

Technologies for engineering repetitive DNA

Shuting Ma^{1,2,3,4} | Yali Cui^{1,2,3,4} | Yi Wu^{1,2,3}

¹State Key Laboratory of Synthetic Biology, Tianjin University, Tianjin, China

²Frontiers Science Center for Synthetic Biology and Key Laboratory of Systems Bioengineering (Ministry of Education), School of Synthetic Biology and Biomanufacturing, Tianjin University, Tianjin, China

³Frontiers Research Institute for Synthetic Biology, Tianjin University, Tianjin, China

⁴School of Chemical Engineering and Technology, Tianjin University, Tianjin, China

Correspondence

Yi Wu.
Email: yi.wu@tju.edu.cn

Funding information

The National Key Research and Development Program of China, Grant/Award Number: 2024YFA0917400; the National Natural Science Foundation of China, Grant/Award Numbers: 32471483, 32551084; the Natural Science Foundation of Tianjin, Grant/Award Number: 23JCYBJC00220

Abstract

Repetitive DNA, a fundamental architectural element of genomes, is widespread across organisms and comprises about 54% of the human genome. With advances in long-read sequencing and bioinformatics approaches, highly repetitive sequences can now be characterized in depth. Nevertheless, engineering repetitive DNA is mainly limited by repeat instability. Despite these challenges, advances have been reported in the construction of short tandem repeats and the large-scale duplication of genes. This review provides a comprehensive overview of current strategies for constructing repetitive DNA, covering both in vitro approaches, including restriction endonuclease ligation, in vitro homologous recombination, and redundancy-minimizing sequences, as well as in vivo methodologies in *Escherichia coli*, *Saccharomyces cerevisiae*, and mammalian cells. We further discuss applications of these methods for developing genetic manipulation tools, constructing models of repetitive DNA, and synthesizing biomaterials. Future progress in engineering repetitive DNA will further advance our capabilities in designing and manipulating complex biological systems and deepen our understanding of them.

KEYWORDS

DNA assembly, genetic manipulation tools, genome synthesis, repetitive DNA engineering, synthetic biology

1 | INTRODUCTION

Repetitive DNA, defined as sequence motifs that recur throughout the genome, comprises about 54% of the human genome [1]. Large amounts of repetitive DNA are likely to be non-functional, decaying remnants of transposable elements (TEs), which have been labeled “junk” or “selfish” DNA [2]. However, growing evidence suggests that repetitive sequences in both coding and noncoding regions play essential roles in evolution, regulation of gene expression, and the generation of genetic variation [3–6]. Over 60 diseases have been identified as being caused by the expansion of simple

repetitive DNA in the genome, with the majority affecting the nervous system, particularly trinucleotide repeat diseases such as Huntington's disease, fragile X syndrome, several spinocerebellar ataxias, myotonic dystrophy, and Friedreich's ataxia [7, 8].

Repetitive DNA is broadly categorized into two types: tandem repeats and interspersed repeats, according to the arrangement of repeat units [9]. Tandem repeats consist of head-to-tail arrays of repeat units. By repeat unit length, tandem repeats are classified into three categories: microsatellites (short tandem repeats, STRs), minisatellites (variable number tandem repeats, VNTRs), and satellite DNA [10]. STRs, defined as

Shuting Ma and Yali Cui contributed equally.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Quantitative Biology* published by John Wiley & Sons Australia, Ltd on behalf of Higher Education Press.

motifs of 1–6 bp, are enriched in regulatory regions, including enhancers, where they can modulate gene expression. Their instability leads to expression variability and underlies neurological repeat-expansion disorders [11, 12]. VNTRs, defined as tandem arrays with repeat-unit lengths of about 10–100 bp, have been shown to influence gene expression, protein function, and complex traits [13–15]. Satellite DNA is enriched in centromeric and pericentromeric heterochromatin and is often present at subtelomeric regions. It exhibits considerable variation in unit size across species, from very short simple satellites in insects to nucleosome-length units in mammals. It contributes to chromosome segregation (kinetochore function), heterochromatin formation, and telomere-associated processes [16, 17]. Interspersed repeats are repetitive DNA derived mainly from TEs, including Class I retrotransposons and Class II DNA transposons. Unlike tandem repeats, whose units are contiguous, interspersed repeats are dispersed throughout the genome and occur as nonadjacent copies [18, 19]. TEs can induce genomic instability through insertional mutations and chromosomal rearrangements. Conversely, they may also contribute to adaptive evolution by providing regulatory elements such as promoters and enhancers that influence gene expression [20–23].

Advances in long-read sequencing, combined with bioinformatics, enable the in-depth characterization of highly repetitive genomic elements by resolving their repeat-unit composition, higher-order organization, and locus-specific variation [24]. Nevertheless, the construction of repetitive DNA remains challenging, primarily stemming from structural instabilities caused by replication slippage during amplification, misassembly driven by high sequence homology, and genetic instability of constructs during serial passaging [25, 26]. To address these challenges, a suite of in vitro amplification-based strategies has been developed, including isothermal amplification-mediated methods, mispairing-guided polymerase extension, restriction endonuclease-mediated ligation, in vitro homologous recombination, and redundancy-minimizing sequence, enabling the construction of tandem repeat tracts spanning kilobases to tens of kilobases in length. In parallel, in vivo construction strategies using model hosts, such as *Escherichia coli* (*E. coli*), *Saccharomyces cerevisiae* (*S. cerevisiae*), and mammalian cells, are being actively developed to yield longer, more structurally complex, and functionally annotated repeat arrays. Notably, amplification editing (AE) has been reported to achieve duplication of 20 bp to 100 Mb with high efficiency and precision [27]. Here, we review current strategies for constructing repetitive DNA, systematically compare in vitro and in vivo approaches, and discuss the potential of engineered repetitive DNA for the development of genetic manipulation tools (specifically multiplex CRISPR-Cas system development

and artificial chromosome engineering), constructing models of repetitive DNA, and biomaterials synthesis, aiming to guide methodological choice and downstream implementation (Figure 1).

2 | CONSTRUCTION OF REPETITIVE DNA

As the sequencing and annotation of repetitive DNA continue to improve, rapid progress has been made in the construction of repetitive DNA. Among current strategies, in vitro generation of kilobase (kb) to tens-of-kb repeat fragments is feasible; however, efficiency declines as fragment length and repeat-unit homogeneity increase. In vivo strategies are advancing, and in situ construction of chromosome-scale repetitive arrays in mammalian cells has been enabled by recent research.

2.1 | In vitro methods for the construction of repetitive DNA

Based on mechanisms of repetitive DNA formation and guiding principles, in vitro strategies for constructing repetitive DNA can be grouped into five broad classes: isothermal amplification-mediated construction of repetitive DNA, mispairing-guided polymerase extension for the construction of repetitive DNA, restriction endonuclease-mediated construction of repetitive DNA, in vitro homologous recombination for the construction of repetitive DNA, and constructing repetitive DNA by reducing sequence redundancy (Figure 2).

2.1.1 | Isothermal amplification-mediated construction of repetitive DNA

In highly homogeneous tandem repeats, conventional PCR is prone to strand slippage and template switching, thereby limiting the accurate amplification of long repetitive sequences. Rolling circle amplification (RCA) utilizes circular templates and strand-displacement DNA polymerases to perform isothermal amplification at around 30°C, producing long single-stranded DNA (ssDNA) concatemers composed of tandem repeats of the template. Product lengths commonly reach tens of kilobases (around 30–70 kb), depending on the system and conditions [28–30]. In the absence of additional primers, RCA proceeds with nearly linear accumulation, adding extra primers produces hyperbranched RCA (HRC) with exponential kinetics [31]. RCA has been widely used to generate simple tandem repeats, functional genetic modules, and scaffolds for DNA nanostructures [32–37].

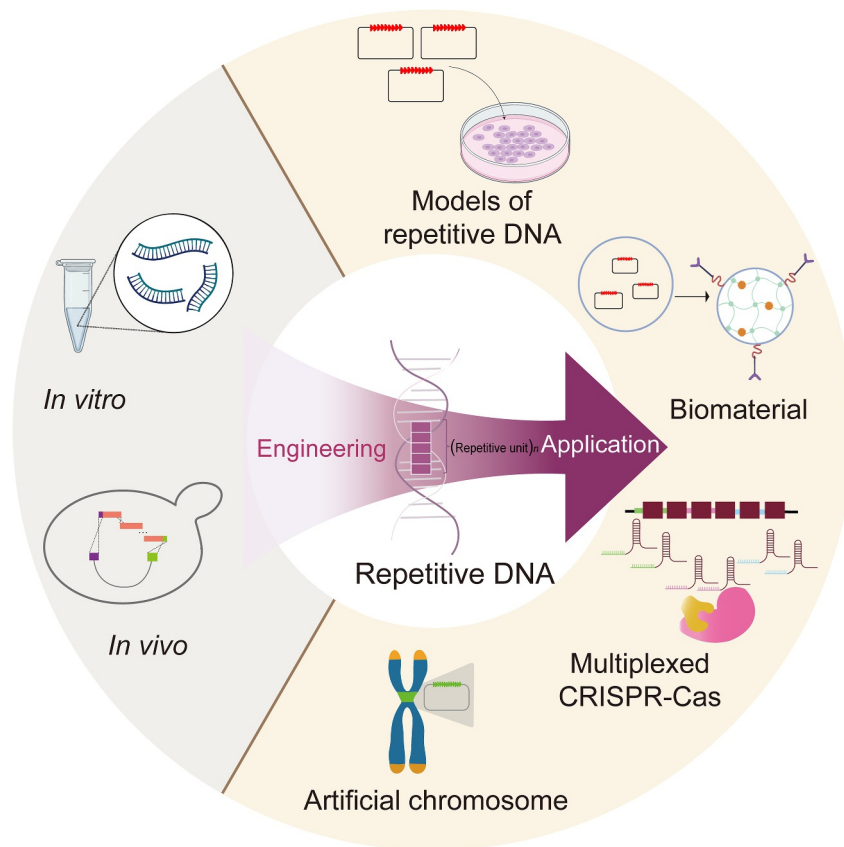


FIGURE 1 Overview of engineering strategies and applications of repetitive DNA.

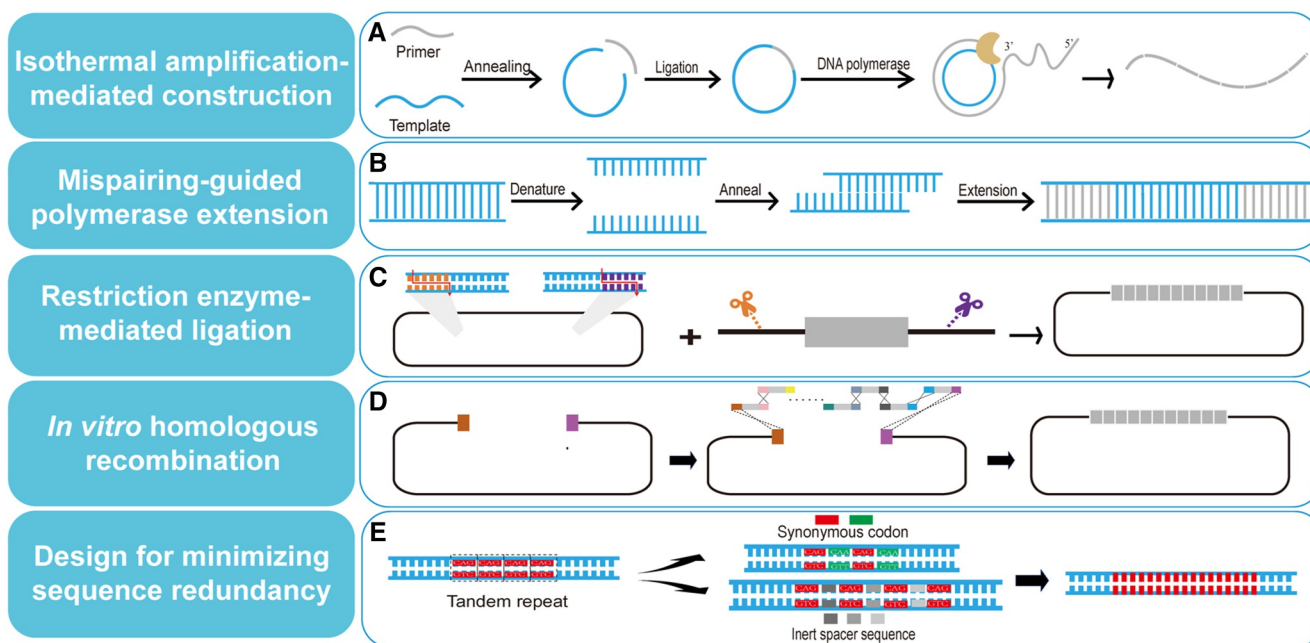


FIGURE 2 Mechanism-based in vitro strategies for constructing repetitive DNA. (A) Isothermal amplification-mediated construction. (B) Mispairing-guided polymerase extension. (C) Restriction enzyme-mediated ligation. (D) In vitro homologous recombination. (E) Design for minimizing sequence redundancy.

In overlap extension rolling circle amplification (OERCA), RCA is combined with overlap-extension PCR to generate double-stranded DNA (dsDNA) libraries with tunable repeat copy numbers in a single-pot reaction. Early studies have shown that OERCA yields a broader, more tunable length distribution than overlap extension PCR (OE-PCR) or ligation-based methods during the synthesis of repetitive genes such as elastin-like polypeptides (ELPs) [38, 39].

Overall, despite these advantages, RCA-based strategies remain constrained by the need for efficient template circularization, variability in product-length distributions, and limited portability across sequence contexts and systems. For constructing ultra-long, high-identity repeat arrays, these methods are typically combined with complementary strategies and workflows.

2.1.2 | Mismatching-guided polymerase extension for the construction of repetitive DNA

During the replication of highly repetitive DNA, DNA polymerase is prone to strand slippage, a phenomenon characterized by the transient dissociation of strands followed by misaligned reannealing. This phenomenon results in the insertion or deletion of repeat units and is a key mechanism underlying microsatellite locus length variation and in vitro amplification artifacts [40].

The synthesis of long iterative polynucleotide (SLIP) exploits this mechanism: efficient expansion of tandem repeats, such as CAG/CTG, is achieved through repeated cycles of denaturation and annealing, followed by gap filling with a high-fidelity DNA polymerase. However, the final product length is heterogeneous and not precisely predictable, and cloning and screening are required to isolate constructs of defined length [41–43]. In enzymatic heat-cool cycling extension, misaligned double-stranded intermediates are generated through repeated thermal cycles, producing 5' overhangs that are extended by DNA polymerase. This enables the synthesis of long repeats from short oligonucleotides, typically yielding products of 0.5–20 kb [44]. For GC-rich or longer repeat units, broader length distributions are observed, and self-complementary or concatemeric products are more likely to form [45].

Repeat-extension PCR (RepEx-PCR) enables the generation of long tandem repeat arrays from two short complementary oligonucleotides by offset annealing, which creates 5' or 3' overhangs. A high-fidelity DNA polymerase fills 5' overhangs and, via its 3–5' proof-reading exonuclease, removes 3' overhangs, producing blunt-ended products that are progressively extended each cycle to yield kb-scale arrays. However, blunt-end ligation is inefficient for long inserts, and long repeats are unstable in *E. coli*, alternative hosts may improve cloning and stability [46, 47]. In blocking PCR for DNA repeat expansion (BPRe), an excess partially

matching blocking primer anneals to the 5' portion of the repeat unit, biasing an offset annealing event in which the 3' overhang of a template primer pairs with the exposed 3' portion of the repeat. Taq polymerase then extends along the template primer to add one repeat unit per productive cycle, yielding double-stranded repeat arrays, whereas a subset of molecules fully reanneal and thus skip expansion in a given round [48].

2.1.3 | Restriction endonuclease-mediated construction of repetitive DNA

Standardized, modular, iterative construction strategies integrate genetic elements into an ordered, hierarchical workflow comprising three levels: component, module, and array, allowing parts to be reused across iterations. Under the BioBrick standard, predefined restriction sites in the prefix and suffix enable idempotent assembly. However, this approach introduces sequence scars, and pairwise ligation steps become inefficient for long tandem arrays [49, 50]. Within this framework, quick gene assembly (QGA) has been used in vitro to construct a tandem array of 25 BioBrick units through iterative *Xba* I/*Spe* I digestion and ligation [51]. In Golden Gate assembly, Type IIS restriction enzymes cleave outside their recognition sites to generate custom overhangs, enabling scarless one-pot, multi-fragment assembly, which is well suited to constructing ordered repeat arrays [52–55]. In PS-Brick, Type IIP and Type IIS enzymes are combined to achieve scarless assembly while preserving part reusability and supporting iterative amplification. However, it is limited by the need to remove internal recognition sites, and its assembly workflow is more complex than that of one-pot Golden Gate methods [49]. Cloning troublesome repeats in loops (CTRL) utilizes BioBrick or Golden Gate-based methods to expand arrays twofold or fourfold per cycle, ultimately reaching 256 binding sites, with accuracy and yield gradually declining as length increases [56].

Beyond these modular schemes, recursive ligation-based approaches have been developed to extend repeat arrays iteratively. In recursive directional ligation (RDL), two restriction enzymes with different recognition sequences are chosen so that their cohesive ends are compatible, enabling iterative, directional ligation to generate repeat arrays. The strategy has been utilized to synthesize libraries of ELPs and to develop seamless cloning workflows for large microsatellite repeat expansions, which are, in principle, applicable to diverse STRs [57, 58]. Although RDL is effective, its utility is limited by sequence constraints imposed by restriction sites embedded within coding regions, substantial background from vector self-ligation and insert circularization, and the burden of multistep cloning. RDL by plasmid reconstruction (PRe-RDL) addresses these

limitations by reconstructing the plasmid using Type IIS cleavage outside the recognition sites, thereby minimizing background and reducing the retention of restriction sites within the repetitive tract, which improves cloning efficiency and experimental control [59, 60].

2.1.4 | In vitro homologous recombination for the construction of repetitive DNA

In overlap-based in vitro construction, about 20–40 bp overlaps are appended to the ends of fragments, enabling directional, scarless joining via base pairing. Overlaps are readily appended by PCR, providing flexibility in experimental design. The Gibson assembly exemplifies this approach: a 5' exonuclease creates 3' ssDNA overhangs that are filled by DNA polymerase and sealed by DNA ligase, allowing multi-fragment assembly in a single isothermal reaction. Although effective for complex constructs, high-copy repeats increase mis-annealing, template switching, and rearrangements, potentially yielding incorrect junctions [61, 62].

To address these limitations, unique nucleotide sequences (UNS) can flank each repetitive module as orthogonal linkers, in conjunction with Gibson or other recombination-based cloning strategies, ensuring the accurate and ordered formation of repeat arrays [63–65]. Similarly, modular overlap-directed assembly with linkers (MODAL) utilizes standardized adapters and orthogonal overlaps to assign unique addresses to modules, improving accuracy, scalability, and modularity in building repeat-containing constructs [66].

2.1.5 | Constructing repetitive DNA by reducing sequence redundancy

The construction of repetitive DNA typically relies on restriction enzyme digestion and homology-directed scarless assembly. However, under high sequence identity, overlaps are prone to mis-annealing and cross-hybridization, incorrect junctions can occur, construct instability increases, and workflows become more complex. To mitigate these challenges, one approach is derepetitization (synonymous recoding), in which the nucleotide sequence is rewritten without altering the encoded protein. For example, the polyglutamine (polyQ) CAG tract can be recoded as a synonymous sequence incorporating CAA interruptions [67]. Computational approaches such as combinatorial codon scrambling [68–70] and tandem repeat DNA sequence redesign (TRSR) [71] can minimize nucleotide-level identity between adjacent repeat units, thereby improving the accuracy of amplification.

When codon recoding is not feasible, algorithmic design of non-repetitive genetic parts, such as non-repetitive parts calculator (NRPC)-generated promoter libraries, offers a validated strategy to reduce local sequence identity while maintaining regulatory function [72]. For co-expression of multiple single-guide RNAs (sgRNAs), non-repetitive part libraries can be generated through directed mutagenesis guided by algorithmic constraints, thereby minimizing internal sequence redundancy [73, 74]. Alternatively, a single-promoter design can transcribe the entire sgRNA array, reducing the need for multiple identical promoter elements [75–77]. In addition, predefined non-repetitive linkers placed at cleavage sites enable the one-pot assembly of multi-site CRISPR arrays while reducing mis-annealing caused by repetitive elements [78, 79] (Table 1).

2.2 | In vivo methods for the construction of repetitive DNA

Engineering repetitive DNA commonly relies on in vitro assembly, a laborious process that is difficult to scale to vast constructs. By contrast, approaches based on host cells are more effective for assembling larger DNA constructs in vivo. Although strategies focused on repeats remain limited, in vivo assembly has been actively explored in *E. coli*, *S. cerevisiae*, and mammalian cells (Figure 3).

2.2.1 | Construction of repetitive DNA in *E. coli*

Leveraging its low endogenous homologous recombination, *E. coli* provides a stable host for large DNA and can stably maintain human fragments of tens to hundreds of kilobases. However, conventional workflows are not readily scalable or iterative. Recently, bacterial artificial chromosome (BAC) stepwise insertion synthesis (BASIS) [80] and conjugation associated linear-bacterial artificial chromosome iterative assembling (CALBIA) [81] have demonstrated truly iterative construction in *E. coli*, extending human genomic segments to the megabase scale and providing a practical route and methodological framework for programmable in vivo construction of repetitive DNA.

BASIS utilizes a BAC scaffold and alternates Cas9 excision with λ -Red recombination and conjugative transfer, relying solely on short terminal homology arms to facilitate iterative, scarless joining of large human DNA blocks. This workflow culminated in a 1.1 Mb segment from human chromosome 21 that retained LINE/SINE elements, G-quadruplex motifs, and STRs, such

TABLE 1 In vitro methods for constructing repetitive DNA.

Mechanism	Method	Maximum repeat sequence size	Reference
Isothermal amplification-mediated	Rolling circle amplification (RCA)	Approaches 70 kb	[27, 28, 31–33]
	Overlap extension rolling circle amplification (OERCA)	1.4 kb	[37, 38]
Mispairing-guided polymerase extension	Synthesis of long iterative polynucleotide (SLIP)	0.57 bp	[41, 42]
	Enzymatic heat-cool cycling extension	> 20 kb	[43, 44]
	Repeat-extension PCR (RepEx-PCR)	0.6 kb (reach tens of kb theoretically)	[45, 46]
	Blocking PCR for DNA repeat expansion (BPPE)	1 kb	[47]
Restriction endonuclease-mediated	Quick gene assembly (QGA)	1.714 kb	[50]
	Golden Gate	4.8 kb (256 repeat units)	[51, 53–55]
	PS-Brick	Up to 3 sgRNA repeats	[48]
	Recursive directional ligation (RDL)	3.6 kb	[56–59]
In vitro homologous recombination	Gibson	3.384 kb	[60, 61]
	Unique nucleotide sequences (UNS)-guided assembly	64 kb	[62–64]
	Modular overlap-directed assembly with linkers (MODAL)	5–20 kb	[65]
Design for minimizing sequence redundancy	Based on codon degeneracy	2.1 kb	[66, 79]
	Combinatorial codon scrambling	2.52 kb	[67–69]
	Tandem repeat DNA sequence redesign (TReSR)	0.534 kb	[70]
	Co-expression of multiple sgRNAs	NA	[53, 74, 76–78, 80]

sequences are comparatively unstable in *S. cerevisiae*, making *E. coli* a more suitable host for carrying and amplifying repeat-rich human DNA [80].

CALBIA, by contrast, builds linear BACs and exploits host RecA-mediated single crossover homologous recombination during bacterial conjugation to achieve day-scale iterations, joining fragments in the hundreds of kb range without the use of Cas9 or λ -Red, thereby reducing the risk of mis-recombination at small repeat tracts after Cas9 cleavage. In practice, CALBIA has been used to construct a 1.07 Mb IgHV cluster and about 1.1–1.2 Mb segments from 18p11.2 in *E. coli*. Free linear BACs carrying near Mb inserts are deletion-prone in AT-rich or repeat-dense regions, whereas linear chromosome integration confers long-term genetic stability [81].

2.2.2 | Construction of repetitive DNA in *S. cerevisiae*

S. cerevisiae is widely used for in vivo DNA construction owing to its highly efficient homologous recombination, mature genetic toolkits, and capacity to maintain

large DNA constructs. Developing more versatile strategies for generating synthetic repeat arrays within the *S. cerevisiae* genome would benefit a broad range of synthetic biology applications.

Transformation-associated recombination (TAR) combined with RCA enables the rapid construction of tandem repeat arrays. Following an initial in vitro RCA step, short repeat units are first concatemerized into multi-kb products, and a single TAR step then assembles these into arrays of tens of kilobases. Additional cycles can extend the arrays to the hundreds-of-kilobases range. However, single-round TAR yields of long arrays are low, and the workflow is relatively complex and time-consuming [82, 83]. Alternatively, CRISPR with *S. cerevisiae* homologous recombination generates targeted double-strand breaks (DSBs), followed by cotransformation of partially overlapping fragments, enabling direct in vivo generation of approximately 1–100 copies. The approach is helpful for medium copy numbers and complementary to in vitro preassembly, but it does not allow the exact number of repeats to be predetermined [84].

Paired nicking-induced amplification (PNAm) places a pair of Cas9 nickases (nCas9) on opposite

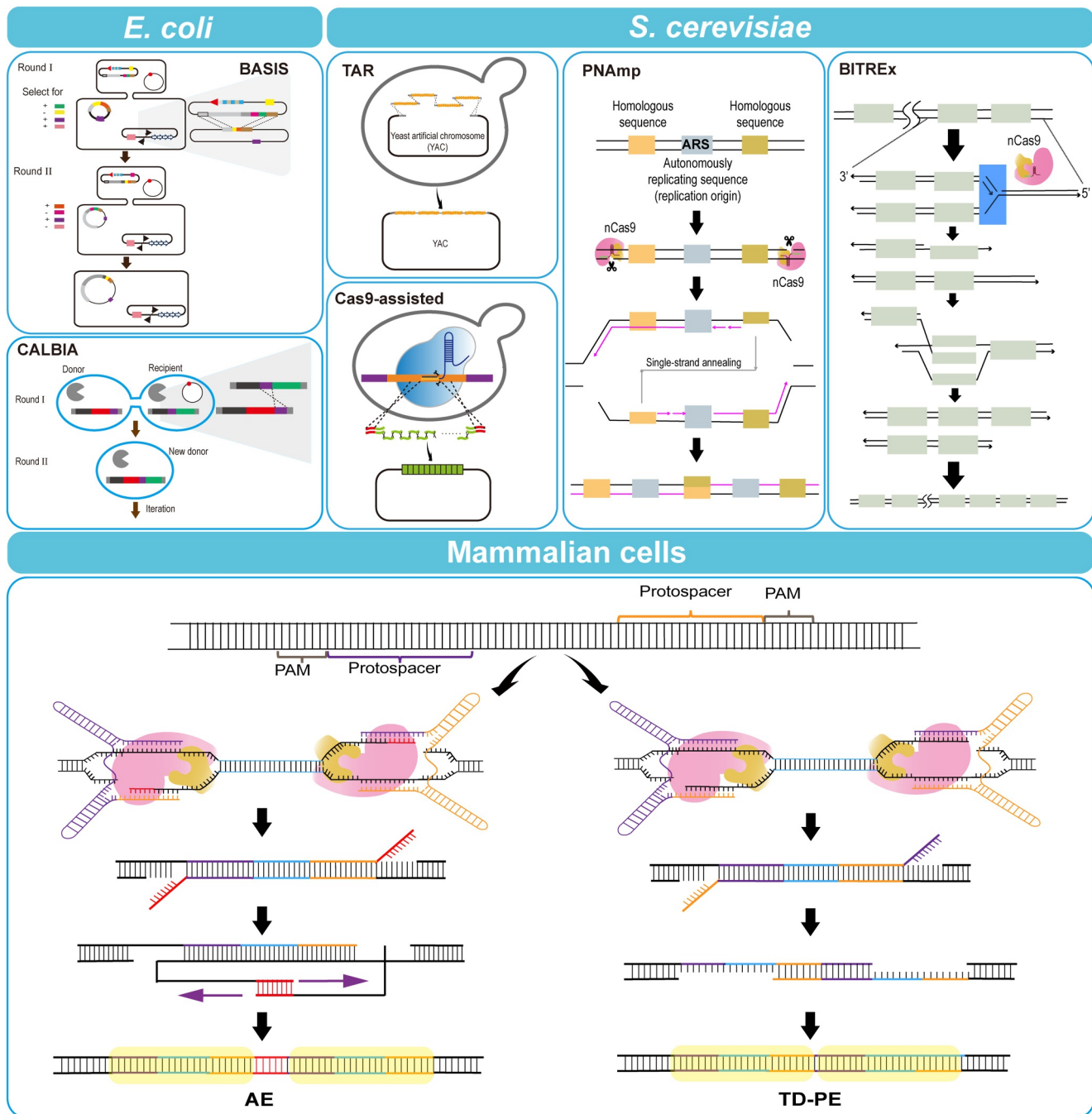


FIGURE 3 Schematic of in vivo strategies for engineering repetitive DNA across *E. coli*, *S. cerevisiae*, and mammalian cells.

strands flanking a replication origin at the boundaries of the target segment so that the two forks emanating from that origin collapse at the Nick sites, generating paired one-ended DSB. After end processing, single-strand annealing (SSA) between co-oriented direct repeats joins the intervening segment head-to-tail, producing a segmental duplication that can be iteratively expanded. This approach induces duplication of segments as large as 1 Mb with efficiencies exceeding 10% in the budding yeast [85]. Building on this foundation, BIR-mediated

tandem repeat expansion (BITREx) likewise couples site-specific nicking to replication-dependent repair, instead invoking break-induced replication (BIR) from an upstream repeat unit to drive head-to-tail copy number gains. However, because this process is replication-coupled, it does not apply to nondividing cells [86].

Beyond repeat-specific strategies, a segmented, iterative scheme combining long homology arms, cleavage mediated by Cas9, and TAR synthesis

resulted in a human AZFa segment of 1.14 Mb with about 70% repetitive content [87]. Construction of a repetitive array by transformation (CAT) introduces linear fragments with flanking homology in a single step, endogenous recombination with in situ tandem amplification yielded ribosomal DNA (rDNA) arrays with about 10–15 copies [88]. The yeast life cycle (YLC) assembly leveraged the YLC to iteratively build a human IGH region of 1.26 Mb, rich in repeats. However, the build remained unstable, and dispersed repeats caused localized deletions [89].

2.2.3 | Construction of repetitive DNA in mammalian cells

Because homology-directed repair (HDR) is relatively inefficient for joining multi-fragment sequences in mammalian cells, the direct construction of repeat-rich constructs remains challenging. To build tandem repeats directly in situ, prime editing-based strategies now enable programmable duplication of sequences. Tandem duplication via prime editing (TD-PE) utilizes paired pegRNAs to program precise duplication segments ranging from 50 bp to 10 kb segments in mammalian cells [90]. AE couples nCas9-reverse transcriptase with paired PAM-out pegRNAs to trigger templated DNA amplification, supporting duplications from tens of base pairs to chromosomal scales without the need for donor DNA [27]. However, these methods mainly produce in situ tandem repeats for pre-existing sequences in the genome, which makes it difficult to amplify exogenous or synthetic sequences. In addition, yeast origin strategies have been adapted for use in mammals, BITREx utilizes BIR from nickase-induced fork breakage to expand tandem arrays in mammalian cells, providing complementary routes to engineer repetitive DNA [86] (Table 2).

3 | APPLICATIONS OF ENGINEERING REPETITIVE DNA

3.1 | Development of genetic manipulation tools

Repetitive DNA engineering is catalyzing the development of next-generation genetic manipulation tools. Multiplex CRISPR-Cas architectures utilize modular repeat arrays to coordinate editing or regulation at multiple loci, whereas artificial chromosome engineering employs precisely constructed centromeric repeats to nucleate the centromere and kinetochore assembly, thereby promoting mitotic stability. Together, these capabilities translate sequence programmability into scalable genomic control, expanding the practical reach of genetic manipulation.

Multiplex CRISPR technology allows for the precise modification of multiple genomic targets within a single experimental setup by integrating multiple guide RNA sequences into a single expression system and pairing them with the corresponding Cas proteins [92]. However, sequence similarity among units promotes mis-annealing, recombination, and misjoining. To address this, non-repetitive array design strategies, together with modular cloning frameworks, are commonly used to ensure stable construct assembly and efficient expression. These advances have broadened the application of multiplex CRISPR across diverse systems, including microbes, mammals, and plants.

Multiplex genome editing is typically achieved using multiple sgRNAs, often organized as sgRNA arrays, to target multiple loci in parallel and generate indels or other editing outcomes [75, 78, 91, 93–98]. Multi-sgRNA arrays enable chromosome-scale engineering (kb to Mb), including deletions, inversions, translocations, and reduction or near elimination of entire chromosomes, providing tools to probe genome architecture and model chromosomal disorders [99]. For transcriptional control, dCas effectors enable the simultaneous targeting of multiple genes via CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa), facilitating pathway rewiring and coordinated upregulation [74, 77, 100–103]. Multiplex base editing uses gBlock or Golden Gate assembled sgRNA cassettes with cytidine or adenine base editors (CBEs/ABEs) to modify dozens of sites in a single transfection, including broad TAG to TAA codon replacement [104].

The de novo construction of artificial chromosomes requires the precise synthesis and efficient assembly of large-scale repetitive DNA arrays to rebuild essential functional domains that nucleate kinetochores, support replication, and ensure stable segregation. In human systems, a synthetic chromosome 17 alpha-satellite higher-order repeat (HOR) unit of 2.7 kb has been directionally multimerized into about 86 kb tandem arrays and cloned into BAC-based vectors. Upon transfection into HT1080 cells, these BAC constructs can seed de novo centromeres and give rise to mitotically stable human artificial chromosomes (HACs) [105]. An alternative pipeline couples RCA in vitro with TAR in yeast to generate about 120–140 kb alphoid DNA arrays. Operator sequences, such as TetO or LacO, can be embedded to permit TetR/LacI tethering for the visualization and manipulation of HAC chromatin and gene expression. De novo HACs formed from these arrays can further expand to megabase scale and exhibit long-term mitotic stability [82, 106, 107].

In plant systems, plant artificial chromosomes can also be constructed de novo by combining the synthesis of repetitive arrays with targeted recruitment of kinetochore proteins. For example, in *Arabidopsis thaliana*, a LacI-CENH3 fusion protein has been used to nucleate neokinetochores at previously inserted LacO tandem

TABLE 2 In vivo methods for constructing repetitive DNA.

In vivo host	Method	Maximum repeat sequence size	References
<i>Escherichia coli</i>	Bacterial artificial chromosome stepwise insertion synthesis (BASIS)	1.1 Mb	[79]
	Conjugation associated linear-bacterial artificial chromosome iterative assembling (CALBIA)	1.07 Mb	[80]
<i>Saccharomyces cerevisiae</i>	Transformation-associated recombination (TAR)	1 Mb	[81, 82, 85]
	Cas9-assisted homologous recombination assembly	Up to 16.5 kb (100 repeats)	[83]
	Paired nicking-induced amplification (PNAmpl)	About 1 Mb	[84]
	BIR-mediated tandem repeat expansion (BITREx)	About 1 Mb	[85]
Mammalian cells	Tandem duplication via prime editing (TD-PE)	10 kb	[91]
	Amplification editing (AE)	100 Mb	[26]

repeats, generating mitotically unstable dicentric chromosomes and providing a proof-of-concept route toward artificial chromosomes in plants [108, 109]. Similarly, in maize, the synthesis of an ABS4 array containing approximately 2400 repeat units, followed by LexA–CENH3-mediated targeting, has resulted in the formation of plant artificial chromosomes that are stably transmitted through meiosis [110, 111].

3.2 | Constructing models of repetitive DNA

More than 60 repeat expansion diseases, most of which primarily affect the nervous system, have been identified to date [7, 8, 112]. Accordingly, engineering in vitro models of repeat expansion diseases is critical for mechanistic dissection, quantitative phenotyping, and scalable testing of candidate interventions.

To address this need, multiple strategies have been employed to reconstruct pathogenic repetitive DNA with precision. For example, restriction enzyme ligation combined with RCA produces CTG arrays at the kb scale, and plasmids bearing hundreds to thousands of CTG repeats reproduce RNA toxicity phenotypes in human cells [113]. SLIP readily yields defined ATXN3 CAG/CAA variants for expression in neuronal models, allowing for the dissection of RNA versus protein toxicity [42]. Additionally, expansion of HTT CAG repeats to approximately 114 units was enabled by Golden Gate cloning combined with CRISPR/Cas9-mediated HDR, followed by site-specific integration into the human genome, yielding isogenic cell lines with varying repeat lengths for Huntington’s disease research [114]. In *Drosophila* models, RDL was used to construct and express vectors containing 3, 36, or 103 pure GGGGCC repeat units, as well as “RNA-only” constructs in which 6-bp stop codon-containing interruptions were inserted every 12 repeats, thereby enabling in vivo modeling of C9ALS/FTD [115]. AE and TD-PE have enabled the precise construction of

disease-relevant repeats at multiple loci in cell and stem-cell contexts, and are emerging as valuable platforms for biological research and disease modeling [27, 90].

Beyond human and animal disease models, engineered repetitive DNA has also been utilized to investigate chromatin regulation in plants and to examine rDNA array dynamics in yeast. Depositing CEN180 arrays into introns nucleates intragenic heterochromatin, marked by CG/CHG/CHH DNA methylation and H3K9me2, in a repeat-dosage-dependent manner. The histone H3K9me2 demethylase IBM1 normally restricts outward spreading, whereas its loss allows this heterochromatic state to be heritably propagated [116]. Inverted repeat RNAs targeting FLOWERING LOCUS T (FT) introns activate intragenic RNA-directed DNA methylation (RdDM) to upregulate FT while preserving its diurnal/spatiotemporal expression pattern [117]. Moreover, tandem booster1 (b1) repeats drive trans-allelic paramutation, which is MOP1-dependent and associated with small interfering RNAs (siRNAs), resulting in transgenerational silencing [118]. Focusing on rDNA, yeast synXII enables the deletion and regeneration of the multigene rDNA array at distinct chromosomal loci and ITS swapping [119], providing a programmable model for nucleolar organization, array copy number homeostasis, and the functional constraints of rRNA processing.

3.3 | Biomaterial synthesis via programmable repeat construction

Biological materials are central to medicine, environmental protection, and industrial manufacturing, and a clear trend toward biodegradability, programmability, and high biocompatibility has been observed. Among fabrication strategies, repetitive sequence construction is used as a key strategy to translate molecular design into functional material performance [120].

In protein materials, recombinant repetitive protein polymers, such as ELPs, amyloid-spider silk proteins, squid ring-teeth proteins, and resilin-like polypeptides (RLPs), have made significant advancements. Using RDL, researchers have expressed ELPs from genes encoding up to about 240 VPGXG repeats [57] and have cloned recombinant spider silk genes of defined length [121]. Golden Gate assembly has been successfully used to construct genes for amyloid-spider silk proteins, ELPs, and squid ring-teeth proteins [122–124]. OERCA enables one-pot, high-throughput synthesis of repetitive genes and has been used to demonstrate *E. coli* expression of ELP libraries and protease-responsive GLP-1 protein polymers [39]. The codon scrambling algorithm was successful in synthesizing a wide range of repeated protein polymers, including ELPs, RLPs, and collagen-like polypeptides, with repeat numbers of up to 150 [68]. Collectively, these methods enable the precise programming of chain length and motif composition in repetitive protein polymers, thereby allowing for tunable thermoresponsive phase transitions and other material functions that underpin controlled drug release and tissue engineering applications [57, 62, 120, 125–127].

In nucleic acid nanomaterials, RCA continuously produces long ssDNA concatemers with tandem repeat units under isothermal conditions. These RCA products can be processed via metal coordination, crosslinking, or mineralization to build DNA nanoflowers (DNFs) [128] and DNA hydrogels [33, 129, 130]. Incorporating chemically modified deoxyribonucleoside triphosphates (dNTPs) enables the direct encoding of fluorescent moieties, ligands, or therapeutics into the structures, allowing for drug and gene delivery, imaging, biosensing, and the creation of degradable scaffolds. These strategies transform programmability at the molecular level of repetitive sequences into manufacturability and functional diversification.

4 | CONCLUSIONS AND PERSPECTIVES

In the era of complete genomes, moving from “reading” to “writing” is not only a natural extension of sequencing but also a necessary test of many genomic inferences. Nevertheless, the precise construction of long repeats remains constrained by both chemistry and biology: on the chemical side, tract length and secondary-structure formation depress synthesis and joining efficiencies [131]; on the biological side, highly homologous repeats are intrinsically unstable during cloning and propagation, leading to copy-number loss and rearrangements [132, 133]. In vitro strategies can generate tens of kb repeat tracts under controlled conditions, but their limitations are broadly predictable. Restriction endonuclease-based routes suffer from the short overhangs produced by many enzymes, which afford insufficient specificity in

repeat-rich contexts and complicate iterative joining. Restriction endonuclease-independent methods also face elevated mis-annealing and chimera risks in high-repeat, GC-rich, or G-quadruplex-forming sequences, and they remain labor-intensive and length-limited. Once constructs exceed about 10–20 kb, shear sensitivity and reduced transformability often translate into host incompatibility or instability [134, 135].

By contrast, in vivo approaches offer clear advantages in achievable length. De novo reconstruction of Mb-scale repeat-rich segments is feasible in yeast [87]. However, the repetitive nature of many human genomic regions can render them unstable in that host [89], underscoring the need to tune recombination pathways and repair balance to stabilize complex architectures. An alternative is to build in hosts with lower endogenous homologous recombination, such as *E. coli* or mammalian cells. In *E. coli*, low natural recombination supports the maintenance of human DNA fragments on the order of hundreds of kilobases. Methods such as BASIS and CALBIA use iterative BAC/linear-BAC workflows to enable programmable in vivo construction of repeat-rich DNA at near-megabase scales, offering a practical route and design playbook [80, 81]. In mammalian cells, tandem duplication via TD-PE and AE can install precise in situ duplications with high efficiency [27, 90]. However, they act on pre-existing genomic sequences; consequently, merging them with synthetic genomics would generalize the approach to the de novo, programmable installation of large-scale synthetic DNA replication.

Overall, most current research primarily circumvents the intrinsic instability of repetitive DNA. Building on existing insights into repetitive sequence stability from molecular dynamics, structural biology, and epigenetics, the next stage should shift toward actively regulating the stability of repetitive DNA. At the host strain, a possible direction is to modulate homologous recombination capacity. This could involve reducing the expression of recombination proteins, constructing a recombination-deficient strain, or selecting naturally low-recombination hosts to enhance the ability of cells to stably maintain repetitive DNA, thereby increasing the attainable length and complexity of engineered repeat constructs. A complementary approach is to modulate the accessibility of repetitive sequences by harnessing epigenetic pathways. Targeted recruitment of heterochromatin-forming factors to repeat domains could lower local accessibility, suppress recombination, and establish boundary-based chromatin compaction as a tunable layer for stabilizing repetitive DNA. In parallel, these strategies can be complemented by advances in computational biology and AI-guided design. Computational and AI-assisted approaches can identify high-risk repetitive segments, generate low-repetition building blocks for redesigning repeat units, and rewrite unstable regions while preserving the encoded protein structure or regulatory function,

thereby facilitating the construction of longer and more stable repetitive DNA arrays. Within this framework, a shift from “difficult to synthesize” toward “precisely controllable” could improve the manufacturability and predictability of complex repeat constructs, providing a firmer methodological footing for higher-order artificial genomes and synthetic life systems.

Advances in repetitive DNA engineering are opening opportunities across diverse domains. In cellular and animal disease models, these methods are shaping a standardized, reproducible framework for mechanistic studies and high-throughput screening. In parallel, they hold promise for converting the molecular programmability of repetitive DNA into manufacturability and functional diversification. When coupled with heterologous expression, they also provide practical routes to the synthesis of high-value biomaterials. At the same time, precise, scalable repeat-rich constructs are a prerequisite for multiplex CRISPR and artificial chromosome engineering. Looking ahead, sequencing and synthesis are likely to continue converging: the costs of building and testing chromosome-scale DNA are expected to decline, whereas methods for engineering repetitive DNA can be anticipated to improve in fidelity, scale, and predictability, thereby accelerating the translation from principles to functional sequences and systems.

AUTHOR CONTRIBUTIONS

Shuting Ma: Conceptualization; investigation; visualization; writing—original draft preparation; writing—review and editing. **Yali Cui:** Investigation; writing—original draft preparation. **Yi Wu:** Conceptualization; project administration; supervision; writing—original draft preparation; writing—review and editing.

ACKNOWLEDGMENTS

This manuscript was supported by the National Key Research and Development Program of China (No. 2024YFA0917400), the National Natural Science Foundation of China (Nos. 32471483 and 32551084), Scientific Research Innovation Capability Support Project for Young Faculty, and the Natural Science Foundation of Tianjin (No. 23JCYBJC00220).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

ETHICS STATEMENT

This article does not contain any studies with human participants or animals performed by any of the authors.

REFERENCES

- [1] Hoyt SJ, Storer JM, Hartley GA, Grady PGS, Gershman A, de Lima LG, et al. From telomere to telomere: the transcriptional and epigenetic state of human repeat elements. *Science*. 2022;376(6588):eabk3112.
- [2] Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet*. 2012;13(1):36–46.
- [3] Zhang X, Meyerson M. Illuminating the noncoding genome in cancer. *Nat Cancer*. 2020;1(9):864–72.
- [4] Mehrotra S, Goyal V. Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. *Genom Proteom Bioinform*. 2014;12(4):164–71.
- [5] Horton CA, Alexandari AM, Hayes MGB, Marklund E, Schaepe JM, Aditham AK, et al. Short tandem repeats bind transcription factors to tune eukaryotic gene expression. *Science*. 2023;381(6664):eadd1250.
- [6] Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. *Genome Biol*. 2018;19(1):199.
- [7] Paulson H. Repeat expansion diseases. *Handb Clin Neurol*. 2018;147:105–23.
- [8] Chen Z, Morris HR, Polke J, Wood NW, Gandhi S, Ryten M, et al. Repeat expansion disorders. *Pract Neurol*. 2025;25(3):204–16.
- [9] Richard G-F, Kerrest A, Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev*. 2008;72(4):686–727.
- [10] Thakur J, Packiaraj J, Henikoff S. Sequence, chromatin and evolution of satellite DNA. *Indian J Manag Sci*. 2021;22(9):4309.
- [11] Kovtun IV, McMurray CT. Features of trinucleotide repeat instability in vivo. *Cell Res*. 2008;18(1):198–213.
- [12] Mirkin SM. Expandable DNA repeats and human disease. *Nature*. 2007;447(7147):932–40.
- [13] Nakamura Y, Koyama K, Matsushima M. VNTR (variable number of tandem repeat) sequences as transcriptional, translational, or functional regulators. *J Hum Genet*. 1998;43(3):149–52.
- [14] Mukamel RE, Handsaker RE, Sherman MA, Barton AR, Hujuel MLA, McCarroll SA, et al. Repeat polymorphisms underlie top genetic risk loci for glaucoma and colorectal cancer. *Cell*. 2023;186(17):3659–73.e3623.
- [15] Zhang S, Song Q, Zhang P, Wang X, Guo R, Li Y, et al. Genome-wide investigation of VNTR motif polymorphisms in 8,222 genomes: implications for biological regulation and human traits. *Cell Genom*. 2024;4(12):100699.
- [16] Willard HF, Wayne JS. Hierarchical order in chromosome-specific human alpha satellite DNA. *Trends Genet*. 1987;3:192–8.
- [17] Novák P, Ávila Robledillo L, Kobližková A, Vrbová I, Neumann P, Macas J. Tarean: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res*. 2017;45(12):e111.
- [18] Hayward A, Gilbert C. Transposable elements. *Curr Biol*. 2022;32(17):R904–9.
- [19] Liao X, Zhu W, Zhou J, Li H, Xu X, Zhang B, et al. Repetitive DNA sequence detection and its role in the human genome. *Commun Biol*. 2023;6(1):954.
- [20] Lee M, Ahmad SF, Xu J. Regulation and function of transposable elements in cancer genomes. *Cell Mol Life Sci*. 2024;81(1):157.
- [21] Ali A, Han K, Liang P. Role of transposable elements in gene regulation in the human genome. *Life*. 2021;11(2):118.
- [22] Bhat A, Ghatage T, Bhan S, Lahane GP, Dhar A, Kumar R, et al. Role of transposable elements in genome stability: implications for health and disease. *Indian J Manag Sci*. 2022;23(14):7802.

- [23] Gebrie A. Transposable elements as essential elements in the control of gene expression. *Mobile DNA*. 2023;14(9):9.
- [24] Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. *Science*. 2022.
- [25] Shi X, Teng H, Sun Z. An updated overview of experimental and computational approaches to identify non-canonical DNA/RNA structures with emphasis on G-quadruplexes and R-loops. *Briefings Bioinf*. 2022;23(6):bbac441.
- [26] Abraham Punnoose J, Ma Y, Hoque ME, Cui Y, Sasaki S, Guo AH, et al. Random formation of G-quadruplexes in the full-length human telomere overhangs leads to a kinetic folding pattern with targetable vacant g-tracts. *Biochemistry*. 2018;57(51):6946–55.
- [27] Zhang R, He Z, Shi Y, Sun X, Chen X, Wang G, et al. Amplification editing enables efficient and precise duplication of DNA from short sequence to megabase and chromosomal scale. *Cell*. 2024;187(15):3936–52.e3919.
- [28] Hutchison CA, Smith HO, Pfannkuch C, Venter JC. Cell-free cloning using ϕ 29 DNA polymerase. *Proc Natl Acad Sci USA*. 2005;102(48):17332–6.
- [29] Dean FB, Nelson JR, Giesler TL, Lasken RS. Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res*. 2001;11(6):1095–9.
- [30] Fire A, Xu SQ. Rolling replication of short DNA circles. *Proc Natl Acad Sci*. 1995;92(10):4641–5.
- [31] Mohsen MG, Kool ET. The discovery of rolling circle amplification and rolling circle transcription. *Acc Chem Res*. 2016;49(11):2540–50.
- [32] Beyer S, Nickels P, Simmel FC. Periodic DNA nanotemplates synthesized by rolling circle amplification. *Nano Lett*. 2005;5(4):719–22.
- [33] Yao C, Zhang R, Tang J, Yang D. Rolling circle amplification (RCA)-based DNA hydrogel. *Nat Protoc*. 2021;16(12):5460–83.
- [34] Hanif W, Yadav I, Hasan E, Alsulaiman D. Programmable all-DNA hydrogels based on rolling circle and multiprimed chain amplification products. *APL Bioeng*. 2023;7(4):046106.
- [35] You X, Wang J, Guo X, Yang D. Advances in rolling circle amplification (RCA)-based DNA-functional materials for cancer diagnosis and therapy. *Advanced NanoBiomed Research*. 2024;4(3):2300108.
- [36] Gu L, Yan W, Liu L, Wang S, Zhang X, Lyu M. Research progress on rolling circle amplification (RCA)-based biomedical sensing. *Pharmaceuticals*. 2018;11(2):35.
- [37] Ali MM, Li F, Zhang Z, Zhang K, Kang D-K, Ankrum JA, et al. Rolling circle amplification: a versatile tool for chemical biology, materials science and medicine. *Chem Soc Rev*. 2014;43(10):3324.
- [38] Amiram M, Luginbuhl KM, Li X, Feinglos MN, Chilkoti A. Injectable protease-operated depots of glucagon-like peptide-1 provide extended and tunable glucose control. *Proc Natl Acad Sci USA*. 2013;110(8):2792–7.
- [39] Amiram M, Quiroz FG, Callahan DJ, Chilkoti A. A highly parallel method for synthesizing DNA repeats enables the discovery of 'smart' protein polymers. *Nat Mater*. 2011;10(2):141–8.
- [40] Viguera E. Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J*. 2001;20(10):2587–95.
- [41] Koscińska E, Witkos TM, Kozłowska E, Wojciechowska M, Krzyżosiak WJ. Cooperation meets competition in microRNA-mediated DMPK transcript regulation. *Nucleic Acids Res*. 2015;43(19):9500–18.
- [42] Figura G, Koscińska E, Krzyżosiak W. In vitro expansion of CAG, CAA, and mixed CAG/CAA repeats. *Indian J Manag Sci*. 2015;16(8):18741–51.
- [43] Takahashi N, Sasagawa N, Suzuki K, Ishiura S. Synthesis of long trinucleotide repeats in vitro. *Neurosci Lett*. 1999;262(1):45–8.
- [44] Whitfield CJ, Turley AT, Tuite EM, Connolly BA, Pike AR. Enzymatic method for the synthesis of long DNA sequences with multiple repeat units. *Angew Chem*. 2015;127(31):9099–102.
- [45] Little RC. Designer DNA for sensing and nanomaterials. Newcastle University; 2019.
- [46] Yang S, Wijegunawardana D, Sheth U, Veire AM, Salgado JMS, Agrawal M, et al. Aberrant splicing exonizes C9ORF72 repeat expansion in ALS/FTD; 2023.
- [47] Isiktas AU, Eshov A, Yang S, Guo JU. Systematic generation and imaging of tandem repeats reveal base-pairing properties that promote RNA aggregation. *Cell Rep Methods*. 2022;2(11):100334.
- [48] Wang X, Yang X. A simple blocking PCR-based method for the synthesis of high-copy dsDNA tandem repeats. *Small*. 2020;16(43):2003671.
- [49] Liu S, Xiao H, Zhang F, Lu Z, Zhang Y, Deng A, et al. A seamless and iterative DNA assembly method named PS-Brick and its assisted metabolic engineering for threonine and 1-propanol production. *Biotechnol Biofuels*. 2019;12(1):180.
- [50] Casini A, Storch M, Baldwin GS, Ellis T. Bricks and blueprints: methods and standards for DNA assembly. *Nat Rev Mol Cell Biol*. 2015;16(9):568–76.
- [51] Yamazaki K-i, De Mora K, Saitoh K. Biobrick-based 'Quick Gene Assembly' in vitro. *Synthetic Biology*. 2017;2(1):ysx003.
- [52] Weber E, Gruetzner R, Werner S, Engler C, Marillonnet S. Assembly of designer tal effectors by Golden Gate cloning. *PLoS One*. 2011;6(5):e19722.
- [53] Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P. Efficient construction of sequence-specific tal effectors for modulating mammalian transcription. *Nat Biotechnol*. 2011;29(2):149–53.
- [54] Vad-Nielsen J, Lin L, Bolund L, Nielsen AL, Luo Y. Golden Gate assembly of CRISPR gRNA expression array for simultaneously targeting multiple genes. *Cell Mol Life Sci: CMLS*. 2016;73(22):4315–25.
- [55] Tasan I, Su CJ, Enghiad B, Zhang M, Mishra S, Zhao H. Two-color imaging of non-repetitive endogenous loci in human cells. *ACS Synth Biol*. 2020;9:2502–14.
- [56] Zoueiri A, Lende-Dorn B, Galloway KE, Ellis T, Ceroni F. Engineered transcription factor-binding arrays for DNA-based gene expression control in mammalian cells. *Trends Biotechnol*. 2025;43(8):2029–48.
- [57] Meyer DE, Chilkoti A. Genetically encoded synthesis of protein-based polymers with precisely specified molecular weight and sequence by recursive directional ligation: examples from the elastin-like polypeptide system. *Biomacromolecules*. 2002;3(2):357–67.
- [58] Rohilla KJ, Ovington KN, Pater AA, Barton M, Henke AJ, Gagnon KT. Systematic microsatellite repeat expansion cloning and validation. *Hum Genet*. 2020;139(10):1233–46.
- [59] Nelson DW, Connor A, Shen Y, Gilbert RJ. Construction of an elastin-like polypeptide gene in a high copy number plasmid using a modified method of recursive directional ligation. *SynBio*. 2024;2:174–89.
- [60] McDaniel JR, MacKay JA, Quiroz FG, Chilkoti A. Recursive directional ligation by plasmid reconstruction allows rapid and seamless cloning of oligomeric genes. *Biomacromolecules*. 2010;11(4):944–52.
- [61] Spakman D, King GA, Peterman EJG, Wuite GJL. Constructing arrays of nucleosome positioning sequences using Gibson assembly for single-molecule studies. *Sci Rep*. 2020;10(1):9903.

- [62] Deyling JK, Turk EM, Holland NB, Udit AK. Preparing genes for repetitive elastin-like polypeptides using Gibson assembly. In: Protein scaffolds: design, synthesis, and applications. New York: Springer; 2018. p. 265–75.
- [63] Torella JP, Lienert F, Boehm CR, Chen J-H, Way JC, Silver PA. Unique nucleotide sequence-guided assembly of repetitive DNA parts for synthetic biology applications. *Nat Protoc*. 2014; 9:2075–89.
- [64] Guye P, Li Y, Wroblewska L, Duportet X, Weiss R. Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Res*. 2013;41(16):e156.
- [65] Torella JP, Boehm CR, Lienert F, Chen J-H, Way JC, Silver PA. Rapid construction of insulated genetic circuits via synthetic sequence-guided isothermal assembly. *Nucleic Acids Res*. 2014;42(1):681–9.
- [66] Casini A, MacDonald JT, Jonghe JD, Christodoulou G, Freemont PS, Baldwin GS, et al. One-pot DNA construction for synthetic biology: the modular overlap-directed assembly with linkers (modal) strategy. *Nucleic Acids Res*. 2014; 42(1):e7.
- [67] Ratnayake D, Newman M, Lardelli M. Degenerate codon mixing for PCR-based manipulation of highly repetitive sequences. *BMC Res Notes*. 2018;11(1):202.
- [68] Tang NC, Chilkoti A. Combinatorial codon scrambling enables scalable gene synthesis and amplification of repetitive proteins. *Nat Mater*. 2016;15(4):419–24.
- [69] Shurer CR, Head SE, Goudge MC, Paszek MJ. Mucin-coating technologies for protection and reduced aggregation of cellular production systems. *Biotechnol Bioeng*. 2019;116(5): 994–1005.
- [70] Alvisi N, Gutiérrez-Mejía FA, Lokker M, Lin Y-T, de Jong AM, van Delft F, et al. Self-assembly of elastin-like polypeptide brushes on silica surfaces and nanoparticles. *Biomacromolecules*. 2021;22(5):1966–79.
- [71] Davey JA, Goto NK. Tresr: a PCR-compatible DNA sequence design method for engineering proteins containing tandem repeats. *PLoS One*. 2023;18(4):e0281228.
- [72] Hossain A, Lopez E, Halper SM, Cetnar DP, Reis AC, Strickland D, et al. Automated design of thousands of non-repetitive parts for engineering stable genetic systems. *Nat Biotechnol*. 2020;38(12):1466–75.
- [73] Yuan Q, Gao X. Multiplex base- and prime-editing with drive-and-process CRISPR arrays. *Nat Commun*. 2022;13(1):2771.
- [74] Reis AC, Halper SM, Vezeau GE, Cetnar DP, Hossain A, Clauer PR, et al. Simultaneous repression of multiple bacterial genes using non-repetitive extra-long sgRNA arrays. *Nat Biotechnol*. 2019;37(11):1294–301.
- [75] Zhang Y, Wang J, Wang Z, Zhang Y, Shi S, Nielsen J, et al. A gRNA-tRNA array for CRISPR-Cas9 based rapid multiplexed genome editing in *Saccharomyces cerevisiae*. *Nat Commun*. 2019;10(1):1053.
- [76] Xie K, Minkenberg B, Yang Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc Natl Acad Sci*. 2015;112(11):3570–5.
- [77] Campa CC, Weisbach NR, Santinha AJ, Incarnato D, Platt RJ. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nat Methods*. 2019; 16(9):887–93.
- [78] Liao C, Ttofali F, Slotkowski RA, Denny SR, Cecil TD, Leenay RT, et al. Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. *Nat Commun*. 2019;10(1):2948.
- [79] Yuan G, Martin S, Hassan MM, Tuskan GA, Yang X. Para: a new platform for the rapid assembly of gRNA arrays for multiplexed CRISPR technologies. *Cells*. 2022;11(16):2467.
- [80] Zürcher JF, Kleefeldt AA, Funke LFH, Birnbaum J, Fredens J, Grazioli S, et al. Continuous synthesis of *E. coli* genome sections and mb-scale human DNA assembly. *Nature*. 2023; 619(7970):555–62.
- [81] Zhong L, Zhang Q, Lu N, Wang T, Xue X, Qin Z. The conjugation-associated linear-BAC iterative assembling (CALBIA) method for cloning 2.1-Mb human chromosomal DNAs in bacteria. *Cell Res*. 2025.
- [82] Ebersole T, Okamoto Y, Noskov VN, Kouprina N, Kim JH, Leem SH, et al. Rapid generation of long synthetic tandem repeats and its application for analysis in human artificial chromosome formation. *Nucleic Acids Res*. 2005;33(15): e130.
- [83] Noskov VN, Lee NC, Larionov V, Kouprina N. Rapid generation of long tandem DNA repeat arrays by homologous recombination in yeast to study their function in mammalian genomes. *Biol Proced Online*. 2011;13(8):8.
- [84] Lancrey A, Joubert A, Boulé J-B. Locus specific engineering of tandem DNA repeats in the genome of *Saccharomyces cerevisiae* using CRISPR/Cas9 and overlapping oligonucleotides. *Sci Rep*. 2018;8(1):7127.
- [85] Sugiyama Y, Okada S, Daigaku Y, Kusumoto E, Ito T. Strategic targeting of Cas9 nickase induces large segmental duplications. *Cell Genom*. 2024;4(8):100610.
- [86] Takesue H, Okada S, Doi G, Sugiyama Y, Kusumoto E, Ito T. Strategic targeting of Cas9 nickase expands tandem gene arrays. *Cell Genom*. 2025;5(4):100811.
- [87] Liu Y, Zhou J, Liu D, Hu X, Yang L, Song X-R, et al. De novo assembly and delivery of synthetic megabase-scale human DNA into mouse early embryos. *Nat Methods*. 2025.
- [88] Jiang S, Cai Z, Wang Y, Zeng C, Zhang J, Yu W, et al. High plasticity of ribosomal DNA organization in budding yeast. *Cell Rep*. 2024;43(2):113742.
- [89] He B, Ma Y, Tian F, Zhao G-R, Wu Y, Yuan Y-J. YLC-assembly: large DNA assembly via yeast life cycle. *Nucleic Acids Res*. 2023;51(15):8283–92.
- [90] Jiao Y, Li M, He X, Wang Y, Song J, Hu Y, et al. Targeted, programmable, and precise tandem duplication in the mammalian genome. *Genome Res*. 2023;33(5):779–86.
- [91] Mans R, Van Rossum HM, Wijsman M, Backx A, Kuipers NGA, Van Den Broek M, et al. CRISPR/Cas9: a molecular swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res*. 2015; 15(2):fov004.
- [92] Xie K, Yang Y. A multiplexed CRISPR /Cas9 editing system based on the endogenous tRNA processing. In: Methods in molecular biology. New York: Springer New York; 2019. p. 63–73.
- [93] Bao Z, Xiao H, Liang J, Zhang L, Xiong X, Sun N, et al. Homology-integrated CRISPR-Cas (HI-CRISPR) system for one-step multigene disruption in *Saccharomyces cerevisiae*. *ACS Synth Biol*. 2015;4(5):585–94.
- [94] Xing H-L, Dong L, Wang Z-P, Zhang H-Y, Han C-Y, Liu B, et al. A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biol*. 2014;14(1):327.
- [95] Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R, et al. A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol Plant*. 2015;8:1274–84.
- [96] Yin L, Maddison LA, Li M, Kara N, LaFave MC, Varshney GK, et al. Multiplex conditional mutagenesis using transgenic expression of Cas9 and sgRNAs. *Genetics*. 2015;200(2): 431–41.
- [97] Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013;339(6121):819–23.
- [98] Žun G, Doberšek K, Petrovič U. Construction and evaluation of gRNA arrays for multiplex CRISPR-Cas9. *Yeast*. 2023; 40(1):32–41.

- [99] Zuo E, Huo X, Yao X, Hu X, Sun Y, Yin J, et al. CRISPR/Cas9-mediated targeted chromosome elimination. *Genome Biol.* 2017;18(1):224.
- [100] Zhao X, Yang L, Li P, Cheng Z, Jia Y, Luo L, et al. High-accuracy crRNA array assembly strategy for multiplex CRISPR. *Mol Ther Nucleic Acids.* 2025;36(1):102428.
- [101] McCarty NS, Shaw WM, Ellis T, Ledesma-Amaro R. Rapid assembly of gRNA arrays via modular cloning in yeast. *ACS Synth Biol.* 2019;8(4):906–10.
- [102] Deaner M, Holzman A, Alper HS. Modular ligation extension of guide RNA operons (LEGO) for multiplexed dCas9 regulation of metabolic pathways in *Saccharomyces cerevisiae*. *Biotechnol J.* 2018;13(9):1700582.
- [103] Cress BF, Toparlak ÖD, Guleria S, Lebovich M, Stieglitz JT, Englaender JA, et al. Crispathbrick: modular combinatorial assembly of Type II-A CRISPR arrays for dCas9-mediated multiplex transcriptional repression in *E. coli*. *ACS Synth Biol.* 2015;4(9):987–1000.
- [104] Chen Y, Hysolli E, Chen A, Casper S, Liu S, Yang K, et al. Multiplex base editing to convert TAG into TAA codons in the human genome. *Nat Commun.* 2022;13(1):4482.
- [105] Basu J. Rapid creation of BAC-based human artificial chromosome vectors by transposition with synthetic alpha-satellite arrays. *Nucleic Acids Res.* 2005;33(2):587–96.
- [106] Pesenti E, Kouprina N, Liskovych M, Aurich-Costa J, Larionov V, Masumoto H, et al. Generation of a synthetic human chromosome with two centromeric domains for advanced epigenetic engineering studies. *ACS Synth Biol.* 2018;7(4):1116–30.
- [107] Kouprina N, Samoshkin A, Erliandri I, Nakano M, Lee H-S, Fu H, et al. Organization of synthetic alphoid DNA array in human artificial chromosome (HAC) with a conditional centromere. *ACS Synth Biol.* 2012;1(12):590–601.
- [108] Kato N, Lam E. Detection of chromosomes tagged with green fluorescent protein in live *Arabidopsis thaliana* plants. *Genome Biol.* 2001;2(research0045.0041):research0045.1.
- [109] Teo CH, Lermontova I, Houben A, Mette MF, Schubert I. De novo generation of plant centromeres at tandem repeats. *Chromosoma.* 2013;122(3):233–41.
- [110] Zhang H, Phan BH, Wang K, Artelt BJ, Jiang J, Parrott WA, et al. Stable integration of an engineered megabase repeat array into the maize genome. *Plant J.* 2012;70(2):357–65.
- [111] Dawe RK, Gent JI, Zeng Y, Zhang H, Fu F-F, Swentowsky KW, et al. Synthetic maize centromeres transmit chromosomes across generations. *Nat Plants.* 2023;9(3):433–41.
- [112] WFA den Dunnen. (2017) Trinucleotide repeat disorders. *Handb Clin Neurol.* 2018;145:383–91.
- [113] Osborne RJ, Thornton CA. Cell-free cloning of highly expanded CTG repeats by amplification of dimerized expanded repeats. *Nucleic Acids Res.* 2007;36(4):e24.
- [114] Malankhanova T, Sorokin M, Medvedev S, Zakian S, Malakhova A. Introducing an expanded trinucleotide repeat tract into the human genome for Huntington's disease modeling in vitro. *CP Human Genetics.* 2020;106(1):e100.
- [115] Mizielinska S, Grönke S, Niccoli T, Ridler CE, Clayton EL, Devoy A, et al. C9orf72 repeat expansions cause neurodegeneration in drosophila through arginine-rich proteins. *Science.* 2014;345(6201):1192–4.
- [116] Liu Z-W, Liu J, Liu F, Zhong X. Depositing centromere repeats induces heritable intragenic heterochromatin establishment and spreading in *Arabidopsis*. *Nucleic Acids Res.* 2023;51(12):6039–54.
- [117] Deng S, Chua NH. Inverted-repeat RNAs targeting FT intronic regions promote FT expression in *Arabidopsis*. *Plant Cell Physiol.* 2015;56(8):1667–78.
- [118] Belele CL, Sidorenko L, Stam M, Bader R, Arteaga-Vazquez MA, Chandler VL. Specific tandem repeats are sufficient for paramutation-induced trans-generational silencing. *PLoS Genet.* 2013;9(10):e1003773.
- [119] Zhang W, Zhao G, Luo Z, Lin Y, Wang L, Guo Y, et al. Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science.* 2017;355(6329):eaaf3981.
- [120] Jeon J, Subramani SV, Lee KZ, Jiang B, Zhang F. Microbial synthesis of high-molecular-weight, highly repetitive protein polymers. *Indian J Manag Sci.* 2023;24(7):6416.
- [121] Dinjaski N, Huang W, Kaplan DL, Nilsson BL, Doran TM. Recursive directional ligation approach for cloning recombinant spider silks. *Methods Mol Biol.* 2018;1777:181–92.
- [122] Kim E, Jeon J, Zhu Y, Hoppe ED, Jun Y-S, Genin GM, et al. A biosynthetic hybrid spidroin-amyloid-mussel-foot protein for underwater adhesion on diverse surfaces. *ACS Appl Mater Interfaces.* 2021;13(41):48457–68.
- [123] Jung H, Pena-Francesch A, Saadat A, Sebastian A, Kim DH, Hamilton RF, et al. Molecular tandem repeat strategy for elucidating mechanical properties of high-strength proteins. *Proc Natl Acad Sci.* 2016;113(23):6478–83.
- [124] Dai B, Sargent CJ, Gui X, Liu C, Zhang F. Fibril self-assembly of amyloid-spider silk block polypeptides. *Biomacromolecules.* 2019;20(5):2015–23.
- [125] Girotti A, Fernández-Colino A, López IM, Rodríguez-Cabello JC, Arias FJ. Elastin-like recombinamers: biosynthetic strategies and biotechnological applications. *Biotechnol J.* 2011;6(10):1174–86.
- [126] Chang MP, Huang W, Mai DJ. Monomer-scale design of functional protein polymers using consensus repeat sequences. *J Polym Sci.* 2021;59(22):2644–64.
- [127] Riziotis IG, Lamprou P, Papachristou E, Mantsou A, Karolidis G, Papi R, et al. De novo synthesis of elastin-like polypeptides (ELPs): an applied overview on the current experimental techniques. *ACS Biomater Sci Eng.* 2021;7(11):5064–77.
- [128] Li C, Wang Y, Li P-F, Fu Q. Construction of rolling circle amplification products-based pure nucleic acid nanostructures for biomedical applications. *Acta Biomater.* 2023;160:1–13.
- [129] Petka WA, Harden JL, McGrath KP, Wirtz D, Tirrell DA. Reversible hydrogels from self-assembling artificial proteins. *Science.* 1998;281(5375):389–92.
- [130] Li F, Tang J, Geng J, Luo D, Yang D. Polymeric DNA hydrogel: design, synthesis and applications. *Prog Polym Sci.* 2019;98:101163.
- [131] Kosuri S, Church GM. Large-scale de novo DNA synthesis: technologies and applications. *Nat Methods.* 2014;11(5):499–507.
- [132] Wells RD, Dere R, Hebert ML, Napierala M, Son LS. Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res.* 2005;33(12):3785–98.
- [133] Blackwood JK, Okely EA, Zahra R, Eykelenboom JK, Leach DRF. DNA tandem repeat instability in the *Escherichia coli* chromosome is stimulated by mismatch repair at an adjacent CAG-CTG trinucleotide repeat. *Proc Natl Acad Sci USA.* 2010;107(52):22582–6.
- [134] Sheng Y, Mancino V, Birren B. Transformation of *Escherichia coli* with large DNA molecules by electroporation. *Nucleic Acids Res.* 1995;23(11):1990–6.
- [135] Siguret V, Ribba AS, Chérel G, Meyer D, Piétu G. Effect of plasmid size on transformation efficiency by electroporation of *Escherichia coli* DH5 alpha. *Biotechniques.* 1994;16(3):422–6.

How to cite this article: Ma S, Cui Y, Wu Y. Technologies for engineering repetitive DNA. *Quantitative Biology.* 2026;e70040. <https://doi.org/10.1002/qub2.70040>