子信息为特色的理工科大学，学科拓展、服务国家重大战略需求是电子科技大学重要的发展战略。因此，《电子科技大学学报》与四川省生物信息学学会合作推出“生物信息”专栏，希望充分展现生物信息基础理论和转化应用上的研究进展和创新成果，提供一个由一线学者主导的开放的学术交流平台。希望通过我们的专栏，来自更多领域的更多青年学子能够投身生物信息研究，并把研究成果写在祖国的大地上，更好地服务“健康中国”这一国家重大战略需求。

专栏编委会主任：沈育荣 教授



沈百荣，教授、博导，四川大学华西医院系统遗传研究院执行院长。中国生物信息学会(筹)副理事长，四川省生物信息学会理事长、四川省医学信息学专委会主任委员。主要从事生物医学信息和智能健康管理，研究方向包括生物标志物发现的理论、生物医学数据共享与安全、智能慢病管理等。2008年回国后主持重大项目和基金委项目等10多项、培养研究生60多人，发表跨学科研究论文150多篇，出版转化信息学系列英文著作4部、中文著作1部。



黄健，教授、博导，教育部新世纪优秀人才，电子科技大学信息生物学研究中心主任、电子科技大学生命科学与技术学院副院长。四川省生物信息学会副理事长，中国运筹学会计算系统生物学会理事，中国细胞生物学会功能基因组信息学与系统生物学会理事，《International Journal of Biological Sciences》《Interdisciplinary Sciences: Computational Life Sciences》等多家国际期刊编委，成都市科技顾问团生物医药领域特聘专家。主要从事生物信息学及转化医学的基础与应用研究，已发表SCI论文80余篇，总引用超过2200次，H指数27。



陈伟，教授、博导，科睿唯安全球高被引学者，爱思唯尔中国高被引学者。《Molecular Therapy-Nucleic Acids》期刊副主编，中国计算机学会生物信息学专委会委员，中国人工智能学会生物信息学专委会委员，四川省生物信息学会理事。主要从事生物信息学和生物医学数据挖掘研究，在《Nucleic Acids Research》《Molecular Therapy》等期刊中发表了SCI收录论文140余篇，多篇论文入选ESI高被引论文和中国百篇最具影响国际学术论文。



林昊，教授、博导，科睿唯安全球高被引学者，爱思唯尔中国高被引学者。主要从事生物医学信息学领域的研究，在SCI期刊上发表论文160余篇，SCI他引超过8000次，超过40篇论文入选ESI高被引论文，构建生物信息在线服务52个，担任《Current Bioinformatics》等多个SCI期刊编委。



邹权，教授、博导。科睿唯安全球高被引学者，爱思唯尔中国高被引学者。主要研究方向为生物信息学、机器学习和字符串算法。目前担任SCI期刊《Current Bioinformatics》主编，《IEEE Access》《Frontiers in Genetics》《Frontiers in Plant Science》副编辑和5个SCI期刊的编委，多次在《Bioinformatics》《PLOS Computational Biology》《RNA》等知名学术期刊上发表文章。

评“CISI2.0: 更好的抗体交叉或自身相互作用预测工具”

沈百荣

侠之大者，为国为民。武侠小说中，大侠往往有奇遇。段誉吞了莽牯朱蛤、郭靖饮下梁子翁的药蛇血，两人于是百毒不侵。凡此种种，类似主动免疫(接种疫苗)或被动免疫(血清疗法)的原始版本。动物免疫血清用于传染病治疗已有上百年的历史，德国科学家冯·贝林因此获得首枚诺贝尔生理学及医学奖。当前，康复患者血浆治疗新冠肺炎依然有效。这些疗法其实质是多克隆抗体药物。从1986年美国食品药品监督管理局批准第1个单克隆抗体药物迄今，全球上市的单抗药已有上100个，广泛用于肿瘤、自身免疫性疾病、传染病等各类疾病的治疗。近十年来，全球最畅销的十大药物中，抗体药物每每占据半壁甚至更多江山。阿达木单抗常年蝉联最畅销药物之首，年度销售额近200亿美元。因此，开发单抗药物，为国为民，不仅缓解人类病痛，还能推动社会经济发展。

虽然开发单抗药物的成功率高于传统化学药物，但仍是万里挑一。近年来，抗体可开发性预测已成为生物信息学研究的一个热点。早期预测抗体可开发性可减少研发成本，加快研发进程，降低市场价格，利国利民。该领域的经典研究是所谓的可开发性指数(DI)计算，相应方法已整合到价格不菲的DS软件平台中。然而，贵不见得好。该方法基于同源建模得到的抗体结构，运算速度慢，结果也常不靠谱。这篇论文基于抗体序列，绕过结构，直接预测抗体的交叉或自身相互作用，而这些互作会直接影响抗体的可开发性。相应预测模型速度飞快，可处理大规模抗体数据，开发的网络程序CISI2.0自由免费。但这还远远不够。

最近，百度旗下的生物信息公司百图生科与清华大学智能产业研究院正在合作举办2021全球抗体亲和力预测大赛，任务是根据抗原和抗体的氨基酸序列预测抗体-抗原结合的亲和力。广义上讲，亲和力是开发抗体药的前提，因此，预测可开发性首先要预测抗体-抗原结合的亲和力。我们期待包括亲和力预测在内的更多更好的抗体可开发性预测方法与工具的问世，让生物信息的学子们能为国为民做出自己的贡献！

评“利用计算生物学方法识别原核启动子的研究进展”

邹 权

当前，生命科学正处于第三次革命中，关注于生物信息存储、传输与表达，是理解“生命是什么”的最佳契机。生物医学也处于一个百花开放、百家争鸣的时代，“生物信息学”“系统生物学”“合成生物学”等多个学科快速发展，“Hi-C”“单细胞”“相分离”“焦亡”等是当前分子生物学领域最火的名词。

分子生物学的研究大都绕不开“基因的表达”。调控基因表达最关键的调控元件就是启动子，它负责和RNA聚合酶的特异性结合，从而转录出RNA。因此，开发能够准确识别启动子的模型，对于研究基因的表达意义重大。该文章对2005年以来开发的39个用于原核启动子识别的计算工作进行了总结和展望。论文对已发表的用于原核启动子识别的基准数据集进行了搜集和整理，整理得到的资料为将来如何构建合理的、具有代表性的训练和测试数据提供了参考；进而，对广泛使用的DNA序列描述符进行了介绍，这些DNA序列特征提取方法不但可以用原核启动子的预测，也可以用于其他DNA调控元件的表征；再次，是关于序列特征的优化算法应用，包括基于信息论的mRMR和基于概率论的F-Score等算法，如何将这些算法的优点结合，是提高启动子识别精度的一个途径；最后是关于机器学习方法的应用。

目前，尽管已经对原核启动子预测的研究取得了较满意的结果，但这些模型仅限于大肠杆菌等少数几个模式生物。将这些方法用于更多的物种上，构建合适的模型，也是未来的发展方向。希望该文章能为更多学者就此问题研究时，提供新思路、新角度。