Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. — Source link ↗

Farren J. Isaacs, Peter A. Carr, Harris H. Wang, Marc J. Lajoie ...+14 more authors

Institutions: Harvard University, Massachusetts Institute of Technology, Stanford University, Yonsei University ...+1 more institutions

Topics: Genome, Stop codon, Genome engineering, Silent mutation and Genetic code

Related papers:

- Programming cells by multiplex genome engineering and accelerated evolution

- Genomically Recoded Organisms Expand Biological Functions

- Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome

- One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products

- RNA-guided editing of bacterial genomes using CRISPR-Cas systems

The following resources related to this article are available online at
www.sciencemag.org (this infomation is current as of July 14, 2011):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/content/333/6040/348.full.html

Supporting Online Material can be found at:
http://www.sciencemag.org/content/suppl/2011/07/13/333.6040.348.DC1.html

This article **cites 35 articles**, 25 of which can be accessed free:
http://www.sciencemag.org/content/333/6040/348.full.html#ref-list-1

This article appears in the following **subject collections:**
Genetics
http://www.sciencemag.org/cgi/collection/genetics

The enhancement of blinking by increased laser power suggested that blinking might form part of the stress response mechanism in *E. coli*. We thus tested whether blinking was associated with cationic efflux, another important mechanism of stress response.

We observed surprising dynamics of a cationic membrane-permeable dye, tetramethyl rhodamine methyl-ester (TMRM), in blinking *E. coli*. As expected for this Nernstian voltage indicator (*16*), TMRM gradually accumulated in the cytoplasm over ~10 min. However, blinks in PROPS fluorescence coincided with precipitous stepwise drops in TMRM fluorescence that showed little or no recovery after the blink (Fig. 4A). The duration of the step in TMRM fluorescence coincided with the duration of the blink: At moderate intensities of red illumination ($I = 10$ W/cm^2) steps lasted less than 200 ms, whereas under little or no red illumination steps typically lasted several seconds (fig. S9). Stepwise disappearance of TMRM was also observed in cells without the PROPS plasmid, when only dim green illumination was used to image the TMRM (30 mW/cm^2; Fig. 4B and fig. S10). The duration of these steps was comparable to that of steps in cells with PROPS under dim red illumination (2 W/cm^2). The rapid disappearance of TMRM during a blink suggested an active-transport mechanism. Dissipation of V_m lowers the thermodynamic barrier to cationic efflux (Fig. 4C) (*6*). A concurrent dissipation of V_m and increase in membrane permeability would be sufficient to induce cationic efflux. PMF-dependent efflux of other cationic dyes has been observed in *E. coli* (*17*) in population-level assays that are insensitive to the dynamics of individual cells.

Bacterial electrophysiology is likely to differ in several key regards from its eukaryotic version due to the comparatively small surface area, yet high surface-to-volume ratio found in bacteria. With a typical membrane capacitance between 10^{-14} and 10^{-13} F, a single ion channel with a conductivity of 100 pS can alter the membrane potential with a time constant of 10^{-3} to 10^{-4} s. In contrast, neurons only fire through the concerted action of a large number of ion channels. Thus, bacterial electrophysiology is likely to be dominated by stochastic opening of individual ion channels and pores. Additionally, the ionic composition of bacteria is less robust than that of eukaryotes. A bacterium with a volume of 1 fl and a cytoplasmic Na$^+$ concentration of 10 mM contains only ~10^7 ions of Na$^+$. A single ion channel passing a current of 2 pA can deplete this store in less than 1 s. These simple estimates show that some of the tenets of neuronal electrophysiology may need rethinking in the context of bacteria.

References and Notes
1. B. Martinac, Y. Saimi, C. Kung, *Physiol. Rev.* **88**, 1449 (2008).
2. O. Béjà, E. N. Spudich, J. L. Spudich, M. Leclerc, E. F. DeLong, *Nature* **411**, 786 (2001).
3. O. Béjà *et al.*, *Science* **289**, 1902 (2000).
4. J. L. Spudich, C. S. Yang, K. H. Jung, E. N. Spudich, *Annu. Rev. Cell Dev. Biol.* **16**, 365 (2000).
5. P. Kolodner, E. P. Lukashev, Y. C. Ching, D. L. Rousseau, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11618 (1996).
6. Materials and methods are available as supporting material on *Science* Online.
7. A. K. Dioumaev, J. M. Wang, Z. Bálint, G. Váró, J. K. Lanyi, *Biochemistry* **42**, 6582 (2003).
8. G. Puchiar, T. Kotnik, D. Miklavčič, *J. Vis. Exp.* **33**, 1659 (2009).
9. G. Miesenböck, D. A. De Angelis, J. E. Rothman, *Nature* **394**, 192 (1998).
10. D. C. Fung, H. C. Berg, *Nature* **375**, 809 (1995).
11. D. Novo, N. G. Perlmutter, R. H. Hunt, H. M. Shapiro, *Cytometry* **35**, 55 (1999).
12. G. D. Sprott, W. G. Martin, H. Schneider, *Photochem. Photobiol.* **24**, 21 (1976).
13. T. Noumi, M. Maeda, M. Futai, *FEBS Lett.* **213**, 381 (1987).
14. A. K. Joshi, S. Ahmed, G. Ferro-Luzzi Ames, *J. Biol. Chem.* **264**, 2126 (1989).
15. W. J. Ingledew, R. K. Poole, *Microbiol. Rev.* **48**, 222 (1984).
16. C. J. Lo, M. C. Leake, T. Pilizota, R. M. Berry, *Biophys. J.* **93**, 294 (2007).
17. J. A. Bohnert, B. Karamian, H. Nikaido, *Antimicrob. Agents Chemother.* **54**, 3770 (2010).

Supporting Online Material
www.sciencemag.org/cgi/content/full/333/6040/345/DC1
Materials and Methods
SOM Text
Figs. S1 to S11
Tables S1 to S6
References (*18–33*)
Movies S1 to S7

24 February 2011; accepted 31 May 2011
10.1126/science.1204763

Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement

Farren J. Isaacs,[1]*†‡ Peter A. Carr,[2,3]*‡§ Harris H. Wang,[1,4,5,6]* Marc J. Lajoie,[1,7] Bram Sterling,[2,3] Laurens Kraal,[1] Andrew C. Tolonen,[1] Tara A. Gianoulis,[1,6] Daniel B. Goodman,[1,5] Nikos B. Reppas,[8] Christopher J. Emig,[9] Duhee Bang,[10] Samuel J. Hwang,[11] Michael C. Jewett,[1,12] Joseph M. Jacobson,[2,3] George M. Church[1,6]

We present genome engineering technologies that are capable of fundamentally reengineering genomes from the nucleotide to the megabase scale. We used multiplex automated genome engineering (MAGE) to site-specifically replace all 314 TAG stop codons with synonymous TAA codons in parallel across 32 *Escherichia coli* strains. This approach allowed us to measure individual recombination frequencies, confirm viability for each modification, and identify associated phenotypes. We developed hierarchical conjugative assembly genome engineering (CAGE) to merge these sets of codon modifications into genomes with 80 precise changes, which demonstrate that these synonymous codon substitutions can be combined into higher-order strains without synthetic lethal effects. Our methods treat the chromosome as both an editable and an evolvable template, permitting the exploration of vast genetic landscapes.

The conservation of the genetic code, with minor exceptions (*1*), enables exchange of gene function among species, with viruses, and across ecosystems. Experiments involving fundamental changes to the genetic code could substantially enhance our understanding of the origins of the canonical code and could reveal new subtleties of how genetic information is encoded and exchanged (*1, 2*). Modifying the canonical genetic code could also lead to orthogonal biological systems with new properties. For instance, a new genetic code could prevent the correct translation of exogenous genetic material and lead to the creation of virus-resistant organisms. Additionally, a recoded genome could enhance the incorporation of unnatural amino acids into proteins, because existing suppressor systems must compete with native translation factors (*3–5*).

The construction of a new genetic code requires methods to manipulate living organisms at a whole-genome scale. Such methods are only now becoming attainable through the advent of advanced tools for synthesizing, manipulating, and recombining DNA (*6*). This has led to a number of impressive genome-scale studies, which include removing transposable elements (*7*), refactoring phage genomes (*8*), genome merging (*9*), whole-genome synthesis (*10*), and transplantation (*11*). Whole-genome de novo synthesis offers the ability to create new genomes without a physical template. Its main limitations are the cost of accurate in vitro DNA assembly and introduction of synthetic DNA into organisms (*12*). For

this reason, de novo synthesis is chosen when trying to create a small number of new DNA constructs of modest size (<10 kb) and high fidelity (*8, 10, 12, 13*). Notably, however, the digital template used in de novo synthesis currently originates almost exclusively from sequences found in nature or minor variants thereof.

Redesigned genomes require approaches that reconcile the desired biological behavior with challenges inherent to biological complexity. Engineering biological systems can be unpredictable, as a single misplaced or misdesigned allele can be lethal. To address these challenges, we have developed approaches that integrate syn-

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [2]Center for Bits and Atoms, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4]Program in Biophysics, Harvard University, Cambridge, MA 02138, USA. [5]Program in Medical Engineering and Medical Physics, Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA 02139, USA. [6]Wyss Institute, Harvard University, Cambridge, MA 02115, USA. [7]Program in Chemical Biology, Harvard University, Cambridge, MA 02138, USA. [8]Joule Unlimited, Cambridge, MA 02139, USA. [9]Department of Bioengineering, Stanford University, Palo Alto, CA 94305, USA. [10]Department of Chemistry, Yonsei University, Shinchon 134, Seoul 120-749, Korea. [11]Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [12]Department of Chemical and Biological Engineering and Chemistry of Life Processes Institute, Northwestern University, Evanston, IL 60208, USA.

*These authors contributed equally to this work.
†Present address: Department of Molecular, Cellular and Developmental Biology, Systems Biology Institute, Yale University, New Haven, CT 06520, USA.
‡To whom correspondence should be addressed. E-mail: farren.isaacs@yale.edu (F.J.I.); carr@mit.edu (P.A.C.)
§Present address: Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02420, USA.

thetic DNA and recombination methods to introduce genome-wide changes dynamically in living cells, thereby engineering the genome through viable intermediates. In recent work, we developed multiplex automated genome engineering (MAGE), which rapidly generates genetic diversity for strain and pathway engineering (*14*). To augment MAGE's ability to introduce nucleotide-scale mutations across the genome, a complementary method was required to assemble modified chromosomes in vivo.

Here, we report the development of a hierarchical conjugative assembly genome engineering (CAGE) method and its integration with MAGE toward reengineering the canonical genetic code of *E. coli* (Fig. 1)—an organism with broad utility in basic and applied research. The *E. coli* genetic code contains three stop codons (UAG, UAA, and UGA) whose translation termination is mediated by two release factors, RF1 and RF2. RF1 recognizes the termination codons UAA and UAG, whereas RF2 recognizes UAA and UGA. We hypothesized that replacing all TAG codons with synonymous TAA codons would abolish genetic dependence on RF1 and permit the newly reassigned TAA codons to be recognized by RF2. This will enable us to test and leverage the redundancy of the genetic code by deleting RF1 (Δ*prfA*), providing a blank TAG codon that could be cleanly reassigned to new function. Given that codon utilization bias has been shown to affect translation efficiency (*15, 16*) and viral infectivity (*13*), we sought to determine whether *E. coli* could maintain viability with the systematic replacement of the 314 TAG codons.
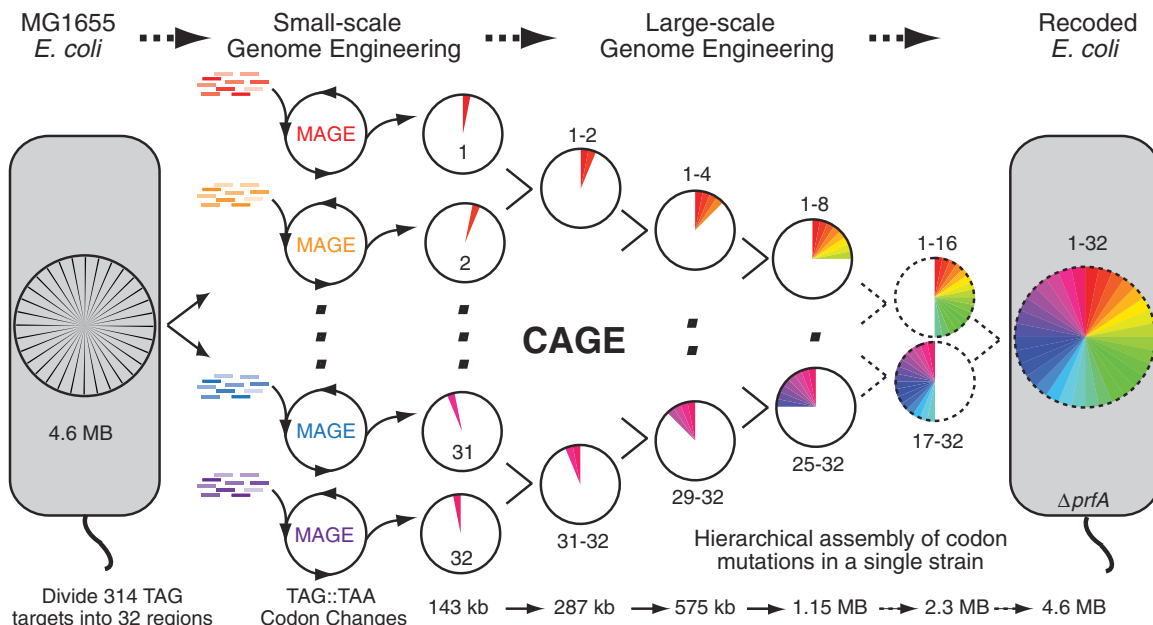
On the basis of the MG1655 genome annotation, we identified 314 *E. coli* genes that con-

tain the TAG stop codon (fig. S1 and table S1). We focused initially on reassigning all 314 stop codons (TAG) to the synonymous stop codon (TAA) in a modified *E. coli* MG1655 strain (EcNR2: *E. coli* MG1655 Δ*mutS*::*cat* Δ(*ybhB-bioAB*)::[λ*cI857* Δ(*cro-ea59*)::*tetR-bla*]). A mismatch repair–deficient (Δ*mutS*) strain was used to achieve high-frequency allelic replacement (*17*). We used MAGE to simultaneously introduce subsets of the TAG-to-TAA codon changes into 32 separate strains (Fig. 1). Specifically, the EcNR2 genome was divided into 32 regions; 31 of these contained 10 targets, and the other contained the remaining four targets. This division was pursued for four reasons. First, pilot experiments (fig. S2) and associated computational predictions showed that the use of pools of 10 or more oligonucleotides (oligos) for MAGE (*14*) achieves highly efficient allelic replacement. Second, limiting the number of MAGE cycles for codon conversions minimizes the total number of cell divisions (six to eight per cycle) in the presence of λ red proteins (which promote recombination and are mutagenic) and deficient mismatch repair (MMR) (*18*). This reduces the number of undesired secondary mutations. Third, the use of smaller oligo pools enabled rapid accumulation of the desired codon conversions in parallel and quantitative measurements of conversion frequencies. Finally, we anticipated that certain codons might be recalcitrant to codon conversion or cause an aberrant phenotype, so it was advantageous to test mutations in small subsets. Candidates included 43 essential genes (*19*) that are terminated by TAG (fig. S1) and 39 genes in which the TAG stop codon overlaps a second reading frame (table S2). Thus, parallel allelic replacement across the 32 regions in

Fig. 1. Strategy for reassigning all 314 TAG codons to TAA in *E. coli*. First, the genome was split into 32 regions each containing 10 TAG stop codons. In parallel, MAGE was used to execute all 10 TAG::TAA codon modifications in a single strain for each genomic region. These partially recoded strains were paired such that a targeted genomic region of one strain (donor) was strategically transferred into a second strain (recipient), permitting the hierarchical consolidation of modified genomic regions using CAGE (see Fig. 4A). This five-stage process transfers genomic fragments ranging in size from ~154 kb to ~2.3 Mb in a controlled manner until a single recoded strain is constructed that lacks the TAG stop codon throughout. Thus far, 28 of 31 conjugations have been completed, where the dotted arrows denote outstanding conjugation steps and dotted genomes represent half- and full-genome strains that have not yet been completed. Once all TAG codons have been converted to TAA, the *prfA* gene will be deleted to inactivate TAG translational termination.

separate strains would enable rapid identification of potentially troublesome alleles.

The 314 oligonucleotides encoding the specified TAG-to-TAA codon mutations (table S3) were computationally designed by means of a software tool (optMAGE, http://arep.med.harvard.edu/optMAGE) on the basis of prior MAGE optimization experiments (14). These oligos were repeatedly applied over 18 MAGE cycles to introduce the codon replacements across 32 cultures (10 targets per culture). We developed two methods based on mismatch amplification mutation assay polymerase chain reaction (MAMA-PCR) (20) to quickly assay target codons. Multiplex allele-specific colony quantitative PCR (MASC-qPCR) (21) (fig. S3) was used to identify clones that contain the greatest number of codon conversions, and multiplex allele-specific colony PCR (MASC-PCR) (21) (fig. S4) was used to measure frequencies of allele replacement at each targeted position. MASC-PCR permitted simultaneous single–base pair (1-bp) measurements (TAG versus TAA) at 10 chromosomal sites per clone (fig. S4).

After 18 MAGE cycles, allelic replacement frequencies were analyzed for all 314 TAG-to-

TAA mutations (Fig. 2) in 1504 clones (47 clones for each of the 32 recoded segments). Allelic replacement frequencies exhibited a high degree of variability among the targets (Fig. 2, outer ring; table S4). The average allelic replacement frequency observed was 37 ± 19% after 18 cycles, and 42% of the population was unconverted after 18 cycles; we observed 1 to 10 converted alleles per clone across the remaining population (Fig. 3A). These measurements suggest the evolution of two types of cells in our mixed cultures: one that appears largely resistant to allelic replacements, and another that readily permits them. With this knowledge, future MAGE methodology could be modified to select highly recombinogenic clones after fewer cycles (e.g., 5 versus 18). Notably, comparable distributions of allelic replacement frequencies were observed for TAG codons present in essential genes, codons overlapping a second reading frame, and codons distributed at various positions throughout both replicating arms (table S5). Moreover, allelic replacement frequencies did not correlate with distance from the origin of replication (oriC), nor with recombination hotspots [e.g., Chi-sites, DNA motifs (5′-GCTGGTGG-3′) in the genome

where homologous recombination could be enhanced] or direction and level of transcription.

All individual TAG-to-TAA conversions were observed, indicating that no TAG stop codon in *E. coli* is required for survival or robust growth. Of 314 codon targets, 298 could be assayed using MASC-PCR, whereas the remaining modifica-
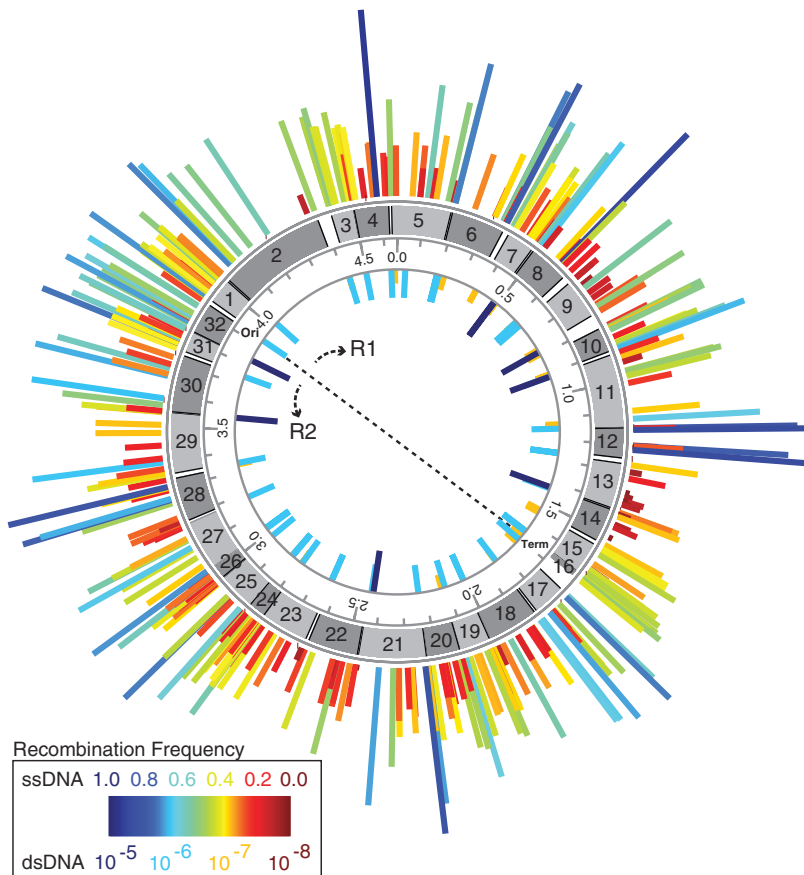

Fig. 3. Clonal rate and distribution of genome modifications after 18 cycles of MAGE. (**A**) Histogram of the frequencies of clones containing 1 to 10 conversions found among 1504 clones screened. A Poisson fit is shown (solid yellow line) for a subpopulation that excludes the zero-conversion group. (**B**) Distribution of modifications among the group of top clones (one for each of the 31 groups of 10 targeted modifications; one additional strain not shown had conversion at all four codon sites). (**C**) Distribution of the 314 90-mer oligos by their extent of total secondary sequence similarity to the *E. coli* genome. Inset box plot: Oligos with a mistarget score of more than 600 bp show, on average, a 32% decrease in allelic replacement (AR) frequency relative to oligos scoring less than 600 bp (25.6% versus 37.6%, *P* < 0.003).

Fig. 2. Frequency map of oligo-mediated TAG::TAA codon replacements and genetic marker integrations across the *E. coli* genome at each replacement position. Circular map illustrates (from inner circle outward): (i) frequency of dsDNA selectable marker integrations; (ii) genome coordinates (in Mb): position of origin (Ori) and terminus (Term) and direction of the two replication forks (R1 and R2); (iii) location of the 32 targeted chromosomal segments; and (iv) frequency of TAG::TAA replacements across all TAG codons—after 18 MAGE cycles—denoted by height- and color-coded bars (scale bar indicates integration frequency).
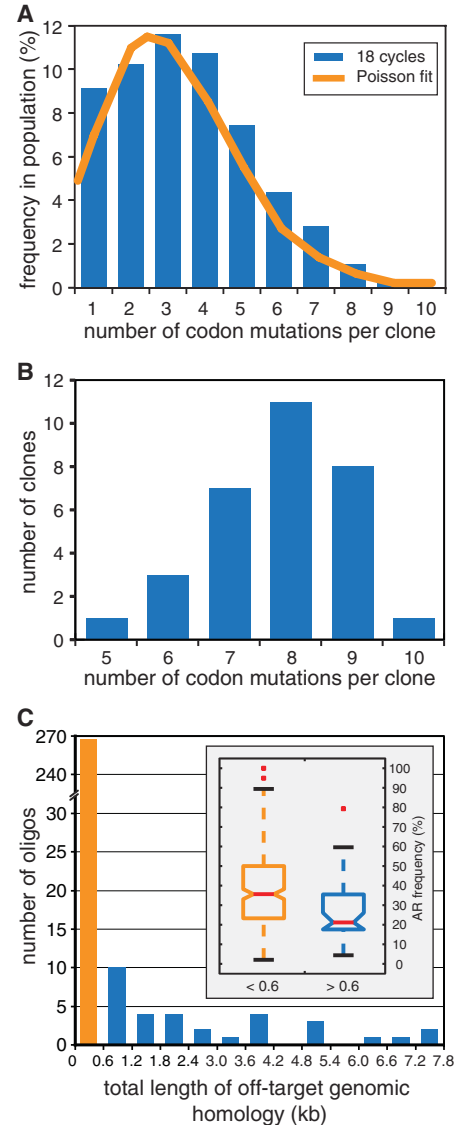
tions were confirmed by direct DNA sequencing. By screening to identify maximally modified cells (Fig. 3B, 5 to 10 modifications per clone with a median of 8) and minimizing aberrant phenotypes (i.e., auxotrophy, decreased fitness) across 1504 clones, we isolated the top clone from each of the 32 populations after 18 MAGE cycles. These clones collectively accumulated 246 of 314 (78%) desired mutations after 18 MAGE cycles. Clones that did not contain all of the codon changes were subjected to an additional 6 to 15 MAGE cycles to convert the remaining TAG codons.

Given that λ Red facilitates highly efficient recombination using short regions of complementary sequences, it was important to assess the potential effects of oligonucleotide hybridization to other (unintended) regions of the genome. We found that 90-mer oligos that have multiple regions of high sequence similarity throughout the genome have a reduced recombination frequency (Fig. 3C) but that these oligos rarely cause mutagenesis at those other locations (see below). To estimate this effect quantitatively, we performed

BLAST alignments of each oligo against the entire genome. To compute a mistarget score, we summed the lengths of the BLAST matches for each oligo sequence against the rest of the *E. coli* genome (blastn, word size = 11, expectation value = 10). Although the majority of oligos (~270) showed only minor sequence similarity to the genome (mistarget score < 600), we found that the score strongly correlated with the frequency of allelic replacement (Fig. 3C). Recombination frequencies were decreased by more than 30% for oligos having many regions of high sequence similarity in the genome (mistarget scores > 600 bp; $P < 0.003$). This information will be useful as a predictor of allele replacement frequency for future oligo designs and can be incorporated into automated design software such as optMAGE.

To directly verify the presence of codon conversions and to obtain a snapshot of secondary mutations accumulated during the MAGE process, we performed Sanger sequencing of ~300 bp surrounding each modified TAG replacement site (96 kb overall, ~3 kb in each of the 32

top strains). Sequencing confirmed the accuracy of our MASC-PCR method and verified the 16 TAG-to-TAA conversions not detected by this assay. Background mutations outside of the 90-bp regions targeted by MAGE oligos consisted of 6 substitutions, 0 insertions, and 3 deletions; in contrast, mutations within the targeted regions included 4 substitutions, 1 insertion, and 28 deletions (fig. S5). The use of a MMR-deficient (Δ*mutS*) strain rendered the expected bias toward substitution mutations in the nontargeted regions. Deletion mutations are probably enriched in oligo-targeted regions because internal deletions are common errors in many oligonucleotide chemical synthesis processes (*22*, *23*). We have developed strategies to minimize both sources of error: (i) optimized oligo synthesis to reduce deletions (e.g., fig. S7), and (ii) the use of chemically modified oligos that are not recognized by MMR to achieve efficient allelic replacement in the presence of a functional MMR pathway.

Because we initially performed codon changes in small subsets, we could easily identify candidate
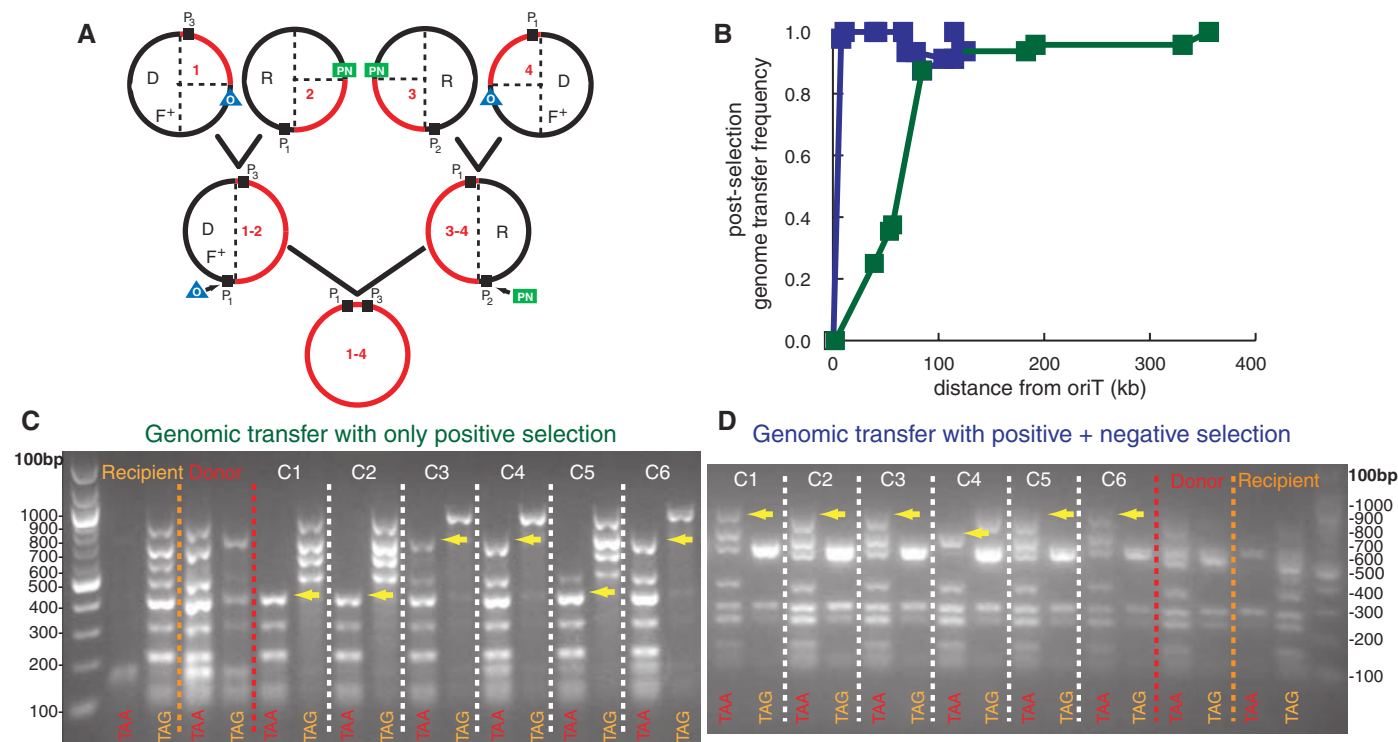

Fig. 4. Hierarchical CAGE methodology for controlled genome transfer. (**A**) Two pairs of strains illustrate the design and methodology of CAGE, with recoded genomic regions in red. Partially recoded strains are split into conjugation pairs. The donor strain (D) contains the following: *oriT-kanamycin* cassette (O, blue triangle); positive selection gene (P$_n$, n = 1, 2, or 3, black rectangle); and F plasmid (F$^+$). The recipient strain (R) contains the following: positive-negative selection gene (PN) and P$_n$, flanking its recoded region. DNA transfer is initiated at O in the donor genome, ensuring transfer of the desired codon mutations and downstream P$_n$. After conjugation, a specific set of three simultaneous selections is applied to yield a recombinant strain that contains the recoded genomic fragment from the donor strain while retaining the other recoded region in the recipient genome. Placement of the PN marker downstream of the *oriT* sequence in the recipient genome ensures that the entire desired region of the donor genome is inherited in the recombinant strain. All conjugation factors are

maintained episomally on F′, so only a ~2-kb *oriT* sequence must be inserted onto the genome to generate a highly controllable *Hfr* donor strain. Because there is no scar between the two recoded regions from the conjugation parents, only one recombination is required to insert the O (donor) or PN (recipient) directly into P$_n$ for the next round of conjugation. This conjugation assembly-selection strategy is implemented in five stages to merge the genomes of 32 recoded strains into a single strain (see Fig. 1). (**B**) Genome transfer frequency as a function of the distance from O. Plots of two conjugations of genome segments illustrating the transfer of ~120 kb with positive and negative selection (blue) and ~360 kb with only positive selection (green) to assemble recoded genomic DNA from donor and recipient genomes. (**C** and **D**) MASC-PCR images of TAA alleles transferred under positive selection alone (C) and positive and positive-negative selection (D). Yellow arrows indicate the genomic point of transfer, which illustrate the inheritance of the donor TAA alleles in the conjugated strain.

mutations that lead to aberrant phenotypes. Growth rates across all 32 top strains (fig. S6 and table S6, average of 47 min per doubling) showed modest deviations from the growth rate of the ancestral strain (42 min per doubling). These changes in growth could be attributed to either the codon changes or the accumulation of secondary mutations in our MMR-deficient strain. Additional phenotypic assays showed a sustained high recombination frequency and a 2.8% frequency of auxotrophy on minimal M9–glucose minimal medium after ~366 generations (table S6). These values compare favorably to previous studies (*24*) in which serial passage of a Δ*mutS* strain resulted in 9% frequency of auxotrophy after ~250 generations.

After converting all TAG codons to TAA across 32 *E. coli* strains, we initiated a five-stage hierarchical assembly (Fig. 1 and table S7) of the modified chromosomal segments into a single strain (Fig. 4). To accomplish this, we developed the hierarchical CAGE method, which is rooted in conjugation, a key mechanism for gene transfer in bacteria (*25, 26*). In contrast to natural mechanisms of conjugal DNA transfer where the *oriT* sequence and conjugal factors act as a contiguous genetic construct, our approach physically decouples the episomally expressed conjugal factors from the chromosomally integrated *oriT* sequence. The *oriT* sequence is fused with a kanamycin resistance gene (*oriT-kan*) so that it can be easily integrated into any permissible locus across the *E. coli* genome via λ Red–mediated double-stranded DNA (dsDNA) recombination (*27*). Thus, we can precisely control the genomic position at which conjugal transfer is initiated (Fig. 4A). This strategy allowed us to use a tractable ~2-kb cassette in place of a cumbersome 30-kb *Hfr* fragment for consecutive manipulations throughout the genome.

Before conjugation, we converted the 32 strains that collectively contain all TAG-to-TAA modifications into 16 pairs of strains primed for large-scale genome transfer (Fig. 1). Within each conjugation pair, a donor strain transfers its recoded genomic region to a recipient strain, which inherits the donor genome and retains its recoded genomic region. Genome transfer is controlled by the precise placement of positive and positive-negative selectable markers integrated with an engineered conjugation strategy to obtain the desired recombinant genomes. Precise placement of these markers into "safe insertion regions" (SIRs: intergenic regions that are not annotated for any coding or regulatory function) by dsDNA recombination (*27*) was intended to maintain genomic integrity and to attain the desired combination of recoded donor and recipient genomes in the recombinant strain (Fig. 4A). In total, two genetic markers were inserted into each of the donor and recipient strains, yielding a total of 64 markers across the 32 modified strains. In the donor strain, the recoded region was flanked by an upstream *oriT-Kan* cassette and a downstream positive selectable marker (P$_1$, e.g., *zeo^R*, *spec^R*,

gent^R). In the recipient strain, the recoded region was flanked with a different positive selectable marker (P$_2$) and a positive-negative selectable marker (PN) such as *tolC* (*28*) or *galK* (*29*). The frequencies of integration among selectable marker cassettes exhibited a high degree of site-specific variability (Fig. 2, inner ring). On average, 59 clones (~10^{-6} frequency) were observed per recombination. However, dsDNA recombination frequencies spanned >3 logs across 81 integration sites tested. Twelve intergenic sites yielded no observable recombinants despite repeated (three or more) attempts. The remaining 69 sites performed as follows: 23 sites at ~10^{-7}, 38 sites at ~10^{-6}, and 8 sites at ~10^{-5} recombination frequencies.

Placement of complementary selectable markers across all 32 strains served as anchor points that enabled hierarchical assembly of recoded genomic fragments. By design, this permits the use of modular *oriT-kan* and *tolC* cassettes throughout the assembly process. Rather than having to prepare a cassette for each SIR, three *oriT-kan* cassettes and three *tolC* cassettes that insert directly into the three positive markers (*zeo^R*, *spec^R*, and *gent^R* genes) are sufficient to guide the remaining four stages of hierarchical assembly. Because *oriT-kan* and *tolC* are not inherited by the recipient strain, each strain can be prepared for subsequent conjugations by simply inserting an *oriT-kan* (donor) or a PN (recipient) directly into one of the strain's inherited positive markers (Fig. 4A).

In the first stage of the hierarchical conjugation strategy (Fig. 1 and table S7), 32 strains each containing 10 codon modifications were merged to produce 16 strains with 20 modifications. Transfer of 1/32 of the genome (~150 kb) occurs at a frequency of ~10^{-4} (Fig. 4B), 2 logs greater than half-genome transfer (*21*). This result supports prior findings that the probability of transferring a specific marker decreases exponentially with its distance from *oriT* (*26*). The relationship between genome transfer efficiency and the distance from *oriT* revealed useful parameters for designing our engineered conjugation scheme. In the absence of a positive-negative marker in the recipient strain, MASC-PCR analysis showed reduced transfer frequency from loci that are in close proximity (<10 kb) to *oriT*, resulting in the uncontrolled transfer of the donor genome—specifically, the loss of mutated TAA codons from the donor genome and the retention of one to four TAG codons from the recipient genome (Fig. 4C). Upon the inclusion of a positive-negative marker [e.g., *tolC* (*28*), *galK* (*29*)] in the recipient genome, desired postconjugal strains were readily selected; that is, full transfer of mutated TAA codons from the donor genome was achieved by selecting for the loss of *tolC* or *galK* placed among the TAG codons in the recipient genome (Fig. 4D). Together, these results demonstrate the requirement for robust positive and positive-negative selectable markers that strategically flank the recoded genomes in

the donor and recipient strains (Fig. 4A). Moreover, MASC-PCR analysis across all codon loci shows that conjugation efficiency is sustained throughout the region of transfer, indicating contiguous transfer of the donor genomic fragment.

Using CAGE, we then consolidated the 32 original strains into eight recoded strains, each with 1/8 of the genome recoded. Two of these eight strains exhibited a dysfunctional *tolC* phenotype (i.e., they simultaneously passed both positive and negative *tolC* selections). Although mutations conferring simultaneous novobiocin sensitivity and colicin E1 resistance have been identified (*28*), there is no literature precedence for the phenotype that we observed. We have discovered two routes to this phenotype. In one strain, the causative allele was present in *tolC*, and we corrected the phenotype by replacing the dysfunctional copy with a functional one. In the other strain, the causative allele appears to be outside of *tolC*. Indeed, *tolC* works in concert with a number of other genes (e.g., *btuB*, *tolA*, *tolQ*, and *tolR*) that have been implicated in dysfunctional negative selection (*30*). Recognizing that the ancestral strains also carried the dysfunctional allele, we reconstructed this 1/8th strain using MAGE, and used it to complete the full set of 1/4 genomes (28 of 31 conjugations). These four strains, which contain up to 80 modifications per genome, can be combined to complete the assembly of a fully recoded strain containing all 314 TAG-to-TAA codon conversions.

In light of the challenges arising from spontaneous point mutations, we sought to assess the effects of MAGE and CAGE on genome stability. Therefore, we performed whole-genome sequencing for the two dysfunctional strains and an additional functional control (figs. S8 and S9 and tables S8 to S11). These strains have 110, 102, and 128 secondary mutations, respectively [total number of single-nucleotide polymorphisms (SNPs) and indels; fig. S8]. After ~960 cell divisions (table S11), the majority of SNPs were transition mutations (98.4% transitions and 1.6% transversions; table S9) and the overall background mutation rate was 2.5 × 10^{-8} per bp per replication (1 error per genome per ~9 replications; table S10). These results are consistent with a Δ*mutS* phenotype (*31*). Our measured error rate was lower than we expected, given that the cumulative potential mutations include contributions from a MMR-deficient strain, repeated exposure to induction of the λ Red recombination system, and conjugation-based genomic manipulations. A mechanistic hypothesis for the lower error rate is that the conjugation process acts as a backcross and removes deleterious secondary mutations through the isolation of clones that maximize fitness. To examine this idea further, we explored the potential functional consequences of these SNPs as indicated by the COG category of the gene or regulatory region associated with the SNP (*32–34*). We used a hypergeometric distribution to determine the enrichment

level of the three main COG categories across all three strains. Both SNPs associated with metabolism (117 SNPs, $P < 4 \times 10^{-4}$) and SNPs associated with information storage and processing (29 SNPs, $P < 0.05$) were shown to be significantly enriched, whereas SNPs associated with cell signaling and transduction (98 SNPs, $P > 0.05$) were not (fig. S9). Future work will be needed to sequence additional strains throughout the ancestral conjugation tree to characterize the frequency, inheritance patterns, and functional bias of such mutations.

This study, which integrates in vivo genome engineering from the nucleotide to the megabase scale, demonstrates the successful replacement of all genomic occurrences of the TAG stop codon in the *E. coli* genome. We found that cells can incorporate all individual TAG-to-TAA codon changes, and that these changes can be assembled into genomes with up to 80 modifications with mild phenotypic consequences. The scarless introduction of codon changes via MAGE enabled the first genome-wide allelic replacement frequency map using single-stranded DNA oligos in *E. coli* (Fig. 2). In addition, our engineered conjugation experiments produced a complementary recombination frequency map of intergenic dsDNA integration sites across the genome (Fig. 2). Together, these experiments revealed both highly accessible and recalcitrant sites for both small- and large-scale chromosomal modifications. These data could serve as valuable resources for future genome engineering efforts. Moreover, synthetic approaches such as the one pursued here may help to refine the existing genome annotation by revealing unannotated functional genetic loci, such as short peptides (*35*) or minigenes (*36*).

Introducing genome-wide changes dynamically in a living cell permits engineering in the

cell's native biological context. In contrast to in vitro genome synthesis (*10*) and transplantation methods (*12*) that introduce discrete and abrupt changes in a single genome, our genome engineering technologies treat the chromosome as an editable and evolvable template and generate targeted and combinatorial modifications across many (~10^9) genomes in vivo (*14*). MAGE is optimal for introducing small modifications in sequence design space, whereas CAGE is designed for taking bigger leaps via large-scale assembly of many modified genomes. Together, these genome editing methods are advantageous when the designed genomes share >90% sequence similarity to existing templates or when many targeted mutations dispersed across the chromosome are desired (e.g., genome recoding).

References and Notes

1. A. Ambrogelly, S. Palioura, D. Söll, *Nat. Chem. Biol.* **3**, 29 (2007).
2. R. D. Knight, S. J. Freeland, L. F. Landweber, *Nat. Rev. Genet.* **2**, 49 (2001).
3. T. S. Young, P. G. Schultz, *J. Biol. Chem.* **285**, 11039 (2010).
4. H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, *Nature* **464**, 441 (2010).
5. T. Mukai *et al.*, *Nucleic Acids Res.* **38**, 8188 (2010).
6. P. A. Carr, G. M. Church, *Nat. Biotechnol.* **27**, 1151 (2009).
7. G. Pósfai *et al.*, *Science* **312**, 1044 (2006); 10.1126/science.1126439.
8. L. Y. Chan, S. Kosuri, D. Endy, *Mol. Syst. Biol.* **1**, 0018 (2005).
9. M. Itaya, K. Tsuge, M. Koizumi, K. Fujita, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15971 (2005).
10. D. G. Gibson *et al.*, *Science* **319**, 1215 (2008); 10.1126/science.1151721.
11. C. Lartigue *et al.*, *Science* **317**, 632 (2007).
12. D. G. Gibson *et al.*, *Science* **329**, 52 (2010); 10.1126/science.1190719.
13. J. R. Coleman *et al.*, *Science* **320**, 1784 (2008).
14. H. H. Wang *et al.*, *Nature* **460**, 894 (2009).
15. B. Irwin, J. D. Heck, G. W. Hatfield, *J. Biol. Chem.* **270**, 22801 (1995).
16. G. A. Gutman, G. W. Hatfield, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 3699 (1989).
17. H. M. Ellis, D. Yu, T. DiTizio, D. L. Court, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6742 (2001).
18. N. Costantino, D. L. Court, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15748 (2003).
19. S. Y. Gerdes *et al.*, *J. Bacteriol.* **185**, 5673 (2003).
20. R. S. Cha, H. Zarbl, P. Keohavong, W. G. Thilly, *PCR Methods Appl.* **2**, 14 (1992).
21. See supporting material on *Science* Online.
22. P. A. Carr *et al.*, *Nucleic Acids Res.* **32**, e162 (2004).
23. A. B. Oppenheim, A. J. Rattray, M. Bubunenko, L. C. Thomason, D. L. Court, *Virology* **319**, 185 (2004).
24. P. Funchain *et al.*, *Genetics* **154**, 959 (2000).
25. H. Ochman, J. G. Lawrence, E. A. Groisman, *Nature* **405**, 299 (2000).
26. G. R. Smith, *Cell* **64**, 19 (1991).
27. D. Yu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5978 (2000).
28. J. A. DeVito, *Nucleic Acids Res.* **36**, e4 (2008).
29. S. Warming, N. Costantino, D. L. Court, N. A. Jenkins, N. G. Copeland, *Nucleic Acids Res.* **33**, e36 (2005).
30. M. Masi, P. Vuong, M. Humbard, K. Malone, R. Misra, *J. Bacteriol.* **189**, 2667 (2007).
31. R. M. Schaaper, R. L. Dunn, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6220 (1987).
32. H. Li, R. Durbin, *Bioinformatics* **25**, 1754 (2009).
33. H. Li *et al.*, 1000 Genome Project Data Processing Subgroup, *Bioinformatics* **25**, 2078 (2009).
34. W. McLaren *et al.*, *Bioinformatics* **26**, 2069 (2010).
35. M. R. Hemm *et al.*, *J. Bacteriol.* **192**, 46 (2010).
36. E. Jacinto-Loeza, S. Vivanco-Domínguez, G. Guarneros, J. Hernández-Sánchez, *Nucleic Acids Res.* **36**, 4233 (2008).

Linking Context with Reward: A Functional Circuit from Hippocampal CA3 to Ventral Tegmental Area

Alice H. Luo,[1]* Pouya Tahsili-Fahadan,[2] Roy A. Wise,[1] Carl R. Lupica,[3] Gary Aston-Jones[2]

Reward-motivated behavior is strongly influenced by the learned significance of contextual stimuli in the environment. However, the neural pathways that mediate context-reward relations are not well understood. We have identified a circuit from area CA3 of dorsal hippocampus to ventral tegmental area (VTA) that uses lateral septum (LS) as a relay. Theta frequency stimulation of CA3 excited VTA dopamine (DA) neurons and inhibited non-DA neurons. DA neuron excitation was likely mediated by disinhibition because local antagonism of γ-aminobutyric acid receptors blocked responses to CA3 stimulation. Inactivating components of the CA3-LS-VTA pathway blocked evoked responses in VTA and also reinstatement of cocaine-seeking by contextual stimuli. This transsynaptic link between hippocampus and VTA appears to be an important substrate by which environmental context regulates goal-directed behavior.

Efficient reward-seeking requires that environmental stimuli be properly interpreted, to predict when and where reward can be expected. The ventral tegmental area (VTA) and its dopaminergic projections are critical components of a reward circuit. Although the influence

of temporal cues on this system has been well studied (*1*), it is unknown how the VTA system relates to other contextual information, such as where reward can be expected. The hippocampus organizes aspects of context into a relational memory network (*2*). Interactions between hippocampus and VTA are important for context-reward associations (*3*). However, the circuitry by which the two interact remains to be elucidated.

We used pseudo-rabies virus (PRV), a retrograde, transsynaptic tracer (*4*), to delineate circuit projections to VTA. Beginning 48 hours after unilateral injection into VTA, PRV (+) neurons

[1]Behavioral Neuroscience Section, Behavioral Neuroscience Research Branch, National Institute on Drug Abuse Intramural Research Program, National Institutes of Health, Baltimore, MD 21224, USA. [2]Department of Neurosciences, Medical University of South Carolina, Charleston, SC 29425, USA. [3]Electrophysiology Research Section, Cellular Neurobiology Branch, National Institute on Drug Abuse Intramural Research Program, National Institutes of Health, Baltimore, MD 21224, USA.

*To whom correspondence should be addressed. E-mail: alice_luo@hotmail.com