University of Nebraska Medical Center
BREAKTHROUGHS FOR LIFE.

University of Nebraska Medical Center
DigitalCommons@UNMC

Journal Articles: Genetics, Cell Biology & Anatomy

Genetics, Cell Biology & Anatomy

4-12-2024

Cross-Contamination of CRISPR Guides and Other Unrelated Nucleotide Sequences Among Commercial Oligonucleotides

Hiroshi Arakawa

Hiromi Miura

Rolen M. Quadros

Masato Ohtsuka

Channabasavaiah B. Gurumurthy

Tell us how you used this information in this short survey.

Follow this and additional works at: https://digitalcommons.unmc.edu/com_gcba_articles

Part of the Medical Anatomy Commons, Medical Cell Biology Commons, and the Medical Genetics Commons

Nucleic Acids Research, 2024, **52**, 3137–3145
https://doi.org/10.1093/nar/gkae068
Advance access publication date: 7 February 2024
Molecular Biology

Cross-contamination of CRISPR guides and other unrelated nucleotide sequences among commercial oligonucleotides

Hiroshi Arakawa [1,*], Hiromi Miura [2], Rolen M. Quadros [3], Masato Ohtsuka [2,4] and Channabasavaiah B. Gurumurthy [3]

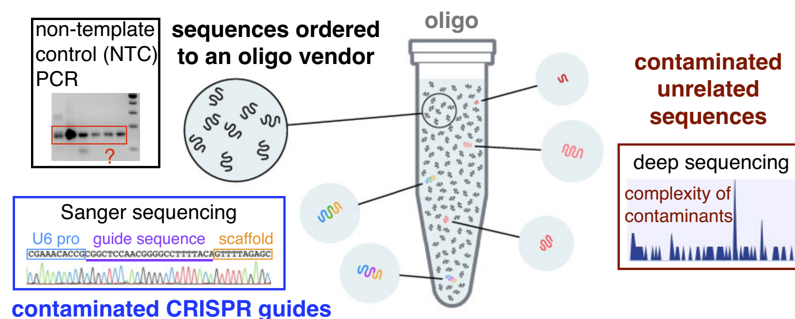
[1]IFOM ETS - The AIRC Institute of Molecular Oncology, Milan, Italy
[2]Division of Basic Medical Science and Molecular Medicine, School of Medicine, Tokai University, Kanagawa Japan
[3]Mouse Genome Engineering Core Facility, University of Nebraska Medical Center, Omaha, NE, USA
[4]The Institute of Medical Sciences, Tokai University, Kanagawa, Japan
[*]To whom correspondence should be addressed. Tel: +39 02 574303306; Fax: +39 02 574303231; Email: hiroshi.arakawa@ifom.eu

Downloaded from https://academic.oup.com/nar/article/52/6/3137/7602849 by guest on 26 April 2024

Abstract

Custom oligonucleotides (oligos) are widely used reagents in biomedical research. Some common applications of oligos include polymerase chain reaction (PCR), sequencing, hybridization, microarray, and library construction. The reliability of oligos in such applications depends on their purity and specificity. Here, we report that commercially available oligos are frequently contaminated with nonspecific sequences (i.e. other unrelated oligonucleotides). Most of the oligos that we designed to amplify clustered regularly interspersed palindromic repeats (CRISPR) guide sequences contained nonspecific CRISPR guides. These contaminants were detected in research-grade oligos procured from eight commercial oligo-suppliers located in three different geographic regions of the world. Deep sequencing of some of the oligos revealed a variety of contaminants. Given the wide range of applications of oligos, the impact of oligo cross-contamination varies greatly depending on the field and the experimental method. Incorporating appropriate control experiments in research design can help ensure that the quality of oligo reagents meets the intended purpose. This can also minimize risk depending on the purposes for which the oligos are used.

Graphical abstract



Introduction

Custom oligonucleotides (oligos) are inexpensive reagents used widely in research, from basic- to translational- research to diagnostics and forensics. In addition to their most common application of polymerase chain reaction (PCR), oligos are used in sequencing, microarray, fluorescence *in situ* hybridization (FISH), and genome-wide functional screens using RNAi and CRISPR knockout technologies. Commercial companies synthesize oligos using standard methods, which involve using four nucleotides supplied via automated liquid handling systems, ultimately generating the customer-requested oligo sequences (1,2). The oligos are shipped as lyophilized reagents or dissolved in buffers to the end users, either in microtubes or multiwell-plate formats. The oligos are synthesized as a standard grade of purity termed ‘desalted’, which is sufficient for most common applications. For some sensitive experiments, researchers opt for high-performance liquid chromatography (HPLC) or polyacrylamide gel electrophoresis (PAGE) purified oligos.

For applications requiring high throughput screening, such as those involving RNAi (3) or CRISPR (4) approaches, researchers procure batches of bar-coded oligos or order oligo-pools and clone them into plasmid vectors to build their knockout libraries. Fishing experiments using such libraries result in several hits, which are then amplified and sequenced to know the clone identities. While a few candidates show expected results when tested using other confirmatory biological assays, it is not uncommon that many hits fail in those assays.

Received: September 26, 2023. **Revised:** January 18, 2024. **Editorial Decision:** January 18, 2024. **Accepted:** January 24, 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Because performing confirmatory biological assays drains a lot of time and resources, it is almost impossible to validate every hit of a screen. If the reagents used for building libraries themselves (or those used for amplifying and confirming the clones following screening experiments) inadvertently contain cross-contaminated oligos, it will make the situation more complex.

We realized this possibility recently, as we observed some unexpected results (presence of unrelated sequences) when setting up a CRISPR library screening, which prompted us to investigate the issue further. Our systematic analyses of various reagents, research equipment, and multiple vendor-supplied oligos suggested that commercially synthesized oligos are the source for such cross-contaminants.

Materials and methods

Oligonucleotides

Oligonucleotides used in this study are shown in Supplementary Table S1.

Oligo suppliers

We ordered oligos from leading commercial oligo-suppliers located in three different geographic regions of the world: three vendors in Europe (for the experiments performed in the Arakawa laboratory), three in the USA (for the experiments performed in the Gurumurthy laboratory), and two in Japan (for the experiments performed in the Ohtsuka laboratory). Because the oligos ordered in these countries were most likely synthesized at different plants/facilities, we considered a total of eight different oligo suppliers as the sources of the oligos we tested in this work.

PCR

In the Arakawa laboratory, most PCRs were performed with Advantage 2 PCR polymerase (TaKaRa), while PrimeSTAR GXL DNA polymerase (TaKaRa) and Q5 DNA polymerase (NEB) were also used as controls. PCR cycling parameters were as follows: various cycles at 98°C for 10 s and 68°C for 10 s. 0.5 μM each of 5′ and 3′ primers were used for PCR.

In the Ohtsuka laboratory, the PCR (for Supplementary Figure S4) was performed in a 10 μl reaction mixture containing 1× GC buffer (TaKaRa), primer set (0.5μM each; '1 and 3' or 'M1159 and 3') and 0.125 U/μl of TaKaRa Taq (TaKaRa) using denaturation (95°C for 5 min), 40 cycles of 95°C for 45 s, 58°C for 30 s and 72°C for 1 min and extension (72°C for 5 min).

In the Gurumurthy laboratory, PCR reactions were performed using Go Taq Promega Hot Start green mix (Promega, Madison, WI, USA). PCR cycling parameters were the same as that of Ohtsuka laboratory, except that the volume of reactions was 50 μl.

Strategies to avoid routine contamination

Aerosol-resistant filter tips were used to prepare PCR samples. Our institute (Arakawa laboratory, where most experiments were performed) offers a service to replace contaminated pipettes with pipettes that have been cleaned, washed, and autoclaved. However, the no template control (NTC) PCR products were confirmed reproducibly with the cleaned or brand-new pipettes. We tested various water- and TE buffer-sources, but the NTC PCR products were reproduced with different water and TE buffers (Supplementary Figure S1G). The same results of NTC PCR were obtained by independent PCR performed on different days.

dsDNase

To test the dsDNA-specificity of dsDNase (Thermo Fisher Scientific), either 10 μg of pBluescript KS (+) plasmid or 1000 pmol of oligo 3 was incubated with 1 μl of dsDNase at 37ºC. After 1, 2, 3, 5, 10, 15, 20, 30 or 60 min, the reaction was stopped by adding EDTA (final concentration of 10 mM). To degrade contaminated dsDNA, either only oligo nucleotides or the whole PCR reaction mixture was incubated at 37 ºC for 2 min and then further incubated at 95 ºC for 5 min to inactivate dsDNase. Thus, dsDNase-treated primer pairs or the PCR reaction mixture were used for PCR reactions.

Cloning and Sanger sequencing

In the Arakawa laboratory, PCR-amplified fragments were cloned into a homemade T-vector by TA cloning. Plasmid DNA was purified using the Wizard Plus SV Minipreps DNA Purification System (Promega) in accordance with the manufacturer's protocol. Gel-purified PCR products and plasmid clones were sequenced with the sequencing primers LEN100 and UC1, respectively. Sequencing was done using Model 373 Automated DNA Sequencer (Applied Biosystems).

In the Ohtsuka laboratory, PCR-amplified fragments were cloned using TOPO® TA Cloning Kit (Life Technologies), and the insert sequences were determined by sequencing of plasmid DNA using M328 primer.

In the Gurumurthy laboratory, the PCR products were separated on a 4% agarose gel and purified using Wizard SV Gel PCR Clean-up System (Promega, cat. no. A9282). Column-purified PCR products were cloned into pGEM®-T Easy Vector Systems (Promega, Madison, WI, USA). Plasmid DNAs were sequenced using one of the M13F or M13R primers.

Deep sequencing

The deep sequencing library was prepared using Accel-NGS 1S Plus DNA Library Kit for Illumina-24 reactions (Swift) and deep-sequenced using MiSeq Reagent Micro Kit v2 for 300 cycles (Illumina). The libraries of 10 samples were deep-sequenced in a flow cell by Illumina MiSeq using ϕX174 DNA as a spike control.

Bioinformatics

FASTQ files demultiplexed by Illumina MiSeq were analyzed using the CLC Genomics Workbench (Ver. 21) (Qiagen). Briefly, 1.9 million of the sequence reads were trimmed to exclude the adapter sequences in the Accel-NGS 1S Plus DNA Library Kit (Swift). From sequence reads, ϕX174 spike DNA was removed, and then the sequences of oligonucleotides were removed by the 'Map Reads to Reference' tool. Sequences commonly appearing in independent sequencing were considered sequences derived from the Accel-NGS 1S Plus DNA Library Kit and, therefore, removed. The rest of the sequence reads were interpreted as contaminated oligo sequences. The 'Map Reads to Reference' tool was also used for mapping oligo contaminants to the reference genome. This tool analyzed the read length of oligo contaminants before mapping.

Downloaded from https://academic.oup.com/nar/article/52/6/3137/7602849 by guest on 26 April 2024

Mapping of reads to the reference sequence can be defined by two parameters in the 'Map Reads to Reference' tool, namely length fraction, which is the minimum percentage of the total alignment length that must match the reference sequence, and the similarity fraction, which is the minimum identity percentage between the alignment region of the read and the reference sequences. To avoid overestimating cross-contaminants, we used the mild parameters to exclude oligonucleotide sequences (30% of the length fraction and 80% of the similarity fraction) and ϕX spikes (50% of the length fraction and 80% of the similarity fraction). We used more stringent parameters to map cross-contaminants to reference genomes (80% of the length fraction and 80% of the similarity fraction). Because MiSeq Reagent Micro Kit v2 for 300 can sequence up to 300 bases, which can cover the entire oligo length produced with current commercial oligo synthesis, the sizes of reads were considered the lengths of oligos.

Because the Accel-NGS 1S Plus DNA Library Kit is single-indexed, a low rate of index hopping could occur during sample preparation for deep sequencing. However, even if index-hopping causes cross-contamination, it could simply mix up the oligo sequences, and such oligo sequences were eliminated to extract oligo contaminants. Therefore, index hopping cannot explain the diverse oligo contaminants.

Reference genomes

The reference genomes are hg38 for Homo sapiens, GRCm39 for Mus musculus, GRCz11 for Danio rerio, BDGP6.32 for Drosophila melanogaster, TAIR10 for Arabidopsis thaliana, ASM584v2 for Escherichia coli, and R64-1-1 for Saccharomyces cerevisiae. The SARS-CoV-2 genome (MN996528) and Human Immunodeficiency Virus 1 genome (NC_001802) were used as reference sequences of the virus genome.

Results

Guide sequence-like PCR products generated in the absence of the template DNA

A few years ago, the Arakawa laboratory developed a novel molecular biological method to convert mRNA into a gRNA library (5). Using this method, a lentiviral CRISPR gRNA library was generated from a mix of several human cell lines and subsequently used to screen high-temperature resistance by transducing the library to haploid myeloid cell line KBM-7 (6).

We designed the primers based on the following two considerations (Figure 1A). (i) We added Illumina adapter tails to each PCR primer to aid in MiSeq deep-sequencing. (ii) Efficient cluster identification of the Illumina sequencing platform depends on the base diversity in the initial 11 cycles. Therefore, a random sequence of 0–3 bases was inserted between the locus-specific sequence and the Illumina tail to generate sequence diversity. We named the oligos as follows: oligo, oligo N, oligo NN and oligo NNN according to the number of additional (random) nucleotides. The mixture of the four Illumina primers was named oligo i (Illumina) (see Figure 1A). Four oligos for each of the forward and reverse primers were mixed and used for respective PCRs.

The guide sequences integrated into the genomic DNA were detected after 27 cycles of PCR (Figure 1B). Unexpectedly, a faint band was noticed in the no template control (NTC). When PCR was extended to 30 cycles, a higher intensity amplicon was detected in the NTC (Figure 1B). The size of this band appeared to be comparable to that of the PCR product of interest rather than a primer dimer suggesting potential contamination of guide sequences within PCR components. Thirty cycles of PCR reaction without primers produced no PCR product in the NTC sample (Figure 1B).

Considering the possibility that the 1i and 2i oligos could have been contaminated, those oligos were re-ordered (as a new batch), and additional primers at the 3′ end (3i and 4i) were designed. Like the first batch of oligos, we added Illumina tails in the second batch. The new set of primers (1i/3i and 1i/4i) also produced PCR products in the absence of a template. Of the three combinations, the 1i/3i combination produced higher quantities of nonspecific amplification (Figure 1C). NTC PCR products were detected even when HPLC and PAGE grade oligos were used. Strikingly, PAGE-grade oligos produced even more PCR products than the desalt-grade oligos (Figure 1D). Furthermore, we concluded that contamination did not arise from any of the components of our PCR, including water, PCR buffers, or polymerases (Supplementary Figure S1).

For the initial experiments, the Arakawa laboratory purchased a collection of different primers with the same sequence specificity as primers 1, 2 and 3 (Figure 1A). Using this collection of primers, 35 cycles of PCR were performed for 84 different primer combinations of 12 distinct forward primers and seven unique reverse primers, all in the absence of a template (Figure 1E). All of the 84 primer combinations generated NTC PCR products. Notably, the amount of PCR product differed depending on the combination of primers. Combinations such as 1NN/3N, 1NN/3NN and 1NN/3NNN generated fewer PCR products (marked by * in Figure 1E). These variations in slightly different sizes and amounts of NTC PCR products suggest the possibility of various contaminants in each oligo stock.

To examine whether such oligo contamination was specific to the vendor, we procured oligos 1 and 2 from two other vendors. The vendor 2 and 3 oligos also generated PCR products in NTC although they were relatively fainter than those from vendor 1; a higher number of PCR cycles (35 or 40 cycles) generated NTC PCR products by using oligos from vendors 2 and 3 (Figure 1F).

The contaminants are single-stranded DNAs.

Electrophoresis of the 2i oligos on denaturing PAGE containing urea revealed several unexpected bands in addition to the main DNA band (Figure 1G). Failure of oligo synthesis typically results in deletions within the oligo sequences due to failed coupling (7). This can result in the formation of a collection of ladder-like bands smaller than the main band. However, we observed several large-sized bands of unknown origin in most of the desalt-grade oligos (Figure 1G). These extra and larger bands were also detected in the highly purified HPLC and PAGE grade oligos from vendor 1 but not from vendor 3′s PAGE grade oligo (Figure 1G). It remains unknown if they are only oligos, other substances, or oligos bound with other substances. Nevertheless, the contaminants amplified by PCR were ruled out for double-stranded DNA molecules because they were resistant to the treatment to degrade double-stranded DNA (Supplementary Figure S2).

Downloaded from https://academic.oup.com/nar/article/52/6/3137/7602849 by guest on 26 April 2024

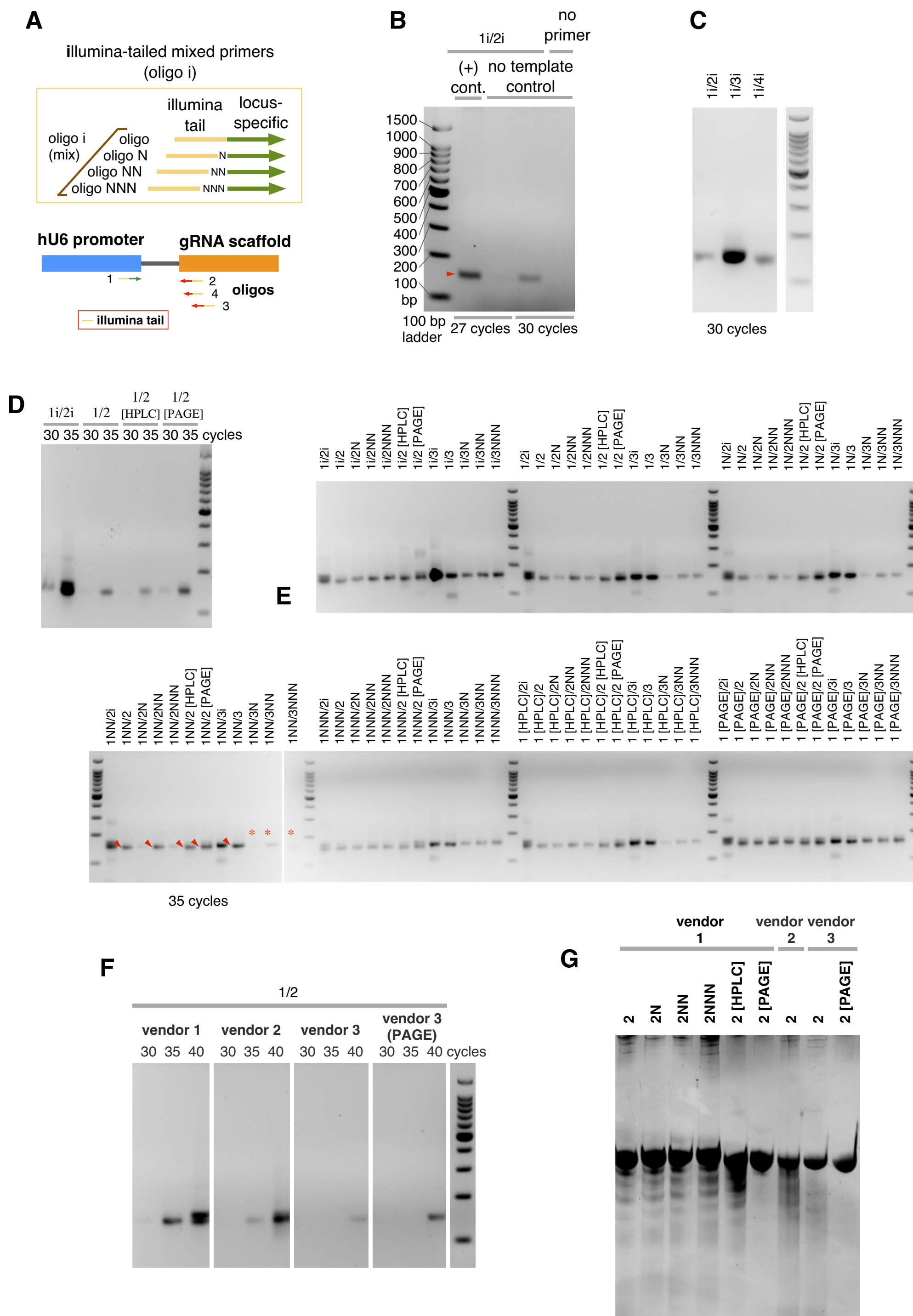


Figure 1. CRISPR guide sequence-like artifacts in no template control (NTC) PCR. (**A**) Schematic of PCR primers to amplify CRISPR guide sequences. Oligo i is the mix of four Illumina-tailed oligos. Positions and directions of oligos are shown by arrows. Two-colored arrows are Illumina-tailed oligos. (**B**) PCR in the presence or absence of template DNA. A genomic DNA sample containing guide sequences was used as a positive control. This genomic DNA was extracted from the KBM-7 myeloid cell line transduced by a single guide library. Genomic DNA samples containing the guide sequences produced a clear PCR product after 27 cycles. However, 27 cycles of PCR without a template also had a slight PCR product (NTC PCR product), which increased after 30 cycles. NTC PCR product was not observed after 30 cycles of PCR without primers. Triangles represent the expected size of guide-sequence PCR products. 100 bp ladder size marker is shown together. (**C**) NTC PCR with three different PCR primer pairs. NTC PCR products were observed with three primer pairs, i1/2i, 1i/3i and 1i/4i, the mixture of 8 oligos. (**D**) NTC PCR products by high-performance liquid chromatography and polyacrylamide gel electrophoresis (HPLC and PAGE, respectively) grade primers. HPLC and PAGE grade primers also generated NTC PCR products. 1i/2i (8 oligos) produced more NTC products than 1/2 (2 oligos). (**E**) NTC PCR products produced by a total 84 combinations of primer pairs. NTC PCR products were observed from all 84 primer pairs, including HPLC and PAGE grade primers. Because a few primers paired with the 1NN did not generate significant NTC PCR products (marked by *), NTC PCR products of 1NN/2, 1NN/2NN, 1NN/2 (HPLC), 1NN/2 (PAGE) (marked by the arrowhead) were further analyzed by Sanger sequencing in Figure 2. (**F**) NTC PCR with PCR primers of three different vendors. Primers from two other vendors, including PAGE grade, also generated NTC PCR products. (**G**) Oligo DNAs in urea-PAGE. 100 pmol of oligos were run on 20% PAGE containing 50% urea. All PCRs were performed without a template except for (B).

Downloaded from https://academic.oup.com/nar/article/52/6/3137/7602849 by guest on 26 April 2024

The contaminants include CRISPR guide sequences

While a few primers paired with the 1NN did not generate significant NTC PCR products (marked by * in Figure 1E), 1NN/2-, 1NN/2NN-, 1NN/2- (HPLC), 1NN/2- (PAGE) and 1NN/3-combinations generated NTC PCR products (marked by the arrowhead in Figure 1E). This suggested that most of the contamination in these primer pairs were derived from oligos other than from 1NN. Another independent PCR with these primer pairs produced the same sized PCR products, which were gel-extracted and used for direct sequencing (i.e. without cloning). Each nucleotide sequence showed a different characteristic chromatogram (Figure 2A). Importantly, unexpected sequences of approximately 20 bp were observed between the primer pairs used for PCR (Figure 2A). The chromatogram of 1NN/3 appeared as a single specific sequence (Figure 2A), also confirmed by other PCR polymerases (Supplementary Figure S3). In contrast, the other four NTC PCR products appeared as a mix of two or more sequences (Figure 2A).

To investigate further, the above PCR products were cloned into a plasmid vector, and six independent plasmid clones from each cloning reaction were sequenced. Probable guide sequences were detected in all six NTC PCR product clones derived from 1NN/2 (PAGE) as shown in Figure 2B. Of these, two different guide sequences were derived from *Gallus gallus* fragile X mental retardation 1 (FMR1), one was from *Homo sapiens* G protein-coupled receptor kinase 6 (GRK6), and one was from a non-coding sequence on the mouse chromosome (Figure 2B). SpCas9 PAM (NGG) sequences were confirmed in the potential target sequences in those genomes. This was intriguing because guide sequences from different species and two guide sequences for the same gene (unrelated to the CRISPR library) were detected in an NTC PCR product. Given the diversity of the contaminates, it is highly unlikely that they originated in our laboratory. Likewise, many different guide sequences from various species were detected in the other four NTC PCR products (Figure 2C).

Contamination of CRISPR guide sequences was detected in every batch of research-grade oligos procured from all suppliers tested.

Observations made in the Arakawa laboratory using three different oligo manufacturers in Europe prompted us to investigate whether this issue is common in other regions. Some of the experiments presented in Figure 2 were repeated by procuring oligos from manufacturers in Asia and the USA (experiments performed at Ohtsuka and Gurumurthy laboratories, respectively). Oligos (1, 3 and M1159) were procured from two major manufacturers in Asia, and three in the USA and tested for contamination; all commercial oligos tested showed the presence of unrelated guide RNA contaminations (Supplementary Figure S4).

Contaminants other than CRISPR guide sequences

The oligos we investigated in this work for cross-contamination were those intended for amplifying CRISPR guide sequences. To understand if those oligos contained heterologous type (i.e. non-CRISPR guide) contaminants, we performed deep-sequencing analysis for 10 oligos (9 oligos shown in Figure 1G and oligo 3 analyzed in Figure 2A and C) using Illumina MiSeq. Since the oligo DNAs are single-stranded, the deep-sequencing libraries were prepared using a specialized kit for library preparation from single-stranded DNA (the Accel-NGS 1S Plus DNA Library Kit for ssDNA).

All 10 oligos analyzed harbored various contaminants, a total of 15 154 reads. The level of oligo contamination was variable, ranging from 0.17% to 2.88% (Figure 3A). The lengths of contaminants ranged from 20 to 200 nucleotides (Figure 3B), indicating the presence of a variety of contaminants that are not resolved by PAGE (Figure 1G). This size distribution reflects the standard oligos synthesized at most commercial vendors. The sequence of the contaminant guide sequence DNA was successfully identified from deep sequencing reads (Figure 3C). This sequence contained the guide sequence together with parts of the human U6 promoter and gRNA scaffold, confirming the result of Sanger sequencing (1NN/3 in Figure 2A and C). The limited length of the U6 promoter and guide scaffold within oligos may explain why the primer pairs closer to the guide sequence produce more NTC PCR products (Supplementary Figure S1E).

To explore the origin of the remaining contaminants, we first mapped the contaminants to the human genome. For example, 162 reads were mapped through human chromosome 1 (Figure 3D), and contaminants were distributed to other chromosomes as well. Oligo contaminants were also mapped on the genome of model organisms such as mice, zebrafish, flies, yeast, plants (*Arabidopsis*) and *Escherichia coli* (Figure 3E). Minor oligos were also found on pathogenic viral genomes such as SARS-CoV-2 and human immunodeficiency virus (Figure 3F) and on a plasmid vector, pBluescript KS (+) (Figure 3G). Of note, we found a relatively much higher number of HIV sequences (10 reads) than SARS-CoV-2 (only 1 read). This could probably be because the companies have been synthesizing HIV oligos for a much longer period than for SARS-CoV-2, or possibly because HIV sequences were being synthesized in parallel at the same time on the same equipment for another customer while our oligos were synthesized.

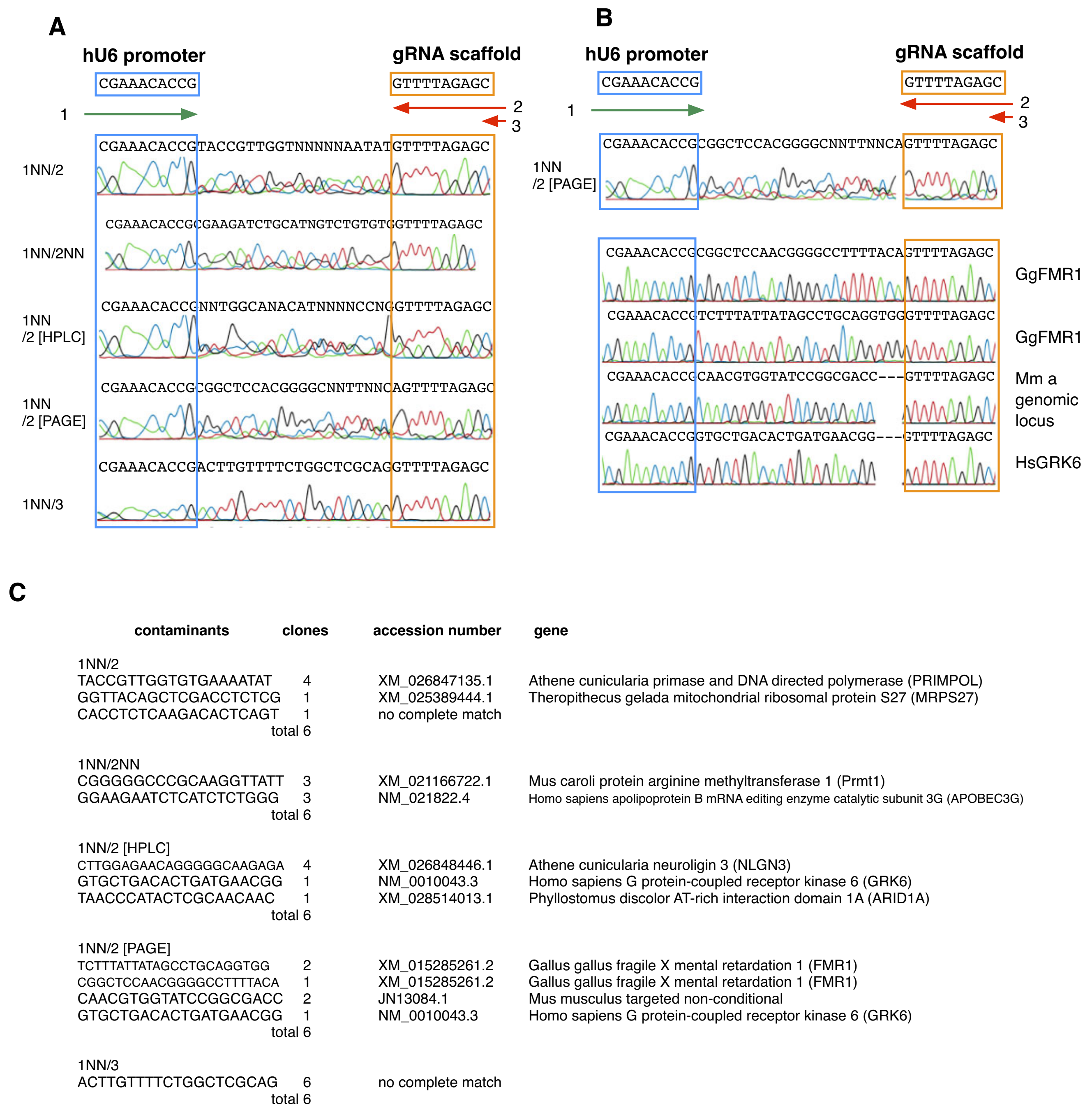
Discussion

Our observation of PCR products in no template control (NTC) samples of CRISPR library screening experiments prompted us to investigate the cross-contamination issue. Our systematic analysis of the issue to amplify guide sequences led us to detect the contamination issue. Based on our observations, we conclude that commercially purchased custom oligo DNAs are frequently cross-contaminated with nonspecific oligos, including the CRISPR guide sequences (Figure 3H).

We suggest that the presence of near-universal cross-contamination of CRISPR guides and other oligonucleotides could be due to the following reasons. While the PCR template of guide sequences can be designed as oligos because of their short length, we used sensitive PCR primers close to the guide sequence, and the mix of eight different primers for amplifying guide sequences has further increased sensitivity for detecting contamination. Sanger sequencing of the guide sequences revealed the diversity of the guide sequence contaminants. CRISPR technology is widely used, increasing the chance that cross-contaminated oligos could contain unrelated guide sequences. Finally, deep sequencing of oligos revealed the many other types of oligo contamination beyond guide sequences.

Due to the following reasons, the typical contamination during the lab work is not considered the primary reason for the cross-contamination in this study. For instance,

Downloaded from https://academic.oup.com/nar/article/52/6/3137/7602849 by guest on 26 April 2024



contaminants	clones	accession number	gene
1NN/2			
TACCGTTGGTGTGAAAATAT	4	XM_026847135.1	Athene cunicularia primase and DNA directed polymerase (PRIMPOL)
GGTTACAGCTCGACCTCTCG	1	XM_025389444.1	Theropithecus gelada mitochondrial ribosomal protein S27 (MRPS27)
CACCTCTCAAGACACTCAGT	1	no complete match	
	total 6		
1NN/2NN			
CGGGGGCCCGCAAGGTTATT	3	XM_021166722.1	Mus caroli protein arginine methyltransferase 1 (Prmt1)
GGAAGAATCTCATCTCTGGG	3	NM_021822.4	Homo sapiens apolipoprotein B mRNA editing enzyme catalytic subunit 3G (APOBEC3G)
	total 6		
1NN/2 [HPLC]			
CTTGGAGAACAGGGGGCAAGAGA	4	XM_026848446.1	Athene cunicularia neuroligin 3 (NLGN3)
GTGCTGACACTGATGAACGG	1	NM_0010043.3	Homo sapiens G protein-coupled receptor kinase 6 (GRK6)
TAACCCATACTCGCAACAAC	1	XM_028514013.1	Phyllostomus discolor AT-rich interaction domain 1A (ARID1A)
	total 6		
1NN/2 [PAGE]			
TCTTTATTATAGCCTGCAGGTGG	2	XM_015285261.2	Gallus gallus fragile X mental retardation 1 (FMR1)
CGGCTCCAACGGGGCCTTTTACA	1	XM_015285261.2	Gallus gallus fragile X mental retardation 1 (FMR1)
CAACGTGGTATCCGGCGACC	2	JN13084.1	Mus musculus targeted non-conditional
GTGCTGACACTGATGAACGG	1	NM_0010043.3	Homo sapiens G protein-coupled receptor kinase 6 (GRK6)
	total 6		
1NN/3			
ACTTGTTTTCTGGCTCGCAG	6	no complete match	
	total 6		

Figure 2. Contamination of CRISPR guide sequences in oligos. (**A**) Sequences of NTC PCR products without cloning. NTC PCR products were purified from agarose gel and sequenced without cloning. Single or mixed sequences were present within the region flanked by primer pairs. Arrows show the location of primers. (**B**) Sequences of NTC PCR products cloned in plasmid vectors. The 1NN/2 (PAGE) NTC product was cloned into a plasmid vector, and each clone was sequenced. Sequences before and after cloning are shown as chromatograms. The genes corresponding to the guide sequence are also shown. (**C**) Specificity of guide sequences in NTC PCR products. NTC PCR products of 5 primer pairs were cloned into a plasmid vector, and each clone was sequenced. The genes corresponding to the guide sequence are also shown with accession numbers.

contamination of pipettes, buffers, or water occurs in the lab; the contaminating sequence will be limited to one or a few sequences studied locally in a given lab. It may be possible for commercial PCR enzymes to be contaminated with small amounts of bacterial genomic DNA from which the enzymes are purified (8). CRISPR is a mechanism of acquired immunity in bacteria (9), so bacterial genomic DNA can contain guide sequences from their CRISPR locus. Because the guide sequence in oligo contaminants comprises the human U6 promoter, guide sequence, and gRNA scaffold, such a combination of sequences does not exist in bacterial genomes. If the oligos are cross-contaminated with each other due to index hopping during deep sequencing, the sequences detected would be those of oligonucleotide sequences synthesized at commercial facilities. Furthermore, we removed the oligonucleotide sequences to extract contaminant sequences. Therefore, cross-contamination between oligos by index hopping cannot explain the diverse contaminant sequences revealed

Downloaded from https://academic.oup.com/nar/article/52/6/3137/7602849 by guest on 26 April 2024

Figure 3. Ratios and complexity of oligo contaminants. (**A**) Ratios of oligo contaminants. The deep sequencing library was prepared from 10 oligos using Accel-NGS 1S Plus DNA Library Kit for Illumina and deep-sequenced in a flow cell using Illumina MiSeq. From the 1.9 million total sequence reads, sequences of oligonucleotides were removed to identify oligo contaminants of 15154 reads. The ratios of oligo contaminants per each oligo's total sequence reads are shown. (**B**) Size distribution of oligo contaminants. Because the sequencing kit can sequence up to 300 bases, which covers the entire length of commercially available oligos, the read lengths of oligo contaminants were considered oligo lengths. (**C**) The sequence of CRISPR guide sequence contaminated in oligo 3. The full-length sequence of the guide sequence oligo contaminated within oligo 3 was aligned with the sequence identified by Sanger sequencing (Figure 2A). Deep sequencing revealed its full-length sequence, whereas Sanger sequencing could not clarify the extent of the overlap between the PCR primers and the contaminated oligo. The locations of primers 1 and 3 are indicated. (**D**) Distribution of oligo contaminants on *Homo sapiens* chromosome 1. 15154 reads of whole oligo contaminants were used for mapping on human chromosome 1. (**E**) Oligo contaminants on the genome of model organisms. Whole oligo contaminants were mapped on the genome of 7 model species. Some oligos may map to the genome of more than one species and may have been counted multiple times. (**F**) Distribution of oligo contaminants on pathogenic viral genomes. Whole oligo contaminants were mapped on the genome of SARS-CoV-2 and HIV. (**G**) Distribution of oligo contaminants on a plasmid vector. Whole oligo contaminants were mapped on the sequence of a plasmid vector, pBluescriptKS(+). (**H**) Schematic image of contaminations in commercial oligos. Commercial oligonucleotides contain diverse cross-contaminants, including CRISPR guide sequences. The percentages in (A) indicate contaminants in each oligo, while those in (D)–(G) indicate reads among whole contaminants (15 154 reads).

Downloaded from https://academic.oup.com/nar/article/52/6/3137/7602849 by guest on 26 April 2024

by deep sequencing. Therefore, it is unlikely that the source of the contamination in this study was a manipulation error in the laboratory, DNA in the environment, bacterial DNA in the enzyme, or contamination between oligos by index hopping, which is generally assumed to be the common sources of contamination.

These results indicate that the contamination we observed was highly likely derived from the commercial oligos. Oligo contamination was detected from all vendors we tested from different regions of the world (Supplementary Figure S4), and therefore, we conclude that this could be a universal technical problem. Multiple methods confirmed contamination: PCR and electrophoresis (Figure 1), Sanger sequencing (Figure 2) and deep sequencing (Figure 3). Each method has its strengths and weaknesses: (i) Our PCR protocol is sensitive to detect guide sequences (Figure 1, Supplementary Figures S1 and S3A), and Sanger sequencing can identify each guide sequence (Figure 2, Supplementary Figures S3B and S4) but cannot detect non-CRISPR sequences. (ii) PAGE gel can visualize oligos regardless of sequences (Figure 1G), but whether each band shows an oligo remains unknown. 3) Deep sequencing identified diverse sequences other than the guide sequence (Figure 3). However, since almost all of the sequences identified in the deep sequencing differed from guide sequences, we presume that the guide sequences may be a minor fraction of types of contaminants. Nevertheless, deep sequencing characterized the full-length sequence of the guide sequence oligo highly contaminated in the oligo 3 (Figure 3C), confirming the PCR (Figure 1C, E, Supplementary Figures S1 and S3A) and Sanger sequencing (Figure 2A, C, Supplementary Figure S3B) approaches.

The level of contamination, as assessed by the intensity of PCR amplifications or the number of PCR cycles used for amplification, differed between manufacturers. Interestingly, HPLC- or PAGE-grade oligos also contained cross-contaminants. Of note, we observed the highest levels of contamination in the HPLC-grade oligos (Figure 3A). Although PAGE and HPLC address the purity of oligos in reducing synthesis truncation products, these purification processes may likely add the risk of secondary contamination due to increased handling at facilities (i.e. while the molar purity of the main product is improved with purification, the additional handling offers an opportunity for sample contamination from other sequences being processed in parallel). While quality control by mass spectrometry is used to verify the correct mass of the desired main product and identify synthesis failure products, such quality control methods may not identify low-level contaminations in such preparations.

The estimated market size for oligos is billions of USD. Commercial companies worldwide synthesize oligos round the clock in most facilities to meet the high demand. We observed cross-contamination among all vendors we tested from three different continents. Our observation of cross-contaminated oligos could be expected because the synthesis facilities produce oligos using the same set of equipment back-to-back to meet the high demand, and the traces of the oligos remaining in the equipment may become a source of cross-contamination of the oligos synthesized subsequently. While it is practically impossible for companies to extensively clean equipment between each synthesis step, the deep sequencing approach shown in this study could help manufacturers identify the source of contaminants to ensure a higher quality of their oligo preparations. Widespread use of single-molecule detection techniques such as next-generation sequencing and digital droplet PCR have already led to major changes in manufacturing Standard Operating Procedures (SOPs) (at least for some vendors).

End users desire fast delivery of inexpensive oligonucleotides to help accelerate the pace of their experiments and to do as much research as possible on limited budgets. Hence commercial oligonucleotide manufacturers employ highly parallelized, high-throughput manufacturing systems that focus on the quality of synthesis while risking low levels of cross-contamination. A single oligonucleotide synthesis event can easily produce 10^{17} or more molecules of that oligonucleotide. Even if the level of contamination is only 1 part per billion, 10^{8} molecules of contaminant would be present in the final preparation, which can be readily detected using modern PCR or NGS methods. While some level of contamination may be inevitable, in our analyses, it is worth noting that relative contamination levels varied between suppliers. Zero cross-contamination can likely be achieved, but such results are probably only possible in the setting of GMP manufacturing of oligonucleotides (for example, products used for human therapeutics) where only a single oligonucleotide is made on a single synthesizer in its cleanroom at a very high cost and slow throughput.

Our findings caution that commercial oligo DNAs can contain contaminated nonspecific sequences, which could affect research studies, particularly if the reagents are used for high throughput library preparation or screening. It would be wise to perform control experiments to validate oligo purity before proceeding with sensitive studies. For example, control experiments, such as the no template control in this study or independently ordered oligos of the same sequence, should be helpful. The risk of contamination can be ignored if the end goal of the experiment is not compromised. For instance, contaminants can be okay for an experiment involving amplifying specific genes from a given species and subsequently cloned to determine the nucleotide sequence. Similarly, a few percent of unrelated oligonucleotides are unlikely to impact the results of an antisense or siRNA experiment. In the deep sequencing of low-cycle amplified samples, the contamination could be overlooked in the noise behind the real signal. Thus, depending on the purpose, the influence of trace amounts of oligo contaminants is limited. In contrast, the contamination of oligos may have caused unexpected artifacts in specialized experiments such as metagenomic analysis, de novo sequencing, ancient genetic material sequencing, rare sequence screening and other experiments currently documented in published papers due to the high number of amplification cycles or reference-independent analysis.

Cross-contamination, almost omnipresent in commercial oligos, has not been reported thus far. This knowledge should be shared among the scientific community, which is the aim of this paper.

Data availability

The deep sequencing data are publically available in DDBJ Sequence Read Archive (DRA). DRR452922–DRR452931 represent the original sequence reads, and DRR452932 - DRR452941 represent the contaminant reads extracted from the original sequence reads.

Downloaded from https://academic.oup.com/nar/article/52/6/3137/7602849 by guest on 26 April 2024

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We are grateful to Benjamin Hershey, Holger Neecke, Ivan Psakhye and Chaithra Sanjeevamurthy for critically reading the manuscript and to Simone Minardi and Mirko Riboni for deep sequencing. We are grateful to Akiko Arakawa for the graphical abstract. We thank Aki Kurosaki for plasmid isolation in the Ohtsuka laboratory and the Support Center for Medical Research and Education staff at Tokai University for sequencing in the Ohtsuka laboratory.

Funding

No funding was received.

Conflict of interest statement

The authors declare no conflicts of interest.

References

1. Caruthers,M.H., Barone,A.D., Beaucage,S.L., Dodds,D.R., Fisher,E.F., McBride,L.J., Matteucci,M., Stabinsky,Z. and Tang,J.Y. (1987) Chemical synthesis of deoxyoligonucleotides by the phosphoramidite method. *Methods Enzymol.*, **154**, 287–313.
2. Horvath,S.J., Firca,J.R., Hunkapiller,T., Hunkapiller,M.W. and Hood,L. (1987) An automated DNA synthesizer employing deoxynucleoside 3′-phosphoramidites. *Methods Enzymol.*, **154**, 314–326.
3. Moffat,J., Grueneberg,D.A., Yang,X., Kim,S.Y., Kloepfer,A.M., Hinkle,G., Piqani,B., Eisenhaure,T.M., Luo,B., Grenier,J.K., *et al.* (2006) A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell*, **124**, 1283–1298.
4. Shalem,O., Sanjana,N.E., Hartenian,E., Shi,X., Scott,D.A., Mikkelson,T., Heckl,D., Ebert,B.L., Root,D.E., Doench,J.G., *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, **343**, 84–87.
5. Arakawa,H. (2016) A method to convert mRNA into a gRNA library for CRISPR/Cas9 editing of any organism. *Sci. Adv.*, **2**, e1600699.
6. Kotecki,M., Reddy,P.S. and Cochran,B.H. (1999) Isolation and characterization of a near-haploid human cell line. *Exp. Cell. Res.*, **252**, 273–280.
7. Yang,J., Stolee,J.A., Jiang,H., Xiao,L., Kiesman,W.F., Antia,F.D., Fillon,Y.A., Ng,A. and Shi,X. (2018) Solid-phase synthesis of phosphorothioate oligonucleotides using sulfurization byproducts for in situ capping. *J. Org. Chem.*, **83**, 11577–11585.
8. Stinson,L.F., Keelan,J.A. and Payne,M.S. (2019) Identification and removal of contaminating microbial DNA from PCR reagents: impact on low-biomass microbiome analyses. *Lett. Appl. Microbiol.*, **68**, 2–8.
9. Barrangou,R., Fremaux,C., Deveau,H., Richards,M., Boyaval,P., Moineau,S., Romero,D.A. and Horvath,P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709–1712.

Downloaded from https://academic.oup.com/nar/article/52/6/3137/7602849 by guest on 26 April 2024

Received: September 26, 2023. **Revised:** January 18, 2024. **Editorial Decision:** January 18, 2024. **Accepted:** January 24, 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com