Recording Temporal Signals with Minutes Resolution Using Enzymatic DNA Synthesis

Namita Bhan[10,*], Alec Callisto[1,*], Jonathan Strutz[1], Joshua Glaser[2], Reza Kalhor[3], Edward Boyden[4,5,6], George Church[3,4], Konrad Kording[2], Keith E.J. Tyo[1,‡]

[1]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA
[2]Center for Theoretical Neuroscience, Columbia University, New York, NY, USA
[3]Center for Epigenetics, Johns Hopkins School of Medicine, Baltimore, MD, USA
[4]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, USA.
[5]McGovern Institute, MIT, Cambridge, MA, USA.
[6]Howard Hughes Medical Institute, Department of Neurobiology, Harvard Medical School, Boston, MA, USA
[7]Department of Genetics, Harvard Medical School, Boston, MA, USA
[8]Wyss Institute for Biologically Inspired Engineering at Harvard University, Boston, MA, USA
[9]Department of Neuroscience, University of Pennsylvania, Philadelphia, PA, USA
[10]Mitolab, Cambridge, MA, USA

*These authors contributed equally to this work

‡Corresponding author: Keith E.J. Tyo
 Telephone: +1 847 868 0319
 Fax: +1 847 491 3728
 Email: k-tyo@northwestern.edu

Author Contributions
NB, AC and KEJT developed the concept. NB, AC, JS, JG, RK, EB, GC, KK, and KT designed

experiments. NB and AC performed all experiments. NB, AC, JS, and JG did primary analysis on

DNA sequencing data. All authors contributed to data analysis and preparation of the manuscript.

Competing Financial Interests
A utility patent application has been filed for some of the developments contained in this article by KT, NB, AC, JS, JG and KK.

Classification: Biological Sciences/Applied Biological Sciences

Keywords: DNA-based recording, enzymatic DNA synthesis, physiological signal recording, synthetic biology

This file includes:
Main text
Figures 1 to 3

49
50
51 **Abstract:**
52
53 Employing DNA as a high-density data storage medium has paved the way for next-generation
54 digital storage and biosensing technologies. However, the multipart architecture of current DNA-
55 based recording techniques renders them inherently slow and incapable of recording fluctuating
56 signals with sub-hour frequencies. To address this limitation, we developed a simplified system
57 employing a single enzyme, terminal deoxynucleotidyl transferase (TdT), to transduce
58 environmental signals into DNA. TdT adds nucleotides to the 3' ends of single-stranded DNA
59 (ssDNA) in a template-independent manner, selecting bases according to inherent preferences and
60 environmental conditions. By characterizing TdT nucleotide selectivity under different conditions,
61 we show that TdT can encode various physiologically relevant signals like Co^{2+}, Ca^{2+}, Zn^{2+}
62 concentrations and temperature changes *in vitro*. Further, by considering the average rate of
63 nucleotide incorporation, we show that the resulting ssDNA functions as a molecular ticker tape.
64 With this method we accurately encode a temporal record of fluctuations in Co^{2+} concentration to
65 within 1 minute over a 60-minute period. Finally, we engineer TdT to allosterically turn off in the
66 presence of physiologically relevant concentration of calcium. We use this engineered TdT in
67 concert with a reference TdT to develop a two-polymerase system capable of recording a single
68 step change in Ca^{2+} signal to within 1 minute over a 60-minute period. This work expands the
69 repertoire of DNA-based recording techniques by developing a novel DNA synthesis-based system
70 that can record temporal environmental signals into DNA with minutes resolution.

71
72

Main Text

Introduction

DNA is an attractive medium for both long-term data storage and for *in vitro* recording molecular events due to its high information density (1–3) and long-term stability (4). Molecular recording strategies write information into DNA by altering existing DNA sequences (5) or adding new sequences (6). For example, systems have been developed that use methods including differential CRISPR spacer acquisition (5, 7, 8), enzymatic synthesis (9–11) and others (1, 8, 12). By connecting these DNA modifications to a user input (in the case of data storage) or environmental signal of interest (in the case of recording events), these strategies enable *post hoc* recovery of signal dynamics over time by DNA sequencing. To date, molecular recording systems, both *in vitro* and *in vivo*, have connected signals of interest to DNA recordings with transcriptional control, using signal-responsive promoters to drive the expression of molecular writers, such as base-editors, CRISPR-associated systems, or gene-circuits, to record changes in signal. These approaches have yielded accurate recordings, however the time required to transduce signals through a recording apparatus that includes transcription, translation, and DNA modification fundamentally constrains the application of these methods to events on the timescales of hours or days. A recording mechanism that relies only on post-translational elements would be inherently faster as signal transduction would only require sub-second conformational shifts in one enzyme.

In an effort to speed up DNA recording processes, we hypothesized that a DNA polymerase (DNAp), which continually incorporates bases (13), could serve as a candidate for post-translational molecular encoding. In such a system, a DNAp functions as a "ticker-tape" recorder, transforming changes in environmental signals into changes in the composition of the DNA it synthesizes (14) (Fig. 1A). Much faster than transcription and translation, nucleotide incorporation occurs on a timescale on the order of milliseconds to seconds (15), potentially enabling orders of magnitude improvements in the temporal accuracy and resolution of molecular recording. However, prototypical DNAp's replicate the contents of an existing strand, which would prevent recording of new information. A DNAp that does not simply replicate DNA but rather creates a *de novo* sequence could allow for DNA recording.

Terminal deoxynucleotidyl transferase (TdT) is a DNAp that can randomly incorporate bases to the 3' of a DNA strand with biases toward particular bases (16, 17). Shifting the nucleotide bias of TdT could make it a prime candidate for post-translational control of DNA encoding. In fact, *in vitro* experiments have shown cations (e.g., Co^{2+}) can shift the bias of TdT (16, 18). In addition, DNA is synthesized in a sequential manner which provides an estimate of the time a particular base is added. We therefore hypothesized that the environment in which a TdT extends a DNA strand might be encoded by the average base-composition of the extended DNA. Put another way, by combining the change in nucleotide bias in the presence of cations and the time bases are added inferred from sequence, a molecular ticker tape may be possible (13, 14).

Here, we introduce TdT-based Untemplated Recording of Temporal Local Environmental Signals (TURTLES), a polymerase-based molecular recording system that achieves high time resolution *in vitro* by utilizing post-translational control to change the bases incorporated. First, we describe methods to characterize DNA sequences synthesized by TdT and show that cation concentrations can be encoded in populations of TdT-synthesized DNA using an approach that analyzes the average composition of several bases added at similar times on the same or parallel strands of DNA. We next developed an algorithm to accurately estimate the times of signal changes and show that temporal information can be accurately recovered by using estimates of DNA synthesis rates to map DNA sequences to real time. We also describe an expanded TURTLES system that uses an engineered, allosterically modulated TdT to expand the generalizability and tunability of the system. By inserting an exogenous sensing domain, we show that TURTLES can be adapted to arbitrary signals of interest. Taken together, these results establish the feasibility of DNA synthesis-based encoding systems and demonstrate minutes resolution recording of cationic environmental signals for enhanced applications in DNA data storage and DNA recording.

129
130 **Results:**

131 **TdT can encode environmental signals *in vitro* via changes in base selectivity:**

132 The cations present in the reaction environment of TdT affect the rate of incorporation for specific
133 nucleotides (16). For example, previous studies (18–20) and our experiments show that when only
134 one nucleotide is present, the incorporation rates of pyrimidines, dCTP and dTTP, increase in the
135 presence of Co^{2+} (Fig. S1).

136 We sought to examine if these Co^{2+}-dependent changes in kinetics also occurred in the presence
137 of all four nucleotides, dATP, dCTP, dGTP, and dTTP (hereafter referred to as A, C, G, and T). The
138 nucleotide composition of ssDNA extended by bovine TdT in a Cobalt-free reaction buffer or with
139 cobalt added was determined by next generation sequencing. In the presences of Co^{2+}, A
140 incorporation increased, while G, T and C incorporation decreased (Fig. 1B and Fig. S2). Notably,
141 the significant difference in the composition of DNA each condition effectively encodes information
142 about the environmental Co^{2+} concentration at the time of DNA synthesis.

143 Next, we were interested in understanding which conditions could be encoded by TURTLES. We
144 examined Ca^{2+}, Zn^{2+}, and temperature. Ca^{2+} signaling is biologically ubiquitous and functions in
145 neural firing (21), fertilization (22, 23), and neurodevelopment (24); Zn^{2+} is an important signal in
146 the development and differentiation of cells (25); and temperature is relevant in many situations.

147 Each signal altered both the particular dNTPs affected and the magnitude of the change in dNTP
148 selectivity. We were able to encode 20 μM Zn^{2+}, 1 mM Ca^{2+} and the temperature of 20 °C (Fig. 1B
149 and Fig. S2). Both cation addition and temperature change also altered the lengths of ssDNA
150 strands synthesized (Fig. S3-8). For each environmental condition tested, we observed significant
151 differences in the composition of TdT synthesized DNA. We conclude that input-dependent
152 changes in TdT nucleotide selectivity can encode environmental information into DNA. For further
153 analysis we chose to focus on Co^{2+} as the candidate cationic signal due to the large difference in
154 TdT selectivity.

155 **Recording a single step change in Co^{2+} concentration with minutes resolution:**

156 Having shown nucleotide selectivity changes in the presence of Co^{2+}, we attempted to identify the
157 time at which Co^{2+} was added to a TdT-catalyzed ssDNA synthesis reaction based on the changes
158 in the nucleotide composition of the synthesized ssDNA (Fig. 2A). During a 60-minute extension
159 reaction, we created step transitions in cobalt concentration by adding 0.25 mM Co^{2+} at 10, 20, and
160 45 minutes (hereafter referred to as a 0→1 input where '0' is Co^{2+}-free and '1' is with 0.25 mM
161 added Co^{2+}) (Fig. 2B top). For each reaction, we analyzed approximately 500,000 DNA strands by
162 deep sequencing and calculated the dNTP incorporation frequencies over all reads. Because the
163 change in dNTP selectivity is a compositional data type (i.e., all changes in base frequency sum to
164 0%), they are not independent and do not satisfy the independence assumption required for most
165 statistical tests. Therefore, to perform hypothesis testing, base composition was transformed into
166 Aitchison space, where the proportion of each base becomes independent of the other three. The
167 output signal for each reaction was calculated as the normalized distance in Aitchison space
168 between the 0 and 1 controls.

169 After normalizing each sequence by its own length, the Aitchison location along the extended
170 strands showed that later addition of Co^{2+} resulted in dNTP selectivity changes farther down the
171 extended strand (Fig. 2B center). To estimate the real time at which changes occurred, we
172 calculated the average location along the strand for all sequences in each condition at which a
173 distance halfway between the 0 and 1 control output signal was reached. To translate this location
174 into a particular time in the experiment, we calculated the average rate of dNTP addition in each
175 state (Fig. S9) and derived an equation that adjusted for the change in rate of DNA synthesis

176 between the 0 and 1 controls (Equation 5, Supplementary Methods 3). Using this information, we
177 estimated that the Co^{2+} additions were made at 11.9, 24.4 and 49.2 minutes (Fig. 2B bottom). We
178 were also able to estimate the time within 4 minutes of the unit input step function for the reverse,
179 a $1\rightarrow 0$ condition (Fig. S10 &11).

180 While we were able to accurately estimate the times of Co^{2+} addition ($0\rightarrow1$) and removal ($1\rightarrow0$),
181 simultaneously synthesizing ~500,000 strands of DNA will be infeasible for certain applications. To
182 determine the number of strands needed for reasonable statistical certainty, we randomly sampled
183 smaller groups of strands from the experiment and evaluated our ability to predict when Co^{2+} was
184 added (Fig. 2C top and bottom). With about 1,000 strands, we still estimated the time of Co^{2+}
185 additions to within 2 minutes of actual input times (Fig 2C,). Thus, TURTLES-1 recordings are
186 robust and encode high resolution temporal information even with a limited number of ssDNA
187 substrates.

Recording multiple fluctuations in Co^{2+} concentration onto DNA with minutes resolution:

189 In contrast to current DNA-based recorders, which rely on time-integrated recording methods (i.e.,
190 accumulation of mutations) or slow signal transducing steps, DNA synthesis-based approaches
191 can record the dynamics of multiple fluctuations in real time. While accumulation can tell what
192 fraction of the time a signal was present in a time period, the ability to record multiple temporal
193 changes would enable new levels of insight into dynamical processes such as physiologically
194 signaling, which are poorly captured by time-integrated recording methods.

195 We used TURTLES to record a $0\rightarrow1\rightarrow0$ input cobalt signal. The 0 condition was maintained first
196 20 minutes, 1 for the next 20 minutes, and 0 for the last 20 minutes of the extension reaction (Fig.
197 2D top). Using the same methods as the single step transition, we calculated the output signal (Fig.
198 2D center). To account for the additional complexity of multiple fluctuations, we used an algorithm
199 previously developed by Glaser *et al.*(13) (see Materials and Methods for details) to binarize the
200 value of the output signal every 0.1 min. We were able to accurately reconstruct the input $0\rightarrow1\rightarrow0$
201 signal, estimating transitions between the 0 and 1 signals occurring at 21 and 41 minutes based
202 on sequencing data (Fig. 2D bottom).

203 Based on the measured experimental parameters, we used *in silico* simulations to estimate the
204 performance of TURTLES in more complex recording environments. We investigated how rapidly
205 signals could change and still be detected and how many consecutive condition changes could be
206 recorded accurately. By varying the length of time of each input condition (0 or 1) from 1 to 20
207 minutes (Fig. S12A), we estimated that TURTLES can record 6 consecutive signal changes with 1
208 minute between each with >75% accuracy from $\geq$ 2000 strands of 100bp ssDNA synthesized
209 (>90% accuracy with $\geq$ 60,000 strands of ssDNA of 50 bp length each) (Fig. S12B). By keeping the
210 duration of each input condition (0 or 1) constant at 10 minutes (Fig S12C) and varying the total
211 number of condition changes, we estimated that TURTLES would be capable of recording 10
212 sequential input signal changes with >80% accuracy (Fig. S12D). We thus show that TURTLES
213 has unprecedented temporal precision and can robustly decode signals across a range of
214 frequencies.

215 Having successfully encoded three sequential environmental states with TURTLES-1, we
216 considered the feasibility of TURTLES-based digital data encoding. In this context, the data density
217 of TURTLES-1 recordings is determined by the number of inputs that yield distinguishable outputs.
218 By altering the ratios of nucleotides provided as substrates to a reaction in sequential steps, the
219 composition of DNA synthesized by TdT could be controlled with high resolution. Assuming 5%
220 changes in nucleotide composition can be distinguished, up to 11 bits of data can be encoded in
221 each step for 33 bits in a three-step recording. The bits encoded in each step would increase to 17
222 bits if 1% changes can be distinguished (Fig S13).

TdT can be engineered to allosterically respond to and encode environmental signals:

224 Unlike Co^{2+} and Zn^{2+}, we observed that Ca^{2+} only modestly altered the dNTP selectivity of TdT,
225 precluding temporal recordings of Ca^{2+} concentration. To show that TURTLES could be expanded
226 to signals to which TdT was unresponsive or weakly responsive, we attempted to engineer a TdT
227 to allosterically respond to Ca^{2+}. The structural determinants of base selectivity in TdT are poorly
228 understood, which ruled out directly increasing the dynamic range of Ca^{2+}-responsive dNTP
229 selectivity changes (26). Accordingly, we conceived a modular recording system based on two
230 distinct TdT species with different inherent dNTP selectivity.

231 The two-TdT system, TURTLES-2, uses a reference TdT whose catalytic rate is unaffected by
232 inputs and a sensor TdT that is allosterically activated or deactivated in response to input signals.
233 By choosing a pair of sensor and reference TdTs with distinct nucleotide selectivity, TURTLES-2
234 encodes environmental signals into changes in DNA composition based on the differential activity
235 of the two TdTs (figure 3A). As the sensing and recording functions of the system are distributed
236 between two TdT variants, TURTLES-2 is more accessible to tuning and engineering efforts.

237 We employed the natural calcium sensing protein, Calmodulin (CaM), and a cognate binding
238 peptide, M13, to generate a TdT with allosterically modulated activity. The calcium-dependent
239 interaction between CaM and M13 has been previously utilized to engineer allosteric calcium
240 biosensors (27, 28) and as a platform for generalizable ligand biosensors (29, 30).

241 Here, we generated variants with M13 fused to one of 4 sites in mTdT that were predicted to
242 minimize the structural disruption of inserting the M13 sequence using SCHEMA-RASPP(31) (Fig
243 S14, Table S2). After initial activity screening (Fig S15) of the variants we observed that one variant,
244 mTdT(M13-388), retained polymerase activity. CaM was subsequently fused to the N-terminus of
245 mTdT(M13-388) via linker. Primer extension reactions showed that the resulting CaM-mTdT(m13-
246 388) variant was active in calcium-free conditions and inactive in calcium-added conditions (Fig
247 S16). To confirm that the calcium-dependent interaction between CaM and M13 was responsible
248 for the observed activity modulation, we mutated four essential Ca^{2+}-binding residues in CaM,
249 which ablated the calcium-sensitivity of CaM-mTdT(M13-388) (Fig S16).

250 Depending on the application, sensor polymerases with different calcium affinities may be useful
251 to selectively record Ca^{2+} fluctuations exceeding threshold concentrations. We anticipated that the
252 modular design of CaM-mTdT(M13-388) would allow the properties of the fusion to be rationally
253 modified with CaM variants with known differences in Ca^{2+} affinity. Polymerase activity was
254 determined by the length distributions of primer extensions, CaM-mTdT(M13-388) variants
255 containing the CaM mutants D96V, D130G, and D142L which reduce the calcium affinity of CaM
256 (32). We observed that all variants exhibited greater activity than the unmodified CaM-mTdT(m13-
257 388) in the presence of low concentrations of calcium (Fig. S17). Moreover, the increase in activity
258 correlated with the reported effective Ca^{2+} K_D of the variants, demonstrating that the calcium-
259 sensitivity of CaM-mTdT(M13-388) can be rationally tuned.

260 Next, we tested if the TURTLES-2 system could the Ca^{2+} state into DNA. CaM-mTdT(M13-388)
261 was purified and NGS analysis of extension reactions performed with the polymerase confirmed
262 the calcium-sensitive phenotype (Fig. S18 & 19). We characterized CaM-mTdT(M13-388) in the
263 context of the TURTLES-2 recording system by performing extensions with a mixture of purified
264 bovine TdT and CaM-mTdT(M13-388) in calcium-free and calcium-added conditions. The two-
265 polymerase system exhibited a significantly altered nucleotide selectivity in the calcium-added
266 conditions (Fig S20). As expected, in the calcium-free condition the overall base incorporation
267 preference was approximately the average of the observed preferences of bovine TdT and CaM-
268 mTdT(M13-388) whereas the overall base preference in the calcium added condition was nearly
269 identical to that of bovine TdT (Fig. S21). We conclude that the differential overall base selectivity
270 of the TURTLES-2 system is capable of encoding the environmental calcium state into DNA.
271

272 **Recording a single step change in Ca^{2+} concentration with minutes resolution with two TdT**
273 **system:**

274 We next investigated if the differential calcium response of TURTLES-2 could be used to infer the
275 time at which calcium concentrations changed in an extension reaction. During a 60-minute
276 extension we tested both 0→1 and 1→0 step transitions at 30 minutes, where '0' is calcium-free
277 and '1' is calcium-added (Fig. 3B top, Fig. 3C top, Supplementary text 6). Using a variation of the
278 model developed in Glaser et al. (13) we inferred transition times of 30 minutes for the 1→0
279 transition (Fig. 3B bottom) and 22 minutes for the 0→1 transition (Fig. 3C bottom). The decreased
280 accuracy of the estimated time for the 0→1 transition did not correspond to an increase in
281 measurement variability, suggesting that the offset is due to systematic factors. Time estimations
282 for 0→1 transitions in TURTLES-2 would likely improve with more sophisticated decoding
283 algorithms and deeper characterization of the transition behavior of CaM-mTdT(M13-388). We
284 conclude that TURTLES-2 enables high resolution temporal encoding of calcium signals.
285
286 **Discussion**

287 In this study, we demonstrated two DNA synthesis-based recording systems that encode and
288 record the temporal dynamics of fluctuating environmental signals with minutes accuracy. By
289 coupling sensing and writing functions, TURTLES simplifies the recording apparatus to post-
290 translational system. This gives TURTLES distinct advantages over the temporal constraints of
291 existing tools, enabling heretofore unprecedented temporal accuracy and resolution. While
292 TURTLES-1 can record several physiologically relevant signals, TURTLES-2 lends tunability to the
293 recording system with simple rational engineering. Given the uncomplicated and genetically-
294 encodable design of TURTLES systems, we anticipate that they can be readily adapted for both *in*
295 *vitro* and *in vivo* applications.

296 While many DNA-based biosensors have been deployed for studying physiological signals of
297 interest (33–37), the scalability and spatial resolution of biosensors is intrinsically limiting in some
298 applications (38). By leveraging *post hoc* recovery of biological data, optimized TURTLES systems
299 may be capable of enabling otherwise inaccessible high-resolution spatial and temporal recordings
300 of physiological signaling molecules that fluctuate on the timescale of 10^1-10^3 minutes. Such signals
301 include slow calcium signaling that occurs in neurons (13, 14, 38, 39) and vertebrate development
302 (22). Additional optimization of the TURTLES system will be required to enable the spatiotemporal
303 resolution for characterizing systems with shorter timescales. Beyond biological applications, there
304 has been a sustained interest in biosensors for testing environmental parameters such as water
305 quality. For longer term tracking of contaminating metal ions in water, one could use TURTLES to
306 track cobalt concentration over time (40, 41). In concert with microfluidic reaction control, TURTLES
307 can also serve as a competitive platform for enzymatic DNA synthesis for data archiving (9–11),
308 which is an appealing alternative to phosphoramidite methods due to the low cost and reduced
309 environmental impact (42). Although TURTLES does not control the specific base added, as few
310 as 5-10 bases could encode a bit. While this information density is lower than that of base-specific
311 DNA synthesis but does not require specialized substrates or complex reaction cycling, which may
312 yield cost savings with respect to base-specified methods.

313 Going forward, more sophisticated computational methods will improve the recording accuracy of
314 TURTLES. In this study we utilized simple, intuitive models of TdT activity to transform sequence
315 data into temporal information. By incorporating kinetic models of TdT activity or machine learning
316 to classify signal changes along individual DNA strands, the accuracy of temporal estimations could
317 likely be increased. These methods would also improve the robustness of TURTLES recordings by
318 reducing the required sequencing depth from thousands to hundreds of reads. In this work, both
319 the inputs and outputs for TURTLES were binarized, however the underlying principle can be
320 extended to record continuously varying analog signals with improved decoding algorithms.

321 The quality of TURTLES recordings may also be improved by engineering the properties of TdTs.
322 In particular, the sequencing depth required to accurately decode recordings can be reduced by
323 increasing the magnitude of changes in nucleotide selectivity in response to inputs. Likewise,
324 reference TdTs that have a more distinct nucleotide selectivity from the CaM-mTdT(M13-388)

325 sensor TdT can be engineered or identified among natural TdT diversity. Improvements to the
326 temporal resolution of TURTLES systems can be accomplished by enhancing the nucleotide
327 incorporation rate of TdT. In TURTLES-2, structural optimization of CaM-mTdT(M13-388) may
328 improve temporal resolution by optimizing the kinetics of the CaM-M13 interaction. Notably,
329 fluorescent biosensors based on CaM-M13 interactions can report calcium spikes on the order of
330 seconds(43, 44), suggesting that calcium sensing will not be limiting with respect to temporal
331 resolution in an optimized system. The functionality of TURTLES-2 may be further expanded by
332 employing generalizable sensors based on the CaM-M13 interaction (29), or by probing TdT with
333 sensing domains other than calmodulin such that new signals of interest can be encoded or
334 recorded. In all, we have demonstrated a new methodology for recording dynamic, environmental
335 information into DNA that relies only on allosteric regulation, enabling minutes resolution.

336 **Materials and Methods**
337
338 **Enzymes and ssDNA substrate:**
339
340 Terminal deoxynucleotidyl polymerase, T4 RNA ligase I, Phusion High-Fidelity PCR Master Mix
341 with HF Buffer were purchased through New England Biolabs (NEB). ssDNA substrates used for
342 extension reactions were ordered from Integrated DNA Technologies (IDT) with standard desalting.
343 dNTPs were obtained from Bioline.
344
345 **CaM fusion design and screening**
346 Four fusion proteins were designed that consisted of CaM fused to the N terminus of mTdT by a
347 (GGGGS)$_4$ linker and M13 inserted immediately following the fusion residue (see below) with
348 flanking GS linkers. Fusion sites were selected from crossover sites identified with the
349 SCHEMA/RASPP algorithm based on which sites were in catalytically essential regions and would
350 be sterically available to CaM. SCHEMA crossover sites were calculated according to previously
351 described protocols(31). Sets of crossover points were calculated for 3, 4, 5, 6, and 7 total
352 crossovers. Calculations were performed with the following parameters: minimum fragment length
353 = 4, bin width = 1, parent sequence = NP_001036693.1, parent structure = PDB 4i27 (all ligands,
354 metals, and waters removed), homology sequences = NP_803461.1, AAH12920.1,
355 NP_001012479.1, XP_021064401.1, XP_020136193.1. All sequences were trimmed to only
356 include residues crystallized in the parent structure. Fusion sites were selected from crossover
357 points that were in the DNA-binding region of mTdT (residues 282, 284, 287) or in Loop 1, a
358 catalytically essential structure (residue 388). M13 fusions were screened for activity without N-
359 terminal CaM to validate that the fusion was tolerated.
360
361 **Cloning CaM-TdT(M13) variants**
362 Molecular cloning of DNA constructs was completed under a contract research agreement with the
363 lab of Dr. J. Andrew Jones at Miami University – Oxford, OH. The pET28a-M-CaM-cTdTS-M13-
364 XXX (282, 284, 287, and 388) variants were constructed using a two-part Gibson Assembly
365 method. The approximately 75bp M13 fragment was amplified from linear double stranded DNA
366 template (gBlock-CaM-Linker-M13, IDT) using Accuzyme DNA polymerase (Bioline) using DNA
367 primers P21 – P28 listed in Table 1. The amplicon was then purified using a Cycle Pure Kit (Omega
368 Biotek). The vector backbone fragment was amplified from pET28a-M-CaM-cTdTS plasmid DNA
369 constructed above using *PfuUltra* II Hotstart PCR Master Mix using DNA primers P29 – P36 listed
370 in Table 1. The PCR product was then digested with *Dpn*I to remove DNA template. The
371 approximately 8100bp amplicon was purified using a gel extraction kit (Omega Biotek). DNA
372 concentration of both linear fragments was measured using the Take3 plate coupled with a Biotek
373 Cytation 5 plate reader. Corresponding backbone and M13 fragments were then assembled using
374 the repliQa™ HiFi Assembly Kit (Quanta bio), transformed into chemically competent DH5α, and
375 selected on LB-Kanamycin (50 μg/mL) plates. Individual colonies were then screened via restriction
376 digestion and verified using Sanger sequencing (CBFG – Miami University) with primers S1 – S8,
377 Table 1.
378
379 **CaM-TdT(M13-388) expression and purification**

380 Purification optimizations determined that N-terminal MBP was unnecessary for expression and
381 purification and was not included in the final expression construct. The expression construct
382 (pET28a-CaM-mTdT(m13-388)) was transformed into chemically competent NEB T7Express cells,
383 plated on kanamycin selective plates, and incubated at 37°C. The following day, a single colony
384 was selected and inoculated into 5mL of kanamycin supplemented LB. the culture was incubated
385 for 20 hours at 37°C. Four flasks of 120mL kanamycin supplemented LB were inoculated 1:400
386 (v/v) with the overnight culture. The cultures were incubated with shaking at 250 RPM. Once the
387 OD600 was between 0.5 and 0.6, the cultures were cooled to room temperature and induced with
388 1mM IPTG. Following induction, the cultures were incubated for 18 hours at 15°C. The cells were
389 pelleted at 4°C and the supernatant was discarded. The decanted cell pellets were stored at -80°C.
390
391 The cell pellets were thawed on ice. Lysis and affinity chromatography were performed using the
392 Takara Bio HisTALON gravity column purification kit; all steps were performed according to the
393 manufacturer's native protein extraction protocol. Note that the cell pellets were treated with
394 optional DnaseI and lysozyme during lysis. 1mL of Takara Bio TALON metal affinity resin was used
395 for affinity chromatography. All binding and wash steps were performed on ice with shaking at 250
396 RPM. 15 bed volumes of wash buffer were used for all washes. CaM-mTdT(M13-388) was eluted
397 from the resin in 10, 500uL fractions. Each fraction was analyzed by SDS-PAGE, and the total
398 protein concentration in each fraction was measured by absorbance at 280nm. The first five elution
399 fractions, which contained the majority of eluted protein, were pooled.
400
401 The pooled fractions were diluted in binding buffer (20 mM Tris-HCl, pH 8.3) and further purified by
402 anion exchange chromatography using a Cytiva HiTrap Q HP 5mL column and a 40 CV gradient
403 from 0 mM to 1 mM NaCl in binding buffer with a GE Healthcare AKTAxpress FPLC. The protein
404 eluted in two fractions.
405
406 Both elutions were buffer exchanged by dialysis into a storage buffer consisting of 200mM KH_2PO_4
407 and 100mM NaCl at pH 6.5 and concentrated using Vivaspin 20 columns to a final concentration
408 of 0.37 mg/mL for the first fraction and 0.98 mg/mL for the second fraction. The fractions were
409 aliquoted, and flash frozen on dry ice for storage at -80°C. Notably, PAGE analysis showed that
410 the second elution contained a product at 25kDa in addition to the expected fusion protein at
411 approximately 70kDa. Both CaM-mTdT(M13-388) elutions recapitulated the calcium-sensitive
412 phenotype and exhibited similar nucleotide selectivity (Fig S19-20). As significantly more protein
413 was recovered in the second elution it was used for all subsequent experiments.
414
415 **Cell free protein expression and primer extension assay**
416 For initial activity screening of fusion variants and CaM-mTdT(M13-388) characterization, proteins
417 were expressed in cell-free reactions. Variants were expressed using NEB PURExpress in 25 µL
418 reactions containing 40% (v/v) PURExpress Solution A, 30% (v/v) PURExpress Solution B, 1.6
419 U/µL NEB Rnase I, 10 ng/µL expression vector DNA, and dH_2O to volume. The expression
420 reactions were incubated for 4 hours at 30°C.
421
422 Primer extension reactions were prepared on ice. Primer extensions were performed in 25 µL
423 reactions containing 1X NEB TdT Reaction buffer, 0.8 µM single-stranded, FAM-labelled substrate
424 DNA FAM_NB (Table S2), 1mM dNTPs, polymerase, and dH_2O to 25 µL. For variants expressed
425 in PUREexpress, 2.5 µL of the expression reaction was used immediately after expression; 20U
426 (approximately 0.2 µg) of the NEB bovine TdTwas used for positive control reactions; approximately
427 0.5 µg, of purified CaM-mTdT(M13-388) was used for activity validation reactions after purification.
428 For calcium-added conditions, $CaCl_2$ was added to the reactions to a final concentration of 1mM.
429 Extension reactions were incubated for 2 hours at 37°C
430
431 Completed extensions were analyzed by urea-PAGE. 8 µL of each completed primer extension
432 reaction was combined with 12 µL of BioRad 2x TBE urea sample buffer and boiled for 10 minutes.
433 The boiled samples were loaded onto a 10% polyacrylamide TBE urea gel (bioRad 4566036), and
434 200V was applied to the gel for 40 minutes. The gels were imaged on a GE Healthcare Typhoon

435   9400 laser scanner using a 200um pixel size and λ_{ex}=488nm, λ_{em}=520nm BP40. Imaging gain was
436   adjusted for each experiment to avoid saturation.
437
438   **Extension reaction for calculating effect of Co^{2+}, Ca^{2+}, Zn^{2+}, and temperature on overall dNTP**
439   **preference of TdT:**
440
441   Each extension reaction consisted of a final concentration of 10 µM ssDNA substrate CS1 (Table
442   S2), 1 mM dNTP mix (each dNTP at 1 mM final concentration), 1.4x NEB TdT reaction buffer, and
443   10 units of TdT to a final volume of 50 µL. When testing the effect of cations, $CoCl_2$ was added at
444   a final concentration of 0.25 mM, $CaCl_2$ at 2 mM, or $Zn(Ac)_2$ at 20 µM. It is important to note that
445   reaction initiation was done by adding TdT to the ssDNA substrate mix (ssDNA substrate mix
446   consisted of the ssDNA substrate, dNTPs and the cation). Prior to reaction initiation, the ssDNA
447   substrate mix and TdT were stored in separate PCR strip tubes at 0 °C (on ice). The reaction was
448   run for 1 hour at 37 °C in a Bio-Rad PCR block. When testing the effect of temperature, the same
449   reaction mix was run on a Bio-Rad PCR block set at tested temperatures for 1 hour. Reactions
450   were stopped by freezing at -20 °C. For initial testing, reactions were analyzed by urea-PAGE 2 µL
451   of the reaction was mixed with 12 µL of TBE-Urea (Bio-Rad) loading dye and boiled for 10 minutes
452   at 100 °C. All of the diluted extension reaction was then loaded onto 30 µL, 10 well 10% TBE-Urea
453   Gel (Bio-Rad) and run for 40 minutes at 200 V. Immediately after the run was over, the gel was
454   stained with Sybr Gold for 15 minutes and imaged on an ImageQuant BioRad.
455
456   **TURTLES 0→1 extension reactions:**
457
458   Mg^{2+} only for 1 hour (signal 0) and Mg^{2+}+Co^{2+} for 1 hour (signal 1) were set up as regular extension
459   reaction mentioned above. The 0→1 reactions where the signal changed from 0 to 1 at various
460   times during the 1 hour extension were run starting at a total volume of 45 µL with Mg^{2+} only. 5 µL
461   2.5 mM $CoCl_2$ was added at the time we wanted the signal to change from 0 to 1. Reactions were
462   all run for a total of 1 hour in triplicates. Fresh signal 0 and signal 1 controls were run with each set-
463   up.
464
465   **TURTLES-2 controls and 0→1 and 1→0 extension reactions:**
466   TURTLES-2 extension reactions contained 1X NEB TdT reaction buffer, 0.1 µM ssDNA substrate
467   CS1_5N (Table S2), 1 mM dNTP mixture, and 2.5 µL polymerase mixture. The polymerase mixture
468   contained CaM-mTdT(M13-388) at a concentration of 0.45 mg/mL and NEB TdT at a concentration
469   of 0.002 mg/mL (0.4U). Calcium-free reactions included EGTA at a final concentration of 50 µM.
470   The high calcium control for 0→1 reactions was supplemented with $CaCl_2$ and EGTA to final
471   concentrations of 100 µM and 50 µM, respectively. The high calcium control for 1→0 reactions was
472   neither supplemented with EGTA nor $CaCl_2$ (supplementary text 6). Reactions were brought to a
473   final volume of 25 µL with nuclease-free water. Reactions were assembled on ice and initiated by
474   adding TdT to the substrate mixture. Reactions were incubated for 1 hour at 37 °C in a Bio-Rad
475   PCR block and terminated by heating reactions to 80 °C for 10 minutes.
476
477   Signal transitions were performed by 1uL additions at 30 minutes. for 1→0 reactions, the addition
478   contained 1X NEB TdT buffer and 1.3 mM EGTA (50 mM EGTA in final reaction post- addition). for
479   1→0 reactions, the addition contained 1X NEB TdT buffer and 2.6 mM $CaCl_2$ (100 µM $CaCl_2$ in final
480   reaction post- addition).
481
482   **Extension reactions for 0→1→0 set-up:**
483
484   Mg^{2+} only for 1 hour (signal 0) and Mg^{2+}+Co^{2+} for 1 hour (signal 1) were set up as regular extension
485   reaction mentioned above. The 0→1→0 reactions where the signal changed from 0 to 1 at 20
486   minutes and back to 0 at 40 minutes were run starting at a total volume of 45 µL with Mg^{2+} only. 5
487   µL 2.5 mM $CoCl_2$ was added at the time we wanted the signal to change from 0→1. For changing
488   the signal from 1→0, since the ssDNA was suspended in reaction buffer for these set-ups, we used
489   a ssDNA clean up kit (methods mentioned below) to remove the reaction buffer, TdT, cation and
490   dNTPs from each reaction. All of the ssDNA collected from the ssDNA clean up kit (20 µL) was

491 then prepared for the last part of the extension reaction. Collected ssDNA was mixed with a dNTP
492 mix at a final concentration of 1 mM (each dNTP at 1 mM final concentration), 1.4x TdT reaction
493 buffer and 10 units of TdT to a final volume of 50 µL. All reactions were always initiated by adding
494 TdT in the end. Signal 0 and signal 1 controls were run for 1 hour for each set-up in triplicates and
495 also put through the ssDNA wash step at 40 minutes. Six replicates were run for 0→1→0 reactions.
496
497 **ssDNA wash for replacing buffers for 0→1→0 reactions:**
498
499 For changing cation concentration from 1 to 0 we utilized the ssDNA clean-up kit (ssDNA/RNA
500 clean/concentrator D7010) from Zymo Research such that all the extended ssDNA synthesized in
501 the initial part of the experiment was retained on the column and the TdT, reaction buffer, cation
502 and dNTPs were washed away. Each 50 µL extension reaction was individually loaded into a
503 separate column. Protocol was followed as mentioned in the kit. ssDNA was eluted into 20 µL
504 ddH$_2$O. We noticed in initial tests that after using the ssDNA clean-up kit, there was little to no TdT-
505 based extension in some replicates (data not included). We presume this is due some ethanol
506 getting carried forward into the eluted ssDNA. Thus we extended the dry spin time based on
507 suggestion from Zymo Research to 4 minutes. We also utilized two other ways to evaporate any
508 remaining ethanol after the column dry spin step based on protocol mentioned in Cold Spring
509 Harbor Protocols(45). We either kept the columns open in a biohood for 15 minutes to allow for
510 evaporation, or after elution of ssDNA we kept the 1.5 mL eppendorf tubes containing the eluted
511 ssDNA open at 45 °C for 3 minutes. Both methods gave better ethanol removal than just dry spin,
512 and they were tried in triplicates and averaged and plotted for the time prediction analysis (Fig. 3C).
513
514 **Illumina library preparation and sequencing:**
515
516 Our sample preparation pipeline for NGS was adapted from a previous protocol(46)·(47). After
517 extension reaction, 2 µL of the product was utilized for a ligation reaction. 22 bp universal tag,
518 common sequence 2 (CS2) of the Fluidigm Access Array Barcode Library for Illumina Sequencers
519 (Fluidigm), synthesized as ssDNA with a 5' phosphate modification and PAGE purified (Integrated
520 DNA Technologies), was blunt-end ligated to the 3' end of extended products using T4 RNA ligase.
521 Ligation reactions were carried out in 20 µL volumes and consisted of 2 µL of extension reaction,
522 1 uM CS1 ssDNA, 1X T4 RNA Ligase Reaction Buffer (NEB), and 10 units of T4 RNA Ligase 1
523 (NEB). Ligation reactions were incubated at 25 °C for 16 hours. Ligated products were stored at
524 −20 °C until PCR that was carried out on the same day. Ligation products were never stored at -20
525 °C for more than 24 hours.
526
527 PCR was performed with barcoded primer sets from the Access Array Barcode Library for Illumina
528 Sequencers (Fluidigm) to label extension products from up to 96 individual reactions. Each PCR
529 primer set contained a unique barcode in the reverse primer. From 5'-3' the forward PCR primer
530 (PE1 CS1) contained a 25-base paired-end Illumina adapter 1 sequence followed by CS1. The
531 binding target of the forward PCR primer was the reverse complement of the CS1 tag that was
532 used as the starting DNA substrate. From 5'-3' the reverse PCR primer (PE2 BC CS2) consisted
533 of a 24-base paired-end Illumina adapter 2 sequence (PE2), a 10-base Fluidigm barcode (BC), and
534 the reverse complement of CS2. CS2 DNA that had been ligated onto the 3' end of extended
535 products served as the reverse PCR primer-binding site. Each PCR reaction consisted of 2 µL of
536 ligation product, 1X Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB), and 400 nM
537 forward and reverse Fluidigm PCR primers in a 20 µL reaction volume. Products were initially
538 denatured for 30 s at 98 °C, followed by 20 cycles of 10 s at 98 °C (denaturation), 30 s at 60 °C
539 (annealing), and 30 s at 72 °C (extension). Final extensions were performed at 72 °C for 10 min.
540 Amplified products were stored at −20 °C until clean up and pooling. QC for individual sequencing
541 libraries was performed as follows. 2 µL of each library was pooled into a QC pool and the size and
542 approximate concentration was determined using Agilent 4200 Tapestation. Pool concentration
543 was further determined using Qubit and qPCR methods. Sequencing was performed on an Illumina
544 MiniSeq Mid Output flow cell and sequencing was initiated using custom sequencing primers
545 targeting the CS1 and CS2 conserved sites in the library linkers. Additionally phiX control library
546 was spiked into the run at 15-20% to increase diversity of the library clustering across the flow cell.

547 After demultiplexing, the percent seen for each sample was used to calculate a new volume to pool
548 for a final sequencing run with evenly balanced indexing across all samples. This pool was
549 sequenced with metrics identical to the QC pool. Library preparation and sequencing were
550 performed at the University of Illinois at Chicago Sequencing Core (UICSQC).

551 **NGS Data Preprocessing:**

552 For each sample, the NGS reads were first trimmed and filtered using cutadapt (v1.16). Only NGS
553 read pairs with both Illumina Common Sequence adapters, CS1 and CS2, were kept. Of these,
554 CS2 was trimmed off each R1 sequence and CS1 was trimmed off each R2 sequence. Cutadapt
555 parameters were set as following: a minimum quality cutoff (-q) of 30, a maximum error rate (-e) of
556 0.05, a minimum overlap (-O) of 10, and a minimum extension length (-m) of 1. The minimum
557 overlap was set to be higher than the default value of 3 because extended sequences in this case
558 are random, and we did not want to filter out sequences where the final 1-10 bases just happen to
559 look like the first 10 bases of CS2 (the read must still contain a full CS2 sequence for it to be kept
560 and subsequently trimmed, however). The 3' (-a) adapter trimmed from the R1 reads was
561 5'AGACCAAGTCTCTGCTACCGTA3' (CS2 reverse complement), and the 5' (-A) adapter trimmed
562 from the R2 reads was 5'TGTAGAACCATGTCGTCAGTGT3' (CS1 reverse complement). FastQC
563 was used to quickly inspect the output trimmed .fastq files before downstream analysis. See
564 *filter_and_trim_TdT.sh* at https://github.com/tyo-nu/turtles for an example preprocessing script. All
565 runs were trimmed using this script. All initial preprocessing was done on Quest, Northwestern
566 University's high-performance computing facility, using a node running Red Hat Enterprise Linux
567 Server release 7.5 (Maipo) with 4 cores and 4 GB of RAM, although only 1 core was used.
568 Preprocessing took between 5 and 30 minutes depending on the number of conditions, replicates,
569 and reads per replicate in a given run.

570 Finally, for each analysis, we did further preprocessing locally. We cut off bases that were still
571 present in the reads but not added during the experiment. Degenerate bases (if any) that are part
572 of the 5' ssDNA substrate (at its 3' end before the extension) were removed from the beginning of
573 each sequence. Then, we cut off 5.8 bases off the end of every sequence because we found that,
574 on average, 5.8 bases were being added after the extension reaction during the 16 hour ligation
575 step (Figure S14). Because 5.8 is not an integer value, we cut 5 bases off of 80% of the sequences
576 and 6 bases off of 20% of the sequences. We also filtered out sequences with length less than 6
577 bases.
578
579 **Timepoint prediction for 0→1 single step change experiment:**
580
581 All further analysis was done in python using Jupyter Notebooks. You can find all the Jupyter
582 Notebooks used for this publication at https://github.com/tyo-nu/turtles. The following algorithm was
583 applied in order to (1) read and normalize each sequence by its own length, (2) calculate a distance
584 metric using the relative dATP, dCTP, dGTP, and dTTP percent incorporation changes between
585 each condition and the 0 control, and (3) transform distances for all conditions into 0 → 1 space
586 based on the 0 and 1 control distance values.
587
588 We first normalize each sequence by length, such that all bases in each sequence are counted
589 across 1000 bins. For example, for a sequence of length 10, the first base would get counted in the
590 first 100 bins, the next base in bins 100-200, and so on.
591
592 We then calculate base composition, X_{ij}, in the sequence for condition, i, at each bin with position,
593 j, using the formula for a closure (equation 1). Note that i is unique for each (condition, replicate)
594 pair if multiple replicates are present for a given experimental condition.
595

596 $$X_{ij} = \left[\frac{n_{ijA}}{\sum_{k \in N} n_{ijk}}, \frac{n_{ijC}}{\sum_{k \in N} n_{ijk}}, \frac{n_{ijG}}{\sum_{k \in N} n_{ijk}}, \frac{n_{ijT}}{\sum_{k \in N} n_{ijk}} \right] \quad (1)$$

597

598   Here, n_{ijk} is the total count of dATP, dCTP, dGTP, or dTTP depending on the value of k ($k \in N =$
599   $\{A, C, G, T\}$) across all sequences for condition, i, at bin, j.
600

601   To calculate distance between two compositions at a given bin location (e.g. between the 0 and 1
602   controls at every bin), we have to first transform the compositional data. We cannot simply take the
603   L2 norm difference of each compositional element because the elements of a composition violate
604   the principle of normality due to the total sum rule (all elements add up to 100%). Thus, the data is
605   first transformed by using the center log-ratio (clr) transformation which maps this 4-component
606   composition from a 3-dimensional space to a 4-dimensional space. We then take the L2 norm of
607   these transformed normal elements. This distance metric is known as the Aitchison Distance, which
608   is used here to calculate the base composition distance, $d_j(0, i)$, from the 0 control to each
609   condition, i, at each bin, j (equation 2).
610

611
$$d_j(0, i) = \sqrt{\sum_{k \in N} \left[\ln\left(\frac{X_{ijk}}{g(X_{ij})}\right) - \ln\left(\frac{X_{0jk}}{g(X_{0j})}\right) \right]} \quad (2)$$

612
613   $N = \{A, C, G, T\}$ and $g(X_{ij})$ is the geometric mean for condition, i, and bin, j, across all four bases
614   in N (equation 3).
615

616
$$g(X_{ij}) = \sqrt[4]{\prod_{k \in N} X_{ijk}} \quad (3)$$

617
618   For condition, i, and bin j, the output signal, s_{ij}, is calculated as
619

620
$$s_{ij} = \frac{d_j(0, i) - d_j(0,0)}{d_j(0,1) - d_j(0,0)} = \frac{d_j(0, i)}{d_j(0,1)} \quad (4)$$

621
622   where $d_j(0,1)$ is the Aitchison distance between the 0 control base composition and 1 control base
623   composition at bin, j. $d_j(0,0) = 0$ for all j. If there were multiple replicates for the 0 control, their
624   average composition was used for X_{0j} (and X_{0jk}) in equation 2. If there were multiple replicates for
625   the 1 control, their average composition was similarly used to calculate $d_j(0,1)$ in equation 4.
626

627   Next, the switch times were estimated for each condition, i, which contains a change in output
628   signal, s_{ij}, (e.g. via addition of Co halfway through the reaction). For experiments with more than
629   one change (e.g. 0 → 1 → 0), a more sophisticated approach was used and is detailed below.
630   However, the following simpler, more intuitive approach was used to predict switch times for 0 →
631   1 and 1 → 0.
632

633   Switch times were estimated for a given condition, i, by (1) finding j_i^*, the average location across
634   all the sequences (bin position, j) at which half the 1 control output signal is reached (i.e. $s_{ij} = 0.5$),
635   (2) calculating α, the ratio of the average rate of nucleotide addition for the 0 and 1 controls, and
636   (3) using j_i^* and α to calculate the switch time, t_i^*, using equations 5 and 6. For a derivation of
637   equation 5, see supplementary methods.
638

639
$$t_i^* = \frac{\alpha \, t_{expt}}{\frac{1}{j_i^*} + \alpha - 1} \quad (5)$$

640
641   where
642

643
$$\alpha = \frac{\overline{r_{a,ctrl}}}{\overline{r_{b,ctrl}}} \quad (6)$$

644
645   $\overline{r_{a,ctrl}}$ is the average synthesis rate of the first environmental condition before the switch. For
646   example, $\overline{r_{a,ctrl}}$ would be calculated using the 0 control for the condition, 0 → 1, but the 1 control
647   for the condition, 1 → 0. The average synthesis rate is calculated by dividing the average extension
648   length by the duration of the experiment. $\overline{r_{b,ctrl}}$ is the average synthesis rate for the second
649   environmental condition (after the switch).
650
651   **Timepoint estimation for 0→1→0 multiple fluctuations experiment:**
652
653   To predict the Co^{2+} condition in the 0→1→0 experiment, we used the algorithm we developed in
654   Glaser et al. for decoding continuous concentrations(13). The input to this algorithm is the amount
655   of output signal on every nucleotide. Here, the output signal is s_{ij} from the previous section. The
656   algorithm uses this information to predict continuous values of Co^{2+} between 0 and 1 for all time
657   points that are most likely to produce the amount of output signal on the nucleotides. To binarize
658   these predictions, we then set a threshold of 0.5. To be able to predict the values of Co^{2+}, the
659   algorithm requires knowledge of the expected amount of output signal in the 0 and 1 control
660   conditions. Here, this is the average output signal across nucleotides in the 0 or 1 control
661   experiments. The algorithm also requires knowledge of the rate of nucleotide addition. Here, we fit
662   an inverse Gaussian distribution to the average experimental dNTP addition rate distribution (the
663   distribution of the sequence lengths divided by the experiment time) from the control experiments.
664   Note that this algorithm also assumes that the rate of dNTP addition is independent of the cation
665   concentration. Thus, when making predictions in the 0→1→0 experiment, we do not account for
666   differences in the rate of dNTP addition distributions between the 0 and 1 conditions. A future
667   algorithm that takes this difference into account could yield more accurate predictions.
668
669   ***In silico* simulations of recording faster and higher number of input signal changes:**
670
671   Using the average dNTP incorporation rate from experiments, and the amount of output signal in
672   the control conditions, we simulated additional experiments *in silico*. Each simulated experiment
673   had at least 6 signal changes (instances of a single signal change from 0→1 or 1→0), where each
674   condition was randomly chosen to be 0 or 1. All nucleotides that were added during the 0 or 1
675   condition had the signal associated with these control conditions. More specifically, to account for
676   the experimental variability in signals within a given control condition, nucleotide signals were
677   sampled from a Normal distribution determined by the experimental variability of nucleotide signals
678   within the control conditions. We calculated the variability in two ways, corresponding to the two
679   representative curves in Fig. S13A and S13C. In one, the variability was calculated across the first
680   100 nucleotides, in which there were at least 2000 recordings of all base numbers. In the second,
681   the variability was calculated across the first 50 nucleotides, in which there were at least 60000
682   recordings of all base numbers. Using the output signal of the simulated nucleotides, we used the
683   algorithm we developed in Glaser et al. for decoding binary concentrations (13). Accuracy
684   corresponds to the percentage of conditions correctly classified as 0 or 1 over the duration of the
685   entire recording experiment.
686
687   **Time point estimation for 0→1, 1→0 single step change for TURTLES-2 using an inverse model:**
688
689   To predict the Co^{2+} condition in the 0→1→0 experiment, we used a variation of the algorithm we
690   developed in Glaser et al. for decoding continuous concentrations (13). This algorithm will predict
691   continuous values of Co^{2+} between 0 and 1 for all time points that are most likely to produce the
692   amount of signal. Here, instead of using the amount of signal on every nucleotide to predict the
693   continuous concentrations, we use the normalized signal.
694
695   Let s_{ij} be the signal as a function of condition, i, and normalized position, j. Let $\gamma_{ij}(t)$ be the
696   probability that a nucleotide corresponding to normalized position j was written at time t. Let C_i be
697   the normalized cation concentration for condition i. Like in (13), our model is that $s_{ij} \sim \sum_t \gamma_{ij}(t) C_i(t)$.

698   We use maximum likelihood estimation to find C_i that minimizes $\sum_j(s_{ij} - \sum_t \gamma_{ij}(t)C_i(t))^2$ subject to
699    $C_i(t) \in [0,1]$ and given condition, i. To binarize the predictions, we then set a threshold for C of 0.5.
700
701   Here, for an experiment of duration t_{expt} (e.g. 60 minutes), we let $\gamma_{ij} = \mathrm{N}((j-0.5)/t_{expt}, \sigma)/Z$,
702   where Z renormalizes the probability distribution after values outside the domain of $[0, t_{expt}]$ are
703   set to 0. We set σ for each experiment so that $\gamma_{i1}(t_{expt})$ is equal to the frequency of strands with a
704   single nucleotide divided by t_{expt} (because a normalized position of 1 would generally only be
705   written in the t_{expt}th minute when there is a single nucleotide strand). Note that future work that
706   more accurately models the kinematics of the polymerase to get a more accurate estimate of γ will
707   provide improved results.
708
709   When running this algorithm on the Ca^{2+} data which has different rates when Ca^{2+} is or isn't present,
710   following the prediction of Ca^{2+} over time with the above algorithm, we used the ratio of
711   incorporation rates between the 0 and 1 condition, as described by Equation 5 and 6, to rescale
712   the results.
713
714
715
716   **Data and code availability**
717   All data generated during this study are included in this published article and its Supplementary
718   Information. NGS data are available from Sequence Read Archive
719   https://www.ncbi.nlm.nih.gov/sra/PRJNA542184
720
721
738
739

References

1. G. M. Church, Y. Gao, S. Kosuri, Next-Generation Digital Information Storage in DNA. *Science (80-.).* **337**, 1628–1628 (2012).

2. N. Goldman, *et al.*, Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **494**, 77–80 (2013).

3. Y. Erlich, D. Zielinski, DNA Fountain enables a robust and efficient storage architecture. *Science (80-.).* **355**, 950–954 (2017).

4. R. N. Grass, R. Heckel, M. Puddu, D. Paunescu, W. J. Stark, Robust Chemical Preservation of Digital Information on DNA in Silica with Error-Correcting Codes. *Angew. Chemie Int. Ed.* **54**, 2552–2555 (2015).

5. R. U. Sheth, S. S. Yim, F. L. Wu, H. H. Wang, Multiplex recording of cellular events over time on CRISPR biological tape. *Science* **358**, 1457–1461 (2017).

6. T. B. Loveless, *et al.*, Lineage tracing and analog recording in mammalian cells by single-site DNA writing. *Nat. Chem. Biol.* **17**, 739–747 (2021).

7. S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, Molecular recordings by directed CRISPR spacer acquisition. *Science (80-.).* **353** (2016).

8. S. L. Shipman, J. Nivala, J. D. Macklis, G. M. Church, CRISPR–Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature* **547**, 345–349 (2017).

9. S. Palluk, *et al.*, De novo DNA synthesis using polymerase-nucleotide conjugates. *Nat. Biotechnol.* **36**, 645–650 (2018).

10. H. H. Lee, R. Kalhor, N. Goela, J. Bolot, G. M. Church, Terminator-free template-independent enzymatic DNA synthesis for digital information storage. *Nat. Commun.* **10** (2019).

11. H. Lee, *et al.*, Photon-directed multiplexed enzymatic DNA synthesis for molecular digital data storage. *Nat. Commun.* **11**, 1–9 (2020).

12. S. S. Yim, *et al.*, Robust direct digital-to-biological data storage in living cells. *Nat. Chem. Biol.* **17**, 246–253 (2021).

13. J. I. Glaser, *et al.*, Statistical Analysis of Molecular Signal Recording. *PLoS Comput. Biol.* **9** (2013).

14. K. P. Kording, Of toasters and molecular ticker tapes. *PLoS Comput. Biol.* **7**, 1–5 (2011).

15. Z. Kelman, M. O'Donnell, DNA polymerase III holoenzyme: Structure and function of a chromosomal replicating machine. *Annu. Rev. Biochem.* **64**, 171–200 (1995).

16. E. A. and A. J. B. Motea, Terminal Deoxynucleotidyl Transferase: The Story of a Misguided DNA Polymerase. **21**, 253–260 (2015).

17. J. B. Boulé, E. Johnson, F. Rougeon, C. Papanicolaou, High-level expression of murine terminal deoxynucleotidyl transferase in Escherichia coli grown at low temperature and overexpressing argU tRNA. *Mol. Biotechnol.* **10**, 199–208 (1998).

18. M. S. Chang, F. J. Bollum, Multiple Roles of Divalent Deoxynucleotidyltransferase Cation in the Terminal Reaction *. **265**, 17436–17440 (1990).

19. J. D. Fowler, Z. Suo, Biochemical , Structural , and Physiological Characterization of Terminal Deoxynucleotidyl Transferase. 2092–2110 (2006).

20. M. R. Deibel, M. S. Coleman, Biochemical properties of purified human terminal deoxynucleotidyltransferase. *J. Biol. Chem.* **255**, 4206–12 (1980).

21. C. Grienberger, A. Konnerth, Imaging Calcium in Neurons. *Neuron* **73**, 862–885 (2012).

22. M. Whitaker, Calcium at fertilization and in early development. *Physiol. Rev.* **86**, 25–88 (2006).

23. S. A. Stricker, Comparative Biology of Calcium Signaling during Fertilization and Egg Activation in Animals. *Dev. Biol.* **211**, 157–176 (1999).

24. S. S. Rosenberg, N. C. Spitzer, Calcium Signaling in Neuronal Development https:/doi.org/10.1101/cshperspect.a004259 (June 24, 2020).

25. C. J. Frederickson, J.-Y. Koh, A. I. Bush, The neurobiology of zinc in health and disease. *Nat. Rev. Neurosci.* **6**, 449–462 (2005).

26. E. A. Motea, A. J. Berdis, Terminal deoxynucleotidyl transferase: The story of a misguided DNA polymerase. *Biochim. Biophys. Acta - Proteins Proteomics* **1804**, 1151–1166 (2010).

27. J. Nakai, M. Ohkura, K. Imoto, A high signal-to-noise ca2+ probe composed of a single green fluorescent protein. *Nat. Biotechnol.* **19**, 137–141 (2001).

28. A. E. Palmer, *et al.*, Ca2+Indicators Based on Computationally Redesigned Calmodulin-Peptide

796   Pairs. *Chem. Biol.* **13**, 521–530 (2006).

797   29.   S. Edwardraja, *et al.*, Caged activators of artificial allosteric protein biosensors (2020)
798   https://doi.org/10.1021/acssynbio.9b00500.

799   30.   Z. Guo, *et al.*, Generalizable Protein Biosensors Based on Synthetic Switch Modules. *J. Am. Chem.*
800   *Soc.*, jacs.8b12298 (2019).

801   31.   M. A. Smith, F. H. Arnold, Designing libraries of chimeric proteins using SCHEMA recombination
802   and RASPP. *Methods Mol. Biol.* **1179**, 335–343 (2014).

803   32.   L. Crotti, *et al.*, Calmodulin mutations associated with recurrent cardiac arrest in infants.
804   *Circulation* **127**, 1009–1017 (2013).

805   33.   S. N. Zulkifli, H. A. Rahim, W. J. Lau, Detection of contaminants in water supply: A review on
806   state-of-the-art monitoring technologies and their applications. *Sensors Actuators, B Chem.* **255**,
807   2657–2689 (2018).

808   34.   S. Slomovic, K. Pardee, J. J. Collins, Synthetic biology devices for in vitro and in vivo diagnostics.
809   *Proc. Natl. Acad. Sci. U. S. A.* **112**, 14429–14435 (2015).

810   35.   X. Liu, *et al.*, Design of a transcriptional biosensor for the portable, on-demand detection of
811   cyanuric acid. *ACS Synth. Biol.* **9**, 84–94 (2020).

812   36.   J. K. Jung, *et al.*, Cell-free biosensors for rapid detection of water contaminants HHS Public Access
813   Author manuscript. *Nat Biotechnol* **38**, 1451–1459 (2020).

814   37.   W. Thavarajah, *et al.*, Point-of-Use Detection of Environmental Fluoride via a Cell-Free
815   Riboswitch-Based Biosensor. *ACS Synth. Biol.* **9**, 10–18 (2020).

816   38.   A. H. Marblestone, *et al.*, Physical principles for scalable neural recording. *Front. Comput.*
817   *Neurosci.* **7**, 1–34 (2013).

818   39.   A. H. Marblestone, *et al.*, Rosetta Brains: A Strategy for Molecularly-Annotated Connectomics
819   (2014) (December 3, 2018).

820   40.   P. J. Vikesland, Nanosensors for water quality monitoring. *Nat. Nanotechnol.* **13**, 651–660 (2018).

821   41.   F. Long, A. Zhu, H. Shi, H. Wang, J. Liu, Rapid on-site/in-situ detection of heavy metal ions in
822   environmental water using a structure-switching DNA optical biosensor. *Sci. Rep.* **3**, 2308 (2013).

823   42.   M. Eisenstein, Enzymatic DNA synthesis enters new phase. *Nat. Biotechnol.* **38**, 1113–1115
824   (2020).

825   43.   J. Akerboom, *et al.*, Optimization of a GCaMP Calcium Indicator for Neural Activity Imaging. *J.*
826   *Neurosci.* **32**, 13819–13840 (2012).

827   44.   T.-W. Chen, *et al.*, Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**,
828   295–300 (2013).

829   45.   M. R. Green, J. Sambrook, Precipitation of DNA with Ethanol. *Cold Spring Harb. Protoc.* **2016**,
830   pdb.prot093377 (2016).

831   46.   A. M. de Paz, *et al.*, High-resolution mapping of DNA polymerase fidelity using nucleotide
832   imbalances and next-generation sequencing. *Nucleic Acids Res.* **46**, e78–e78 (2018).

833   47.   B. M. Zamft, *et al.*, Measuring cation dependent DNA polymerase fidelity landscapes by deep
834   sequencing. *PLoS One* **7** (2012).

835

836

837
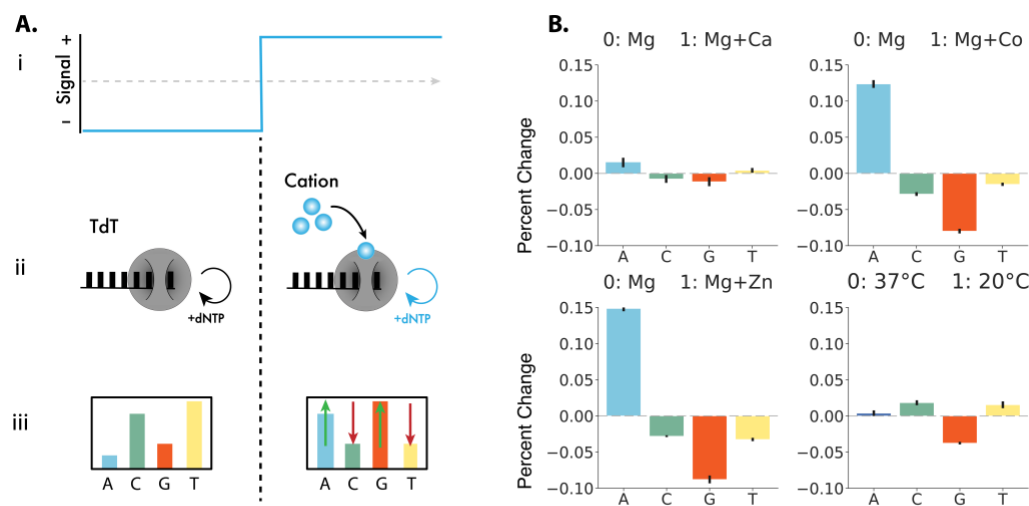
838
839 **Figure 1: TURTLES device architecture and environmental signal responses. (A.i)** A
840 representative time-varying input signal. **(A.ii)** TdT interacts directly with the signal of interest (small
841 blue circles) **(A.iii)** resulting in different average DNA compositions in each condition. **(B)** Change
842 in frequency of nucleotide selectivity by TdT in the presence of various environmental signals
843 tested. Signal 0 is 10 mM Mg^{2+} at 37 °C for 1 hour. Signal 1 was: (top-right) 10 mM Mg^{2+} + 1 mM
844 Ca^{2+} at 37 °C for 1 hour A increased by 1.5%, G decrease by 1.2%, T increased by 0.4%, and C
845 decreased by 0.8%; (top-left) 10 mM Mg^{2+} + 0.25 mM Co^{2+} at 37 °C for 1 hour, A incorporation
846 increased by 12.4%, while G decreased by 8.0% and T and C decreased by 1.5% and 2.9%
847 respectively ; (bottom-left) 10 mM Mg + 20 µM Zn^{2+} at 37 °C for 1 hour, A increased by 14.9%, G
848 8.8% decreased by, T decreased by 3.3%, and C decreased by 2.8%; and (bottom-right) 10 mM
849 Mg^{2+} at 20 °C for 1 hour, 0.4% increase in A, 3.8% decrease in G, 1.5 % increase in T and 1.8%
850 increase in incorporation of C. Error bars show two standard deviations of the mean. Statistical
851 significance was assessed after first transforming the data into Aitchison space which makes each
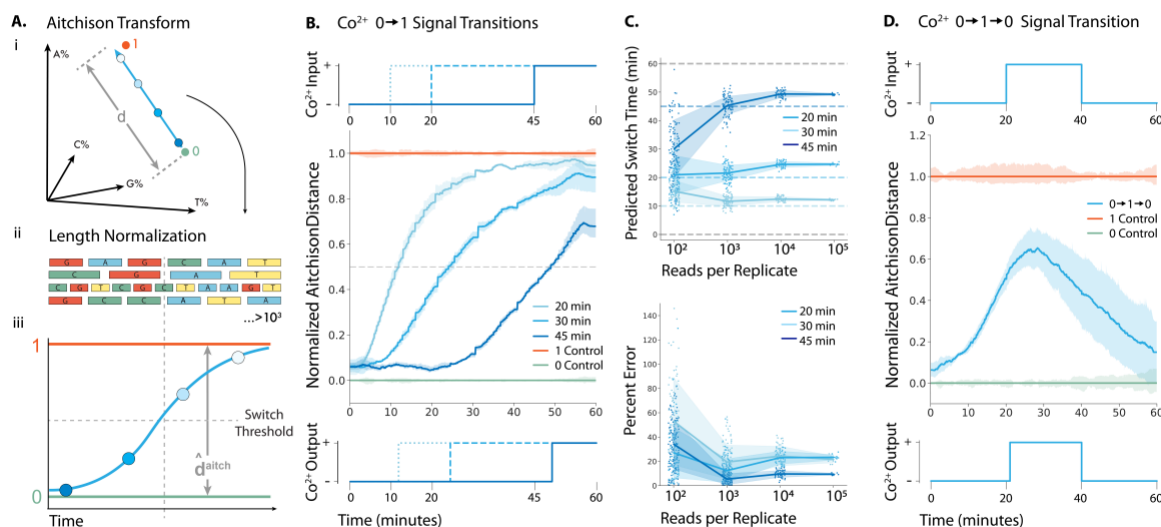852 dNTP frequency change statistically independent of the others (Fig. S2).
853
854
855

856
857 **Figure 2: Recording a Co²⁺ fluctuations into ssDNA with minutes resolution *in vitro*. (A.i.)**
858 Representation of how percent incorporation of each nucleotide is dependent on each of the
859 nucleotide incorporated. **(A.ii.)** Sequences were normalized by length before the nucleotide
860 composition at each time point was calculated. **(A.iii.)** By transforming the percent incorporation of
861 each nucleotide to the Atichison distance we can calculate the total "output signal". We plot the
862 Aitchison distance for the recording experiment between the 0 (green) and 1 (orange) signal. **(B)**
863 **Top:** 0.25 mM Co²⁺ was added at time 10, 20 and 40 minutes to generate a 0→1 transition. **Center:**
864 Mean output signal across 3 biological replicates. Vertical lines are drawn at the inferred transition
865 time. **Bottom:** Predicted output signal transition times were 11.9, 24.4 and 49.2 minutes. **(C) Top:**
866 Predicted switch times for each 0→1 transition calculated from randomly sampled subsets of
867 sequences. **Bottom:** Time prediction error for each 0→1 transition calculated from randomly
868 sampled subsets of sequences. **(D) Top:** 0.25 mM Co²⁺ was added at 20 minutes and then removed
869 at 40 minutes to generate a 0→1→0 transition. **Center:** Mean output signal across 3 biological
870 replicates. **Bottom:** Using the algorithm detailed in Glaser et al. (13), the signal was deconvoluted
871 into a binary response, with vertical lines drawn at the predicted switch times of 21 minutes and 41
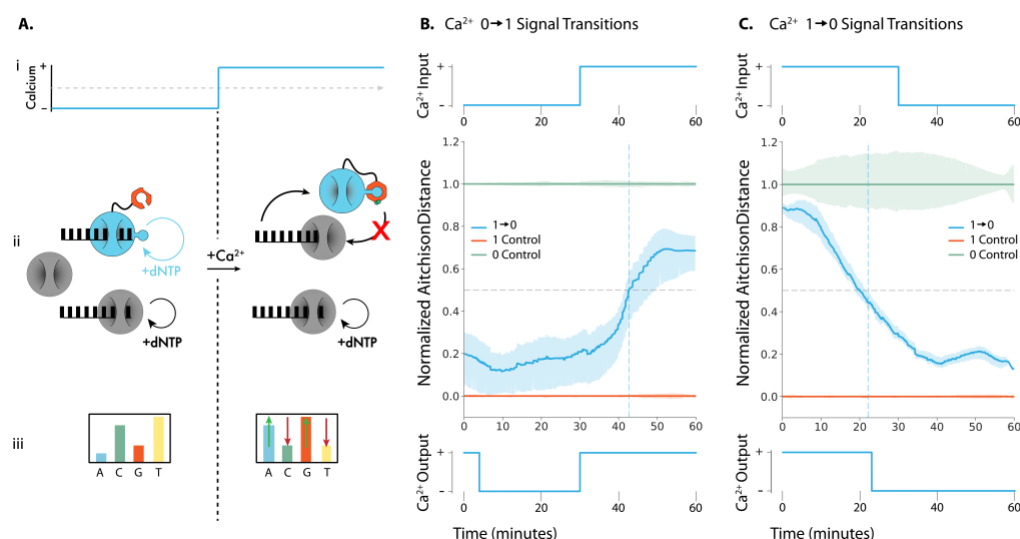872 minutes.

Figure 3: Recording Ca^{2+} changes into ssDNA with TURTLES-2. (A.i.) Representative calcium transition changing concentration from calcium-free to calcium-added conditions during a TdT-based DNA synthesis reaction. Mg^{2+} concentration and reaction temperature are held constant. **(A.ii.)** The CaM subunit (orange) of engineered TdT (teal) binds with the fused M13 peptide in the presence of Ca^{2+} to allosterically turn off DNA synthesis. The activity of the reference TdT (grey) is not affected by Ca^{2+}. In the absence of Ca^{2+} both the engineered TdT and reference TdT are carrying out DNA synthesis. In the presence of Ca^{2+} only the reference TdT synthesizes DNA. **(A.iii.)** This results in a change in the overall nucleotide incorporation preference upon a change in Ca^{2+}. **(B) Top:** 100 µM CaCl$_2$ was added to the extension reaction at 30 minutes to generate a 0→1 transition. **Center:** Mean output signal across 3 biological replicates. **Bottom:** Using a modified version of the algorithm detailed in Glaser et al. (13), the signal was deconvoluted into a binary response, with the predicted switch time of 30 minutes **(C) Top:** 50 µM EGTA was added to the extension reaction at 30 minutes to generate a 1→0 transition. **Center:** Mean output signal across 3 biological replicates. **Bottom:** Using a modified version of the algorithm detailed in Glaser et al. (13), the signal was deconvoluted into a binary response, with the predicted switch time of 22 minutes.

891
892 **Supplementary Information Text**
893 **Supplementary Methods**: Extended description of methods

894
895 **1. Extension reaction with individual dNTPs for testing effect of Co^{2+}:**

896
897 For initial testing to show Co^{2+} dependent dNTP preference change the ssDNA substrate used was
898 AMD006 (Table S2). Total reaction volume was 25 μL with 0.1 μM ssDNA substrate, 1x NEB TdT
899 reaction buffer, and 0.1 mM of each dNTP tested. Final concentration of $CoCl_2$ in the test reaction
900 was 0.25 mM. Reactions were initiated by addition of 5 units of TdT per reaction. Reactions were
901 run for 30 minutes at 37 °C and stopped by boiling at 70 °C for 10 minutes. Then, 8 μL of the
902 reaction was mixed with 12 μL of TBE-Urea loading dye and boiled for 10 minutes at 100 °C. All of
903 the diluted extension reaction was then loaded onto 30 μL, 10 well 10% TBE-Urea Gel (Bio-Rad)
904 and run for 40 minutes at 200 V. Immediately after the run was over, the gel was stained with Sybr
905 Gold for 15 minutes and imaged on ImageQuant BioRad.

906
907 **2. Extension reactions for 1→0 set-up:**

908
909 Mg^{2+} only for 1 hour (signal 0) and $Mg^{2+}+Co^{2+}$ for 1 hour (signal 1) were set-up as regular extension
910 reactions mentioned in Materials and Methods. The 1→0 reactions where the signal changed from
911 1 to 0 at 40 minutes were put through a ssDNA was step at 40 minutes. ssDNA wash to remove
912 cations, TdT and dNTPs was done exactly as mentioned in Materials and Methods. Reactions were
913 all run for 1 hour in triplicates. Signal 0 and signal 1 controls were run for 1 hour for each set-up in
914 triplicates and also put through the ssDNA wash step at 40 minutes.

915
916 **3. Derivation of Equation 5**

917
918 We start by deriving the equations for the average rate before the switch (r_A) and after the switch
919 (r_B) for condition, i:
920

$$r_{a,i} = \frac{j_i^*}{t_i^*} \quad (1a)$$

921
922

$$r_{b,i} = \frac{1 - j_i^*}{t_{expt} - t_i^*} \quad (2a)$$

923
924
925 where j_i^* is the average location in the sequences (length fraction, 0 to 1) at which the output
926 signal, s_{ij}, reaches 0.5 (Equation 4), t_i^* is the switch time, and t_{expt} is the total duration of the
927 experiment. Because we can estimate $r_{a,i}$ and $r_{b,i}$ from average rates of the 0 and 1 controls
928 across replicates ($\overline{r_{a,ctrl}}$ and $\overline{r_{b,ctrl}}$), we can use their ratio to combine equation 1a and 2a, above
929 to write
930

$$\frac{\overline{r_{a,ctrl}}}{\overline{r_{b,ctrl}}} \approx \frac{r_{a,i}}{r_{b,i}} = \frac{j_i^*}{t_i^*}\left(\frac{t_{expt} - t_i^*}{1 - j_i^*}\right) \quad (3a)$$

931
932
933 Solving for t_i^*, we get equation 5:
934

$$t_i^* = \frac{\alpha \, t_{expt}}{\frac{1}{j_i^*} + \alpha - 1} \quad (5)$$

935
936
937 where
938

939
$$\alpha = \frac{\overline{r_{b,ctrl}}}{\overline{r_{a,ctrl}}} \quad (4a)$$

940
941 We use equation 5 for time prediction (t_i^*) after calculating j_i^* for a given condition and α from the
942 0 and 1 controls. In equation 4a, a is the first condition before the switch (0 or 1) and b is the
943 condition after the switch (1 or 0).

944
945 **4. Extensions reaction set-up for calculating rate of dNTP addition:**
946
947 Each extension reaction consisted of a final concentration of 10 μM initiating ssDNA substrate, 1
948 mM dNTP mix (each dNTP at 1 mM final concentration), 1.4x NEB TdT reaction buffer, and 10
949 units of TdT to a final volume of 50 μL. The ssDNA substrate used for this extension reaction was
950 CS1_5N. We have shown (data not included) that the identity of the last 5 bases on the 3' end of
951 the substrate affects the identity of the dNTP added to the ssDNA substrate. Thus, we purchased
952 a ssDNA substrate (CS1_5N) with the last 5 bases having the base composition same as TdT
953 dNTP preference under signal 0 (25% dATP, 15% dCTP, 45% dGTP and 15% dTTP). The
954 reactions were initiated upon addition of TdT and run at 37 °C for 2 hours. 2 μL of sample was
955 collected and immediately frozen (on ice, 0 °C) at 30 s, 1 min, 2 min, 3 min, 4 min, 5 min, 10 min,
956 20 min, 30 min, 45 min, 60 min, 92 min and 120 min. Subsequently, each sample was put through
957 the ligation and Illumina library generation process as mentioned in Materials and Methods.

958
959 **5. Test set-up for checking ssDNA clean-up kit bias:**
960
961 Mg^{2+} only for 1 hour (signal 0) and $Mg^{2+}+Co^{2+}$ for 1 hour (signal 1) were set up as regular extension
962 reactions mentioned in Materials and Methods. The 0→1 reactions where the signal changed from
963 0 to 1 during the 1 hour extension were run starting with 45 μL with Mg^{2+} only. 5 μL of 2.5 mM $CoCl_2$
964 was added at 10 min. Reactions were all run for 1 hour in triplicates. Fresh signal 0 and signal 1
965 controls were run for 1 hour with each set-up. 2 μL of extension reaction was used for ligation ("No
966 Wash" set of samples). Ligation and subsequent PCR steps for Illumina library generation were
967 followed as mentioned in Materials and Methods. Rest of the 48 uL of extension reaction was
968 washed using the ssDNA clean-up kit. Protocol was followed as mentioned in the kit. ssDNA was
969 eluted into 25 μL of ddH_2O and 2 μL of that was used for ligation ("Wash" set of samples). Ligation
970 and subsequent PCR steps for Illumina library generation were followed as mentioned in Materials
971 and Methods. Data obtained from Illumina sequencing was analyzed for the "No Wash" and "Wash"
972 set of samples. Further, switch time calculations were carried out as mentioned previously (Fig.
973 S12).

974
975 **6. High calcium conditions for TURTLES-2 reactions.** High calcium conditions for 0→1 and 1→0
976 reactions had different compositions to enable transitions without employing intermediate column
977 washes. Although commercial TdT reaction buffer contains no added calcium, we observed that
978 CaM-mTdT(M13-388) was inactivated in reactions that were not supplemented with at least 50 μM
979 EGTA (Figure S18), very likely due to calcium present in water or other reagents used. By titrating
980 EGTA into a TdT extension reaction supplemented with 7 μM fura-2 until the fura-2 signal
981 plateaued, we estimated that most free calcium could be eliminated from the reactions by the
982 addition of 50 μM EGTA. By employing the un-supplemented reaction buffer as the high calcium
983 condition for 1→0 reactions, we were able to transition to a low calcium condition with the addition
984 of 50 μM EGTA.

985