

Intramolecular "Chimerization" of *Escherichia coli* RNAs as a New Type of Transcript Editing

Shvyreva U.S.¹, Boyko L.E.², Kolzhetsov N.P.¹, Shavkunov K.S.¹,
Glazunova O.A.¹, Alikina O.V.¹, Markelova N.Yu.¹, Panyukov V.V.^{1,3},
Ozoline O.N.¹

¹*Institute of Cell Biophysics, Pushchino Scientific Center for Biological Research, Russian Academy of Sciences, Pushchino, Russia*

²*Moscow State University. M.V. Lomonosov, Moscow, Russia*

³*Institute of Mathematical Problems of Biology RAS, Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Pushchino, Russia*

Abstract. Chimeric RNAs, composed of sequences from different genes, are the least studied category of bacterial RNAs, and their very existence has not been universally recognized so far. The primary aim of this study was to test the hypothesis that unusual RNA chimeras can arise from transcription priming by a specific 8/9-nucleotide RNAs, GCCAAGGC(G). Applying single-molecule sequencing of *Escherichia coli* cDNA on Oxford Nanopore MinIon platform, we confirmed the expected presence of these 8/9-mers at the 5'-ends of numerous transcripts. Surprisingly, we also discovered thousands of transcripts from diverse genes containing an internal insertion of the GCCAAGGC 8-mer as a replacement of native genomic 3–15 nucleotide sequences. Given that only $1.52 \pm 0.47\%$ of oligonucleotides, containing this 8-mer as a genomic sequence, carry a single-nucleotide substitution within it, the revealed intramolecular "chimeras" likely represent genuine post-transcriptional RNA modifications not compromised by potential artifacts of the experiment. Notably, the substitutions occurred predominantly at sites where the replaced genomic sequences contained GC dinucleotides at both borders. We detected no consensus sequences in the 20-nucleotide flanks near the modification site, implying that signals for intramolecular modifications are located within the replaceable segments. Although detected in over two hundred distinct gene products, most of the chimeric transcripts originated from 16S or 23S rRNAs and were often located within antisense transcripts. The discovered modification has no analogues. While this article discusses its possible biological role, further investigation will require new approaches and fresh conceptual ideas.

Key words: *chimerization of bacterial RNAs, post-transcriptional RNA editing, bacterial secreted RNAs, antisense RNAs, nanopore transcriptomics*

INTRODUCTION

Transcripts of many genes across all domains of life undergo a variety of modifications that enhance their functional potency. Chemical modifications, in particular, are an obligatory component of the maturation of ribosomal and transfer RNAs, while forming secondary structures untranslated regions at the ends of messenger RNAs (5'- and 3'-UTRs), after endonucleolytic detachment, can play an independent regulatory role. Chimeric RNAs are molecules formed by the joining of two separate transcripts. In eukaryotes, such RNAs are often generated by intergenic splicing events, including such as long-range intrachromosomal and interchromosomal trans-splicing, as well as cis-splicing between adjacent genes [1]. Whether they retain protein-coding ability or not, these chimeric RNAs can interfere with

numerous molecular processes and affect cellular metabolism. Eukaryotic chimeric RNAs have attracted significant attention because of their apparent link to cancer [2–5] and a number of hereditary diseases. A classic example is chronic myeloid leukemia, caused by a reciprocal inter-chromosomal translocation that creates the BCR::ABL hybrid gene with constitutively active tyrosine kinase [6]. While initially thought to be exclusive to cancers with chromosomal rearrangements, chimeric RNAs are now known to form in normal cells as well [7] and appear herein through several non-canonical mechanisms [8, 9]. These pathways include the fusion of RNA transcripts from opposite DNA strands [10], junction with transposable elements [11] or gene leader sequences [12], and chimerization via circulization [9].

The functional implications of eukaryotic chimeric RNAs are remarkably diverse [13]. They can act as competing endogenous RNAs (ceRNA) or function as microRNA (miRNA) sponges [1]. By targeting the entire transcript for nonsense-mediated decay, chimeric RNAs can critically reduce their intramolecular level [9]. Several lines of evidence suggest that trans-spliced chimeric RNA may serve as a template for DNA rearrangement [14, 15] or act as a scaffold to bring two genomic loci into close proximity [9, 16]. Consequently, chimeric RNAs are increasingly recognized as components of basic physiology, involved in regulating multiple cellular processes.

Recent advances in high-throughput sequencing, which provide single nucleotide resolution and long read coverage [11], have enabled the systematic discovery of novel chimeric RNAs. Dozens of computational approaches have been developed to identify them in various transcriptomic data [17]. Most algorithms begin with an alignment step to identify reads that map discordantly to two different genes, following by a series of filtering and/or realignment steps are applied to compile candidate chimeric RNAs [18]. Several databases consolidate this information. ChimerDB [19] is one of the primary repositories that provides a gateway to vast amount of data from The Cancer Genome Atlas (TCGA) project. It integrates deep sequencing data analysis (ChimerSeq), text mining of PubMed publications (ChimerPub), and extensive manual annotations (ChimerKB). By the end of 2020, ChimerDB v.4 contained 65945 fusion genes [19], a number that increased to 67610 by the end of 2025 (www.kobic.re.kr/chimerdb/). Another resource, ChiTaRS v.8.0, is a specialized repository for human chimeric transcripts, containing 47445 curated RNA transcripts and over 100000 chimeric sequences [20]. The hhrtBase, in contrast, is a database for transcription readthrough (TRT) events in healthy human tissues, containing data for 75248 TRT events across 11692 genes from 43 tissues [21]. However, none of these resources compile chimeric RNAs from bacteria, and the history of their study is very short.

While RNA-seq has identified many mRNA chimeras in prokaryotes [22–24], most of them were presumed to be spurious artifacts. This assumption was based on the fact that bacterial genomes are largely intron-less, and their transcripts were not expected to undergo splicing. To assess the validity of bacterial RNA chimeras, Lorens-Rico and colleagues [24] analyzed the transcriptomes of *Pseudomonas aeruginosa* and *Mycoplasma pneumoniae* belonging to gram negative and gram-positive bacteria, respectively. The experimental design involved preparing samples in two ways: mixing the RNAs of two species at different ratios immediately after RNA extraction and mixing the bacterial cells prior to RNA extraction. Using three different software tools, the authors detected large numbers of putative chimeric RNAs. Crucially, in the mixed-RNA samples, the authors found a high frequency of inter-species chimeras. For example, at a 1:1 mixing ratio, approximately 20 % of the chimeric reads contained sequences from both *M. pneumoniae* and *P. aeruginosa*. By comparing the expected and observed frequencies of different chimeras across the various mixing ratios, the authors concluded that the chimeric RNAs were likely artifacts generated during library preparation.

It was, therefore, surprising to detect non-canonical chimeras in our previous study [25], whose primary aim was to identify potentially signaling RNAs involved in intercellular

interactions. The search was conducted using eight sets of RNA-seq reads obtained from an Ion Torrent PGM sequencer where reads ranging from 12 to 50 nt in length were analyzed. This analysis revealed an 8/9-mer GCCAAGGC(G) at the 5'-end of reads originating from the fimA-fimI intergenic region (GTTCAGGGACGTCATTACGGGCA- -). Such chimeras constituted 0,029 % of the abundant reads (472844) containing transcripts from the fimA-fimI intergenic space. Since the 9-mer could be produced as an abortive transcript from an antisense promoter within *ravA* gene; shares its terminal G with the first guanine of the fimA-fimI sequence and was found in the same junction at the 5'-ends of many other RNAs, we hypothesized that these chimeras were products of primed transcription [25]. This follow-up study was undertaken to check that hypothesis using long-reads sequencing on the Oxford Nanopore platform, which provides the advantage of analyzing junction points within much longer sequence context [11, 13]. This approach led to the discovery and statistical characterization of an unexpected intramolecular chimerization.

MATERIALS AND METHODS

Bacterial strain and growth condition

We used the wild-type *Escherichia coli* K-12 strain MG1655 (*E. coli*) as a source of intracellular RNA. Two bacterial samples were grown in M9 mineral medium (3 g/l KH_2PO_4 , 6 g/l Na_2HPO_4 , 0.5 g/l NaCl, 1 g/l NH_4Cl , 2 mM MgSO_4 , 0.1 mM CaCl_2 , 0.5 % glucose, pH 7.0). The media were inoculated with an overnight *E. coli* culture at a 1:1000 ratio and bacteria were incubated with shaking at 37°C for 24 hours. The first sample (6 ml) was cultured in twelve 0.5 mL wells of 48-well clear flat-bottom TC-treated cell culture plate (Falcon®, Corning, USA), with a Synergy H1 Hybrid Multi-Mode Reader (BioTek Instruments, USA). To increase RNA yield, the second sample was cultivated under similar conditions but in a water bath with shaker using 30 mL of M9 medium in 100 ml flasks.

Total RNA purification

In the first experiment, cells from the culture grown to an optical density (OD_{600}) of 0.8–1.0 were harvested by centrifugation at 5000 rpm (4°C), rinsed with PBS, and resuspended in lysis buffer (20 mM Sodium acetate pH 5.0, 1 % SDS, 20 mM EDTA). Total RNA was isolated following the manufacturer's protocol for TRIzol reagent (Invitrogen, Carlsbad, CA, USA). In the second experiment, cells from the culture with OD_{600} of 1.0–1.2 were harvested, rinsed and resuspended directly in TRIzol. The sample was then subjected to three freeze-thaw cycles in liquid nitrogen and at 37°C water bath before proceeding with the standard TRIzol protocol. The quantity and quality of the isolated RNA were estimated with NanoDrop spectrophotometer ($\text{A}_{260}/\text{A}_{280} \geq 1.8$ and $\text{A}_{260}/\text{A}_{230} \geq 1.8$) (Thermo Fisher Scientific, Singapore). Then, both RNA samples were treated with 2 units of the DNase I (New England Biolabs, USA) per approximately 1 μg DNA. Heating and adding EDTA to a final concentration of 2mM stopped the reaction. Afterward, RNA fractions were purified using the CleanRNA Standart kit (Evrogen, Russia), their concentrations were measured using the Qubit 3.0 fluorometer (Thermo Fisher Scientific, Singapore) and Qubit RNA BR Assay Kit (Thermo Fisher Scientific, Eugene, OR, USA). The concentration and quality of RNA were also assessed using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Singapore). Samples meeting the following criteria: $\text{A}_{260}/\text{A}_{280} \geq 1.8$ and $\text{A}_{260}/\text{A}_{230} \geq 1.8$ were treated with DNAase I (New England Biolabs, USA) at a ratio of 2 units per approximately 1 μg of DNA. The reaction was stopped by heating and adding EDTA to a final concentration of 2 mM. Following DNase treatment, RNA was purified using the CleanRNA Standard kit (Evrogen, Russia). Final RNA concentrations were measured with a Qubit 3.0 fluorometer (Thermo Fisher Scientific, Singapore) using the Qubit RNA BR Assay Kit (Thermo Fisher Scientific, Eugene, OR, USA).

cDNA preparation

To prepare cDNA libraries, RNA samples were polyadenylated with Poly(A) polymerase (TransGen Biothech CO, China). We used 0.3 µg of RNA from the first sample and 1.5 µg from the second sample per reaction. The polyadenylated RNA was cleaned up with AMPure XP beads (Beckman Coulter, Brea, CA, USA). Its concentrations was measured with a Qubit 3.0 fluorometer (Thermo Fisher Scientific, Singapore) and the Qubit RNA BR Assay Kit (Thermo Fisher Scientific, Eugene, OR, USA). cDNA synthesis was performed using the Mint kit (Evrogen, Russia). First strand was generated with a 3' primer comprising an oligo(dT) sequence, complementary to the poly(A) tail. A unique feature of the MMLV-based reverse transcriptase in this kit is the addition of several non-templated nucleotides (PlugOligo) to the 5' end of the first strand to prevent spurious annealing. Subsequently, double-stranded cDNA was synthesized by PCR amplification using adapters with identical sequences at both ends. After 24–26 amplification cycles, prior to reaction saturation, the double stranded cDNA libraries were fractionated on a 1.5 % agarose gel alongside a Quick-Load 50 bp DNA Ladder (New England Biolabs, Ipswich, MA, USA). cDNA of the desired size (≤ 500 bp in the first experiment and ≤ 350 –400 bp in the second experiment) were excised and purified using the Monarch Gel Cleanup Kit (New England Biolabs, USA). The final cDNA concentration was measured using the Qubit 3.0 fluorometer (Thermo Fisher Scientific, Singapore) and QuDye dsDNA HS (Lumiprobe, Russia).

Sequencing library preparation, sequencing and preprocessing for targeted analysis

cDNA libraries were prepared using the Native Barcoding Kit 96 V14 (Oxford Nanopore Technologies, United Kingdom). Sequencing was performed on a MinION Mk1B platform (Oxford Nanopore Technologies, United Kingdom) with compatible flow cells and managed with MinKNOW software (v.25.05.14). Basecalling was conducted with the dna_r10.4.1_e8.2_400bps_sup@v5.0.0 model. All downstream analyses were performed on Ubuntu 24.04.3 LTS. Reads with a mean Q-score below 20 were discarded. Poly-A tails, as well as PlugOligo and CDS adapters (AAGCAGTGGTATCAACGCAGAGTACGCGGG and AAGCAGTGGTATCAACGCAGAGTACT₃₀VN, respectively) were trimmed using Fastp v. 0.4.1.

Table 1. Sequencing and evaluation statistics

Categories of obtained and analyzed reads	Number of reads obtained in 2 experiments	
	Culturing in Synergy H1 plates	Culturing in falcon tubes
Total number of sequence reads	3,633,292	3,780,025
Number of reads passed Q=20 control	3,179,836	3,124,321
Mean length of oligos shorter than 500 nt	341	152
Number of obtained reads with 8-mer	7513	30,414
Number of reads with 5'-terminal 8-mers	34 (0.51%)	37 (0.53 %)
Number of reads derived from 4.5S RNA	43	19,839
Number of analyzed internal reads with 8-mer on both sides	6624	7002
Number of reads with 9-mer	1515	903

The search for reads containing specific *k*-mers was performed using grep 3.11 in a case-insensitive manner. To simplify the analysis of sequences flanking the *k*-mers, a custom Python script was implemented. Briefly, the script splits each read at the target *k*-mer using

the split() function, extracts the flanking sequences, and trims them to a uniform length of 20 bp. Flanks shorter than 20 bp or longer than 500 bp were excluded from further analysis. However, reads that did not meet the flanking-length criteria were retained as supporting information for sequence annotation. In total, over 3 million high-quality reads ≥ 50 nt in length were obtained from both experiments (Table 1) and subsequently used to identify oligonucleotides containing the 8-mer GCCAAGGC and the 9-mer GCCAAGGCG.

Search for chimeric RNA molecules and their analysis

Reads containing the target 8- or 9-mers were assembled from the two sequencing libraries as outlined in Figure 1. Observing that the dataset containing the 8-mers was substantially larger in both experiments (Table 1) compared to that of 9-mers, only oligonucleotides with the GCCAAGGC 8-mer were analyzed in this study. Within each experiment, these reads were sorted into two groups. Group 1 comprised reads, where the octamer was located at the 5'-end (Fig. 1). This group was compiled to test whether this positioning, previously observed in short (< 50 nt) *E. coli* oligonucleotides sequenced on an Ion Torrent PGM platform, is reproducible and may occur within longer transcripts. Reads with the 8-mer at the 5' end were segmented into the 8-mer itself, the adjacent 20-mer on its 3'-side (right flank) and the remaining part of the read. The flanking 20-mers were then mapped to the genome of *E. coli* K-12 MG1655 using the PrMis3MatcherOutDiffLen option of Matcher software (mathcell.ru/prog/ReadTools.rar).

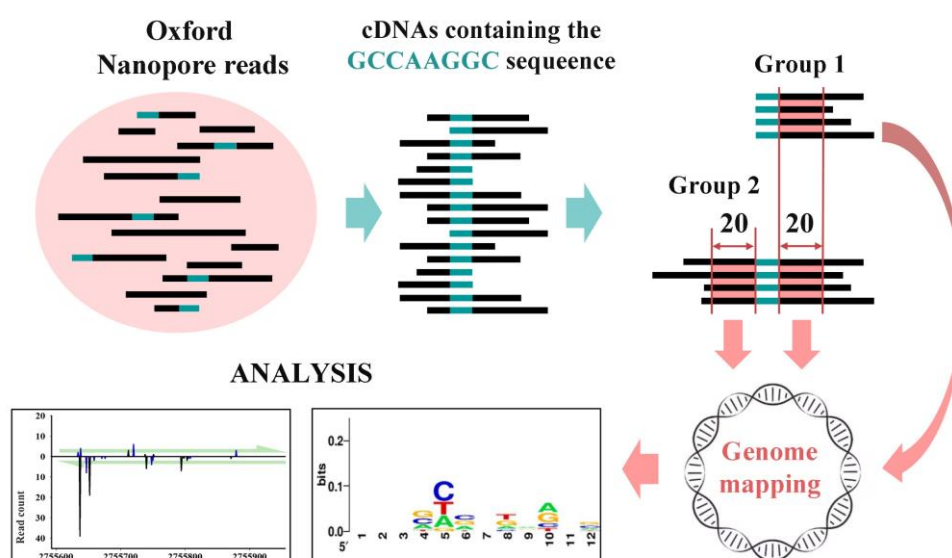


Fig. 1. Scheme illustrating the strategy used in the study.

Group 2 contained the remaining reads, where 8-mer was internal and flanked by at least 20 nucleotides on each side. They were divided such that the GCCAAGGC of interest was surrounded by 20-mers on both sides. For these reads, the genuine sequences replaced by GCCAAGGC were identified using neighboring 20-mers as markers of their origin. Only exact matches to the genomic sequence were allowed. However, when evaluating potential chimeras with 5'-terminal 8-mers, fusions containing a short (up to 2 bases) insertion between the octamer and its adjacent flank were also examined.

The in-house program DNATool was then used to navigate along the annotated genome and determine the position and orientation of the mapped 20-mers relative to the overlapping genes. From the set of reads where both flanking 20-mers were successfully mapped, we removed those containing genomic sequences from the 45 loci where the GCCAAGGC octamer occurs naturally. This purified set was then applied for subsequent analysis to make a

first step in characterization of intramolecular chimerization. A sequence Logos were generated using WebLogo (weblogo.berkeley.edu/logo.cgi) [26].

Statistics

To evaluate the statistical significance of the discovered intramolecular substitutions and to estimate risk of potential artefacts, caused by reverse transcriptions, PCR [27] or sequencing errors, we analyzed reads derived from five efficiently transcribed genomic loci containing the GCCAAGGC sequence. Within these reads, we specifically searched for substitutions occurring in the genomic octamers located between 20-mers that perfectly matched the reference genome.

RESULTS

It is intuitively believed, that intermolecular joining is essentially a random event. While several chimerization mechanisms in eukaryotes can account for certain selectivity, methods like RIL-seq (RNA-interaction by ligation and sequencing) [28, 29], which are widely used in RNA interactomics [30], are fundamentally based on the randomness of ligation events. It was therefore surprising to detect multiple copies of a chimeric oligonucleotide with 5'-GCCAAGGC-3' at the 5'-end of the transcript from the fimA-fimI intergenic intra-operonic region at the 3'-end [25]. Given that this octamer could originate from 45 genomic loci, 14 of which lie within transcribed regions, its involvement in chimerization is quite probable and consistent with definition of this phenomenon. The highly abundant 3'-terminal segment, however, is derived from an intergenic region. Its fusion to other transcripts likely does not represent typical intermolecular chimerization. Nevertheless, we identified numerous other gene products carrying the octamer GCCAAGGC at the 5'-end, often with an additional guanine at the junction point. Therefore, the ligation of either the 8-mer GCCAAGGC or the 9-mer GCCAAGGCG with products from other genes was considered a genuine chimerization event.

Proposing that such chimeras might occur during transcription initiation, we searched and identified promoters for 28 distinct types of 3'-terminal oligonucleotide segments, containing the junction point with 9-mers located within regions typically associated with transcription “bubble” formation (positions -11+1 relative to the transcriptional start point [31]). Priming of RNA synthesis by extraneous oligonucleotides in this region is highly probable. An additional 34 oligonucleotide types flanking the 5'-terminal 9-mers were located 2–20 bp downstream, an observation that remains consistent with a promoter-based chimerisation hypothesis. However, 50.4 % of the remaining 3'-terminal oligonucleotides originated from regions significantly distant from the nearest promoter. It appears that many of these were prone to form stable secondary structures precisely at the chimerization site. At such regions, transcription pausing, involving RNA polymerase backtracking, may occur, allowing short oligonucleotides to enter the complex and compete with the nascent RNA strand in the polymerization site of the enzyme. Thus, based on the data, derived from Ion Torrent PGM sequencing of transcripts shorter than 50 nt, we propose two molecular mechanisms for chimerization with short oligonucleotides: one involving transcriptional priming at promoters and the other occurring at RNA polymerase pause sites. The primary goal of this study was to test these hypotheses using long-read sequencing data obtained from the Oxford Nanopore MinION platform.

Oligonucleotide 5'-GCCAAGGC-3' can prime transcription at promoter sites and chimerize fragments of rRNAs

Using different modes of culturing for *E. coli* MG1655 in mineral medium M9, we obtained two large sets of intracellular RNAs and collected reads containing 8-mer 5'-GCCAAGGC-3' in their sequence. In both experiments, we detected oligonucleotides,

containing extraneous octamer at the 5'-end and their number appeared almost similar (34 and 37 reads). Notably, that this similarity was expected, as a great difference between experiments in the numbers of reads containing indicated octamer was mainly due to the exceptional abundance of 4.5S signal recognition particle RNA encoded by the *ffs* gene in the second set (Table 1). The sizes of the remaining sets were comparable, as well as the percentage of 5'-terminally fused oligonucleotides.

We precisely identified the origins of the partners in chimeric molecules bearing the 5'-terminal octamer for 22 molecules in the first set and for 18 chimeras in the second. When we allowed for a small gap at the junction point, these numbers increased to 28 and 26 reads, respectively. In both experiments, the majority of chimeras consisted of 8-mers fused to fragments of large ribosomal RNAs. Being clearly overrepresented (27 and 22 fusions in the first and second experiments, respectively), such products hypothetically may result from the transcription-mediated rescue of stalled RNA polymerase complexes at genomic loci that encode highly structured oligonucleotides. It is significant that identical chimeric molecules were detected in multiple copies across both experiments (Fig. 2,a). It is even more significant that along with RNA-fragments derived from sequence encoding 16S rRNA (chimeras from position 224047), we also observed chimerization of their antisense products (aRNAs) (chimeras from position 225155).

a) Fusions potentially formed in stalled transcription complexes

Gene	Position of fusion	Str	Str	Copy numbers	
				Exp 1	Exp 2
<i>rrsH</i>	224047	+	In RNA: <u>GCCAAGGC</u> GACGATCCCTAGCTGGTCTGAGAGGATGACC In DNA: <u>ACCTAGGCGACGATCCCTAGCTGGTCTGAGAGGATGACC</u>	7	4
<i>rrsH</i>	225155	+	In aRNA: <u>GCCAAGGC</u> CCGGGAACGTATTACCGTGGCATTCTGATC In DNA: <u>TACAAGGCCCGGGAACGTATTACCGTGGCATTCTGATC</u>	8	13

b) Transcription priming

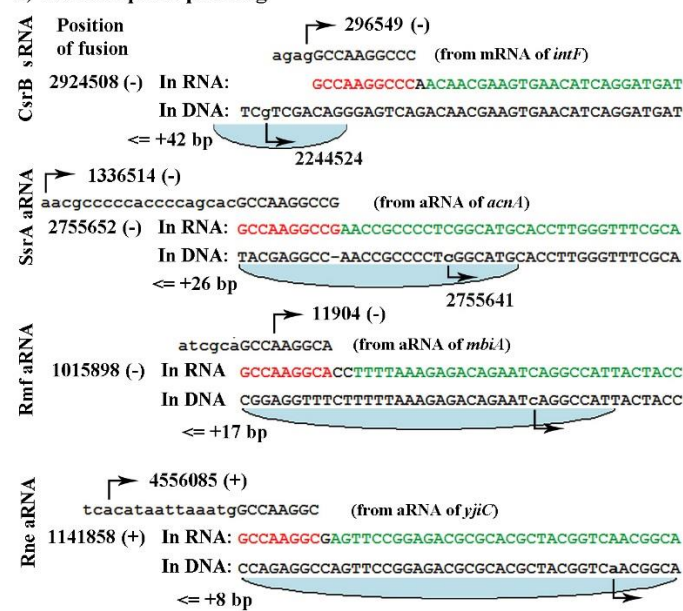


Fig. 2. Intermolecular fusions supporting the hypothesis of transcription-mediated chimerization. **a)** Two types of transcripts derived from 16S rRNA genes with similarly modified, but sequence-different 5'-ends (underlined). **b)** Potential 5'-terminal fusions, mediated by transcription priming at promoter regions. Sequences from reads are shown in red and green. The red segments highlight sequence similarity with octamer GCCAAGGC. Arrows mark transcription start points predicted by PlatProm [32]. Green parts correspond to genomic sequences with their 5'-end positions noted on the left. Rightward arrows show the predicted TSPs in the modification site. Leftward arrows with numbers, accompanied by blue half-ellipses, outline the promoter region. Here and below, all RNA sequences are shown with thymine (T) instead of uridine (U) to facilitate their comparison with DNA.

The remaining five molecules bearing the 5'-terminal octamers were present as single-copy samples. As expected, the junction points with the 8-mer in four of these molecules were located within promoter regions (Fig. 2,b). An ideal configuration of the fusion and transcription start point (TSP) was observed for the chimerization event involving the small regulatory RNA CsrB. Its synthesis is initiated from the TSP at position 2244524, which is accurately predicted by PlatProm [32] and identified experimentally [33]. The modifying oligonucleotide (GCCAAGGCC, Fig. 2,b) could be produced as a potential abortive product from the intragenic promoter within the *intF* gene. The proximity of the nearest promoter to the junction point in SsrA (tmRNA) also supports a chimerization mechanism via transcription priming with GCCAAGCCG (Fig. 2,b).

Although the last two reads shown in Fig. 2,b, which resulted from octamer fusion with transcripts from the *rmf* and *rne* genes, are less likely products of priming-mediated chimerization, this is also possible due to plasticity of promoter DNA and potential errors of *in silico* promoter prediction. Thus, based on the data obtained, we concluded that the emergence of chimeras terminally modified by linking to a specific octamer in promoter regions is possible, but optimal combinations of promoter architecture and priming oligonucleotides are rare.

Transcriptome of *E. coli* contains thousands of oligonucleotides with intramolecular substitution of genome-encoded sequences with GCCAAGGC

Obtaining a relatively small number of long oligonucleotides with the 5'-terminally fused octamer, we collected all sequence reads containing the internal motif of GCCAAGGC and found that many derived from genomic loci absent among the 45 genes known to contain this motif in their genomic sequences. In the first experiment, 20-nucleotide sequences flanking the left borders of the 8-mers were predominantly derived from 16S and 23S ribosomal RNAs, while in the second experiment, the main contribution came from 4.5S RNAs. For all that, approximately 52 % of reads identified by the 5'-ends of the left-flanking 20-mers were found in at least one copy in both experiments. Assuming that 8-mers join different transcripts, we combined reads with the internal 8-mer from both experiments and mapped to the genome sequences of the 20-mers flanking octamers on both sides. Surprisingly, all belonged to the same types of transcripts, likely bearing a short replacement of the genomic sequence.

We then analyzed the sequences of the flanking 20-mers (Fig. 3,a and 3,f), as well as the substituted genomic motifs between them (Figures 3,b and 3,e). This analysis was initially focused on reads where either only the left flank (Fig. 3,a and 3,b) or only the right 20-mer (Fig. 3,e and 3,f) were precisely matched to the genome. In both cases, sequencing logos were generated for reads with a copy number of at least 3. Reflecting the presence of modified oligonucleotides derived from different genomic loci, both 20-mers flanking the internal octamers showed no significant sequence preferences (Fig. 3,a and 3,f).

In contrast, octamer-substituted genomic 8-mers exhibited a clear bias towards GC dinucleotides present at the octamer ends (Fig. 3,b and 3,e). The central part of the replaced genomic motifs showed a high diversity without any similarity to the octamer (Fig. 3,c and 3,d) in both sense (black) and antisense (red) transcripts. To determine whether the unexpected alignment dependence (Fig. 3b and 3e) might result from variability in the length of the substituted genomic sequences, we collected 1032 sequence reads in which both flanking 20-mers matched the genome precisely and evaluated intermediate sequences. This set comprised fifty distinct types of substituted oligonucleotides, each represented by at least two copies in the combined experimental data. As hypothesized, the length of these segments varied considerably within this subset. The shortest substituted segment was derived from 23S rRNA and contained only 3 bases. Identified in only one copy (Fig. 4,a), it was not represented by its flanking 20-mers in sets compiled for the sequence logo generation (Fig. 4, b–d). Newly generated sequence logos (Fig. 4,b) were intentionally obtained for sequences

that did not overlap with those used for Figure 3, for which either only the left flank or only the right 20-mer were precisely identified. For all that, they again showed no significant nucleotide preferences in the flanking 20-mers (Fig. 4,b). However, the sequence logo generated from the replaced 8-mers in 17 non-redundant oligonucleotides where both adjacent 20-mers matched the genome exactly, revealed the expected non-randomness at both termini. Therefore, the asymmetry observed in the logos of Figure 3 may be due to partial alignment of substituted segments with high length variability (Fig. 4,c), whereas intramolecular integration of identical octamers likely arises from the replacement of genomic sequences by 8-mers rather than their bilateral ligation to different transcripts.

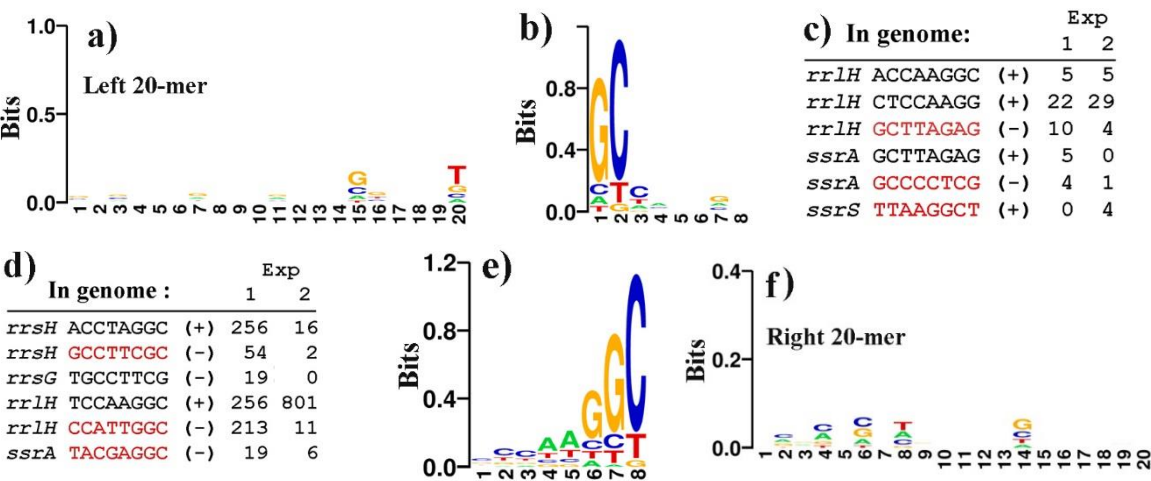


Fig. 3. Reads containing the 8-mer GCCAAGGC instead of genomic sequences exhibit random nucleotide distribution in the 20-mers surrounding the insertion site, while substituted oligonucleotides showed a GC preference at both borders. **a, f)** Sequence Logos representing the 56 (**a**) and 67 (**f**) non-redundant genomic sequences of 20-mers flanking the left and right sides, of the internal 8-mers respectively. **b)** Sequence Logo created for the 56 non-redundant genomic sequences of 8-mers located to the right of the precisely genome-aligned left 20-mer. **e)** Same as b, but for 67 non-redundant 8-mer genomic sequences located to the left of the precisely annotated right 20-mer. Sequence logos were created using WebLogo at <https://weblogo.berkeley.edu/logo.cgi> [26]. Panels **c** and **d** exemplify diversity and multiplicity of 8-mers replaced by octamer GCCAAGGC on the right of left 20-mer and on the left of the right 20-mer, respectively. Sequences marked in red correspond to antisense products of genes.

To assess the error rate in the obtained sequencing reads, we used transcripts that inherently contain GCCAAGGC in their sequence. Although most of these are located in antisense orientation, we found genes (*ffs*, *hns*, *ptsI*, *rpmJ*, and *rpoA*) that produced a sufficient number of transcripts for statistical analysis. Reads containing their products were extracted using 20-mers flanking the intrinsic GCCAAGGC and the changes between these flanks were analyzed. The frequency of single base-pair substitutions, estimated from reads of these five genes, yielded on average 1.52 ± 0.47 %. While this substitution rate is sufficient to explain the occasional derivation of the GCCAAGGC motif from sequences differing by one or two nucleotides (Fig. 4,c), it cannot explain targeted mutagenesis toward an invariant sequence. More detailed insight came from analysis of 16604 transcripts derived from the *ffs* gene. In this set only 168 erroneous sequences were detected within the context of the model octamer (1.0 %). However, 62 oligonucleotides contained a shortened version (3–7 nt) of GCCAAGGC (0.37 %), while 38 reads contained an extended octamer, reaching up to 13 nt in three reads (0.22 %). Of these derivatives, 80 preserved GGC at the 3'-end and 31 retained GCC at the 5'-end. Thus, the specific pattern of GCC and GGC dispositioning likely guides intramolecular chimerization mediated by GCCAAGGC.

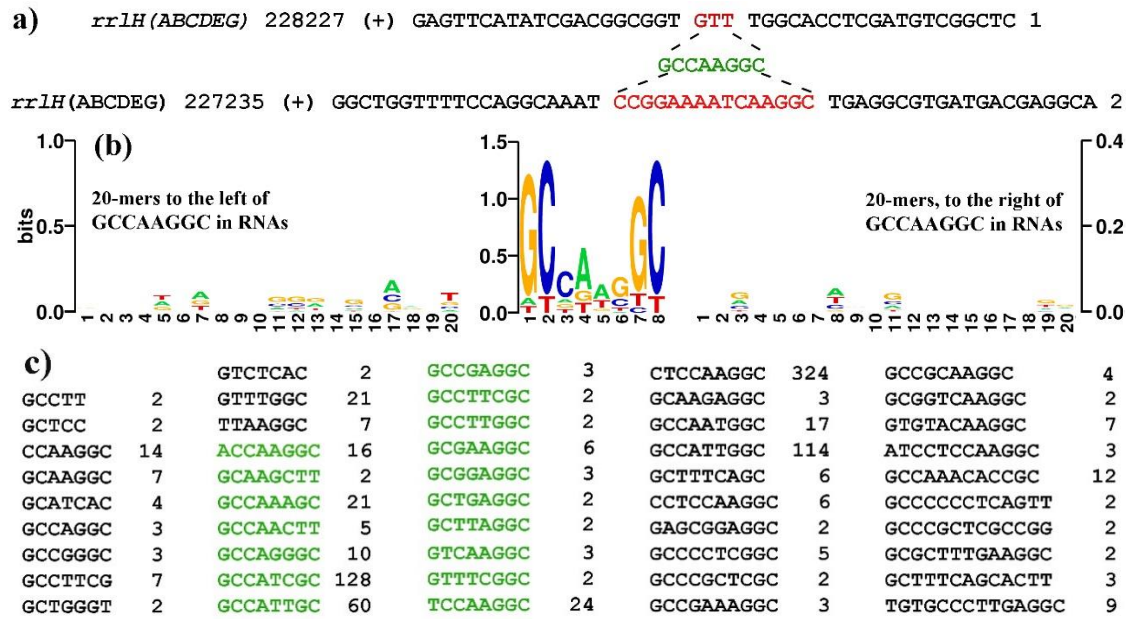


Fig. 4. The length of modules post-transcriptionally substituted by the 8-mer GCCAAGGC varied from 3 to 15 nucleotides, with a preference for length of 8 nt, which contained dominant GC dinucleotides on both borders. **a)** Genomic sequences of 23S rRNA fragments that are identical across all seven operons (*rrlH*, A, B, C, D, E and G) where the 8-mer replaced either the smallest (position 228227) or longest (position 228235) segment, indicated by the 5'-ends of the left flanking 20-mers, respectively. **b)** Sequence logos generated from non-redundant genomic sequences of 17 substituted 8-mers (center), 47 left flanking 20-mers (left panel) and 45 right flanking 20-mers (right panel). **c)** List of genomic motifs replaced by 8-mer. Octamers used for logo generation are highlighted in green. Numerals indicate the copy count of reads containing each identified motif.

No convincing evidence was found supporting participation of GCCAAGGC in intermolecular fusion

Although chimeric RNAs are classified into different categories by various criteria [5], all are defined by the joining of transcripts from different genes. In this study, however, we identified only seven fusion types mediated by GCCAAGGC that formally satisfy the requirement for intramolecular junctions. Six of these are shown in Figure 5. They include the two most abundant fusions, which originate from the same genomic site (positions 2730907–2730909), as well as cases in which nearly identical fragments of the *rrsH* gene antisense product join to transcripts from the homologous *rrsA* and *rrsC* genes.

Apparent bilateral fusions		GCCAAGGC	
RrsG (-) 2730907	GT T GGTGGGGTAACGGCTC	CCAAGGC	GACGATCCCTAGCTGGTCT 224048 RrsH (+) 23
RrsG (-) 2730908	TGT T GGTGGGGTAACGGCTC	ACCAAGGC	GACGATCCCTAGCTGGTCT 224048 RrsH (+) 27
RrsG (-) 2730909	TTGT T GGTGGGGTAACGGCT	CACCAAGGC	GACGATCCCTAGCTGGTCT 224048 RrsH (+) 1
RrsH (+) 224018	CTAGTAGGTGGGGTAACGGCT	CCTAGGC	GACGATCCCTAGCTGGTCT 224048 RrsH (+) -
RrlG (-) 2728939	CGGCGAGCGAACGGGGAGCA	CACCAAGGC	CTGAATCAGT A TGTGTGTT 3943978 RrlC (+) 1
RrlG (-) 2728939	CGGCGAGCGAACGGGGAGCA	GCCCAGAGC	CTGAATCAGTGTGTGTGTT 2728910 RrlG (-) -
RrsH (-) 224010	CCATCTGGGCACATCCGATG	GCAAGAGGC	CCGAAGGT A CCCCCTCTTTG 4035147 RrsA (-) 2
RrsC (-) 3942047	CCATCTGGGCACATC T GATG	GCAAGAGGC	CCGAAGGTCCCCCTCTTTG 223981 RrsH (-) 3
RrsH (-) 224010	CCATCTGGGCACATCCGATG	GCAAGAGGC	CCGAAGGTCCCCCTCTTTG 223981 RrsH (-) -

Fig. 5. Examples of sequencing reads where the internal 8-mer separated oligonucleotides attributed to different genomic loci (coordinates on left and right). The single-nucleotide difference distinguishing homologous genes is marked in red. The octamer was replaced by the corresponding genomic sequence in each read and is highlighted in green to demonstrate the sequence variation.

However, in each instance, the annotation of the fused genes relies on a single-nucleotide difference (marked in red), distinguishing the otherwise identical sequences of homologous genes across the seven ribosomal operons of *E. coli*. Consequently, we conclude that our analysis do not support the involvement of GCCAAGGC in genuine intermolecular coupling.

Sense and antisense transcripts share non-random sites of intramolecular chimerization

The Figure 6 shows the distribution of sites modified by GCCAAGGC in the context of 16S and 23S ribosomal RNAs (RrsH and RrlH), as well as tmRNA encoded by *ssrA* gene. tmRNA releases stalled ribosomes by functioning as both a tRNA and a mRNA.

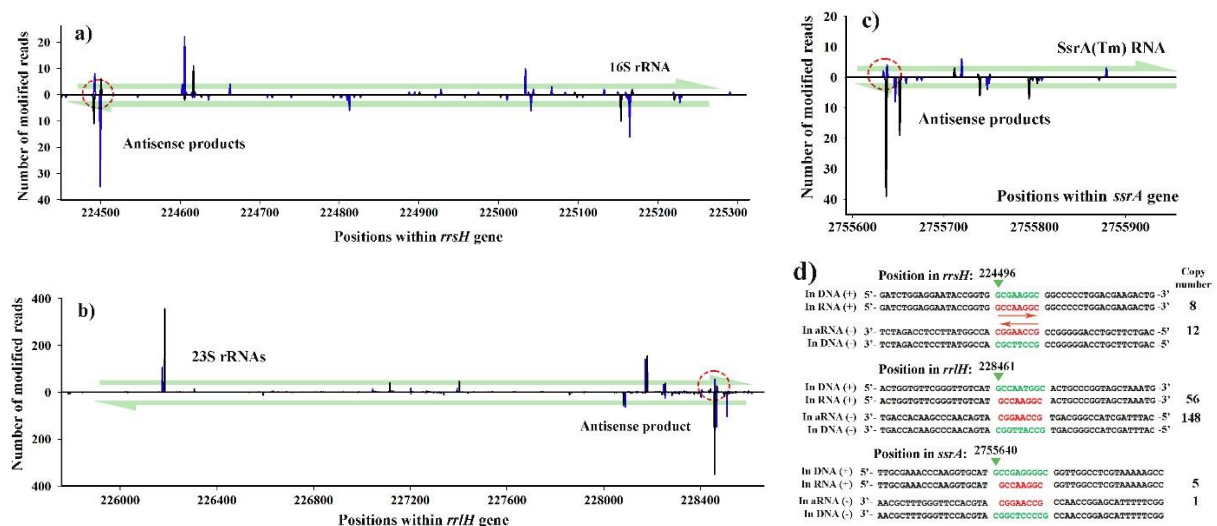


Fig. 6. Targets of intramolecular chimerization are distributed non-randomly across translation associated RNAs and frequently coincide in sense and antisense gene products. **a-c)** Distribution of sequencing reads containing GCCAAGGC substitutions along the 16S (**a**), 23S (**b**) and tmRNAs (**c**), as well as across their complementary antisense RNAs (aRNAs, indicated). The blue and black bars represent the left and right boundaries of the modifying octamer, respectively. Regions exemplified in panel **d** are marked with circles. **d)** Representative examples of RNAs transcribed in sense and antisense orientations. Genomic sequences are highlighted in green for both sense products and aRNAs. The green arrowhead indicates the nucleotide corresponding to the genome position specified nearby.

Thus, the distribution of modified sites across homological rRNAs is represented using only the products of the first operon in the *E. coli* genome. This operon (*rrsH* – *ileV* – *alaV* – *rrlH* – *rrfH*) also contains genes for IleV and AlaV tRNAs, as well as for 5S RNA RrfH. However, all octamer-mediated modification detected in both experiments were located exclusively in either RrsH (Fig. 6,a) or RrlH (Fig. 6,b). In both RNAs, modifications were distributed non-randomly, with preferred sites located near the beginning or end of each RNA molecule. tmRNA, which is considerably shorter (363 nucleotides), was modified at only a few sites and relatively small number of reads containing the GCCAAGGC sequence were detected (Fig. 6,c). Nevertheless, modifications in regions corresponding to genomic coordinates 2755667, 2755679 and 2755768 were observed in both experiments. An unexpected finding of this study was the large number of RNAs transcribed as antisense products (aRNAs). Such transcripts, derived from tmRNA (Fig. 6,c), and from the beginning of RrsH (Fig. 6,a) were even more abundant than their sense-transcribed counterparts. Moreover, when modified aRNAs originated from regions overlapping with modifications sites in rRNAs or tmRNA (red circles in figures 6,a-c), the observed substitutions precisely match those in the corresponding rRNAs and tmRNA (Fig. 6,d). These results suggest that antisense RNAs may play a role in the biogenesis of intra-molecularly modified RNAs.

DISCUSSION

Although chimeric RNAs can be classified into multiple categories depending on their biogenesis, coding potentials, the locations of fusion points relative to parental gene exons, and others factors [5], none of the existing classification aligns with the phenomenon observed in our study. Cross-strand chimerization, which results from convergent transcription of sense and antisense RNAs [10], likely shares certain features with our observations, due to the involvement of aRNAs in both processes. However, while identifying numerous cases of overlapping modification sites in complementary sense and antisense products, we detected no evidence of their direct cross-molecular fusion. Therefore, it is likely that intramolecular substitutions reported here are more closely related to post-transcriptional editing than to conventional RNA chimerization.

Although post-transcriptional editing in bacteria is already a subject of focused study [34], research has primarily centered on adenine-to-inosine (A-to-I) conversion. This modification is subsequently recognized as guanosine by cellular machinery, enabling bacteria to generate functional protein diversity without altering genomic DNA. In contrast, our study revealed the replacement of variable-length RNA segments with conserved octamer sequence, i.e. a mechanism more akin to chimerization through recombination than to canonical editing. The predominance of GC dinucleotides at both ends of the replaced sequences, which are also present in the octamer, complicates the mechanistic interpretation, as it likely necessitates an intermediate that facilitates sequence exchange despite the absence of complementarity at the junction.

While some of the oligonucleotides containing the internal GCCAAGGC substitution originated from mRNA (RxC, HdeA, RplV and others), the vast majority are derived from structured RNAs. These include ribosomal RNAs (16S and 18S rRNA), tmRNA, and small regulatory RNAs such as CsrB and SsrS. Intramolecular modifications that integrate the octamer with terminal self-complementarity (GCC- - GGC) should influence RNA folding and, consequently, alter molecular function. At the same time, the observed variations in the length of substituted segments within mRNAs like RxC and RplV may be even more interesting, as they may result in single amino acid substitutions and frameshifting. A detailed analysis of such mRNA modifications, alongside the structural impact of the inserted octamer on RNA conformation, will be a primary focus of future research.

The most intriguing yet least predictable implication of this work concerns the potential functionality of the observed intramolecular chimerization. The most realistic current hypothesis posits that the post-transcriptional substitutions of GCCAAGGC into fragmented or otherwise damaged RNAs could serve as a decay signal. If true, this would imply the existence of multifunctional cellular hubs capable of assembling these “tagged” molecules, likely via specific base-pairing interactions. However, the *E. coli* genome does not contain regions enriched in sequences complementary to GCCAAGGC. This absence presents a significant mechanistic puzzle, indicating that further progress in this field will require novel, heuristic ideas and approaches.

The study was supported by a grant from the Russian Science Foundation No. 24-14-00433, <https://rscf.ru/project/24-14-00433/>.

REFERENCES

1. Shi X., Singh S., Lin E., Li H. Chimeric RNAs in cancer. *Adv. Clin. Chem.* 2021. V. 100. P. 1–35. doi: [10.1016/bs.acc.2020.04.001](https://doi.org/10.1016/bs.acc.2020.04.001)
2. Kannan K., Wang L., Wang J., Ittmann M.M., Li W., Yen L. Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. *Proc. Natl. Acad. Sci. U S A.* 2011. V. 108. No. 22. P. 9172–9177. doi: [10.1073/pnas.1100489108](https://doi.org/10.1073/pnas.1100489108)
3. Qin F., Song Z., Babiceanu M., Song Y., Facemire L., Singh R., Adli M., Li H. Discovery of CTCF-sensitive Cis-spliced fusion RNAs between adjacent genes in

- human prostate cells. *PLoS Genet.* 2015. V.11. No. 2. Article No. e1005001. doi: [10.1371/journal.pgen.1005001](https://doi.org/10.1371/journal.pgen.1005001)
4. Kumar-Sinha C., Kalyana-Sundaram S., Chinnaiyan A.M. Landscape of gene fusions in epithelial cancers: seq and ye shall find. *Genome Med.* 2015. V. 7. P. 129. doi: [10.1186/s13073-015-0252-1](https://doi.org/10.1186/s13073-015-0252-1)
 5. Li Z., Qin F., Li H. Chimeric RNAs and their implications in cancer. *Curr. Opin. Genet. Dev.* 2018. V. 48. P. 36–43. doi: [10.1016/j.gde.2017.10.002](https://doi.org/10.1016/j.gde.2017.10.002)
 6. Rowley J.D. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature.* 1973. V. 243. No. 5405. P. 290–293. doi: [10.1038/243290a0](https://doi.org/10.1038/243290a0)
 7. Liu D., Xia J., Yang Z., Zhao X., Li J., Hao W., Yang X. Identification of chimeric RNAs in pig skeletal muscle and transcriptomic analysis of chimeric RNA TNNI2-ACTA1 V1. *Front. Vet. Sci.* 2021. V. 8. Article No. 742593. doi: [10.3389/fvets.2021.742593](https://doi.org/10.3389/fvets.2021.742593)
 8. Khan S., Tang Y., Song Z., Chen X., Wang L., Zhang C., Qin F. Beyond chromosomal rearrangements: The expanding landscape of gene fusions and chimeric RNAs. *Gene & Protein in Disease.* 2025. V. 4. No. 1. Article No. 3641. doi: [10.36922/gpd.3641](https://doi.org/10.36922/gpd.3641)
 9. Chwalenia K., Facemire L., Li H. Chimeric RNAs in cancer and normal physiology. *Wiley Interdiscip Rev RNA.* 2017. V. 8. No 6. Article No. e1427. doi: [10.1002/wrna.1427](https://doi.org/10.1002/wrna.1427)
 10. Wang Y., Zou Q., Li F., Zhao W., Xu H., Zhang W. Deng H., Yang X. Identification of the cross-strand chimeric RNAs generated by fusions of bi-directional transcripts. *Nat. Commun.* 2021. V. 12. Article No. 4645 (2021). doi: [10.1038/s41467-021-24910-2](https://doi.org/10.1038/s41467-021-24910-2)
 11. Li Y., Liu Y., Xie Y., Wang Y., Wang J., Wang H., Xia L., Xie D. Long-read RNA sequencing enables full-length chimeric transcript annotation of transposable elements in lung adenocarcinoma. *BMC Cancer.* 2025. V. 25. No. 1. Article No. 482. doi: [10.1186/s12885-025-13888-5](https://doi.org/10.1186/s12885-025-13888-5)
 12. Zhang H., Hou Y., Miranda L., Campbell D.A., Sturm N.R., Gaasterland T., Lin S. Spliced leader RNA trans-splicing in dinoflagellates. *Proc. Natl. Acad. Sci. U S A.* 2007. V. 104. P. 4618–4623. doi: [10.1073/pnas.0700258104](https://doi.org/10.1073/pnas.0700258104)
 13. Sun Y., Li H. Chimeric RNAs Discovered by RNA Sequencing and Their Roles in Cancer and Rare Genetic Diseases. *Genes (Basel).* 2022. V. 13. No. 5. Article No. 741. doi: [10.3390/genes13050741](https://doi.org/10.3390/genes13050741)
 14. Nowacki M., Vijayan V., Zhou Y., Schotanus K., Doak T.G., Landweber L.F. RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature.* 2008. V. 451. P. 153–158. doi: [10.1038/nature06452](https://doi.org/10.1038/nature06452)
 15. Fang W., Landweber L.F. RNA-mediated genome rearrangement: hypotheses and evidence. *Bioessays.* 2013. V. 35. No. 2. P. 84–87. doi: [10.1002/bies.201200140](https://doi.org/10.1002/bies.201200140)
 16. Yan Z., Huang N., Wu W., Chen W., Jiang Y., Chen J., Huang X., Wen X., Xu J., Jin Q., et al. Genome-wide colocalization of RNA-DNA interactions and fusion RNA pairs. *Proc. Natl. Acad. Sci. U S A.* 2019. V. 116. No. 8. P. 3328–3337. doi: [10.1073/pnas.1819788116](https://doi.org/10.1073/pnas.1819788116)
 17. Singh S., Li H. Comparative study of bioinformatic tools for the identification of chimeric RNAs from RNA Sequencing. *RNA Biol.* 2021. V. 18. P. 254–267. doi: [10.1080/15476286.2021.1940047](https://doi.org/10.1080/15476286.2021.1940047)
 18. Ma C., Shao M., Kingsford C. SQUID: transcriptomic structural variation detection from RNA-seq. *Genome Biol.* 2018. V. 19. No. 1. Article No. 52. doi: [10.1186/s13059-018-1421-5](https://doi.org/10.1186/s13059-018-1421-5)
 19. Jang Y.E., Jang I., Kim S., Cho S., Kim D., Kim K., Kim J., Hwang J., Kim S., Kim J., et al. ChimerDB 4.0: an updated and expanded database of fusion genes. *Nucleic Acids Res.* 2020. V. 48. No. D1. P. D817–D824. doi: [10.1093/nar/gkz1013](https://doi.org/10.1093/nar/gkz1013)

20. DSouza D., Bik L., Giwa O., Cohen S., Barazany H.L., Siegal T., Frenkel-Morgenstern M. ChiTaRS 8.0: the comprehensive database of chimeric transcripts and RNA-seq data with applications in liquid biopsy. *Nucleic Acids Res.* 2025. V. 53. No. D1. P. D1302–D1312. doi: [10.1093/nar/gkae1126](https://doi.org/10.1093/nar/gkae1126)
21. Mei Y., Cheng Z., Lu Y., Wu S., Chen X. Comprehensive resource for transcription read through events in healthy human tissues. *Sci. Data.* 2025. V. 12. No.1. Article No. 1176. doi: [10.1038/s41597-025-05557-w](https://doi.org/10.1038/s41597-025-05557-w)
22. Haas B.J., Gevers D., Earl A.M., Feldgarden M., Ward V.D., Giannoukos G., Ciulla D., Tabbaa D., Highlander K.S., Sodergren E., et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* 2011. V. 21. No. 3. P. 494–504. doi: [10.1101/gr.112730.110](https://doi.org/10.1101/gr.112730.110)
23. Doose G., Alexis M., Kirsch R., Findeiß S., Langenberger D., Machné R., Mörl M., Hoffmann S., Stadler P.F. Mapping the RNA-Seq trash bin: unusual transcripts in prokaryotic transcriptome sequencing data. *RNA Biol.* 2013. V. 10. No. 7. P. 1204–1210. doi: [10.4161/rna.24972](https://doi.org/10.4161/rna.24972)
24. Lloréns-Rico V., Serrano L., Lluch-Senar M. Assessing the hodgepodge of non-mapped reads in bacterial transcriptomes: real or artifactual RNA chimeras? *BMC Genomics.* 2014. V. 15. No. 1. Article No. 633. doi: [10.1186/1471-2164-15-633](https://doi.org/10.1186/1471-2164-15-633)
25. Shavkunov K.S., Markelova N.Y., Alikina O.V., Glazunova O.A., Panyukov V.V., Kolzhetsov N.P., Kiselev S.S., Ozoline O.N. Products of abortive transcription can prime synthesis of chimeric oligonucleotides. *Mat. Biol. Bioinform.* 2024. V. 19. No. 2. P. 453–471. doi: [10.17537/2015.10.453](https://doi.org/10.17537/2015.10.453)
26. Crooks G.E., Hon G., Chandonia J.M., Brenner S.E. WebLogo: a sequence logo generator. *Genome Res.* 2004. V. 14. No. 6. P. 1188–1190. doi: [10.1101/gr.849004](https://doi.org/10.1101/gr.849004)
27. Peng Z., Yuan C., Zellmer L., Liu S., Xu N., Liao D.J. Hypothesis: artifacts, including spurious chimeric RNAs with a short homologous sequence, caused by consecutive reverse transcriptions and endogenous random primers. *J. Cancer.* 2015. V. 6. No. 6. P. 555–567. doi: [10.7150/jca.11997](https://doi.org/10.7150/jca.11997)
28. Bar A., Argaman L., Altuvia Y., Margalit H. Prediction of novel bacterial small RNAs from RIL-Seq RNA-RNA interaction data. *Front. Microbiol.* 2021. V. 12. Article No. 635070. doi: [10.3389/fmicb.2021.635070](https://doi.org/10.3389/fmicb.2021.635070)
29. Melamed S., Peer A., Faigenbaum-Romm R., Gatt Y.E., Reiss N., Bar A., Altuvia Y., Argaman L., Margalit H. Global mapping of small RNA-target interactions in bacteria. *Mol. Cell.* 2016. V. 63. No. 5. P. 884–897. doi: [10.1016/j.molcel.2016.07.026](https://doi.org/10.1016/j.molcel.2016.07.026)
30. Liu F., Chen Z., Zhang S., Wu K., Bei C., Wang C., Chao Y. In vivo RNA interactome profiling reveals 3'UTR-processed small RNA targeting a central regulatory hub. *Nat. Commun.* 2023. V. 14. No. 1. Article No. 8106. doi: [10.1038/s41467-023-43632-1](https://doi.org/10.1038/s41467-023-43632-1)
31. Vishwakarma R.K., Brodolin K. The σ subunit-remodeling factors: An emerging paradigms of transcription regulation. *Front. Microbiol.* 2020. V. 11. Article No. 1798. doi: [10.3389/fmicb.2020.01798](https://doi.org/10.3389/fmicb.2020.01798)
32. Shavkunov K.S., Masulis I.S., Tutukina M.N., Deev A.A., Ozoline O.N. Gains and unexpected lessons from genome-scale promoter mapping. *Nucleic Acids Res.* 2009. V. 37. No. 15. P. 4919–4931. doi: [10.1093/nar/gkp490](https://doi.org/10.1093/nar/gkp490)
33. Salgado H., Gama-Castro S., Lara P., Mejia-Almonte C., Alarcón-Carranza G., López-Almazo A.G., López-Almazo A.G., Betancourt-Figueroa F., Peña-Loredo P., Alquicira-Hernández S. et al. RegulonDB v12.0: A comprehensive resource of transcriptional regulation in *E. coli* K-12. *Nucleic Acids Res.* 2024. V. 52. P. D255–D264. doi: [10.1093/nar/gkad1072](https://doi.org/10.1093/nar/gkad1072)
34. Bar-Yaacov D., Pilpel Y., Dahan O. RNA editing in bacteria: occurrence, regulation and significance. *RNA Biol.* 2018. V. 15. No. 7. P. 863–867. doi: [10.1080/15476286.2018.1481698](https://doi.org/10.1080/15476286.2018.1481698)

Received 10.10.2025. Revised 04.12.2025. Published 11.12.2025.