

# Non-complementary strand commutation as a fundamental alternative for information processing by DNA and gene regulation

Received: 8 July 2021

Accepted: 16 November 2022

Published online: 05 January 2023

 Check for updates

Maxim P. Nikitin <sup>1,2</sup> 

The discovery of the DNA double helix has revolutionized our understanding of data processing in living systems, with the complementarity of the two DNA strands providing a reliable mechanism for the storage of hereditary information. Here I reveal the ‘strand commutation’ phenomenon—a fundamentally different mechanism of information storage and processing by DNA/RNA based on the reversible low-affinity interactions of essentially non-complementary nucleic acids. I demonstrate this mechanism by constructing a memory circuit, a 5-min square-root circuit for 4-bit inputs comprising only nine processing ssDNAs, simulating a 572-input AND gate (surpassing the bitness of current electronic computers), and elementary algebra systems with continuously changing variables. Most importantly, I show potential pathways of gene regulation with strands of maximum non-complementarity to the gene sequence that may be key to the reduction of off-target therapeutic effects. This Article uncovers the information-processing power of the low-affinity interactions that may underlie major processes in an organism—from short-term memory to cancer, ageing and evolution.

The structure of DNA<sup>1</sup>, discovered in 1953, implied that every strand of DNA is meant to have the ‘perfect’ companion—the complementary strand. This unique feature instantaneously suggested nature’s design of data storage within DNA. Indeed, it is straightforward to encode and then decipher the information that is laid out as linear sequences of A/T/G/C-monomers on two intertwined strands of DNA, where A on one strand binds with T on the other strand, and similarly G complements C. The selectivity of interactions between the respective monomers leads to a remarkably high mutual specificity of the two strands, while multi-point binding ensures extraordinary affinity and tight association, even at single-molecule concentrations. Non-covalent (but essentially non-reversible) binding of the two precise copies of the encoded information guarantees efficient long-term storage, readout and duplication of the data in both living and synthetic systems<sup>2</sup>.

Although this perception of DNA has formed the fundamental basis for biology research, many riddles about DNA—such as the purpose of the noncoding parts of the genome—remain unsolved.

Although we are constantly discovering new epigenetics mechanisms that affect genetic data readout (such as on/off switching of transcription and translation), the question arises of whether any other data are held by DNA in a non-traditional form—that is, unreadable within the current paradigm of the double helix, complementarity and the genetic code? If yes, then what would be the underlying mechanisms?

Because no such form has been identified *in vivo* over the past 65 years, perhaps, first, it needs to be recreated artificially *in vitro*. The young field of biocomputing<sup>3</sup> is devoted to synthetic systems that can mathematically process data using biomolecules (for example, for advanced therapeutics and drug delivery<sup>4,5</sup>). With relatively few extracellular systems based on small molecules<sup>6,7</sup> and/or proteins<sup>8,9</sup>, numerous DNA-based approaches<sup>10–13</sup> have been proposed, following the pioneering work of Adleman<sup>14</sup>.

Despite the variety of published DNA-based systems, all of them rely on the fundamental ability of DNA to form a double-stranded complex of two complementary strands (for example, for strand

<sup>1</sup>Sirius University of Science and Technology, Sochi, Russia. <sup>2</sup>Abisense LLC, Dolgoprudny, Moscow Region, Russia.  e-mail: [max.nikitin@gmail.com](mailto:max.nikitin@gmail.com)

displacement<sup>15</sup> or to ensure the specificity of DNAzymes<sup>16,17</sup>) and hence lie within the traditional double-helix paradigm.

In this Article I demonstrate a 'strand commutation' mechanism for data processing with DNA that lies outside that paradigm, that is, a mechanism based on essential non-complementarity between the interacting DNA strands.

## Results

I noted one remarkable feature of DNA, the beauty of which has been long overshadowed by the elegance of the double helix. Let us look at a DNA strand  $X$ . Generally, its complementary strand,  $\bar{X}$ , will have the highest affinity among all other strands. However, if we generate variations of  $\bar{X}$  by sequential substitution of nucleotides for random ones, then we will get a set of oligonucleotides with a variety of irregular complementarity patterns (Fig. 1a) and a full range of affinities for  $X$ —from extremely affine (like the original  $\bar{X}$ ) to the absolutely non-binding ones. Here I will refer to such strands as 'undercomplementary' to make a distinction from other patterns of partial complementarity with different biological meaning, such as those with alternating domains of complete complementarity and non-complementarity as used in strand displacement, single-nucleotide polymorphisms and so on.

To provide an understanding of the obtained 'affinity continuum', we can use state-of-the-art algorithms that relatively accurately predict the affinity between single-stranded DNA molecules (ssDNA) based on their sequence. Throughout this manuscript, NUPACK<sup>18,19</sup> is used for this.

For example, we can compute the affinity for a short 10-base random ssDNA  $X = \text{GCAGTATTCG}$  with all  $4^{10}$  ( $\sim 10^6$ ) possible 10-nt DNA sequences. Figure 1a shows that, even for such short strands, the computed dissociation constants are packed so tightly that an undercomplementary sequence can be chosen that has affinity within 10% of any given dissociation constant (excluding  $\sim 25$  min/max values). Figure 1b shows that a similar effect can be observed if we look at the affinity of all 10-base strands simultaneously towards two oligos:  $X$  and the complementary  $\bar{X}$ , or  $X$  and another random  $Y$  (for example, TAGCGCAGTA, which has little affinity to  $X$ ). Aside from affinity extrema, a close-packed multidimensional affinity continuum is formed such that an undercomplementary oligo can be selected with on-demand affinities towards multiple preselected oligos.

Now let us look at what happens in a random mix of short ssDNA molecules when we deliberately remove all complementary strands and leave only weakly interacting undercomplementary ones. Figure 1 illustrates the differences between such a system (Fig. 1d) and those based on the traditional complementarity idea (Fig. 1c). In the absence of their 'perfect' companions, each ssDNA will reversibly associate with other strands, constantly interchanging partners. Such complexes will coexist simultaneously, governed by the law of mass action. Mathematically speaking, in equilibrium, the system will comply with a set of nonlinear equations, as shown in Fig. 1d.

This strand commutation process (together with the affinity continuum feature of nucleic acids) offers a fundamental alternative to the double helix for the storage and processing of data within DNA/RNA.

### Basics of the phenomenon and illustration with Boolean logic

Although strand commutation provides a great variety of ways to process different sorts of data, its power will be first demonstrated here for the construction of systems that implement Boolean logic. I begin with this illustration to simplify the explanation as well as to allow comparison with other state-of-the-art biocomputing systems, most of which operate in terms of logic gates.

Before specific logic gates and circuits are described, general rules need to be postulated for consistency and for the correct assessment of the computational results. As noted, each system represents a set of weakly interacting oligonucleotides. Although other options

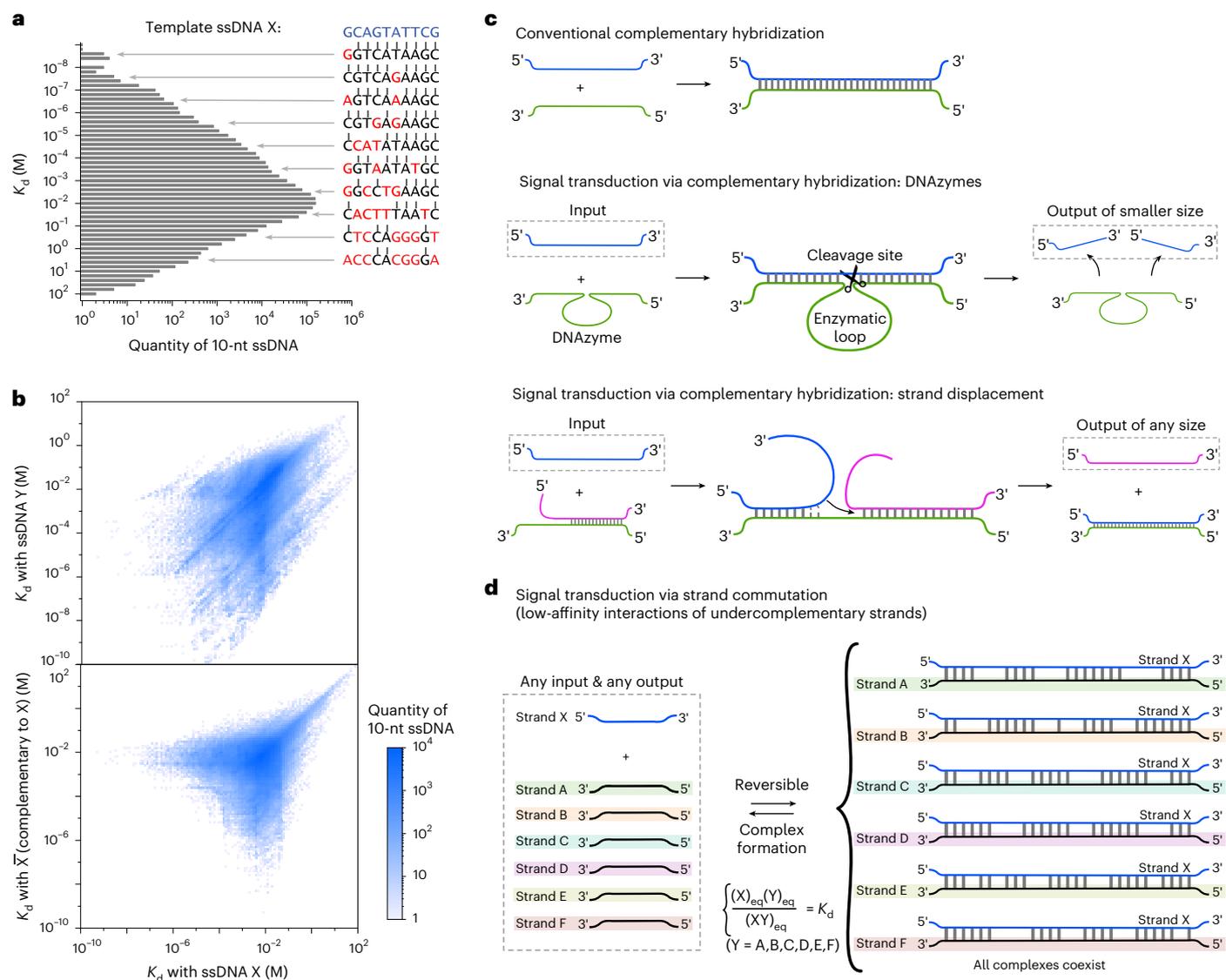
are available (Discussion), here the information will be encoded into concentration levels of the complexes that are formed within an oligo-nucleotide set. For simplicity, the system's state (the result of computation) will be assessed by looking at the complexes formed by a single 'signal' oligo  $S$ . To report the state of its complexes, the signal oligo will be labelled by a fluorophore, and its counterparts (oligos  $Q_n$ ,  $n = 1, 2, \dots$ ) will be labelled with a quencher. The detectable level of fluorescence will indicate the result of the computation: 'logic truth (1)' in the case of unquenched fluorescence of the free  $S$  and 'logic false (0)' in the case of quenched fluorescence of the formed complexes. Addition of input oligos ( $I_m$ ,  $m = 1, 2, \dots$ ) will initiate the computation, with the presence of the input oligo considered 'logic truth (1)' and absence indicates 'logic false (0)'. The threshold between true and false signals will be the average between the minimum and maximum output signals observed among all input combinations.

Let us start with the most basic single-input YES gate, the output value of which equals the input value. Such a gate can be built with only three oligos, as shown in Fig. 2a: input  $I$ , signal  $S$  and one intermediary strand—quencher  $Q$ . If  $Q$  has certain weak affinities towards both  $S$  and  $I$ , but the affinity between  $I$  and  $S$  is much weaker (ideally negligible), then the system will function as follows. When no  $I$  is present, some of  $S$  is bound with  $Q$ , and fluorescence is quenched to a certain low level. Once  $I$  is added to the system, the balance of the  $Q$  complexes will shift;  $I$  will bind some of the  $Q$ , effectively pulling  $Q$  out of the complex with  $S$  (partially), thereby unquenching the fluorescence. Hence, when  $I$  is present (input = 1), the detectible signal increases (output = 1). To achieve a higher dynamic range of the true/false signals, we can adjust the affinity and concentrations (A&C) in such a way that  $Q$  prefers to bind with  $I$  over  $S$  (dissociation constants  $K_d[IQ] < K_d[QS]$ ) and/or concentrations  $[I] > [Q] > [S]$ . In the following, such tuning of the system will be referred to as 'A&C adjustment'.

For the experimental implementation (let us begin with RNA), three ssRNAs were designed (the sequences for all used gates are shown in Supplementary Note 1, as well as concentration and mutual affinity data) with the following NUPACK-predicted affinities:  $K_d[QS] = 4 \times 10^{-9}$  M,  $K_d[IQ] = 2 \times 10^{-9}$  M and  $K_d[IS] = 2 \times 10^{-2}$  M (hereafter,  $K_d$  is the approximate dissociation constant; details are provided in Supplementary Note 1).  $S$  was synthesized with the Cy5 label on its 5' end, and  $Q$  with the BHQ2 quencher on its 3' end. Note that the fluorophore and quencher are known to slightly change the affinity of the labelled strands. Also, having non-complementary bases near the quencher–fluorophore pair may lower the efficiency of fluorescence quenching (thereby causing discrepancies between prediction and experiment). However, this influence was neglected during design of the systems and the calculations of  $K_d$ . Figure 2a shows the correct YES-gate performance of the system using these three oligos with  $[S] = 1 \mu\text{M}$ ,  $[Q] = 1 \mu\text{M}$  and  $[I] = 10 \mu\text{M}$ . Supplementary Fig. 1 shows that correct gating can also be observed with different A&C adjustment with equal concentrations  $[S] = [Q] = [I] = 1 \mu\text{M}$ .

The inverted NOT gate (the output is opposite to the input) can be realized with strand commutation between four oligos:  $S$ ,  $Q$ , mediator  $M$  and  $I$  (Fig. 2b).  $I$  should interact only with  $M$ ;  $M$  with  $I$  and  $Q$ ;  $Q$  with  $I$  and  $S$ ; and  $S$  only with  $Q$ . Furthermore, the A&C adjustment should be such that  $M$  prefers to bind with  $I$  over  $Q$ , and  $Q$  favours binding with  $M$  over  $S$ . Accordingly, when no input is present,  $M$  pulls most of the  $Q$  from its bond with  $S$ , and the free  $S$  fluoresces intensely. Addition of  $I$  leads to it binding with  $M$ . This frees up  $Q$ , which now binds with  $S$  and quenches the signal. Hence, when the input is 1, the output is 0. Figure 2b shows the performance of the experimental implementation of the RNA-based NOT gate.

A similar idea can be used to construct more complex gates. For diversity, in the following examples, ssDNA ensembles will be used. A three-input OR gate and a three-input AND gate using 12-nt and 15-nt ssDNA, respectively, are shown in Fig. 2c,d (Supplementary Figs. 2 and 3 show the gate kinetics). In the OR gate, all three inputs ( $I_1, I_2, I_3$ ) act on



**Fig. 1 | Signal transduction—conventional complementarity paradigm versus non-complementary strand commutation.** **a, b**, Affinity continuum for ssDNA molecules: number of 10-nt ssDNAs for a given  $K_d$  with an arbitrary ssDNA X (per 0.2 orders of  $K_d$ ) (**a**) and two-dimensional distribution for the number of 10-nt ssDNAs for a given  $K_d$  towards arbitrary X and Y (**b**, top) or towards X and X (ssDNA complementary to X; **b**, bottom) per 0.1×0.1 orders of  $K_d$ . **c**, Examples of signal transduction between interacting oligonucleotide within the conventional

complementarity paradigm: overlapping regions of various strands have identical sequences. **d**, The idea behind signal transduction through strand commutation, that is, low-affinity interactions of essentially non-complementary strands: a single strand can simultaneously interact with a multitude of undercomplementary strands with different non-matching sequences, thereby passing the information to multiple recipients.

the same  $Q$ . In the AND gate, each input acts on its own  $Q_n$  ( $n = 1, 2, 3$ ), and all  $Q_n$  simultaneously act on  $S$ . In the OR gate, A&C adjustment should allow each  $I_n$  to take up most of the  $Q$  on its own; that is, each  $I_n$ - $Q$  complex should be much more favourable than  $Q$ - $S$ . In this case, once any of the inputs is present, then all  $Q$  becomes bound to  $I_n$ , and  $S$  is unquenched. In the AND gate, A&C adjustment should allow each  $Q_n$  (in the absence of the respective input) to quench most of the  $S$  on its own.

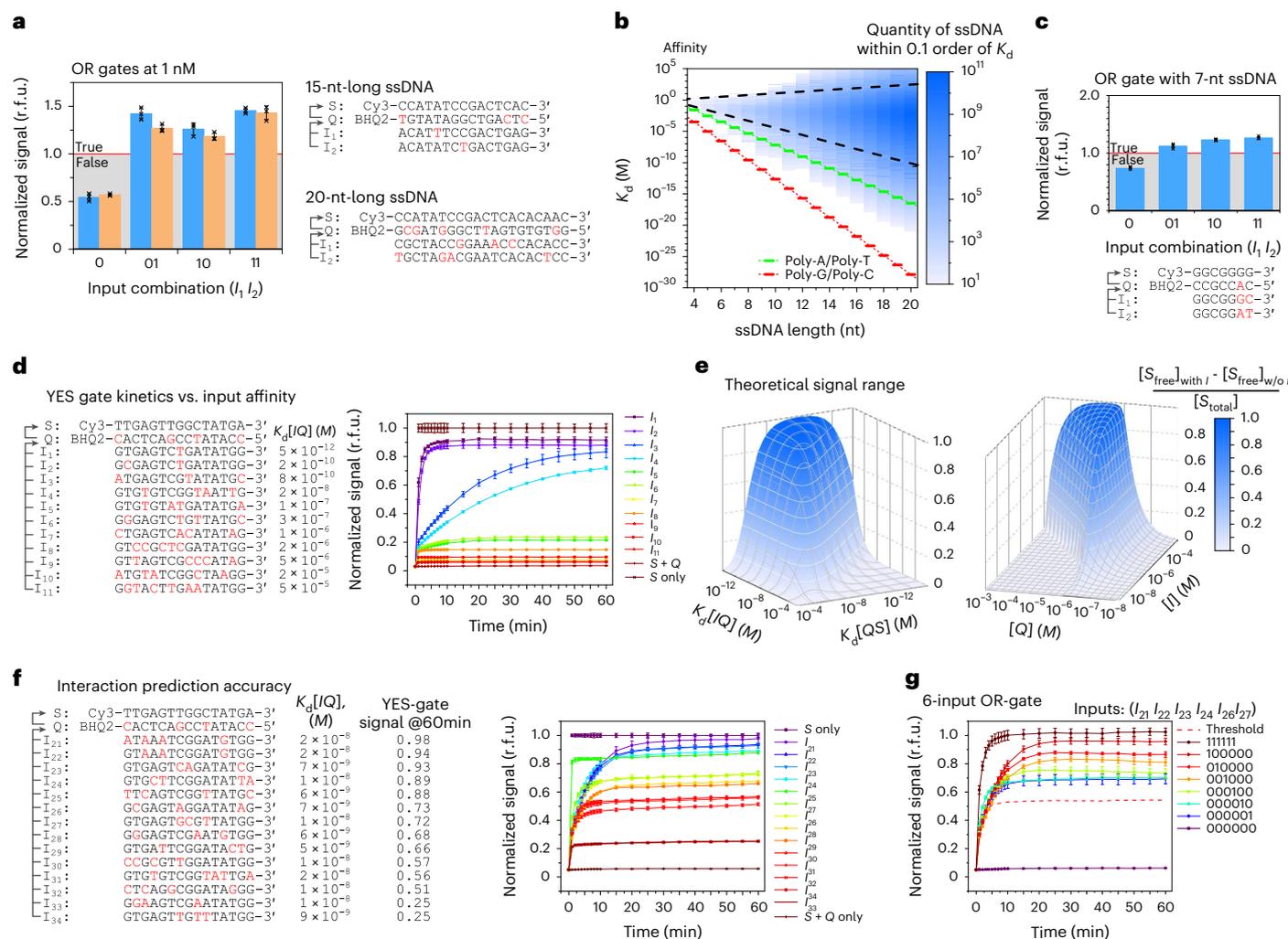
By combining such YES/NOT/AND/OR gates we can program the systems for more complex behaviour. For example, Fig. 2e shows the construction and performance of a NAND gate as the OR[NOT[Input]] combination (which is equivalent to the NOT[AND[Input]] combination). Before demonstrating the more complex logic circuits, let us look at the parameters that influence strand commutation at the level of the basic YES/OR gates. The parameters, discussed in more detail in the following, are strand concentration, strand size, kinetics, balance

between low affinity and specificity, dynamic range and NUPACK prediction accuracy.

First, the key issue is whether the phenomenon can be observed at the cell-level concentrations of small DNA/RNA molecules (for example, microRNA, miRNA), which are generally much lower than 1  $\mu\text{M}$ , but can be as high as tens and even hundreds of nanomoles/litre<sup>20,21</sup>. As dictated by the law of mass action, there is no concentration limit to the effect. OR gates with 15-base and 20-base ssDNA could be experimentally constructed at all tested concentrations down to 1 nM (where fluorescently labelled oligonucleotides can still be detected by standard plate readers; Fig. 3a and Supplementary Figs. 4 and 5). Therefore, concentration-wise, it is feasible that strand commutation plays an important role in living systems.

Second, what is the minimal size of the compatible strands? Figure 3b and Supplementary Fig. 6 show the relatively smooth





**Fig. 3 | Performance aspects of the basic YES/OR logic gates.** **a**, Two-input OR gates with nanomolar 15-nt (blue) or 20-nt (orange) ssDNA ( $[S] = [Q] = 1$  nM;  $[I_1] = [I_2] = 10$  nM) ( $n = 3$ ). **b**, Density of affinity continuum for the different lengths of ssDNA. The graph plots the NUPACK simulation of interactions of the randomly generated ssDNA pairs: simulated data are shown within the black dashed lines and extrapolated outside the region, as explained in detail in Supplementary Fig. 6. Red and green dashed lines demonstrate affinities of poly-G/poly-C or poly-A/poly-T duplexes, respectively. **c**, Two-input OR gate with 7-nt ssDNA ( $n = 3$ ). **d**, YES-gate kinetics for inputs of various affinities towards  $Q$  ( $n = 3$ ). **e**, Theoretical dynamic range of the YES-gate output, that is, free  $S$  concentration with input minus free  $S$  concentration without input relative

to the total  $S$  concentration: for the case of  $[S] = [Q] = [I] = 1$   $\mu$ M for various  $K_d$  [Q],  $K_d$  [QS] or for the case of  $K_d$  [Q] =  $K_d$  [QS] =  $10^{-6}$  M ( $[S] = 1$   $\mu$ M) for various concentrations. **f**, Correlation of the NUPACK prediction with experiment: kinetics of various inputs with similar affinity to  $Q$  but with different undercomplementarity patterns. **g**, Six-input OR gate ( $n = 3$ ). In **a** and **c**, the true/false threshold is the average of the maximum and minimum outputs. The outputs of each gate are normalized by the respective threshold signal. In **a**, **c**, **d** and **f**, nucleotides that are complementary to the corresponding sequence (shown with arrows) are written in black, and non-complementary ones are in red. Data are presented in relative fluorescence units (r.f.u.) as mean  $\pm$  s.d., and the  $n$  values indicate the number of independent samples.

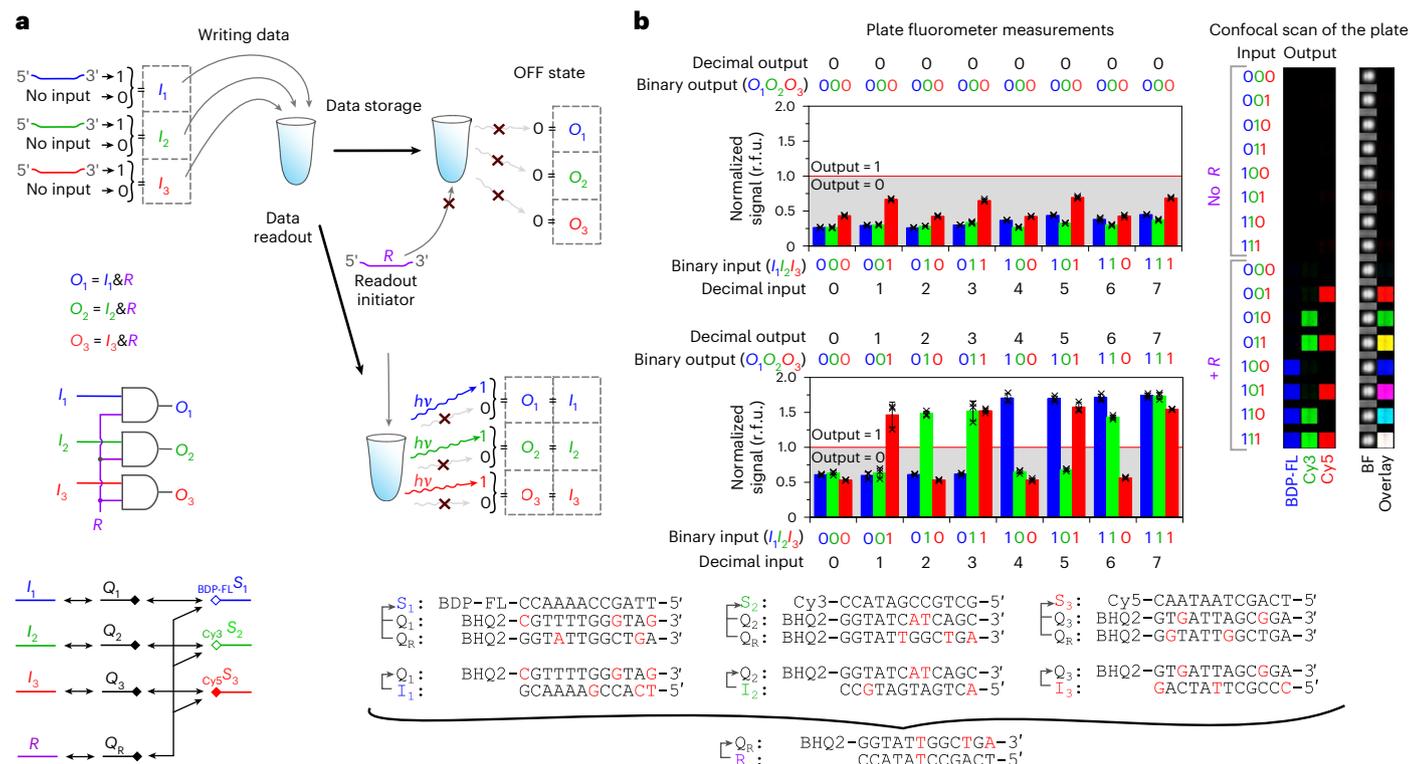
different (Fig. 3f). Although NUPACK greatly facilitated this study, the design of sophisticated circuits nevertheless required tedious experimental testing/selection of the oligos. For example, by evaluating the performance of these 14 inputs, I was able to select only six of those with the highest degree of unquenching that showed similar enough behaviour to construct a correctly functioning six-input OR gate (Fig. 3g). Evidently, NUPACK has been designed primarily to handle fully complementary strands. Substantially better accuracy can perhaps be achieved by creating a similar algorithm that is trained by using undercomplementary strands or by building an adjustment add-on for NUPACK (see the example by Bee et al.<sup>23</sup>).

### Biocomputing circuits

Now, let us look at how strand commutation deals with more complex data processing. The first example is an artificial 3-bit 'short-term'

memory circuit with write-and-read function. The system is fed with a set of three inputs that it needs to 'remember'. Next, to read out the stored information at an arbitrary time point, the fourth readout input is introduced into the system, and a fluorescent signal at three different wavelengths (each corresponding to a specific input) reports the values of the initial inputs.

Figure 4 shows the general scheme of the system, which has three signal oligos:  $S_1$  is labelled with BDP-FL,  $S_2$  with Cy3, and  $S_3$  with Cy5. These fluorophores can be easily distinguished from each other according to their fluorescence spectra. Each signal oligo has a corresponding quencher oligo ( $Q_1, Q_2, Q_3$ ) to process inputs  $I_1, I_2, I_3$  to be 'remembered', respectively. However, there is also one common quencher,  $Q_R$ , that can bind with each of the signal oligos and processes the readout input  $R$ . In other words, these are three different AND gates (with independent outputs) that share one common readout input.



**Fig. 4 | Memory circuit and its design and performance. a**, Experimental set-up, schematics and sequences of ssDNA used in the memory circuit, as an example of direct storage of information via strand commutation. The memory cell is first written with a 3-bit number encoded by three input oligos that are omitted (=0) or added (=1) to the test tube. Upon addition of the input, the circuit's output remains off (0). However, once the readout initiator oligo  $R$  is added to the test tube, the circuit changes its state in the three fluorescence channels (BDP-FL, Cy3, Cy5) so that the 3-bit output shows the stored input value. Nucleotides that are complementary to the corresponding sequence

(shown with arrows) are written in black, and non-complementary ones in red. **b**, Experimental performance of the circuit. Computation results are measured using a plate fluorometer, 5 min after addition of the inputs ( $n = 3$  independent samples). The true/false threshold is the average of the maximum and minimum outputs. The outputs of each gate are normalized by the respective threshold signal. Data are presented in relative fluorescence units (r.f.u.) as mean  $\pm$  s.d. The confocal scan of the samples was taken 1 h after addition of the inputs, and is representative of  $n = 3$  independent experiments (additional images are presented in Supplementary Fig. 9).

Until the readout input is added to the system, all outputs are 0. When the readout is initiated ( $R = 1$ ), output of each gate will be equal to the value of the remembered input ( $I_1$ ,  $I_2$  or  $I_3$ ). Figure 4 shows the experimentally measured signals for the as-designed system as well as the confocal scan of the microplate (for visual demonstration of the readout).

The next example demonstrates the strand commutation efficiency by direct comparison with other biomolecular computing approaches. So far, one of the toughest benchmark tests in the realm of biocomputing has been calculation of the square root of 4-bit numbers. In 2010, Qian and Winfree tackled the problem using a DNA strand-displacement circuit that featured a complex signal restoration mechanism<sup>15</sup>. The circuit required 130 oligos and 10 h for computation. In 2019, Song et al.<sup>24</sup> presented a circuit based on the strand-displacing DNA polymerase; this lacked signal restoration but could solve the problem using 37 oligos in 25 min.

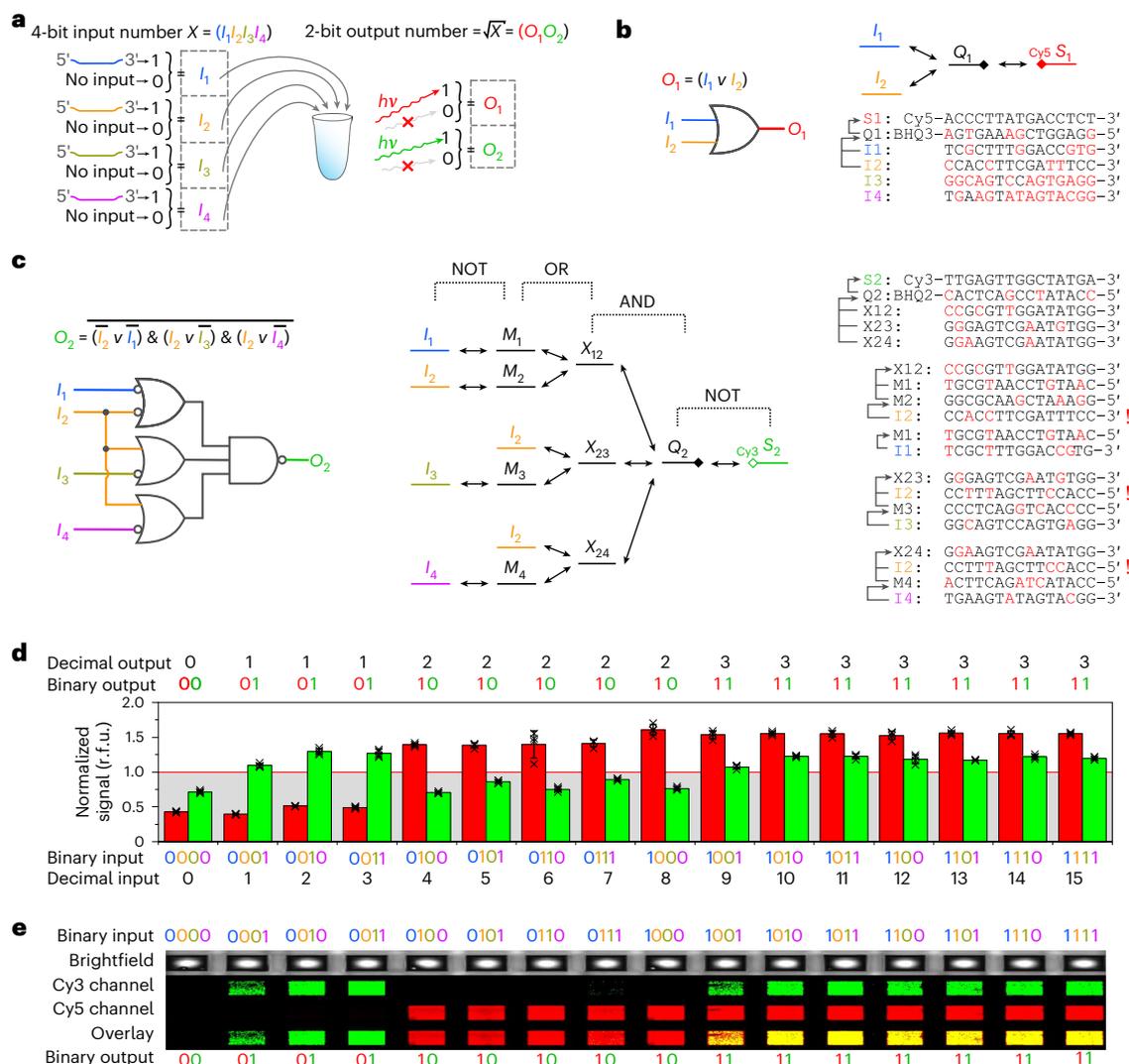
Here it is demonstrated that, with strand commutation, the same task can be realized with only nine short 15-nt oligonucleotides (+4 inputs) and less than 5 min of computation time. Figure 5 shows the scheme of the implemented logic functions and the design of the systems based on the combination of the described above YES/NOT/AND/OR gates. Computations by the high and low registers of the output ( $O_1$  and  $O_2$ , respectively) proceed simultaneously in a single test tube, with output readout at different wavelengths (two corresponding signal oligos  $S_1$  and  $S_2$  are labelled with Cy5 and Cy3, respectively). The  $O_1$  circuit is an OR gate (Fig. 5b), but the  $O_2$  circuit is much more complex (Fig. 5c).

It was possible to generate many sets of oligos that implement correct performance of the  $O_2$  circuit in simulation. However, as suggested by Fig. 3f, with increased complexity, NUPACK simulation of the strand interactions substantially deviated from the experimental behaviour. Therefore, the in silico designed oligo set needed extensive manual tuning, both in terms of sequences and concentrations. First, parasite cross-reactivity (that is, non-predicted) needed to be minimized. Second, the concentrations of the oligos had to be regulated so that for all strands at the same distance from the signal (see scheme in Fig. 5c), the variation between minimum and maximum signals (corresponding to the present or absent strand) became the same.

With such adjustments, the working circuit for calculating the square root was constructed (the sequences and concentrations are shown in Supplementary Note 1). Figure 5d,e shows the performance of the circuit for all possible 4-bit numbers 5 min post input addition (Supplementary Fig. 10 shows the temporal stability of the gate's output signal).

### Simulation of the highly complex information-processing systems

Next the potential scope of strand commutation will be demonstrated by means of in silico simulations of advanced biocomputing systems that are far superior to those shown before. These examples also provide an insight into the vast amounts of information potentially hidden within the low-affinity biomolecular interactions of living organisms. Regrettably, such complex systems cannot yet be tested experimentally until next-generation algorithms are developed that predict affinities between molecules with dramatically higher precision.



**Fig. 5 | Square-root circuit and its design and performance.** **a**, Experimental set-up of the circuit. **b**, Design of the circuit for computing high bit output  $O_1$ . **c**, Design of the circuit for computing low bit output  $O_2$ . In **b** and **c**, nucleotides that are complementary to the corresponding sequence (shown with arrows) are written in black, and non-complementary ones in red. Exclamation marks flag different orientations of the  $I_2$  input. **d**, Computation results were measured using a plate fluorometer, 5 min after addition of the inputs. The true/false

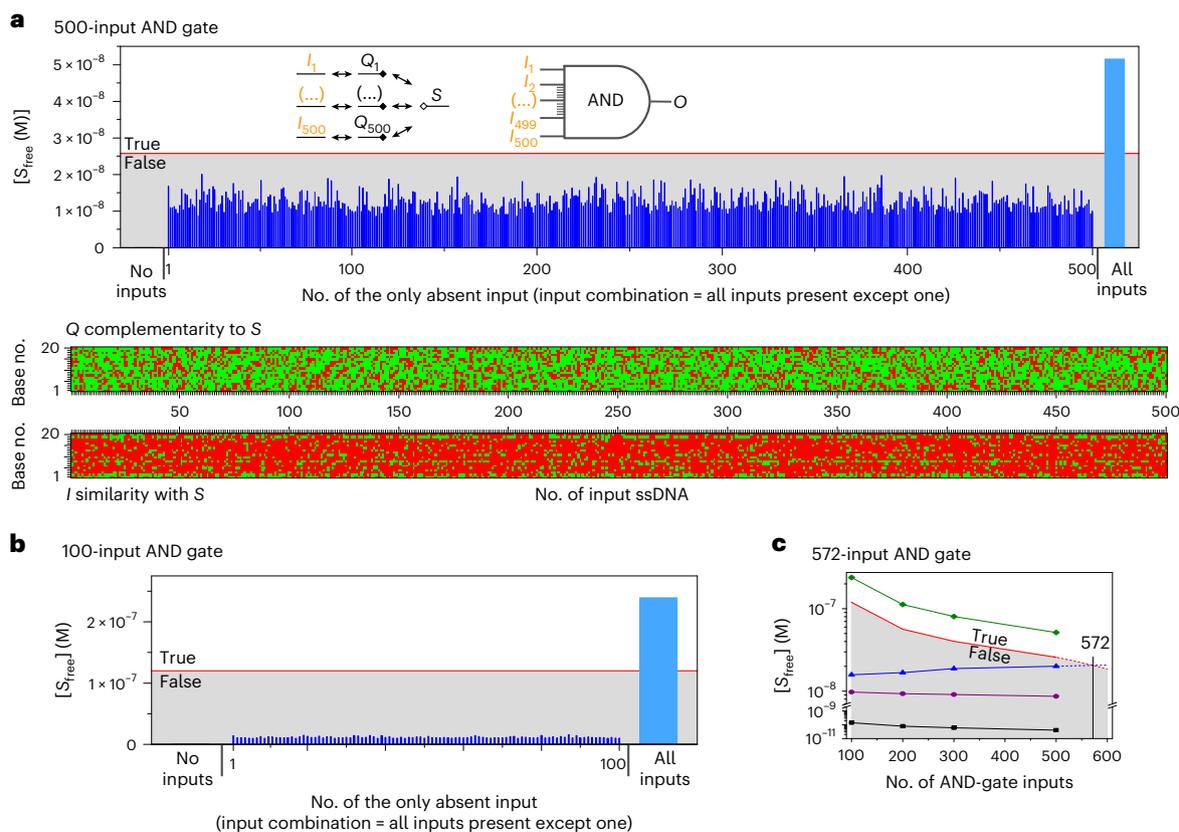
threshold is the average of the maximum and minimum outputs. The outputs of each gate are normalized by the respective threshold signal. Data are presented in relative fluorescence units (r.f.u.) as mean  $\pm$  s.d. ( $n = 4$  independent samples). **e**, Confocal scan of the samples 1 h after addition of the inputs (the image is scaled along the x axis for easier comparison with **d**), representative of  $n = 4$  independent experiments (additional images in the genuine aspect ratio are provided in Supplementary Fig. 11).

I should explicitly note that the following simulations have two distinct levels that have separate limitations for translation to experiment. The first is the NUPACK sequence-based prediction of strand affinities, which needs further improvement to account for all complexities of DNA interactions<sup>25</sup>. The second is the analysis of reversible reactions within the set of weakly interacting molecules according to the well-established fundamental law of mass action. Therefore, in principle, if a set of molecules is identified with the same affinity constants as in the examples below, they should show the predicted behaviour in experiments.

Let us imagine that we have a gene with a certain 20-base sub-sequence that can be affected by associating with a same-sized single-stranded oligo, for example, via the RNase-H-mediated antisense mechanism (that is, cleavage of messenger RNA (mRNA) upon its hybridization with ssDNA<sup>26</sup>). Within the standard complementarity paradigm, this 20-base-long part offers a maximum of  $4^{20} \approx 10^{12}$  possibilities that can affect it; that is, there are  $10^{12}$  ssDNAs that can either bind or not bind the target sequence. Strand commutation offers at

least  $2^{572} \approx 10^{172}$  possibilities for affecting such a gene in the same set-up. That is substantially more than the number of all elementary particles in the visible universe by any published estimate.

I demonstrate this feature by in silico simulation of a 500-input AND gate and show that up to 572-input gates can be achieved in this simulation. Interestingly, so far, such input bitness has not yet been accomplished, even in the most powerful electronic computers. To construct such an AND gate, which was designed similarly to the three-input AND gate shown in Fig. 2c, the following algorithm was developed. First, the sequence of the signal oligo  $S$  was fixed, then  $I$ - $Q$  pairs were sequentially selected that (1) had a predefined range of  $I$ - $Q$  and  $Q$ - $S$  affinities and (2) were checked to have 'parasite' affinity with all previously selected oligos below a certain limit. The optimized thresholds that could generate more than 1,000  $I$ - $Q$  pairs had the following approximate (for a definition see Supplementary Note 1) dissociation constants:  $5 \times 10^{-7} \text{ M} < K_d[QS] < 9 \times 10^{-7} \text{ M}$ ;  $4.1 \times 10^{-10} \text{ M} < K_d[I/Q] < 2.6 \times 10^{-9} \text{ M}$ ; parasite crosstalk threshold  $K_d > 1.8 \times 10^{-5} \text{ M}$ .



**Fig. 6 | High bitness systems—500-input and 572-input AND gates.**

**a**, Schematics and simulated performance of the 500-input AND gate. The ‘all inputs = 1’ and ‘all inputs = 0’ values that determine the threshold (the average of maximum and minimum output signals) are shown, together with 500 input combinations corresponding to cases with all inputs = 1 except the  $X$ th input = 0. All other  $(2^{500} - 502) \approx 3 \times 10^{150}$  combinations must produce a smaller output than these 500 input combinations. The outputs of each gate are normalized by the respective threshold signal. The ssDNA sequences and their concentrations are listed in Supplementary Table 1. The complementarity of mediator  $Q$  nucleotides to  $S$  as well as the similarity of inputs  $I$  with  $S$  are shown schematically below the

graph: green squares represent matching bases and red squares non-matching bases. All oligos are 20-nt long. **b**, Simulated performance of the 100-input AND gate. Note a higher value for ‘all inputs = 1’ than for the 500-input AND gate.

**c**, Values for the AND gate outputs depending on the number of the gate’s inputs: black squares, ‘all inputs = 0’; green diamonds, ‘all inputs = 1’; violet circles and blue triangles, the minimum and maximum output values among ‘all inputs = 1 except one’ combinations, respectively. Data extrapolation shows that the maximum output values among ‘all inputs = 1 except one’ combinations are still below threshold for the 572-input AND gate.

The set was optimized by removing ‘bad’  $I$ - $Q$  pairs (those that yielded worse-than-average outputs) by sequential testing of AND gates with an increasing number of inputs (100, 200, 300 and 500). Such escalation was performed to reduce the computation time: a 100-input gate requires ~3 min-CPUcore processing, a 200-input gate 10 min-CPUcore, and a 500-input gates 216 days-CPUcore. Finally, 500 pairs were identified for correct performance of the 500-input AND gate shown in Fig. 6a (Supplementary Table 1 lists the sequences and concentrations of the oligos).

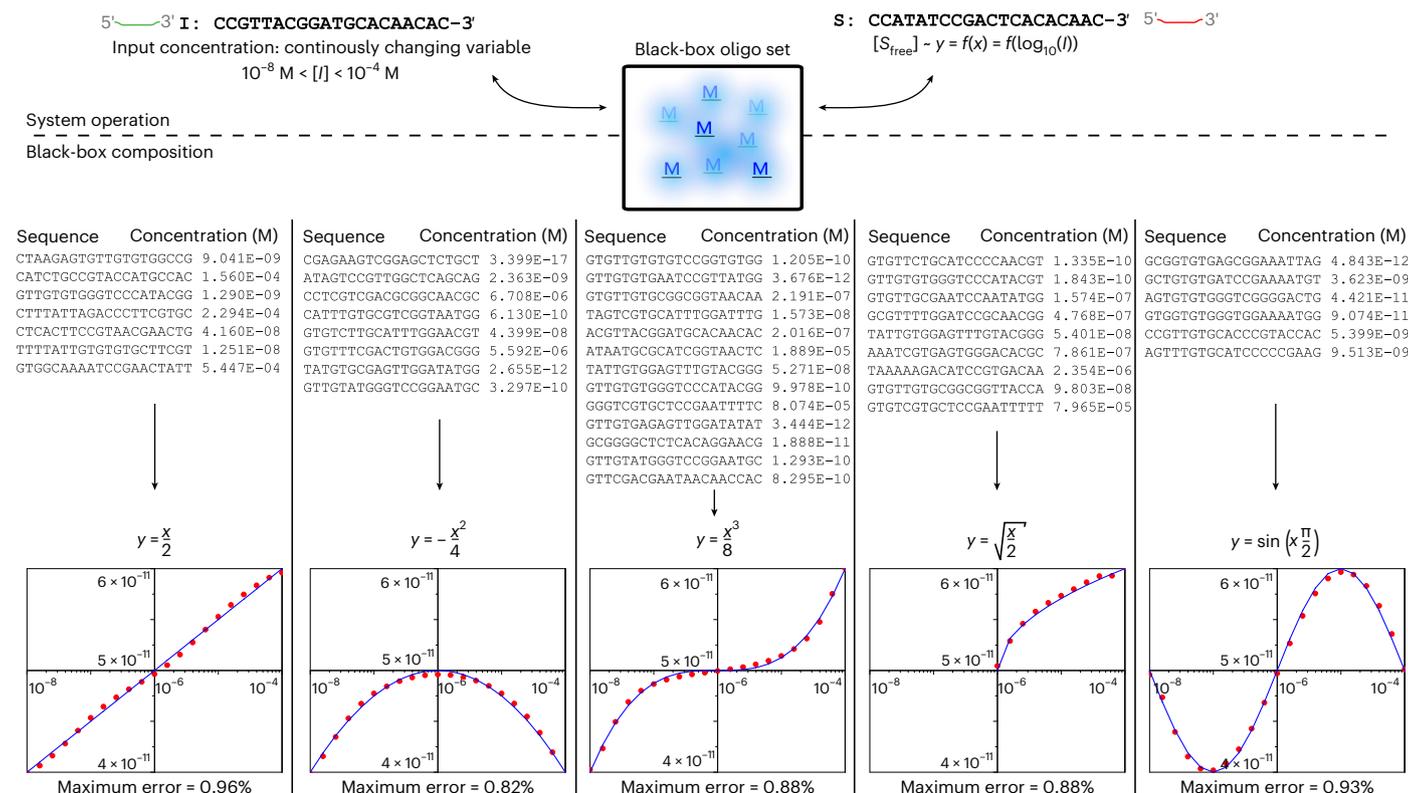
It should be noted that, because even 100-input gates have an enormous number of different input combinations ( $2^{100} \approx 10^{30}$ ), to confirm correct logic gating of an  $N$ -input AND gate, only the following  $N + 2$  input combinations were tested: (1) all inputs = 0, that is,  $S + \{\text{all } Qx\}$ ; (2) all inputs = 1, that is,  $S + \{\text{all } Qx\} + \{\text{all } Ix\}$ ; (3) all except one input = 1, that is,  $S + \{\text{all } Qx\} + \{I_1, \dots, I_{x-1}, I_{x+1}, I_N\}$ , for  $1 \leq x \leq N$ . Indeed, if all single-false input combinations have output less than the threshold (determined by ‘all 0’ or ‘all 1’ input conditions), the gate should operate correctly for all other input combinations, as they exhibit even less output.

Interestingly, Fig. 6c shows that, although the output for the all-true input condition gradually decreases with increasing number of inputs in the constructed gates, the maximum value for the single-false input condition stays almost the same for all gates (compare the performance of the 500-input and 100-input AND gates in Fig. 6a,b). If we

extrapolate these data, we can see that the all-true input values pass above the threshold (twice the maximum output for the single-false input) up to the 572-input AND gate, which is thus also possible.

Although Boolean algebra perfectly serves the needs of digital electronic computing, it is unlikely to do the same for many natural processes within living systems. On the contrary, living systems mostly rely on analogue information-processing, because the concentrations of participating molecules usually change non-discretely, that is, continuously.

Accordingly, the remarkable abilities of strand commutation will now be demonstrated for complex modulation of analogue signals by simulating ssDNA ensembles that solve algebraic equations for continuously changing variables, for example,  $y = Ax^2$  for any  $-2 < x < 2$ . To this end, an evolutionary algorithm is developed here that finds the set of ssDNAs that generates the best fit to the algebraic equations. The system was designed as follows. Similarly to the systems above, one strand was fixed as the input  $I$  (CCGTTACGGATGCACAACAC) and another one as the output signal  $S$  (CCATATCCGACTCACACAAC). The total concentration of  $S$  was fixed at  $10^{-10}$  M. The concentration of free  $S$  would indicate the output  $y$  coordinate value as  $y = ([S_{\text{free}}] - [y_{\text{offset}}]) / [\text{coeff}]$ , where  $[y_{\text{offset}}] = 5 \times 10^{-11}$  M and  $[\text{coeff}] = 10^{-11}$  M are the arbitrary offset and normalization coefficients, respectively. As the  $x$ -coordinate value, the logarithm of the normalized concentration of  $I$  was used, that



**Fig. 7 | Analogue systems solving elementary algebra problems.** Simulation of the non-Boolean black-box systems for solving various problems with continuously changing variables. Sequences of the processing ssDNA of each system are shown, together with their respective concentrations. The NUPACK-simulated performance of each system is shown as red dots and the theoretical

solution of the equation is shown as a blue line. The  $x$  values of the red dots were used within the evolutionary algorithm to minimize the system's output error with respect to the theoretical value. The optimization was stopped once the maximum error (displayed below each graph) at each point dropped below 1%.

is,  $x = \log([I]/[x_{\text{offset}}])$ , where  $[x_{\text{offset}}] = 10^{-6} \text{ M}$  is an arbitrary offset concentration. The ranges of  $x, y$  are standardized so that  $10^{-8} \text{ M} < [I] < 10^{-4} \text{ M}$  and  $4 \times 10^{-11} \text{ M} < [S] < 6 \times 10^{-11} \text{ M}$ , so  $-2 \leq x \leq 2$  and  $-1 \leq y \leq 1$ .

The designed algorithm was first given an initial set of signal-processing ssDNAs. At each evolutionary epoch, the algorithm sequentially generated mutations within the sequences and altered the concentrations of the strands. For each set, a loss function value was calculated (using the mean-square method). At the end of the epoch, the set that yielded the closest fit to the correct solution to the problem was selected as the base for the next epoch. The algorithm stopped once a solution with a maximum error of 1% at each tested  $x$  value was found. If no solution could be found after a certain period, the number of oligos in the set was increased.

Figure 7 shows the identified sets of oligos, which induce changes in free  $S$  concentration (output) as a response to variation of input  $I$  concentration according to the predefined functions normalized to fit the  $x, y$  region defined by  $y = 0.5x, y = (-0.25)x^2, y = (0.125)x^3, y = \sqrt{x/2}$  and  $y = \sin(x\pi/2)$ . Evidently, these sets can solve mathematical equations of elementary algebra. For example, Supplementary Fig. 12 demonstrates the solution of the equation  $(-0.25)x^2 = \sin(x\pi/2)$ .

### Gene regulation and the natural importance of strand commutation

Finally, as a glimpse into the importance of strand commutation as a natural phenomenon, I experimentally demonstrate a disruptive idea: an arbitrary unstructured ssDNA can specifically regulate expression of a given gene, irrespective of the mutual complementarity or similarity between their sequences. By specific regulation, I do not mean that gene expression becomes susceptible to any input regardless of

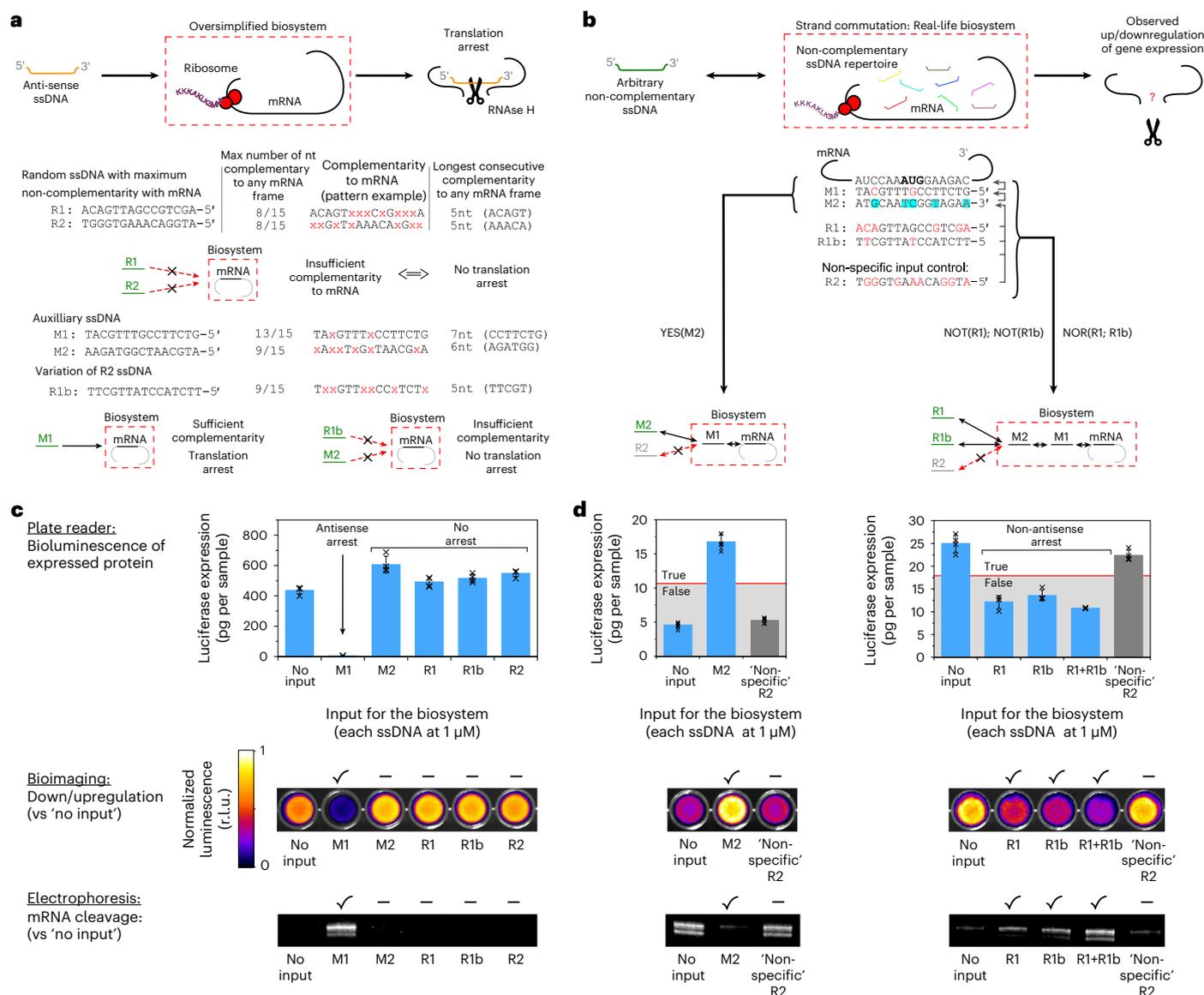
its sequence, but that the expression can be made dependent on the specific sequence of the given arbitrary ssDNA while staying unresponsive to other sequences.

Although such an idea lies outside the current paradigm (as it has become established over the decades to search for gene-regulating oligos based on their complementarity to the target gene), it becomes evident from the above-described logic circuits.

To prove the point, let us again consider how strand commutation influences one of the most explored natural mechanisms of gene expression regulation, namely, RNase-H-mediated degradation of mRNA upon its binding with the antisense ssDNA, which results in translation arrest (scheme in Fig. 8a)<sup>26</sup>. Accordingly, as the output oligo  $S$  we will now use a physiologically meaningful nucleic acid, specifically, mRNA.

Because we are dealing with low-affinity interactions, for an unobstructed proof we need a reaction system that has absolute minimum components that can interfere with the circuitry, even with the lowest affinities. Accordingly, the following experimental set-up is used. First, the model gene mRNA (Firefly Luciferase, Supplementary Fig. 13) is subjected to ssDNA inputs in the presence of the purified RNase H, and the results are passed to a rabbit reticulocyte extract cell-free translation system for gene expression. Successful translation with luminescent product is considered output = 1, and translation arrest with diminished luminescence is considered output = 0.

It is next shown that an arbitrary unstructured ssDNA, even of the maximum possible non-complementarity to the gene's sequence, can be made into a specific input that regulates the gene expression output result of such a circuit. To generate short oligos of the ultimate non-complementarity to mRNA, two evaluation parameters were used:



**Fig. 8 | Gene regulation circuits, and the strand commutation mechanism versus the conventional antisense concept.** **a**, Schematics for the antisense mechanism dependent on RNase H. Design of the random 15-nt sequences R1, R2 of maximum possible non-complementarity to Firefly Luciferase mRNA, mediators M1, M2 and R1b input variant. Nucleotides that are complementary to mRNA are written in black, and non-complementary ones are each shown as a red x (a full analysis of complementarity and similarity is provided in Supplementary Figs. 14 and 15). **b**, Strand commutation mechanism behind the regulation of gene expression with ssDNA of maximum non-complementarity to the target mRNA, and schematics of YES(M2), NOT(R1), NOT(R1b) and NOR (R1, R1b) gates. Nucleotides that are complementary to the corresponding sequence (shown with arrows) are written in black, and non-complementary ones in red. M2 bases that do not match the mRNA sequence are highlighted in cyan. **c,d**, Effect of addition of the designed ssDNA to the translation reaction: inputs are first incubated at 1 μM with biosystems containing RNase H with mRNA only (**c**) or mRNA + M1/mRNA + M1 + M2 (**d**), then mixed with the cell-free expression system. Investigation of the produced effect is carried out

with plate-reader-measured luminescence of the generated Luciferase upon addition of the substrate, plotted in mass units based on the recombinant protein calibration curve. Data are presented as mean ± s.d. (**a,b**, NOR gate,  $n = 4$  independent samples; **b**, YES gate,  $n = 6$  independent samples). Bioimaging photos visualize the normalized luminescence of the generated Luciferase (maximum observed signal within each biosystem is normalized to ~90% of the scale range;  $n = 3$  independent samples; study performed independently from the plate-reader study with increased reaction volumes; Supplementary Fig. 16 provides photographs of all samples). Electrophoresis data show cleaved mRNA fragments of the expected size (Supplementary Fig. 13 shows the cleavage site) for samples without addition of cell-free expression system components and increased concentrations of mRNA and RNase H for better detection (inputs were kept at 1 μM; Supplementary Fig. 17 and the source data file provide the full photographs; representative of  $n = 3$  independent experiments). In all cases, 'No input' represents the addition of the same volume of buffer; ticks denote observed differences in protein expression/mRNA cleavage as compared to the 'No input' condition. R.l.u., relative luminescence units.

(1) the maximum number of the oligo's bases complementary to mRNA within any linear sub-sequence of mRNA (not considering potential bulges); (2) the longest consecutive sub-sequence of the oligo complementary to mRNA. Among 100,000 randomly generated 15-nt-long ssDNAs, the minimum observed value for the first parameter was 8 and for

the second was 5. Accordingly, as the model inputs, two unstructured oligos R1 and R2 were chosen with both parameters simultaneously at these minima (the sequences are shown in Fig. 8 and Supplementary Note 1, and a detailed analysis of complementarity in Supplementary Fig. 14). Although it is possible that even less complementary ssDNA

could be found by screening all 15-nt oligos, it would hardly affect the conclusions. As shown in Fig. 8a, addition of either R1 or R2 at a concentration of 1  $\mu\text{M}$  does not inhibit mRNA translation. Accordingly, they do not exhibit the ‘conventional’ antisense effect and are indeed non-complementary to mRNA.

However, a specific logic gate NOT(R1) can be constructed such that addition of R1 (but not R2) leads to translation inhibition. As in the NOT gates described above, we need to add two mediator oligos, M1 and M2, between R1 and S (as the target S sub-sequence we use an mRNA sub-sequence around the AUG codon).

In this example, for added value, we refrain from changing the concentrations of the oligos during A&C adjustment of the circuit and operate only by varying the affinities. The concentrations of all input oligos are fixed at 1  $\mu\text{M}$ , which is at the level of the half-maximum inhibitory concentration ( $\text{IC}_{50}$ ) of the common antisense ssDNA.

After designing and screening various M1/M2 candidates, the functional sequences shown in Fig. 8 were found. Additionally, to demonstrate that R1 is not unique or unusual, an R1b input of similar affinity to M2 was designed with close to minimum possible complementarity to mRNA (Fig. 8a and Supplementary Fig. 14).

Of all the mentioned ssDNA, only M1 has sufficient, but not full, complementarity to mRNA to induce mRNA cleavage by RNase H (Fig. 8a, bottom half, lower) and inhibit protein expression (Fig. 8a, bottom half, higher). R1, R1b and R2 are inert.

However, due to strand commutation, the presence of M1 and M2 in the mRNA mix makes the R1 and R1b efficient inhibitors of the observed translation, as shown in Fig. 8b. Moreover, such inhibition is specific (the ‘non-specific’ R2 still has no effect) and pronounced, with 2.1- and 1.8-fold reductions in protein production for R1 and R1b, respectively. Their combined action, each at 1  $\mu\text{M}$ , within the NOR(R1, R1b) gate results in 2.3-fold inhibition. Figure 8 also explicitly confirms that R1 and R1b (but not R2) induce additional mRNA cleavage by RNase H through strand commutation with M1 and M2.

Accordingly, for any observer who does not take into account the presence of M1 and M2 (which are also non-complementary to mRNA), R1 and R1b will have an ‘enigmatic’ off-target translation inhibition action. It should also be noted that the used oligos have no noticeable similarity to the mRNA sequence such as to expect a ‘conventional’ influence on gene expression as sense oligonucleotides<sup>27</sup> (a detailed analysis of similarity to the mRNA sequence is presented in Supplementary Fig. 15).

This Article shows that the overall ssDNA/ssRNA repertoire in the microenvironment may be no less important for gene regulation with oligonucleotides than their complementarity to the gene sequence. Due to the high abundance of various nucleic acids in cells, the strand commutation phenomenon may have an immense effect on maintaining the homeostasis of gene expression. At the same time, this study has been performed in an extremely simplified biosystem containing only a small set of nucleic acids. It is premature to judge whether strand commutation significantly affects gene expression in vivo. However, if it does, it may be a useful as a source of drug candidates and for understanding brain activity, evolution and other complex processes of living systems.

## Discussion

There are several aspects for discussion regarding the strand commutation phenomenon: (1) its advantages and drawbacks as a biocomputing approach, (2) its immediate importance for biomedical research and applications and (3) a general overview of its potential involvement in natural processes.

## Biocomputing

Although two mathematical regimes of computing (Boolean logic and elementary algebra) are demonstrated here, it is entirely possible that larger volumes of data can be processed more quickly if a different

mathematics is used. The affinity continuum allows the generation of an ssDNA set with virtually any predetermined affinity matrix. This unique feature can offer experimental implementations of many mathematical ideas proposed within reaction network theory<sup>28,29</sup>.

As compared to other biocomputing methods that allow complex computing, the advantage of the strand commutation concept is the virtual absence of restrictions regarding molecules that are compatible with the method. No limitations are imposed on the relationships of the input, output and mediating entities; that is, they can be of the same or different sizes (unlike in the DNAzyme concept), they can be very short (unlike in the strand-displacement concept), and they can all be of a non-DNA nature.

I believe that strand commutation can achieve high computation speeds, first because signal propagation along the circuit is determined by the lifetime ( $k_{\text{off}}$ ) of each complex in the system, which can be designed to be short due to (1) low affinity between the strands in general and (2) an unlimited flexibility in the choice of strand sequences and their undercomplementarity patterns. Second, with strand commutation, systems can operate without concerns about unwanted crosstalk between participating oligos at high concentrations, which means faster speeds in general (Supplementary Figs. 3 and 4 and ref. 30). Indeed, because of the low-affinity regime, oligos in strand commutation systems can barely bind even their intended partners, let alone non-desired strands. In contrast, the strand-displacement approach usually deals with much longer strands that form high-affinity complexes. Their high concentrations may result in spurious reactions that would destabilize signal transduction and limit circuit complexity. At the same time, the kinetics of toehold-mediated strand exchange is complex and adjustable<sup>24,30,31</sup>, so it is possible that computation speeds may increase in the future.

In the demonstrated systems, data are encoded within the concentrations of the molecular complexes. However, other system parameters can also be used for this purpose. For example, storage of information as the rates of concentration changes would allow even higher speeds of data processing. However, prediction of molecular behaviour at equilibrium remains far from ideal. The design of dynamic systems thus has to wait for advances in molecular modelling methods<sup>32</sup>.

Regarding the limitations of the approach, in most of the above systems the input and output oligos were not simultaneously selected arbitrarily; the oligos were designed to roughly fit the predefined matrix of mutual affinities between all oligos within the system. This is not as convenient as the strand-displacement approach. Nevertheless, I believe that, with sufficient strand length and proper A&C considerations, arbitrary oligos can be employed as inputs and outputs. Indeed, in the gene-expression control study, the input by design had minimal nucleotides complementary to the gene sequence. Moreover, the issue can be easily addressed by adding more commutating oligos between input and output, though at the cost of higher prediction complexity.

At this point, the presented systems have no built-in error correction mechanisms (as has been realized in the strand-displacement approach<sup>15</sup>). On the one hand, they are unnecessary at this point, as accurate and faster computation could be achieved without them. On the other hand, their incorporation could make systems more robust.

As concerns the scalability of this approach, it is unclear what complexity of computation can be achieved in a single homogeneous sample until more accurate interaction prediction algorithms are developed. With an increase in ssDNA lengths, the affinity continuum expands so rapidly that unwanted crosstalk seems to be easily avoidable. At fixed oligo lengths, a substantial increase in the number of participating oligos would generally require decreasing their concentrations and raising their affinities to sustain the reversible binding regime. This may decrease computation speeds.

At the same time, a true path to scaling up lies in compartmentalization, as nature does with multicellular organisms. Complexity

would substantially increase if one could achieve physical separation of the ‘computation units’ on a circuit board, either with semipermeable membranes or by implementing strand commutation on the solid phase (via immobilization of some oligos). Such a solution would completely eliminate the unwanted crosstalk problem. However, signal transduction between such units remains the grand challenge for all biocomputing approaches. Furthermore, compartmentalization concepts will be faced by long-term stability issues—they may require most of the natural cell machinery to maintain molecular consistency.

### Immediate importance

The straightforward implications of these findings are primarily related to applications of relatively small RNA and DNA (sRNA/sDNA), including the long noncoding RNA. Although little is still known about the whole spectrum of sRNA activity, they have been shown to regulate transcription, chromosome replication, RNA processing and translation, mediate communication between cells through exosome transport, and so on<sup>33</sup>.

The results of this Article change the perspective for analysis of such nucleic acids. Previously, the pool of gene-regulating strands has been established primarily based on complementarity to the target. The present study suggests that, due to strand commutation, multiple tiers of consecutively interacting nucleic acids (including those absolutely non-complementary to the target) need to be analysed to understand the full extent of the regulation processes. Such analysis may offer new insights into complex phenomena that involve nucleic-acid turnover, such as cancer<sup>34</sup>, ageing<sup>35</sup> and even memory<sup>36</sup>. It would be particularly interesting to understand whether the long noncoding RNAs participate in gene regulation via strand commutation.

Understanding the sRNA/sDNA repertoire and their interactions may allow better prediction and minimization of off-target effects<sup>37</sup> of gene therapies, including gene editing and antisense and siRNA treatments of various diseases<sup>38</sup>. Furthermore, reduction of the unexpected long-term side effects of next-generation vaccines based on nucleic acids may be especially important in the post-COVID era.

The reported phenomenon can also be employed to design gene therapy with fine multifactor tuning based on cellular cues. In particular, strand commutation allows the incorporation of additional levels of control over gene therapy to increase the specificity of therapeutic circuitry. For example, a gene transfection agent could be delivered systemically together with a combination of sRNA that allows the expression of the gene only in a predefined profile of the cell-produced sRNA.

### Outlook of a general natural phenomenon

Since its emergence about a century ago, molecular biology has redefined our understanding of living systems. Numerous research efforts have been devoted to reveal the vast pathways of molecular interactions within organisms. The high specificity of the discovered interactions forms the fundamental reliability basis of life. The concept of a step-by-step information transfer along strictly defined molecular pathways seems perfect to ensure the predictability and sustainability of living processes.

The present work exposes a different but complementary idea—reliable data transition through a continuum of reversible low-affinity/low-specificity molecular interactions. As shown above, the discovered molecular commutation phenomenon features a relatively high degree of reliability of data processing and storage, with a remarkable flexibility of tuning.

Importantly, the commutation phenomenon is based entirely on the fundamental law of mass action and is not restricted to nucleic acids. Although I believe that this phenomenon will be of primary importance for the DNA/RNA realm (due to the unique straightforward sequence–affinity relationship), in principle it may manifest itself with any other molecule type (proteins, small molecules, carbohydrates, lipids and so on). It may thus be useful to look for commutation in the

realm of proteins and other molecules participating in ‘promiscuous’ interactions<sup>39</sup>, that is, when a receptor can bind to many diverse ligands. For example, in the case of olfactory receptors, an enormous variety of smell sensations can be generated by profiling binding affinities of each ligand with a few receptor types, and the promiscuity of bone morphogenetic protein enables wider cellular addressing, as shown in ref. 40.

At the same time, although the present work uncovers the value of the promiscuous interactions of undercomplementary nucleic acids, the demonstrated data-processing power stems not from the mere ability of the ssDNA to bind to many other undercomplementary strands, but from the reversibility of the low-affinity interactions, that is, from the continuous exchange and rebalancing of the molecules between their complexes, which results in signal transduction (hence, the term ‘commutation’). This is conceptually different from olfactory receptors, which use the low affinity of molecular binding primarily for a constant system reset to sense new odours.

In general, if molecular commutation significantly manifests itself *in vivo*, it may be involved in two ways: (1) to uphold homeostasis due to the immense number of weakly contributing molecules and (2) quite the opposite, to generate randomness for evolution, which may also result in complex failures within the organism. In particular, a single mutation or concentration change of a far-tier effector can disturb multiple important pathways via a butterfly effect. For example, consider the memory circuit above: the readout of three different inputs can be disturbed by a concentration change or mutation of a single  $Q_R$  strand.

As noted above, I believe it would be hard to directly observe and prove manifestation of the molecular commutation phenomenon *in vivo*. In contrast to high-affinity molecular pathway systems, any measurement within a low-affinity system will have a nonlinear impact on multiple states (or pathways) of the interactions. Consider, for example, how measurement (hence altering concentration) of the  $I_2$  input in the square-root circuit would affect complexes of the  $X_{12}$ ,  $X_{23}$ ,  $X_{24}$  and  $S_2$  output. Observing the phenomenon and reconstructing the causal relationships in such complex and ‘concentrated’ systems such as living organisms will be a challenging task. It will require substantial advancements in biosensing techniques, molecular modelling and big data analysis.

### Conclusion

Further investigation of strand commutation (and molecular commutation in general) could open new perspectives in fundamental science and biomedicine. Even if no natural processes are meant to utilize the mechanism, there is zero chance that it cannot spontaneously manifest itself. Representing the dark matter of molecular biology, low-affinity interactions are much more abundant and unpredictable than high-affinity ones. In any case, the existence of the demonstrated phenomenon substantially complicates the study of molecular interactions within living systems. Perhaps the low-affinity ‘affinome’ (as part of the interactome) can give us the key to those complex matters that we struggle to understand within the current paradigm of high-specificity and high-affinity interactions.

### Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, 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/s41557-022-01111-y>.

### References

1. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
2. Ceze, L., Nivala, J. & Strauss, K. Molecular digital data storage using DNA. *Nat. Rev. Genet.* **20**, 456–466 (2019).

3. Benenson, Y. Biomolecular computing systems: principles, progress and potential. *Nat. Rev. Genet.* **13**, 455–468 (2012).
4. Tregubov, A. A., Nikitin, P. I. & Nikitin, M. P. Advanced smart nanomaterials with integrated logic-gating and biocomputing: dawn of theranostic nanorobots. *Chem. Rev.* **118**, 10294–10348 (2018).
5. Nikitin, M. P. et al. Enhancement of the blood-circulation time and performance of nanomedicines via the forced clearance of erythrocytes. *Nat. Biomed. Eng.* **4**, 717–731 (2020).
6. De Silva, P. A., Gunaratne, N. H. Q. & McCoy, C. P. A molecular photoionic and gate based on fluorescent signalling. *Nature* **364**, 42–44 (1993).
7. Erbas-Cakmak, S. et al. Molecular logic gates: the past, present and future. *Chem. Soc. Rev.* **47**, 2228–2248 (2018).
8. Nikitin, M. P., Shipunova, V. O., Deyev, S. M. & Nikitin, P. I. Biocomputing based on particle disassembly. *Nat. Nanotechnol.* **9**, 716–722 (2014).
9. Katz, E. & Privman, V. Enzyme-based logic systems for information processing. *Chem. Soc. Rev.* **39**, 1835–1857 (2010).
10. Friedland, A. E. et al. Synthetic gene networks that count. *Science* **324**, 1199–1202 (2009).
11. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* **333**, 1307–1311 (2011).
12. Xie, M. & Fussenegger, M. Designing cell function: assembly of synthetic gene circuits for cell biology applications. *Nat. Rev. Mol. Cell Biol.* **19**, 507–525 (2018).
13. Fan, D., Wang, J., Wang, E. & Dong, S. Propelling DNA computing with materials' power: recent advancements in innovative DNA logic computing systems and smart bio-applications. *Adv. Sci.* **7**, 2001766 (2020).
14. Adleman, L. M. Molecular computation of solutions to combinatorial problems. *Science* **266**, 1021–1024 (1994).
15. Qian, L. & Winfree, E. Scaling up digital circuit computation with DNA strand displacement cascades. *Science* **332**, 1196–1201 (2011).
16. Stojanovic, M. N. & Stefanovic, D. A deoxyribozyme-based molecular automaton. *Nat. Biotechnol.* **21**, 1069–1074 (2003).
17. Elbaz, J. et al. DNA computing circuits using libraries of DNAzyme subunits. *Nat. Nanotechnol.* **5**, 417–422 (2010).
18. Zadeh, J. N. et al. NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173 (2011).
19. Dirks, R. M., Bois, J. S., Schaeffer, J. M., Winfree, E. & Pierce, N. A. Thermodynamic analysis of interacting nucleic acid strands. *SIAM Rev.* **49**, 65–88 (2007).
20. Bissels, U. et al. Absolute quantification of microRNAs by using a universal reference. *RNA* **15**, 2375–2384 (2009).
21. Calabrese, J. M., Seila, A. C., Yeo, G. W. & Sharp, P. A. RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **104**, 18097–18102 (2007).
22. Zhang, J. X. et al. Predicting DNA hybridization kinetics from sequence. *Nat. Chem.* **10**, 91–98 (2018).
23. Bee, C. et al. Molecular-level similarity search brings computing to DNA data storage. *Nat. Commun.* **12**, 4764 (2021).
24. Song, T. et al. Fast and compact DNA logic circuits based on single-stranded gates using strand-displacing polymerase. *Nat. Nanotechnol.* **14**, 1075–1081 (2019).
25. Camunas-Soler, J., Alemany, A. & Ritort, F. Experimental measurement of binding energy, selectivity and allostery using fluctuation theorems. *Science* **355**, 412–415 (2017).
26. Bennett, C. F. & Swayze, E. E. RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu. Rev. Pharmacol. Toxicol.* **50**, 259–293 (2010).
27. Modarresi, F. et al. Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. *Nat. Biotechnol.* **30**, 453–459 (2012).
28. Viswa Virinchi, M., Behera, A. & Gopalkrishnan, M. A reaction network scheme which implements the EM algorithm. In *DNA Computing and Molecular Programming. DNA 2018. Lecture Notes in Computer Science* (eds Doty, D. & Dietz, H.) Vol. 11145, 189–207 (Springer, 2018).
29. Cappelletti, D., Ortiz-Muñoz, A., Anderson, D. F. & Winfree, E. Stochastic chemical reaction networks for robustly approximating arbitrary probability distributions. *Theor. Comput. Sci.* **801**, 64–95 (2020).
30. Zhang, D. Y. & Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **131**, 17303–17314 (2009).
31. Srinivas, N. et al. On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Res.* **41**, 10641–10658 (2013).
32. Zhang, J. X. et al. Predicting DNA hybridization kinetics from sequence. *Nat. Chem.* **10**, 91–98 (2017).
33. Storz, G. An expanding universe of noncoding RNAs. *Science* **296**, 1260–1263 (2002).
34. Janssen, A., Colmenares, S. U. & Karpen, G. H. Heterochromatin: guardian of the genome. *Annu. Rev. Cell Dev. Biol.* **34**, 265–288 (2018).
35. Tiwari, V. & Wilson, D. M. DNA damage and associated DNA repair defects in disease and premature aging. *Am. J. Human Genet.* **105**, 237–257 (2019).
36. Bédécarrats, A., Chen, S., Pearce, K., Cai, D. & Glanzman, D. L. RNA from trained aplysia can induce an epigenetic engram for long-term sensitization in untrained aplysia. *eNeuro* **5**, 38–56 (2018).
37. Manghwar, H. et al. CRISPR/Cas systems in genome editing: methodologies and tools for sgRNA design, off-target evaluation, and strategies to mitigate off-target effects. *Adv. Sci.* **7**, 1902312 (2020).
38. Gong, H., Liu, C.-M., Liu, D.-P. & Liang, C.-C. The role of small RNAs in human diseases: potential troublemaker and therapeutic tools. *Med. Res. Rev.* **25**, 361–381 (2005).
39. Nobeli, I., Favia, A. D. & Thornton, J. M. Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol.* **27**, 157–167 (2009).
40. Su, C. J. et al. Ligand-receptor promiscuity enables cellular addressing. *Cell Syst.* **13**, 408–425 (2022).

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

Springer Nature or its licensor (e.g. a society or other partner) 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 2023

## Methods

### Materials

All oligonucleotides were purchased from Lumiprobe (Russia). Fluorophore- and quencher-labelled oligonucleotides, as well as ssRNA, were purified by polyacrylamide gel electrophoresis (PAGE). All other oligonucleotides were desalted (C18 cartridge). A list of the used oligonucleotides and their properties is provided in Supplementary Note 1.

Other materials were obtained from the following suppliers. Promega: rabbit reticulocyte lysate, nuclease-treated (cat. no. L4960); Luciferase control RNA (cat. no. L4561), ONE-Glo EX Luciferase assay system (cat. no. E8110), QuantiLum recombinant Luciferase (cat. no. E1701). New England BioLabs: RNase H (cat. no. M0297S). Thermo Fisher: SYBR Gold nucleic-acid gel stain (cat. no. S11494), RNA gel loading dye (cat. no. R0641). Sigma: tris(hydroxymethyl)aminomethane (cat. no. Sigma 7–9), ethylenediaminetetraacetic acid disodium salt dihydrate, Tween 20. SPL Life Sciences: 384-well plates (cat. no. 33384). All other chemical reagents were of analytical grade and were used without further purification.

### Logic gates and circuits

All gates (except gene regulation ones) were performed in 0.1 M Tris buffer, pH 7.0, with 1 M NaCl and 0.01% Tween 20 (TBST buffer) at room temperature (–22–25 °C). Generally, investigation of logic-gate performance was performed by mixing 50 µl of 2× concentrations of all oligonucleotides that resemble the logic-gate system with 50 µl of 2× concentrations of the inputs. Supplementary Note 1 lists the concentrations of the oligonucleotides in the final solution for each gate.

### Fluorescence measurements

Fluorescence was measured using a Clariostar plate reader (BMG Labtech) with the following excitation/emission settings: 530 ± 20/580 ± 30 nm and 610 ± 30/675 ± 50 nm for single Cy3 or Cy5, respectively. The same were used for samples with both Cy3 and Cy5. For the samples with three fluorophores, the following settings were used to minimize crosstalk: BDP-FL, 440 ± 20/510 ± 20 nm; Cy3, 560 ± 10/595 ± 10 nm; Cy5, 610 ± 20/690 ± 50 nm. The gain was adjusted for each experiment to achieve maximum signal at 90% of the reader's signal range.

### Confocal microscopy

Confocal microscopy images were obtained with a Fluoview FV3000 confocal laser scanning microscope (Olympus), a UPLSAPO ×4 objective, 561-nm and 640-nm lasers with 570–620-nm and 650–750-nm detection wavelengths, respectively, and a transmitted light detector at 640 nm. Multiple images were stitched during acquisition within the Fluoview software.

### Cell-free translation study

**Plate reader.** For this study, 1.5 µl of input solution in RNase H reaction buffer (NEB; 50 mM Tris-HCl, 75 mM KCl, 3 mM MgCl<sub>2</sub>, 10 mM DTT, pH 8.3 @ 25 °C) was mixed with 1 µl of Luciferase mRNA and RNase H solution in the same buffer to the following final concentrations: 1 µM inputs, 5.5 µg ml<sup>-1</sup> Luciferase mRNA and 33 U ml<sup>-1</sup> RNase H (note that deterioration of RNase H activity during storage may decrease the ratio of true/false output signals). The mixing was performed in a cool room (–16–18 °C) to maintain the originally simulated values of the oligo mutual affinities. RNase H reaction buffer has a lower salt concentration than used during simulation (1 M Na<sup>+</sup>, 25 °C); see Supplementary Note 1 for the affinities under both conditions. The mix was incubated for 40 min and then combined with 5.5 µl of Promega nuclease-treated rabbit reticulocyte lysate translation mix (per 100 µl of rabbit reticulocyte lysate, 1.4 µl of amino acids minus leucine, 1.4 µl of amino acids minus methionine and 100 µl of Tween 20 0.01% were added). After 1 h of incubation at 30 °C, 2.5 µl of each sample was transferred to a black 384-well plate and mixed with 20 µl of ONE-Glo EX Luciferase

assay system substrate. Luminescence signals were then measured using a Clariostar plate reader (BMG Labtech) in whole spectrum with a gain of 3,000, converted to the corresponding protein concentration based on the recombinant Luciferase calibration curve (plotted as bar charts in Fig. 8).

**Bioimaging.** The same experiment as above was independently performed using twice larger volumes of reagents, a 96-well plate and 100 µl of the same substrate. The plate was imaged with the LumoTrace FLUO bioimaging system (Abisense) equipped with a Retiga Lumo camera (Photometrics) using Micromanager<sup>41</sup> on Icy<sup>42</sup> software with Abisense plugins. Bioluminescence was captured with a 3-min exposure and no emission filter, and bright-field with a 100-ms exposure.

**Electrophoresis.** A 3-µl volume of input solution in RNase H reaction buffer, as above, was mixed with 2 µl of Luciferase mRNA and RNase H solution in the same buffer to the following final concentrations: 1 µM inputs, 16 µg ml<sup>-1</sup> Luciferase mRNA and 66 U ml<sup>-1</sup> RNase H (higher concentrations of mRNA and RNase H were used here to allow better detection of the cleavage results). The samples were incubated at 16–18 °C for 40 min, mixed with 7 µl of RNA gel loading dye (2×), incubated at 80 °C for 1 min, and subjected to 12% denaturing PAGE with 8 M urea. The gels were stained with SYBR Gold according to the manufacturer's protocol and imaged using a ChemiDoc MP system (Bio-Rad).

### NUPACK simulations

The performance of the 500-input gate was assessed by running NUPACK scripts on Amazon Elastic Compute Cloud (EC2) multiple parallel t3.medium spot-instances. Design and performance testing of all other systems was run on a personal computer.

### Statistical analysis

All studies were evaluated with  $n \geq 3$  independent samples at each data point to ensure reproducibility. Sample sizes are noted in the figure legends. No statistical methods were used to predetermine sample size. No specific blinding and randomization method was used. Data are presented as individual measurements as well as means ± s.d.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this Article.

### Data availability

The data that support the findings of this study are provided in the Article and its Supplementary Information, and are also available from the author on request. Source data are provided with this paper.

### Code availability

The NUPACK (Ubuntu 14.04 BASH) and MATLAB scripts used to design the systems and analyse their performance are too numerous to be readily shared publicly, but can be made available from the corresponding author on reasonable request.

### References

- Edelstein, A., Amodaj, N., Hoover, K., Vale, R. & Stuurman, N. Computer control of microscopes using µManager. *Curr. Protoc. Mol. Biol.* **14**, 14.20 (2010).
- De Chaumont, F. et al. Icy: an open bioimage informatics platform for extended reproducible research. *Nat. Methods* **9**, 690–696 (2012).

### Acknowledgements

I express deep gratitude to all developers of and contributors to the NUPACK algorithm, without which this study would be far less comprehensive. I thank I. L. Nikitina for assistance with manuscript

preparation and the Cell Technologies Center core facility of the Institute of Cytology of the Russian Academy of Sciences for the confocal images in Figs. 4 and 5.

### Author contributions

M.P.N. conceived the idea, designed and performed the study, and wrote the manuscript.

### Competing interests

M.P.N. has filed patent applications RU2019145384 (granted) and PCT/RU2020/050402 covering aspects of these findings. M.P.N. is the founder of the Abisense company, which manufactures the LumoTrace bioimaging system.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41557-022-01111-y>.

**Correspondence and requests for materials** should be addressed to Maxim P. Nikitin.

**Peer review information** *Nature Chemistry* thanks Anne Condon, Grigory Tikhomirov 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](http://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	BMG Labtech CLARIOstar control 5.20 R5 (plate reader acquisition software); Olympus FLUOView Software (confocal image acquisition); Micromanager 1.4.23_20160628 on ICY 2.1.4.0 with Abisense hardware control plugins (acquisition software for Abisense LumoTrace FLUO bioimaging system); Biorad Image Lab 4.1 build 16 (gel documentation software).
Data analysis	Custom BASH scripts for UBUNTU 14.04 with NUPACK 3.0.4. (generation of ssDNA sequences and analysis of circuit performance); <a href="http://www.nupack.org">www.nupack.org</a> (generation of RNA sequences for YES/NOT gates); Custom script for MATLAB R2013a (Fig. 3e); BMG Labtech MARS 3.02 R2 (plate reader data analysis); OriginPro 2015 (graphs); ImageJ 1.48v (image brightness/contrast; lookup table; overlay); Inkscape 0.92 (drawing of the figures).

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

The data that support the findings of this study are provided in the article and its Supplementary Information, and also available from the authors on request. Source data are provided with this paper.

NUPACK (Ubuntu 14.04 BASH) and MATLAB scripts used to design the systems and analyze their performance are too numerous to be readily shared publicly but can be made available from the corresponding author on reasonable request.

## 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](https://www.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	All studies were evaluated with $n \geq 3$ independent samples at each data point to ensure reproducibility. Sample sizes are noted in figure legends. No statistical methods were used to predetermine sample size.
Data exclusions	No data were excluded from analysis.
Replication	All experiments were performed at least twice, with similar results.
Randomization	No specific randomization method was used.
Blinding	No specific blinding method was used.

## 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

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging