

Transformation-associated recombination (TAR) cloning and its applications for gene function; genome architecture and evolution; biotechnology and biomedicine

Natalay Kouprina¹ and Vladimir Larionov¹

¹Developmental Therapeutics Branch, National Cancer Institute, Bethesda, MD 20892, USA

Correspondence to: Natalay Kouprina, **email:** kouprinn@mail.nih.gov

Keywords: transformation-associated recombination; TAR; microbes; biomedicine; biotechnology

Received: November 16, 2023

Accepted: November 27, 2023

Published: December 22, 2023

Copyright: © 2023 Kouprina and Larionov. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#) (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Transformation-associated recombination (TAR) cloning represents a unique tool to selectively and efficiently recover a given chromosomal segment up to several hundred kb in length from complex genomes (such as animals and plants) and simple genomes (such as bacteria and viruses). The technique exploits a high level of homologous recombination in the yeast *Saccharomyces cerevisiae*. In this review, we summarize multiple applications of the pioneering TAR cloning technique, developed previously for complex genomes, for functional, evolutionary, and structural studies, and extended the modified TAR versions to isolate biosynthetic gene clusters (BGCs) from microbes, which are the major source of pharmacological agents and industrial compounds, and to engineer synthetic viruses with novel properties to design a new generation of vaccines. TAR cloning was adapted as a reliable method for the assembly of synthetic microbe genomes for fundamental research. In this review, we also discuss how the TAR cloning in combination with HAC (human artificial chromosome)- and CRISPR-based technologies may contribute to the future.

INTRODUCTION

TAR cloning is the method that enables selective, rapid, and efficient capture of genes or chromosomal regions of choice from total genomic DNA of organisms ranging from microbes and viruses to plants, and animals as circular YAC/BAC (yeast artificial chromosome/bacterial artificial chromosome) molecules which can propagate in yeast as well as in bacterial cells. TAR cloning exploits the efficient non-meiotic homologous recombination of the yeast *Saccharomyces cerevisiae* [1–4]. A desired region of choice is captured from genomic DNA using TAR vector containing either two unique sequences (hooks) homologous to the 5'- and 3'-ends of the target region (see chapter 3.1) or one unique hook and a common repeat (see chapter 3.3) or using a counter selectable marker (see chapter 3.2) [5–9].

For gene functional studies and potentially for gene therapy, it is important that TAR cloning isolates full-length genes that contain coding (exons) as well as

non-coding (introns) regions, including their regulatory elements, that can reproduce a physiological gene expression. For the past two decades, TAR cloning has become a reliable method with multiple applications in functional and structural genomics, characterization of chromosomal rearrangements, evolutionary studies, and engineering synthetic microbial genomes and viruses with contribution in biotechnology and biomedicine (Figure 1).

Recently TAR cloning has been adapted and widely used to clone BGCs from microbes and environmental DNA samples (see chapter 4.4) [10–12]. Typically, natural products (NPs) BGCs are produced by co-expression of multiple genes involved in regulation, biosynthesis, transport, and resistance to the drug(s) [13]. A direct capture of intact BGCs by TAR cloning has proven to be important for discovery of novel NPs to understand their biosynthesis and molecular mechanisms [11].

The previous methods of isolation of mammalian genes and BGCs relied on the construction of YAC, BAC, or cosmid libraries followed by PCR screening of

thousands of transformants to identify the ones of your interest. In such libraries, the frequency of region-positive clones was less than 0.0003% [14]. If we compare, for example, the efficiency of TAR cloning of a human gene (~35%), with the frequency of gene recovery from YAC or BAC libraries (0.0003%), TAR cloning efficiency is higher than ~100,000 times. Moreover, from the libraries a desired region is often recovered as a set of DNA fragments. In such cases, they need to be pieced together to reassemble the entire region that makes this method technically challenging and time-consuming.

An alternative technique, Cas9-Assisted Targeting of Chromosome segments (CATCH), has been developed and applied to recover BGCs from microorganisms directly in *E. coli* [15]. This method is based on the *in-vitro* cleavage of target DNA from a native bacterial chromosome using RNA-guided Cas9 nuclease and subsequent ligation into a cloning vector via Gibson assembly [16]. In principle, this method may be considered as an alternative to TAR cloning when the goal is to isolate BGC from the individual microbe. However, it has never been applied to isolate BGCs from environmental samples (e.g., from soil or gut microbiota) that typically contain thousands of bacterial species or from complex genomes such as animals and plants.

In this review, we focus on the main TAR cloning parameters, TAR cloning variations, and multiple applications, i.e., isolation of intact genes and gene clusters for functional, evolutionary, and structural studies and capture of intact BGCs for heterologous expression and natural product discovery. In addition, we describe the application of TAR for cloning and assembly of Mb-scale genome genomes. Finally, we discuss the progress in applying the TAR technology to construct HACs for functional and structural studies of human kinetochore. The benefit of combining the TAR cloning technology

with the HAC gene delivery system for therapeutic gene delivery and expression studies is also discussed.

PARAMETERS OF TAR CLONING

Features of TAR vector

A TAR vector contains a YAC cassette (a yeast selectable marker and a yeast centromere) for proper propagation, segregation, and selection of the cloned material in yeast, and a BAC cassette (a bacterial origin of replication and a bacterial selectable marker) that allows TAR isolates to propagate in bacterial cells. TAR vectors used for isolation of DNA fragments from complex genomes does not contain a yeast origin of replication and therefore cannot propagate in yeast cells producing no background. In this case, the TAR cloning method requires a presence of at least one autonomously replicating sequence (*ARS*) that can function as yeast origin of replication in the cloned genomic DNA fragment. Potential *ARS*-like sequences have a 17-bp *ARS* core consensus, WWWWT TAYRTTTWGTT, in which W = A or T, Y = T or C, and R = A or G [17]. Such sequences occur at a frequency of approximately one per 20–40 kb in all eukaryotic genomes thus far examined [18]. This suggests that TAR cloning can readily isolate most chromosomal regions with a vector that lacks an *ARS* because it relies on the acquisition of an *ARS* element from the targeted chromosomal DNA fragment. Some genomic regions in microbes also contain *ARS*-like sequences and may be captured by TAR vector lacking *ARS*. So far, several hundred different genomic regions were isolated using TAR vectors without *ARS* [19]. Features of TAR vector used for isolation of DNA fragments from simple genomes, such as bacterial or viral, are described in chapter 3.2.

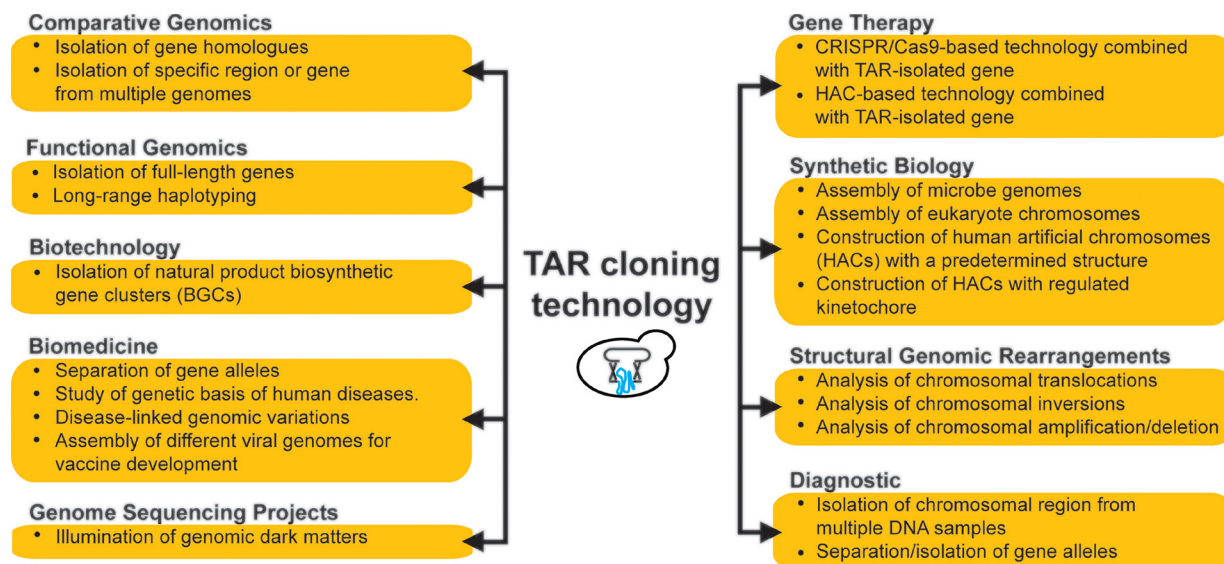


Figure 1: Applications of the TAR cloning technology.

Size and divergence requirements for the targeting sequences (hooks) in TAR vectors

A genomic region of choice is targeted by TAR vector containing two unique guiding sequences (hooks) homologous to the 5' and 3' flanks of the target. The minimal size of these hooks may be as short as 60 bp, though longer hooks can also be used [20]. Hooks should be unique sequences, which can be assured by blasting candidate sequences against a genome reference sequence (<http://genome.ucsc.edu/cgi-bin/hgBlat>). The hooks are cloned into TAR vector in the same orientation as they occur in the targeted genome. Prior to TAR cloning experiments, a vector is linearized between the hooks to make them highly recombinogenic (Figure 2).

It was found that a divergence of up to 15% between the targeting hooks and the target genomic sequences does not prevent isolation of specific genes or regions from different organisms using TAR vector with the hooks developed from the human genome sequence [21]. The yield of region-positive clones with TAR vector containing the unique sequence hooks is comparable to that when the hooks are divergent by 15%. Such robustness with respect to DNA sequence enables TAR cloning to be applied to isolate gene orthologs and paralogs.

The size of TAR-isolated and TAR-assembled genomes and genomic regions

So far, TAR cloning allowed isolating genomic fragments up to 300 kb in size, which is sufficient for successful cloning of most mammalian genes and microbial gene clusters. However, this limit is not absolute, as some YAC libraries contain YACs ranging from 430 kb to 1,200 kb [22]. To TAR-isolate a genomic DNA fragment with the size bigger than 50 kb, DNA should be protected from shearing by preparing a high-molecular weight DNA in agarose blocks [9]. Note that after isolation of a region of choice in yeast, if necessary, this region may be easily modified using homologous recombination in yeast [23, 24] or transferred directly to another host, for example bacterial cells.

As described, there is almost no size limitation for the genomic fragment to propagate in yeast cells. Assembly of individual chromosomes or whole genomes by transformation-associated recombination in yeast was successfully accomplished for species such as *M. pneumoniae* (0.8 Mb) [25], *Acholeplasma laidlawii* (1.5 Mb) [26], *Prochlorococcus marinus MED4* (1.6 Mb) [27], and eukaryotic algal *Phaeodactylum tricorutum* (27.4 Mb) [28]. In the latter case, two individual chromosomes were assembled using TAR. Recently, several groups assembled the entire viral genomes using a TAR platform [29, 30].

Anticipated results

Once TAR vector with its specific hooks is constructed and genomic DNA is prepared, the entire procedure to isolate a gene or a region of choice takes approximately three weeks. With 0.5–1 mg of TAR vector, 1–2 mg of genomic DNA and 1×10^8 yeast spheroplasts, the yield of transformants varies from 10 to 150 colonies on one Petri dish. The yield of region-positive clones from mammalian or individual microbial genomes or environmental DNA samples varies from 35% to 95% [20]. In the case of complex genomes, it is preferable that genomic DNA is pre-treated before spheroplast transformation by CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats) that are recognized by Cas9 nuclease [31–35] that is designed in such a way to cut near the targeted 5' and 3' end sequences making them highly recombinogenic [9, 36]. In this case, for example, the yield of region-positive clones isolated from complex genomes increases from 1% up to 35% [36].

VARIATIONS OF TAR CLONING APPROACHES

CRISPR/Cas9-mediated TAR cloning using a vector with two unique targeting hooks

For the past decade, the cost, quality, and efficacy of the TAR cloning method has been improved significantly by rigorous testing for the accuracy of isolation of loci from different organisms, shifting the technology from unusual to routine. The updated TAR protocol does not require significant experience with yeast, because screening of approximately 20–30 yeast transformants is typically enough to find a clone containing the region of choice [9]. Figure 2 shows a step-by-step general scheme of TAR cloning of a region of choice from genomic DNA using TAR vector containing two targeting hooks [9]. Genomic DNA is pre-treated with the specifically designed programmable endonuclease CRISPR/Cas9 (Step 1 and Step 2 in Figure 2) that creates double-strand breaks (DSBs) bracketing the target genomic DNA sequence leading to increase of region-positive clones 35 times [9, 36]. Step 3 and Step 4 in Figure 2 include co-transformation of TAR vector linearized between the hooks and genomic DNA into competent yeast spheroplasts followed by recombination between the target hook sequences in the vector and targeted 3' and 5' ends of a genomic segment and rescue of a region of choice as a circular YAC/BAC molecule. Selection of region-positive clones is carried out by PCR using diagnostic primers. The TAR-cloned material may be directly moved from yeast cells to bacterial cells (Step 5) that facilitates BAC DNA isolation (Step 6 in Figure 2) for a further analysis.

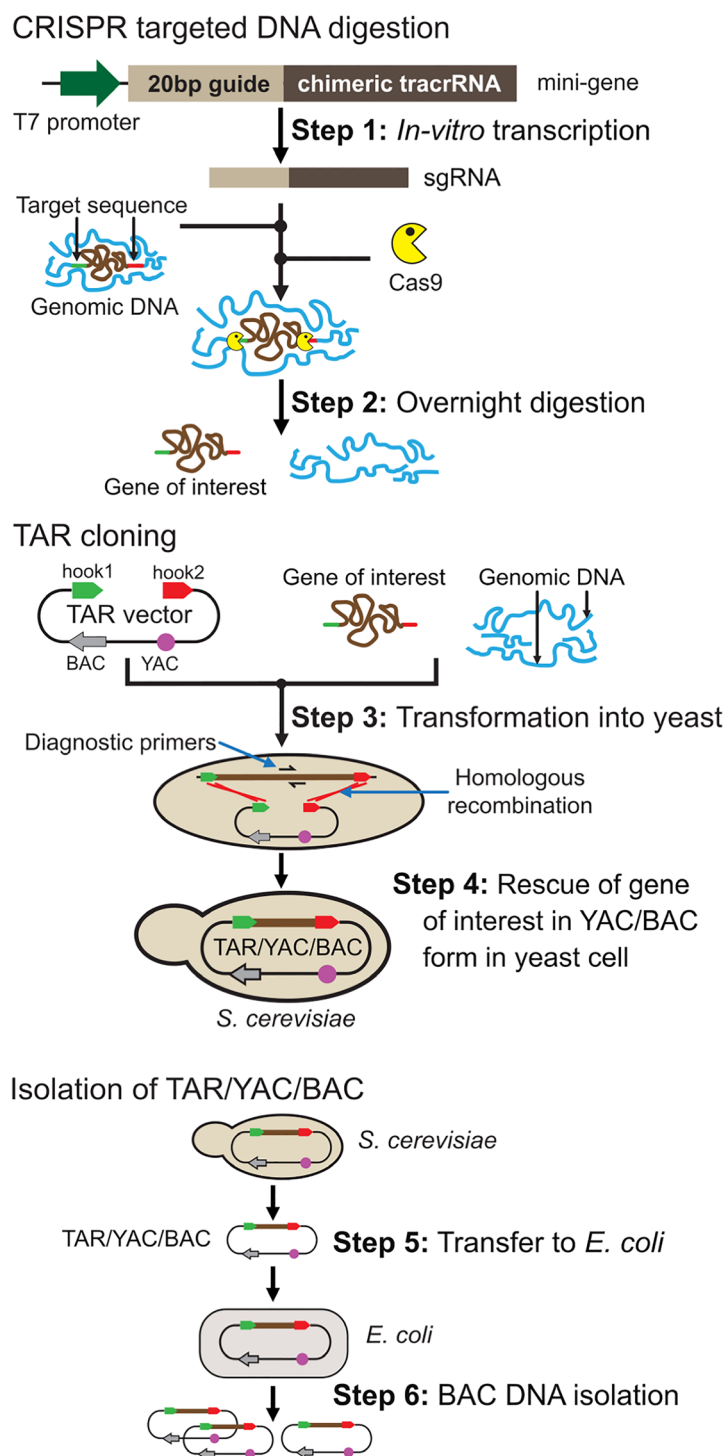


Figure 2: The experimental consecutive steps of TAR isolation of a gene/region of choice from the total genome. Preparation of genomic DNA for TAR cloning. **Step 1:** Genomic DNA is cleaved at the positions close to the 5' and 3' ends of a region of choice by Cas9-gRNA complexes. The CRISPR minigene may be generated by PCR or artificially synthesized. **Step 2:** Overnight digestion of the mixture containing gRNAs, Cas9 and genomic DNA. A TAR cloning of a region of choice from genomic DNA. A TAR vector contains YAC (in purple) and BAC (in grey) cassettes and two unique targeting sequences (hook1 in green and hook2 in red) homologous to the 5' (in green) and 3' (in red) ends of a region of choice. Before cloning, TAR vector is linearized between the hooks to expose targeting hooks. **Step 3:** CRISPR/Cas9-treated genomic DNA and a linearized TAR vector are co-transformed into yeast *Saccharomyces cerevisiae* cells. **Step 4:** Recombination between targeting hooks in the vector and the target sequences in the genomic DNA region leads to rescue of the genomic region as a circular YAC/BAC molecule. (c) Isolation of YAC molecules containing a region of choice from yeast cells [148]. **Step 5:** Transferring TAR-isolated YAC/BAC molecules from yeast cells to bacterial cells. **Step 6:** BAC DNA is isolated for further sequencing and functional analyses.

TAR cloning using a counter-selectable marker

For chromosomal GC-rich regions such as centromeres and telomeres of mammalian genomes or simple genomes like microbes and viruses that have few or no *ARS*-like sequences, the TAR method described in chapter 3.1 is not applicable. To overcome that limitation, a substantially different version of the TAR method has been developed [10–12]. A general scheme of this version using a counter-selectable marker is shown in Figure 3. The TAR vector contains a yeast *ARS* element and a negative-selectable marker *URA3* that represents a hybrid gene containing the open reading frame of the *S. cerevisiae URA3* gene and the promoter of the *S. pombe ADHI* gene (Figure 3A), which has strict spacing requirements for its function, i.e., the distance between the TATA element of the promoter and the transcription initiation site must be no more than 130 bp [37, 38]. Accordingly, the combined length of the targeting hooks in the TAR vector should not exceed 130 bp. The hooks are placed between the TATA box and the transcription initiation site of *URA3*. As a result of such a design, an insertion of any genomic fragment between the hooks due to homologous recombination between the hooks and the target genomic sequences (Figure 3B) leads to the inactivation of *URA3* expression (5-FOA^R) (Figure 3C). Thus, because yeast cells expressing *URA3* are sensitive to 5-fluoroorotic acid (5-FOA^S), the proper TAR clones containing a region of choice should be selected against the background (5-FOA^R) arising from vector recircularization (5-FOA^S) (Figure 3C). It is worth noting that sometimes TAR cloning of GC-rich bacterial DNA regions can be challenging even with the TAR *ARS*-containing vector. In these cases, only the fragments of approximately 100 kb or bigger are typically recovered [39].

Radial TAR cloning using a vector with a unique targeting hook and a common repeat

For many animal and plant genomes, only limited sequence information is still available. Therefore, TAR cloning with the vector containing two targeting hooks could be infeasible. To circumvent this limitation, another version of the TAR method was developed [7]. This version, branded as a radial TAR cloning, uses a vector as described in chapter 3.1 except that one specific hook has a unique sequence while another one has a common repeat sequence (i.e., *Alu* for the primate genomes or *BI* for the mouse genome) (Figure 4). Such a vector construction makes possible to isolate a region of choice as a set of nested overlapping fragments of different size that extend from the unique hook to the different recombination sites of a given repeat (*Alu1*, *Alu3* and *Alu5*) (Figure 4; above). By changing orientation of the unique targeting hook (from 3' end

to 5' end), it becomes possible to isolate overlapping genomic regions that extend from the unique hook to recombination sites of a repeat located on the opposite side along the chromosome (*Alu7*, *Alu8* and *Alu10*) (Figure 4; below). The size of clones obtained by radial TAR cloning varies from 30 kb to 300 kb, reflecting the frequency and position of a repeat [7, 40–43]. It is worth noting that the yield of region-positive clones for radial TAR cloning is comparable to the yield obtained with the vector containing two targeting hooks. Radial TAR cloning has been applied to close the gaps in the human genome sequence and to isolate several specific regions from human and mouse genomes [7, 41, 43–46].

APPLICATIONS OF TAR CLONING

Functional genomics: isolation of full-size single-copy genes from complex genomes

Whereas cDNA clones are still widely used for gene expression [47–50], the full-size genes containing all exons, introns, and flanking regulatory elements become preferable because the scientific and especially biomedical communities show a keen interest in the mechanism regulating gene or gene cluster expression by means of alternative splicing, alternative promoter-enhancer usage, expression of non-coding RNAs from intronic regions, and 3D genome folding. TAR cloning of individual genes, containing all the necessary cis regulatory regions, provides a unique material for functional, structural (chapter 4.2), and population studies; for comparative genomics (chapter 4.3); long-range haplotyping (chapter 4.2); for biotechnology (chapter 4.4) and biomedicine (chapter 4.2). In addition, the TAR cloning technology may assist in designing diagnostics for genomic disorders caused by chromosomal rearrangements.

Over the past 25 years, TAR cloning has been used to isolate hundreds of full-size genes and gene clusters from genomes of humans, nonhuman primates, mice, and microbes [8, 20]. For example, functional analysis of several human genes, including 84 kb and 90 kb breast cancer genes *BRCA1* and *BRCA2* [6, 51], the 50 kb 3' hypoxanthine phosphoribosyltransferase (*HPRT*) gene [7] that is mutated in Lesch-Nyhan syndrome, the 80 kb tumor suppressor gene *KAI1* [8], the 60 kb *NBS1* gene that is mutated in Nijmegen breakage syndrome, and the 30 kb *VHL* gene that is mutated in von Hippel–Lindau syndrome [8], demonstrated a high fidelity of the TAR-cloned genomic material. Accordingly, TAR-isolated genes were successfully used in transgenesis. In one example, a transgenic mouse carrying the entire 50 kb human *TERT* locus [52] was used to show that *in vivo* expression of human and mouse *TERT* genes differ significantly, raising awareness about the use of mouse models for human cancer and aging [53].

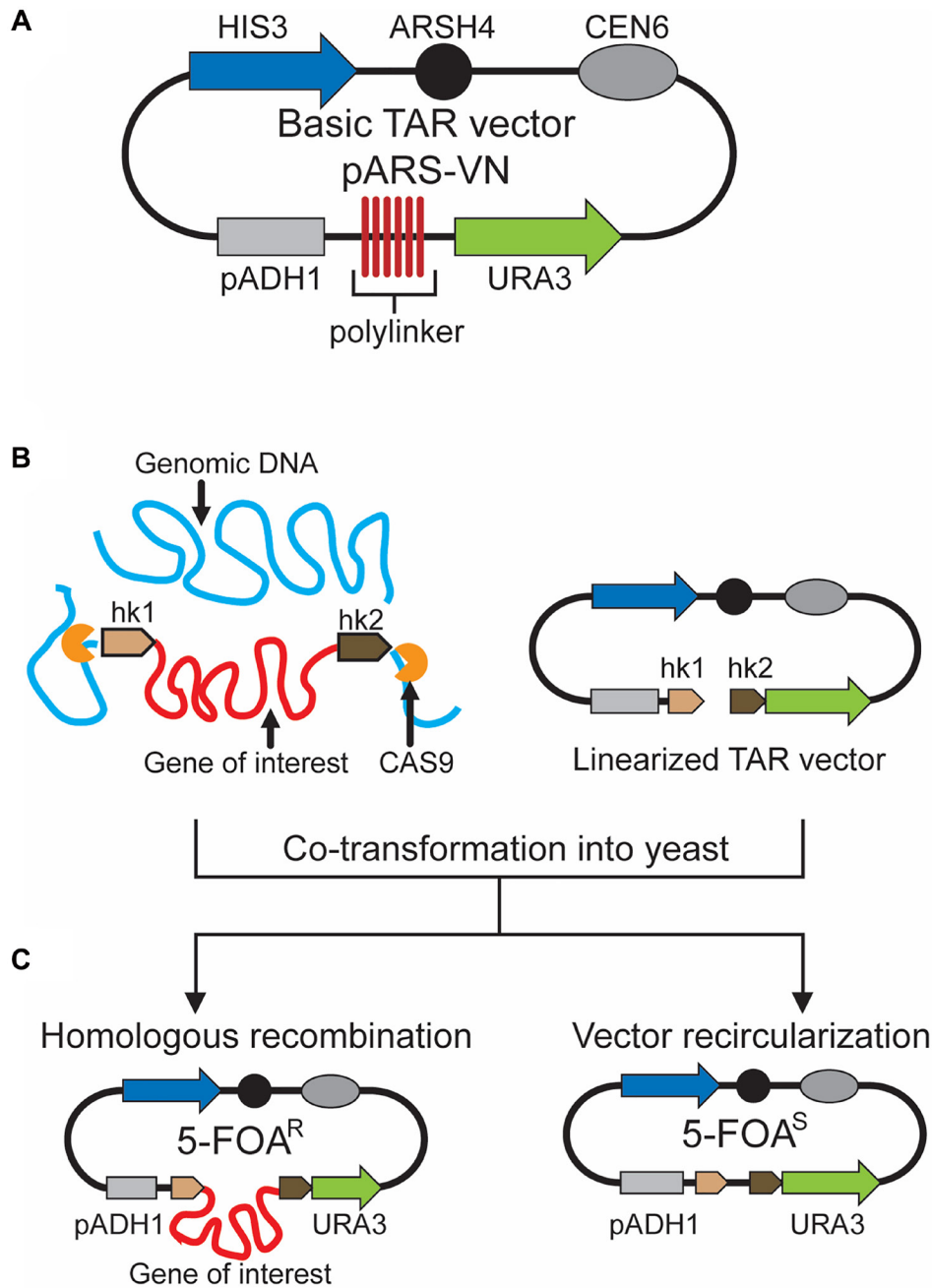


Figure 3: TAR cloning of genomic regions lacking ARS sequences from microbial and viral genomes. (A) Scheme of the basic TAR vector pARS-VN. The plasmid carries a yeast positive-selectable marker (*HIS3*) (in blue), a yeast centromere (*CEN6*) (in dark grey), a negative-selectable marker (*URA3*) (in green), a pADH1 promoter of the *ADH1* gene (in light grey), a yeast origin of replication (*ARSH4*) (in black), and a polylinker (a multiple cloning site) (in red). *URA3* is a hybrid gene containing the open reading frame of the *S. cerevisiae URA3* gene and the promoter of the *S. pombe pADH1* gene, which only tolerates insertion of up to 130 bp of sequence between the TATA box and the transcription initiation site. The combined length of the targeting hooks should be less than 130 bp in order not to disrupt *URA3* expression. Such configuration allows selection of TAR cloning events rather than vector recircularization. (B, C) A diagram of TAR cloning of a region of choice (in red) from genomic DNA (in blue). A TAR vector contains two unique targeting hooks (hk1 in beige and hk2 in brown) homologous to the 5' and 3' ends of a region of choice. The TAR vector DNA is linearized between the hooks to expose targeting sequences. If necessary, genomic DNA may be treated by CRISPR-Cas9 endonuclease before yeast transformation [36] to increase the yield of region-positive clones. Homologous recombination between the hooks and the target genomic sequence leads to insertion of the cloned material between the TATA box and the transcription initiation site ATG, which abolishes *URA3* expression. Such clones can, therefore, be selected by their ability to grow on media containing 5-FOA (5-FOA^R). Recircularization of TAR vector alone due to nonhomologous end-joining leads to reconstruction of the counter-selectable *URA3* marker. In this case, the distance between the TATA element and the transcription initiation site ATG is not changed, and the *URA3* marker is normally expressed. Therefore, they do not grow on media containing 5-FOA (5-FOA^S). (R = resistant, S = sensitive to 5-FOA).

A substantial progress in gene functional studies was made upon combining the TAR-isolated full-size genes with the HAC-based gene delivery and expression vectors (chapter 4.7) [54, 55]. For example, the entire human *HPRT* locus was TAR-isolated as a 100 kb YAC/BAC clone [7], loaded into the HAC vector and then shown to complement the genetic defect of Hprt-deficient hamster CHO cells [56]. The examples of correction of genetic deficiencies in human patient-derived cells include the TAR-isolated genomic copies of *NBS1*, *BRCA1*, *VHL*, and *PKD1* genes loaded into the HAC that allows expression of the genes in target cells under conditions that recapitulate the physiological regulation of endogenous loci [57, 58]. More examples, indicating the accuracy of TAR cloning, include TAR-isolation and long-read sequencing of the 14 rDNA gene copies covering ~0.82 Mb of the human chromosome 21 rDNA cluster that enabled the accurate reconstruction of a high-quality 44,838 bp reference sequence [59]. TAR cloning and sequencing of the entire rDNA array end-to-end, including proximal and distal junction sequences, from the human chromosome 22 facilitated the reconstruction of the entire NOR (nucleolar organizer region) [60].

Genetic basis of human diseases: separation of alleles and long-range molecular haplotyping

The word “haplotype” is derived from the word “haploid”, which describes cells with only one set of chromosomes, and from the word “genotype, which refers to the genetic markers of the organism. A haplotype is a group of genes within the organism that is inherited together from a single parent. Each haplotype has a frequency, which represents the proportion of chromosomes with the adjacent markers in the population. The haplotype frequency, a measure of the coordinated distribution of adjacent markers in the population, represents the correlation between those markers during inheritance.

In principle, the haplotypes or individual homologous chromosomes may be separated by pedigree analysis. However, this approach is laborious and restricted by the need to collect DNA samples from family members or different population groups. Other approaches such as microdissection of chromosomes or amplification of spermatocyte DNA are also time-consuming and labor-intensive making them unacceptable when analyzing many individuals. Modern computational methods do allow for

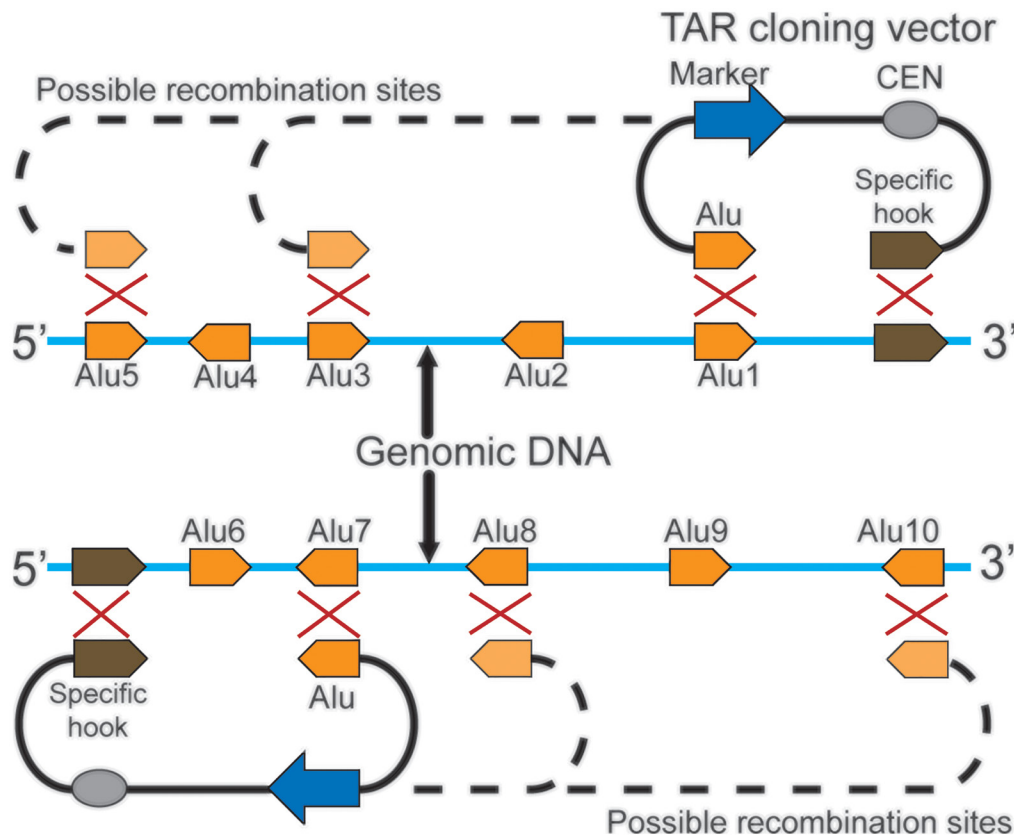


Figure 4: Isolation of a genomic region of choice by radial TAR cloning. The possible positions and orientations of the *Alu* sequences (in orange) in the genomic DNA are shown (from *Alu1* to *Alu10*). A unique hook in the TAR vector is marked in brown. TAR vector contains *CEN* (in grey) corresponds to the yeast chromosome VI centromere while a marker (in blue) is a selectable marker *HIS3*. Recombination during transformation between the hooks and genomic DNA leads to formation of the circular overlapping YAC molecules that extend from the unique hook to various *Alu* positions (*Alu1* to *Alu5*). Changing an orientation of the specific hook in the TAR vector leads to cloning of the genomic DNA in the opposite orientation (*Alu6* to *Alu10*).

haplotype analysis of samples of unrelated individuals [61–63]. However, all the approaches listed above have limitations, especially in resolving the phase of paternal and maternal chromosomes.

TAR cloning represents a simple and reliable method used to resolve the haplotype characterization problem. Because recombination between the targeting hooks in the TAR vector and the homologous target sequences in the genome occurs at equal frequencies in both chromosomes, the parental alleles of a gene from multiple DNA samples can be simultaneously isolated in a single TAR cloning experiment (Figure 5). A representative example of application of TAR cloning for such a purpose is the separation of alleles of the 50 kb human *TERT* gene [64]. The *TERT* gene contains four VNTR (a variable number of tandem repeats) blocks, two of them located in intron 2 and two others in intron 6. To identify the parental *TERT* alleles, VNTR sequences in TAR isolates were examined and showed a specific allele-identifying combination of

microsatellites at each of the polymorphic sites. Further sequencing of individual TAR isolates and analysis of segregation of these VNTRs in families revealed that all of them followed a Mendelian inheritance pattern [64]. Thus, TAR cloning allows separation of the haplotypes in individuals and has a potential to identify haplotypes that may contribute to disease(s).

More impressive that TAR cloning is suitable for large-scale analysis of long-range haplotypes in multiple, inherently heterozygous, individuals. An example of reconstructing long-range haplotypes in the cluster of the *SPANX-A/D* gene sub-family, located within a 750 kb region at Xq27-q28 that is presumable involved in the hereditary prostate cancer locus HPCX1 [65, 66], is TAR isolation of individual *SPANX* genes. The *SPANX-A/D* gene sub-family consists of five genes: *SPANX-C*, *SPANX-B*, *SPANX-A1*, *SPANX-A2*, and *SPANX-D* [67–69] (Figure 6A), with *SPANX-C* and *SPANX-D* genes separated by approximately 500 kb. Note that

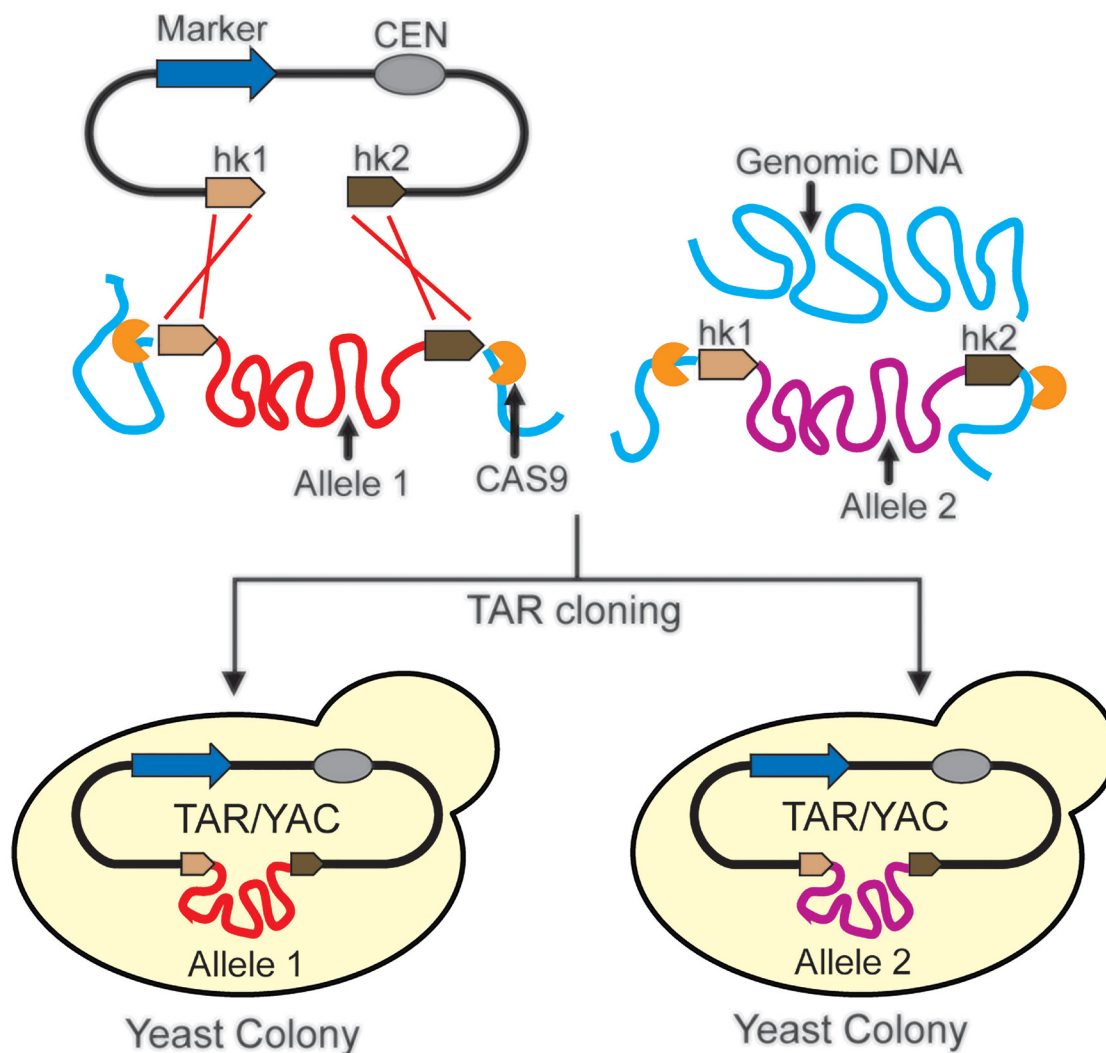


Figure 5: Separation of the alleles of the gene by TAR cloning. TAR vector contains a yeast selectable marker (in blue), a yeast centromere (*CEN*) and two unique hook sequences (hk1 in beige and hk2 in brown). TAR cloning results in multiple gene-positive transformants. Therefore, different alleles of a gene become separated into different yeast cells, allowing for further independent analysis.

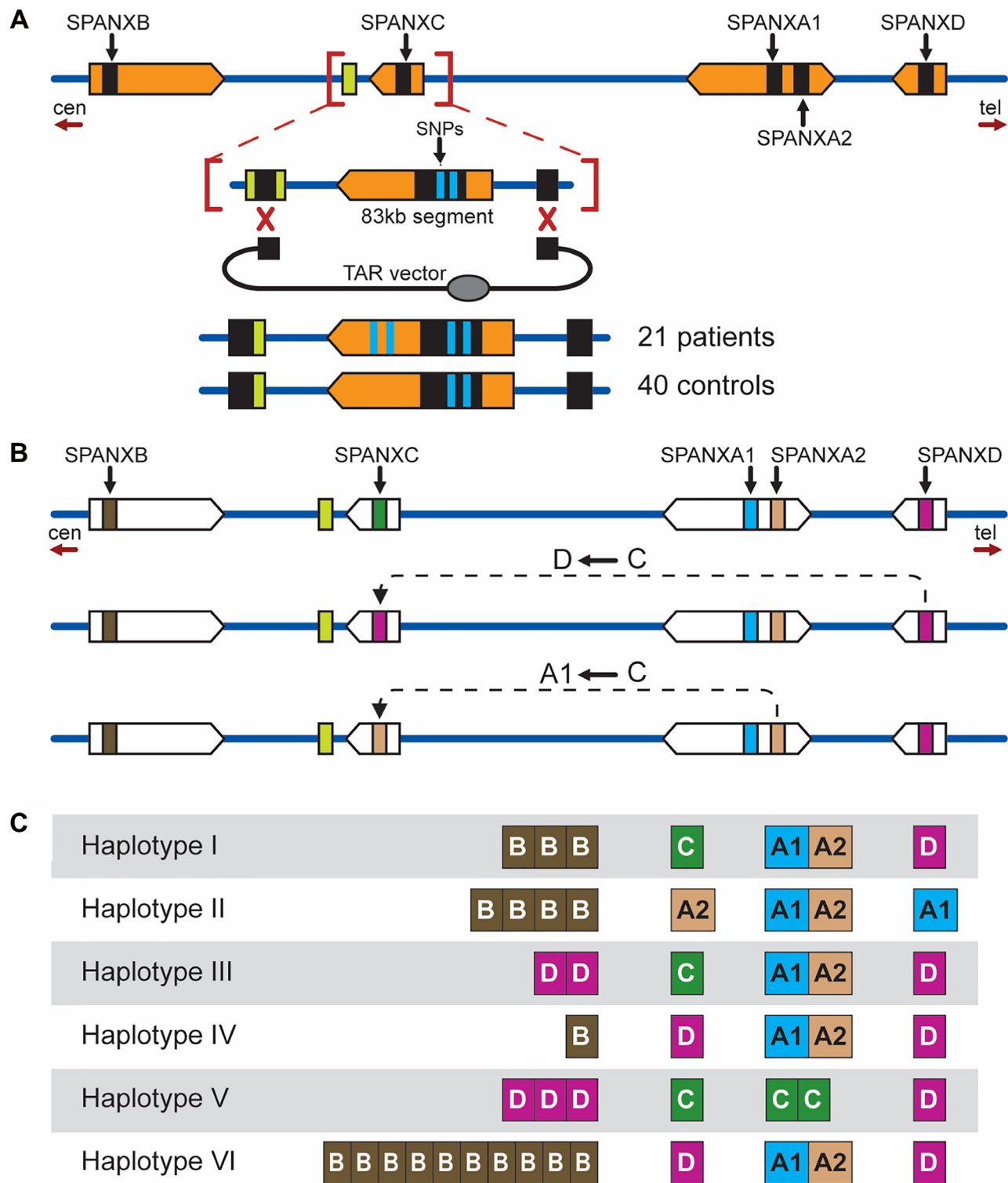


Figure 6: Long-range haplotyping of the Xq27 region in prostate cancer patients and healthy individuals. (A) TAR isolation of the *SPANXC* gene. The positions of *SPANX* genes are shown. The genes (in black) share 98% identity and reside within large segmental duplications (SDs) (in orange) with >95% identity. The *SPANXC* gene was TAR-isolated as an 83 kb segment from 21 patients and 40 healthy individuals in one TAR cloning experiment. The targeting hooks were chosen from the sequences outside the SD. The same strategy was used in mutational analysis of other gene family members. Orientation of the genes in the SDs towards the centromere (cen) and telomere (tel) is indicated: *SPANXD* to *SPANXC* and *SPANXA1* to *SPANXC*. **(B)** Gene conversion events between the *SPANX*-A/D gene family members (indicated by black arrows). **(C)** Six (I–VI) *SPANX* long-range haplotypes identified in the patients. Each column represents one of *SPANX* genes. For example, analysis of TAR isolates from individual VI revealed 10 copies of *SPANXB* at the *SPANXB* locus and conversion of *SPANXD* to *SPANXC* at the *SPANXC* locus.

SPANX-A/D gene members have a level of homology close to 95% and reside within large segmental duplications (SDs) with >95% identity [70] (Figure 6A) that excludes a conventional PCR for gene mutational analysis. TAR cloning enabled the isolation of each member of this sub-family from dozens of normal individuals and patients during a quite short time [67–69] (Figure 6A). Further sequencing analysis of the TAR isolates revealed a high frequency of recombination between the genes due to gene conversion (for example, *SPANX-C* to *SPANX-A1* or *SPANX-C* to *SPANX-D*) (Figure 6B). As seen, the corresponding recombinational interaction operates over a long distance (~500 kb) [71]. Sequencing data allowed to reconstruct long-range *SPANX* haplotypes [68] (see examples in Figure 6C). Moreover, sequence analysis and long-range haplotyping in normal individuals and patients revealed no disease-specific mutations or genomic alterations within the *SPANX* gene cluster that excluded a 750-kb region at Xq27-q28 as a candidate locus for prostate malignancy [69].

To summarize, TAR cloning is a unique tool to rapidly and accurately isolate both alleles of a gene and build long-range haplotypes for multiple heterozygous individuals. In addition, TAR cloning can help to develop diagnostics for disorders caused by genomic rearrangements that would provide the important foundation for the biomedicine of the future.

Comparative genomics and evolutionary studies: isolation of gene homologues

As described in chapter 2.2, 15% divergence between the targeting hook sequences in the TAR vector and target 5' and 3' ends of the genomic region does not prevent recombination allowing TAR isolation of homologous regions from different species [21]. For example, the efficiency of cloning of the mouse *HPRT* gene using TAR vector containing the human targeting hooks having 14% divergence with the 5' and 3' ends of the mouse *HPRT* gene was the same as cloning of the human *HPRT* gene [8]. Other examples of application of TAR technology of the large genomic regions from evolutionary close species include: the *BRCA1* tumor suppressor gene [72], the microcephaly gene *ASPM* controlling brain size [73], the *SPANX* gene sub-family [70], and the *NBS1* and *ATM* genes involved in DNA repair [20]. These genes were isolated from human and non-human primate species using TAR vectors containing the human targeting hooks corresponding to 5' and 3' gene-flanking regions with the hook sequences being 14–15% diverged. The subsequent sequence analysis then enabled the reconstruction of the evolutionary history of these genes [67, 72, 73].

Sequence analysis of the *ASPM* gene, which encodes for a mitotic spindle protein [74], revealed a high conservation in both coding and noncoding regions and allowed to infer that evolution of this gene was under

positive selection in hominoids [73]. That study also suggested that the evolutionary selection of *ASPM* in the African hominoid clade preceded hominid brain expansion by several million years and strongly correlated with differences in cerebral cortical size [73].

The *BRCA1* gene is involved in many cellular functions, including DNA replication, cell-cycle checkpoint activation, gene transcriptional regulation, DNA damage repair, kinetochore function, and centrosome function. Analysis of the synonymous versus non-synonymous substitution ratio in the coding region of *BRCA1* revealed that the coding (internal) sequence has evolved under positive selection while the terminal regions of *BRCA1*, which encode the BRCT domain and RING finger, are almost identical in all primates [72]. Interestingly, the human *BRCA1* gene contains 129 *Alu* elements, accounting for ~42% of the entire gene sequence. It was shown that a significant fraction of germline *BRCA1* mutations in hereditary breast and ovarian cancers are deletions and duplications caused by homologous *Alu-Alu* recombination [75], resulting in gene inactivation [76]. Sequence analysis of TAR-isolated *BRCA1* homologues revealed that the *Alu* repeats involved in disease-associated genomic rearrangements are conserved in nonhuman primates, suggesting their functional significance. Additionally, *Alu*-mediated rearrangements, including *Alu*-associated deletions and *Alu* transpositions, are the major force of evolutionary changes in noncoding *BRCA1* sequences [72].

The most unexpected example is reconstruction of the evolutionary history of the *SPANX-A/D* gene sub-family in primates [70, 71]. As described above (see chapter 4.2), these genes are located within 95% identical SDs. The latter precludes the detection of lineage-specific amplification of these genes by routine PCR or next generation sequencing analyses of syntenic chromosomal segments. TAR cloning enabled to overcome that problem, and the syntenic fragments from human, chimpanzee, bonobo, gorilla, orangutan, and macaque were isolated and sequenced. Remarkably, the corresponding TAR clones from syntenic regions of chimpanzee, bonobo, and gorilla genomes did not contain the *SPANX-C* gene that means that this gene is human-specific [70] (Figure 7). Analysis of the *SPANX-B* containing duplication revealed a variable number of a 12 kb tandem repeat carrying *SPANX-B* within SD, ranging from 1 to 14 copies, that is present only in humans [71]. More interesting, further analysis of the TAR isolates revealed that the *SPANX-A/D* gene sub-family is absent in orangutan and macaque (Figure 7), making this gene sub-family specific for the human lineage.

To summarize, despite the progress in genome sequencing, a quick, an efficient and a simultaneous TAR isolation of gene homologues from different species provides an opportunity to address fundamental questions in environmental evolution.

Biotechnology: selective isolation and assembly of biosynthetic gene clusters (BGCs) from individual microbial genomes and environmental DNA samples

NPs BGCs and their derivatives are the major source of pharmacological agents and industrial compounds [77]. With ineffectiveness of most antibiotics and the spread of drug-resistant pathogens, the discovery of new BGCs and developing them into drugs has become an urgent necessity. So, it is not surprising that TAR cloning became a widely used, an effective, rapid, and accurate tool for capture of BGCs from microbial genomes and for their cloning and assembly from collections of overlapping environmental eDNA clones (summarized in [78, 79]).

Over the past decade, there are many successful examples of TAR cloning of BGCs from bacteria and collections of soil-derived eDNA clones for commercial purposes as well as for functional studies [11, 13, 78, 80–104]. In many cases, cloning of BGCs from cultured microorganisms is possible using the protocol with TAR vector lacking an *ARS* (chapter 3.1) (Figure 2) [105]. However, a lot of microbial genomes are low in the density of *ARS*-elements and therefore it is highly likely that some

BGCs do not possess *ARS*-like sequences. For such cases, a TAR cloning protocol, that uses a counter-selectable marker, adapted for genomes that have few or no *ARS*-like sequences is preferable (chapter 3.2) (Figure 3) [12]. Figure 8 describes a general scheme of TAR isolation of a natural product biosynthetic gene cluster from a microbe with its following transfer to a bacterial host strain for production of the natural compound or basic research.

One of the examples of TAR capture of BGCs from eDNA samples is screening of eDNA megalibraries [106]. Isolation and further structure elucidation of metabolites obtained through heterologous expression of these gene clusters identified three new fluostatins (F, G, H) that had not been characterized before from studies of cultured species. Two other groups described TAR capture of overlapping soil-derived eDNA clones followed by their re-assembly into ~90 kb BGCs [85, 87]. TAR capture of BGC to yield a new antibiotic was described by Yamanaka et al. [96]. More specifically, a nonribosomal peptide synthetase cluster 73 kb in size was isolated from the marine actinomycete *Saccharomonospora sp.* CNQ-490 to produce lipopeptide antibiotic taromycin A in the model expression host *Streptomyces coelicolor* [96]. Later TAR cloning of the 6-demethylchlortetracycline BGC from

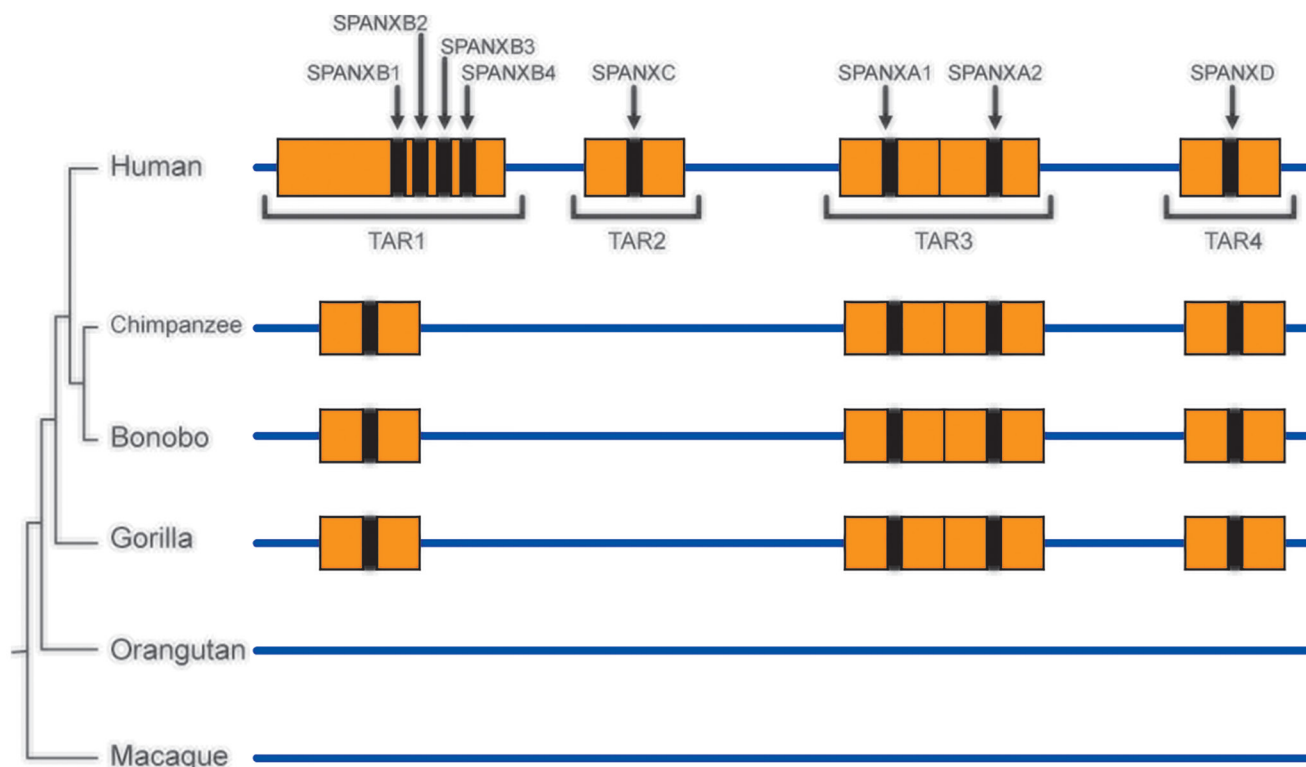


Figure 7: TAR cloning of the *SPANX-A/D* genes and their organization in primates. Syntenic genomic fragments containing different members of the *SPANX-A/D* gene family from the human, chimpanzee, bonobo, gorilla, orangutan, and macaque were TAR-isolated. Unique targeting hooks in the TAR vectors were chosen from the human genome regions flanking the SDs where the *SPANX-A/D* genes are located. Note that the divergence of these regions between primates is approximately 15%. Sequencing of the TAR isolates (TAR2) from chimpanzee, bonobo, gorilla, orangutan, and macaque showed that these primates do not contain the *SPANX-C* gene along with the SDs. Sequencing of the TAR isolates (TAR1) from primates showed that the *SPANXB* gene along with the SDs is absent in the orangutan and macaque. In human, the *SPANXB* gene is amplified from one up to 14 copies that is human specific.

Streptomyces aureofaciens was described [107]. A similar approach was used to isolate the 54 kb aromatic polyketide antitumor agent cosmomycin BGC from *Streptomyces* bacteria [108] and the putative thioviridamide-like gene

cluster, including up and downstream flanking regions, from *Streptomyces* sp. NRRL S-4 [109]. TAR cloning also allowed discovery and isolation of the 67 kb malacidins as calcium-dependent antibiotics with activity against

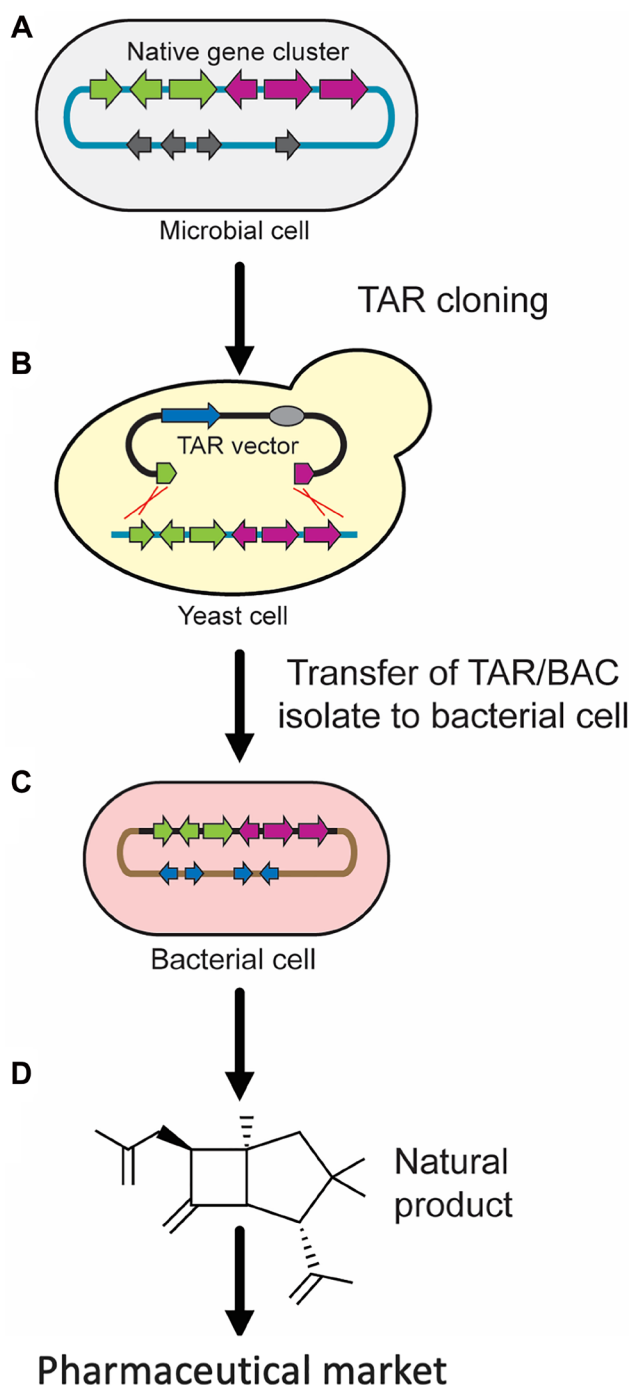


Figure 8: TAR cloning of natural product biosynthetic gene clusters (BGCs) from microbial genomes. (A) Organization of a native gene cluster located on a chromosome propagated in a microbial cell. (B) Isolation of BGC by TAR vector where the hooks (in green and in purple) correspond to the ends of the gene cluster. TAR is based on homologous recombination between the hooks of TAR vector and co-transformed genomic DNA isolated from host microbial cells that leads to formation of a circular YAC/BAC construct. In the TAR vector, the hooks or homology arms (in green and in purple) correspond to the ends of the cluster. (C, D) TAR-cloned gene cluster is transferred to bacterial cells and, if necessary, integrated into the chromosome of a host strain for basic research or production of natural compounds.

multidrug-resistant Gram-positive pathogens [110]. In another work, a 33 kb genomic region that includes a cryptic antibiotic biosynthesis gene locus was identified and TAR-captured from human pathogenic *Nocardia* strain and then expressed in the *Streptomyces* host revealing it to be a source of the brasiliquinones and benz(a)anthraquinone antibiotics [111].

More recent examples of TAR application is capture of large BGCs with high G+C content, including 98 kb tylosin (tyl), 128 kb daptomycin (dpt), and 127 kb salinomycin (sal) with their further heterologous expression in *Streptomyces coelicolor* M1146 to produce tylosins in the resulting recombinant strains [104] and 127 kb stictamycin (sal), an aromatic polyketide antibiotic isolated from a New Zealand Lichen-Sourced *Streptomyces* species with activity against *Staphylococcus aureus* that is the most pathogenic (it is typically causes skin infections and sometimes pneumonia, endocarditis, and osteomyelitis) [112]. Another group reported a yeast-based platform that exploits TAR cloning for capture, expression, and analysis of a BGC encoding a nonribosomal peptide eponemycin, a novel antibiotic, and TAR capture of TMC-86A that belongs to a family of peptide natural products to clarify the biosynthesis of these important proteasome inhibitors [113].

Recently Awal and co-authors [114] have performed an extraordinary work on isolation of 30 specific genes, that are comprised within the compact magnetosome gene cluster (MGCs), from the *Magnetospirillum* bacteria using the TAR cloning technique. In species of *Magnetospirillum*, biosynthesis of magnetosomes is a complex process, governed by these 30 genes. A further reconstruction, transfer, and analysis of this entire magnetosome cluster is promising for engineering the biomineralization of magnetite crystals with different morphologies that would be valuable for biotechnical applications [114].

Another notable example was demonstrated by Santos-Aberturas and co-authors [115]. Thioviridamide is a structurally novel ribosomally synthesized and post-translational modified peptide (RiPP) produced by *Streptomyces olivoviridis* NA005001. This peptide is characterized by a series of thioamide groups and possesses potent antiproliferative activity in cancer cell lines. The authors investigated the diversity of thioviridamide-like pathways across sequenced bacterial genomes and three diverse members of this family were TAR-captured from the genetically intractable *Streptomyces sclerotialis* bacterial strain [115].

CRISPR/Cas9-mediated TAR cloning (chapter 3.1) was applied to isolate the core genes for plipastatin biosynthesis from *B. amyloliquefaciens* HYM12 followed by their highly efficient expression in a heterologous system of *Bacillus subtilis* [116]. The same strategy was applied for isolation of staurosporine BGC 22.5 kb

in size from the native producer and then introduction into heterologous hosts *Streptomyces avermitilis* [117]. Staurosporine is the most well-known member of the indolocarbazole alkaloid family. It can induce apoptosis of many types of cells as a strong protein kinase inhibitor and is used as an important compound for the synthesis of the antitumor drug [117].

A recent example of reconstruction of BGC is TAR cloning and further reassembly of the gene cluster of microcystin-LR from *Microcystis aeruginosa*, a species of freshwater cyanobacteria that can form harmful blooms of economic and ecological importance [118]. *Microcystis aeruginosa* produces microcystin-LR (MC-LR), the most common cyanotoxin. Isolation of this cluster allowed to study the biosynthetic pathways and molecular mechanisms of MC-LR.

To summarize, TAR method provides a powerful, effective, and accurate tool for isolation of natural product biosynthetic gene clusters for biomedicine, biotechnology, and fundamental research.

Biomedicine: assembly and cloning of synthetic viruses and bacteriophages

TAR cloning is used to genetically engineer synthetic viruses with novel properties to design a new generation of vaccines. Recently Kurhade and co-authors summarized the status of TAR-assembled viral genomes, including SARS-CoV-2, and their further applications for studying the pathogenesis and replication of viruses and the development of vaccines [119]. Figure 9 shows an example of construction of a synthetic genomic RNA for the respiratory syndrome coronavirus 2 (SARS-Cov2) [120]. Step 1 includes PCR amplification or chemically synthesized 12 small viral fragments having overlapping ends. F1 fragment is fused with T7 promoter at its 5' terminus. F12 fragment is fused with polyA (pA) at its 3' terminus. Step 2 describes viral genome assembly in yeast as YAC molecules containing the full-length viral DNA using TAR vector containing two targeting hooks with the homology to the 5' and 3' ends of the PCR-amplified F1 and F12 fragments. Step 3 includes *in vitro* transcription of the assembled genome with T7 RNA polymerase to generate the infectious full-length viral genomic RNA.

Using a similar TAR-based approach, two other groups reconstructed RNA viruses, including members of the Coronaviridae, Flaviviridae, and Pneumoviridae families [29, 30]. They used sub-genomic fragments of diverse origins: viral isolates, cloned viral DNA, clinical viral samples, or synthetic DNA fragments. These fragments were reassembled as YAC molecules in one step in yeast using transformation-associated recombination. Then using T7 RNA polymerase infectious virus RNAs were generated for the rescue of a viable viruses. The same group also engineered the SARS-CoV-2 virus using chemically synthesized synthetic DNA fragments of the

virus. Only a week was required to generate the full-length viral genomic RNA [29, 30].

Another impressive example of TAR-based assembly of the entire viral genome is reconstruction of infectious laryngotracheitis virus (ILTV), known as Gallid alphaherpesvirus-1 [121]. In this case, the authors generated overlapping cosmid clones, that encompassed 90% of the 151 kb ILTV genome. Homologous recombination between the clones in yeast allowed to develop the full-length genome of the ILTV virus [121].

TAR-based approach allowed to manipulate the constructs by modifying the genes encoding virulence factors that facilitated the development of the improved virus vaccines and establishing ILTV-based viral vectors for expressing immunogens of other avian pathogens.

Recently the TAR technology allowed to rescue different strains of feline infectious peritonitis virus without multiple cloning steps [122]. That virus causes a deadly disease in cats for which there is no effective vaccine. In this study, the authors provided an improved TAR-based

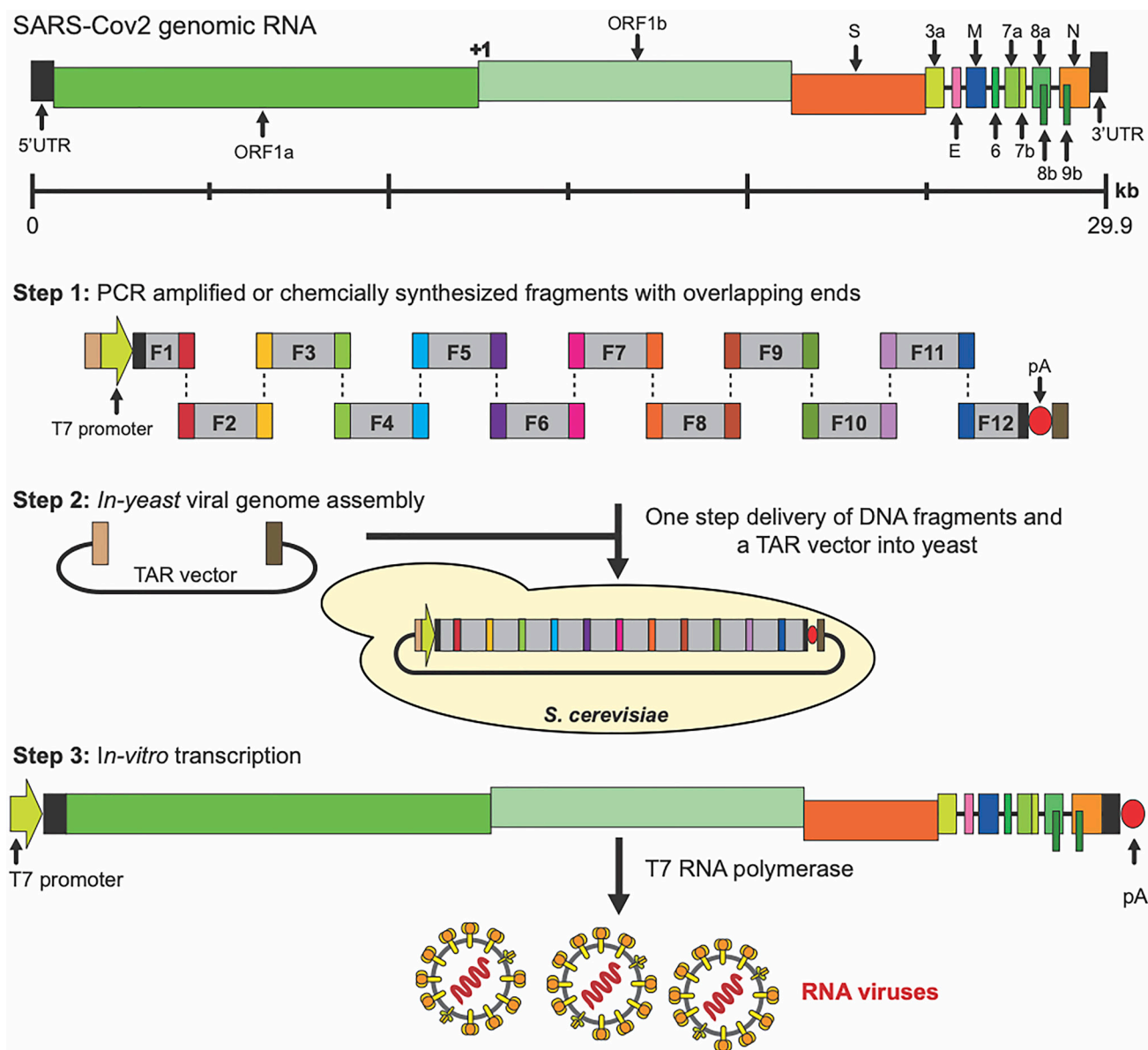


Figure 9: One-step SARS-CoV-2 genome assembly in yeast *S. cerevisiae* using transformation-associated recombination (TAR) in yeast [120]. **Step 1:** Genomic DNA of SARS-CoV-2 was divided into 12 overlapping fragments. F1 fragment was fused with T7 promoter at its 5' terminus (in green and in light brown rectangle) and F12 fragment was fused with poly(A) (pA) at its 3' terminus (in dark brown). TAR vector has targeting hooks (in light brown and in dark brown) having homology to T7 promoter (the light brown part) and the sequence fused with F12 (in dark brown). **Step 2:** One-step delivery of 12 fragments and the linearized TAR vector into yeast cells and following assembly of the viral full-length cDNA by homologous recombination. **Step 3:** The EagI site at the 5' end of the fragment F12 was cleaved to linearize the molecule. The linearized molecule was used as the template to synthesize the viral infectious full-length RNA of SARS-CoV-2 using T7 RNA polymerase.

system and constructed infective cDNA in one week. This allowed them to construct an infectious virus that would benefit for the vaccine development and pathogenic mechanism research [122]. Similarly, a combination of PCR and TAR cloning allowed to construct the genome of the *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) [123]. TAR cloning was used to assemble the overlapping fragments into a complete herpes simplex virus type 1 genome (HSV-1) [124] and has been also adapted to directly clone the genome of large human cytomegalovirus (HCMV) [125].

Bacteriophages, also known as phages, are viruses that infect and replicate only in bacterial cells. They are ubiquitous in the environment and are recognized as the most abundant type of organisms on earth. Recently they have received renewed attention for their potential to address the rise of multidrug-resistant bacteria resulting from the overuse of antibiotics [126, 127]. Therefore, modification of phages by homologous recombination in yeast or their assembly by TAR cloning provides a promising approach against antibiotic-resistant bacteria. Though the phages have a limited range of hosts that hinder their effectiveness, TAR engineering may expand the bacterial host range, and improve phage pharmacological efficacy. One among multiple examples of TAR engineering of phages is a full phiX174 genome assembly with the yield 44% of the required clones [128]. More examples of TAR-based assembly of phage DNA fragments into complete genomes in yeast, followed by transformation into the hosts to produce activated phages for drug production are summarized by Jia et al. [127].

Synthetic biology: assembly of microbe genomes and Mb-scale human DNA fragments

An era of assembly of synthetic microbial genomics began in 2008. A team of the J. Craig Venter Institute described the first synthesized and assembled bacterial *Mycoplasma genitalium* genome (JCV1-1) approximately 590 kb in size [82]. The *M. genitalium* genome was assembled using a combination of *in vitro* enzymatic and *in vivo* TAR cloning approaches. First, 25 DNA fragments with an average length of 24 kb were assembled *in vitro* into four 144 kb fragments having the homologous ends to each other. Then TAR method was applied to assemble the whole *M. genitalium* genome. Later, using the recombination machinery in yeast, this team assembled the 25 overlapping DNA fragments in yeast into a complete microbial genome JCV1-1 ~590 kb in size in a single step (Figure 10) [129].

For the past decade TAR cloning promoted a significant progress in synthetic biology. Another example of TAR application is assembly of eleven ~100 kb overlapping DNA fragments into the complete 1.1 Mb *M. mycoides* genome that was then transferred into closely related *M. capricolum* cells to form chimeric *M. mycoides*

cells. The novel chimeric cells were capable of self-replication and revealed the expected phenotypes [83]. At present, TAR cloning is routinely used to assemble the whole genomes from either synthetic or natural DNA molecules; for example, 0.8 Mb *M. pneumoniae* [25], 1.5 Mb *Acholeplasma laidlawii* [26], and 1.6 Mb *Prochlorococcus marinus* MED4 genomes [27].

The potential of TAR cloning in yeast as a universal host for *in vivo* assembly of large eukaryotic chromosomes has been also demonstrated [130]. One of the examples describes assembly of two eukaryotic chromosomes, each ~500 kb in size, of the algal *Phaeodactylum tricornerutum* genome (27.4 Mb) [28]. The TAR strategy has been also applied to assembly four-, five-, and six-gene complex pathways to generate yeast cells synthesizing beta-carotene, and violacein [131] and to engineer complex pathways, such as the synthesis of amorphaadiene and vanillin [132, 133]. Thus, TAR cloning is adapted as a reliable and accurate method for the assembly of synthetic genomes and pathways [134, 135].

Synthetic biology: construction of human artificial chromosomes with a defined structure

For the past two decades, HAC-based vectors have been widely used for gene delivery, new anticancer drug screening, discovery of novel genes involved in chromosome transmission, and for the study of centromere assembly and function [54, 55, 136, 137]. The HACs have a potential for gene therapy, and regenerative medicine [138]. The HACs are maintained stably as an additional 47th chromosome in human cells over multiple generations due to the presence of functional kinetochore [137]. Because the HACs can carry the genes with all regulatory elements, this allows the genes to mimic the pattern of the natural gene expression.

In 2005 Ebersole and co-authors described a method to construct synthetic alphoid DNA arrays with a predetermined structure [139]. Using *transformation-associated* recombination in yeast, it was shown that 2mer or 4mer or 5mer alphoid DNA repeats consisting of alphoid 170 bp monomers and having the ends homologous to each other may be one-step TAR-assembled into long synthetic alphoid DNA arrays varying in size from 50 to 140 kb (Step 1: Figure 11). After transfection of such arrays into human cells, *de novo* HACs are generated, ranging in size from 1 to 10 Mb due to amplification of the input alphoid DNA arrays (Steps 2 and 3) [139–143]. Because any nucleotide in the original dimer can be easily changed before its amplification, this TAR-based method allows to identify the critical regions of the alphoid repeat for *de novo* centromere seeding.

A decade ago, a HAC termed as the alphoid^{tetO}-HAC was constructed, using the TAR-based method. The first step included one-step assembly in yeast of 34 overlapping 348 bp in size alphoid DNA dimers to form

a 120 kb synthetic array. In each dimer, one monomer contained a tetO sequence in place of the CENP-B box that allowed these dimers to be targeted specifically with tetR-fusion proteins. The 120 kb array was transformed into human fibrosarcoma HT1080 cells forming a 1.1

Mb HAC [141, 144]. Further analysis of the HAC revealed that heterochromatin is incompatible with centromere function and that centromeric transcription is important for centromere assembly and maintenance [55, 137, 144]. The alphaoid^{tetO}-HAC was adapted for

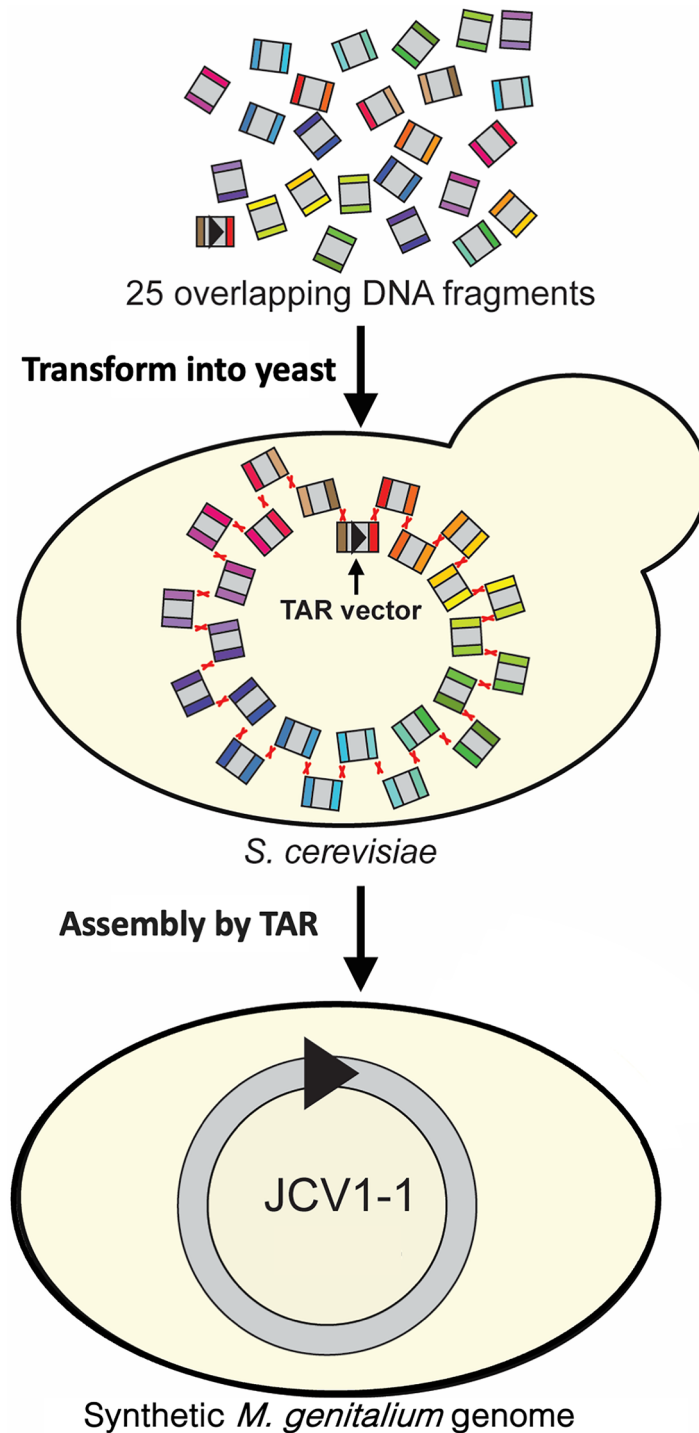


Figure 10: Assembly of a synthetic *Mycoplasma genitalium* genome from 25 overlapping DNA fragments in the *Saccharomyces cerevisiae* yeast cell [129]. DNA fragments from 17 kb up to 35 kb each along with the TAR vector were transformed into yeast cell. Homologous recombination between the ends of the fragments led to one-step assembly of the synthetic *M. genitalium* genome. This assembled genome called JCV1-1, 590,011 bp in size, includes the TAR vector sequence (marked by black triangle). TAR vector has targeting hooks having homology to the first and the last fragments of *M. genitalium* genome.

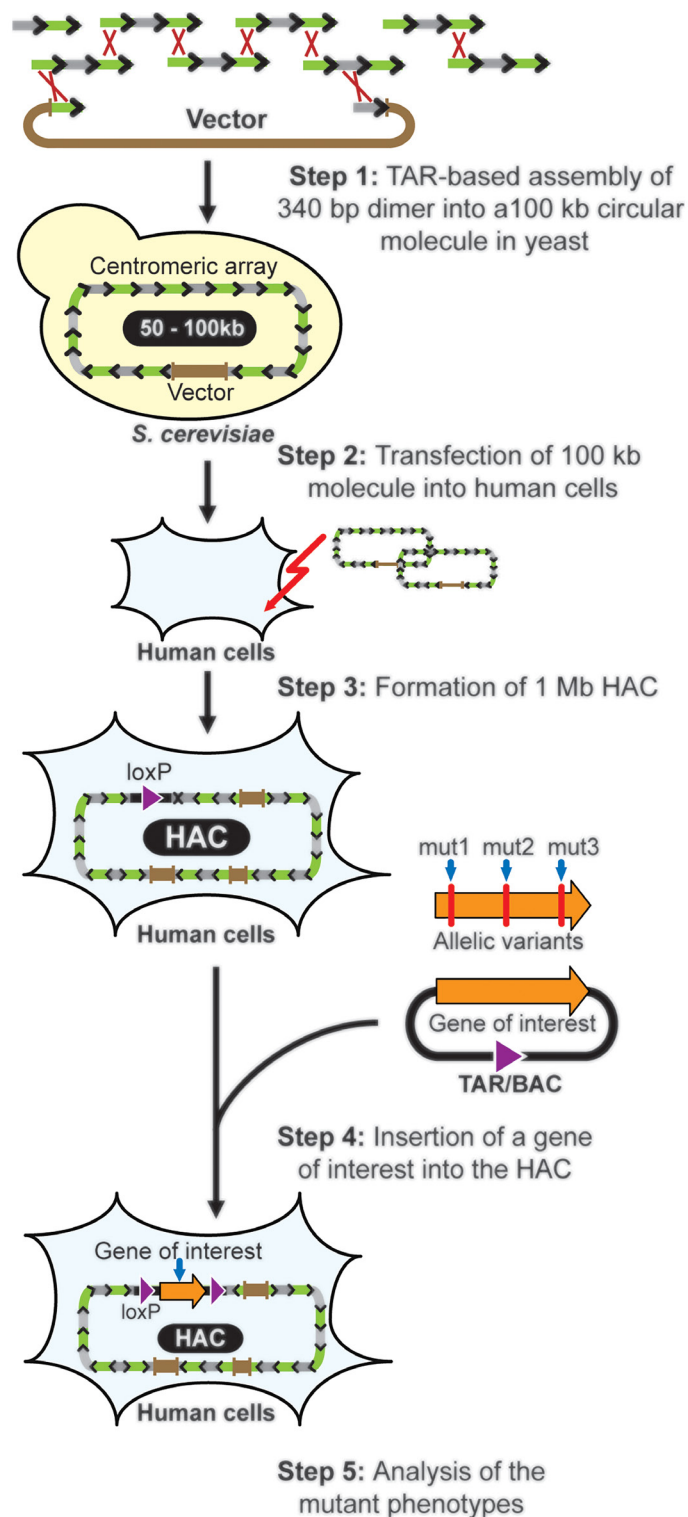


Figure 11: HAC-based gene delivery vector formation. **Step 1:** Assembly of the 340 bp DNA dimers into 50–100 kb centromeric arrays by TAR cloning in yeast *S. cerevisiae* cells. TAR vector contains yeast and bacterial cassettes, a mammalian selectable marker (the Blasticidin resistance gene), and the 340 bp dimers as the targeting hooks. TAR vector linearized between the hooks and a mixture of dimers are transformed into yeast cells. End-to-end recombination of the DNA dimers followed by interaction of the recombined fragments with the vector hooks results in formation of the 50–100 kb centromeric arrays as circular YAC/BAC molecules. **Step 2:** TAR vector contains a mammalian selectable marker (the Blasticidin resistance gene) that allows a TAR/BAC clone containing a 100 kb centromeric array to be transferred to and maintained in human cells. **Step 3:** Formation of the HAC accompanied by input alphoid DNA multimerization up to ~1 Mb [141]. **Step 4:** A TAR-isolated gene of choice or an allelic variant of a gene (mut1, mut2 or mut3) is loaded into the single gene acceptor loxP site of the HAC by Cre-loxP mediated recombination. **Step 5:** Analysis of the mutant phenotypes of the gene.

gene delivery and expression studies that allowed the TAR-isolated genomic copies of the genes (*HPRT*, *VHL*, *NBS1*, *BRCA1*, *PDK1*, *ATM*, rDNA) to be inserted into a unique gene loading *LoxP* site for further analysis of allelic variants [56–58, 145, 146] (Steps 4 and 5). Importantly, the complementation of mutant phenotypes arising from stable gene expression can be reversed by inactivating HAC's kinetochore in proliferating cell populations, a feature that provides a control for phenotypic changes attributed to expression of HAC-encoded genes [57, 58, 146].

To conclude, the TAR-based method has a general application in elucidating the role of other tandem repeats in chromosome organization and dynamics. It is worth noting that in 2008 Gibson and co-authors applied a similar strategy to assemble a complete synthetic *Mycoplasma* genome from 25 overlapping DNA fragments [129].

CONCLUSIONS

TAR cloning has become a valuable procedure for the selective and efficient isolation and manipulation of large DNA molecules. Its ability to isolate unperturbed native genomic regions provides a basis for a multitude of practical applications in biomedicine and biotechnology. The availability of full-length genes containing exons and introns with 5' upstream and 3' downstream regulatory sequences will catalyze major breakthroughs in functional, structural, and comparative genomics, diagnostics, gene replacement, and generation of animal models for human diseases. The ability to isolate individual gene alleles will help to clarify whether a particular allele is associated with predisposition to different diseases, including cancer. Accumulated comprehensive knowledge of the genetic basis of human diseases provides a foundation for future medical research and has far-reaching implications for basic, clinical, and commercial efforts to understand, prevent, and treat diseases and develop new strategies for their diagnostic and treatment. The TAR technology is applicable to capture BGCs directly and efficiently from microbe organisms and environmental DNA samples, which are the source of the natural products for the pharmaceutical market. In addition, this will help us to understand BGCs biosynthesis and molecular mechanisms. TAR cloning is used to genetically engineer synthetic viruses with novel properties that may be used for the development of new vaccines. In perspective, given the potential of TAR-engineered HACs to deliver a TAR-cloned therapeutic gene(s) into cells with its associated regulatory elements offers a tremendous potential for gene therapy applications. Note that the HAC with the ability to carry an unlimited number of TAR-isolated genes [147] allows the development of multiple-gene humanized models, disease models, and the reprogramming and investigation of complex biomedical pathways.

Author contributions

Conceptualization, N.K. and V.L.; writing – original draft preparation, N.K.; writing - review and editing, N.K.; figure making- N.K.; references collecting - N.K. and V.L.; funding acquisition, V.L. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

Authors have no conflicts of interest to declare.

FUNDING

This research was funded by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research, USA (N.K. and V.L.; ZIA BC010413).

REFERENCES

1. Orr-Weaver TL, Szostak JW, Rothstein RJ. Yeast transformation: a model system for the study of recombination. *Proc Natl Acad Sci U S A*. 1981; 78:6354–58. <https://doi.org/10.1073/pnas.78.10.6354>. [PubMed]
2. Kunes S, Botstein D, Fox MS. Transformation of yeast with linearized plasmid DNA. Formation of inverted dimers and recombinant plasmid products. *J Mol Biol*. 1985; 184:375–87. [https://doi.org/10.1016/0022-2836\(85\)90288-8](https://doi.org/10.1016/0022-2836(85)90288-8). [PubMed]
3. Ma H, Kunes S, Schatz PJ, Botstein D. Plasmid construction by homologous recombination in yeast. *Gene*. 1987; 58:201–16. [https://doi.org/10.1016/0378-1119\(87\)90376-3](https://doi.org/10.1016/0378-1119(87)90376-3). [PubMed]
4. Larionov V, Kouprina N, Eldarov M, Perkins E, Porter G, Resnick MA. Transformation-associated recombination between diverged and homologous DNA repeats is induced by strand breaks. *Yeast*. 1994; 10:93–104. <https://doi.org/10.1002/yea.320100109>. [PubMed]
5. Larionov V, Kouprina N, Graves J, Chen XN, Korenberg JR, Resnick MA. Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc Natl Acad Sci U S A*. 1996; 93:491–96. <https://doi.org/10.1073/pnas.93.1.491>. [PubMed]
6. Larionov V, Kouprina N, Solomon G, Barrett JC, Resnick MA. Direct isolation of human BRCA2 gene by transformation-associated recombination in yeast. *Proc Natl Acad Sci U S A*. 1997; 94:7384–87. <https://doi.org/10.1073/pnas.94.14.7384>. [PubMed]
7. Kouprina N, Annab L, Graves J, Afshari C, Barrett JC, Resnick MA, Larionov V. Functional copies of a human gene can be directly isolated by transformation-associated recombination cloning with a small 3' end target sequence. *Proc Natl Acad Sci U S A*. 1998; 95:4469–74. <https://doi.org/10.1073/pnas.95.8.4469>. [PubMed]

8. Kouprina N, Larionov V. TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nat Rev Genet.* 2006; 7:805–12. <https://doi.org/10.1038/nrg1943>. [PubMed]
9. Kouprina N, Kim JH, Larionov V. Highly Selective, CRISPR/Cas9-Mediated Isolation of Genes and Genomic Loci from Complex Genomes by TAR Cloning in Yeast. *Curr Protoc.* 2021; 1:e207. <https://doi.org/10.1002/cpz1.207>. [PubMed]
10. Noskov VN, Kouprina N, Leem SH, Ouspenski I, Barrett JC, Larionov V. A general cloning system to selectively isolate any eukaryotic or prokaryotic genomic region in yeast. *BMC Genomics.* 2003; 4:16. <https://doi.org/10.1186/1471-2164-4-16>. [PubMed]
11. Zhang JJ, Yamanaka K, Tang X, Moore BS. Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination. *Methods Enzymol.* 2019; 621:87–110. <https://doi.org/10.1016/bs.mie.2019.02.026>. [PubMed]
12. Kouprina N, Noskov VN, Larionov V. Selective isolation of large segments from individual microbial genomes and environmental DNA samples using transformation-associated recombination cloning in yeast. *Nat Protoc.* 2020; 15:734–49. <https://doi.org/10.1038/s41596-019-0280-1>. [PubMed]
13. Covington BC, Xu F, Seyedsayamdost MR. A Natural Product Chemist's Guide to Unlocking Silent Biosynthetic Gene Clusters. *Annu Rev Biochem.* 2021; 90:763–88. <https://doi.org/10.1146/annurev-biochem-081420-102432>. [PubMed]
14. Asakawa S, Abe I, Kudoh Y, Kishi N, Wang Y, Kubota R, Kudoh J, Kawasaki K, Minoshima S, Shimizu N. Human BAC library: construction and rapid screening. *Gene.* 1997; 191:69–79. [https://doi.org/10.1016/s0378-1119\(97\)00044-9](https://doi.org/10.1016/s0378-1119(97)00044-9). [PubMed]
15. Jiang W, Zhao X, Gabrieli T, Lou C, Ebenstein Y, Zhu TF. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat Commun.* 2015; 6:8101. <https://doi.org/10.1038/ncomms9101>. [PubMed]
16. Gibson DG. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res.* 2009; 37:6984–90. <https://doi.org/10.1093/nar/gkp687>. [PubMed]
17. Theis JF, Newlon CS. The ARS309 chromosomal replicator of *Saccharomyces cerevisiae* depends on an exceptional ARS consensus sequence. *Proc Natl Acad Sci U S A.* 1997; 94:10786–91. <https://doi.org/10.1073/pnas.94.20.10786>. [PubMed]
18. Noskov V, Kouprina N, Leem SH, Koriabine M, Barrett JC, Larionov V. A genetic system for direct selection of gene-positive clones during recombinational cloning in yeast. *Nucleic Acids Res.* 2002; 30:E8. <https://doi.org/10.1093/nar/30.2.e8>. [PubMed]
19. Kouprina N, Larionov V. Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma.* 2016; 125:621–32. <https://doi.org/10.1007/s00412-016-0588-3>. [PubMed]
20. Noskov VN, Koriabine M, Solomon G, Randolph M, Barrett JC, Leem SH, Stubbs L, Kouprina N, Larionov V. Defining the minimal length of sequence homology required for selective gene isolation by TAR cloning. *Nucleic Acids Res.* 2001; 29:E32. <https://doi.org/10.1093/nar/29.6.e32>. [PubMed]
21. Noskov VN, Leem SH, Solomon G, Mullokandov M, Chae JY, Yoon YH, Shin YS, Kouprina N, Larionov V. A novel strategy for analysis of gene homologues and segmental genome duplications. *J Mol Evol.* 2003; 56:702–10. <https://doi.org/10.1007/s00239-002-2442-x>. [PubMed]
22. Xie YG, Han FY, Peyrard M, Ruttledge MH, Fransson I, DeJong P, Collins J, Dunham I, Nordenskjöld M, Dumanski JP. Cloning of a novel, anonymous gene from a megabase-range YAC and cosmid contig in the neurofibromatosis type 2/meningioma region on human chromosome 22q12. *Hum Mol Genet.* 1993; 2:1361–68. <https://doi.org/10.1093/hmg/2.9.1361>. [PubMed]
23. Loots GG. Modifying yeast artificial chromosomes to generate Cre/LoxP and FLP/FRT site-specific deletions and inversions. *Methods Mol Biol.* 2006; 349:75–84. <https://doi.org/10.1385/1-59745-158-4:75>. [PubMed]
24. Kouprina N, Larionov V. Exploiting the yeast *Saccharomyces cerevisiae* for the study of the organization and evolution of complex genomes. *FEMS Microbiol Rev.* 2003; 27:629–49. [https://doi.org/10.1016/S0168-6445\(03\)00070-6](https://doi.org/10.1016/S0168-6445(03)00070-6). [PubMed]
25. Benders GA, Noskov VN, Denisova EA, Lartigue C, Gibson DG, Assad-Garcia N, Chuang RY, Carrera W, Moodie M, Algire MA, Phan Q, Alperovich N, Vashee S, et al. Cloning whole bacterial genomes in yeast. *Nucleic Acids Res.* 2010; 38:2558–69. <https://doi.org/10.1093/nar/gkq119>. [PubMed]
26. Karas BJ, Tagwerker C, Yonemoto IT, Hutchison CA 3rd, Smith HO. Cloning the *Acholeplasma laidlawii* PG-8A genome in *Saccharomyces cerevisiae* as a yeast centromeric plasmid. *ACS Synth Biol.* 2012; 1:22–28. <https://doi.org/10.1021/sb200013j>. [PubMed]
27. Tagwerker C, Dupont CL, Karas BJ, Ma L, Chuang RY, Benders GA, Ramon A, Novotny M, Montague MG, Venepally P, Bami D, Schwartz A, Andrews-Pfannkoch C, et al. Sequence analysis of a complete 1.66 Mb *Prochlorococcus marinus* MED4 genome cloned in yeast. *Nucleic Acids Res.* 2012; 40:10375–83. <https://doi.org/10.1093/nar/gks823>. [PubMed]
28. Karas BJ, Molparia B, Jablanovic J, Hermann WJ, Lin YC, Dupont CL, Tagwerker C, Yonemoto IT, Noskov VN, Chuang RY, Allen AE, Glass JI, Hutchison CA 3rd, et al. Assembly of eukaryotic algal chromosomes in yeast. *J Biol Eng.* 2013; 7:30. <https://doi.org/10.1186/1754-1611-7-30>. [PubMed]

29. Thi Nhu Thao T, Labroussaa F, Ebert N, V'kovski P, Stalder H, Portmann J, Kelly J, Steiner S, Holwerda M, Kratzel A, Gultom M, Schmied K, Laloli L, et al. Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature*. 2020; 582:561–65. <https://doi.org/10.1038/s41586-020-2294-9>. [PubMed]
30. Khan D, Terenzi F, Liu G, Ghosh PK, Ye F, Nguyen K, China A, Ramachandiran I, Chakraborty S, Stefan J, Khan K, Vasu K, Dong F, et al. A viral pan-end RNA element and host complex define a SARS-CoV-2 regulon. *Nat Commun*. 2023; 14:3385. <https://doi.org/10.1038/s41467-023-39091-3>. [PubMed]
31. Mali P, Esvelt KM, Church GM. Cas9 as a versatile tool for engineering biology. *Nat Methods*. 2013; 10:957–63. <https://doi.org/10.1038/nmeth.2649>. [PubMed]
32. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. 2014; 157:1262–78. <https://doi.org/10.1016/j.cell.2014.05.010>. [PubMed]
33. Wijshake T, Baker DJ, van de Sluis B. Endonucleases: new tools to edit the mouse genome. *Biochim Biophys Acta*. 2014; 1842:1942–50. <https://doi.org/10.1016/j.bbadis.2014.04.020>. [PubMed]
34. Kim JM, Kim D, Kim S, Kim JS. Genotyping with CRISPR-Cas-derived RNA-guided endonucleases. *Nat Commun*. 2014; 5:3157. <https://doi.org/10.1038/ncomms4157>. [PubMed]
35. Karvelis T, Gasiunas G, Siksnys V. Programmable DNA cleavage in vitro by Cas9. *Biochem Soc Trans*. 2013; 41:1401–6. <https://doi.org/10.1042/BST20130164>. [PubMed]
36. Lee NC, Larionov V, Kouprina N. Highly efficient CRISPR/Cas9-mediated TAR cloning of genes and chromosomal loci from complex genomes in yeast. *Nucleic Acids Res*. 2015; 43:e55. <https://doi.org/10.1093/nar/gkv112>. [PubMed]
37. Furter-Graves EM, Hall BD. DNA sequence elements required for transcription initiation of the *Schizosaccharomyces pombe* ADH gene in *Saccharomyces cerevisiae*. *Mol Gen Genet*. 1990; 223:407–16. <https://doi.org/10.1007/BF00264447>. [PubMed]
38. Miret JJ, Pessoa-Brandão L, Lahue RS. Orientation-dependent and sequence-specific expansions of CTG/CAG trinucleotide repeats in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A*. 1998; 95:12438–43. <https://doi.org/10.1073/pnas.95.21.12438>. [PubMed]
39. Noskov VN, Karas BJ, Young L, Chuang RY, Gibson DG, Lin YC, Stam J, Yonemoto IT, Suzuki Y, Andrews-Pfannkoch C, Glass JI, Smith HO, Hutchison CA 3rd, et al. Assembly of large, high G+C bacterial DNA fragments in yeast. *ACS Synth Biol*. 2012; 1:267–73. <https://doi.org/10.1021/sb3000194>. [PubMed]
40. Kouprina N, Graves J, Cancilla MR, Resnick MA, Larionov V. Specific isolation of human rDNA genes by TAR cloning. *Gene*. 1997; 197:269–76. [https://doi.org/10.1016/s0378-1119\(97\)00271-0](https://doi.org/10.1016/s0378-1119(97)00271-0). [PubMed]
41. Cancilla MR, Tainton KM, Barry AE, Larionov V, Kouprina N, Resnick MA, Sart DD, Choo KH. Direct cloning of human 10q25 neocentromere DNA using transformation-associated recombination (TAR) in yeast. *Genomics*. 1998; 47:399–404. <https://doi.org/10.1006/geno.1997.5129>. [PubMed]
42. Humble MC, Kouprina N, Noskov VN, Graves J, Garner E, Tennant RW, Resnick MA, Larionov V, Cannon RE. Radial transformation-associated recombination cloning from the mouse genome: isolation of Tg.AC transgene with flanking DNAs. *Genomics*. 2000; 70:292–99. <https://doi.org/10.1006/geno.2000.6384>. [PubMed]
43. Kim J, Noskov VN, Lu X, Bergmann A, Ren X, Warth T, Richardson P, Kouprina N, Stubbs L. Discovery of a novel, paternally expressed ubiquitin-specific processing protease gene through comparative analysis of an imprinted region of mouse chromosome 7 and human chromosome 19q13.4. *Genome Res*. 2000; 10:1138–47. <https://doi.org/10.1101/gr.10.8.1138>. [PubMed]
44. Kouprina N, Leem SH, Solomon G, Ly A, Koriabine M, Otstot J, Pak E, Dutra A, Zhao S, Barrett JC, Larionov V. Segments missing from the draft human genome sequence can be isolated by transformation-associated recombination cloning in yeast. *EMBO Rep*. 2003; 4:257–62. <https://doi.org/10.1038/sj.embor.embor766>. [PubMed]
45. Grimwood J, Gordon LA, Olsen A, Terry A, Schmutz J, Lamerdin J, Hellsten U, Goodstein D, Couronne O, Tran-Gyamfi M, Aerts A, Altherr M, Ashworth L, et al. The DNA sequence and biology of human chromosome 19. *Nature*. 2004; 428:529–35. <https://doi.org/10.1038/nature02399>. [PubMed]
46. Leem SH, Kouprina N, Grimwood J, Kim JH, Mullokandov M, Yoon YH, Chae JY, Morgan J, Lucas S, Richardson P, Detter C, Glavina T, Rubin E, et al. Closing the gaps on human chromosome 19 revealed genes with a high density of repetitive tandemly arrayed elements. *Genome Res*. 2004; 14:239–46. <https://doi.org/10.1101/gr.1929904>. [PubMed]
47. Tanaka H, Tsujimura A. Pervasiveness of intronless genes expressed in haploid germ cell differentiation. *Reprod Med Biol*. 2021; 20:255–59. <https://doi.org/10.1002/rmb2.12385>. [PubMed]
48. Qi M, Nayar U, Ludwig LS, Wagle N, Rheinbay E. cDNA-detector: detection and removal of cDNA contamination in DNA sequencing libraries. *BMC Bioinformatics*. 2021; 22:611. <https://doi.org/10.1186/s12859-021-04529-2>. [PubMed]
49. Medina-Santos R, Fernandes Costa TG, Silva de Assis TC, Kalapothakis Y, de Almeida Lima S, do Carmo AO, Gonzalez-Kozlova EE, Kalapothakis E, Chávez-Olórtegui C, Guerra-Duarte C. Analysis of NGS data from Peruvian *Loxosceles laeta* spider venom gland reveals toxin diversity. *Comp Biochem Physiol Part D Genomics Proteomics*. 2022; 43:101017. <https://doi.org/10.1016/j.cbd.2022.101017>. [PubMed]

50. Zheng HC, Xue H, Zhang CY. REG4 promotes the proliferation and anti-apoptosis of cancer. *Front Cell Dev Biol.* 2022; 10:1012193. <https://doi.org/10.3389/fcell.2022.1012193>. [PubMed]
51. Annab LA, Kouprina N, Solomon G, Cable PL, Hill DE, Barrett JC, Larionov V, Afshari CA. Isolation of a functional copy of the human BRCA1 gene by transformation-associated recombination in yeast. *Gene.* 2000; 250:201–8. [https://doi.org/10.1016/s0378-1119\(00\)00180-3](https://doi.org/10.1016/s0378-1119(00)00180-3). [PubMed]
52. Leem SH, Londoño-Vallejo JA, Kim JH, Bui H, Tubacher E, Solomon G, Park JE, Horikawa I, Kouprina N, Barrett JC, Larionov V. The human telomerase gene: complete genomic sequence and analysis of tandem repeat polymorphisms in intronic regions. *Oncogene.* 2002; 21:769–77. <https://doi.org/10.1038/sj.onc.1205122>. [PubMed]
53. Horikawa I, Chiang YJ, Patterson T, Feigenbaum L, Leem SH, Michishita E, Larionov V, Hodes RJ, Barrett JC. Differential cis-regulation of human versus mouse TERT gene expression in vivo: identification of a human-specific repressive element. *Proc Natl Acad Sci U S A.* 2005; 102:18437–42. <https://doi.org/10.1073/pnas.0508964102>. [PubMed]
54. Kouprina N, Tomilin AN, Masumoto H, Earnshaw WC, Larionov V. Human artificial chromosome-based gene delivery vectors for biomedicine and biotechnology. *Expert Opin Drug Deliv.* 2014; 11:517–35. <https://doi.org/10.1517/17425247.2014.882314>. [PubMed]
55. Kouprina N, Petrov N, Molina O, Liskovych M, Pesenti E, Ohzeki JI, Masumoto H, Earnshaw WC, Larionov V. Human Artificial Chromosome with Regulated Centromere: A Tool for Genome and Cancer Studies. *ACS Synth Biol.* 2018; 7:1974–89. <https://doi.org/10.1021/acssynbio.8b00230>. [PubMed]
56. Ayabe F, Katoh M, Inoue T, Kouprina N, Larionov V, Oshimura M. A novel expression system for genomic DNA loci using a human artificial chromosome vector with transformation-associated recombination cloning. *J Hum Genet.* 2005; 50:592–99. <https://doi.org/10.1007/s10038-005-0300-6>. [PubMed]
57. Kim JH, Kononenko A, Erliandri I, Kim TA, Nakano M, Iida Y, Barrett JC, Oshimura M, Masumoto H, Earnshaw WC, Larionov V, Kouprina N. Human artificial chromosome (HAC) vector with a conditional centromere for correction of genetic deficiencies in human cells. *Proc Natl Acad Sci U S A.* 2011; 108:20048–53. <https://doi.org/10.1073/pnas.1114483108>. [PubMed]
58. Kononenko AV, Bansal R, Lee NC, Grimes BR, Masumoto H, Earnshaw WC, Larionov V, Kouprina N. A portable BRCA1-HAC (human artificial chromosome) module for analysis of BRCA1 tumor suppressor function. *Nucleic Acids Res.* 2014; 42:e164. <https://doi.org/10.1093/nar/gku870>. [PubMed]
59. Kim JH, Dilthey AT, Nagaraja R, Lee HS, Koren S, Dudekula D, Wood Iii WH, Piao Y, Ogurtsov AY, Utani K, Noskov VN, Shabalina SA, Schlessinger D, et al. Variation in human chromosome 21 ribosomal RNA genes characterized by TAR cloning and long-read sequencing. *Nucleic Acids Res.* 2018; 46:6712–25. <https://doi.org/10.1093/nar/gky442>. [PubMed]
60. Kim JH, Noskov VN, Ogurtsov AY, Nagaraja R, Petrov N, Liskovych M, Walenz BP, Lee HS, Kouprina N, Phillippy AM, Shabalina SA, Schlessinger D, Larionov V. The genomic structure of a human chromosome 22 nucleolar organizer region determined by TAR cloning. *Sci Rep.* 2021; 11:2997. <https://doi.org/10.1038/s41598-021-82565-x>. [PubMed]
61. Salem RM, Wessel J, Schork NJ. A comprehensive literature review of haplotyping software and methods for use with unrelated individuals. *Hum Genomics.* 2005; 2:39–66. <https://doi.org/10.1186/1479-7364-2-1-39>. [PubMed]
62. Kuleshov V, Xie D, Chen R, Pushkarev D, Ma Z, Blauwkamp T, Kertesz M, Snyder M. Whole-genome haplotyping using long reads and statistical methods. *Nat Biotechnol.* 2014; 32:261–66. <https://doi.org/10.1038/nbt.2833>. [PubMed]
63. Wang M, Beck CR, English AC, Meng Q, Buhay C, Han Y, Doddapaneni HV, Yu F, Boerwinkle E, Lupski JR, Muzny DM, Gibbs RA. PacBio-LITS: a large-insert targeted sequencing method for characterization of human disease-associated chromosomal structural variations. *BMC Genomics.* 2015; 16:214. <https://doi.org/10.1186/s12864-015-1370-2>. [PubMed]
64. Kim JH, Leem SH, Sunwoo Y, Kouprina N. Separation of long-range human TERT gene haplotypes by transformation-associated recombination cloning in yeast. *Oncogene.* 2003; 22:2452–56. <https://doi.org/10.1038/sj.onc.1206316>. [PubMed]
65. Schleutker J, Matikainen M, Smith J, Koivisto P, Baffoe-Bonnie A, Kainu T, Gillanders E, Sankila R, Pukkala E, Carpten J, Stephan D, Tammela T, Brownstein M, et al. A genetic epidemiological study of hereditary prostate cancer (HPC) in Finland: frequent HPCX linkage in families with late-onset disease. *Clin Cancer Res.* 2000; 6:4810–15. [PubMed]
66. Xu J, Meyers D, Freije D, Isaacs S, Wiley K, Nusskern D, Ewing C, Wilkens E, Bujnovszky P, Bova GS, Walsh P, Isaacs W, Schleutker J, et al. Evidence for a prostate cancer susceptibility locus on the X chromosome. *Nat Genet.* 1998; 20:175–79. <https://doi.org/10.1038/2477>. [PubMed]
67. Kouprina N, Noskov VN, Pavlicek A, Collins NK, Schoppee Bortz PD, Ottolenghi C, Loukinov D, Goldsmith P, Risinger JI, Kim JH, Westbrook VA, Solomon G, Sounders H, et al. Evolutionary diversification of SPANX-N sperm protein gene structure and expression. *PLoS One.* 2007; 2:e359. <https://doi.org/10.1371/journal.pone.0000359>. [PubMed]
68. Kouprina N, Noskov VN, Solomon G, Otsot J, Isaacs W, Xu J, Schleutker J, Larionov V. Mutational analysis of SPANX genes in families with X-linked prostate cancer. *Prostate.* 2007; 67:820–28. <https://doi.org/10.1002/pros.20561>. [PubMed]

69. Kouprina N, Lee NC, Pavlicek A, Samoshkin A, Kim JH, Lee HS, Varma S, Reinhold WC, Otstot J, Solomon G, Davis S, Meltzer PS, Schleutker J, Larionov V. Exclusion of the 750-kb genetically unstable region at Xq27 as a candidate locus for prostate malignancy in HPCX1-linked families. *Genes Chromosomes Cancer*. 2012; 51:933–48. <https://doi.org/10.1002/gcc.21977>. [PubMed]
70. Kouprina N, Mullokandov M, Rogozin IB, Collins NK, Solomon G, Otstot J, Risinger JI, Koonin EV, Barrett JC, Larionov V. The SPANX gene family of cancer/testis-specific antigens: rapid evolution and amplification in African great apes and hominids. *Proc Natl Acad Sci U S A*. 2004; 101:3077–82. <https://doi.org/10.1073/pnas.0308532100>. [PubMed]
71. Kouprina N, Pavlicek A, Noskov VN, Solomon G, Otstot J, Isaacs W, Carpten JD, Trent JM, Schleutker J, Barrett JC, Jurka J, Larionov V. Dynamic structure of the SPANX gene cluster mapped to the prostate cancer susceptibility locus HPCX at Xq27. *Genome Res*. 2005; 15:1477–86. <https://doi.org/10.1101/gr.4212705>. [PubMed]
72. Pavlicek A, Noskov VN, Kouprina N, Barrett JC, Jurka J, Larionov V. Evolution of the tumor suppressor BRCA1 locus in primates: implications for cancer predisposition. *Hum Mol Genet*. 2004; 13:2737–51. <https://doi.org/10.1093/hmg/ddh301>. [PubMed]
73. Kouprina N, Pavlicek A, Mochida GH, Solomon G, Gersch W, Yoon YH, Collura R, Ruvolo M, Barrett JC, Woods CG, Walsh CA, Jurka J, Larionov V. Accelerated evolution of the ASPM gene controlling brain size begins prior to human brain expansion. *PLoS Biol*. 2004; 2:E126. <https://doi.org/10.1371/journal.pbio.0020126>. [PubMed]
74. Kouprina N, Pavlicek A, Collins NK, Nakano M, Noskov VN, Ohzeki J, Mochida GH, Risinger JI, Goldsmith P, Gunsior M, Solomon G, Gersch W, Kim JH, et al. The microcephaly ASPM gene is expressed in proliferating tissues and encodes for a mitotic spindle protein. *Hum Mol Genet*. 2005; 14:2155–65. <https://doi.org/10.1093/hmg/ddi220>. [PubMed]
75. Puget N, Torchard D, Serova-Sinilnikova OM, Lynch HT, Feunteun J, Lenoir GM, Mazoyer S. A 1-kb Alu-mediated germ-line deletion removing BRCA1 exon 17. *Cancer Res*. 1997; 57:828–31. [PubMed]
76. Petrij-Bosch A, Peelen T, van Vliet M, van Eijk R, Olmer R, Drüsedau M, Hogervorst FB, Hageman S, Arts PJ, Ligtenberg MJ, Meijers-Heijboer H, Klijn JG, Vasen HF, et al. BRCA1 genomic deletions are major founder mutations in Dutch breast cancer patients. *Nat Genet*. 1997; 17:341–45. <https://doi.org/10.1038/ng1197-341>. [PubMed]
77. Newman DJ, Cragg GM. Natural Products as Sources of New Drugs from 1981 to 2014. *J Nat Prod*. 2016; 79:629–61. <https://doi.org/10.1021/acs.jnatprod.5b01055>. [PubMed]
78. Wang W, Zheng G, Lu Y. Recent Advances in Strategies for the Cloning of Natural Product Biosynthetic Gene Clusters. *Front Bioeng Biotechnol*. 2021; 9:692797. <https://doi.org/10.3389/fbioe.2021.692797>. [PubMed]
79. Wang X, Zhou N, Wang B. Bacterial synthetic biology: tools for novel drug discovery. *Expert Opin Drug Discov*. 2023; 18:1087–97. <https://doi.org/10.1080/17460441.2023.2239704>. [PubMed]
80. Becker M, Aitcheson N, Byles E, Wickstead B, Louis E, Rudenko G. Isolation of the repertoire of VSG expression site containing telomeres of *Trypanosoma brucei* 427 using transformation-associated recombination in yeast. *Genome Res*. 2004; 14:2319–29. <https://doi.org/10.1101/gr.2955304>. [PubMed]
81. Young R, Taylor JE, Kurioka A, Becker M, Louis EJ, Rudenko G. Isolation and analysis of the genetic diversity of repertoires of VSG expression site containing telomeres from *Trypanosoma brucei gambiense*, *T. b. brucei* and *T. equiperdum*. *BMC Genomics*. 2008; 9:385. <https://doi.org/10.1186/1471-2164-9-385>. [PubMed]
82. Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, Zaveri J, Stockwell TB, Brownley A, Thomas DW, Algire MA, Merryman C, Young L, Noskov VN, et al. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*. 2008; 319:1215–20. <https://doi.org/10.1126/science.1151721>. [PubMed]
83. Lartigue C, Vashee S, Algire MA, Chuang RY, Benders GA, Ma L, Noskov VN, Denisova EA, Gibson DG, Assad-Garcia N, Alperovich N, Thomas DW, Merryman C, et al. Creating bacterial strains from genomes that have been cloned and engineered in yeast. *Science*. 2009; 325:1693–96. <https://doi.org/10.1126/science.1173759>. [PubMed]
84. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*. 2010; 329:52–56. <https://doi.org/10.1126/science.1190719>. [PubMed]
85. Kim JH, Feng Z, Bauer JD, Kallifidas D, Calle PY, Brady SF. Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. *Biopolymers*. 2010; 93:833–44. <https://doi.org/10.1002/bip.21450>. [PubMed]
86. Gaida A, Becker MM, Schmid CD, Bühlmann T, Louis EJ, Beck HP. Cloning of the repertoire of individual *Plasmodium falciparum* var genes using transformation associated recombination (TAR). *PLoS One*. 2011; 6:e17782. <https://doi.org/10.1371/journal.pone.0017782>. [PubMed]
87. Feng Z, Kallifidas D, Brady SF. Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *Proc Natl Acad Sci U S A*. 2011; 108:12629–34. <https://doi.org/10.1073/pnas.1103921108>. [PubMed]
88. Ross AC, Gulland LE, Dorrestein PC, Moore BS. Targeted capture and heterologous expression of the *Pseudoalteromonas alterochromide* gene cluster in *Escherichia coli* represents a promising natural product

- exploratory platform. *ACS Synth Biol.* 2015; 4:414–20. <https://doi.org/10.1021/sb500280q>. [PubMed]
89. Cano-Prieto C, García-Salcedo R, Sánchez-Hidalgo M, Braña AF, Fiedler HP, Méndez C, Salas JA, Olano C. Genome Mining of *Streptomyces* sp. Tü 6176: Characterization of the Nataxazole Biosynthesis Pathway. *Chembiochem.* 2015; 16:1461–73. <https://doi.org/10.1002/cbic.201500153>. [PubMed]
90. Bonet B, Teufel R, Crüsemann M, Ziemert N, Moore BS. Direct capture and heterologous expression of *Salinispora* natural product genes for the biosynthesis of enterocin. *J Nat Prod.* 2015; 78:539–42. <https://doi.org/10.1021/np500664q>. [PubMed]
91. Li Y, Li Z, Yamanaka K, Xu Y, Zhang W, Vlamakis H, Kolter R, Moore BS, Qian PY. Directed natural product biosynthesis gene cluster capture and expression in the model bacterium *Bacillus subtilis*. *Sci Rep.* 2015; 5:9383. <https://doi.org/10.1038/srep09383>. [PubMed]
92. Tang X, Li J, Millán-Aguíñaga N, Zhang JJ, O'Neill EC, Ugalde JA, Jensen PR, Mantovani SM, Moore BS. Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining. *ACS Chem Biol.* 2015; 10:2841–49. <https://doi.org/10.1021/acscchembio.5b00658>. [PubMed]
93. Li ZR, Li Y, Lai JY, Tang J, Wang B, Lu L, Zhu G, Wu X, Xu Y, Qian PY. Critical Intermediates Reveal New Biosynthetic Events in the Enigmatic Colibactin Pathway. *Chembiochem.* 2015; 16:1715–19. <https://doi.org/10.1002/cbic.201500239>. [PubMed]
94. Shao Z, Zhao H, Zhao H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* 2009; 37:e16. <https://doi.org/10.1093/nar/gkn991>. [PubMed]
95. Ongley SE, Bian X, Neilan BA, Müller R. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat Prod Rep.* 2013; 30:1121–38. <https://doi.org/10.1039/c3np70034h>. [PubMed]
96. Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC, Moore BS. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc Natl Acad Sci U S A.* 2014; 111:1957–62. <https://doi.org/10.1073/pnas.1319584111>. [PubMed]
97. Agarwal V, El Gamal AA, Yamanaka K, Poth D, Kersten RD, Schorn M, Allen EE, Moore BS. Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat Chem Biol.* 2014; 10:640–47. <https://doi.org/10.1038/nchembio.1564>. [PubMed]
98. Mizutani K. High-throughput plasmid construction using homologous recombination in yeast: its mechanisms and application to protein production for X-ray crystallography. *Biosci Biotechnol Biochem.* 2015; 79:1–10. <https://doi.org/10.1080/09168451.2014.952614>. [PubMed]
99. Yuan Y, Andersen E, Zhao H. Flexible and Versatile Strategy for the Construction of Large Biochemical Pathways. *ACS Synth Biol.* 2016; 5:46–52. <https://doi.org/10.1021/acssynbio.5b00117>. [PubMed]
100. Zhao Q, Wang L, Luo Y. Recent advances in natural products exploitation in *Streptomyces* via synthetic biology. *Eng Life Sci.* 2019; 19:452–62. <https://doi.org/10.1002/elsc.201800137>. [PubMed]
101. Xu Y, Du X, Yu X, Jiang Q, Zheng K, Xu J, Wang P. Recent Advances in the Heterologous Expression of Biosynthetic Gene Clusters for Marine Natural Products. *Mar Drugs.* 2022; 20:341. <https://doi.org/10.3390/md20060341>. [PubMed]
102. Zhgun AA. Fungal BGCs for Production of Secondary Metabolites: Main Types, Central Roles in Strain Improvement, and Regulation According to the Piano Principle. *Int J Mol Sci.* 2023; 24:11184. <https://doi.org/10.3390/ijms241311184>. [PubMed]
103. Moore SJ, Lai HE, Li J, Freemont PS. *Streptomyces* cell-free systems for natural product discovery and engineering. *Nat Prod Rep.* 2023; 40:228–36. <https://doi.org/10.1039/d2np00057a>. [PubMed]
104. Tian L, Shi S, Zhang X, Han F, Dong H. Newest perspectives of glycopeptide antibiotics: biosynthetic cascades, novel derivatives, and new appealing antimicrobial applications. *World J Microbiol Biotechnol.* 2023; 39:67. <https://doi.org/10.1007/s11274-022-03512-0>. [PubMed]
105. Kouprina N, Larionov V. Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat Protoc.* 2008; 3:371–77. <https://doi.org/10.1038/nprot.2008.5>. [PubMed]
106. Feng Z, Kim JH, Brady SF. Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type II PKS gene cluster. *J Am Chem Soc.* 2010; 132:11902–3. <https://doi.org/10.1021/ja104550p>. [PubMed]
107. Wu N, Huang H, Min T, Hu H. TAR cloning and integrated overexpression of 6-demethylchlortetracycline biosynthetic gene cluster in *Streptomyces aureofaciens*. *Acta Biochim Biophys Sin (Shanghai).* 2017; 49:1129–34. <https://doi.org/10.1093/abbs/gmx110>. [PubMed]
108. Larson CB, Crüsemann M, Moore BS. PCR-Independent Method of Transformation-Associated Recombination Reveals the Cosmomycin Biosynthetic Gene Cluster in an Ocean *Streptomyces*. *J Nat Prod.* 2017; 80:1200–4. <https://doi.org/10.1021/acs.jnatprod.6b01121>. [PubMed]
109. Frattaruolo L, Lacret R, Cappello AR, Truman AW. A Genomics-Based Approach Identifies a Thioviridamide-Like Compound with Selective Anticancer Activity. *ACS Chem Biol.* 2017; 12:2815–22. <https://doi.org/10.1021/acscchembio.7b00677>. [PubMed]
110. Hover BM, Kim SH, Katz M, Charlop-Powers Z, Owen JG, Ternei MA, Maniko J, Estrela AB, Molina H, Park S,

- Perlin DS, Brady SF. Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat Microbiol.* 2018; 3:415–22. <https://doi.org/10.1038/s41564-018-0110-1>. [PubMed]
111. Herisse M, Ishida K, Porter JL, Howden B, Hertweck C, Stinear TP, Pidot SJ. Identification and Mobilization of a Cryptic Antibiotic Biosynthesis Gene Locus from a Human-Pathogenic *Nocardia* Isolate. *ACS Chem Biol.* 2020; 15:1161–68. <https://doi.org/10.1021/acscchembio.9b00763>. [PubMed]
 112. Hou P, Woolner VH, Bracegirdle J, Hunt P, Keyzers RA, Owen JG. Stictamycin, an Aromatic Polyketide Antibiotic Isolated from a New Zealand Lichen-Sourced *Streptomyces* Species. *J Nat Prod.* 2023; 86:526–32. <https://doi.org/10.1021/acs.jnatprod.2c00801>. [PubMed]
 113. Huang C, Zabala D, de Los Santos ELC, Song L, Corre C, Alkhalaf LM, Challis GL. Parallelized gene cluster editing illuminates mechanisms of epoxyketone proteasome inhibitor biosynthesis. *Nucleic Acids Res.* 2023; 51:1488–99. <https://doi.org/10.1093/nar/gkad009>. [PubMed]
 114. Awal RP, Lefevre CT, Schüler D. Functional expression of foreign magnetosome genes in the alphaproteobacterium *Magnetospirillum gryphiswaldense*. *mBio.* 2023; 14:e0328222. <https://doi.org/10.1128/mbio.03282-22>. [PubMed]
 115. Santos-Aberturas J, Chandra G, Frattaruolo L, Lacroix R, Pham TH, Vior NM, Eyles TH, Truman AW. Uncovering the unexplored diversity of thioamidated ribosomal peptides in Actinobacteria using the RiPPER genome mining tool. *Nucleic Acids Res.* 2019; 47:4624–37. <https://doi.org/10.1093/nar/gkz192>. [PubMed]
 116. Hu Y, Nan F, Maina SW, Guo J, Wu S, Xin Z. Clone of plipastatin biosynthetic gene cluster by transformation-associated recombination technique and high efficient expression in model organism *Bacillus subtilis*. *J Biotechnol.* 2018; 288:1–8. <https://doi.org/10.1016/j.jbiotec.2018.10.006>. [PubMed]
 117. Zhang Z, Yang S, Li Z, Wu Y, Tang J, Feng M, Chen S. High-titer production of staurosporine by heterologous expression and process optimization. *Appl Microbiol Biotechnol.* 2023; 107:5701–14. <https://doi.org/10.1007/s00253-023-12661-7>. [PubMed]
 118. Zheng Y, Xue C, Chen H, Jia A, Zhao L, Zhang J, Zhang L, Wang Q. Reconstitution and expression of mcy gene cluster in the model cyanobacterium *Synechococcus* 7942 reveals a role of MC-LR in cell division. *New Phytol.* 2023; 238:1101–14. <https://doi.org/10.1111/nph.18766>. [PubMed]
 119. Kurhade C, Xie X, Shi PY. Reverse genetic systems of SARS-CoV-2 for antiviral research. *Antiviral Res.* 2023; 210:105486. <https://doi.org/10.1016/j.antiviral.2022.105486>. [PubMed]
 120. Wang W, Peng X, Jin Y, Pan JA, Guo D. Reverse genetics systems for SARS-CoV-2. *J Med Virol.* 2022; 94:3017–31. <https://doi.org/10.1002/jmv.27738>. [PubMed]
 121. Spatz S, García M, Fuchs W, Loncoman C, Volkening J, Ross T, Riblet S, Kim T, Likens N, Mettenleiter T. Reconstitution and Mutagenesis of Avian Infectious Laryngotracheitis Virus from Cosmid and Yeast Centromeric Plasmid Clones. *J Virol.* 2023; 97:e0140622. <https://doi.org/10.1128/jvi.01406-22>. [PubMed]
 122. Cao H, Gu H, Kang H, Jia H. Development of a rapid reverse genetics system for feline coronavirus based on TAR cloning in yeast. *Front Microbiol.* 2023; 14:1141101. <https://doi.org/10.3389/fmicb.2023.1141101>. [PubMed]
 123. Shang Y, Wang M, Xiao G, Wang X, Hou D, Pan K, Liu S, Li J, Wang J, Arif BM, Vlcek JM, Chen X, Wang H, et al. Construction and Rescue of a Functional Synthetic Baculovirus. *ACS Synth Biol.* 2017; 6:1393–402. <https://doi.org/10.1021/acssynbio.7b00028>. [PubMed]
 124. Oldfield LM, Grzesik P, Voorhies AA, Alperovich N, MacMath D, Najera CD, Chandra DS, Prasad S, Noskov VN, Montague MG, Friedman RM, Desai PJ, Vashee S. Genome-wide engineering of an infectious clone of herpes simplex virus type 1 using synthetic genomics assembly methods. *Proc Natl Acad Sci U S A.* 2017; 114:E8885–94. <https://doi.org/10.1073/pnas.1700534114>. [PubMed]
 125. Vashee S, Stockwell TB, Alperovich N, Denisova EA, Gibson DG, Cady KC, Miller K, Kannan K, Malouli D, Crawford LB, Voorhies AA, Bruening E, Caposio P, Früh K. Cloning, Assembly, and Modification of the Primary Human Cytomegalovirus Isolate Toledo by Yeast-Based Transformation-Associated Recombination. *mSphere.* 2017; 2:e00331-17. <https://doi.org/10.1128/mSphereDirect.00331-17>. [PubMed]
 126. Song J, Liu Z, Zhang Q, Liu Y, Chen Y. Phage Engineering for Targeted Multidrug-Resistant *Escherichia coli*. *Int J Mol Sci.* 2023; 24:2459. <https://doi.org/10.3390/ijms24032459>. [PubMed]
 127. Jia HJ, Jia PP, Yin S, Bu LK, Yang G, Pei DS. Engineering bacteriophages for enhanced host range and efficacy: insights from bacteriophage-bacteria interactions. *Front Microbiol.* 2023; 14:1172635. <https://doi.org/10.3389/fmicb.2023.1172635>. [PubMed]
 128. Jaschke PR, Lieberman EK, Rodriguez J, Sierra A, Endy D. A fully decompressed synthetic bacteriophage øX174 genome assembled and archived in yeast. *Virology.* 2012; 434:278–84. <https://doi.org/10.1016/j.virol.2012.09.020>. [PubMed]
 129. Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA 3rd. One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc Natl Acad Sci U S A.* 2008; 105:20404–9. <https://doi.org/10.1073/pnas.0811011106>. [PubMed]
 130. Annaluru N, Muller H, Mitchell LA, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM, Cai Y, Zeller K, Agmon N,

- et al. Total synthesis of a functional designer eukaryotic chromosome. *Science*. 2014; 344:55–58. <https://doi.org/10.1126/science.1249252>. [PubMed]
131. Mitchell LA, Chuang J, Agmon N, Khunsriraksakul C, Phillips NA, Cai Y, Truong DM, Veerakumar A, Wang Y, Mayorga M, Blomquist P, Sadda P, Trueheart J, Boeke JD. Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S.cerevisiae*. *Nucleic Acids Res*. 2015; 43:6620–30. <https://doi.org/10.1093/nar/gkv466>. [PubMed]
 132. Brochado AR, Patil KR. Overexpression of O-methyltransferase leads to improved vanillin production in baker's yeast only when complemented with model-guided network engineering. *Biotechnol Bioeng*. 2013; 110:656–59. <https://doi.org/10.1002/bit.24731>. [PubMed]
 133. Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R, Horning T, Tsuruta H, Melis DJ, Owens A, Fickes S, Diola D, Benjamin KR, et al. Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc Natl Acad Sci U S A*. 2012; 109:E111–18. <https://doi.org/10.1073/pnas.1110740109>. [PubMed]
 134. Gibson DG. Programming biological operating systems: genome design, assembly and activation. *Nat Methods*. 2014; 11:521–26. <https://doi.org/10.1038/nmeth.2894>. [PubMed]
 135. Karas BJ, Suzuki Y, Weyman PD. Strategies for cloning and manipulating natural and synthetic chromosomes. *Chromosome Res*. 2015; 23:57–68. <https://doi.org/10.1007/s10577-014-9455-3>. [PubMed]
 136. Kouprina N, Earnshaw WC, Masumoto H, Larionov V. A new generation of human artificial chromosomes for functional genomics and gene therapy. *Cell Mol Life Sci*. 2013; 70:1135–48. <https://doi.org/10.1007/s00018-012-1113-3>. [PubMed]
 137. Molina O, Kouprina N, Masumoto H, Larionov V, Earnshaw WC. Using human artificial chromosomes to study centromere assembly and function. *Chromosoma*. 2017; 126:559–75. <https://doi.org/10.1007/s00412-017-0633-x>. [PubMed]
 138. Sinenko SA, Ponomartsev SV, Tomilin AN. Pluripotent stem cell-based gene therapy approach: human de novo synthesized chromosomes. *Cell Mol Life Sci*. 2021; 78:1207–20. <https://doi.org/10.1007/s00018-020-03653-1>. [PubMed]
 139. Ebersole T, Okamoto Y, Noskov VN, Kouprina N, Kim JH, Leem SH, Barrett JC, Masumoto H, Larionov V. Rapid generation of long synthetic tandem repeats and its application for analysis in human artificial chromosome formation. *Nucleic Acids Res*. 2005; 33:e130. <https://doi.org/10.1093/nar/gni129>. [PubMed]
 140. 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. <https://doi.org/10.1186/1480-9222-13-8>. [PubMed]
 141. Kouprina N, Samoshkin A, Erliandri I, Nakano M, Lee HS, Fu H, Iida Y, Aladjem M, Oshimura M, Masumoto H, Earnshaw WC, Larionov V. Organization of synthetic alphoid DNA array in human artificial chromosome (HAC) with a conditional centromere. *ACS Synth Biol*. 2012; 1:590–601. <https://doi.org/10.1021/sb3000436>. [PubMed]
 142. Pesenti E, Kouprina N, Liskovych M, Aurich-Costa J, Larionov V, Masumoto H, Earnshaw WC, Molina O. Generation of a Synthetic Human Chromosome with Two Centromeric Domains for Advanced Epigenetic Engineering Studies. *ACS Synth Biol*. 2018; 7:1116–30. <https://doi.org/10.1021/acssynbio.8b00018>. [PubMed]
 143. Pesenti E, Liskovych M, Okazaki K, Mallozzi A, Reid C, Abad MA, Jeyaprakash AA, Kouprina N, Larionov V, Masumoto H, Earnshaw WC. Analysis of Complex DNA Rearrangements during Early Stages of HAC Formation. *ACS Synth Biol*. 2020; 9:3267–87. <https://doi.org/10.1021/acssynbio.0c00326>. [PubMed]
 144. Nakano M, Cardinale S, Noskov VN, Gassmann R, Vagnarelli P, Kandels-Lewis S, Larionov V, Earnshaw WC, Masumoto H. Inactivation of a human kinetochore by specific targeting of chromatin modifiers. *Dev Cell*. 2008; 14:507–22. <https://doi.org/10.1016/j.devcel.2008.02.001>. [PubMed]
 145. Iida Y, Kim JH, Kazuki Y, Hoshiya H, Takiguchi M, Hayashi M, Erliandri I, Lee HS, Samoshkin A, Masumoto H, Earnshaw WC, Kouprina N, Larionov V, Oshimura M. Human artificial chromosome with a conditional centromere for gene delivery and gene expression. *DNA Res*. 2010; 17:293–301. <https://doi.org/10.1093/dnares/dsq020>. [PubMed]
 146. Liskovych M, Petrov NS, Noskov VN, Masumoto H, Earnshaw WC, Schlessinger D, Shabalina SA, Larionov V, Kouprina N. Actively transcribed rDNA and distal junction (DJ) sequence are involved in association of NORs with nucleoli. *Cell Mol Life Sci*. 2023; 80:121. <https://doi.org/10.1007/s00018-023-04770-3>. [PubMed]
 147. Lee NCO, Kim JH, Petrov NS, Lee HS, Masumoto H, Earnshaw WC, Larionov V, Kouprina N. Method to Assemble Genomic DNA Fragments or Genes on Human Artificial Chromosome with Regulated Kinetochore Using a Multi-Integrase System. *ACS Synth Biol*. 2018; 7:63–74. <https://doi.org/10.1021/acssynbio.7b00209>. [PubMed]
 148. Noskov VN, Chuang RY, Gibson DG, Leem SH, Larionov V, Kouprina N. Isolation of circular yeast artificial chromosomes for synthetic biology and functional genomics studies. *Nat Protoc*. 2011; 6:89–96. <https://doi.org/10.1038/nprot.2010.174>. [PubMed]