
科学家提出DNA信息存储阴阳双编码新方法

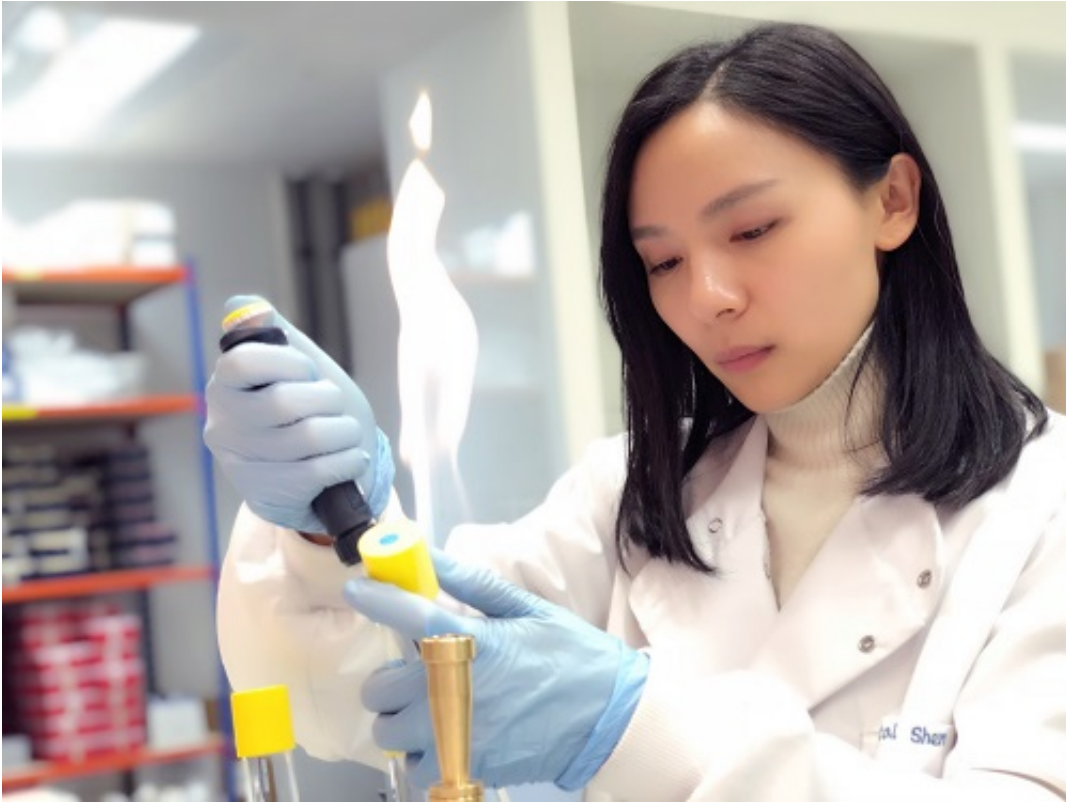
作者：writer 来源：爱科学

本文原地址：<https://www.iikx.com/news/progress/18159.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！

科学家提出DNA信息存储阴阳双编码新方法。近日，深圳华大生命科学研究院研究员沈玥团队与合作者在《自然计算科学》上发表了封面文章，为DNA信息存储的应用提供了一种高密度、高稳定性的比特-碱基编解码方法，并完成了体内外两种模式的信息存储实验验证。

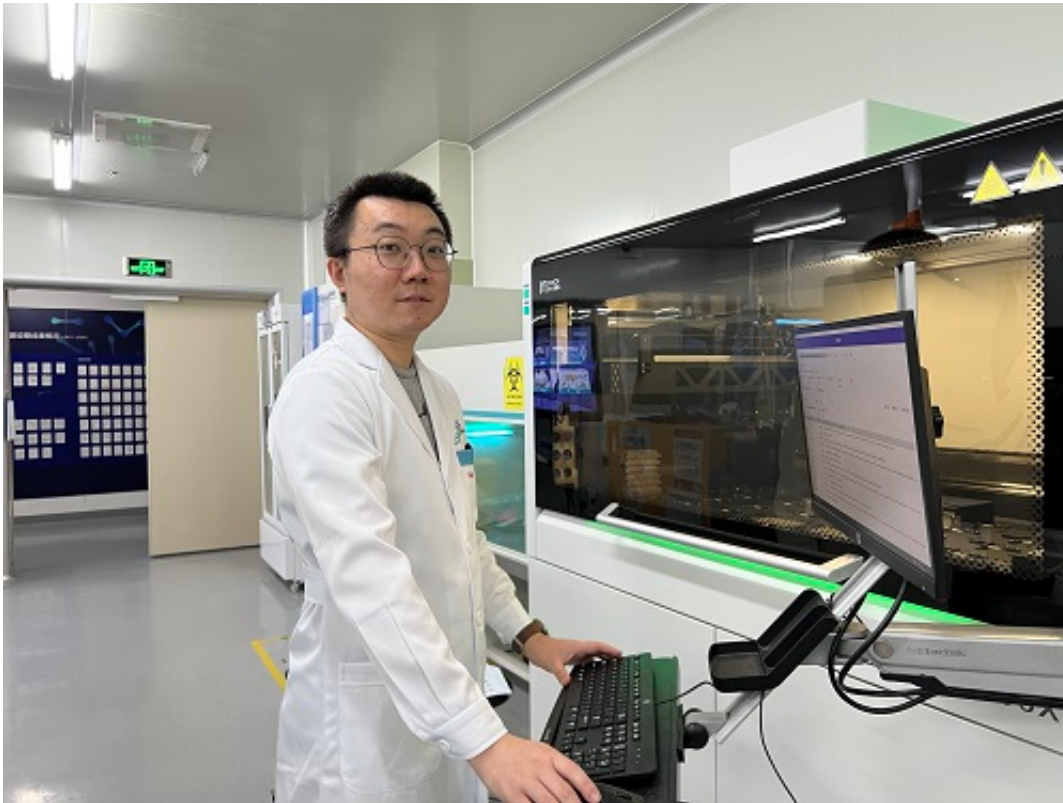
DNA存储作为生物技术与信息技术相融合的主要方向之一，近年来受到广泛关注。DNA存储的编解码，即比特与碱基之间的转换，是DNA存储中最重要的一环之一，不仅决定信息转换的效率（信息密度），还直接影响存储信息的稳定性及可靠恢复性。从2012年起，编解码技术的发展主要聚焦于提升信息密度，而技术兼容性和原始信息的稳定恢复方面的考虑尚不全面。2017年以前，编解码技术都未能实现完全的技术兼容，产生序列的GC含量很大程度上还是依赖于原始数据的0/1分布情况。2017年，美国哥伦比亚大学研究团队开发的DNA喷泉码几乎解决了这一问题，但直接套用的信道编码技术有较强的数据类型偏好性，因此在实际的存储应用中存在较高的数据无法恢复风险的问题。



沈玥在实验室 研究团队供图

该论文的通讯作者沈玥告诉《中国科学报》，为解决这一问题，他们将从DNA双链模型中受到的启发，与中华文化中阴阳对立统一的思想相结合，巧妙地应用于DNA编解码系统，以两套不同的规则，分别对两条二进制信息进行一对一编译转换，再取两者统一交集的部分为最终解，实现将两条独立的信息组合统一为一串DNA序列；另一方面，通过引入筛选机制，他们将与现有合成测序技术兼容性不佳的序列通过预先设置的筛选条件进行过滤。根据不同的组合方法，该系统共能提供1536种不同的编码规则组合，大大扩展了其应用场景范围。

研究人员还通过编码学的理论推导以及不同数据类型文件的模拟编码，证明了该系统在保证信息密度的前提下，在数据恢复稳定性方面体现显著的性能提升(存储数据的平均恢复率较DNA喷泉码现有水平提升近两个数量级)。



平质在实验室 研究团队供图

该论文的共同第一作者、深圳华大生命科学研究院助理研究员平质告诉记者，近年来用细胞进行DNA信息存储也受到了极大关注，为此，他们还测试了该系统在酵母细胞内存储、传代后的数据恢复稳定性。结果证明，作为载体的酵母菌株经过1000代以上的传代，信息仍可以被完整恢复，该存储方式接近天然DNA分子存储物理信息密度的理论极限，每克DNA能存储的信息量约为432.2EB。

该研究开发了一种全新的DNA存储编码方法，并提出1536种不同编码规则组合的方案，为DNA存储的多类型应用提供了重要工具，有望在海量数据长期存储的新型介质研究中起到积极的推动作用。（来源：中国科学报田瑞颖）

相关论文信息：

<https://doi.org/10.1038/s43588-022-00231-2>

<https://doi.org/10.1038/s43588-022-00235-y>

版权声明：凡本网注明来源：中国科学报、科学网、科学新闻杂志的所有作品，网站转载，请在正文上方注明来源和作者，且不得对内容作实质性改动；微信公众号、头条号等新媒体平台，转载请联系授权。邮箱：shouquan@stimes.cn。

作者：沈玥等 来源：《自然—计算科学》

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