CrossMark

MINI-REVIEW

Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology

Natalay Kouprina[1] · Vladimir Larionov[1]

Abstract Transformation-associated recombination (TAR) cloning represents a unique tool for isolation and manipulation of large DNA molecules. The technique exploits a high level of homologous recombination in the yeast *Sacharomyces cerevisiae*. So far, TAR cloning is the only method available to selectively recover chromosomal segments up to 300 kb in length from complex and simple genomes. In addition, TAR cloning allows the assembly and cloning of entire microbe genomes up to several Mb as well as engineering of large metabolic pathways. In this review, we summarize applications of TAR cloning for functional/ structural genomics and synthetic biology.

Keywords Transformation-associated recombination · TAR cloning · Human artificial chromosome · HAC · Synthetic biology

Introduction

Transformation-associated recombination (TAR) cloning is based on the well-known property of free DNA ends as efficient substrates for homologous recombination in yeast (Orr-Weaver et al. 1981). Assembly of two DNA molecules by recombination

✉ Natalay Kouprina
  kouprinn@mail.nih.gov

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

in yeast was first demonstrated by Botstein and colleagues (Kunes et al. 1985; Ma et al. 1987). This team reported that a yeast vector DNA, which is broken in a sequence absent in the yeast genome, is efficiently rescued by recombination with a homologous restriction fragment included during yeast transformation (Kunes et al. 1985). A couple of years later, the authors suggested using this in vivo bimolecular reaction as a novel method for plasmid construction as an alternative to the more common method of in vitro DNA ligation (Ma et al. 1987). The breakthrough step was the demonstration that at specific conditions, transformation-associated recombination in yeast can be utilized for selective isolation of large genomic regions from total mammalian genomic DNA as linear or circular yeast artificial chromosomes (YACs) (Larionov et al. 1996, 1997; Kouprina et al. 1998; Kouprina and Larionov 2003). While the principles of this technology were developed in the late 1990s, the main parameters of TAR cloning have been modified during the past few years, which has allowed for highly efficient and reproducible gene targeting. Under such optimized conditions, any desirable chromosomal fragment up to 300 kb can be isolated in yeast from multiple samples within 2 weeks with a yield of gene/genomic fragment-positive clones as high as 32 % (Kouprina and Larionov 2006, 2008; Lee et al. 2015). For comparison, the frequency of gene-positive clones in genomic Bacterial Artificial Chromosome (BAC) libraries that are generated by DNA ligation is less than 0.003 % (Asakawa et al. 1997). The selectivity of gene isolation by TAR cloning from complex genomes was achieved by use of highly competent yeast spheroplasts and exclusion of a yeast origin of replication (*ARS* element) from a TAR vector. Propagation of TAR-generated YACs in yeast cells absolutely depends on acquisition of mammalian DNA fragments with *ARS*-like sequences that can function as an origin of replication in yeast. Because these sequences are common in mammalian DNA, with approximately one *ARS*-like sequence per 20–30 kb (Stinchcomb et al. 1980), most

mammalian genes can be isolated by TAR cloning. For chromosomal regions with multiple repetitive elements (such as the centromere and telomere) for GC-rich regions that are poor in *ARS*-like sequences, and for very simple genomes, the *ARS* frequency might be reduced, which precludes their isolation by the standard method. For such cases, a modified TAR method has been developed (Noskov et al. 2003a). In this system, the TAR vector contains an *ARS* and a counter-selectable marker. Negative genetic selection eliminates the background caused by vector recircularization that results from end-joining during yeast transformation. It is worth noting that cloning of large GC-rich bacterial DNA fragments may sometimes be challenging even in the presence of an *ARS*-containing vector, and in this case, only fragments of approximately 100 kb can be recovered (Noskov et al. 2012). A general scheme of TAR cloning of a single copy gene from total genomic DNA is presented in Fig. 1.

TAR cloning has found many applications in the postgenomic era (Fig. 2). For example, it serves as a tool for selective isolation of a specific chromosomal segment or gene from an individual. It can also be used to isolate rearranged chromosomal regions, such as translocations and inversions, from patients and model organisms. Here, we provide an overview of the recent applications of TAR cloning for structural and functional genomics as well as for synthetic biology. We also discuss the role of TAR cloning in construction of gene delivery vectors based on human artificial chromosomes (HACs), along with the benefits of coupling TAR gene cloning technology with the HAC gene delivery and expression system.

Applications of TAR cloning

Selective isolation of full-size single copy genes and gene clusters from total genomes

More and more publications have appeared describing complex mechanisms regulating gene or gene cluster expression by means of alternative splicing, alternative promoter-enhancer usage, and expression of non-coding RNAs from intronic regions. Therefore, full-size genes containing all of the necessary regulatory regions become preferable for gene functional studies because only in such configurations can "physiological" expression be achieved. The efficient homologous recombination machinery of the budding yeast host allows for the selective isolation of any full-length gene or gene cluster from entire complex genomes through recombination between a TAR cloning vector, containing targeting sequences homologous to a region/gene of interest and homologous sequences in the cotransformed genomic DNA (Fig. 1). This results in rescue of the desired chromosomal fragment or gene in a circular or linear form that is able to propagate, segregate, and be selected in yeast cells.
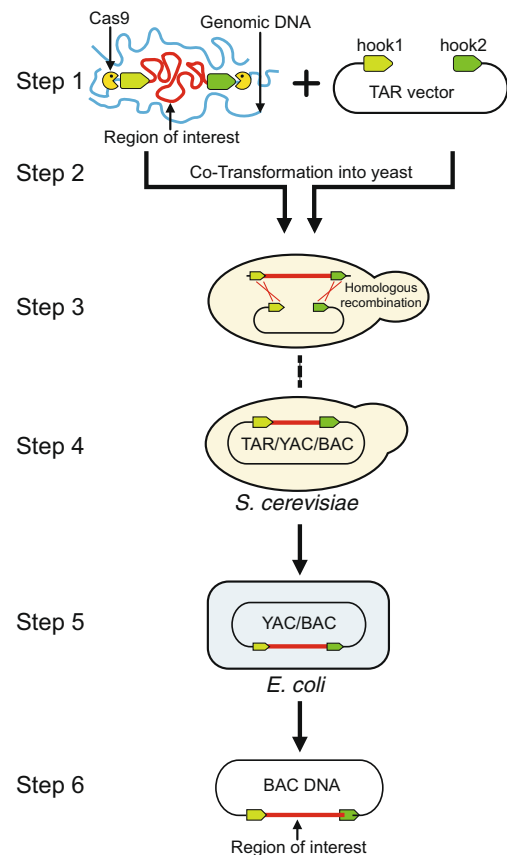

Fig. 1 The consecutive experimental steps of selective gene isolation. Step 1: The diagram shows TAR cloning of a gene of interest from total genomic DNA with a TAR vector containing YAC and BAC cassettes and two unique targeting sequences (hook1 and hook2) (in *green*) homologous to the 5′ and 3′ end of a gene of interest. The size of hooks may be as small as 60 bp (Noskov et al. 2001). The hook sequences may either be both unique, or one of the hook sequences may be a common repeat (for example, an *Alu* repeat for cloning from human genomic DNA). The TAR vector DNA is linearized by a unique endonuclease located between the hooks to expose targeting sequences. If necessary, genomic DNA may be treated by CRISPR/Cas9 endonuclease before yeast transformation, which will greatly increase the yield of gene/region-positive TAR clones. Step 2: Genomic DNA and a linearized TAR vector are cotransformed into yeast *Saccharomyces cerevisiae* cells. Steps 3 and 4: Recombination between targeting sequences in the vector and the targeted sequences in the genomic DNA fragment leads to rescue of the fragment (or gene) as a circular TAR/YAC/BAC molecule. Step 5: Transferring TAR-isolated molecules containing a region of interest from yeast cells to bacterial cells by electroporation. Step 6: BAC DNA is isolated by a standard procedure for further sequencing or functional analysis

For complex genomes such as human, the basic TAR cloning protocol yields between 0.5 and 5 % of the target gene-positive clones (Kouprina and Larionov 2006), while for simple microbe genomes such as the protist *Trypanosoma brucei*, *Trypanosoma equiperdum*, or *Plasmodium falciparum*, the yield is much higher. For example, the yield of the TAR-cloned *P. falciparum* var genes reached up to 48 % (Gaida et al. 2011), while TAR cloning of the repertoires of VSG expression site containing telomeres isolated from

Fig. 2 Multiple applications of the TAR cloning technology
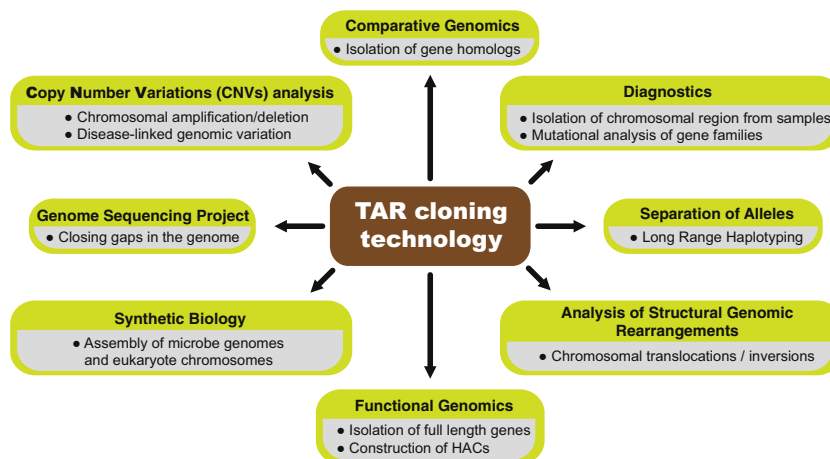
T. equiperdum resulted in a yield as high as 18 % (Young et al. 2008). Recently, we significantly increased the fraction of gene-positive clones isolated from complex genomes up to 32 % by using Cas9 nucleases that create specific double-strand breaks bracketing the target genomic DNA sequence, which makes the ends of the targeted region highly recombinogenic (Lee et al. 2015) (Fig. 1).

TAR cloning of genes and gene clusters from simple genomes

At present, TAR cloning is routinely used for isolation of genes and gene clusters from simple genomes such as bacteria, protist, and virus. The method was also adopted for rapid construction of biochemical pathways. An example of engineering of viral genomes in yeast is assembly of modular viral scaffolds for targeted bacterial population editing (Ando et al. 2015). Examples of direct capture of genes or gene clusters for functional studies or commercial purposes include cloning of the repertoire of individual *P. falciparum var* genes up to 30 kb in size (Gaida et al. 2011), the *Pseudoalteromonas alterochromide* gene cluster of 34 kb in size controlling the production of alterochromide lipopeptides (Ross et al. 2015), the nataxazole biosynthesis gene cluster for heterologous expression into *S. albus* J1074 and *S. lividans* JT46 (Cano-Prieto et al. 2015), the lipopeptide gene cluster of 73 kb in size from the marine actinomycete *Saccharomonospora sp.* to yield the antibiotic taromycin A (Yamanaka et al. 2014), the *Salinispora pacifica* natural product gene cluster of 21.3 kb in size for the biosynthesis of enterocin (Bonet et al. 2015), the amicoumacin biosynthetic gene cluster from the marine *Bacillus subtilis* 1779 for antibiotic production in *Bacillus* host (Li et al. 2015a), as well as identification of thiotetronic acid antibiotic biosynthetic pathways in *Salinispora* (Tang et al. 2015), isolation of genes controlling production of colibactin in *E. coli* to clarify the biosynthetic pathway to (pre)colibactin (Li et al. 2015b), and isolation of the repertoire of VSG expression site containing telomeres from *Trypanosoma brucei gambiense*, *T. b. brucei*, and

T. equiperdum (Becker et al. 2004; Young et al. 2008). One very interesting work presents a general experimental framework that permits the recovery of large natural product biosynthetic gene clusters on overlapping soil-derived eDNA (environmental) clones and the resulting reassembly of these large gene clusters (~90 kb) using transformation-associated recombination (TAR) in yeast (Kim et al. 2010; Feng et al. 2011). The application of TAR cloning for rapid cloning and assembly of biosynthetic gene clusters from collections of overlapping eDNA clones is an important step toward being able to functionally study larger natural product gene clusters from uncultured bacteria. Large-scale functional cloning followed by screening of resulting eDNA clones should be a productive strategy for generating previously structurally uncharacterized chemical entities for use in future drug development efforts. TAR cloning has also been adapted to serve as a general platform for the engineering of large metabolic pathways from basic parts, which is summarized in the referenced publications (Shao et al. 2009; Ongley et al. 2013; Agarwal et al. 2014; Mizutani 2015; Yuan et al. 2016). To summarize, TAR cloning from microbial genomes helps to better understand the evolutionary and physiological role of natural products, their structural/functional features, mechanisms of action, and biosynthesis, which can be used in the future to address fundamental questions in environmental evolution and biotechnology (Giordano et al. 2015).

TAR cloning of full-size genes and entire loci from complex genomes

Complex genomes include animal and plant genomes. Because TAR cloning is applicable for multiple genomic samples and clinical material, it is becoming a powerful tool for functional studies in complex genomes. Over the past few years, TAR cloning has been used to isolate unique regions, full-size genes, and gene clusters up to 250 kb from genomes of the human, nonhuman primate, and mouse (Kouprina and Larionov 2006; Kim et al. 2011; Kononenko et al. 2014). Physical analysis has demonstrated a high fidelity of the

material isolated by TAR cloning. For example, by transfecting the isolates into mammalian cells, the functionality of several genes was confirmed, including the breast cancer gene *BRCA1* (Annab et al. 2000; Kononenko et al. 2014), the human *HPRT* gene (Kouprina et al. 1998), the tumor suppressor gene *KAI1* (Kouprina and Larionov 2006), the metastasis suppressor gene *TEY1* (Nihei et al. 2002), the human *TERT* gene coding for a catalytic subunit of human telomerase (Leem et al. 2002), and the cancer-associated genes *VHL* (mutated in von Hippel–Lindau syndrome) and *NBS1* (mutated in Nijmegen breakage syndrome (Kim et al. 2011). Further evidence of the accuracy of the TAR cloning was obtained by isolation and shot-gun sequencing of the genomic regions containing the sperm protein associated with the nucleus on the X chromosome (*SPANX*) genes (Kouprina et al. 2004a, 2005a, 2007a, b, 2012). The sequences of one of the *SPANX* gene family members, *SPANX-C*, isolated from several individuals differed only by a single nucleotide substitution due to natural polymorphism (Kouprina et al. 2005a). TAR-isolated genes can also be success-fully used in transgenesis. For example, transgenic mice carrying the entire 50 kb *hTERT* locus were constructed and used to show that in vivo expression of human and mouse *TERT* genes differ significantly, raising a concern about the use of mouse models for human cancer and aging (Horikawa et al. 2005).

Separation of gene alleles and long-range molecular haplotyping

Haplotype analysis is a common tool used in association stud-ies. The joint distribution of adjacent markers in a population, which is the haplotype frequency, represents the correlation between those markers. A major difficulty in using haplotype markers lies in determining the haplotype phase for individ-uals heterozygous for more than one marker. Classically, hap-lotypes are resolved by pedigree analysis, but this approach is limited by having to collect DNA samples from appropriate family members and by meiotic recombination. Separation of individual chromosomes by construction of hybrid cell lines, microdissection of chromosomes, or amplification of sper-matocyte DNA are also laborious and time-consuming and cannot be used for a large number of individuals. At present, many computer programs are available for haplotype analysis of samples of unrelated individuals (Salem et al. 2005; Kuleshov et al. 2014; Wang et al. 2015), though all of them have limitations and are difficult in resolving the phase of maternal and paternal chromosomes. TAR cloning can be used to unambiguously resolve the "haplotype problem." Both parental alleles of a gene can be quickly isolated in one TAR cloning experiment because homologous recombination between targeting hooks and the target sequence occurs at equal frequencies in both chromosomes (Fig. 3).

One good example of the application of TAR cloning for this purpose is separation of the alleles of the *hTERT* gene (Kim
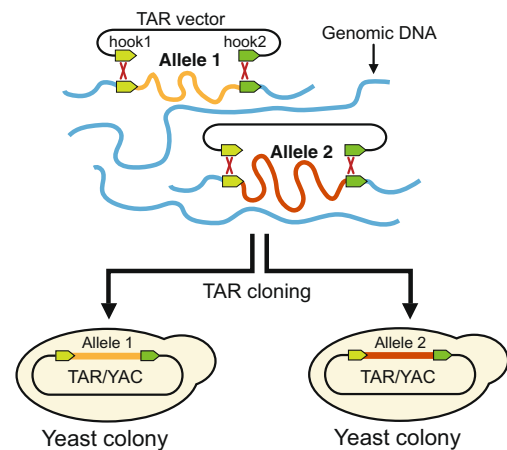

Fig. 3 TAR cloning for separation of alleles and long-range haplotyping. TAR cloning experiments result in isolation of multiple gene-positive clones. Thus, the genomic region corresponding to different alleles becomes separated into different yeast cells, allowing for independent analysis

et al. 2003). To prove that different alleles were indeed isolated, we examined the variable number of tandem repeat (VNTR) blocks of the TAR isolates. As known, the *hTERT* gene con-tains four VNTR blocks—two in intron 2 and two in intron 6. Earlier analysis of the segregation of these VNTRs in families revealed that all were transmitted through meiosis and followed a Mendelian inheritance pattern. Analysis of TAR isolates con-taining parental *hTERT* alleles determined the specific combi-nation of minisatellites at each of the polymorphic sites.

An example of the application of TAR cloning for long-range molecular haplotypying is isolation of the *SPANX* genes from different individuals (Kouprina et al. 2007a, b, 2012). This case is very complicated because sequences of these gene members have a high level of homology and reside within large segmental duplications with >95 % identity. Therefore, their mutational analysis could not be carried out by a routine PCR analysis or by the next-generation sequencing. We ap-plied a TAR-cloning-based approach by cloning each member of this *SPANX* gene family (*SPANX-A1*, *SPANX-A2*, *SPANX-B*, *SPANX-C*, and *SPANX-D*). Analysis of the TAR clones re-vealed that roughly two thirds of *SPANX* sequence variants resulted from gene conversion suggesting a high frequency of recombination between the genes. Based on traces of gene conversion detected in the *SPANX-C* locus, it was concluded that recombinational interaction operates over a long distance (*SPANX-D* and *SPANX-C* genes are separated by more than 500 kb). In most cases, reports of gene conversion in the human genome are based on comparison of DNA sequences obtained from different individuals. In the case of *SPANX* genes, donor and targeted sequences were isolated from the same individual. Therefore, our data provided a unique oppor-tunity to analyze the mechanism of gene conversion in humans and build long-range *SPANX* haplotypes in individ-uals. Analogously, TAR cloning may be applicable to other

gene families and large segmental duplications (SDs), whose analysis is prohibited by a high level of homology between family members.

For a long time, the cluster of *SPANX* genes was considered to play a role in prostate cancer. Direct isolation of a set of overlapping genomic segments of the 750-kb region in X-linked families with predisposition to prostate cancer revealed no disease-specific alterations within the *SPANX* gene cluster (Kouprina et al. 2012). Thus, TAR cloning allowed exclusion of the 750-kb genetically unstable region at Xq27 as a candidate locus for prostate malignancy.

To summarize, TAR cloning is well suited for separation of alleles and for a large-scale analysis of long-range haplotypes in multiply heterozygous individuals and can also be used to identify haplotypes that may contribute to disease.

Isolation of gene homologues for evolutionary studies

Since up to 15 % divergence in DNA sequences does not prevent recombination during transformation in yeast (Noskov et al. 2003b), TAR cloning can be used to isolate gene homologs. This technique was used to reconstruct the evolutionary history of several disease-associated genes: the breast cancer gene *BRCA1* (Pavlicek et al. 2004), the microcephaly gene *ASPM* (Kouprina et al. 2004b), the cancer/testis-specific antigen gene family *SPANX* (Kouprina et al. 2004a, 2005a), and the *ATM* and *NBS1* genes involved in DNA repair (NK, unpublished observations). Complete genomic copies of these genes from human, chimpanzee, gorilla, bonobo, orangutan, and rhesus macaque genomes were isolated using TAR vectors containing human-specific targeting sequences developed from 5′ and 3′ gene-flanking regions. Analysis of cloned sequence regions revealed that some genes have evolved under strong positive selection. For example, *ASPM*, which encodes a centrosomal protein (Kouprina et al. 2005b), showed accelerated evolution in the African hominoid clade, preceding hominid brain expansion by several million years (Kouprina et al. 2004b). These data suggest that the evolutionary selection of specific segments of the *ASPM* sequence strongly relates to differences in cerebral cortical size. Analysis of the non-synonymous/synonymous ratio in the *BRCA1* gene revealed that most of the internal *BRCA1* sequence is variable between primates and has evolved under positive selection. In contrast, the terminal regions of *BRCA1*, which encode the RING finger and BRCT domains, experienced negative selection, which left them almost identical between the compared primates (Pavlicek et al. 2004). Finally, on the basis of protein sequence conservation, we have been able to identify missense changes that are likely to compromise BRCA1 function.

Comparison of TAR-isolated, full-length gene homologs also provides information on the evolution of noncoding regions. An increasing body of evidence indicates the importance of noncoding regions in regulating gene expression and their involvement in genomic rearrangements resulting in gene inactivation. For example, a significant fraction of germ-line *BRCA1* mutations, which cause hereditary predisposition to breast and ovarian cancers, are deletions and duplications that involve one or more exons. Most of them are caused by recombination between *Alu* repeats, which are particularly numerous in *BRCA1*. Sequence analysis of full-size *BRCA1* homologs isolated by TAR cloning from a representative group of nonhuman primates revealed that *Alu*-mediated rearrangements, including *Alu* transpositions and *Alu*-associated deletions, are the major forces of evolutionary changes in noncoding *BRCA1* sequences (Pavlicek et al. 2004). Moreover, analysis of *BRCA1* TAR-isolated clones indicated that the structural instability of the locus might be an intrinsic feature of anthropoids. Most of the *Alu* repeats involved in disease-associated genomic rearrangements have been retained in nonhuman primates, suggesting that the repeats are of functional significance.

A good example of the application of TAR cloning for evolutionary analysis is reconstruction of evolutionary history of the *SPANX-A/D* gene family in primates (Kouprina et al. 2004a, 2005a). Rapid evolution of these genes and their location within segmental duplications (SDs) impede a routine PCR or next-generation sequencing analyses of syntenic chromosomal segments that is required to detect lineage-specific amplification of these genes. To overcome this problem, we selectively isolated syntenic segments from human, chimpanzee, bonobo, gorilla, orangutan, and macaque. Sequence comparison revealed that TAR clones from syntenic regions of chimpanzee, bonobo, and gorilla genomes do not contain the *SPANX-C* gene, meaning that the *SPANX-C* gene is human specific. Analysis of the *SPANX-B* locus revealed a variable number of gene copies ranging from 1 to 14 due to the tandem duplication of a 12-kb DNA segment carrying the *SPANX-B* gene present only in humans. Moreover, the analysis revealed that the *SPANX-A/D* gene family is absent in orangutan and macaque (Fig. 4).

Isolation of chromosomal regions which are unclonable in bacterial vectors

It is well known that some genomic regions are missing in the existing BAC libraries because they cannot be cloned efficiently, if at all, in *Escherichia coli* cells. Such regions may include long inverted repeats and sequences with Z-DNA-like structures that are extremely unstable in *E. coli*. Poorly clonable human sequences include both non-coding and coding regions. Human DNA sequences that are unstable or unclonable in *E. coli* could be cloned and propagated in yeast using TAR-generated YACs. Two genes that are toxic for bacterial cells (*MUC2* and *KAI1*) have been TAR-cloned in yeast and then sequenced. Resequencing of these regions
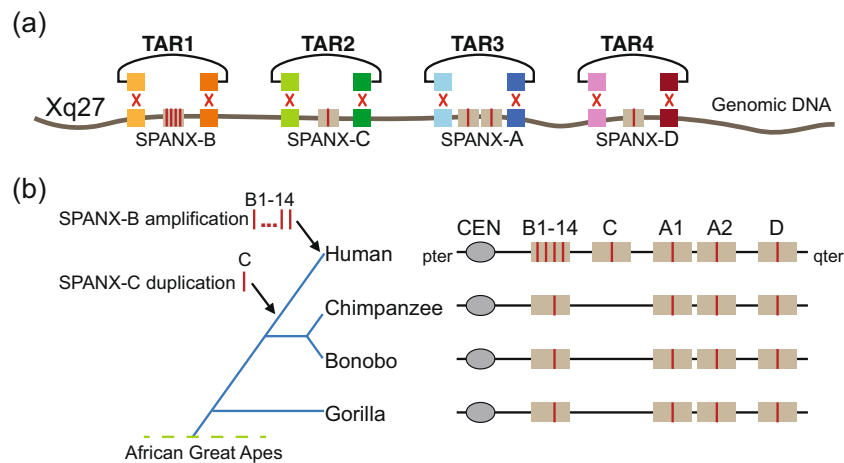
Fig. 4 Organization of the *SPANX* genes in primates. Syntenic genomic fragments containing different members of the *SPANX* gene family from the human, chimpanzee, bonobo, gorilla, orangutan, and macaque were isolated by TAR cloning. Unique targeting hooks in the TAR vectors were chosen from the available human genome sequences. Comparison of the sequences showed that TAR clones from chimpanzee, bonobo, and gorilla do not contain *SPANX-C* along with the duplication. In addition, the amplification of *SPANX-B* from one to 14 copies in different individuals is also human specific. The *SPANX-A/D* gene family is absent in the orangutan and macaque

showed that the errors in the draft genome sequence were the results of both missassembly and loss of specific DNA sequences during cloning in *E. coli* (Kouprina et al. 2003a).

TAR cloning was put to use in the final phase of the human genome sequencing and significantly contributed to closing the gaps. As an example, several gap regions were selectively recovered from human chromosome 19 (Leem et al. 2004; Grimwood et al. 2004) by TAR cloning using sequence information of the flanking contigs. Analysis of the gap sequences revealed that they contain several abnormalities that could result in instability of the sequences in microbe hosts, including large blocks of microsatellites and minisatellites and a high density of *Alu* repeats. In addition, sequencing of these gap regions allowed us to generate a complete sequence of four genes, including the neuronal cell signaling gene *SCK1/ SLI* that contains a record number of minisatellites, most of which are polymorphic and transmitted through meiosis in a Mendelian fashion (Leem et al. 2004). The same recombinational cloning approach may be applied for isolation of genomic fragments from other genomes that are not clonable in *E. coli* vectors.

Construction of HACs for gene functional studies

Progress in functional genomics depends on both the availability of full-length genes and a suitable system for gene delivery. TAR cloning solved the problem of full-length gene isolation. During the past few years, there has also been great progress in construction of new gene delivery vectors, i.e., human artificial chromosomes or HACs (Harrington et al. 1997; Ikeno et al. 1998; Henning et al. 1999; Ebersole et al. 2000; Mejía et al. 2002; Ikeno et al. 2002; Kouprina et al. 2003a; Katoh et al. 2004; Kazuki and Oshimura 2011; Kouprina et al. 2013, 2014;

Oshimura et al. 2015). HACs are extra chromosomes carrying all of the required components of a functional kinetochore assembled on a large block of centromeric alphoid DNA repeats. Therefore, they are maintained stably as an additional 47th chromosome in human cells, thus avoiding insertional mutagenesis due to integration into host chromosomes. They do not cause severe immunogenic responses, which have been a serious problem with viral-based vectors (Lufino et al. 2008; Buchholz et al. 2009; Wanisch and Yáñez-Muñoz 2009; Epstein 2009; Maier et al. 2010; Mingozzi et al. 2011; Zhang et al. 2014). HACs have essentially unlimited cloning capacity and may carry one or several therapeutic genes surrounded by their long-range controlling elements that should confer physiological levels of fully regulated gene expression (Tedesco et al. 2011; Kim et al. 2011; Kononenko et al. 2014). When regulatory elements such as enhancers, for example, are too far away from the gene or even on a different chromosome, recapitulation of physiological regulation of endogenous loci can be achieved using a HAC vector containing several gene acceptor sites. A gene of interest may be inserted into one acceptor site, while the enhancer element is inserted into another (Suzuki et al. 2014). Several studies have demonstrated the efficacy of de novo assembled HACs for delivery and expression of full-size genes, including those isolated by TAR cloning, in human or embryonic cells with their further differentiated progeny in mice (Kim et al. 2011; Kononenko et al. 2014; Liskovykh et al. 2015; Oshimura et al. 2015).
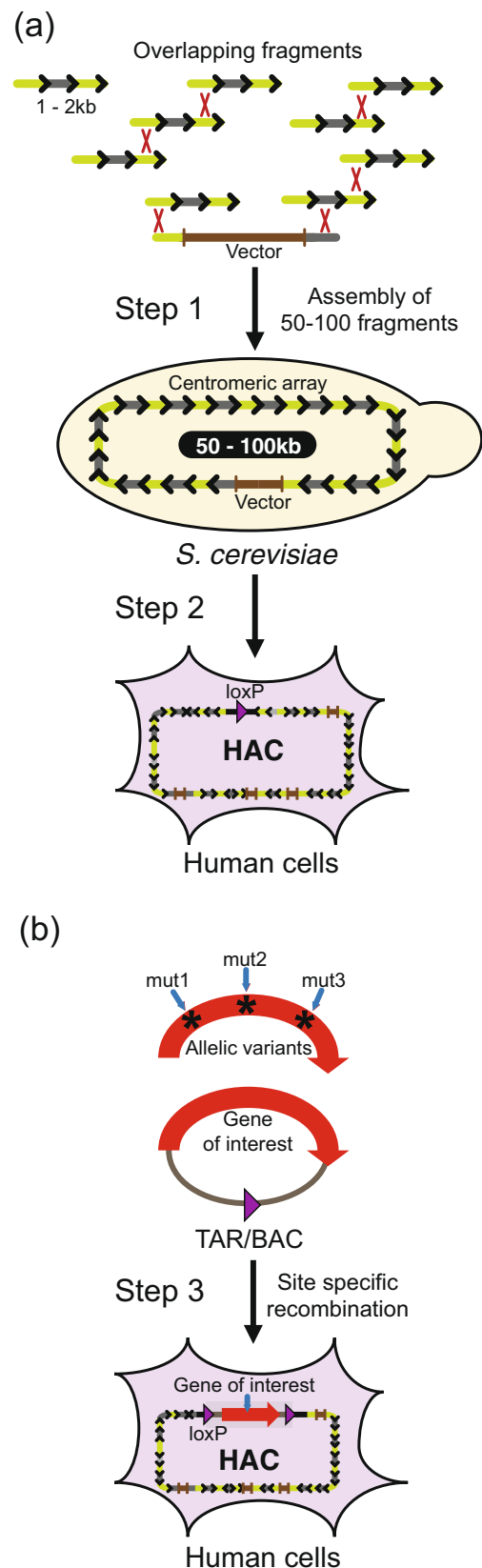
Existing strategies for HAC construction can be divided into two classes: top down, based on the truncation of an existing chromosome into a much smaller minichromosome suitable for further manipulation, and bottom up, based on transfection of large blocks of centromeric alphoid DNA into human cells with resulting HAC formation (Oshimura et al.

Fig. 5 Coupling of TAR gene cloning technology with the HAC gene delivery and expression system. **a** Amplification of synthetic alphoid DNA repeats for HAC construction by TAR cloning. Step 1: Amplification involves two substeps: rolling-circle amplification (RCA) of 340 bp alphoid DNA dimers to 1–2 kb tandem repeats (direction of repeats is shown by *arrows*) and cotransformation of the TAR vector and a mixture of the overlapping tandem repeats followed by end-to-end recombination through interaction of the recombined fragments with the TAR vector hooks that results in the rescue of large alphoid arrays up to 100 kb in size as circular molecules. The vector contains a yeast cassette, a mammalian selectable marker (the *Neo* or *BS* gene), and a BAC replicon. Step 2: After transfection into human cells, both the synthetic and natural arrays are capable of de novo HAC formation. To convert a HAC into a gene delivery vector, a unique loxP gene loading site is inserted. **b** Step 3: Loading of a TAR-isolated gene allele (mut1, mut2, or mut3) of a gene of interest into the loxP site of the HAC gene delivery vector by Cre-loxP-mediated recombination

2015; Katona 2015; Moralli and Monaco 2015). TAR cloning allows for the construction of synthetic arrays of alphoid DNA with a predetermined structure, which may be then used as substrates for HAC formation (Ebersole et al. 2005). This approach starts with a short alphoid DNA multimer (e.g., a dimer) that is assembled by in vivo homologous recombination in yeast (TAR cloning) into alphoid arrays up to 150 kb in size (Fig. 5a) (Ebersole et al. 2005). The minimum size of the array that is required for de novo HAC formation is ~40 kb. After transfection of such synthetic arrays (or natural alphoid arrays cloned by TAR cloning (Kouprina et al. 2003b)) into human cells, de novo HACs are generated, ranging in size from 1 to 10 Mb due to amplification of the input alphoid DNA (Fig. 5a). Thus, the use of TAR cloning advances construction of HACs with a predetermined structure, which is very useful for analysis of the human kinetochore (Bergmann et al. 2012; Ohzeki et al. 2015).

The most advanced de novo constructed HAC is the alphoid^tetO^-HAC engineered from a ~50-kb synthetic alphoid DNA array, in which the tetracycline operator (tet-O) sequences were embedded (Nakano et al. 2008). Therefore, this HAC contains a conditional kinetochore that can be inactivated by expression of tet-repressor fusion proteins, resulting in loss of the HAC from populations of dividing cells. It was subsequently demonstrated that the alphoid^tetO^-HAC may carry a genomic copy of a TAR-cloned gene (either *VHL*, *NBS1*, *BRCA1*, or *HPRT*) (Fig. 5b) (Kim et al. 2011; Kononenko et al. 2014), which if necessary may be successfully eliminated from cells by loss of the HAC following kinetochore inactivation.

In the recent study, the HAC vector containing a regulated centromere was used for delivery and expression of the TAR-cloned 90-kb genomic copy of the *BRCA1* gene into the gene-deficient human cells (Kononenko et al. 2014). The use of this system demonstrated that deficiency in BRCA1 results in an elevated level of transcription of diverged pericentromeric repeats, forming constitutive heterochromatin, as well as higher-order alphoid DNA repeats (HORs), forming a functional kinetochore. These data support the hypothesis that epigenetic alterations of these regions initiated in the absence of BRCA1

could result in kinetochore disassembly, leading to chromosome instability and cell transformation.

Thus, the alphoid^tetO-HAC vector has significant advantages because it provides a mechanism to compare the phenotype of human target cells with or without a functional copy of any cloned gene of interest, a feature that provides a control for phenotypic changes attributed to the expression of HAC-encoded genes. This generation of human artificial chromosomes from synthetic alphoid arrays seems to be the most suitable for studies of gene function and therapeutic applications.

TAR cloning for synthetic biology

The recent progress in synthetic genomics was determined to a certain extent by the use of TAR cloning. A few years ago, the J Craig Venter Institute had attempted complete synthesis and assembly of a whole bacterial (*Mycoplasma genitalium*) genome (582,970 bp), a culmination of about 10 years of work (Gibson et al. 2008a). The *M. genitalium* genome was assembled using a combination of in vitro enzymatic and in vivo TAR cloning methods. In the early stage, an in vitro recombination method was utilized to assemble 25 DNA cassettes with an average length of 24 kb into eight 72-kb assemblies and subsequently assembled into four 144-kb assemblies. It was found that the efficiency of the in vitro procedure greatly declined as the assemblies became larger, and the half-genomes, each of 290 kb in size, were unable to be assembled at all. Therefore, the in vivo recombination in yeast was exploited to complete the final whole genome assembly. In the meantime, it was also found possible to directly assemble 25 DNA cassettes from the earliest stages into a complete genome in a single step using TAR cloning in yeast (Gibson et al. 2008b). Thus, use of TAR cloning greatly simplifies the assembly of large DNA molecules from both synthetic and natural fragments.

Subsequently, Gibson et al. (2010) reported the creation of a bacterial cell controlled by a chemically synthesized genome. A 1.1 Mb *M. mycoides* genome was assembled by TAR cloning from 11 ~100-kb DNA segments in yeast (Lartigue et al. 2009); it was then transplanted into a closely related *M. capricolum* recipient cell to form a new *M. mycoides* cell that was controlled solely by the synthetic genome. The new cells have expected phenotypic properties and are capable of continuous self-replication. Thus, the full circle starting from de novo synthesis of a whole mycoplasma genome and ending with its transfer into a bacterial cell has been accomplished. These organisms were first selected for genome cloning/replacements because of their small genome size and also because of the expectation that the genomes were unlikely to produce toxic gene products in yeast as a result of their alternative genetic code.

So far, cloning in yeast of whole genomes or individual chromosomes was successfully accomplished for species such as *M. pneumoniae* (0.8 Mb) (Benders et al. 2010), *Acholeplasma laidlawii* (1.5 Mb) (Karas et al. 2012), *Prochlorococcus marinus MED4* (1.6 Mb) (Tagwerker et al. 2012), and eukaryotic algal *Phaeodactylum tricornutum* (27.4 Mb) (Karas et al. 2013). In the last case, two individual chromosomes with a size of ~500 kb were assembled by TAR cloning in yeast. At present, propagation in yeast has been hampered for the *Acholeplasma laidlawii* genome due to the presence of a single gene that is toxic to yeast cells. Also, a high G+C content (55 %) prevents propagation in yeast of bacterial fragments larger than 200 kb. It becomes a possibility if an additional *ARS* sequence(s) is added (Noskov et al. 2012; Karas et al. 2013).

In contrast to most other eukaryotic organisms, the introduction of targeted and specific modifications such as gene deletions, insertions, or replacements into the genome of *Saccharomyces cerevisiae* has been standard for decades (Hinnen et al. 1978; Scherer and Davis 1979). Therefore, the reported assembly of microbe genomes in yeast provides an unprecedented opportunity for their further modifications in order to address key questions of synthetic biology.

The yeast *S. cerevisiae* is the most intensively studied unicellular eukaryote and one of the main industrial microorganisms used in the production of biochemicals. The potential of yeast as a powerful host for synthetic biology has already been successfully demonstrated by both basic research, namely the de novo synthesis of a complete chromosome (Annaluru et al. 2014), and the application-oriented engineering of complex pathways, such as the synthesis of amorphadiene and vanillin (Brochado and Patil 2013; Westfall et al. 2012). Recently, the TAR cloning strategy has been applied for assembling genetic pathways for expression in *S. cerevisiae* (Mitchell et al. 2015). The authors demonstrated the assembly of four-, five-, and six-gene pathways to generate *S. cerevisiae* cells synthesizing β-carotene and violacein.

Currently, the most ambitious project in yeast synthetic biology is the complete de novo synthesis of all 16 chromosomes, Sc2.0. Although this project is at its very beginning, with one synthetic chromosome (chromosome III) completed (Annaluru et al. 2014), the consortium plans to finish all additional chromosomes by 2019. It is worth noting that the assembly of overlapping DNA fragments by TAR cloning was a critical step in de novo synthesis of the first yeast chromosomes (Annaluru et al. 2014).

To conclude, TAR cloning has emerged an essential method for the assembly of new microbe genomes and synthetic biology. Additional information about strategies for cloning and manipulation of natural and synthetic chromosomes can be found in recent publications (Gibson 2014; Karas et al. 2015).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest. This article does not contain any studies with human participants or animals performed by any of the authors.

References

Agarwal V, El Gamal AA, Yamanaka K, Poth D, Kersten RD, Schorn M, Allen EE, Moore BS (2014) Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. Nat Chem Biol 10: 640–U182

Ando H, Lemire S, Pires DP, Lu TK (2015) Engineering modular viral scaffolds for targeted bacterial population editing. Cell Syst 1:187–196

Annab L, Kouprina N, Solomon G, Cable L, Hill D, Barrett JC, Larionov V, Afshari C (2000) Isolation of functional copy of the human BRCA1 gene by TAR cloning in yeast. Gene 250:201–208

Annaluru N, Muller H, La M, Ramalingam S, Stracquadanio G, Richardson SM, Dymond JS, Kuang Z, Scheifele LZ, Cooper EM, Cai Y, Zeller K, Agmon N, Han JS, Hadjithomas M, Tullman J, Caravelli K, Cirelli K, Guo Z, London V, Yeluru A, Murugan S, Kandavelou K, Agier N, Fischer G, Yang K, Martin JA, Bilgel M, Bohutski P, Boulier KM, Capaldo BJ, Chang J, Charoen K, Choi WJ, Deng P, DiCarlo JE, Doong J, Dunn J, Feinberg JI, Fernandez C, Floria CE, Gladowski D, Hadidi P, Ishizuka I, Jabbari J, Lau CY, Lee PA, Li S, Lin D, Linder ME, Ling J, Liu J, Liu J, London M, Ma H, Mao J, McDade JE, McMillan A, Moore AM, Oh WC, Ouyang Y, Patel R, Paul M, Paulsen LC, Qiu J, Rhee A, Rubashkin MG, Soh IY, Sotuyo NE, Srinivas V, Suarez A, Wong A, Wong R, Xie WR, Xu Y, Yu AT, Koszul R, Bader JS, Boeke JD, Chandrasegaran S (2014) Total synthesis of a functional designer eukaryotic chromosome. Science 344:55–58

Asakawa S, Abe I, Kudoh Y, Kishi N, Wang Y, Kubota R, Kudoh J, Kawasaki K, Minoshima S, Shimizu N (1997) Human BAC library: construction and rapid screening. Gene 191:69–79

Becker M, Aitcheson N, Byles E, Wickstead B, Louis E, Rudenko G (2004) Isolation of the repertoire of VSG expression site containing telomeres of Trypanosoma brucei 427 using transformation-associated recombination in yeast. Genome Res 14:2319–2329

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, Merryman C, Venter JC, Smith HO, Glass JI, Hutchison CA 3rd (2010) Cloning whole bacterial genomes in yeast. Nucleic Acids Res 38:2558–2569

Bergmann JH, Martins NM, Larionov V, Masumoto H, Earnshaw WC (2012) HACking the centromere chromatin code: insights from human artificial chromosomes. Chromosome Res 20:505–519

Bonet B, Teufel R, Crüsemann M, Ziemert N, Moore BS (2015) Direct capture and heterologous expression of Salinispora natural product genes for the biosynthesis of enterocin. J Nat Prod 78:539–542

Brochado AR, Patil KR (2013) Overexpression of O-methyltransferase leads to improved vanillin production in baker's yeast only when complemented with model-guided network engineering. Biotechnol Bioeng 110:656–659

Buchholz CJ, Mühlebach MD, Cichutek K (2009) Lentiviral vectors with measles virus glycoproteins—dream team for gene transfer? Trends Biotechnol 27:259–265

Cano-Prieto C, García-Salcedo R, Sánchez-Hidalgo M, Braña AF, Fiedler HP, Méndez C, Salas JA, Olano C (2015) Genome mining of streptomyces sp. Tü 6176: characterization of the nataxazole biosynthesis pathway. Chembiochem 16:1461–1473

Ebersole TA, Ross A, Clark E, McGill N, Schindelhauer D, Cooke H, Grimes B (2000) Mammalian artificial chromosome formation from circular alphoid input DNA does not require telomere repeats. Hum Mol Genet 9:1623–1631

Ebersole T, Okamoto Y, Noskov VN, Kouprina N, Kim JH, Leem SH, Barrett JC, Masumoto H, Larionov V (2005) Rapid generation of long synthetic tandem repeats and its application for analysis in human artificial chromosome formation. Nucleic Acids Res 33, e130

Epstein AL (2009) Progress and prospects: biological properties and technological advances of herpes simplex virus type 1-based amplicon vectors. Gene Ther 16:709–715

Feng Z, Kallifidas D, Brady SF (2011) Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. Proc Natl Acad Sci U S A 108: 12629–12634

Gaida A, Becker MM, Schmid CD, Bühlmann T, Louis EJ, Beck HP (2011) Cloning of the repertoire of individual Plasmodium falciparum var genes using transformation associated recombination (TAR). PLoS One 6, e17782

Gibson DG (2014) Programming biological operating systems: genome design, assembly and activation. Nat Methods 11:521–526

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, Glass JI, Venter JC, Hutchison CA 3rd, Smith HO (2008a) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science 319:1215–1220

Gibson DG, Benders GA, Axelrod KC, Zaveri J, Algire MA, Moodie M, Montague MG, Venter JC, Smith HO, Hutchison CA 3rd (2008b) 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 105:20404–20409

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, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA 3rd, Smith HO, Venter JC (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329:52–56

Giordano D, Coppola D, Russo R, Denaro R, Giuliano L, Lauro FM, di Prisco G, Verde C (2015) Marine microbial secondary metabolites: pathways, evolution and physiological. Adv Microb Physiol 66: 357–428

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, Bajorek E, Black S, Branscomb E, Caenepeel S, Carrano A, Caoile C, Chan YM, Christensen M, Cleland CA, Copeland A, Dalin E, Dehal P, Denys M, Detter JC, Escobar J, Flowers D, Fotopulos D, Garcia C, Georgescu AM, Glavina T, Gomez M, Gonzales E, Groza M, Hammon N, Hawkins T, Haydu L, Ho I, Huang W, Israni S, Jett J, Kadner K, Kimball H, Kobayashi A, Larionov V, Leem SH, Lopez F, Lou Y, Lowry S, Malfatti S, Martinez D, McCready P, Medina C, Morgan J, Nelson K, Nolan M, Ovcharenko I, Pitluck S, Pollard M, Popkie AP, Predki P, Quan G, Ramirez L, Rash S, Retterer J, Rodriguez A, Rogers S, Salamov A, Salazar A, She X, Smith D, Slezak T, Solovyev V, Thayer N, Tice H, Tsai M, Ustaszewska A, Vo N, Wagner M, Wheeler J, Wu K, Xie G, Yang J, Dubchak I, Furey

TS, DeJong P, Dickson M, Gordon D, Eichler EE, Pennacchio LA, Richardson P, Stubbs L, Rokhsar DS, Myers RM, Rubin EM, Lucas SM (2004) The DNA sequence and biology of human chromosome 19. Nature 428:529–535

Harrington JJ, Van Bokkelen G, Mays RW, Gustashaw K, Willard HF (1997) Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. Nat Genet 15:345–355

Henning KA, Novotny EA, Compton ST, Guan XY, Liu PP, Ashlock MA (1999) Human artificial chromosomes generated by modification of a yeast artificial chromosome containing both human alpha satellite and single-copy DNA sequences. Proc Natl Acad Sci U S A 96:592–597

Hinnen A, Hicks JB, Fink GR (1978) Transformation of yeast. Proc Natl Acad Sci U S A 75:1929–1933

Horikawa I, Chiang JC, Patterson T, Feigenbaum L, Leem SH, Michishita E, Larionov V, Hodes RJ, Barrett JC (2005) Differential cis-regulation of human versus mouse TERT gene expression in vivo: identification of a human-specific repression element. Proc Natl Acad Sci U S A 102:18437–18442

Ikeno M, Grimes B, Okazaki T, Nakano M, Saitoh K, Hoshino H, McGill NI, Cooke H, Masumoto H (1998) Construction of YAC-based mammalian artificial chromosomes. Nat Biotechnol 16:431–439

Ikeno M, Inagaki H, Nagata K, Morita M, Ichinose H, Okazaki T (2002) Generation of human artificial chromosomes expressing naturally controlled guanosine triphosphate cyclohydrolase I gene. Genes Cells 7:1021–1032

Karas BJ, Tagwerker C, Yonemoto IT, Hutchison CA 3rd, Smith HO (2012) Cloning the Acholeplasma laidlawii PG-8A genome in Saccharomyces cerevisiae as a yeast centromeric plasmid. ACS Synth Biol 1:22–28

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, Smith HO, Venter JC, Weyman PD (2013) Assembly of eukaryotic algal chromosomes in yeast. J Biol Eng 7:30

Karas BJ, Suzuki Y, Weyman PD (2015) Strategies for cloning and manipulating natural and synthetic chromosomes. Chromosome Res 23:57–68

Katoh M, Ayabe F, Norikane S, Okada T, Masumoto H, Horike S, Shirayoshi Y, Oshimura M (2004) Construction of a novel human artificial chromosome vector for gene delivery. Biochem Biophys Res Commun 321:280–290

Katona RL (2015) De novo formed satellite DNA-based mammalian artificial chromosomes and their possible applications. Chromosome Res 23:143–157

Kazuki Y, Oshimura M (2011) Human artificial chromosomes for gene delivery and the development of animal models. Mol Ther 19:1591–1601

Kim JH, Leem SH, Sunwoo Y, Kouprina N (2003) Separation of long-range human TERT gene haplotypes by transformation-associated recombination cloning in yeast. Oncogene 22:2443–2456

Kim JH, Feng Z, Bauer JD, Kallifidas D, Calle PY, Brady SF (2010) Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. Biopolymers 93:833–844

Kim JH, Kononenko A, Erliandri I, Kim TA, Nakano M, Iida Y, Barrett JC, Oshimura M, Masumoto H, Earnshaw WC, Larionov V, Kouprina N (2011) Human artificial chromosome (HAC) vector with a conditional centromere for correction of genetic deficiencies in human cells. Proc Natl Acad Sci U S A 108:20048–20053

Kononenko AV, Bansal R, Lee NC, Grimes BR, Masumoto H, Earnshaw WC, Larionov V, Kouprina N (2014) A portable BRCA1-HAC (human artificial chromosome) module for analysis of BRCA1 tumor suppressor function. Nucleic Acids Res 42(21)

Kouprina N, Larionov V (2003) Exploiting the yeast Saccharomyces cerevisiae for the study of the organization and evolution of complex genomes. FEMS Microbiol Rev 27:629–649

Kouprina N, Larionov V (2006) TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. Nat Rev Genet 7:805–812

Kouprina N, Larionov V (2008) Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast Saccharomyces cerevisiae. Nat Protoc 3:371–377

Kouprina N, Annab L, Graves J, Afshari C, Barrett JC, Resnick MA, Larionov V (1998) Functional copies of a human gene can be directly isolated by TAR cloning with a small 3′ end target sequence. Proc Natl Acad Sci U S A 95:4469–4474

Kouprina N, Leem SH, Solomon G, Ly A, Koriabine M, Otstot J, Pak E, Dutra A, Zhao S, Barrett JC, Larionov V (2003a) Segments missing from the draft human genome sequence can be isolated by TAR cloning in yeast. EMBO Rep 4:257–262

Kouprina N, Ebersole T, Koriabine M, Pak E, Rogozin IB, Katoh M, Oshimura M, Ogi K, Peredelchuk M, Solomon G, Brown W, Barrett JC, Larionov V (2003b) Cloning of human centromeres by transformation-associated recombination in yeast and generation of functional human artificial chromosomes. Nucleic Acids Res 31:922–934

Kouprina N, Mullokandov M, Rogozin IB, Collins NK, Solomon G, Otstot J, Risinger JI, Koonin EV, Barrett JC, Larionov V (2004a) 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 101:3077–3082

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 (2004b) Accelerated evolution of the ASPM gene controlling brain size begins prior to human brain expansion. PLoS Biol 2:653–663

Kouprina N, Pavlicek A, Noskov VN, Solomon G, Otstot J, Isaacs W, Carpten JD, Trent JM, Schleutker J, Barrett JC, Jurka J, Larionov V (2005a) Dynamic structure of the SPANX gene cluster mapped to the prostate cancer susceptibility locus HPCX at Xq27. Genome Res 15:1477–1486

Kouprina N, Pavlicek A, Collins NK, Nakano M, Noskov VN, Ohzeki JI, Mochida GH, Risinger JI, Goldsmitt P, Gunsior M, Solomon G, Gersch W, Kim JH, Barrett JC, Walsh CA, Jurka J, Masumoto H, Larionov V (2005b) The microcephaly ASPM gene is expressed in proliferating tissues and encodes for a mitotic spindle protein. Hum Mol Genetics 14:2155–2165

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, Herr JC, Jurka J, Lobanenkov V, Schlessinger D, Larionov V (2007a) Evolutionary diversification of SPANX-N sperm protein gene structure and expression. PLoS One 2, e359

Kouprina N, Noskov VN, Solomon G, Otstot J, Isaacs W, Xu J, Schleutker J, Larionov V (2007b) Mutational analysis of SPANX genes in families with X-linked prostate cancer. Prostate 67:820–828

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 (2012) Exclusion of the 750-kb genetically unstable region at Xq27 as a candidate locus for prostate malignancy in HPCX1-linked families. Genes Chromosomes Cancer 51:933–948

Kouprina N, Earnshaw WC, Masumoto H, Larionov V (2013) A new generation of human artificial chromosomes for functional genomics and gene therapy. Cell Mol Life Sci 70:1135–1148

Kouprina N, Tomilin AN, Masumoto H, Earnshaw WC, Larionov V (2014) Human artificial chromosome-based gene delivery vectors

for biomedicine and biotechnology. Expert Opin Drug Deliv 11: 517–535

Kuleshov V, Xie D, Chen R, Pushkarev D, Ma Z, Blauwkamp T, Kertesz M, Snyder M (2014) Whole-genome haplotyping using long reads and statistical methods. Nat Biotechnol 32:261–266

Kunes S, Botstein D, Fox MS (1985) Transformation of yeast with linearized plasmid DNA. Formation of inverted dimers and recombinant plasmid products. J Mol Biol 184:375–387

Larionov V, Kouprina N, Graves J, Chen XN, Korenberg JR, Resnick MA (1996) Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. Proc Natl Acad Sci U S A 93:491–496

Larionov V, Kouprina N, Solomon G, Barrett JC, Resnick MA (1997) Direct isolation of human BRCA2 gene by transformation-associated recombination in yeast. Proc Natl Acad Sci U S A 94: 7384–7387

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, Hutchison CA 3rd, Smith HO, Venter JC, Glass JI (2009) Creating bacterial strains from genomes that have been cloned and engineered in yeast. Science 325:1693–1696

Lee NC, Larionov V, Kouprina N (2015) Highly efficient CRISPR/Cas9-mediated TAR cloning of genes and chromosomal loci from complex genomes in yeast. Nucleic Acids Res 43, e55

Leem SH, Londono-Vallejo JA, Kim JH, Bui H, Tubacher E, Solomon G, Park JE, Horikawa I, Kouprina N, Barrett JC, Larionov V (2002) Cloning of the human telomerase gene: complete genomic sequence and analysis of tandem repeats polymorphisms in intronic regions. Oncogene 21:769–777

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, Barrett JC, Larionov V (2004) Closing the gaps on human chromosome 19 revealed genes with a high density of repetitive tandemly arrayed elements. Genome Res 14:239–246

Li Y, Li Z, Yamanaka K, Xu Y, Zhang W, Vlamakis H, Kolter R, Moore BS, Qian PY (2015a) Directed natural product biosynthesis gene cluster capture and expression in the model bacterium Bacillus subtilis. Sci Rep 5:9383

Li ZR, Li Y, Lai JY, Tang J, Wang B, Lu L, Zhu G, Wu X, Xu Y, Qian PY (2015b) Critical intermediates reveal new biosynthetic events in the enigmatic colibactin pathway. Chembiochem 16:1715–1719

Liskovykh M, Ponomartsev S, Popova E, Bader M, Kouprina N, Larionov V, Alenina N, Tomilin A (2015) Stable maintenance of de novo assembled human artificial chromosomes in embryonic stem cells and their differentiated progeny in mice. Cell Cycle 14: 1268–1273

Lufino MM, Edser PA, Wade-Martins R (2008) Advances in high-capacity extrachromosomal vector technology: episomal maintenance, vector delivery, and transgene expression. Mol Ther 16: 1525–1538

Ma H, Kunes S, Schatz PJ, Botstein D (1987) Plasmid construction by homologous recombination in yeast. Gene 58:201–216

Maier P, von Kalle C, Laufs S (2010) Retroviral vectors for gene therapy. Future Microbiol 5:1507–1523

Mejía JE, Alazami A, Willmott A, Marschall P, Levy E, Earnshaw WC, Larin Z (2002) Efficiency of de novo centromere formation in human artificial chromosomes. Genomics 79:297–304

Mingozzi F, Katherine A, High KA (2011) Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. Nat Rev Genet 12:341–356

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 (2015) Versatile genetic assembly system (VEGAS) to assemble pathways for expression in S. cerevisiae. Nucleic Acids Res 43:6620–6630

Mizutani K (2015) High-throughput plasmid construction using homologous recombination in yeast: its mechanisms and application to protein production for X-ray crystallography. Biosci Biotechnol Biochem 79:1–10

Moralli D, Monaco ZL (2015) Developing de novo human artificial chromosomes in embryonic stem cells using HSV-1 amplicon technology. Chromosome Res 23:105–110

Nakano M, Cardinale S, Noskov VN, Gassman R, Vagnarelli P, Kandels-Lewis S, Earnshaw WC, Larionov V, Masumoto H (2008) Inactivation of a human kinetochore by a specific targeting of chromatin modifiers. Dev Cell 14:507–522

Nihei N, Kouprina N, Larionov V, Oshima J, Martin GM, Ichikawa T, Barrett JC (2002) Functional evidence for a metastasis suppressor gene for rat prostate cancer within a 60 kb region on human chromosome 8p21-p12. Cancer Res 62:367–370

Noskov V, Koriabine M, Solomon G, Randolph M, Barrett JC, Leem SH, Stubbs L, Kouprina N, Larionov V (2001) Defining the minimal length of sequence homology required for selective gene isolation by TAR cloning. Nucleic Acids Res 29, E62

Noskov VN, Kouprina N, Leem SH, Ouspenski I, Barrett JC, Larionov V (2003a) A general cloning system to selectively isolate any eukaryotic or prokaryotic genomic region in yeast. BMC Genomics 4:16

Noskov VN, Leem SH, Solomon G, Mullokandov M, Chae JY, Yoon YH, Shin YS, Kouprina N, Larionov V (2003b) A novel strategy for analysis of gene homologues and segmental genome duplications. J Mol Evol 56:702–710

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, Venter JC, Weyman PD (2012) Assembly of large, high G+C bacterial DNA fragments in yeast. ACS Synth Biol 1:267–273

Ohzeki J, Larionov V, Earnshaw WC, Masumoto H (2015) Genetic and epigenetic regulation of centromeres: a look at HAC formation. Chromosome Res 23:87–103

Ongley SE, Bian X, Neilan BA, Müller R (2013) Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. Nat Prod Rep 30:1121–1138

Orr-Weaver TL, Szostak JW, Rothstein RJ (1981) Yeast transformation: a model system for the study of recombination. Proc Natl Acad Sci U S A 78:6354–6358

Oshimura M, Uno N, Kazuki Y, Katoh M, Inoue T (2015) A pathway from chromosome transfer to engineering resulting in human and mouse artificial chromosomes for a variety of applications to biomedical challenges. Chromosome Res 23:111–133

Pavlicek A, Noskov N, Kouprina N, Barrett JC, Jurka J, Larionov V (2004) Evolution of the tumor suppressor BRCA1 locus in primates: implications for cancer predisposition. Hum Mol Genet 13:1–15

Ross AC, Gulland LE, Dorrestein PC, Moore BS (2015) Targeted capture and heterologous expression of the Pseudoalteromonas alterochromide gene cluster in Escherichia coli represents a promising natural product exploratory platform. ACS Synth Biol 4:414–420

Salem RM, Wessel J, Schork NJ (2005) A comprehensive literature review of haplotyping software and methods for use with unrelated individuals. Hum Genomics 2:39–66

Scherer S, Davis RW (1979) Replacement of chromosomal segments with altered DNA sequences constructed in vitro. Proc Natl Acad Sci U S A 76:4951–4955

Shao Z, Zhao H, Zhao H (2009) DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res 37, e16

Stinchcomb DT, Thomas M, Kelly J, Selker E, Davis RW (1980) Eukaryotic DNA segments capable of autonomous replication in yeast. Proc Natl Acad Sci U S A 77:4559–4563

Suzuki T, Kazuki Y, Oshimura M, Hara T (2014) A novel system for simultaneous or sequential integration of multiple gene-loading

vectors into a defined site of a human artificial chromosome. PLoS One 9(10), e110404

Tagwerker C, Dupont CL, Karas BJ, Ma L, Chuang RY, Benders GA, Ramon A, Novotny M, Montague MG, Venepally P, Brami D, Schwartz A, Andrews-Pfannkoch C, Gibson DG, Glass JI, Smith HO, Venter JC, Hutchison CA 3rd (2012) Sequence analysis of a complete 1.66 Mb Prochlorococcus marinus MED4 genome cloned in yeast. Nucleic Acids Res 40:10375–10383

Tang X, Li J, Millán-Aguiñaga N, Zhang JJ, O'Neill EC, Ugalde JA, Jensen PR, Mantovani SM, Moore BS (2015) Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. ACS Chem Biol 10:2841–2849

Tedesco FS, Hoshiya H, D'Antona G, Gerli MF, Messina G, Antonini S, Tonlorenzi R, Benedetti S, Berghella L, Torrente Y, Kazuki Y, Bottinelli R, Oshimura M, Cossu G (2011) Stem cell-mediated transfer of a human artificial chromosome ameliorates muscular dystrophy. Sci Transl Med 3: 96ra78

Wang M, Beck CR, English AC, Meng Q, Buhay C, Han Y, Doddapaneni HV, Yu F, Boerwinkle E, Lupski JR, Muzny DM, Gibbs RA (2015) PacBio-LITS: a large-insert targeted sequencing method for characterization of human disease-associated chromosomal structural variations. BMC Genomics 16:214

Wanisch K, Yáñez-Muñoz RJ (2009) Integration-deficient lentiviral vectors: a slow coming of age. Mol Ther 17:1316–1332

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, Keasling JD, Leavell MD, McPhee DJ, Renninger NS, Newman JD, Paddon CJ (2012) Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. Proc Natl Acad Sci U S A 109:E111–E118

Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC, Moore BS (2014) Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc Natl Acad Sci U S A 111:1957–1962

Young R, Taylor JE, Kurioka A, Becker M, Louis EJ, Rudenko G (2008) 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 9:385

Yuan Y, Andersen E, Zhao H (2016) Flexible and versatile strategy for the construction of large biochemical pathways. ACS Synth Biol 5:46–52

Zhang W, Hagedorn C, Schulz E, Lipps HJ, Ehrhardt A (2014) Viral hybrid-vectors for delivery of autonomous replicons. Curr Gene Ther 14:10–23