Nanopore microscope identifies RNA isoforms with structural colours

Filip Bošković and Ulrich Felix Keyser

Identifying RNA transcript isoforms requires intricate protocols that suffer from various enzymatic biases. Here we design three-dimensional molecular constructs that enable identification of transcript isoforms at the single-molecule level using solid-state nanopore microscopy. We refold target RNA into RNA identifiers with designed sets of complementary DNA strands. Each reshaped molecule carries a unique sequence of structural (pseudo)colours. Structural colours consist of DNA structures, protein labels, native RNA structures or a combination of all three. The sequence of structural colours of RNA identifiers enables simultaneous identification and relative quantification of multiple RNA targets without prior amplification. Our Amplification-free RNA Target Multiplex Isoform Sensing (ARTEMIS) method reveals structural arrangements in native transcripts in agreement with published variants. ARTEMIS discriminates circular and linear transcript isoforms in a one-step, enzyme-free reaction in a complex human transcriptome using single-molecule read-out.

Single-molecule identification of multiple transcript isoforms in parallel without preamplification is critical for understanding transcriptome diversity and gene expression networks. Identification and quantification of structural arrangements in native transcripts are both challenging, and current methods do not necessarily yield results reflecting innate transcriptome diversity. Although identification of long RNA molecules is possible with existing nucleic acid detection methods, these methods lack specificity and simplicity. In addition, common approaches mainly rely on enzymatic reactions and require preamplification. These lead to inevitable biases and loss of information. RNA sequencing approaches require extensive and intricate adaptations to achieve the sequencing of transcript variants and to test their circularity. These widely used techniques face amplification and reverse transcription biases, and detection of transcript variants is affected by short reads in RNA sequencing. Recently, nanopore sequencing introduced direct RNA read-out, however, access to single-molecule information of gene expression level in combination with low-quality reads and uncertainty about the 5’ end of the transcript remain major challenges. As in previously established RNA sequencing methods, nanopore RNA-seq also suffers from enzymatic biases. Additionally, secondary structures in both RNA and complementary DNA (cDNA) contribute to biases by obstructing the binding of primers and sequencing adaptors. In a vast majority of cases for identifying disease-specific transcripts, looking for a transcriptome subset is sufficient rather than sequencing the whole transcriptome. Accurate annotation of transcripts is still challenging, and numerous transcripts remain uncatalogued.

We are in need of an enzyme-free method for targeted identification of native transcript isoforms, that avoids notoriously laborious protocols and that reveals structural arrangements on the single-molecule level.

Here, we introduce our ARTEMIS method. ARTEMIS relies on the molecular design of identifiers (IDs) for native RNA targets that reshapes the RNA ‘scaffolds’ into unique structures inspired by the DNA origami technique. IDs in ARTEMIS consist of an RNA–DNA nanostructure. Each ID is composed of sequence-specific structural (pseudo)colours that identify the target RNA and its isoforms.

With ten colours we can potentially create ten billion unique identifiers that are adaptable to both nanopore microscope and super-resolution microscopies that can be based on DNA-based point accumulation for imaging in nanoscale topography (DNA-PINT). Nanopore microscopy works via a voltage-driven translocation of negatively charged RNA IDs through a small orifice towards a positively charged electrode in an electrolyte solution. It translates the designed RNA ID into a current signal. Colours placed on each RNA ID produce downward current spikes that correspond to the size of a colour. Larger colours induce a larger blockage of current. Notably, we identified RNA targets of interest in a complex human transcriptome enabling targeted enzyme-free transcriptomics. In addition, ARTEMIS identifies physical rearrangements of transcript isoforms including the order of their structural elements and their orientation, length and circularity by using short DNA oligonucleotides (also called oligos) as inexpensive scaffolding material. In addition to DNA, we used targeted RNA origami self-assembly, that is, uncomplemented RNA as additional structural colours that can be located along the molecule or at its ends. ARTEMIS solves long-read sequencing problems of strongly structured regions or too long transcripts since our designed RNA IDs allow for efficient identification of full-length RNA. We identified isoforms of a messenger RNA and a long non-coding RNA and show that we can perform relative isoform quantification. ARTEMIS will enable studies of native RNA diversity, gene expression analysis, RNA motif nanofolding and RNA interactions with proteins and small molecules.

Results and discussion

Structural colours enable RNA isoform identification with nanopore microscope. As a first example, we introduce the concept of structural colours for three RNA target isoforms that are complemented with DNA oligonucleotides as shown in Fig. 1a. In this example, the molecular design of each isoform-specific ID is represented by a molecular pattern with two sites that have structural colour ‘1’, ‘2’ or ‘3’. We create a structural colour by interspersing an integer number of structural units. Each structural colour has a linearly increasing molecular weight with its ‘colour number’ (design

Cavendish Laboratory, University of Cambridge, Cambridge, UK. E-mail: ufk20@cam.ac.uk
Fig. 1 | ARTEMIS identifies multiple RNA targets using structural colours and nanopore microscopy. 

**a**, RNA isoform-specific ID fabrication using structural colours (red, 1; orange, 2; and green, 3). Three RNA isoforms are linearized with complementary DNA oligos and labelled with an ‘exon-specific’ structural colour, thus creating RNA IDs (RNA–DNA hybrid) ‘12’, ‘13’ or ‘23’. Each structural colour is composed of an integer number of structural units. 

**b**, The ID with ten different structural colours is read by passing it through the nanopore microscope. A structural colour consists of an integer number of structural units (0–10) that are placed sequentially and read as one structural colour. Each structural unit is composed of the part that binds to a target and has the overhang (docking strand, black) and the imaging strand (red) that is complementary to the overhang and has a terminal structure (monovalent streptavidin or DNA cuboid). 

**c**, The nanopore microscope detects up to ten structural colours within the same molecular ID. An example nanopore event in which each structural colour is identified by the nanopore microscope is shown. 

**d**, Single-molecule read-out of structural colours and their identity as assigned in example nanopore events. 

**e**, Ten-colour ID heatmap indicates error rates for assigning structural colours with >85% accuracy. Sample size (N) is 60 nanopore events.

details of structural colours are illustrated in Supplementary Fig. 1). The basic design of the structural unit (Fig. 1b and Supplementary Fig. 1) consists of a docking strand (shown in black), complementary to the specific sequence on the target RNA, and the overhang sequence (red). The imaging strand is complementary to the docking strand’s overhang and contains a structure of fixed molecular
weight. For ARTEMIS, we employed monovalent streptavidin19 or a DNA cuboid nanostructure (designs are depicted in Supplementary Figs. 1 and 2a,b; DNA cuboid oligonucleotides are listed in Supplementary Table 1).

The sequence of structural colours made from RNA–DNA hybrids (RNA ID) can be conveniently read using a solid-state nanopore microscope20 (Fig. 1b) as a rapid, enzyme-free and affordable alternative for both short- and long-read sequencing. ARTEMIS with its direct nanopore read-out avoids the technical artefacts of RNA-seq and the imperfections of the motor proteins used in nanopore sequencing21. RNA ID fabrication and identification do not require reverse transcription or preamplification; hence, it offers rapid one-step direct detection of native RNA targets in the whole transcriptome.

Using the well-characterized ability of nanopore microscopes to detect molecular weight22,23, we show, as in fluorescent microscopy, identification of multiple colours (Fig. 1b–c). We designed and tested structural colours with up to ten levels as shown in the schematic in Fig. 1b (a four-colour ID is shown in Supplementary Figs. 2 and 4; oligonucleotides are shown in Supplementary Tables 2 and 3). Ourten-colour palette is shown in Fig. 1b–c, demonstrating the simultaneous detection of ten colours at the single-molecule level in a nanopore microscope (Fig. 1c,e). Example events and design details for the ten-colour ID are shown in Supplementary Figs. 3 and 4; oligonucleotides are listed in Supplementary Tables 2 and 4. The number of possible colour combinations depends on the length of the RNAs. We estimate that up to 1010 colour combinations are possible combining all structural colours (Fig. 1b–e), highlighting the feasibility of ARTEMIS for transcriptome profiling. We validated the fabrication of IDs with biotinylated ‘imaging strand’ using polyacrylamide gel electrophoresis (PAGE) with and without the addition of neutravidin (Supplementary Fig. 4). Additionally, we verified the correct assembly of ten colours with fluorescence quenching using fluorescein (6-FAM)-labelled structural units (Supplementary Fig. 5a). IDs were fabricated in equimolar concentration, each containing only one colour from ‘1’ to ‘10’ (Supplementary Figs. 5 and 6a). The fluorescence output of each separate colour indicates accurate fabrication of structural colours (Supplementary Figs. 5b and 6). We show that assigning the accuracy of a detected colour to a designed colour is >85% accurate and that for colours ‘1’, ‘2’, ‘3’ and ‘4’ is 97% (Fig. 1e). Our ability to design IDs with specific colours at the predicted positions enables the low error rate (Supplementary Fig. 7).

Multiplexed RNA identification in one pot reaction. We next used ARTEMIS to identify various transcripts in a one pot reaction as

---

**Fig. 2 | One pot reaction for multiplexed amplification-free identification of 18S rRNA, 28S rRNA and MS2 RNA.**

- **a**, Simultaneous ID fabrication for multiple RNA targets in a complex mixture of human total RNA. We assembled IDs by mixing universal human reference RNA with MS2 RNA and short complementary oligonucleotides targeting 18S rRNA, 28S rRNA and MS2 RNA. Six DNA nanostructures (double-hairpins; red) form a single structural colour (‘1’).
- **b**–**d**, Typical 18S rRNA ID ‘1111’ (crimson), 28S rRNA ID ‘11111’ (green) and MS2 RNA ID control (grey) as measured from a single nanopore are presented in **b**, **c** and **d**, respectively.
Fig. 3 | ARTEMIS discriminates engineered, alternative splicing isoforms resulting from any physical transcript arrangement. **a**, Isoform-specific labelling is achieved by labelling each synthetic exon (I, II, or III) with an asymmetric sequence of structural colours that results in unique IDs. **b**, Example events of the ordered RNA isoforms that differ in the order and combination of structural elements (i.e. synthetic exons) as illustrated in **a**.RNA isoforms can differ in length, and so successful discrimination of the two length isoforms is shown. **d**, Detection of circular isoforms is shown by creating IDs on DNA scaffolds. The nanopore microscope can distinguish between circular and linear configurations. **c**, The nanopore microscope discriminates the linear and circular populations based on the translocation time ($\Delta t$), which is about two times larger for the linear than for the circular ID. Sample size is 168 nanopore events.

schematized in Fig. 2a. As a proof-of-concept, we created distinctive IDs in a complex nucleic acid mixture (Fig. 2a and Supplementary Fig. 8) for human 18S, 28S ribosomal RNA (rRNA; oligonucleotides used for fabrication of 18S rRNA ID and 28S rRNA ID are listed in Supplementary Tables 5 and 6, respectively) and an external RNA ID control from MS2 bacteriophage with a known concentration (oligonucleotides are listed in Supplementary Tables 7 and 8 with the design details in Supplementary Fig. 9a). We successfully identified 18S and 28S rRNA in human total universal RNA (composition is listed in Supplementary Table 9) and human total cervical adenocarcinoma RNA.

Each RNA ID was identified with the nanopore microscope; events for 18S rRNA ID with four sites (‘1111’), 28S rRNA ID with five sites (‘11111’) and an external RNA ID control with three sites (‘111’) are depicted in Fig. 2b–d, respectively (additional events Supplementary Fig. 10). Expected velocity fluctuations during translocation play a minor role in measuring distances between sites as we achieve the correct read-out and position sequencing of sites along the target RNA (Supplementary Fig. 9c).

Two main obstacles for general RNA analysis are (1) degradation by nucleases assisted by magnesium ions and (2) structured regions that terminate the amplification or block hybridization. The former issue we addressed by replacing divalent ions with various monovalent ions. The removal of magnesium provides the added benefit of reducing the RNA structure stabilization and fragmentation for RNA ID fabrication (Supplementary Fig. 11). We also determined the optimum salt concentration for RNA ID fabrication in our experimental conditions (Supplementary Fig. 12). The stability and purity of the fabricated RNA IDs over time were assessed using nanopores and agarose gel electrophoresis (Supplementary Figs. 13 and 14). We find that RNA IDs show no to minimal degradation with standard storage conditions, in agreement with the previous observations for RNA–DNA hybrids.

ARTEMIS discriminates RNA isoforms by their order, directionality, length and circularity. Based on the multiplexed detection, ARTEMIS discriminated transcript variants, which are a result of alternative transcript processing and structural arrangements in a premature transcript (pre-messenger RNA (pre-mRNA); Fig. 3). As a proof-of-principle, we designed exons and their respective isoforms (Fig. 3). We designed a sequence of three structural colours per exon (Fig. 3a), thus creating asymmetric and isoform-specific RNA IDs (designs with example events are presented in Supplementary Fig. 15, and oligonucleotides used are listed in Supplementary Tables 10 and 11). ARTEMIS identified order, length and circular isoforms (Fig. 3b–d, respectively). The combination of exons results in multiple transcript isoforms with the same length but different sequences, that is, IDs (Supplementary Fig. 16). In Fig. 3b, we show
three correctly identified isoforms with the same length but a different order of exons that contain either exons I and II (‘211312’), exons I and III (‘123112’) or exons II and III (‘312123’) as shown in Fig. 3b. Even more, we can clearly determine the directionality of matching exons in isoforms by our asymmetric design of the three structural colours. Besides this, we demonstrate the identification of length isoforms for exon I (Fig. 3c and Supplementary Fig. 17).

Another critical feature that is hardly achievable with RNA-seq is the discrimination of circular and linear isoforms (Fig. 3d,e). As a proof-of-concept, we used M13 phage DNA to create circular and linear IDs with the sequence of colours ‘111’ using the same oligonucleotide mixture (Fig. 3d; event examples shown in Supplementary Fig. 18, while oligonucleotides are in Supplementary Table 12). The scatter plot in Fig. 3e shows minimal overlap of two populations indicating the circular and linear IDs ‘111’. We fixed the position of colours in the circular ID conformation by using interlock oligonucleotides (Supplementary Fig. 18c,d). Circularity discrimination is also assessed by in vitro RNA circularization (Supplementary Fig. 19) of linear MS2 RNA ID ‘111’ using T4 RNA ligase I (ref. 25).

Programmable self-assembled RNA origami. Up to this point in our study, ARTEMIS has depended on the reshaping of RNA to a linear sequence of structural colours. Nevertheless, some transcripts contain strong RNA secondary structures that are challenging to remove with short oligo hybridization, or long transcripts greater than 10 kb (refs. 15,26). Especially the latter transcripts would require oligos that complement the whole RNA, which may be cost prohibitive. Instead, we decided to use these parts of the RNA as alternative structural colours present in the native target. We extended the ARTEMIS ID functionality by using short RNA motifs (Fig. 4a), long regions (Fig. 4b) or a combination of DNA and RNA structural colours (Fig. 4c).
The asymmetry in the RNA ID is directly obvious during RNA ID carcinoma total RNA (Fig. 5a,b). Finally, we use ARTEMIS for the targeted identification of enolase RNA isoform relative quantification in a complex transcriptome.

More specifically, we assembled RNA origami IDs by employing secondary structure formation in prespecified locations (Fig. 4a; oligos are listed in Supplementary Table 13). Three structural colours were assembled by the nanoscale folding of 114-nucleotide (nt)-long, 190-nt-long and 342-nt-long single-stranded RNA to form structural colours ‘1’, ‘U’ and ‘Y’, respectively (predicted two-dimensional and three-dimensional structures are shown in Fig. 4d; more details in Supplementary Fig. 20). As above, each self-assembled RNA ‘origami’ has a specific current signature that can be identified from nanopore events, as shown in Fig. 4a (additional events are presented in Supplementary Fig. 21). The accuracy of each structural colour identification is over 99%, as displayed in the summary in Fig. 4e.

Interestingly, RNA IDs can even be realized when only the middle part of a long RNA is linearized, as shown in Fig. 4b (oligonucleotides are listed in Supplementary Table 14). The two terminal RNA structures are 1,230 nt and 401 nt in length (‘A’ and ‘Q’, respectively). The asymmetry in the RNA ID is directly obvious during RNA ID translocation through a nanopore. The two terminal downward signals directly correspond to the terminal RNA colours ‘A’ and ‘Q’ (additional events are presented in Supplementary Fig. 22) with an accuracy in all the unfolded identification of ~100% (Fig. 4f). Finally, we designed a combination of ID ‘111’ with a terminal RNA origami on both ends, as shown in Fig. 4c. The RNA ID traces in Fig. 4c show that DNA and RNA structural colours can be combined for RNA IDs. Hence, ARTEMIS can deal with secondary structures, and RNA length and accurate read-out is possible for any combination of structural colours. It is important to note that the addition of colours expands the number of unique RNA IDs.

RNA isoform relative quantification in a complex transcriptome.

Finally, we use ARTEMIS for the targeted identification of enolase (ENO) isoforms in commercially available human cervix adenocarcinoma total RNA (Fig. 5a,b). The ENO gene is known to have multiple transcript isoforms that differ in length or sequence as a result of alternative splicing of pre-mRNA. We employed three colours to identify four transcript isoforms (Fig. 5a; oligonucleotides are listed in Supplementary Table 15). RNA isoform ID designs and example events are illustrated in Fig. 5a. We determined the expression level of each ENO1 transcript isoform (Fig. 5b). An internal reference ID can further improve the transcript isoform-level quantification (more details in Supplementary Fig. 8), and so we chose 18S rRNA as an intersample reference. We confirmed that the nanopore event frequency is independent of the level of complementarity between the target RNA and oligonucleotides (Supplementary Fig. 23). The absolute concentration may be calculated from the nanopore frequency of RNA ID events using a previously introduced model.

As further demonstration, we used ARTEMIS to target X-inactive specific transcript long non-coding RNA (Xist lncRNA). Here we used terminal RNA, internal RNA motifs and DNA structural colours to identify length isoforms in the native transcriptome (Fig. 5c–e). We targeted part of Xist RNA to fabricate ID ‘111111’ (the design of Xist lncRNA ID is schematized in Supplementary Fig. 24, and oligonucleotides used for its fabrication are enumerated in Supplementary Table 16). The part of the sequence that differs among long (L-isoform) and short (S-isoform) isoforms is left unpaired (Fig. 5c). The expected ID read-out should depict the six sites with a structural colour ‘1’, the terminal unpaired RNA coil ‘A’ and an internal self-assembled RNA origami colour ‘L’ as predicted from the sequence (Fig. 5c–e). We show typical examples of Xist lncRNA isoform IDs that match the predicted design and previously identified Xist lncRNA isoforms (Supplementary Fig. 24).

In this study, we introduced ARTEMIS, an approach that reshapes an RNA target into a sequence of structural colours that we call an ID, using the subnanometre precision of DNA nanotechnology. ARTEMIS omits amplification and enzyme-based steps and identifies multiple native RNA transcripts and alternative events.
splicing variants in parallel using the nanopore microscope. As an electric measurement device, a nanopore microscope has a spatial resolution comparable to that of complex optical microscopes with higher throughput and straightforward origami assembly. Most diseases are classified by a change of a few transcripts. Thus, accurate identification of RNAs of interest has to bypass prior amplification and reverse transcription biases. Our approach has the potential to identify extensive RNA diversity from a gene of interest without the need to align to reference transcriptomes. It is known that reference transcriptomes neglect unrefereed RNA diversity, and hence the extensive RNA variation is lost. ARTEMIS is complementary to RNA-sequencing-based approaches as RNA IDs rely on RNA or genomic sequence information and do not provide de novo sequence information. We demonstrated multiple isoform identification by using the same oligo mix to identify order isoforms and length isoforms, as well as transcript circularity. ARTEMIS features are promising for characterizing therapeutic RNA uniformity and circularity using minimal sample amounts. Amplification-based RNA sequencing technologies enable identification of RNA transcripts at the single-cell level. Further developments are required to enable single-cell RNA characterization with ARTEMIS. Nanopore read-out of RNA at the single-cell level would require the integration of ARTEMIS and RNA handling with droplet-based techniques.

We employed RNA origami self-assembly as an additional way to identify transcripts of interest or to overcome some challenges of RNA analysis. Transcript IDs are assembled by using stable RNA structures as structural colours that can be either within the molecule or at its ends. We believe that ARTEMIS opens avenues for single-molecule mapping of RNA motifs. In addition, our study demonstrates that highly abundant natural RNAs may serve as scaffolds for DNA origami assembly with a wider length range and yield.

Our multicolour palette paves the way towards targeted isoform profiling in the whole transcriptome that excludes enzymatic and amplification biases. ARTEMIS has the potential to create ~10^10 unique RNA IDs that are readable using nanopore microscopy or imaging methods that rely on super-resolution microscopy, including DNA-PAINT.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, data extended, supplementary information, author acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41557-022-01037-5.

Received: 24 November 2021; Accepted: 4 August 2022; Published online: 19 September 2022

References
1. Ozsolak, F. & Milos, P. M. RNA sequencing: advances, challenges and opportunities. Nat. Rev. Genet. 12, 87–98 (2011).
2. Jeck, W. R. & Sharpless, N. E. Detecting and characterizing circular RNAs. Nat. Biotechnol. 32, 453–461 (2014).
3. Harvey, S. E. & Cheng, C. Methods for characterization of alternative RNA splicing. Methods Mol. Biol. 1402, 229–241 (2016).
4. Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR. Genome Res. 6, 986–994 (1996).
5. Stark, R., Grzech, M. & Hadfield, J. RNA sequencing: the teenage years. Nat. Rev. Genet. 20, 631–656 (2019).
6. Gootenberg, J. S. et al. Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a and Casm6. Science 360, 439–448 (2018).
7. Mamanova, L. et al. FRT-seq: amplification-free, strand-specific transcriptome sequencing. Nat. Methods 7, 130–132 (2010).
8. McGettigan, P. A. Transcriptomics in the RNA-seq era. Curr. Opin. Chem. Biol. 17, 4–11 (2013).
9. Meyer, C. A. & Liu, X. S. Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat. Rev. Genet. 15, 709–721 (2014).
10. Garalde, D. R. et al. Highly parallel direct RNA sequencing on an array of nanopores. Nat. Methods 15, 201–206 (2018).
11. Conn, V. & Conn, S. J. SplintQuant: a method for accurately quantifying circular RNA transcript abundance without reverse transcription bias. RNA 25, 1202–1210 (2019).
12. Workman, R. E. et al. Nanopore native RNA sequencing of a human poly(A) transcriptome. Nat. Methods 16, 1297–1305 (2019).
13. Guo, Y., Li, C. I., Ye, F. & Shyr, Y. Evaluation of read count based RNAseq analysis methods. BMC Genomics 14, S2 (2013).
14. Soneson, C. et al. A comprehensive examination of nanopore native RNA sequencing for characterization of complex transcriptomes. Nat. Commun. 10, 3359 (2019).
15. Li, J., Jiang, H. & Wong, W. H. Modeling non-uniformity in short-read rates in RNA-Seq data. Genome Biol. 11, R50 (2010).
16. Devinson, J. W. et al. Universal alternative splicing of noncoding exons. Cell 65, 245–255.e5 (2018).
17. Lagarde, J. et al. High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. Nat. Genet. 49, 1731–1740 (2017).
18. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. Nature 440, 297–302 (2006).
19. Howarth, M. et al. A monovalent streptavidin with a single fentomol biotin binding site. Nat. Methods 3, 267–273 (2006).
20. Li, J., Gershshow, M., Stein, D., Brandin, E. & Golovchenko, J. A. DNA molecules and configurations in a solid-state nanopore microscope. Nat. Mater. 2, 611–615 (2003).
21. Deamer, D., Akeson, M. & Branton, D. Three decades of nanopore sequencing. Nat. Biotechnol. 34, 518–524 (2016).
22. Dekker, C. Solid-state nanopores. Nat. Nanotechnol. 2, 209–215 (2007).
23. Plesa, C., Van Loo, N., Ketterer, P., Dietz, H. & Dekker, C. Velocity of DNA during translocation through a solid-state nanopore. Nano Lett. 15, 732–737 (2015).
24. Roberts, R. W. & Crothers, D. M. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. Science 258, 1463–1466 (1992).
25. Müller, S. & Appel, B. In vitro circularization of RNA. RNA Biology 14, 1018–1027 (2017).
26. Amarasinghe, S. L. et al. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 21, 30 (2020).
27. Popenda, M. et al. Automated 3D structure composition for large RNAs. Nucleic Acids Res. 40, e112 (2012).
28. Uhlén, M. et al. A pathology atlas of the human cancer transcriptome. Science 357, eaau2507 (2017).
29. Howe, K. L. et al. Ensembl 2021. Nucleic Acids Res. 49, D884–D981 (2021).
30. Kuchipudi, S. V. et al. 18S rRNA is a reliable normalisation gene for real time PCR based on influenza virus infected cells. Virol. J. 9, 230 (2012).
31. Bell, N. A. W., Muthukumar, M. & Keyser, U. F. Translocation frequency of double-stranded DNA through a solid-state nanopore. Phys. Rev. E 93, 022401 (2016).
32. Eyras, E., Caccamo, M., Curwen, V. & Clamp, M. ESTGenes: alternative splicing from ESTs in Ensembl. Genome Res. 14, 976–987 (2004).
33. Steinhauer, C., Jungmann, R., Sobey, T. L., Simmel, F. C. & Tinnefeld, P. DNA origami as a nanosopic ruler for superresolution microscopy. Angew. Chem. Int. Ed. 48, 8870–8873 (2009).
34. Huang, G., Voet, A. & Maglia, G. Frac nanopores with adjustable diameter identify the mass of opposite-charge peptides with 44 dalton resolution. Nat. Commun. 10, 835 (2019).
35. Morrill, A. & Gautheret, D. Bridging the gap between reference and real transcriptomes. Genome Biol. 20, 112 (2019).
36. Wesselhoef, R. A., Kowalski, P. S. & Anderson, D. G. Engineering circular RNA for potent and stable translation in eukaryotic cells. Nat. Commun. 9, 2629 (2018).
37. Zhang, X. et al. Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-seq systems. Mol. Cell 73, 130–142.e5 (2019).
38. Han, D. et al. Single-stranded DNA and RNA origami. Science 358, eaao2648 (2017).
39. Liu, D. et al. Branched kissing loops for the construction of diverse RNA homooligomeric nanostructures. Nat. Chem. 12, 249–259 (2020).
40. Geary, C., Grossi, G., McRae, E. K. S., Rothemund, P. W. K. & Anderson, E. S. RNA origami design tools enable cotranscriptional folding of kilobase-sized nanoscaffolds. Nat. Chem. 13, 549–558 (2021).

Publisher’s note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.
© The Author(s), under exclusive licence to Springer Nature Limited 2022.
Methods

RNA identifier fabrication. To fabricate an RNA ID, we used oligonucleotide mixes that contain docking strands and an imaging strand. We prepared a 40 μl reaction by mixing RNA sample (to 20 nM or 800 fmol) and oligonucleotides (to 60 nM each or 2,400 fmol) in 100 mM LiCl (or 10 mM MgCl₂), 1× TE buffer (10 mM Tris-HCl buffer, 1 mM ethylenediaminetetraacetic acid, pH 8.0) and nuclease-free water was added to the final reaction volume. Buffers were filtered with the MF-Millipore membrane filter with 0.22 μm pore size. The reaction was mixed by pipetting and spun down. The mixture was heated to 70 °C for 30 s and then gradually cooled (−0.5 °C per cycle, 90 cycles of 30 s each) over 45 minutes to room temperature and held at 4 °C. RNA IDs were filtered using 0.5 ml 100 kDa cut-off Amicon filter units. The washing buffer used for filtration was composed of filtered 10 mM Tris-HCl (pH 8.0) with 0.5 mM MgCl₂.

Nanopore fabrication. We fabricated 10–15 nm nanopores using a laser-assisted capillary puller (P2000F, Sutter Instruments). Glass capillaries with an outer diameter of 0.5 mm and an inner diameter of 0.2 mm with filament were purchased from Sutter Instruments. The heat protocol had the following values: HEAT, 480; VEL, 25; DEL, 170; and PUL, 200.

Nanopore microscopy. Nanopore measurements were performed in 4 M LiCl, 1× TE, pH 9.4 using Axopatch 200B, and data were collected under a constant voltage of 600 mV. Single events in ionic current recordings were first isolated according to threshold parameters such as duration, current drop and event charge deficit using home-built LabVIEW codes. Events were filtered from the background using the mean event current of the ID (RNA/DNA hybrid level) that discriminates events with the duplex level from random RNA blobs. In addition, by setting the range of the event duration and event charge deficit (that is, the surface area of the event), we excluded all events that did not fit to an RNA ID. We selected only unfolded (that is, linear) RNA ID events for further analysis. Then, we identified the number of downward spikes or peaks in an event, which should match the potential expected number of sites. Finally, by relative comparison of the ionic current blockage of the identified spikes, structural colours were assigned at the level of single molecules.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data supporting the findings of this study are available in the main text and the Supplementary Information. Additional raw data are available at https://doi.org/10.17863/CAM.87123. Source data are provided with this paper.

Acknowledgements

We thank J. Zhu and M. Fletcher for the critical reading of the manuscript and useful suggestions. We thank the Howarth Lab from the University of Oxford for the monovalent streptavidin. U.F.K. acknowledges funding from a European Research Council Consolidator grant (DesignerPores no. 647144) and European Research Council Proof-of-Concept grant (PoreDetect no. 899538). F.B. acknowledges funding from George and Lilian Schiff Foundation Studentship, the Winton Programme for the Physics of Sustainability Ph.D. Scholarship and St John's College Benefactors’ Scholarship.

Author contributions

F.B. conceived the idea. F.B. and U.F.K. designed the study. F.B. performed the experiments and analysed the data. F.B. and U.F.K. wrote the manuscript.

Competing interests

F.B. and U.F.K. are inventors for the ARTEMIS method (United Kingdom patent application no. 2113935.7, in process) submitted by Cambridge Enterprise on the behalf of the University of Cambridge. U.F.K. is a cofounder of Cambridge Nucleomics.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41557-022-01037-5.

Correspondence and requests for materials should be addressed to Ulrich Felix Keyser.

Peer review information Nature Chemistry thanks Sergii Pud, Adam Hall and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a  | Confirmed

☐  ☒ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

☐  ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☐  ☒ The statistical test(s) used AND whether they are one- or two-sided

Only common tests should be described solely by name; describe more complex techniques in the Methods section.

☐  ☒ A description of all covariates tested

☐  ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☐  ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

☐  ☒ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

Give P values as exact values whenever suitable.

☐  ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

☐  ☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

☐  ☒ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection  The data was collected using software written with National Instruments LabVIEW 2013.

Data analysis  Data analysis was performed using software written with National Instruments LabVIEW 2013.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw ionic current translocation data are shown in the manuscript or the supporting information.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size  Each nanopore device was run continuously to collect as many as possible RNA ID translocations. Typically this is due to clogging of the nanopore or in some cases due to the long measurements (>10 hours) air passing through a PDMS chip. Therefore there is no set sample size. However we only used devices where we measured a minimum of 100 unfolded RNA IDs or for more than 10 h for native human total RNA samples.

Data exclusions  In our measurements we only used unfolded RNA ID translocations for further analysis. In this way, we can more precisely assign IDs.

Replication  The measurements shown in the manuscript were obtained using at least three different nanopores to exclude potential pore-to-pore variability.

Randomization  Randomization was not relevant to this study since we were not comparing experimental groups. Here, we used commercially available total RNA samples from either human total universal RNA or cervical adenocarcinoma total RNA (Invitrogen).

Blinding  Blinding was not relevant to this study since we were not comparing experimental groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|----------------------------------|---------|
| n/a involved in the study        | n/a involved in the study |
| ☒ Antibodies                     | ☒ ChIP-seq |
| ☒ Eukaryotic cell lines          | ☒ Flow cytometry |
| ☒ Palaeontology and archaeology  | ☒ MRI-based neuroimaging |
| ☒ Animals and other organisms    |         |
| ☒ Human research participants    |         |
| ☒ Clinical data                  |         |
| ☒ Dual use research of concern   |         |