Published in final edited form as:

Nat Chem. 2020 August 01; 12(8): 683–690. doi:10.1038/s41557-020-0502-8.

Discovery and evolution of RNA and XNA reverse transcriptase function and fidelity

Gillian Houlihan¹, Sebastian Arangundy-Franklin¹, Benjamin T. Porebski¹, Nithya Subramanian¹, Alexander I. Taylor^{1,2}, Philipp Holliger^{1,*}

¹MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Francis Crick Avenue, Cambridge CB2 0QH, UK

²Cambridge Institute of Therapeutic Immunology & Infectious Disease (CITIID), Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Puddicombe Way, Cambridge CB2 0AW, UK

Abstract

The ability of reverse transcriptases (RTs) to synthesize a complementary DNA (cDNA) from natural RNA and a range of unnatural xeno nucleic acid (XNA) template chemistries, underpins key methods in molecular and synthetic genetics. However, RTs have proven challenging to discover and engineer, in particular for the more divergent XNA chemistries. Here we describe a general strategy for the directed evolution of RT function for any template chemistry called compartmentalized bead labelling (CBL), and demonstrate it by the directed evolution of efficient RTs for 2'-O-methyl-RNA (2'OMe-RNA) and hexitol nucleic acids (HNA) and the discovery of RTs for the orphan XNA chemistries D-altritol nucleic acid (AtNA) and 2' Methoxyethyl-RNA (MOE-RNA), for which previously no RTs existed. Finally, we describe the engineering of XNA RTs with active exonucleolytic proofreading as well as the directed evolution of RNA RTs with very high cDNA synthesis fidelities, even in the absence of proofreading.

Introduction

Reverse transcriptases (RTs) are of central importance in the biology of retroviruses, mobile retroelements and in telomere maintenance. RT function, the conversion of RNA template sequences into complementary cDNAs, has also enabled key methods in molecular and synthetic biology such as RT-qPCR¹, RNAseq² and ribosome display³. Increasingly, RTs have also become a key enabling tool in the nascent field of synthetic genetics⁴, enabling the synthesis, replication and evolution of synthetic genetic polymers (xeno nucleic acids,

*Correspondence and requests for materials should be addressed to P.H. ph1@mrc-lmb.cam.ac.uk.

Author contributions

GH, SA-F and PH conceived and designed experiments. GH performed all experiments except structural models (with SA-F, BTP), deep sequencing and data analysis (with BTP) and RT characterization (with AIT, NS). All authors analyzed data, discussed results and co-wrote the manuscript.

Competing interests

The authors declare no competing interest.

XNAs) including nucleic acids with altered sugar rings (congeners)^{5, 6}, modified bases⁷ and backbones⁸.

Most commonly used RTs derive from variants of retroviral enzymes such as Moloney Murine Leukaemia Virus (M-MuLV RT)⁹ (e.g. Superscript III (SSIII)), avian myoblastis virus (AMV-RT), from *Bacillus stearothermophilus* (Bst) DNA polymerase (RTx)¹⁰ or from bacterial retrotransposons (MarathonRT)¹¹. Some fortuitously display RT activity for some XNA chemistries such as Bst for α -L-threofuranosyl nucleic acids (TNA)¹² and 2'-fluoroarabino nucleic acids (FANA)¹³. However, for most XNA templates there are either no RTs available or existing RT enzymes are inefficient. To fully realize novel applications arising from modified nucleic acids and more generally to enhance RT applications in synthetic biology new strategies for RT discovery and development are needed.

Here we describe the development and application of compartmentalized bead labelling (CBL), a general strategy for the engineering of RT function and fidelity that is customizable to any template chemistry. We demonstrate the utility of CBL by the directed evolution of a wide range of new RTs starting from RT521K⁵, a variant of Tgo, the replicative DNA polymerase from *Thermococcus gorgonarius*. We describe the evolution and characterization of efficient RTs for 2'O-methyl (2'OMe)-RNA and hexitol nucleic acids (HNA) - which outperform previously described RTs - and the *de novo* discovery of RTs for the orphan XNA template chemistries D-altritol nucleic acids (AtNA), 2' Methoxyethyl-RNA (2'MOE-RNA) and P- α -S-phosphorothioate 2'MOE-RNA (PS-2'MOE-RNA), for which no previous RT enzymes had been described. Finally, we demonstrate the ability of CBL as a selection platform for more complex phenotypes with the discovery of the highest fidelity RNA RT enzymes described to date.

Results

The greatest challenge to overcome during RT evolution is to link the gene (genotype) encoding the RT to the cDNA product (phenotype) arising from the reverse transcription of a (chemically divergent (XNA)) template, which must be supplied extraneously. CBL addresses this genotype-phenotype linkage challenge by co-encapsulation of single *E. coli* cells (containing RT-gene encoding plasmid and expressed RT enzyme) together with microbeads decorated with XNA (or RNA) templates and plasmid capture probes inside the aqueous compartments of a water-in-oil (w/o) emulsion (Fig. 1). Heat lysis of *E. coli* cells releases both the RT enzyme and encoding plasmids into the compartments, where both on-bead reverse transcription and on-bead capture of RT-encoding plasmids takes place, linking the cDNA product of reverse transcription (phenotype) to the RT-encoding plasmid (genotype).

Following emulsion breaking and bead recovery, on-bead cDNA synthesis is scored using fluorescent anti-sense probes and this signal can be further amplified using the hybridization chain reaction (HCR)¹⁴. This allows a sensitive and scalable detection of RT activity and the subsequent isolation of fluorescent beads comprising the most active RT activities by fluorescent activated bead sorting (FACS) with enrichment factors as high as 500-fold fold per round (Fig.1, Extended data Fig. 1).

We initiated CBL RT selection experiments from a previously described variant of the DNA polymerase from the hyperthermophilic archaeon *T. gorgonarius* (Tgo). This variant, RT521K (Tgo: V93Q, D141A, E143A, E429G, F445L, A485L, I521L, E664K, K726R) comprises mutations that inactivate its 3'-5' exonuclease (D141A, E143A), abrogate stalling at template uracil (V93Q) and enhance template binding and promiscuous RT activity across a range of chemistries (A485L, E429G, I521L, E664K, K726R)⁵. As a first target, we chose 2'OMe-RNA (in which the 2'-hydroxyl group of RNA is substituted by a methoxy group). 2'OMe-RNA occurs naturally as a sporadic post-transcriptional rRNA and mRNA modification¹⁵ and due to its high resistance to nuclease degradation and obvious non-toxicity is widely considered an attractive chemistry in aptamer discovery and therapeutic nucleic acids^{16, 17}. While RT521K as well as some commercial RTs (specifically Superscript III (SSIII) (Thermo Fisher Scientific, USA)¹⁶ and RTx (New England Biolabs, USA)) as well as some engineered forms of Taq polymerase Stoffel fragment¹⁸ can reverse transcribe 2'OMe-RNA, the efficiency of RT521K was limited, especially on complex (e.g. random sequence 2'OMe N₄₀ sequence) templates. We sought to test the ability of CBL to evolve an improved reverse transcriptase for 2'OMe-RNA that would outperform the currently best performing RT enzyme SSIII.

Previous efforts in engineering DNA-templated XNA and RNA synthesis had revealed the importance of optimizing binding of the non-cognate primer-template duplex to the polymerase^{8, 19}. To this end we examined the high-resolution ternary structure of the homologous (92% sequence identity with Tgo) DNA polymerase from *T. kodakaraensis* (KOD (PDB ID 5OMF))²⁰ and identified 7 potentially suboptimal template contacts. We mutated equivalent residues in RT521K to positively charged amino acids (K, R), reasoning that increasing electrostatic interactions would be the most effective way to generally enhance polymerase binding to a non-cognate template strand. We screened 8 mutants for 2'OMe RT activity using a plate-based RT activity assay (Extended data Fig. 2). While most mutations were either neutral or deleterious, two mutations S383K and N735K both improved 2'OMe RT activity (Extended data Fig. 2) possibly by enhancing the electrostatic surface potential of the polymerase (Extended data Fig. 3).

In order to map putative polymerase regions involved in 2'OMe RT activity, we initiated CBL selections from a random mutant RT521K library. After three rounds of selection I114T, located in the uracil binding pocket (UBP)²¹, was strongly enriched, pointing to the UBP as a critical region for RT activity. We therefore also screened a further 3 mutations in the UBP or its vicinity for increased 2'OMe RT activity but none proved superior to the selected I114T mutation (Extended data Fig. 2).

Next we combined the mutations obtained by both CBL RT selection and RT activity screening in one polymerase, this yielded RT-TKK (RT521K: I114T, S383K, N735K) (Extended data Fig. 2, 3), with improved 2'OMe-RNA RT activity (Fig. 2).

Encouraged by this, we used CBL to test 7 libraries in the TKK framework to map polymerase subdomains involved in further improving RT activity (Supplementary Fig. 1). The library with greatest activity, library P, was diversified in the polymerase finger domain at positions phylogenetically variable in the polB family (library P: RT-TKK: 493, 496, 497,

499-501 chosen to relieve potential steric conflicts between the bulky 2'-methoxy groups in the 2'OMe-RNA template and the polymerase (Supplementary Fig. 2). After three rounds of CBL selection of library P, we screened for improved 2'OMe-RNA RT activity using a plate-based RT activity assay^{7, 26} (Extended data Fig. 2) and identified RT-C8, -B12 and -D5 as the best performing 2'OMe-RNA RTs (Fig. 2). In all three cases, selected mutations involve mutation of large, aromatic Phe and Tyr side-chains towards smaller, more flexible groups reducing the steric envelope of the P-helix (Fig. 2c, Supplementary Fig. 2). The most active variant RT-C8 (RT-TKK: F493V, Y496N, Y497L, Y499A, A500Q, K501H), yielded ~12-fold more full-length RT product than RT521K and 3-fold more than the best commercially available RTs SSIII and RTx (Fig. 2d) on both a defined sequence 100 nucleotide (nt) 2'OMe-RNA template and a challenging random-sequence (N₄₀) 2'OMe-RNA template (Fig. 3c).

RT-C8 derives from an inherently thermostable polymerase (Tgo) but differs from it by 18 (potentially destabilizing) mutations, yet maintained its thermostability and could also efficiently reverse transcribe a 2'OMe-RNA N₄₀ template, typically employed during aptamer selections at temperatures up to 90°C (with optimal activity up to 80°C), the highest RT temperatures ever described (Fig. 3b). Such an activity may prove advantageous e.g. for the reverse transcription of sequences with highly stable secondary structures (likely present in a N₄₀ sequence pool) as 2-OMe-RNA secondary structures are even more stable than RNA, which again forms more stable duplexes than DNA²². Such high temperature RT activity is expected to be generally useful for *in vitro* evolution experiments as cDNA synthesis on other non-canonical N₄₀ library oligonucleotide templates as many XNAs (including HNA, AtNA)²³ but also 2'MOE-RNA and others^{24, 25} show highly stable base-pairing and secondary structure formation, which can impede cDNA synthesis and *in vitro* evolution.

Finally, RT-C8 outperformed all other enzymes in RT-PCR sensitivity (S_j) on 2'OMe-RNA templates, with S_j defined as the minimal number of template molecules that can be reliably detected. While S_j can reach single-molecule levels for DNA in PCR²⁶, RT-PCR generally requires 10 - 100 (or more) copies of a target RNA²⁷ (and XNA), which limits detection sensitivity of RT-PCR assays and likely leads to an undersampling of sequence diversity during selection experiments^{5, 28}. We determined the sensitivity for (reliable) detection of 2'OMe-RNA templates (S_{2OMe}) of different RTs as judged by the Cq (cycle quantification) value in RT-qPCR. This showed that RT-C8 had a >1000-fold higher S_{2OMe} than the parental RT521K, which could not detect 2'OMe templates at levels of fewer than 10⁶ copies. Furthermore, at the lowest concentration of 2'OMe template (10³ copies) the mean Cq of RT-C8 was 31 compared to 33 for SSIII. Thus RT-C8 requires two fewer PCR cycles (corresponding to 4-fold higher S_{2OMe} sensitivity in RT-qPCR) than SSIII (Fig. 3d).

We sought to explore whether CBL could select for improved RT activity in other XNA chemistries and performed CBL selections for an HNA RT. HNA is an acid stable, nuclease resistant, XNA chemistry that has been previously shown to be capable of forming ligands (HNA aptamers)⁵ and catalysts (HNAzymes)²⁸. Compared with the promiscuous HNA RT activity previously observed for RT521K, the CBL-selected RT-H4, showed >5-fold

improved activity (Fig. 2e, as judged by the quantity of full-length cDNA product) including on challenging HNA N₄₀ templates (Supplementary Fig. 3).

Next we examined the activities of the selected RT enzymes on two “orphan” XNA chemistries, D-altritol nucleic acid (AtNA) and 2'-Methoxyethyl RNA (2'MOE-RNA)), for which previously no RT existed and which therefore could neither be reverse transcribed back into cDNA nor analysed by sequencing. We discovered that RT-TKK enabled efficient cDNA synthesis on AtNA templates²³ (Fig. 2f), which allowed a demonstration of an innate potential of AtNA random eicosamer (N₂₀) pools for spontaneous recombination akin to RNA (published elsewhere²⁹).

2'MOE-RNA, in which the 2'OH of RNA is substituted by a bulky, hydrophobic methoxyethyl group (Fig. 4a) is a chemistry widely used in therapeutic anti-sense oligonucleotides (ASO), but no RT had been available that could reverse transcribe 2'MOE-RNA templates. This may be due to the fact that it presents an even more challenging steric envelope than 2'OMe-RNA (Fig. 4b). Indeed, no available RTs (including the commercial RTs RTx or SSIII that can reverse transcribe 2'OMe-RNA) could reverse transcribe a chemically synthesized 2'MOE-RNA template. In contrast, RT-C8 was able to reverse transcribe both MOE-RNA templates and enable RT-PCR detection (Extended data Fig. 4) as well as a modified P- α -S-phosphorothioate 2'MOE-RNA (PS-2'MOE-RNA) template (Fig. 3c) comprising Nusinersen / Spinraza (an ASO composed entirely of PS-2'MOE-RNA clinically approved for the treatment of spinal muscular atrophy)³⁰ albeit with lower efficiency compared to 2'OMe, the chemistry it was evolved for.

It had recently been shown that RNA reverse transcriptases could use exonucleolytic proof reading during cDNA synthesis, which improved fidelity³¹. However, on challenging XNA templates proofreading by the polymerase 3'-5' exonuclease activity presents a kinetic barrier to forward synthesis and is therefore usually disabled by mutation (Tgo: D141A, E143A) to maximize yield. In contrast, the newly selected RT-C8 (and -H4) could synthesize full-length cDNAs on 2'OMe-RNA (and to a lesser extent on HNA) with comparable yields with the 3'-5' exonuclease activity re-activated (A141D, A143E: *exo*⁺)(Fig. 5, Supplementary Fig. 4). Furthermore, RT-C8^{exo}⁺ could edit a 3' mismatched primer on a 2'OMe-RNA template and extend to full-length cDNA product, while an exonuclease-deficient RT-C8 stalled cDNA synthesis (Fig. 5c). Deep sequencing revealed that proofreading enhanced the fidelity of cDNA synthesis on 2'OMe-RNA by ~3-fold (and >5-fold on HNA) (Fig. 5b, d, Supplementary Table 1). RT-C8^{exo}⁺ and -H4^{exo}⁺ therefore represent the first XNA-RTs with active proofreading during cDNA synthesis.

While reactivation of the exonuclease in the parent RT521K also improved fidelity (10-fold) (Fig. 5b, Supplementary Table 1, Supplementary Fig. 5), fidelity for cDNA synthesis on RNA templates remained below that of the best available RT enzymes. This prompted us to explore if CBL could also be used to evolve higher fidelity RTs.

Selection for high fidelity requires a rigorous mechanism for mismatch discrimination as mutations become rarer and rarer in the product population as the selection progresses and fidelity increases. We explored a number of strategies, but only two proved sufficiently

stringent: 1) comprising an RNA template containing a single G and a nucleotide mix replacing dCTP with the terminator ddCTP (dG/dA/dT/ddCTP). Thereby low fidelity RTs bypass template G upon misincorporation, while high fidelity RTs only incorporate ddC and terminate cDNA synthesis (Fig. 6a) allowing a positive selection for increased fidelity using a dual fluorophore detection approach; 2) comprising a DNA template devoid of dA together with a dNTP mix replacing dTTP with dUTP (dC/dG/dA/dUTP). Thereby low fidelity RTs (occasionally) misincorporate dUTP (e.g. opposite template G), while high fidelity RTs do not incorporate dUTP. This allows stringent post-synthetic discrimination as exposure to both Uracil DNA glycosylase and Endonuclease VIII (USER enzyme mix) efficiently excises dU and thereby cleaves any cDNA containing a misincorporated dU (Fig. 6b).

Using both strategies in parallel we started CBL selections from a focused library around the nucleotide binding pocket (library HF: 406 - 408, 410 - 412)(Supplementary Fig 1), which had previously been shown to contain “anti-mutator” mutations in related polB polymerases (e.g. phage T4 and yeast Pol δ , Pol ϵ) as well as the structurally distant *E. coli* Pol I & III³² hinting at a general structural determinant of polymerase fidelity. After two rounds of CBL, we screened selected RTs using a plate-based RT activity and fidelity assay (Supplementary Fig. 6), which again scored fidelity as a measure of correct ddCTP incorporation and the avoidance of dUTP misincorporation. This screen identified a range of RTs with apparent improvements in fidelity deriving from both CBL fidelity selection strategies (1) ddC: RT-E3, -C9 (2) dU: RT-H11)(Fig. 6c; Supplementary Fig. 7).

A true measurement of high fidelity requires sequencing of the cDNA products as the commonly used lacZ reversion assay is known to significantly underestimate error rates¹¹ (Supplementary Table 3). To precisely determine the fidelities of the newly selected RTs and also correct for the inherent error frequency of the sequencing process we adapted a previously described barcoded primer strategy³³, that uniquely labels every cDNA product and allows an unambiguous discrimination of errors which occurred during cDNA synthesis from those introduced during the sequencing or library workup (Supplementary Fig. 8). Using this approach, we reverse transcribed human mRNA using the selected RTs as well as commercially available RT enzymes such as M-MuLV-RT and SSIII (with known fidelity) as a benchmark and sequenced the barcoded cDNA products to a depth of 10^5 - 10^7 base calls.

For commercial RTs our method yielded fidelity values in excellent agreement with previously published values obtained by both PacBio sequencing³⁴ and Illumina sequencing^{11, 31, 35} (Supplementary Tables 1, 2, 3). This provides high confidence for the fidelities measured for the selected RTs, which indicated median error rates for the selected RTs more than 230-fold lower than RT521K (Fig. 6e, Supplementary Table 2) even in the absence of proofreading. The highest fidelity RT-H11 displayed a median error rate of 1.08×10^{-5} , which is nearly 10-fold lower than M-MuLV RT and even lower than a recently described proofreading RT (RTX)^{31, 35} (Supplementary Table 3) making RT-H11 one of the highest fidelity RTs described so far.

CBL selections also revealed two key mutations (P410T / S411R) important for fidelity (Fig. 6d). When introduced into the parent RT521K (RT-TR), these two mutations alone improved fidelity to almost the same as the highest fidelity RT H11, as well as shifting the mutational

spectrum (in particular reducing C / T > T / C and A / G > G / A transition mutations) (Fig. 6e, f, Supplementary Fig. 7), which dominate the error spectrum of M-MuLV (and derived) RTs.

Discussion

Reverse transcriptases (RTs) derived from retroviruses and other sources are often suboptimal for many applications. Therefore, RT function has been engineered by screening^{5, 36, 37, 38} and directed evolution^{18, 31, 39, 40}, but the discovery (or optimization) of RTs for XNAs or modified RNAs has remained challenging. Here we have described CBL (compartmentalized bead labelling) as a general strategy for the engineering of bespoke RT enzymes with complete freedom-of-choice of substrate / template chemistry (Fig. 1, Extended data Fig. 1). CBL comprises further key advances over alternative methods including 1) clonal on-bead capture of plasmids from a single bacterial cell (~10 copies per bead (Extended data Fig 1)), 2) efficient on-bead capture of any custom RNA (or XNA) template (requiring as little as 10 pmol per round) permitting RT selections for chemistries not (easily) accessible by solid-phase synthesis (e.g. HNA), 3) highly sensitive fluorescent readout of the cDNA signal by hybridization chain reaction (HCR) amplification (Extended data Fig. 1), 4) bead isolation by FACS with tuneable selection stringency and 5) rapid plasmid recovery and selection turnover avoiding a subcloning step by random-primed rolling circle amplification (RCA) directly from beads (Fig. 1).

We demonstrate the potential of CBL by the rapid (~3 rounds) evolution of improved RT enzymes for RNA, 2'OMe-RNA, HNA and the *de novo* discovery of first-in-class RTs for AtNA, 2' MOE and P- α -S-phosphorothioate (PS)-2'MOE-RNA. In all cases, novel RTs were not only superior to the starting enzyme (RT521K), but in the case of the 2'OMe-RNA RT-C8 superior to highly developed commercial enzymes such as SSIII and RTx with regards to yield (Fig. 2, 3c), high temperature activity (up to 90 °C)(Fig. 3b) and RT-PCR detection sensitivity (Fig. 3d). CBL has also begun to reveal key structural features enabling RNA and XNA RT activity, which will aid further RT engineering. These are pinpointed by mutations in 1) the template uracil binding pocket UBP (I114T) (Extended data Fig. 1), 2) the template binding region (S383K, N735K) (Extended data Fig. 1, 2), 3) the P-helix (Y496N, Y497L, Y499A) (Fig. 2c, Supplementary Fig. 2) and 4) the nucleotide binding pocket (P410T, S411R) (Fig. 6d), presumably acting by 1) excluding template U from the UBP preventing premature stalling, 2) increasing the affinity of the polymerase for the non-cognate RNA / XNA-DNA duplexes, 3) decreasing steric conflicts between the polymerase template binding surface and bulkier, non-cognate templates and the polymerase and 4) by increasing the accuracy of substrate selection through increased geometric constraints in the active site⁴⁶. High-fidelity RTs are of value for increased precision in the analysis of cellular transcriptomes and viral quasispecies in RNAseq as well as the faithful transmission of genetic information during cycles of *in vitro* evolution of RNA and XNA. We explored two approaches to enhance cDNA synthesis fidelity: a) exonucleolytic proofreading and b) selection for fidelity. For exonucleolytic proofreading, a large conformational change is required to unwind 4-5 base pairs to allow the 3'-end of the nascent strand to partition to the exonuclease site is located >30Å from the polymerase (primer extension) active site^{41, 42, 43}. While inhibited in the parental RT521K (presumably by the increased duplex stability of

2'OMe-RNA²² and HNA²³), reactivation of the proofreading exonuclease in RT-C8 (and to a lesser extent RT-H4) yielded RTs with an ability to excise DNA-XNA mismatches and enhance cDNA synthesis fidelity (Fig. 5c, Supplementary Fig. 4, Supplementary Table 1), the first proofreading XNA-RTs. However, even though reactivation of exonucleolytic proofreading on RNA templates (as described by others³¹) also improved fidelity (by ~10-fold (Fig. 5b, Supplementary Fig. 5, Supplementary Table 1), it remained below that of the best commercially available RTs. We therefore explored the *de novo* selection of high-fidelity cDNA synthesis by CBL which yielded novel RTs with median error rates as low as 1.1×10^{-5} (Fig. 6), the highest fidelity in cDNA synthesis described so far (over 230-fold higher than the parental RT521K and 5-10 fold higher than best commercial RT enzymes (Fig. 6e, Supplementary Table 2)) even without proofreading and even higher than a proofreading RT³¹ (Supplementary Table 3). In future, even higher fidelity RTs may be attainable, but their utility may be constrained by the fidelity of RNA synthesis and transcription (Supplementary Table 4)) and the pervasive presence of mutagenic post-transcriptional modifications in eukaryotic RNAs^{34, 44, 45}.

In conclusion, CBL is a new, general and modular strategy for directed evolution of RT function and fidelity, which should be extendable to any enzymatic activity linked to nucleic acids as well as expanding the toolbox for enzymatic transactions and forms of “transliteration” among different genetic polymers.

Methods

Detailed methods for molecular biology procedures, XNA template generation, polymerase screening and next generation sequencing are available in the Supplementary Information.

Compartmentalized bead labelling (CBL)

1 Library preparation—Mutant polymerase libraries were prepared essentially as described (ref Pinheiro). Briefly, *E. coli* cells were transformed with the polymerase library and plated on 500 cm² bioassay dishes. Transformations were scraped from the plate and were grown in 30 mL 2x TY medium at 37 °C until OD₅₉₅ ~ 0.6 was reached. Expression was induced with 0.4 µg/ml anhydrotetracycline (Sigma) for 4 h at 37°C followed by overnight incubation at 4°C. The cells were spun down (5,000 rpm) for 3 minutes at 4°C, the pellet was resuspended in 1 mL of 1x ThermoPol minus Triton X-100 buffer (1x TP-T: 20 mM Tris-HCl, 10 mM (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, pH 8.8) (New England Biolabs (NEB)), spun down at 5,000 rpm for 3 min, resuspended in 500 µl 1x TP-T, spun down and finally resuspended in 200 µl 1x TP-T.

2 Bead preparation—40 pmol each of XNA / RNA template and dual-biotinylated FD or test7 primers (Supplementary Table 5) were annealed (50 mM NaCl, 1x ThermoPol (TP) buffer (20 mM Tris-HCl pH 8.8, 10 mM (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, 0.1% (v/v) Triton X-100,)) at 80°C for 2 min in 40 µl reaction volumes and cooled slowly at 0.1°C/sec to 4°C. Streptavidin coated paramagnetic beads (~5 × 10⁷) (M-280 Streptavidin dynabeads Thermo Fisher Scientific) were washed twice in 1x bead binding and wash buffer (BWBS) (10 mM Tris.HCl pH 7.4, 1 M NaCl, 0.1% v/v Tween 20, 1 mM EDTA) buffer. The annealed XNA / RNA template and biotinylated primer duplexes were bound to M-280

streptavidin beads in 1x BWBS buffer along with 10 pmol of dual biotinylated capture oligo, bbcapture (Supplementary Table 7), rotating for 1 hour at room temperature or overnight at 4°C. The beads were washed in BWBS, BWBT (low salt BWBS) (10 mM Tris-HCl, pH 7.4, 20 mM NaCl, 1 mM EDTA, 0.1% v/v Tween 20) and finally 1x TP buffer.

3 Emulsion preparation and selection— 2.10^8 induced cells in 1x TP-T (see above) were added to the aqueous phase (150 μ l) containing 0.125 mM dNTPs, 2 mM MgSO_4 , 1 mM DTT, glycerol (10% v/v), formamide (2% v/v), 1.5 $\mu\text{g}/\mu\text{l}$ yeast tRNA (Roche) in 1x TP buffer with $\sim 5 \times 10^7$ streptavidin beads coated with primer, XNA/RNA template and capture oligonucleotide (see above) and were emulsified as described⁴⁶. Emulsions were transferred to PCR tubes with 50 μ l of emulsion in each well. Emulsions were heated to 75°C for 3 minutes to lyse *E. coli* in emulsion and RT reaction was allowed to proceed according to the detailed selection parameters listed in Supplementary Table 11. After the RT reaction was complete, beads were extracted from emulsions, by pooling emulsions in a 2 ml Eppendorf tube, 500 μ l of 100% glycerol was added, mixed thoroughly by vortexing at full speed (~ 30 sec) and spun at 13,000 rpm (13 krpm) for 1 min. The top oil layer was carefully removed and 500 μ l of Vogelstein break buffer (10 mM Tris-HCl (pH 7.4), 1% v/v Triton-X 100, 1% SDS, 100 mM NaCl, 1 mM EDTA) was added, mixed by vortexing and spun at 13 krpm for 1 min. The Eppendorf tube was turned 180 degrees in the centrifuge and spun again at 13 krpm for 1 min. The top oil layer was again removed, 500 μ l of BWBT was added, mixed by vortexing and spun at 13 krpm for 1 min. The beads were captured on a magnet and consecutively washed in 500 μ l BWBT, 500 μ l BWBS, 500 μ l TBT2 (10 mM Tris.HCl pH 7.4, 20 mM NaCl, 0.1% v/v Tween 20, 0.1 mg/ml BSA) and finally 500 μ l BWBT.

4 Template removal from beads—For CBL selections containing templates with a section of RNA at the 5' end, the RNA section of the template was degraded for probing of the cDNA product. Beads were washed once in 1x RNase H buffer (NEB), resuspended in 1x RNase buffer with 1.25 units of RNase A, 5 units of RNase T1 (Thermo Fisher Scientific), 1.25 units RNase H (NEB) in 50 μ l volume per $\sim 5 \times 10^7$ beads and incubated at 37°C for 2 h rotating in a hybridization oven. For CBL fidelity selections using a template devoid of A's USER digestion was performed to remove low fidelity clones. Beads were washed once in 1x Cutsmart buffer (NEB), resuspended in 50 μ l of 1x Cutsmart buffer with 4 units USER (NEB) and incubated at 37°C overnight. The beads were washed in 1x lambda exonuclease buffer and the 5' phosphorylated DNA template was removed using lambda exonuclease (NEB) at 37°C for 1 h rotating in a hybridization oven.

Hybridization Chain Reaction (HCR) on beads, sorting and plasmid amplification

HCR initiator template (10 pmol) was diluted in 50 μ l of 50 mM NaCl, 10 mM Tris, Tween 0.1% and used to resuspend beads and incubated rotating overnight at room temperature. The beads were washed with 2x in BWBT and 1x in 5x SSCT (5xSSC, 0.1% v/v Tween 20). Alexa647 labelled hairpin 1 and hairpin 2 (Supplementary Table 8) were annealed at 3 μM in 5xSSC at 95°C for 90 sec and slowly cooled to room temperature for 30 min. 1 μ l of each hairpin was added to 48 μ l amplification buffer (5x SSC, 0.1% Tween20, 0.1% dextran sulfate). Beads were washed twice in 50 μ l 5x SSCT and 50 μ l of the hairpin mix was added

to beads and rotated at room temperature for 4-6 h. Beads were washed twice in 5x SSCT, and finally resuspended in PBST (PBS, 0.1% Tween 20) for FACS. For RNA fidelity selections, a dual probe and HCR approach was adopted whereby 10 pmol of FITC labelled probe was bound to beads in 10 mM Tris.HCl, 50 mM NaCl and 0.1% Tween20. Beads were washed 2x in BWBT before adding the initiator and performing the HCR as described above.

Beads were sorted using a Synergy high speed sorter (Sony). The total number of sorted beads varied from 10^4 - 10^6 depending on the gate applied. Sorted beads were spun down for 30 minutes at 13 krpm. The supernatant was discarded and beads resuspended in 5 μ l elution buffer (10 mM Tris-HCl, pH 8.5) and plasmid were amplified by addition of beads to a phi29 amplification reaction using random RNA 10mer primers a with 5' thiol group (SH-N10). 100 μ M of SH-N10 (Supplementary Table 7) were annealed to the plasmids at 80°C for 3 minutes in phi29 DNA polymerase reaction buffer (NEB). Phi29 polymerase (10 units) (NEB), 0.1 mg/ml BSA and 0.25 mM dNTPs were added to the reaction (20 μ l final volume) and incubated at 30°C for 16 h followed by heat inactivation for 10 min at 65°C. DNA concatamer products were digested with NdeI (NEB) for 3 h at 37°C. The digested plasmid band was gel extracted and purified using a PCR purification column (Bioline), the products were ligated for 2 h at 22°C, purified using a PCR purification column (Bioline) and transformed into 10-beta electrocompetent cells (NEB). Transformed cells were used in the following round of selection and / or stored as a glycerol stock at -80°C.

Extended Data

Extended Data Fig. 1. CBL model selections, signal amplification and plasmid capture.

a, Cells expressing RT521K were spiked into an excess of cells expressing inactive RT (RT-DTD) at 1:100 (top) and 1:1000 (bottom) ratios and encapsulated in w/o emulsion with beads coated with DNA primer bbFd and template TnotTest7 (Supplementary Table 7). After in-emulsion cell lysis and primer extension, beads were recovered and sorted for fluorescence using FACS (middle panel, dashed box). Plasmids recovered from sorted beads

were amplified by qPCR and the amount of RT521K and RT- DTD plasmids post-CBL selection quantified. Enrichment was calculated by determining the ratio of RT521K: RT- DTD plasmids in the input (before CBL selection) (left panels) and comparing it with the output ratio (after CBL selection)(right panels) yielding enrichment factors of ~500-fold in both cases. **b**, HCR signal amplification. Cells expressing RT521K were encapsulated with beads coated with primer bbTest7 annealed to RNA template tnog_full (Supplementary Table 7) and RNA reverse transcription was performed in emulsion. After recovering the beads from emulsion, the sample was split in half and beads were treated with either a nucleic acid probe (top bead) or underwent HCR (bottom bead). Flow cytometry analysis of the beads labelled with a direct probe (top plot) or HCR (bottom plot) shows approximately an order of magnitude increase in signal and a greater percentage of fluorescently labelled beads in HCR conditions. **c**, Plasmid capture on microbeads. Plasmids bound to beads during CBL were quantified by qPCR. Cells expressing RT521K were combined with beads in bulk (in solution) or were encapsulated in w/o emulsions and underwent reverse transcription. Beads in w/o emulsion were extracted and bulk and emulsion treated beads were sorted for fluorescence by HCR. The number of plasmids per bead was quantified before and after FACS. Purified plasmid was bound to beads without capture oligo (untreated beads) and with capture oligo as controls showing approximately 10 plasmids captured per bead in emulsion and stably bound (surviving post-sort).

Extended Data Fig. 2. RT mutation screen.

a, Space filling surface model of the ternary structure of KOD polymerase (PDB ID 5OMF) with primer (nascent) strand (red) and template strand (green) shown. The position of

mutations selected for screening are shown in blue. **b**, ELONA (enzyme-linked oligonucleotide assay)-based RT activity assay scheme: (from left to right) RT reactions are performed with a biotinylated primer bbTest7 (Supplementary Table 7), bound to wells in a streptavidin plate and hybridized to 2'-O-Me RNA template (cyan) TFRst 2'-O-Me (Supplementary Table 7). RT synthesized cDNA (red), which remains bound to the plate after template removal. The presence of the (correct) cDNA can be detected by a specific oligonucleotide probe FitcFd (Supplementary Table 7) labelled with FITC (green), which in turn is detected by an anti-FITC antibody (blue) conjugated to horse-radish peroxidase (yellow star). **c**, RT mutation activity screen: only mutations S383K, N735K and I114T show an improved signal and when combined (RT-TKK) show more than double the signal of wt (RT521K) (NTC, no template negative control, n'=3).

Extended Data Fig. 3. RT-TKK: mutations and electrostatic surface.

a, Sequence alignment of RT521K and RT-TKK with mutations shown in blue. **b**, Space filling model of the ternary structure of KOD polymerase (PDB ID 5OMF) with RT-TKK mutations I114T, S383K and N735K (blue), primer strand (red), template (green) and incoming deoxynucleotide triphosphate (orange). **c**, Zoom in of the uracil binding pocket (UBP) with uracil base bound (PDB ID 2VWJ) and V93Q mutation (orange). (present in both RT-521K and RT-TKK) and I114T mutation (blue spheres) (RT-TKK). Note how V93Q narrows the UBP (compared to wild-type UBP shown in Extended Data Fig. 1a) and sterically excludes uracil from the binding pocket. The mechanistic basis of the I114T mutation improvement of cDNA synthesis on 2'-O-Me RNA is currently unclear. The main chain NH of I114 hydrogen bonds with uracil in the wild-type UBP. Mutation to I114T may alter main chain conformation to disrupt this interaction and this may further improve uracil exclusion. **d**, Electrostatic potential of the primer/template binding surface of KOD (left) and its change upon N735K and S383K mutations. Note the increase of positively charged surface potential in proximity to the template strand (green), which is likely to enhance template binding.

Extended Data Fig. 4. A reverse transcriptase for 2'-MOE RNA.

a, Chemical structure of 2'-MOE RNA. **b**, denaturing urea-PAGE of cDNA synthesis on 2'-MOE RNA template (sequence: 10–27 Spinraza (Supplementary Table 7)) and **c**, RT-PCR of cDNA synthesis. Only RT-C8 can reverse transcribe 2'-MOE RNA (as well as PS 2'-MOE RNA (Fig. 4)).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the Medical Research Council (MRC) program grant program no. MC_U105178804 (PH, GH, AIT), the Biotechnology and Biological Sciences Research Council (BBSRC) UK (09-EuroSYNBIO-OP-013) (BTP), and a research collaboration between AstraZeneca UK Limited and the Medical Research Council-MRC-Astra Zeneca Blue Sky Grant (SA-F, NS).

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

Custom scripts used for RT fidelity analysis in this study can be found at GitHub: <https://github.com/holliger-lab/fidelity-analysis>

References

1. White AK, VanInsberghe M, Petriv OI, Hamidi M, Sikorski D, Marra MA, et al. High-throughput microfluidic single-cell RT-qPCR. *Proc Natl Acad Sci USA*. 2011; 108(34):13999–14004. [PubMed: 21808033]
2. Stark R, Grzelak M, Hadfield J. RNA sequencing:the teenage years. *Nat Rev Genet*. 2019; 20(11):631–656. [PubMed: 31341269]
3. Schaffitzel C, Hanes J, Jermutus L, Pluckthun A. Ribosome display:an in vitro method for selection and evolution of antibodies from libraries. *J Immunol Methods*. 1999; 231(1-2):119–135. [PubMed: 10648932]
4. Taylor AI, Houlihan G, Holliger P. Beyond DNA and RNA:The Expanding Toolbox of Synthetic Genetics. *Cold Spring Harb Perspect Biol*. 2019; 11(6)
5. Pinheiro VB, Taylor AI, Cozens C, Abramov M, Renders M, Zhang S, et al. Synthetic genetic polymers capable of heredity and evolution. *Science*. 2012; 336(6079):341–344. [PubMed: 22517858]
6. Yu H, Zhang S, Chaput JC. Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor. *Nat Chem*. 2012; 4(3):183–187. [PubMed: 22354431]
7. Biondi E, Lane JD, Das D, Dasgupta S, Piccirilli JA, Hoshika S, et al. Laboratory evolution of artificially expanded DNA gives redesignable aptamers that target the toxic form of anthrax protective antigen. *Nucleic Acids Res*. 2016; 44(20):9565–9577. [PubMed: 27701076]
8. Arangundy-Franklin S, Taylor AI, Porebski BT, Genna V, Peak-Chew S, Vaisman A, et al. A synthetic genetic polymer with an uncharged backbone chemistry based on alkyl phosphonate nucleic acids. *Nat Chem*. 2019; 11(6):533–542. [PubMed: 31011171]
9. Kotewicz ML, Dalessio JM, Driftmier KM, Blodgett KP, Gerard GF. Cloning and Overexpression of Moloney Murine Leukemia-Virus Reverse-Transcriptase in Escherichia-Coli. *Gene*. 1985; 35(3):249–258. [PubMed: 2412939]
10. Ong, JL; Evans, JT; Tanner, N. U.S. Patent WO 2013/033528. 2013.
11. Zhao C, Pyle AM. Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution. *Nat Struct Mol Biol*. 2016; 23(6):558–565. [PubMed: 27136328]
12. Dunn MR, Chaput JC. Reverse Transcription of Threose Nucleic Acid by a Naturally Occurring DNA Polymerase. *Chembiochem*. 2016; 17(19):1804–1808. [PubMed: 27383648]
13. Wang Y, Ngor AK, Nikoomanzar A, Chaput JC. Evolution of a General RNA-Cleaving FANA Enzyme. *Nat Commun*. 2018; 9(1):5067. [PubMed: 30498223]
14. Choi HM, Beck VA, Pierce NA. Next-generation in situ hybridization chain reaction:higher gain, lower cost, greater durability. *ACS Nano*. 2014; 8(5):4284–4294. [PubMed: 24712299]
15. Ayadi L, Galvanin A, Pichot F, Marchand V, Motorin Y. RNA ribose methylation (2'-O-methylation):Occurrence, biosynthesis and biological functions. *Bba-Gene Regul Mech*. 2019; 1862(3):253–269.
16. Burmeister PE, Lewis SD, Silva RF, Preiss JR, Horwitz LR, Pendergrast PS, et al. Direct in vitro selection of a 2'-O-methyl aptamer to VEGF. *Chem Biol*. 2005; 12(1):25–33. [PubMed: 15664512]
17. Voit T, Topaloglu H, Straub V, Muntoni F, Deconinck N, Campion G, et al. Safety and efficacy of drisapersen for the treatment of Duchenne muscular dystrophy (DEMAND II):an exploratory,

- randomised, placebo-controlled phase 2 study. *Lancet Neurol.* 2014; 13(10):987–996. [PubMed: 25209738]
18. Chen T, Hongdilokkul N, Liu Z, Adhikary R, Tsuen SS, Romesberg FE. Evolution of thermophilic DNA polymerases for the recognition and amplification of C2'-modified DNA. *Nat Chem.* 2016; 8(6):556–562. [PubMed: 27219699]
 19. Cozens C, Pinheiro VB, Vaisman A, Woodgate R, Holliger P. A short adaptive path from DNA to RNA polymerases. *Proc Natl Acad Sci U S A.* 2012; 109(21):8067–8072. [PubMed: 22566643]
 20. Kropp HM, Betz K, Wirth J, Diederichs K, Marx A. Crystal structures of ternary complexes of archaeal B-family DNA polymerases. *Plos One.* 2017; 12(12)
 21. Firbank SJ, Wardle J, Heslop P, Lewis RJ, Connolly BA. Uracil recognition in archaeal DNA polymerases captured by X-ray crystallography. *J Mol Biol.* 2008; 381(3):529–539. [PubMed: 18614176]
 22. Rozners E, Moulder J. Hydration of short DNA, RNA, and 2-OMe oligonucleotides determined by osmotic stressing. *Nucleic Acids Res.* 2004; 32(20):6153–6153.
 23. Allart B, Khan K, Rosemeyer H, Schepers G, Hendrix C, Rothenbacher K, et al. D- Altritol Nucleic Acids (ANA):Hybridisation Properties, Stability, and Initial Structural Analysis. *Chemistry.* 1999; 5(8):2424–2431.
 24. Freier SM, Altmann KH. The ups and downs of nucleic acid duplex stability:structure-stability studies on chemically-modified DNA:RNA duplexes. *Nucleic Acids Res.* 1997; 25(22):4429–4443. [PubMed: 9358149]
 25. Wilds CJ, Damha MJ. Duplex recognition by oligonucleotides containing 2'-deoxy-2'-fluoro-D-arabinose and 2'-deoxy-2'-fluoro-D-ribose. Intermolecular 2'-OH-phosphate contacts versus sugar puckering in the stabilization of triple-helical complexes. *Bioconjugate Chem.* 1999; 10(2):299–305.
 26. Ohuchi S, Nakano H, Yamane T. In vitro method for the generation of protein libraries using PCR amplification of a single DNA molecule and coupled transcription/translation. *Nucleic Acids Res.* 1998; 26(19):4339–4346. [PubMed: 9742233]
 27. Okello JB, Rodriguez L, Poinar D, Bos K, Okwi AL, Bimenya GS, et al. Quantitative assessment of the sensitivity of various commercial reverse transcriptases based on armored HIV RNA. *Plos One.* 2010; 5(11):e13931. [PubMed: 21085668]
 28. Taylor AI, Pinheiro VB, Smola MJ, Morgunov AS, Peak-Chew S, Cozens C, et al. Catalysts from synthetic genetic polymers. *Nature.* 2015; 518(7539):427–430. [PubMed: 25470036]
 29. Mutschler H, Taylor AI, Porebski BT, Lightowlers A, Houlihan G, Abramov M, et al. Random-sequence genetic oligomer pools display an innate potential for ligation and recombination. *Elife.* 2018; 7:e43022. [PubMed: 30461419]
 30. Ottesen EW. Iss-N1 Makes the First Fda-Approved Drug for Spinal Muscular Atrophy. *Transl Neurosci.* 2017; 8(1):1–6. [PubMed: 28400976]
 31. Ellefson JW, Gollihar J, Shroff R, Shivram H, Iyer VR, Ellington AD. Synthetic evolutionary origin of a proofreading reverse transcriptase. *Science.* 2016; 352(6293):1590–1593. [PubMed: 27339990]
 32. Herr AJ, Williams LN, Preston BD. Antimutator variants of DNA polymerases. *Crit Rev Biochem Mol.* 2011; 46(6):548–570.
 33. Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA. Detection of ultra-rare mutations by next-generation sequencing. *Proc Natl Acad Sci U S A.* 2012; 109(36):14508–14513. [PubMed: 22853953]
 34. Potapov V, Fu XQ, Dai N, Correa IR, Tanner NA, Ong JL. Base modifications affecting RNA polymerase and reverse transcriptase fidelity. *Nucleic Acids Res.* 2018; 46(11):5753–5763. [PubMed: 29750267]
 35. Yasukawa K, Iida K, Okano H, Hidese R, Baba M, Yanagihara I, et al. Next-generation sequencing-based analysis of reverse transcriptase fidelity. *Biochem Bioph Res Co.* 2017; 492(2):147–153.
 36. Baranauskas A, Paliksa S, Alzbutas G, Vaitkevicius M, Lubiene J, Letukiene V, et al. Generation and characterization of new highly thermostable and processive M-MuLV reverse transcriptase variants. *Protein Eng Des Sel.* 2012; 25(10):657–668. [PubMed: 22691702]

37. Arezi B, Hogrefe H. Novel mutations in Moloney Murine Leukemia Virus reverse transcriptase increase thermostability through tighter binding to template-primer. *Nucleic Acids Res.* 2009; 37(2):473–481. [PubMed: 19056821]
38. Sauter KB, Marx A. Evolving thermostable reverse transcriptase activity in a DNA polymerase scaffold. *Angew Chem Int Ed Engl.* 2006; 45(45):7633–7635. [PubMed: 17054304]
39. Skirgaila R, Pudzaitis V, Paliksa S, Vaitkevicius M, Janulaitis A. Compartmentalization of destabilized enzyme-mRNA-ribosome complexes generated by ribosome display: a novel tool for the directed evolution of enzymes. *Protein Eng Des Sel.* 2013; 26(7):453–461. [PubMed: 23667164]
40. Ong JL, Loakes D, Jaroslawski S, Too K, Holliger P. Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. *J Mol Biol.* 2006; 361(3):537–550. [PubMed: 16859707]
41. Franklin MC, Wang JM, Steitz TA. Structure of the replicating complex of a pol alpha family DNA polymerase. *Cell.* 2001; 105(5):657–667. [PubMed: 11389835]
42. Reha-Krantz LJ. DNA polymerase proofreading: Multiple roles maintain genome stability. *Bba-Proteins Proteom.* 2010; 1804(5):1049–1063.
43. Gouge J, Ralec C, Henneke G, Delarue M. Molecular Recognition of Canonical and Deaminated Bases by P. abyssus Family B DNA Polymerase. *J Mol Biol.* 2012; 423(3):315–336. [PubMed: 22902479]
44. Tzelepis K, Rausch O, Kouzarides T. RNA-modifying enzymes and their function in a chromatin context. *Nat Struct Mol Biol.* 2019; 26(10):858–862. [PubMed: 31582848]
45. Zhou HQ, Rauch S, Dai Q, Cui XL, Zhang ZJ, Nachtergaele S, et al. Evolution of a reverse transcriptase to map N-1-methyladenosine in human messenger RNA. *Nat Methods.* 2019; 16(12):1281–1288. [PubMed: 31548705]
46. Diehl F, et al. BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nat Methods.* 2006; 3:551–559. [PubMed: 16791214]

Fig. 1. CBL selection scheme.

i) A library of *E. coli* cells expressing different RT mutants (top) and microbeads (bottom) displaying plasmid-capture oligonucleotides (curved line), pre-bound primer (half arrow) and XNA templates (green) with a 5'-RNA end (black) are co-encapsulated ii) with RT reagents (dNTPs, buffer) in the aqueous compartments of a water-in-oil (w/o) emulsion. Cell lysis releases both the RT protein (spheres) and its encoding plasmid into the lumen of the w/o compartment, where iii) the RT is challenged to reverse transcribe the XNA template, and plasmids are captured, creating a population of clonal beads comprising RT genes linked to their cDNA products. While the red RT can reverse transcribe the XNA template (cDNA: red dashed line), the cyan RT is inactive on the XNA template. iv) After emulsion-breaking and template removal (5'-RNA end), a complementary initiator strand (**a** (purple)) is annealed to the synthesized cDNA and amplified by the hybridization chain reaction (HCR) using mutually complementary fluorescent hairpins **a'-b** (black) and **b'-a** (grey), which v) allows bead sorting by FACS with fluorescence proportional to the extent of cDNA synthesis and vi) amplification of bead-bound plasmids (and encoded RT genes) using random-primed rolling circle amplification (RCA) and retransformation for another round of selection (vi -> i). CBL thus links the phenotype / function of an RT (cDNA synthesis) directly to its genotype (encoding plasmid).

Fig. 2. XNA reverse transcriptases.

a, Sequence alignment of the first-generation RT-521K and CBL evolved reverse transcriptases RT-TKK, -C8, -H4 (mutations in blue, purple, green). **b**, Space filling model of the ternary structure of KOD DNA polymerase (PDB ID 5OMF) with mutations identified in RT-TKK in blue, and region of library P (from which RT-C8, -H4 derive) in purple. **c**, Ribbon model of active site of KOD with key mutations in RT-C8 displayed in purple (sticks) with wild-type side-chain residues shown as space-filling envelope highlighting the reduction in steric bulk upon mutation of four Y residues to smaller side-chains. **d-f**, Chemical structures (left panel) and denaturing urea-PAGE (right) of reverse transcription products (red) from templates (green): **(d)** 2'OMe-RNA (template: TFR7, primer: Cy5fd (Supplementary Table 7) 65°C / 1h). While Superscript III RT (and to a lesser

extent RTx and RT-TKK) show some 2'OMe-RNA RT activity, RT-C8 yields 3-fold more full length 80 nt cDNA product. **(e)** hexitol nucleic acids (HNA template: TempNpure, primer: Cy5test7 (Supplementary Table 7) 65°C / 1h). The selected RT-H4 yields full length 72 nt cDNA product. **(f)** D-altritol nucleic acids (AtNA) template: HR16_1, primer: Cy5test7 (Supplementary Table 7) (65°C / 12 h). Only the new RT-TKK can reverse transcribe this orphan XNA chemistry and yield full-length cDNA product.

Fig. 3. Characterization of 2'OMe-RNA reverse transcriptase C8.

a, Sequence alignment of the parental RT-521K and engineered 2'OMe-RNA reverse transcriptases RT-TKK, -D5, -B12, -C8. **b**, cDNA synthesis by RT-C8 on a challenging 2'OMe-RNA N₄₀ template (typically used for aptamer selections (N40OMelib, Supplementary Table 7)) as a function of temperature (x °C / 1h). Near 100% yield is obtained at temperatures as high as 80 °C. **c**, Short burst cDNA synthesis on the same 2'OMe-RNA template (65 °C / 10 min) showing RT-C8 has superior cDNA synthesis activity on 2'OMe-RNA compared to starting RT-521K and other selected RTs (RT-TKK, -D5, -B12). **d**, 2'OMe-RNA RT-qPCR sensitivity ($S_{2\text{OMe}}$) (as Cq (cycle threshold value : qPCR cycle number when PCR product fluorescence can be detected above the background signal)) of parent (RT521K), engineered (RT-TKK, RT-C8) and commercial RT (Superscript III (SS III)) as a function of template copy number (2'OMe-RNA template H9 (Supplementary Table 9)). While no product can be detected at $<10^6$ 2'OMe templates for RT521K, as few as 10^3 2'OMe-RNA template molecules could reliably be detected by both RT-C8 and SSIII, with RT-C8 showing ca. 4-fold higher $S_{2\text{OMe}}$ as judged by mean Cq (error bars show standard deviation, n=12)(Extended data Table 1).

Nat Chem. Author manuscript; available in PMC 2021 February 01.

PS-2'MOE-RNA template 10-27PS (red) and primer Test7 (Supplementary Table 7) sequence and denaturing urea-PAGE of cDNA synthesis on PS-2'MOE-RNA template sequence: Spinraza ASO (underlined)) (left panel) and RT-PCR of cDNA synthesis (65°C / 2 h). Only RT-C8 can reverse transcribe PS-2'MOE-RNA (as well as 2'MOE-RNA templates (Extended data Figure 3)).

Fig. 5. Proofreading XNA RTs.

a, Sequence alignment of parent RT-521K and reverse transcriptases RT521Kexo⁺, RT-C8 and -C8exo⁺ (exonuclease reactivating mutations in red). **b**, Error frequencies of RNA and XNA RTs and improvement with exonucleolytic proofreading (right). Reactivation of the 3'-5' exonuclease improves cDNA synthesis fidelity on both RNA and XNA templates. **c**, Left panel: cDNA synthesis activity of parent RT521K, -521Kexo⁺, RT-C8 and -C8exo⁺ on a 2'OMe-RNA template with a matched DNA primer (TFRst 2'OMe / Cy5Test7, Supplementary Table 7)(left panel) or a 3'-A:G mismatched primer (Cy5Test7mismatchA Supplementary Table 7) (50°C 1 min, 65°C 2 min)(right panel). Only RT-C8exo⁺ can correct the mismatched primer and extend to the end of the 2'OMe template. **d**, Error spectrum of cDNA synthesis on a 2'OMe RNA template for RT-C8 and RT-C8exo⁺.

a-b, High-fidelity RT selection strategies. **a**, Selection against misincorporation and mismatch extension opposite template dG (left panel, from top): i) high-fidelity (HF) RTs (blue sphere) terminate cDNA synthesis after correct ddCTP insertion and the resulting truncated cDNA therefore only binds to a FITC-probe (green star). In contrast low-fidelity (LF) RTs (grey sphere) perform an error-prone readthrough and synthesize full length cDNA, which binds both FITC- and Alexa647-probe (yellow star) enabling ii) gating and isolation of HF-RT beads by 2-colour FACS. iii) Denaturing urea-PAGE gel of RT reactions (50°C 2h / RNA template tnoG_full ; primer Cy5Fd (Supplementary Table 7)). RT521K extends to full length cDNA, while HF RT-E3, -H11 and -TR stall after correct ddCTP incorporation. **b**, Selection against misincorporation opposite template dG, dC and dT (right panel from top): i) on template lacking dA low-fidelity (LF) RTs misincorporate dUTP resulting and fragmentation of cDNA by Uracil DNA glycosylase and the DNA glycosylase-lyase Endonuclease VIII (USER enzyme mix). In contrast, HF RTs do not misincorporate dUTP and therefore cDNA is not cleaved and can bind Alexa647 probe allowing ii) selective isolation of HF RTs by FACS. iii) Denaturing urea-PAGE gel of RT reactions (50°C 1h /

DNA template TnotTest7; primer Cy5Fd (Supplementary Table 7)) with and without USER digestion. RT521K cDNA synthesis product is fragmented by USER, while HF RT products remain intact. **c**, Sequence alignment of RT-521K and high-fidelity reverse transcriptases RT-E3, -H11, -TR (selected mutations in cyan, TR mutation motif in magenta). **d**, High-fidelity mutations P410T and S411R (magenta (RT-TR)) in the nucleotide binding pocket of KOD polymerase (PDB 5OMF) are shown as sticks with diversified residues (library HF) shown as a space-filling envelope (cyan) with finger subdomain (yellow). Note the increased steric bulk introduced by the HF mutations. **e-f**, cDNA synthesis fidelity of parent RT521K, RT-H11, RT-TR and the commercial RTs M-MuLV RT and Superscript III (SSIII) as median error frequency (**e**) and mutational profiles (**f**) (mean error frequencies, see Supplementary Tables 2, 3), showing a >200-fold improvement of fidelity over starting RT-521K and a >4-fold improvement over SSIII.