

# A DNA-of-things storage architecture to create materials with embedded memory

Julian Koch<sup>1</sup>, Silvan Gantenbein<sup>2</sup>, Kunal Masania<sup>2</sup>, Wendelin J. Stark<sup>1</sup>, Yaniv Erlich<sup>3,4\*</sup> and Robert N. Grass<sup>1,4\*</sup>

**DNA storage offers substantial information density<sup>1-7</sup> and exceptional half-life<sup>3</sup>. We devised a 'DNA-of-things' (DoT) storage architecture to produce materials with immutable memory. In a DoT framework, DNA molecules record the data, and these molecules are then encapsulated in nanometer silica beads<sup>8</sup>, which are fused into various materials that are used to print or cast objects in any shape. First, we applied DoT to three-dimensionally print a Stanford Bunny<sup>9</sup> that contained a 45 kB digital DNA blueprint for its synthesis. We synthesized five generations of the bunny, each from the memory of the previous generation without additional DNA synthesis or degradation of information. To test the scalability of DoT, we stored a 1.4 MB video in DNA in plexiglass spectacle lenses and retrieved it by excising a tiny piece of the plexiglass and sequencing the embedded DNA. DoT could be applied to store electronic health records in medical implants, to hide data in everyday objects (steganography) and to manufacture objects containing their own blueprint. It may also facilitate the development of self-replicating machines.**

The world's data are growing at exponential rates in today's information age<sup>5</sup>. However, attempts to further miniaturize traditional storage architectures, such as hard-drives and magnetic tapes, are becoming increasingly difficult<sup>10,11</sup>. These devices are reaching their physical limitations and can hardly keep pace with digital storage requirements. Due to these challenges, there is great interest in the potential of using DNA molecules as an architecture for long-term cold storage. Previous studies have reported that DNA storage can reach 215 PB g<sup>-1</sup>, orders of magnitude higher physical densities than traditional devices<sup>1,5</sup>, and can have a half-life of thousands of years<sup>3,12</sup>.

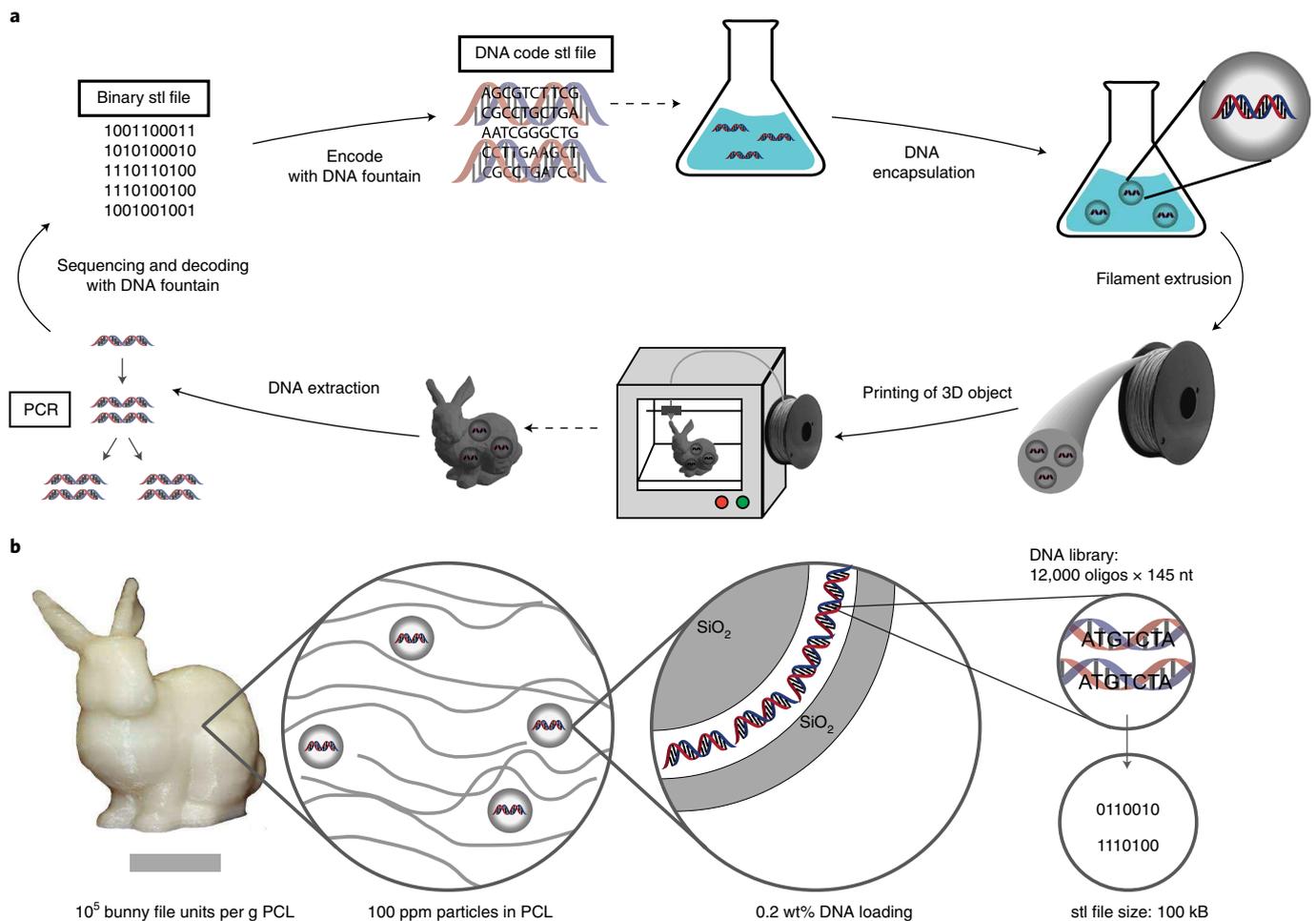
Beyond exceptional density and endurance, DNA storage is virtually the only storage architecture that can take any shape. This stands in stark contrast to traditional storage architectures such as tapes and hard-drives, where the actual shape of the device is often critical to its functionality. Previous studies of DNA storage have largely overlooked the virtue of the absence of shape constraints. However, this property allows the realization of new storage architectures beyond today's conventional designs.

Here, we explored a storage architecture, dubbed DoT, in which DNA molecules are fused to a functional material to create objects with immutable memory. The DoT architecture starts with encoding the data in DNA molecules in a manner that is robust to errors. Due to the low density of the DNA molecules in the embedding material, we expect that some DNA fragments will not be recovered from the object, creating fragment dropouts. To ameliorate

that, we used DNA Fountain<sup>5</sup> as our encoding scheme. This scheme provides high flexibility in setting virtually any redundancy level that can correct dropout errors and has been shown to perfectly retrieve data from minute quantities of material<sup>5</sup>. One challenge for the DoT architecture is that simply mixing the DNA with the functional material results in quickly degraded DNA due to hydrolysis stress and elevated temperatures during the preparation of the mixture (Supplementary Fig. 1). To mitigate this, the DoT architecture first encapsulates the DNA in silica nanoparticles<sup>8</sup>, resulting in silica particle-encapsulated DNA (SPED). The SPED shell sequesters the DNA molecules, prolonging their half-life and facilitating the mixing of DNA with the embedding material. Finally, we mix the SPED with the desired material and shape it using either three-dimensional (3D) printing or casting techniques.

To empirically test data storage using the DoT architecture, we created a 3D object that embeds DNA that encodes the blueprint for creating itself (Fig. 1a and Supplementary Video 1). This configuration is reminiscent of biological organisms, in which the instructions for making an object reside within the matter itself. As for the object, we selected the Stanford Bunny<sup>9</sup>, which is a common computer graphics 3D test model. First, we compressed the binary stereolithography (stl) file of the bunny from 100 kB to 45 kB. Next, we used DNA Fountain to encode the file in 12,000 DNA oligonucleotides (oligos), which is the maximal number of oligos produced by a single CustomArray chip (Supplementary Note 1). With this number of oligos compared with the file size, DNA Fountain encoding yields a redundancy level of 5.2×, meaning that we can tolerate a dropout of even 80% of the DNA oligos and still correctly decode the file. The length of the oligos was 145 nucleotides (nt), consisting of 104 nt of payload and 41 nt for PCR annealing sites (Supplementary Fig. 2). Next, we loaded the PCR-amplified oligos into SPED beads (Supplementary Fig. 3 and Supplementary Note 2) and embedded the beads in polycaprolactone (PCL). PCL is a biodegradable thermoplastic polyester that offers low melting temperature and high solubility in various organic solvents, making it an ideal material for blending and printing under mild conditions. To prepare 3D-printing filaments, we mixed the SPED capsules with dissolved PCL and extruded the mix into a 2.85 mm filament compatible with desktop 3D printers. Notably, the filament contained SPED beads in a concentration of 100 mg kg<sup>-1</sup> (100 ppm), which did not create any detectable changes to the mechanical properties, weight or color of the filament<sup>13</sup>. The DNA loading within SPED was 2 mg of DNA per gram of SPED beads, which translates to a DNA concentration of 0.2 mg kg<sup>-1</sup> (0.2 ppm) of the PCL filament (Fig. 1b and Supplementary Note 3), well below the concentration

<sup>1</sup>Functional Materials Laboratory, Department of Chemistry and Applied Biosciences, ETH Zurich, Zurich, Switzerland. <sup>2</sup>Complex Materials, Department of Materials, ETH Zurich, Zurich, Switzerland. <sup>3</sup>Erlich Lab LLC, Raanana, Israel. <sup>4</sup>These authors contributed equally: Yaniv Erlich, Robert N. Grass. \*e-mail: [erlichya@gmail.com](mailto:erlichya@gmail.com); [robert.grass@chem.ethz.ch](mailto:robert.grass@chem.ethz.ch)



**Fig. 1 | DoT workflow and proof-of-principle 3D printing of a Stanford Bunny. a**, The digital file is encoded into a DNA oligo library using DNA Fountain encoding. The synthesized library is encapsulated by a sol-gel synthesis method into small glass particles and blended into PCL, which is then extruded into a standard 3D-printing filament. The object, as defined by the initial digital file, is printed with the PCL filament that contains DNA. The DNA library can be extracted from any part of the printed object and amplified by PCR. By sequencing the DNA and decoding the DNA Fountain, the original .stl file can be retrieved to 3D-print new objects. **b**, Schematic of the hierarchical architecture of the 3D-printed Stanford Bunny, which contains all the instructions needed to reprint the object. Scale bar, 1 cm.

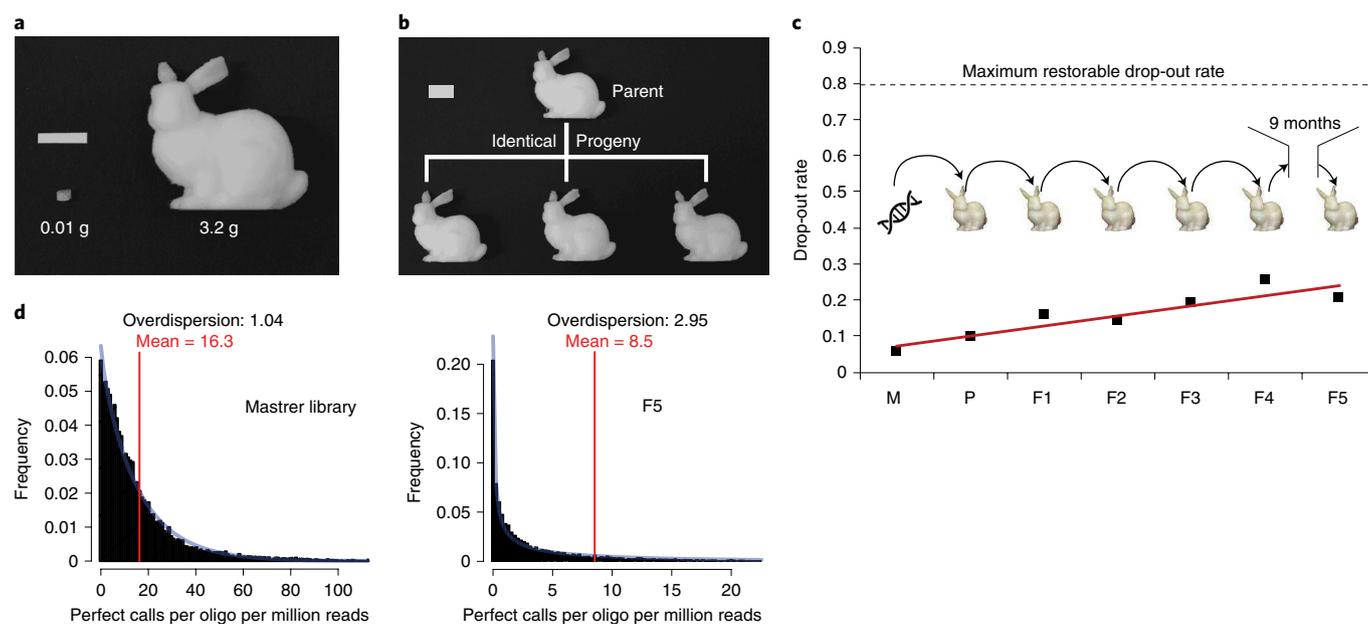
of DNA in biological organisms compared with their body weight (~1,000 ppm in *Escherichia coli*). Finally, we proceeded with 3D printing of the Stanford Bunny using the same file as stored in the DNA-containing PCL filament.

Our results show that the data can be perfectly and rapidly retrieved from the 3D object by consuming a minute quantity of material using a portable sequencer. We clipped ~10 mg of the printed PCL from the ear of the bunny, which is 0.3% of the total material of the bunny that weighed 3.2 g (Fig. 2a). Next, we released the SPED beads from the embedding PCL using tetrahydrofuran (THF), extracted the DNA from the SPED beads using buffered oxide etch (BOE) and purified the library using a standard PCR cleaning kit. The recovered DNA library weighed 25 pg in 50  $\mu$ l volume, corresponding to about 14,000 copies of the encoded file, including the 5.4 $\times$  redundancy (Supplementary Note 4). This entire process took 4 h end-to-end. We then amplified 1  $\mu$ l of the recovered DNA, equivalent to 1/50 of the recovered DNA, using ten PCR cycles and sequenced the library using an iSeq, a portable Illumina sequencer (Supplementary Note 5). This process took 17 h and yielded 1,046,118 reads. Finally, we processed the data using the DNA Fountain decoder and perfectly retrieved the stored .stl file despite missing 5.9% of the original oligos and being subject to

sequencing errors. This took a few minutes on a standard laptop (Supplementary Fig. 4).

Encouraged by these results, we conducted multi-round replication experiments using the DoT architecture. In the first replication round, we fused the PCR-amplified DNA from the parent Stanford Bunny (dubbed 'P' following the common notation in genetics) to a nascent PCL filament using the DoT procedure. Next, we created three offspring 3D structures (F1) using the PCL filament and the retrieved .stl file (Fig. 2b). Subsequently, we clipped about 10 mg from one of the F1 bunnies, extracted the DNA, sequenced the library with iSeq to retrieve the .stl file and 3D printed the next generation. We repeated the same procedure for a total of five generations (Fig. 2c), where in each generation a new PCL filament was created by fusing the PCR-amplified DNA molecules of the previous experiment. Finally, to demonstrate the ability to store the DNA long term, we sequenced the F4 9 months after its synthesis, sequenced the information and used it to generate a further product generation, F5.

We were able to perfectly retrieve the file from all five generations of progeny, including the retrieval of information 9 months after synthesis. In nearly every replication round, we saw an increase in the fraction of dropout molecules, from a level of 5.9% for the



**Fig. 2 | DoT objects can be replicated for multiple generations.** **a**, Just 10 mg of the parent 3D-printed object was sufficient to recover enough DNA with accurate building instructions (.stl file) to generate identical progeny. Scale bar, 1 cm. **b**, The initial bunny and three ‘children’ printed with the DNA instructions embedded in the parent DNA. The progeny also hold DNA instructions using the original DNA library. Scale bar, 1 cm. **c**, For every new generation, the file corruption rate increased, yet for the six experimentally implemented generations, it remained well below the file corruption rate that the Fountain error correction code allowed. **d**, Oligo coverage frequency distribution for the master library before embedding the DNA and the fifth descendant generation (F5). M, master library; P, parent.

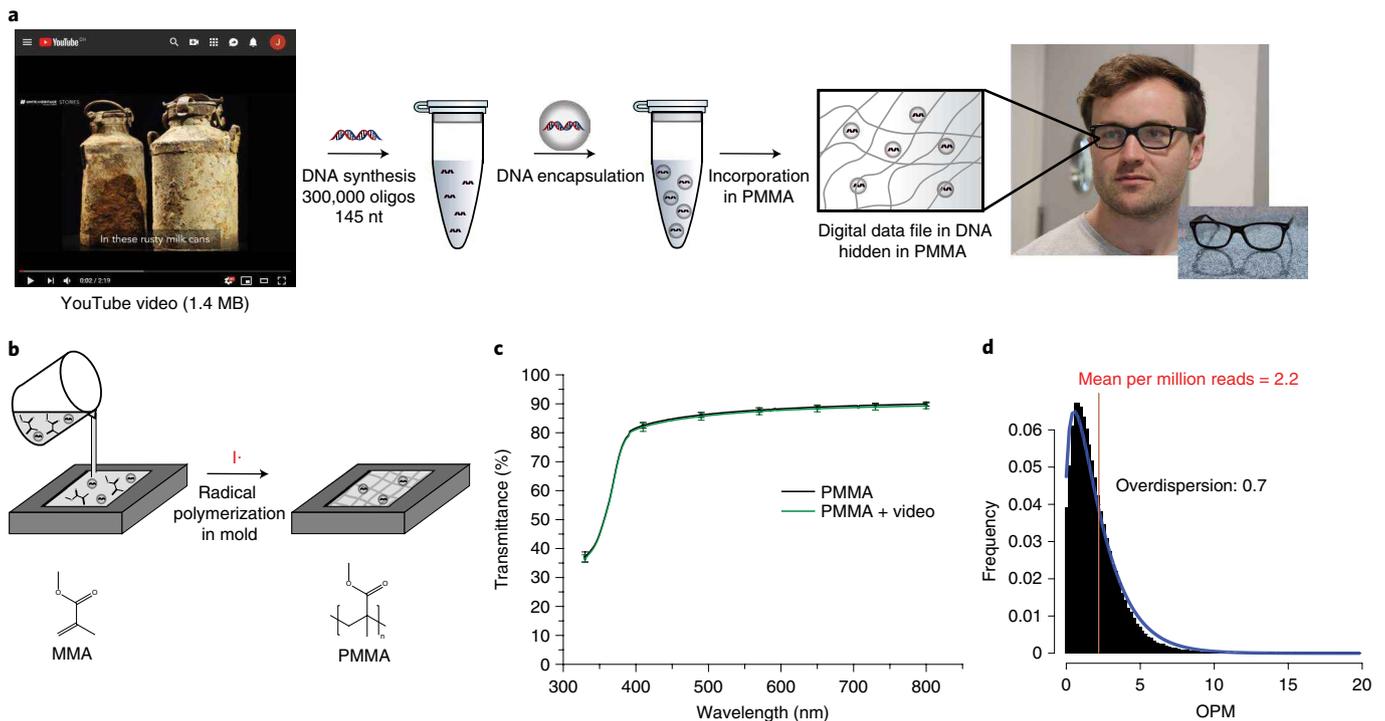
original library before inserting the DNA into the PCL filament to a level of over 20% in the final generations (Fig. 2c and Supplementary Table 1). To better understand the effect of replication on the library composition, we also modeled the number of correct sequence reads per oligo using a negative binomial distribution. This analysis showed an escalation in the negative binomial over-dispersion parameter from 1.0 in original library to 2.95 in F5 (Fig. 2d,e). This increase in over-dispersion means that the oligos are becoming less equally represented, which may be ascribed to the added thermal stress during polymer extrusion/printing and increasing numbers of PCR rounds<sup>5,14</sup>. In addition, we observed a significant increase (Spearman  $\rho = 0.24$ ,  $n = 600,000$ ,  $P < 10^{-9}$ ) in the rate of substitution errors as a function of the generation and a significant but modest decrease in the rate of deletions and increase in the rate of insertions (Spearman  $\rho_{\text{deletion}} = -0.05$ , Spearman  $\rho_{\text{insertion}} = 0.008$ ,  $n = 600,000$ ,  $P < 10^{-9}$ ) (Supplementary Fig. 5). The length distribution of the DNA oligos was quite similar and we did not observe a clear trend (Supplementary Fig. 6). However, we did find a slight underrepresentation in later generations (Pearson  $r = -0.0375$ ,  $n = 4.9 \times 10^6$ ,  $P < 10^{-9}$ ) of DNA oligos with higher numbers of guanine nucleotides (Supplementary Fig. 7), which are known to be sensitive for extrinsic stressors<sup>15</sup>. Despite these myriad imperfections, the DNA Fountain strategy retrieved the file correctly for all generations, showing the robustness of our strategy.

The economy of the DoT architecture is consistent with mass production of goods with memory at negligible per-unit costs. Each replication consumes only 0.3% from each bunny and yields sufficient DNA material to create 29 offspring bunnies. Therefore, even if we restrict the number of replications to five generations, it is theoretically possible to create at least  $(29/0.003)^5 = 8.44 \times 10^{19}$  bunnies without resynthesizing the DNA library. These results indicate that despite the relatively high upfront costs to synthesize the file for the first time (\$2,500 in our case), the per-unit costs of synthesis are likely to be negligible as mass production will require only one synthesis event. To better understand the total costs per unit, we

conducted an analysis that also includes the costs of the other molecular procedures, such as SPED preparation and PCR, using current laboratory costs and an industrial scale (Supplementary Note 6). Our results show that with the industrial scale costs, manufacturing over 1,000 units of the same item, the cost of 1 MB DoT is smaller than the cost of the polymer filament (Supplementary Fig. 8).

We envisage that the DoT architecture could be compatible with a wide range of embedding materials. For such compatibility, the DNA library should be miscible with the final material and survive the melting temperature of the embedding material and the material formation chemistry, which may include radical oxygen species (RoS) or other types of aggressive chemical reactions (for example, radical polymerization). Since DNA is not dissolvable in most organic solvents, or polymeric systems, the encapsulation of the DNA library into a dispersible carrier enables the mixing of the DNA with such materials. As silica particles have a long use history as a polymer filler<sup>16</sup>, the usage of SPEDs is advantageous<sup>17</sup>.

We sought to empirically test the robustness of our architecture for various polymer preparation techniques. We first used quantitative PCR (qPCR) to measure the amount of DNA retrieved from the SPED beads in PCL using various extrusion and 3D-printing temperatures (Supplementary Fig. 9 and Supplementary Note 7). We found that the amount of recovered DNA follows Arrhenius expression trends (Supplementary Note 8). For the extrusion process, for each 30 °C increase in temperature, there is a drop of approximately one order of magnitude in the recovered DNA. On the other hand, we observed only a relatively slow decline in the recovered DNA with the increase of the 3D-printing temperature, which is expected, as this process is much faster than extrusion and therefore exposes the library to high heat only for a short amount of time. The results also indicate that it is possible to use elevated temperatures for short cycle times during polymer processing, while suffering some losses of DNA<sup>13</sup>, which could be compensated by loading more DNA into the polymer preforms. Such temperatures will allow the use of DoT for selected polymer injection molding processes,



**Fig. 3 | A video is concealed in reading glasses.** **a**, The Oneg Shabbat video file is encoded into a DNA oligo library using DNA Fountain encoding. The DNA library is encapsulated in SPED and incorporated into transparent PMMA glasses. **b**, For the formation of the information-bearing polymer, the monomer methyl methacrylate (MMA) is premixed with SPED and radical initiator and poured into a mold. Polymerization at room temperature yields solid and transparent objects. **c**, The transparency of the DoT glasses was measured with UV-visible spectroscopy and compared with the transparency of PMMA glasses without SPED. Mean values and error bars (s.d.) from  $n=3$  independent experiments; three different spots on the same glass. **d**, Oligo coverage frequency distribution of the video file after extraction from the PMMA glasses.

and pre-impregnated carbon fiber composite autoclaving. Next, we also evaluated the robustness of our library to RoS<sup>8</sup> using the Fenton-like copper reaction<sup>18</sup> and bleach oxidation (Supplementary Note 9). We found that the encapsulation of the DNA is sufficient to protect against these extreme conditions without any detectable change in the amount of DNA. As such, this robustness to RoS and oxidation agents should enable DoT to be embedded in materials manufactured by radical polymerization, such as polystyrene, polyacrylates and acrylamides. Taken together, these results suggest that a wider range of materials can be used to embed the DNA library.

Next, we conducted a larger scale DoT architecture experiment in a distinct setting, using a different material, DNA synthesis method, material shaping technique and application (Fig. 3a). We started by encoding in DNA a 2 min video (1.4 MB) about the Oneg Shabbat archive, a collection of documents from the Warsaw Ghetto that were concealed during the Holocaust in milk cans; they are now part of the UNESCO (United Nations Educational, Scientific and Cultural Organization) Memory of the World Register and are a prime example of physical steganography<sup>19</sup>. Next, we used DNA Fountain to encode the file in 300,000 DNA oligos of 104 nt, yielding a redundancy level of 4.2×, and synthesized these oligos with Twist Biosciences. We then fused them to SPED beads and loaded them at a concentration of 100 ppm to methyl methacrylate, which was polymerized using dicumyl peroxide as initiator at room temperature to polymethyl methacrylate (PMMA, plexiglass), a widely used transparent plastic material (Fig. 3b). As we were interested in testing our approach to conceal data, we cast the plexiglass in the shape of a lens and mounted it on a glasses frame. The plexiglass with the concealed data had highly similar optical properties as ordinary plexiglass with a similar transmittance pattern between

400 nm and 700 nm (Fig. 3c). While DNA absorbs light of shorter wavelengths (230–260 nm), even ordinary plexiglass has virtually zero transmittance<sup>20</sup> in UV and therefore cannot expose the presence of DNA. The end result was a pair of ordinary-looking glasses with ordinary-transmittance lenses (Fig. 3a) that secretly stored a video message.

The DoT architecture could easily retrieve the video message concealed in the glasses. We collected a small piece of 10 mg from one lens and isolated the SPED beads by dissolving the plexiglass in acetone. We then harnessed the ability to spin-down the SPED particles by centrifugation to recover them from the dissolved polymer solution. We then followed our standard extraction procedure, yielding 0.4 ng of DNA, corresponding to about 6,000 copies of the encoded file. Sequencing the library with an iSeq resulted in 5.7 million paired-end reads (Supplementary Note 10). The average coverage was about 11 reads per correct oligo, the overdispersion parameter was 0.7 and 288,253 of the 300,000 oligos were detected in the library at least once (96.0%) (Fig. 3d). We applied the sequence reads to the DNA Fountain decoder and were able to correctly recover the library after processing the first 73,512 oligos. All lines of evidence suggest that larger file sizes can be stored with the DoT architecture, as future experiments can reduce either the logical redundancy from 4.2× to closer to the Shannon limit or the physical redundancy from 6,000 copies to a much smaller number.

We posit that the DoT architecture is useful in applications where blueprint level information is required to be accessible, either in a personalized item or a mass-manufactured object. For example, in the field of 3D medical or dental implants, each structure is unique and customized for the precise anatomical structure of the patient. As SPED beads are nontoxic, DoT allows storage of the design of

the implant together with other medical background information in the implant, offering a long-range back-up alternative to traditional electronic medical records, which are usually required to be kept only for 5–10 yr. Similarly, DoT might be applicable for building materials, pharmaceutical products and electronic components. All of these are areas that require product control information, which currently is not available at the point of usage. While for personalized items the cost of DNA synthesis may still currently be a limiting factor, for mass-manufactured objects the cost of synthesizing the DNA library becomes marginal as the number of devices increases (Supplementary Fig. 8). Having relevant information directly integrated via DoT into the product in an invisible, distributed and nonremovable manner will consequently allow tighter and better controlled product quality measures. Of course, this vision will require faster DoT retrieval protocols and ubiquitous DNA sequencers, which may be available in the future<sup>21</sup>.

Another area of interest for the DoT architecture is information steganography. Previous studies have suggested hiding data in DNA as a form of microdot<sup>22</sup>. The DoT architecture extends this concept and enables a wide range of everyday objects, from a keychain to a bottle lid, to be turned into concealed storage devices that can secretly carry data. An adversary trying to intercept the data will face multiple barriers: first, the SPED beads do not change the properties of the material, meaning that the adversary will have to test multiple objects to reveal the concealing object. Second, the DNA is sequestered in the SPED, so common DNA-sensing techniques such as UV light will not be helpful. Third, even if the library is recovered, the adversary will have to know the annealing sites to amplify the message by PCR. Taken together, these properties suggest that DoT is highly resistant to malicious interception, but more research is needed to test various scenarios to detect and hide the data.

Finally, a more futuristic application of DoT could be in the area of self-replicating devices. Our results suggest that one DoT library has sufficient replicating capacity to produce storage material for a virtually unlimited supply of objects. This can simplify blueprint transmission in self-replicating devices. More generally, the increasing data flow leads to increased dependency on delocalized data storage and large-scale data centers. The DoT approach provides a powerful alternative for localized and off-the-grid storage technology, capable of storing sensitive and precious data for multiple generations. It could also be used for integrating fabrication plans and conditions into complex products such as houses and art installations, as a way of preserving information for later generations.

### Online content

Any methods, additional references, Nature Research 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/s41587-019-0356-z>.

Received: 27 February 2019; Accepted: 11 November 2019;  
Published online: 09 December 2019

### References

1. Church, G. M., Gao, Y. & Kosuri, S. Next-generation digital information storage in DNA. *Science* **337**, 1628–1628 (2012).
2. Goldman, N. et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **494**, 77–80 (2013).
3. Grass, R. N., Heckel, R., Puddu, M., Paunescu, D. & Stark, W. J. Robust chemical preservation of digital information on DNA in silica with error-correcting codes. *Angew. Chem. Int. Ed.* **54**, 2552–2555 (2015).
4. Yazdi, S. M. H. T., Gabrys, R. & Milenkovic, O. Portable and error-free DNA-based data storage. *Sci. Rep.* **7**, 5011 (2017).
5. Erlich, Y. & Zielinski, D. DNA fountain enables a robust and efficient storage architecture. *Science* **355**, 950–954 (2017).
6. Organick, L. et al. Random access in large-scale DNA data storage. *Nat. Biotechnol.* **36**, 242–248 (2018).
7. Carmean, D. et al. DNA data storage and hybrid molecular–electronic computing. *Proc. IEEE* **107**, 63–72 (2019).
8. Paunescu, D., Puddu, M., Soellner, J. O. B., Stoessel, P. R. & Grass, R. N. Reversible DNA encapsulation in silica to produce ROS-resistant and heat-resistant synthetic DNA ‘fossils’. *Nat. Protoc.* **8**, 2440–2448 (2013).
9. Turk, G. & Levoy, M. Zippered polygon meshes from range images. in *Proc. 21st Annual Conference on Computer Graphics and Interactive Techniques* 311–318 (ACM, 1994).
10. Evans, R. F. L., Chantrell, R. W., Nowak, U., Lyberatos, A. & Richter, H. J. Thermally induced error: density limit for magnetic data storage. *Appl. Phys. Lett.* **100**, 102402 (2012).
11. Charap, S. H., Lu, P.-L. & He, Y. Thermal stability of recorded information at high densities. *IEEE Trans. Magn.* **33**, 978–983 (1997).
12. Zhirnov, V., Zadeegan, R. M., Sandhu, G. S., Church, G. M. & Hughes, W. L. Nucleic acid memory. *Nat. Mater.* **15**, 366–370 (2016).
13. Paunescu, D., Fuhrer, R. & Grass, R. N. Protection and deprotection of DNA—high-temperature stability of nucleic acid barcodes for polymer labeling. *Angew. Chem. Int. Ed.* **52**, 4269–4272 (2013).
14. Heckel, R., Mikutis, G. & Grass, R. N. A characterization of the DNA data storage channel. *Sci. Rep.* **9**, 9663 (2019).
15. Bonnet, J. et al. Chain and conformation stability of solid-state DNA: implications for room temperature storage. *Nucleic Acids Res.* **38**, 1531–1546 (2010).
16. Bandyopadhyay, A., de Sarkar, M. & Bhowmick, A. K. Polymer–filler interactions in sol–gel derived polymer/silica hybrid nanocomposites. *J. Polym. Sci. B* **43**, 2399–2412 (2005).
17. Paunescu, D., Stark, W. J. & Grass, R. N. Particles with an identity: tracking and tracing in commodity products. *Powder Technol.* **291**, 344–350 (2016).
18. Pham, A. N., Xing, G., Miller, C. J. & Waite, T. D. Fenton-like copper redox chemistry revisited: hydrogen peroxide and superoxide mediation of copper-catalyzed oxidant production. *J. Catal.* **301**, 54–64 (2013).
19. Schmeh, K. *Versteckte Botschaften: Die faszinierende Geschichte der Steganografie* (Heise Verlag, 2017).
20. Abramoff, B. & Covino, J. Transmittance and mechanical properties of PMMA-fumed silica composites. *J. Appl. Polym. Sci.* **46**, 1785–1791 (1992).
21. Erlich, Y. A vision for ubiquitous sequencing. *Genome Res.* **25**, 1411–1416 (2015).
22. Clelland, C. T., Risca, V. & Bancroft, C. Hiding messages in DNA microdots. *Nature* **399**, 533–534 (1999).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019

## Methods

**DNA encoding for the bunny experiment.** The original binary stl file was downloaded from <http://graphics.stanford.edu/pub/3Dscanrep/bunny.tar.gz> and was reduced in resolution from 22 MB to 100 kB by reducing the vertex count of the mesh with Blender 2.79b (Blender Foundation). Next, we compressed the file using gzip and encoded the data using DNA Fountain software with the following parameters: --max\_homopolymer 3 --gc 0.05 --rs 2 --size 20 --delta 0.001 --c\_dist 0.025. These parameters correspond to a maximal homopolymer of three nucleotides, a GC content of  $50 \pm 5\%$ , 2 B of Reed–Solomon error-correcting code, 20 B of data in each oligo and robust soliton distribution<sup>23</sup>.

**Library synthesis for the bunny experiment.** The DNA Fountain output was a library with 12,000 oligos of length 104 nt. We added the following PCR annealing sequences to flank each oligo: 5'-ACACGACGCTCTTCCGATCT-3' (20 nt) and 5'-AGATCGGAAGAGCACACGTCT-3' (21 nt), so the resulting oligos were of length 145 nt. We synthesized the oligos with CustomArray. From the synthesized library, 47 ng was PCR amplified using the following primers: 5'-ACACGACGCTCTTCCGATCT-3' and 5'-AGACGTGTGCTCTTCCGATCT-3; and KAPA SYBR FAST qPCR Master Mix (KAPA Biosystems) using the following cycling parameters: 95 °C for 15 s, 54 °C for 30 s, 72 °C for 30 s for 30 cycles. We then loaded 12.2 µg of DNA onto 1.75 mg of silica nanoparticles (Supplementary Fig. 10), functionalized with *N*-trimethoxysilylpropyl-*N,N,N*-trimethylammonium chloride and subsequently coated with an additional silica layer by adding tetraethoxysilane using a protocol described by Paunescu et al.<sup>8</sup>.

**Incorporation into PCL and 3D printing.** First, 2 mg of SPED was centrifuged at 15,000 r.p.m. for 3 min and the supernatant was removed. The particles were then washed and centrifuged twice in 2-ml Eppendorf tubes with 2 ml of acetone (>99.8%, ProLabo). The particles were then suspended in 2 ml of THF (VWR, >99.5%) and added to THF-dissolved PCL (MakerBot, 20 wt% PCL in THF). To distribute the particles in the polymer, the suspension was mixed with a high-speed mixer (SpeedMixer DAC 150 FVZ). After solvent evaporation under a fume hood, the solid polymer was extruded (Filastruder) at 60 °C to form a 2.85 mm filament. Objects were then printed with a commercial 3D printer (Ultimaker 2+, Ultimaker) modified with a geared direct drive extruder and a V6 hotend (E3D). The parts were printed onto a heated build-plate (45 °C) that was coated with a layer of adhesive spray (3DLac) at a speed of 20 mm s<sup>-1</sup> and forced convection was applied to cool the extrudate. The Stanford Bunny was printed with a nozzle temperature of 100 °C. An open-source fused deposition modeling slicer (Cura, Ultimaker) was used to generate the print paths.

**Sequencing and product cycle generation for the bunny experiments.** We clipped ~10 mg from the bunny ear using a scalpel. This piece was dissolved in 250 µl of THF (VWR) before adding 10 µl of BOE solution. The BOE was added to release the DNA molecules from the SPED beads. The DNA was then purified by PCR Purification Kit (Qiagen) and PCR using the same qPCR protocol as for the library synthesis. The different DNA samples from each generation were pooled and diluted to 1 nM and mixed with an additional 2% spike of PhiX added for base diversity. Finally, the DNA pool was sequenced with Illumina iSeq100 using the 150 nt paired-end protocol. A detailed procedure for sequencing preparation can be found in Supporting Note 5.

**Decoding for the bunny experiments.** For each generation, the paired-end sequencing data were stitched together using Pear<sup>24</sup>. We then removed sequence reads whose length was not equal to the original design length 104 nt using awk. Next, we continued with a decoding strategy similar to the original DNA Fountain scheme<sup>8</sup>. In short, reads were first collapsed and sorted based on their appearance from the most abundant to the least abundant, breaking ties lexicographically. Then, we decoded the file back using DNA Fountain, which employs a message-passing algorithm scheme. We compared the input SHA256 of the compressed stl file (314d00155dcb259683cfc02f2be9edbbf205a978f573569c79ab2d1b5e6bdc53) with the SHA256 of the retrieved file for each generation. In all cases, the SHA256 signatures were identical. Subsequently, the extracted file was printed as described in the previous section. Fitting of the negative binomial distribution was done in R using the `tdistr` command. For the analysis of the number of errors as a function of generation time, we scanned each sequence read against all possible 12,000 designed oligos. To this end, we started by splitting the 12,000 oligos into *k*-mers and associating each unique *k*-mer with a list of the original oligos. Then, we took each sequence read, split into 1-base pair overlapping *k*-mers and enumerating all possible oligos. In practice, we used *k*=9. The best-matching oligo had to have at least two more matching *k*-mers than the second best and at least four distinct *k*-mer matches. After finding the best-matching oligo, we performed an alignment between the read and the best-matching oligo using muscle v.3.8.31 (ref. <sup>25</sup>) to find insertions, deletions and substitutions. We used the following alignment parameters: `match = 1`, `mismatch = -1`, `gap_opening = -1`, `gap_extension = -0.5`. Statistical analyses were carried out with Python SciPy and *P* values smaller than  $10^{-9}$  were reported as  $P < 10^{-9}$  in the main text.

**Storage and retrieval of the Oneg Shabbat video.** The video published by UNESCO on YouTube was entitled 'The incredible and moving story of Oneg

Shabbat' (<https://www.youtube.com/watch?v=yqcLITbSXUg>). It was accessed 14 May 2019 and was reformatted (AAC, H.264) to a resolution of 240 pixels at 12 frames per second with 60 kilobase mono audio yielding an mp4 video file of 1.4 MB. The digital bitstream was encoded in DNA using DNA Fountain software with the same parameters as for the bunny experiment with two exceptions: first, we let the DNA Fountain software generate 300,000 DNA oligos. Second, we allowed a slightly larger GC content of  $50 \pm 20\%$ . After appending adapter information, 300,000 DNA sequences (length 145 nt) were ordered from Twist Bioscience. The DNA sample (6 µg) was dissolved in 120 µl of Milli-Q grade water to reach a concentration of 50 ng µl<sup>-1</sup> and stored at -20 °C.

**Incorporation of Oneg Shabbat library into a PMMA cast.** SPED (2 mg), which was manufactured from the DNA library using published methods<sup>8</sup>, was centrifuged at 15,000 r.p.m. for 3 min and the supernatant aqueous solution was removed. The particles were then washed and centrifuged twice in 2-ml Eppendorf tubes with 2 ml of acetone (>99.8%, ProLabo). The particles were then suspended in 2 ml of acetone by ultrasonication. For casting one pair of PMMA glasses, 0.66 mg of particles were added to 7 ml of methyl methacrylate (SKresin 1702, S u. K Hock). The suspension was mixed in a 150-ml polypropylene beaker with a high-speed mixer (SpeedMixer DAC 150 FVZ) for 30 s at 3,500 r.p.m. To initiate the radical polymerization, 70 mg of Peroxan (PM25S, S u. K Hock) was added before mixing again for 30 s at 3,500 r.p.m. with the high-speed mixer. The reaction mixture was left overnight in the beaker, resulting in a transparent PMMA disk. With a scalpel, the form of the glasses lens was cut out of the disk and inserted into the eyeglass frame. For comparing the visibility of PMMA with and without SPED, the same procedure was used to produce a PMMA glass without adding SPED.

**Decoding of the Oneg Shabbat video.** We used very similar steps to those used for decoding the bunny experiment to decode the Oneg Shabbat video. The only exception was that during the first round of decoding we noticed that a large number of oligos were sequenced from both their forward and reverse strands. The source of this issue remained unclear but might have stemmed from a parsing error in obtaining the reverse and forward strands from the sequencer. To overcome this issue and to find the correct orientation for each read, we reverse-complemented each sequence read and generated a tandem of two reads sorted lexicographically. For example, if the original read was TTTA then we created the reverse complement, which is AAAT, and then stitched the two reads together in a lexicographical order, resulting in the tandem read AAAT-TTTA. We carried out this process for all reads in the file after the stitching step. Then, we continued with the normal preprocessing steps of collapsing the tandem reads and sorting them based on their appearances. Next, we took each part of the tandem read and evaluated the Reed–Solomon code. We then moved the part that showed zero errors based on the Reed–Solomon code to the decoder. If both parts showed zero errors, it signified that we could not distinguish the orientation and so we discarded the entire tandem read. If none of the parts showed zero errors, we also discarded the entire tandem read. For example, if AAAT showed zero errors based on the Reed–Solomon code and TTTA showed one error, we progressed the AAAT to the decoder and discarded the TTTA. This process resulted in 287,344 unique oligos for decoding. Finally, we decoded the file back using the 287,344 oligos. We compared the input SHA256 of the video (ef819b4f0fe855becb53a3a39183d6b854f24c9c5963f474ec019bbdfef913c) with the SHA256 of the retrieved file, which was identical. We could also play the video using QuickTime player on a standard Mac laptop.

**Thermal and chemical stability assays.** DNA-infused PCL was extruded and 3D printed at temperatures ranging from 60 °C to 120 °C. Assessing long-term stability, dried pure and SPED DNA libraries were stored at 60 °C for 10 d and then re-suspended and quantified by qPCR. Chemical stability was tested by a radical treatment stability assay, where a combination of ascorbic acid, H<sub>2</sub>O<sub>2</sub> and CuCl<sub>2</sub> produced reactive oxidative species and tested the protective properties of SPED (see Supporting Notes 7 and 8 for details).

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

## Data availability

The input files are available at <https://figshare.com/s/283543df2b734b7988c6>. The sequencing files are available on the European Nucleotide Archive website, under accession code PRJEB35217.

## Code availability

DNA Fountain is available at <https://github.com/TeamErich/dna-fountain>.

## References

- Luby, M. LT codes. in *Symposium on Foundations of Computer Science 2002* 271–280 (IEEE, 2002).
- Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate illumina Paired-End read mergeR. *Bioinformatics* **30**, 614–620 (2013).

25. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113–113 (2004).

### Acknowledgements

The authors thank the Christen group at ETH for giving access to the iSeq 100 sequencing device. Y.E. thank B. Zaks for helpful discussions. Financial support by the ETH Zurich is kindly acknowledged.

### Author contributions

The idea for the study was by conceived by Y.E. and R.N.G. The study was conceptualized by R.N.G., Y.E., W.J.S. and K.M. Coding was done by Y.E. Experimental work was performed by J.K. 3D printing was carried out by S.G. Visualization was performed by J.K. and R.N.G. The first draft was written by Y.E. All authors contributed

to reviewing, editing and providing additional text for the manuscript. The order of the first two authors, who contributed equally, was determined lexicographically.

### Competing interests

Y.E. is an employee of MyHeritage, owner of Erlich Lab LLC, and holds patents in the area of DNA storage. ETH Zurich holds patents on DNA encapsulation. Y.E. and R.N.G. are listed as inventors on a patent application in the field of DNA of things.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41587-019-0356-z>.

**Correspondence and requests for materials** should be addressed to Y.E. or R.N.G.

**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 [Authors & Referees](#) 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

iSeq 100 Control Software 1.3 (illumina), Generate FASTQ Analysis Module 1.0.0 (illumina), LightCycler 480 Software Version 1.1 (Roche)

Data analysis

Blender 2.79b (Blender Foundation), DNA Fountain Software, PEAR (Zhang et al.), R statistical software

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/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

All data available from the from the corresponding author upon 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://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	<input type="text" value="We encoded one file into 12,000 oligos. The file was decoded in total 6 times from 6 generations of plastic bunnies."/>
Data exclusions	<input type="text" value="No data was excluded from this study"/>
Replication	<input type="text" value="Multiple generations of bunnies were tested (6) and from all generations, the bunny file could be retrieved."/>
Randomization	<input type="text" value="Randomization is not necessary in our work since we encoded only one file and the goal was to retrieve the same file again."/>
Blinding	<input type="text" value="The decoding software is not aware to the input file"/>

## 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                                 | 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               |
| <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               |

- | 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 |