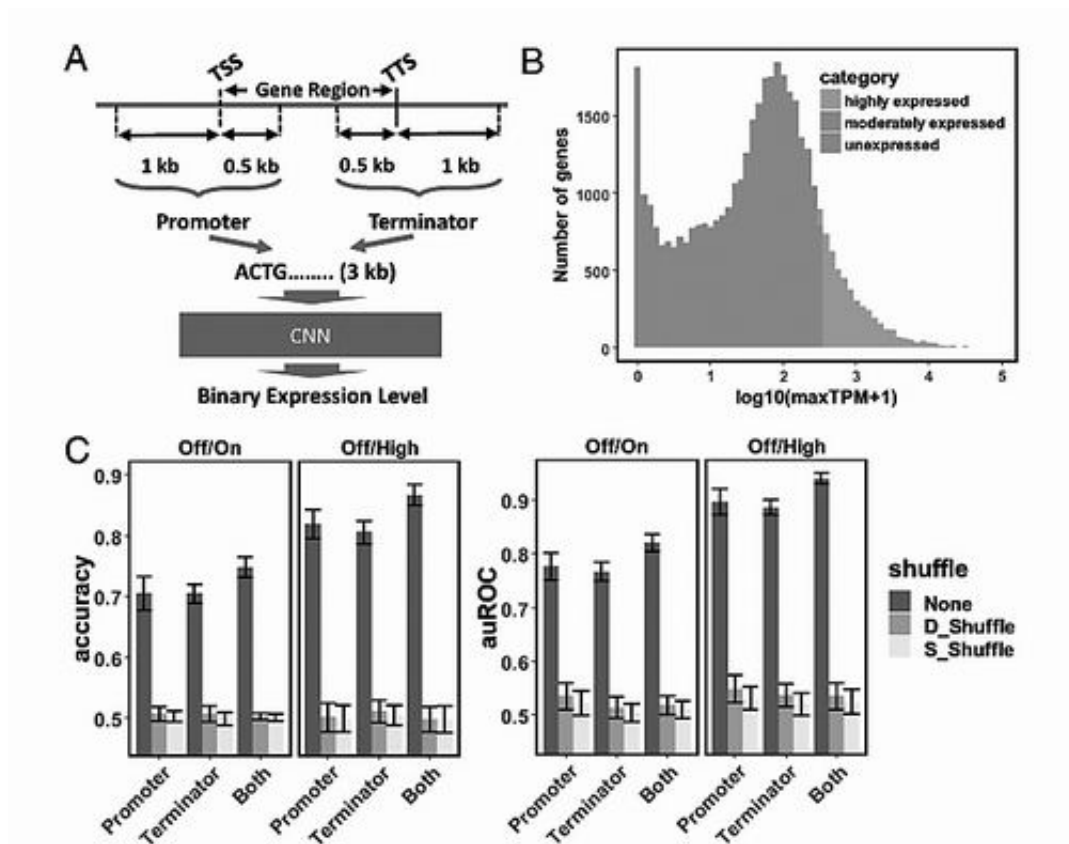


人工智能让育种“物美价廉”

作者：韩扬眉 来源：中国科学报

本文原地址：<https://www.iikx.com/news/progress/4342.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！



预测二分化基因表达量的卷积神经网络模型建立

人工智能让育种“物美价廉”。自从作物被驯化以来，培育集抗性强、优质、高产等性状为一体的作物品种一直是育种家的梦想。DNA分子结构模型的发现推动了分子生物学的发展，让育种家们能够从基因和分子水平上解码作物的生命秘密，通过调控基因获得特定表型，以期培育出最想要的作物品种。

然而，如何调控作物基因才能培育优良品种？如何不用大规模田间试验就能预测基因变异后的作物生长状况？时至今日，这些问题依然困扰着育种学家们。

近日，中国农业科学院生物技术研究所副研究员汪海与合作者共同开发出从基因组DNA序列预测基因表达调控模式的人工神经网络模型，有望借助人工智能(AI)技术实现定向育种。相关成果

发表在《美国科学院院刊》上。

从经验到精准定向

育种，从某种意义上来说，是把来自不同种质资源的优良等位基因聚合起来。

作物育种经历了漫长的改良之路。传统育种是耕作者对作物表型变异的肉眼观察，通过主观判断选出高产优质抗性强的育种材料。后来，职业育种家出现，他们根据对作物遗传规律的认识，通过预先设计杂交育种试验，再从后代中筛选出优良栽培品种。

这些方法曾为作物改良、有效解决粮食安全问题作出了巨大的贡献。但在某种程度上，却都是基于经验和观察，完全根据表型对育种材料进行选育的经验育种。科学家曾无奈而又形象地将其形容为一把尺子一杆秤，用牙咬，用眼瞪。

作物表型易受环境、气候等因素影响，依赖于经验育种效率低，且成本高、田间管理难度大。过去几十年甚至上百年来，基本是沿用这种方式，并无大的突破。华南农业大学生命科学学院教授王海洋告诉《中国科学报》。

直到20世纪50年代，分子生物学与基因工程的诞生，打开了人类认识生命本质的大门。作物育种从经验育种时代进入了分子定向育种时代。这个时期，育种家可在明确基因型的表型效应的情况下，有的放矢地把符合预期要求的基因型进行组合。

找到控制作物最佳性状的基因，对其进行标记，在后代中监测追踪，从而有目的地对单一目标性状进行基因改良，大大提高了育种效率和精确度。王海洋说。

然而，伴随着高通量基因组测序技术的发展，越来越多的作物全基因组密码被解开。在海量的基因组数据面前，控制优良性状的基因是哪些？怎样的基因组组合才能产出最优的作物品种？上述分子标记有效利用与定向育种的先决条件，人们却不得而知。

汪海表示，明确哪些分子标记和哪些性状相关联，需要借助机器学习模型或深度学习模型帮助育种家根据基因型预测表型。人工智能技术突破了人的经验，使作物育种更加精准而高效。

深度学习模型帮助预测优势品种

机器学习是借助计算机算法建立模型并解析数据，通过不断学习数据的自身特征并训练模型，从而实现对目标对象的判断和预测。

汪海告诉《中国科学报》，传统的基于线性模型的机器学习方法由于不考虑生物学过程背后的分子机制，造成模型不会举一反三，在某个基因上学习到的特征不能运用到相似分子机制的基因，而且不能有效预测低频、罕见变异的表型效应。以玉米为例，玉米自然群体中就有超过50%的变异属于低频、罕见变异。

以基因组序列为预测变量的深度学习模型可以克服这一难点。

研究人员以基因家族代替单个基因为单位随机分配训练集和测试集数据，以解决进化依赖造成的模型过拟合问题。接着进一步利用多种算法对模型进行解析，获得了调控基因表达的关键DNA

基序。在此模型基础上，研究人员利用进化上亲缘关系较近的两个物种，成功预测了同源基因的相对表达量，并进一步获得了调控同源基因相对表达量的关键DNA基序。

汪海表示，深度学习模型通过模拟分子生物学过程，可在自然群体中预测直接造成表型的因果变异，而非和因果变异紧密连锁的变异。未来可以针对因果变异进行基因组编辑，直接将有利自然变异引入现有的育种材料。

此外，与传统高投入、大规模的田间试验相比，人工神经网络模型可在计算机中对基因组DNA序列进行虚拟诱变，并利用模型预测变异的后果。从而再挑选符合预期目标的变异序列进行实验验证，实现低成本定点定向设计育种。汪海说。

智能化育种4.0时代

这是作物优良基因挖掘方法的突破，也代表了未来的发展方向。中国农业大学农学与生物技术学院植物遗传育种学系教授、国家玉米改良中心主任李建生告诉《中国科学报》。

以人工智能为代表的新一代人工智能技术具有更强大的数据挖掘能力，正推动作物育种走向智能化的4.0时代。

中国农业大学作物基因组与生物信息学系教授王向峰撰文以玉米为例，对育种4.0时代进行了详细的阐释：依托人工智能、基因组测序、基因编辑等相关技术，实现玉米组学基因型与表型大数据的快速积累，通过遗传变异等数据的整合，实现作物性状调控基因的快速挖掘与表型的精准预测，通过人工改造基因元器件与人工合成基因回路，使作物具备新的抗逆、高效等生物学性状，并通过在全基因组层面上建立机器学习预测模型，创建智能组合优良等位基因的自然变异、人工变异、数量性状位点的育种设计方案，最终实现智能、高效、定向培育新品种。

在人工智能技术辅助育种方面，美国农业公司已有应用。比如原孟山都公司，通过人工智能筛选，只需对最具开发潜力的品种分子进行田间测试，即可帮助农民增收。此外，借助机器学习和预测建模技术，快速为农民提供数字化解决方案。

中国要实现应用还有一段路程要走。李建生表示，与国外农业公司种业集中度高、规模大相比，中国种业公司多为作坊式生产且分布分散，要实现高通量的基因筛选与预测，需要改良适合中国种业发展的模型和方法。

在研究方面，汪海坦承，目前，把深度学习等人工智能技术应用于基因组学领域在国内外都刚刚起步。

在他看来，阻碍人工智能技术在基因组学中广泛应用的因素之一是跨领域人才缺乏。基因组学领域的人需要学习和掌握人工智能技术方法，并根据基因组学领域问题的特殊性，对人工智能技术进行改造。

除此之外，训练深度学习模型需要大量的数据。然而在农业领域，作物的基因型和表型数据量却积累不足。

王海洋建议，研究人员在育种后，除了留下优质品种数据，也要保存非理想型品种的全套基因组和表型数据，以便数据建模时进行优劣比较，找出调控优良表型性状的基因。

大数据时代下智能化育种的前提是标准化大数据体系。而农业数据采之不易且不统一，王海洋表示，作物表型数据差异性较大，不同人采集的数据真实可靠性与准确性也难以控制。除此之外，彼此数据不开放共享，使得研究中可比较的数据量少。有数据是第一步。对数据进行规范化采集处理、存储与管理，并建立开放共享的数据库更重要。

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](http://www.iikx.com)转发