中国科学: 生命科学 2025 年 第 55 卷 第 10 期: 1968 ~ 1978
SCIENTIA SINICA Vitae lifecn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS
评 述 中国知名大学及研究院所专栏 天津大学合成生物学20周年专辑

哺乳动物染色体构建及应用

钟源, 刘培源, 朱婧怡, 付宗恒, 周龙飞, 李霞, 刘悦, 元英进*, 赵广荣*

合成生物技术全国重点实验室, 教育部合成生物学前沿科学中心, 天津大学合成生物与生物制造学院, 天津 300072
* 联系人, E-mail: yjyuan@tju.edu.cn; grzhao@tju.edu.cn

收稿日期: 2025-07-02; 接受日期: 2025-09-08; 网络版发表日期: 2025-09-25
国家重点研发计划(批准号: 2024YFA0916600)和天津市自然科学基金(批准号: 24JCQNJC01250)资助

摘要 设计构建染色体是合成生物学的底层技术. 过去二十年基因组设计合成研究实现从微生物基因组到哺乳动物染色体序列的设计构建技术跃迁. 从染色体结构元件和功能序列到染色体疾病、人源化免疫系统等模型的独特应用, 哺乳动物染色体的设计构建为解析和重构高等生物基因组提供全新的技术范式. 本文围绕设计-构建-应用的工程化路线, 综述哺乳动物染色体相关设计原则, 组装方法与实际应用进展, 讨论了哺乳动物染色体设计构建的挑战并展望未来的发展方向.

关键词 合成基因组学, 染色体设计, 染色体构建, 人工染色体, 哺乳动物细胞

合成基因组学借助对DNA大片段进行人工设计与构建的工程化手段, 达成对基因组的理性改造乃至从头合成的科学目标, 从而揭示遗传信息和生物表型之间的内在联系, 实现对特定生命系统的构建和解析[1]. 设计和构建哺乳动物染色体, 应用于基础生物学研究, 必将深化人们对高等生物复杂基因组的认知并提升其功能重塑的能力[2]; 同时应用于研发抗体、基因与细胞治疗药物、构建复杂疾病模型, 为应对生命健康领域的重大难题提供全新的解决思路(图1).

1 哺乳动物染色体设计

基因组设计的目标, 是在系统性解析天然染色体结构与功能的基础上, 通过精简、扩展、重构、泛基因组化、柔性化等手段, 实现对人工染色体模块结构与功能的重新设计[3~7]. 这一过程不仅旨在检验和扩充人们对天然染色体的理解, 更致力于创造自然界中不存在的染色体, 从而在基因组水平上赋予生命体全新的结构与功能. 哺乳动物基因组庞大, 全长达到Gb级, 单个基因长度达到百kb甚至Mb级别[8], 基因间调控复杂, 受多层级染色质空间结构以及远端调控网络影响[9], 维持染色体存在的着丝粒及端粒等结构元件是重复嵌套的复杂序列, 使其人工设计在长度和复杂程度上面临着区别于其他生物基因组的独特挑战[10~12], 当前哺乳动物基因组设计主要有基于天然序列的有模板设计和基于人造序列的无模板设计两种策略.

1.1 有模板设计

有模板的基因组设计, 是以天然染色体的序列和结构为参考, 通过有针对性和理性的修改(如删除、插

引用格式: 钟源, 刘培源, 朱婧怡, 等. 哺乳动物染色体构建及应用. 中国科学: 生命科学, 2025, 55: 1968–1978
Zhong Y, Liu P Y, Zhu J Y, et al. Construction and application of mammalian chromosomes (in Chinese). Sci Sin Vitae, 2025, 55: 1968–1978, doi: 10.1360/SSV-2025-0211

© 2025《中国科学》杂志社

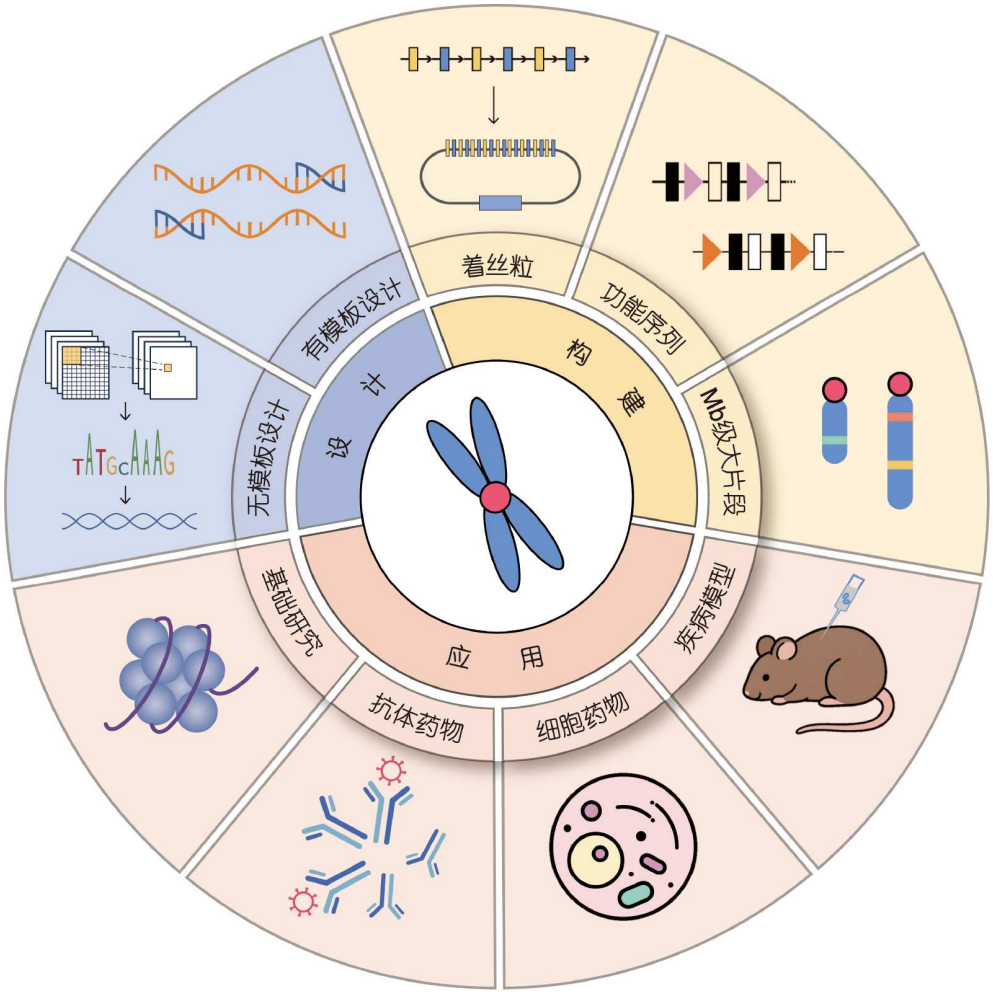

图 1　哺乳动物染色体设计构建及应用
Figure 1　Design, construction, and application of mammalian chromosomes

入、替换或重排)获得新染色体[13]. 这些修改通常起始于基因水平, 通过基因内部序列与方向的调整、基因间重排元件插入和基因顺序的调整使其有别于纯天然序列[14,15]. 但哺乳动细胞基因组的大尺度、复杂性使基因注释、功能理解不完全, 难以进行高度精简有模板设计哺乳动物染色体[4,16]. 经过几十年的发展, 哺乳动物染色体的有模板设计形成三种主要技术路线. 第一种技术路线是“自上而下”的染色体保守删减设计, 也是最早的设计模式, 它最大程度地保留染色体级别的天然序列以维持其功能[17], 如人21号染色体HAC除保留天然着丝粒和端粒外, 还人工增加嘌呤霉素抗性基因、潮霉素抗性基因、组胺醇脱氢酶基因、LoxP位点等序列[18]. 第二种技术路线是“自下而上”的染色体重复单元扩增设计, 重点关注着丝粒结构元件的功能与机制解析[19], 如在着丝粒α-卫星序列单体中插入微生物的操纵基因序列, 扩增构建出在动物细胞中发挥功能的人造着丝粒[20]. 第三种技术路线是人源化替换与功能模块移植的染色体设计, 将人源大片段DNA序列或特定基因簇精准替换动物染色体片段或整合到基因组上, 在保留受体生物染色体结构的同时实现功能模块的“人源化”改造[21], 如人源化*HPRT1*小鼠. 这些有模板设计在维持染色体功能、解析结构元件作用及实现功能模块精准移植等方面具有一定的优势, 但随着人工智能的发展和生物学大数据的爆炸式增长, 无模板设计将成为未来哺乳动物染色体的工程化改造与功能研究的重要手段.

1.2 无模板设计

人工智能正在为哺乳动物染色体的无模板设计注入新动力. 通过学习来自基因组、转录组、蛋白质组和代谢组等多层次、大规模实验数据, 这些无模板设计模型能预测调控序列功能和三维基因组结构, 从而减少对天然序列片段裁剪和拼接的依赖, 从头设计出满足特定功能需求的调控元件, 甚至完整的染色体序列.

自2015年以来, 基于深度学习的多层卷积神经网络(convolutional neural network, CNN)和循环神经网络(recurrent neural network, RNN)开始应用于功能元件DNA序列的设计[22,23]. CNN侧重提取DNA序列的局部特征, 如转录因子结合位点, RNN侧重捕捉序列内和序列间依赖关系, 在启动子和增强子设计方面取得重要进展. 2023年, 本实验室设计构建、表征合成型B细胞启动子库强度, 鉴定启动子活性的关键基序组合及其位置偏好, 开发了基于CNN的启动子设计模型[24]. DeepSEED运用条件生成对抗网络和长短期记忆网络(一类特殊的RNN)来捕捉序列中的时序依赖关系, 在哺乳动物细胞中设计出多西环素诱导表达的人工启动子, 进行表达活性测试与迭代评价, 获得活性显著优于天然启动子的合成启动子序列[25]. 通过在GC含量校准的随机序列库中定向植入人肢端黑色素瘤MM001细胞系关键转录因子结合位点(如SOX10, MITF, TFAP2), 经预训练CNN预测打分并结合双萤光素酶报告筛选高分序列, 通过反复迭代优化, 设计出MM001细胞特异性人工增强子, 其启动活性与天然增强子相当[26]. 类似地, 2025年, 通过深度学习设计增强子, 获得在人HepG2和K562细胞系中高度特异性的人工增强子序列, 且最短仅50 bp的微型增强子也能维持活性与细胞特异性[27].

CNN和RNN模型一次只能捕捉几十到几百碱基范围内的局部序列特征, 在预测元件活性和设计启动子、增强子等短片段序列时表现良好. 相比之下, 大语言模型(large language models, LLM)在模型结构和算法上发展非常迅速, 能够处理Mb级DNA序列, 并整合多种数据类型. 近年来, 先后报道多种染色体设计的LLM, 从kb级别序列分析工具, 逐步发展为支持哺乳动物基因组多尺度研究的Mb级分析与输出平台, 在长序列的功能预测和注释、人工增强子序列和染色体级别片段生成方面取得显著进展.

2021年, DNABERT模型采用BERT-base架构, 首次将基于Transformer的预训练模型引入基因组功能注释领域, 约有110 M参数[28]. 通过全基因组预训练, 并对转录因子结合位点识别、染色质可及性预测和组蛋白修饰定位等多种任务进行再训练, 显著提升DNABERT模型对应功能注释的准确度. 2023年, HybriDNA融合Transformer-Attention与Mamba2 SSM架构, 能在单碱基分辨率下处理长达131 kb的DNA序列, 可控生成人细胞特异性的人工增强子序列[29]. 同年, DNAGPT模型首次将GPT架构引入DNA序列预训练, 理论上能设计输出10 kb的DNA序列[30]. 2024年, Evo模型首次结合注意力与数据驱动卷积的分层状态空间架构(StripedHyena)进行基因组序列预训练, 应用于原核生物基因组设计[31]. 2025年, Evo 2模型采用多尺度StripedHyena 2架构, 参数量升级到40 B, 其训练数据包括动物基因组序列, 设计的理论窗口扩展到1 Mb染色体片段, 并可生成16 kb人类线粒体基因组序列[32].

AI设计染色体序列的处理能力不断增强, 输入窗口和输出序列长度快速增长. 针对AI设计生成的DNA序列, 急需进行真实场景的生物学实验, 通过比较无模板设计序列、有模板设计序列及野生型序列的生物学功能, 以评估这些设计序列的实际可用性; 并反馈给大语言模型, 从而不断优化无模板AI设计能力. 通过建立AI设计序列与生物学功能之间的模型关系, 最终实现无模板的哺乳动物染色体功能化设计, 这是染色体设计的未来方向.

2 哺乳动物染色体的构建

哺乳动物染色体是长度达数十至数百Mb的线性DNA分子, 每条染色体含有一个由Mb级重复序列组成的功能性着丝粒, 负责维持染色体的稳定分离, 染色体末端则由TTAGGG重复序列构成的端粒保护[33~35]. 此外, 染色体上广泛分布转座元件、基因簇与顺式调控序列, 构成多层级的调控网络, 支撑其复杂的功能执行[36]. 目前哺乳动物染色体构建的研究进展主要集中在三个方面: 着丝粒的构建、功能序列的构建以及Mb级染色体大片段的人工合成组装(图2).

2.1 哺乳动物染色体着丝粒的构建

着丝粒是稳定染色体遗传和正确复制、分配的关键结构元件，哺乳动物天然着丝粒序列特点是高度重复的α卫星(alphoid)DNA，含有着丝粒蛋白B (centromere protein B，CENP-B)特异性识别结合基序(称为CENP-B盒). 基于天然着丝粒序列的特点，构建含有*tetO*序列(大肠杆菌四环素操纵基因)的α卫星DNA单体(简称α*tetO*)阵列的着丝粒[37~40]. 类似地，大肠杆菌乳糖操纵基因*lacO*序列也已经用于构建*lacO* α卫星DNA单体(简称α*lacO*)阵列的着丝粒[19]. 这些几十kb的阵列，在细胞内不稳定，常常自发多聚化形成着丝粒，由此构建长达数Mb的tetO-HAC或lacO-HAC，难以实际应用.

随着对着丝粒组装机制的深入理解，后续构建策略逐渐摆脱对高重复序列α卫星DNA的依赖. 如人类4号染色体的4q21区域，与天然着丝粒DNA序列不同，没有重复串联的α卫星DNA序列，也没有CENP-B盒. 2019年，将*lacO*阵列与4q21 BAC组合，构建lacO-4q21-HAC. 在细胞内过表达LacI-HJURP融合蛋白，通过LacI与*lacO*的特异性结合作用，将HJURP定位至*lacO*序列上，HJURP再募集CENP-A，组装CENP-A核小体，形成着丝粒. 但该着丝粒仍然不稳定，在细胞内多聚化，最终形成5~10 Mb的HAC[41]. 2024年，研究发现，在lacO-4q21-HAC中装载600 kb支原体基因组时，通过酿酒酵母原生质体融合递送到哺乳动物细胞中，能以单拷贝稳定存在，是哺乳动物染色体着丝粒构建的里程碑事件，为染色体大片段递送、功能研究等提供有力工具[42].

2.2 哺乳动物染色体功能序列的构建

在哺乳动物染色体功能序列自下而上的组装构建过程中，几kb到几十kb片段主要通过聚合酶循环组装、Gibson组装、Golden Gate组装等体外组装方式，这些方法为后续构建更大尺度DNA提供基础模块. 如采用酶切和Gibson组装相结合，将53个人类免疫球蛋白重链可变区基因表达单元(每个约3.5 kb)，逐级组装成100 kb左右的压缩型人类免疫球蛋白重链可变区基因座[43]. 然而，随着目标片段长度的增加，高分子量DNA在体外组装过程中易受剪切力损伤而发生断裂，操作困难，体内能组装到百kb级以上DNA片段，更具优势[1]. 酿酒酵母因其高效的同源重组能力、强大的大片段DNA承载能力以及成熟的遗传操作技术体系，已经成为哺乳动物染色体大片段DNA组装构建的首选宿主[44~48].

转化相关的重组(transformation‐associated recombination, TAR)是一种最早开发的酵母体内DNA片段同源重组组装技术[49,50]. 2019年，利用TAR技术将34个平均长度3.8 kb的DNA片段，组装成119 kb的人类*HTRA1/ARMS2*[51]. 2022年，针对鼠*HoxA*基因簇，将长度为5 kb的28个DNA片段组装成134 kb的基础*HoxA*构建体，随后引入增强子模块，构建总长约170 kb的*HoxA*扩展版本[52]. 2024年，利用CRISPR-Cas9技术在酿酒酵母中对构建的*Sox2*基因进行精确的片段删除、倒位或插入等二次编辑操作，构建多种长度不同的*Sox2*变体. 这些构建体可以通过酵母质粒回收后转入大肠杆菌扩增，形成高纯度、大量的重组DNA，为后续在小鼠胚胎干细胞中进行的靶向整合与功能研究提供可靠的DNA大片段[53].

TAR技术可直接从复杂的哺乳动物基因组中捕获特定的染色体DNA片段，也可用于合成序列的多片段迭代组装. 为此，将在合成酵母基因组计划(Sc2.0)中使用的SwAP-In (Switching Auxotrophies Progressively by Integration)技术[13]拓展应用到动物染色体功能序列的构建中，开发eSwAP-In (extrachromosomal Switching Auxotrophies Progressively by Integration)技术[21]. 与SwAP-In技术不同，eSwAP-In的同源重组与迭代整合多个DNA片段、置换酵母营养缺陷标记发生在质粒上. 2021年，将38个带重叠同源臂的3 kb片段，通过三轮eSwAP-In，在酵母中组装出长度为101 kb的人*HPRT1*基因. 2024年，通过eSwAP-In技术，由27个4 kb和1个2 kb的合成片段在酵母中迭代组装100.7 kb的反向*HPRT1*序列，用于后续小鼠胚胎干细胞中的功能研究[54].

为了在哺乳动物细胞内应用SwAP-In技术进行组装，开发mSwAP-In (mammalian switching antibiotic resistance markers progressively for integration)技术[55]. 在小鼠胚胎干细胞中，交替使用两种可互换的标记盒(由不同Cas切割位点序列和筛选标记基因组成)，通过CRISPR-Cas9介导的同源重组，逐轮整合DNA片段，实现115 kb合成型*Trp53*基因座的染色体内构建. 同时，该方法也支持一次性将116或180 kb的人源*ACE2*组装

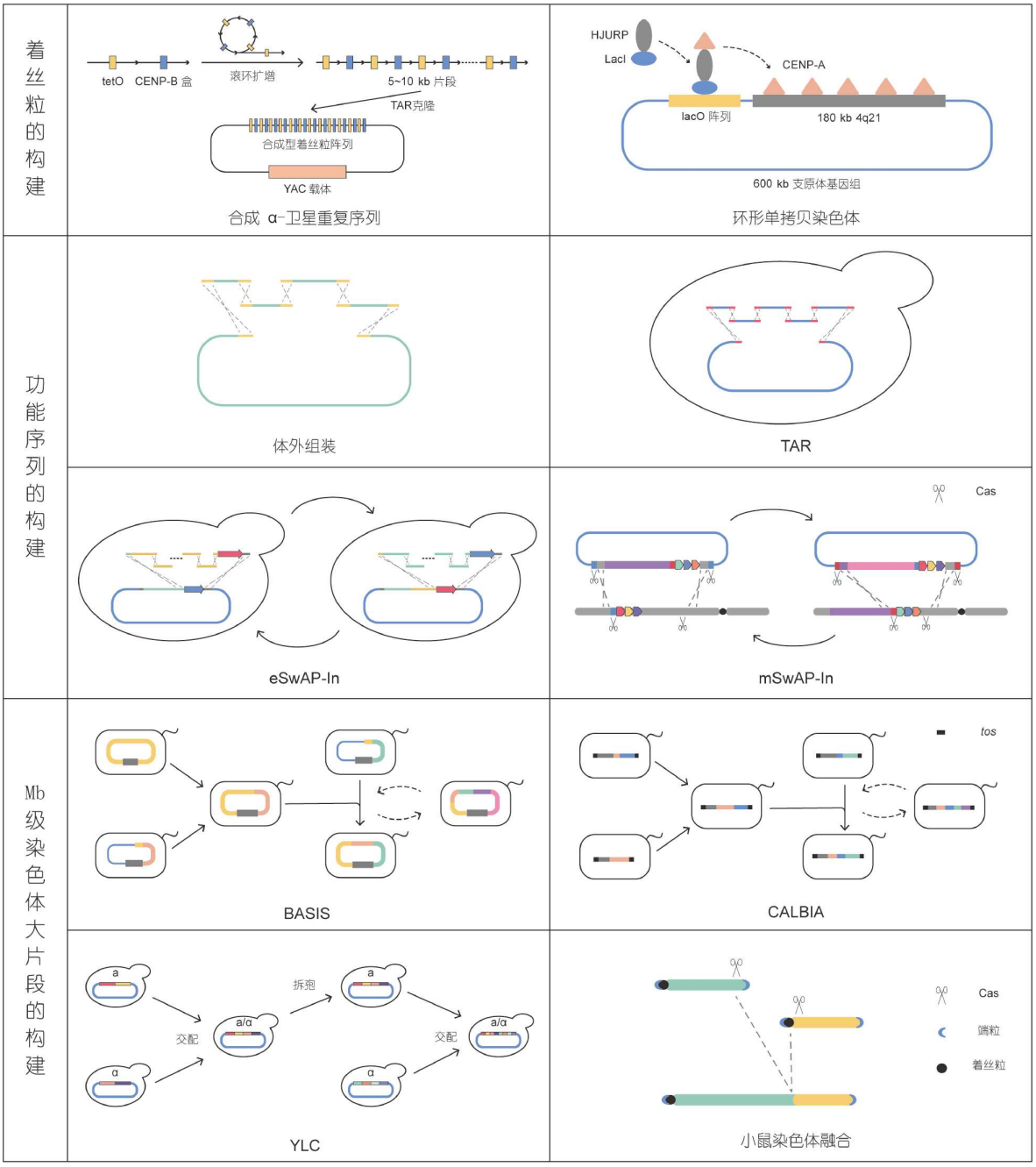

图 2 哺乳动物染色体构建策略
Figure 2 Strategies for mammalian chromosome construction

体定点整合至基因组上. 最后通过CRISPR-Cas9或piggyBAC去除标记盒, 实现染色体功能序列的无痕组装和细胞构建. mSwAP-In在小鼠胚胎干细胞中进行的百kb级染色体片段精准整合构建, 可望扩展到其他细胞类型中.

2.3 哺乳动物Mb级染色体大片段的构建

哺乳动物染色体长度在百Mb级[33,35], 开发Mb级大片段组装技术是染色体构建的关键. 目前已经在大肠杆菌、酿酒酵母和哺乳动物细胞中分别开发人类染色体Mb级DNA组装构建技术.

2023年，英国剑桥大学的Chin团队[56]开发BAC逐步插入合成(BAC Stepwise Insertion Synthesis, BASIS)技术，这是一种大肠杆菌细胞内迭代DNA组装技术. 每一轮组装中，含供体DNA的BAC通过接合转移进入受体菌，利用CRISPR-Cas9辅助切割，在λ-Red系统作用下发生同源重组，实现供体DNA片段与受体DNA片段的组装. 采用此技术，将9个170 kb的DNA片段，迭代组装构建1.1 Mb的人类21号染色体区域片段. 该染色体片段包括多个功能基因、调控元件、G-四链体以及长散在重复序列和短散在重复序列在内的复杂基因组特征.

与BASIS技术不同，接合转移辅助的线性BAC迭代组装(conjugation-associated linear-BAC iterative assembling, CALBIA)技术是以原核线性BAC作为载体，无需CRISPR-Cas9和λ-Red系统辅助，可对大片段DNA的迭代组装[57]. 使用该方法对7个150~200 kb的线性BAC进行组装，获得1.07 Mb的人类免疫球蛋白重链可变区基因簇. CALBIA也能在线性大肠杆菌染色体上组装，经过6轮迭代，将人类18p11.2区域311~845 kb的6个BAC片段组装在大肠杆菌染色体上，构建出2.13 Mb染色体片段.

为提高酵母中超大片段DNA的组装效率，本实验室提出将大片段DNA的组装过程嵌入酿酒酵母生命周期的策略，开发酵母生活史组装技术(Yeast Life Cycle assembly, YLC-组装)[58]. 通过不同型别单倍体酵母细胞的交配，将两个大片段DNA转移到二倍体细胞中. 在CRISPR-Cas9系统的辅助下，两个DNA片段在二倍体细胞内发生高效组装. 经过减数分裂与拆孢，筛选出含组装体的单倍体酵母细胞，即可进入下一轮交配与组装循环. 该方法将大片段DNA的转移、组装一体化，在细胞内实现高效、迭代式的大片段DNA组装. 通过三轮迭代组装，成功将6个200 kb的人类免疫球蛋白重链片段组装成1.26 Mb功能基因座. YLC组装技术突破现有酿酒酵母体内DNA组装的限制，理论上可构建十Mb级超大DNA片段，可用于大规模人工基因组的构建. 相较于TAR, SwAP-In等酵母体内方法，YLC-组装避免体外提取和转化操作对大片段DNA的剪切力损伤，简化实验流程，提高组装效率，降低细胞内组装的技术门槛.

针对现有方法在重复序列组装中易出现错配的问题，本实验室提出组合式层级组装策略，开发SynNICE (nucleus isolation for chromosomes extraction, NICE)技术. 和SwAP-In技术相比，第一轮将5.5 kb的小片段组装成40~71 kb的长片段，使含有末端重复序列的片段数量大大减少，降低后续长片段之间的同源重组频率和错配发生几率；同时增加同源臂的长度(约500 bp)，并利用酵母有性生殖和CRISPR-Cas9技术，提高大片段的正确组装效率. 通过三轮迭代组装，将233个约5.5 kb的DNA序列组装成1.14 Mb、重复序列含量达69.38%的人类Y染色体*AZFa*区域片段，首次实现Mb级人类高度重复序列的从头合成组装. 开发酵母细胞核载体技术，将1.14 Mb的组装体转移到小鼠胚胎细胞中. 该技术实现高重复序列、超大DNA片段的构建和跨物种转移，拓展了哺乳动物Mb级染色体大片段功能重构的技术路径[59].

为开发哺乳动物中染色体级别的重排技术，2022年，在小鼠单倍体胚胎干细胞中，通过CRISPR-Cas9系统引入双链断裂，并利用NHEJ通路将两条非同源染色体融合，构建出308.3 Mb的小鼠4号和5号融合染色体，以及377.6和263.3 Mb两种小鼠1号和2号融合染色体，为Mb级染色体大片段的构建和哺乳动物染色体合成提供新策略[60].

3　哺乳动物染色体设计构建的应用前景

哺乳动物基因组具有高度复杂性，其基因调控网络、顺式作用元件及染色质构象之间的相互作用难以系统解析，常规方法在操控尺度、保真度与调控环境重建方面仍存在局限. 通过合成和精准整合百kb级功能序列，能够实现内源或异位背景下的序列重构与调控解析. 哺乳动物染色体设计构建已广泛应用于生物学基础研究、抗体药物研发、基因细胞药物开发及复杂人类疾病模型的构建，为深入理解哺乳动物基因调控原理与推动精准医疗提供重要技术支撑.

3.1　生物学基础研究应用

哺乳动物染色体设计构建通过合成与操控大片段DNA，能够重构基因调控网络、解析染色质状态，广泛应用于关键基因簇调控机制研究、增强子功能验证及默认表观状态研究，为复杂性状建模与表观遗传研究提供核心技术支撑.

哺乳动物基因组复杂基因簇中多元件的调节作用

难以解析, 如哺乳动物*HoxA*基因簇编码高度保守的转录因子HOXA, 其精确调控依赖于复杂的顺式作用元件、三维染色质结构和信号应答机制, 而各调控元件的相对贡献仍未阐明. 在酵母中设计构建130~170 kb不同大小的*HoxA*基因簇及其变体, 转移整合至小鼠胚胎干细胞的*Hprt1*位点, 构建包含不同组合的转录因子结合位点、增强子及染色质结构元件的细胞模型, 以研究其在视黄酸诱导分化过程中的转录、染色质结构和拓扑域形成[52]. 结果表明, 在*HoxA*基因簇中, 簇内的转录因子结合位点特别是视黄酸反应元件和HOX是响应视黄酸信号激活基因表达并建立染色质边界的核心模块, 远端增强子在激活基因表达的初始特异性中并非必需, 但可与簇内激活子结合协同作用, 从而增强转录活性. *HoxA*基因簇内部的功能性顺式调控元件还能驱动染色质边界形成和组蛋白修饰积累, 构建稳定的拓扑与表观状态, 完成从形态信号到基因表达程序的转换.

增强子作为调控基因表达的顺式作用元件, 其功能已被广泛认可, 但增强子的活性及调控效果高度依赖其基因组上下游的邻近序列、三维染色质结构及相互作用元件等. 这种依赖性导致单一增强子功能难以被独立分析和准确预测, 限制对复杂基因调控网络的系统理解. 小鼠胚胎干细胞中*Sox2*基因是关键发育调控基因, 其表达依赖于远端增强子簇, 该区域由多个DNase I超敏位点(DHSs)构成, 但其单个位点的贡献及相互依赖程度尚不明确, 且传统研究方法难以在内源基因座背景下解析这些复杂调控元件的功能. 设计构建一系列包含定点缺失、倒位、易位及转录因子识别位点精细突变的合成大片段, 并整合到小鼠胚胎干细胞内源*Sox2*基因座获得细胞模型, 分析单个DHS对基因表达的影响[53]. 结果表明, 单个DHS元件的激活活性取决于其所在的序列组合和相对位置关系而非方向, DHS24和DHS26具有自主活性, 而其他元件则完全依赖邻近元件共同作用, 明确核心增强子内部特定转录因子(如雌激素相关受体、SOX等)结合位点对整体调控活性的贡献, 揭示调控元件之间复杂的协同作用网络.

真核基因组中大量非编码区功能尚不明确, 尤其在缺乏显性调控输入的情况下, 染色质的默认转录状态及其物种间的保守性仍未阐明. 采用eSwAP-In技术从头设计构建101 kb大小的反向人类*HPRT1*基因, 删除原有序列中进化形成的编码或调控信息并保留天然序列基本特征. 将该片段分别导入酿酒酵母和小鼠胚胎干细胞, 研究进化上未经选择的大片段序列的默认转录活性[54]. 结果表明, 酵母基因组处于开放、活跃的转录状态, 有利于新生基因的出现和进化. 而小鼠基因组转录状态趋于沉默, 且转录激活需要特定调控元件或精细序列环境, 从而揭示不同真核细胞之间基因组的默认表观遗传状态的重大差异.

3.2 抗体药物研发应用

哺乳动物染色体设计构建可实现对免疫基因座的精确设计, 突破常规基因工程的限制, 可构建具备外源抗体重构、表达能力的模型动物. 针对新冠病毒(SARS-CoV-2)持续进化, 其受体结合域(RBD)突变株导致疫苗预防效果下降, 从多种骆驼科动物中筛选获得30个具有完整开放阅读框的重链抗体的可变区基因, 将每个基因与小鼠免疫球蛋白重链基因座的可变区启动子、前导外显子-内含子序列及下游重组信号序列构建为单一表达单元, 组装成约25 kb的插入盒, 通过CRISPR-Cas9技术定点整合至小鼠免疫球蛋白重链基因座可变区, 构建能表达骆驼来源纳米抗体的小鼠模型, 可用于研发具有中和新冠病毒变异株活性的抗体[61].

3.3 基因细胞药物的开发

通过百kb级基因座的合成与精准操控, 突破传统载体在容量、调控和安全性上的瓶颈, 哺乳动物染色体设计构建为精确模拟人类疾病和开发新型治疗策略提供可能.

奈梅亨断裂综合征(NBS)和冯·希佩尔-林道综合征(VHL)分别由*NBS1*和*VHL*基因突变导致, 均需精确递送全长基因以恢复其生理功能. 然而当前基因治疗载体在容量与调控能力方面不能满足大片段基因的承载和递送. 将*NBS1*(约55 kb)或*VHL*(约25 kb)基因, 定点组装到αtet-HAC上, 导入*NBS1*缺陷细胞和*VHL*缺陷肾癌细胞, 可恢复细胞正常功能. 这为NBS和VHL综合征等单基因病提供潜在治疗策略, 也为干细胞编辑与瞬时表达提供安全可控的新方法[37,40].

P53是关键肿瘤抑制因子, 在小鼠中由肿瘤抑制基因*Trp53*编码, 长达115 kb. 该基因的其突变高度集中于含甲基化CpG二核苷酸的热点位点, 而基因编辑技

术难以实现百kb级高保真序列替换，限制对高风险突变位点的系统性改写及抗突变模型的构建．通过密码子优化将小鼠*Trp53*热点区CpG位点同义突变为不易甲基化的AG，构建抗突变的合成基因*synTrp53*，并采用mSwAP-In技术实现其在内源位点的精准替换．结果表明，*synTrp53*正常激活下游靶基因并维持细胞周期及凋亡响应，热点区C>T/G>A突变率较野生型下降90%以上，这为癌症突变机制解析及预防性基因干预提供新策略，拓展哺乳动物染色体设计构建在细胞治疗中的应用[55].

3.4　疾病模型构建

哺乳动物染色体的设计构建还可以精确设计组装和整合大规模基因组序列，在哺乳动物基因功能验证及疾病模型构建中展现出独特优势.

β-地中海贫血(β-thalassemia)是遗传性红细胞疾病, 主要由β-珠蛋白基因突变或缺失导致. 通过理性设计，可构建α-珠蛋白和β-珠蛋白基因编码区域及调控区域的变体来替换小鼠中的同源基因，以研究它们对血红蛋白产生的影响并揭示地中海贫血的致病机理. 相比于传统研究方法，设计合成的变体可以从新的角度解析致病机理，并为进一步的疾病治疗提供新思路[62].

次黄嘌呤-鸟嘌呤磷酸核糖转移酶(HPRT1)是嘌呤代谢途径中的关键酶，其突变可导致严重的X连锁遗传疾病．由于*HPRT1*基因结构复杂且富含调控元件，基因编辑技术难以实现大规模、高精度的基因插入. 通过基于酵母的eSwAP-In迭代组装体系，可组装包含调控区的101 kb人*HPRT1*基因序列，借助Cre介导将DNA整合至小鼠胚胎干细胞X染色体预设位点．功能验证显示，重组细胞稳定表达人源HPRT1蛋白，且保持多能性及染色体完整性[21].

血管紧张素转化酶2 (ACE2)是新冠病毒刺突S糖蛋白的功能受体，介导新冠病毒的入侵和感染．而小鼠*Ace2*在表达水平和剪接模式上与人类存在显著差异，限制其在新冠病毒感染模型中的应用．通过mSwAP-In方法，用116和180 kb人*ACE2*基因座片段替换小鼠同源位点, 构建人源化*ACE2*小鼠模型. 该模型表现出更接近人类的组织特异性表达模式和剪接异构体生成模式，特别是180 kb *ACE2*人源化模型能够更为精准地模拟人类在新冠病毒感染后肺部病毒载量变化、免疫响应特征及疾病严重程度．进一步地, 在人源化*ACE2*小鼠胚胎干细胞中, 构建更接近人类的双人源化新冠病毒感染小鼠模型, 为新冠病毒致病机制研究和抗体药物开发提供了高度拟人的动物模型[55].

4　展望

哺乳动物染色体的理性设计与合成为解析复杂基因调控与重塑高级生命功能开辟崭新道路．借助有模板路线的自上而下删减与模块化扩增，以及人工智能驱动的元件与染色体无模板序列生成，已从基因元件迈向整染色体雏形的新阶段．基于跨宿主迭代组装、体内同源重组与接合转移等技术原理建立的整合体系, 已能高保真组装与定点整合百 kb~Mb级DNA片段. 由此催生的关键调控网络解析、抗突变基因替换及多位点人源化模型等案例，正在变革基础研究、精准医学与抗体药物开发的模式.

尽管技术突破不断涌现，染色体设计、构建与应用环节仍相互制约，面临系统性挑战．在染色体设计环节，现有AI模型主要基于统计规律，通常仅能捕捉局部序列特征，尚难以准确预测染色质的3D结构和动态表观修饰; 同时, 训练数据多以单一细胞系的短片段为主, 缺乏跨物种、跨发育阶段的多样性信息, 加之缺少功能验证的高通量闭环系统，导致大语言模型设计仍需依赖“合成后试错”．在染色体构建环节，Mb级组装及其转移都存在瓶颈，特别是由于内源的DNA同源重组修复机制，异源重复序列的组装是所有组装宿主的通用问题; 现有的转移技术手段(如电转、显微注射和细胞融合)效率低下, 且有诱发染色体断裂、脱靶整合和免疫激活的致瘤风险，难以支撑临床应用和大规模动物模型的构建.

针对当前染色体设计和构建的瓶颈，需要建立覆盖设计-构建-测试-学习的全流程数据闭环，配合自动化设备，将每轮构建与测试的结果数据实时传回给AI设计模型，以快速迭代，提高设计的准确性和实用性. 针对染色体构建环节的挑战，需要深度开发和优化组装用专属底盘细胞，如减弱内源同源重组修复能力、保护组装序列等，提高细胞内重复序列的组装准确率和组装染色体的长期稳定性，不断拓展组装染色体的长度．针对Mb组装染色体转移难的问题，开发新型生

物材料或无核人工细胞，用于合成染色体的包裹和跨细胞转移，加速临床应用转化. 展望未来，随着基因组和医学数据等生物学数据的积累、设计构建工具的完善和成熟，将促进完整染色体乃至全基因组的设计合成，加速新药研发，服务于生物医药产业和人民健康福祉.

参考文献

1 James J S, Dai J, Chew W L, et al. The design and engineering of synthetic genomes. Nat Rev Genet, 2025, 26: 298–319

2 Venter J C, Glass J I, Hutchison Iii C A, et al. Synthetic chromosomes, genomes, viruses, and cells. Cell, 2022, 185: 2708–2724

3 Hutchison Iii C A, Chuang R Y, Noskov V N, et al. Design and synthesis of a minimal bacterial genome. Science, 2016, 351: aad6253

4 He B, Fu Z H, Wu Y. et al. Research progress of synthetic mammalian genomics. Synbio J, 2022, 3: 78–97 [何博, 付宗恒, 吴毅, 等. 哺乳动物合成基因组学研究进展. 合成生物学, 2022, 3: 78–97]

5 Zhang W, Mitchell L A, Bader J S, et al. Synthetic genomes. Annu Rev Biochem, 2020, 89: 77–101

6 Gerecht K, Freund N, Liu W, et al. The expanded central dogma: genome resynthesis, orthogonal biosystems, synthetic genetics. Annu Rev Biophys, 2023, 52: 413–432

7 Kim K, Choe D, Cho S, et al. Reduction-to-synthesis: the dominant approach to genome-scale synthetic biology. Trends Biotechnol, 2024, 42: 1048–1063

8 Nurk S, Koren S, Rhie A, et al. The complete sequence of a human genome. Science, 2022, 376: 44–53

9 Dixon J R, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature, 2012, 485: 376–380

10 Boeke J D, Church G, Hessel A, et al. The genome project-write. Science, 2016, 353: 126–127

11 Mao Y, Zhao Y, Zhou Q, et al. Chromosome engineering: technologies, applications, and challenges. Annu Rev anim Biosci, 2024, 13: 25–47

12 Ostrov N, Beal J, Ellis T, et al. Technological challenges and milestones for writing genomes. Science, 2019, 366: 310–312

13 Richardson S M, Mitchell L A, Stracquadanio G, et al. Design of a synthetic yeast genome. Science, 2017, 355: 1040–1044

14 Xie Z X, Li B Z, Mitchell L A, et al. “Perfect” designer chromosome V and behavior of a ring derivative. Science, 2017, 355: eaaf4704

15 Wu Y, Li B Z, Zhao M, et al. Bug mapping and fitness testing of chemically synthesized chromosome X. Science, 2017, 355: eaaf4706

16 Miga K H, Koren S, Rhie A, et al. Telomere-to-telomere assembly of a complete human X chromosome. Nature, 2020, 585: 79–84

17 Brown D M, Glass J I. Technology used to build and transfer mammalian chromosomes. Exp Cell Res, 2020, 388: 111851

18 Katoh M, Ayabe F, Norikane S, et al. Construction of a novel human artificial chromosome vector for gene delivery. Biochem Biophys Res Commun, 2004, 321: 280–290

19 Pesenti E, Liskovykh M, Okazaki K, et al. Analysis of complex DNA rearrangements during early stages of HAC formation. ACS Synth Biol, 2020, 9: 3267–3287

20 Pesenti E, Kouprina N, Liskovykh M, et al. Generation of a synthetic human chromosome with two centromeric domains for advanced epigenetic engineering studies. ACS Synth Biol, 2018, 7: 1116–1130

21 Mitchell L A, McCulloch L H, Pinglay S, et al. *De novo* assembly and delivery to mouse cells of a 101 kb functional human gene. Genetics, 2021, 218: iyab038

22 de Almeida B P, Reiter F, Pagani M, et al. DeepSTARR predicts enhancer activity from DNA sequence and enables the *de novo* design of synthetic enhancers. Nat Genet, 2022, 54: 613–624

23 Quang D, Xie X. DanQ: a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences. Nucleic Acids Res, 2016, 44: e107

24 Fu Z H, He S Z, Wu Y, et al. Design and deep learning of synthetic B-cell-specific promoters. Nucleic Acids Res, 2023, 51: 11967–11979

25 Zhang P, Wang H, Xu H, et al. Deep flanking sequence engineering for efficient promoter design using DeepSEED. Nat Commun, 2023, 14: 6309

26 Taskiran I I, Spanier K I, Dickmänken H, et al. Cell-type-directed design of synthetic enhancers. Nature, 2024, 626: 212–220

27 Yin C, Castillo-Hair S, Byeon G W, et al. Iterative deep learning design of human enhancers exploits condensed sequence grammar to achieve cell-type specificity. Cell Syst, 2025, 16: 101302

28 Ji Y, Zhou Z, Liu H, et al. DNABERT: pre-trained bidirectional encoder representations from transformers model for DNA-language in genome.

Bioinformatics, 2021, 37: 2112–2120
29 Ma M, Liu G, Cao C, et al. HybriDNA: a hybrid transformer-mamba2 long-range DNA language model. 2025, arXiv: 2502.10807
30 Zhang D, Zhang W, Zhao Y, et al. DNAGPT: a generalized pre-trained tool for versatile DNA sequence analysis tasks. 2023, arXiv: 2307.05628
31 Nguyen E, Poli M, Durrant M G, et al. Sequence modeling and design from molecular to genome scale with Evo. Science, 2024, 386: eado9336
32 Brixi G, Durrant M G, KU J, et al. Genome modeling and design across all domains of life with Evo 2. BioRxiv, 2025, 2025: 638918
33 Venter J C, Smith H O, Adams M D. The sequence of the human genome. Clin Chem, 2015, 61: 1207–1208
34 Logsdon G A, Vollger M R, Hsieh P H, et al. The structure, function and evolution of a complete human chromosome 8. Nature, 2021, 593: 101–107
35 Lander E S, Linton L M, Birren B, et al. Initial sequencing and analysis of the human genome. Nature, 2001, 409: 860–921
36 Xu H, Chi Y, Yin C, et al. Three-dimensional genome structures of single mammalian sperm. Nat Commun, 2025, 16: 3805
37 Iida Y, Kim J H, Kazuki Y, et al. Human artificial chromosome with a conditional centromere for gene delivery and gene expression. DNA Res, 2010, 17: 293–301
38 Kouprina N, Petrov N, Molina O, et al. Human artificial chromosome with regulated centromere: a tool for genome and cancer studies. ACS Synth Biol, 2018, 7: 1974–1989
39 Kim J H, Kononenko A, Erliandri I, et al. Human artificial chromosome (HAC) vector with a conditional centromere for correction of genetic deficiencies in human cells. Proc Natl Acad Sci USA, 2011, 108: 20048–20053
40 Kouprina N, Samoshkin A, Erliandri I, et al. Organization of synthetic alphoid DNA array in human artificial chromosome (HAC) with a conditional centromere. ACS Synth Biol, 2012, 1: 590–601
41 Logsdon G A, Gambogi C W, Liskovykh M A, et al. Human artificial chromosomes that bypass centromeric DNA. Cell, 2019, 178: 624–639.e19
42 Gambogi C W, Birchak G J, Mer E, et al. Efficient formation of single-copy human artificial chromosomes. Science, 2024, 383: 1344–1349
43 Li H, Xu S, Liu Y, et al. Efficient *de novo* assembly of 100 kb-scale human functional immunoglobulin heavy variable (IGHV) gene fragments *in vitro*. ACS Synth Biol, 2025, 14: 2559–2571
44 Lartigue C, Vashee S, Algire M A, et al. Creating bacterial strains from genomes that have been cloned and engineered in yeast. Science, 2009, 325: 1693–1696
45 DiCarlo J E, Norville J E, Mali P, et al. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res, 2013, 41: 4336–4343
46 Postma E D, Dashko S, van Breemen L, et al. A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. Nucleic Acids Res, 2021, 49: 1769–1783
47 Karas B J, Jablanovic J, Sun L, et al. Direct transfer of whole genomes from bacteria to yeast. Nat Methods, 2013, 10: 410–412
48 Muller H, Annaluru N, Schwerzmann J W, et al. Assembling large DNA segments in yeast. Methods Mol Biol, 2012, 852: 133–150
49 Kouprina N, Larionov V. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. Chromosoma, 2016, 125: 621–632
50 Kouprina N, Larionov V. Transformation-associated recombination (TAR) cloning and its applications for gene function; genome architecture and evolution; biotechnology and biomedicine. Oncotarget, 2023, 14: 1009–1033
51 Laurent J M, Fu X, German S, et al. Big DNA as a tool to dissect an age-related macular degeneration-associated haplotype. Precision Clin Med, 2019, 2: 1–7
52 Pinglay S, Bulajić M, Rahe D P, et al. Synthetic regulatory reconstitution reveals principles of mammalian *Hox* cluster regulation. Science, 2022, 377: eabk2820
53 Brosh R, Coelho C, Ribeiro-dos-Santos A M, et al. Synthetic regulatory genomics uncovers enhancer context dependence at the *Sox2* locus. Mol Cell, 2023, 83: 1140–1152.e7
54 Camellato B R, Brosh R, Ashe H J, et al. Synthetic reversed sequences reveal default genomic states. Nature, 2024, 628: 373–380
55 Zhang W, Golynker I, Brosh R, et al. Mouse genome rewriting and tailoring of three important disease loci. Nature, 2023, 623: 423–431
56 Zürcher J F, Kleefeldt A A, Funke L F H, et al. Continuous synthesis of *E. coli* genome sections and Mb-scale human DNA assembly. Nature, 2023, 619: 555–562
57 Zhong L, Zhang Q, Lu N, et al. The conjugation-associated linear-BAC iterative assembling (CALBIA) method for cloning 2.1-Mb human chromosomal DNAs in bacteria. Cell Res, 2025, 35: 309–312

58 He B, Ma Y, Tian F, et al. YLC-assembly: large DNA assembly via yeast life cycle. Nucleic Acids Res, 2023, 51: 8283–8292
59 Liu Y, Zhou J, Liu D, et al. *De novo* assembly and delivery of synthetic megabase-scale human DNA into mouse early embryos. Nat Methods, 2025, 22: 1686–1697
60 Wang L B, Li Z K, Wang L Y, et al. A sustainable mouse karyotype created by programmed chromosome fusion. Science, 2022, 377: 967–975
61 Xu J, Xu K, Jung S, et al. Nanobodies from camelid mice and llamas neutralize SARS-CoV-2 variants. Nature, 2021, 595: 278–282
62 Suzuki N, Itou T, Hasegawa Y, et al. Cell to cell transfer of the chromatin-packaged human β-globin gene cluster. Nucleic Acids Res, 2010, 38: e33

Construction and application of mammalian chromosomes

ZHONG Yuan, LIU PeiYuan, ZHU JingYi, FU ZongHeng, ZHOU LongFei, LI Xia, LIU Yue, YUAN YingJin* & ZHAO GuangRong*

State Key Laboratory of Synthetic Biology, Frontiers Science Center for Synthetic Biology (Ministry of Education), School of Synthetic Biology and Biomanufacturing, Tianjin University, Tianjin 300072, China
** Corresponding authors, E-mail: yjyuan@tju.edu.cn; grzhao@tju.edu.cn*

The design and construction of chromosomes represent foundational technologies in synthetic biology. Over the past two decades, genomic design and synthesis research has made significant progress, transitioning from the microbial genomes to mammalian chromosomes. From the chromosomal structural elements and functional sequences to the unique application of the chromosomal disease models and humanized immune systems, the design and construction of mammalian chromosomes offer a novel technical paradigm for exploring and reconstructing the genomes of higher organisms. In this article, we focused on the design-construction-application workflow of mammalian chromosomes, and summarized the progress of the design principles, assembly methods, and practical applications. We discussed the challenges and proposed the future prospects for designing and constructing mammalian chromosomes.

synthetic genomics, chromosome design, chromosome construction, artificial chromosome, mammalian cell

doi: 10.1360/SSV-2025-0211