研究基于深度学习算法优化序列特异性的C-to-G单碱基编辑器

作者：writer 来源：中国科学院

本文原地址：https://www.iikx.com/news/progress/15172.html

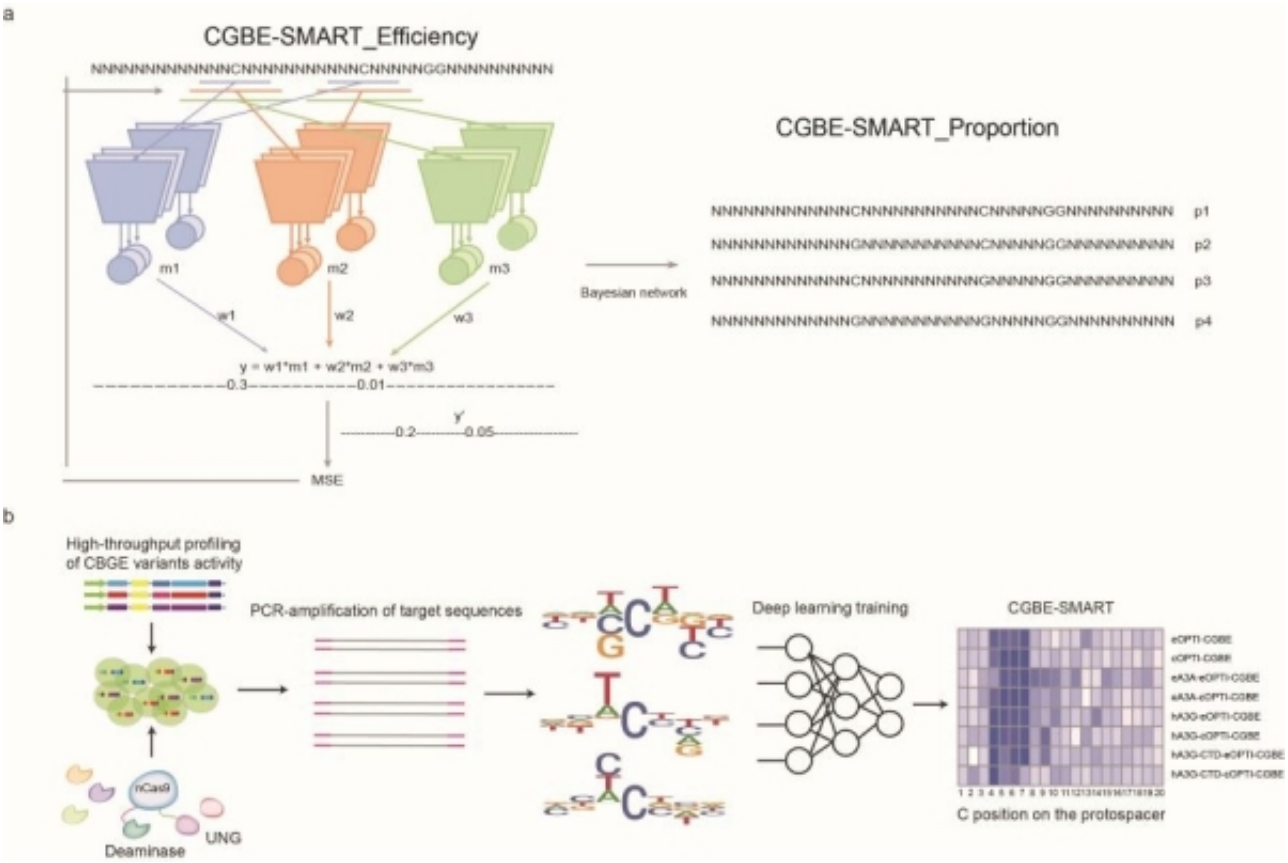
本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

8月12日，*Nature Communications*发表了题为*Optimization of C-to-G base editors with sequence context preference predictable by machine learning methods*的研究论文，该研究由中国科学院脑科学与智能技术卓越创新中心、上海脑科学与类脑研究中心研究员孙怡迪研究组与中国农业科学院深圳农业基因组研究所研究员左二伟研究组合作完成。该研究建立了深度学习模型的算法CGBE-SMART，该方法能够准确预测新型OPTI-CGBEs的单碱基编辑效率和编辑效果。

单碱基编辑技术是基于CRISPR/Cas系统改造发展的新型基因编辑技术，可在不引入DNA双链断裂的情况下，精确地将DNA或RNA中的一个碱基替换为另一个碱基。目前，已开发并得到广泛应用的碱基编辑器包括胞嘧啶碱基编辑器（cytosine base editor，CBE）及腺嘌呤碱基编辑器（adenine base editor，ABE），但这两种碱基编辑器只能实现将C · G碱基对替换为T · A碱基对（C→T），或将A · T替换为G · C（A→G）。因此，CBE或ABE只能修复由C>T或者A>G导致的遗传表型或疾病，而对于其他类型的单碱基突变却束手无策。2020年，科研人员在CBE的基础上，研发了能够将胞嘧啶转换为鸟嘌呤的碱基编辑器（C-to-G base editor，CGBE）。而关于CGBE编辑器的研究仍处于初步阶段，对于其特异性、保真性及编辑特点仍需要进一步研究。David Liu实验室与其合作者对CGBEs系统进行改造与升级，构建了高效的CGBEs编辑器。为能够方便科研人员的日常研究，人工智能与基因编辑结合的愈发紧密，David Liu及Hyongbum Henry Kim等实验室分别建立了可预测单碱基编辑器编辑效果的BE-Hive及DeepBE等深度学习模型。

孙怡迪研究组与左二伟研究组通过筛选不同物种来源的UNGs、密码子优化和全基因组及转录组范围测序，获得了可进行高效C到G碱基颠换以及高保真的OPTI-CGBEs。为了方便其他研究人员选择合适的C-to-G碱基编辑器以及高效预估编辑效率，科研人员建立了预测不同C-to-G碱基编辑器编辑效果的深度学习模型CGBE-SMART（如图）。CGBE-SMART结合了神经网络及概率图模型，为每一个编辑位置独立训练一套参数来预测该位置上的编辑效率。模型使用了大小不同的卷积核，建立一组基础单元网络对编辑位置周围的碱基进行特征提取和效率预测。研究人员将不同基础单元网络的预测结果用一套习得的参数进行加权平均。模型以编辑位点附近的40bp作为输入，通过神经网络预测出guide RNA结合位置1至20的编辑效率，并进一步利用贝叶斯网络预测不同编辑结果的占比（图a）。研究人员将CGBE-SMART用在不同的CGBE编辑器的8个文库数据集上进行实验。在所有的8个数据集上，BE-SMART具有较高的预测准确性（图b）。CGBE-SMART能够准确预测C-to-G编辑效率，且与之前的预测模型相比，在预测C-to-T编辑效果中有更出色的表现。

研究人员全面优化了CGBE碱基编辑器，获得了高编辑效率与低脱靶的OPTI-CGBEs；通过机器学习开发了CGBE-SMART深度学习模型用于预测OPTI-CGBEs编辑结果。该研究将进一步加速CGBE的应用研究。研究工作得到国家自然科学基金委员会、中国农业科学院、深圳市的资助。



CGBE深度学习模型方法构建思路。a、CGBE-SMART深度学习模型示意图；b、CGBE-SMART预测模型的设计

研究团队单位：脑科学与智能技术卓越创新中心

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发