SYNTHETIC BIOLOGY

Efficient formation of single-copy human artificial chromosomes

Craig W. Gambogi[1,2,3,4], Gabriel J. Birchak[1,2,3,5], Elie Mer[1,2,3,4], David M. Brown[6], George Yankson[7], Kathryn Kixmoeller[1,2,3,4], Janardan N. Gavade[1,3,4], Josh L. Espinoza[6], Prakriti Kashyap[1,3,4], Chris L. Dupont[6], Glennis A. Logsdon[1,2,3,4]†, Patrick Heun[7,8], John I. Glass[6], Ben E. Black[1,2,3,4,5]*

Large DNA assembly methodologies underlie milestone achievements in synthetic prokaryotic and budding yeast chromosomes. While budding yeast control chromosome inheritance through ~125-base pair DNA sequence-defined centromeres, mammals and many other eukaryotes use large, epigenetic centromeres. Harnessing centromere epigenetics permits human artificial chromosome (HAC) formation but is not sufficient to avoid rampant multimerization of the initial DNA molecule upon introduction to cells. We describe an approach that efficiently forms single-copy HACs. It employs a ~750-kilobase construct that is sufficiently large to house the distinct chromatin types present at the inner and outer centromere, obviating the need to multimerize. Delivery to mammalian cells is streamlined by employing yeast spheroplast fusion. These developments permit faithful chromosome engineering in the context of metazoan cells.

Yeast artificial chromosomes (YACs) (*1*–*3*) are typically 0.1 to 1 Mb and have allowed for advances in molecular biology including the cloning of large disease genes (*4*) and the generation of entire synthetic prokaryotic genomes (*5*, *6*). They have also provided the foundation for the generation of entirely synthetic budding yeast chromosomes (*7*). Writing new chromosomes, or even entire genomes, is an aspiration for synthetic biologists working in diverse eukaryotes, including in mammalian and plant systems (*8*, *9*) as it would enable applications of genome engineering across research, biotechnology, and health-related landscapes (*8*).

Human artificial chromosomes (HACs) were developed ~25 years ago (*10*–*12*) and are on average 1 to 10 Mb in their functional form after their establishment in cells. They pave the way for advances and insights in eukaryotic systems where a specific key chromosomal locus, the centromere, is typically more than a thousand times larger than a budding yeast point centromere and is functionally defined not by a particular sequence but by an array of nucleosomes containing a histone H3 variant, CENP-A (*13*). Unlike YACs in yeast, an unintended byproduct of de novo HAC formation is multimerization of the initial input DNA construct—typically 100- to 200-kb bacterial artificial chromosomes (BACs)—creating functional HACs with a variable number of multimers (typically >40-fold) (*14*–*16*). The multimerization and the uncontrolled rearrangement of the input DNA that accompanies it during the early steps of HAC formation (often with equally unintended incorporation of portions of natural chromosomes) has severely hindered their development toward the broader promise of synthetic biology and therapeutic applications (*17*). We have now overhauled the design and delivery of HACs; rather than control the multimerization process we bypassed it completely by increasing the size of the input DNA. We report success in forming single-copy HACs at an overall efficiency of de novo establishment that surpasses all earlier versions.

[1]Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [2]Graduate Program in Biochemistry and Molecular Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [3]Penn Center for Genome Integrity, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [4]Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [5]Graduate Program in Cell and Molecular Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [6]J. Craig Venter Institute, La Jolla, CA 92037, USA. [7]Wellcome Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, UK. [8]Department of Biology, Molecular Genetics, Technical University Darmstadt, 64289 Darmstadt, Germany. *Corresponding author. Email: blackbe@pennmedicine.upenn.edu †Present Address: Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

Results

An overhauled platform for efficient HAC formation

We predicted that to remain single-copy and avoid multimerization, the initial construct would need to be larger than the BAC-based HAC constructs of earlier versions (*14*–*16*). This is based on the understanding that centromeres require multiple domains with distinct functions that are spatially separated at mitosis when cohered sister chromatids align on the microtubule-based spindle (*18*). The centromeric region harboring CENP-A nucleosomes (the region for assembling the mitotic kinetochore) typically discontinuously spans ~75 to 300 kb (*19*–*22*), whereas the inner centromere is another largely heterochromatic region that regulates sister chromatid cohesion and a quality control pathway (termed "error correction") that monitors bipolar spindle attachment. We reasoned that BAC-based HAC constructs, which typically start in the 100-300-kb size range (*14*–*16*), can likely only form when multimerization occurs. The larger size is presumably required to accommodate formation of both distinct chromatin domains that define a functional centromere. Conversely, we reasoned that starting with a larger initial construct will bypass this requirement, allowing HACs to form more frequently and without multimerization.

To test our prediction, we devised a scheme that employs three recent technical advances to build and test a single-copy HAC construct (fig. S1A). First, YAC constructs are readily generated in the 0.5 to 2 Mb size range (*5*, *23*) through transformation-associated recombination (TAR) cloning (*24*). Second, bypassing the requirement for long (>40 kb) stretches of highly repetitive centromere DNA (α-satellite) for HAC formation (*16*) permits the use of nonrepetitive DNA. This is critical because TAR cloning is not compatible with long repetitive sequences (*25*). Third, large YAC constructs can be efficiently delivered to mammalian cells through optimized fusion with yeast spheroplasts (*23*), potentially leading to a marked increase in independent HAC formation events relative to what has been achieved with low-efficiency transfection-based delivery of BAC-based HAC vectors in prior versions (*14*–*16*).

The HAC template was constructed through TAR assembly starting with a YAC harboring 550 kb of *Mycoplasma mycoides* genomic DNA (*6*), 4q21 BAC^{LacO} (*16*), and linkers for recombination that also include a yeast auxotrophic marker and a mammalian expression cassette for mCherry (fig. S1A). *M. mycoides* genomic DNA was chosen because it represents a heterologous DNA sequence that is known to be readily propagated in budding yeast (*6*). It serves as a noncoding sequence in the context of a eukaryotic cell and is not expected to elicit unintended or detrimental impact on HAC formation or cell function. It was also attractive because the AT richness of the *M. mycoides* DNA (~75%) is anticipated to support the heterochromatic nature of the inner centromere. Further, *M. mycoides* DNA has already been efficiently delivered to cultured human cells (*23*) and allows for unambiguous detection of HACs as it is comprised of unique, nonhuman sequence. 4q21 BAC^{LacO} was chosen because it is the only HAC construct comprised of nonrepetitive DNA (182 kb of DNA from human chromosome 4 at the 4q21 locus) that has been demonstrated to form functional HACs, instead of the 40 to 200 kb of highly repetitive α-satellite-based BAC constructs that prior HAC studies have used (*14*–*16*, *26*). The LacO array on 4q21 BAC^{LacO} permits subsequent targeting of mCherry-LacI-HJURP for initial seeding of CENP-A nucleosomes (*16*). We termed the new construct YAC-*Mm*-4q21^{LacO}.

For recipient cells, we used the HT1080 Dox-inducible mCherry-LacI-HJURP line in which 4q21 BAC^{LacO}-based HACs were seeded with CENP-A nucleosomes (*16*). The HT1080 background, in general, was chosen because it is the one in which HAC formation has historically been performed (*11*, *12*, *14*–*16*, *26*) due to its chromatin state that is permissive to occasional centromere formation (*27*). We also generated a second recipient line, U2OS$^{Dox\text{-}inducible\ mCherry\text{-}LacI\text{-}HJURP}$, because the U2OS background is established as an efficient recipient of YACs through spheroplast fusion (*23*). Both of our chosen recipient lines were first optimized for spheroplast fusion conditions (fig. S2) and then subjected to HAC formation assays with YAC-*Mm*-4q21^{LacO} (Fig. 1 and figs. S2 and S3). Following spheroplast fusion, we noted that unlike prior HAC assays where only ~40 surviving colonies emerge in 2 to 3 weeks, a nearly confluent monolayer of G418-S-resistant cells was present after 8 days of selection. For both recipient cell types, a substantial proportion (42 ± 9% and 46 ± 5%) of the neomycin-resistant cells harbor HACs. Most or all of these are substantially smaller in size (<1 μm) (Fig. 1, B to D) than the multimerized HACs of prior generations (~2 μm) (*14*–*16*). Without induction of mCherry-LacI-HJURP, there was only a very small proportion of HACs, with most cells possessing detectable FISH signal coming from an integration into a natural chromosome of the recipient cell (Fig. 1C). Our initial findings therefore strongly indicate exceedingly high efficiency of YAC delivery, robust HAC formation rates upon seeding CENP-A nucleosome assembly, essentially uniform avoidance of the stochastic and highly unlikely multimerization that has accompanied prior systems for de novo HAC formation, and no restriction to the specific cell line (HT1080) to which prior generations of HACs were confined.

YAC-*Mm*-4q21^{LacO} HACs harbor multidomain centromeres for faithful inheritance

We next sought to test the degree to which the centromeres on the HACs could support mitotic function. Starting from the polyclonal population of cells that survived G418-S selection (Fig. 1), we screened 24 monoclonal isolates and identified 8 lines harboring HACs (the other lines included 2 in which YAC-*Mm*-4q21^{LacO} had integrated into a natural chromosome and 14 with no detectable HAC) and measured the proportion of cells in each harboring a detectable HAC (Fig. 2, A and B). Sequencing analysis of four of the monoclonal isolates revealed the presence of the HAC but no evidence of any integration of the endogenous yeast chromosomes that were also introduced during the initial spheroplast fusion step (fig. S4). Our findings are consistent with prior studies where yeast chromosome segregation after spheroplast fusion was only observed after strong selection for rare genome integration events (*28*). Overall, most cells upon isolation of a clonal line harbor HACs, matching this property of our prior generations of HACs (*11*, *12*, *14*, *16*). Four of the lines (colored data points in graph in Fig. 2B) were subjected to three independent HAC stability assays over a month of growth in the absence of antibiotic selection (Fig. 2C). The average daily HAC loss rate of 0.011 ± 0.006 (Fig. 2C) was low, similar to prior generations of HACs (*12*, *16*, *29*).

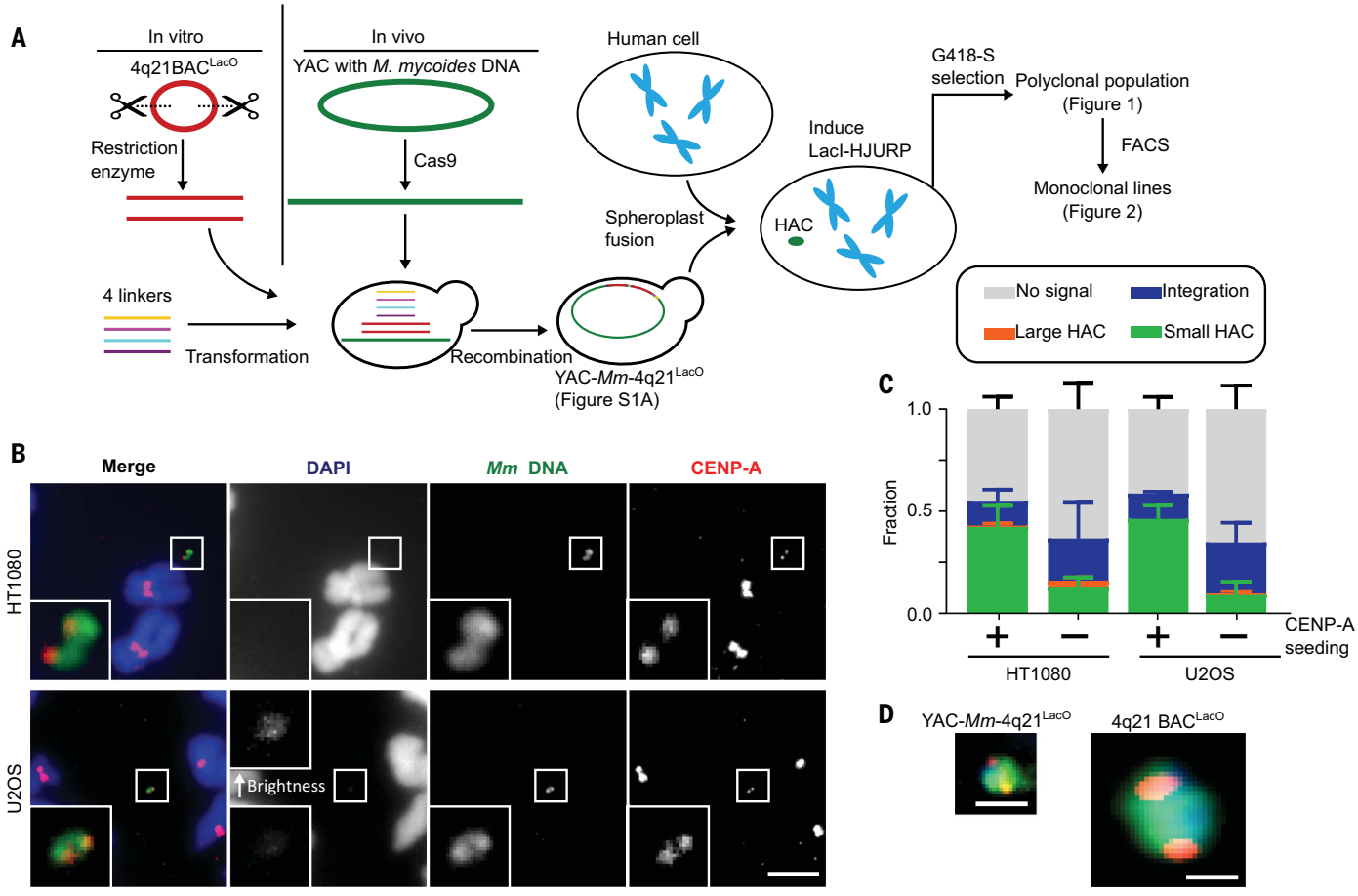

Fig. 1. 760-kb HAC constructs efficiently acquire centromeres and exist as autonomous chromosomes. (**A**) Schematic of HAC-generation approach. (**B**) Representative images of a single-copy HAC generated in HT1080 and U2OS cells. (Insets) 5x magnification; bar, 5 μm. (**C**) Quantification of proportion of "small HACs" (FISH signal spans < 1 μm), "large HACs" (FISH signal spans > 1 μm), and "integrations" and "no signal" spreads generated from HAC formation assays. The mean [± standard deviation (SD)] is shown. (**D**) Size comparison of a HAC made from YAC-*Mm*-4q21^{LacO} and a multimerized HAC made from 4q21 BAC^{LacO}. Both HACs are shown at the same scale. Coloring is identical to what is shown in the merged images in (B). Bar, 1 μm.

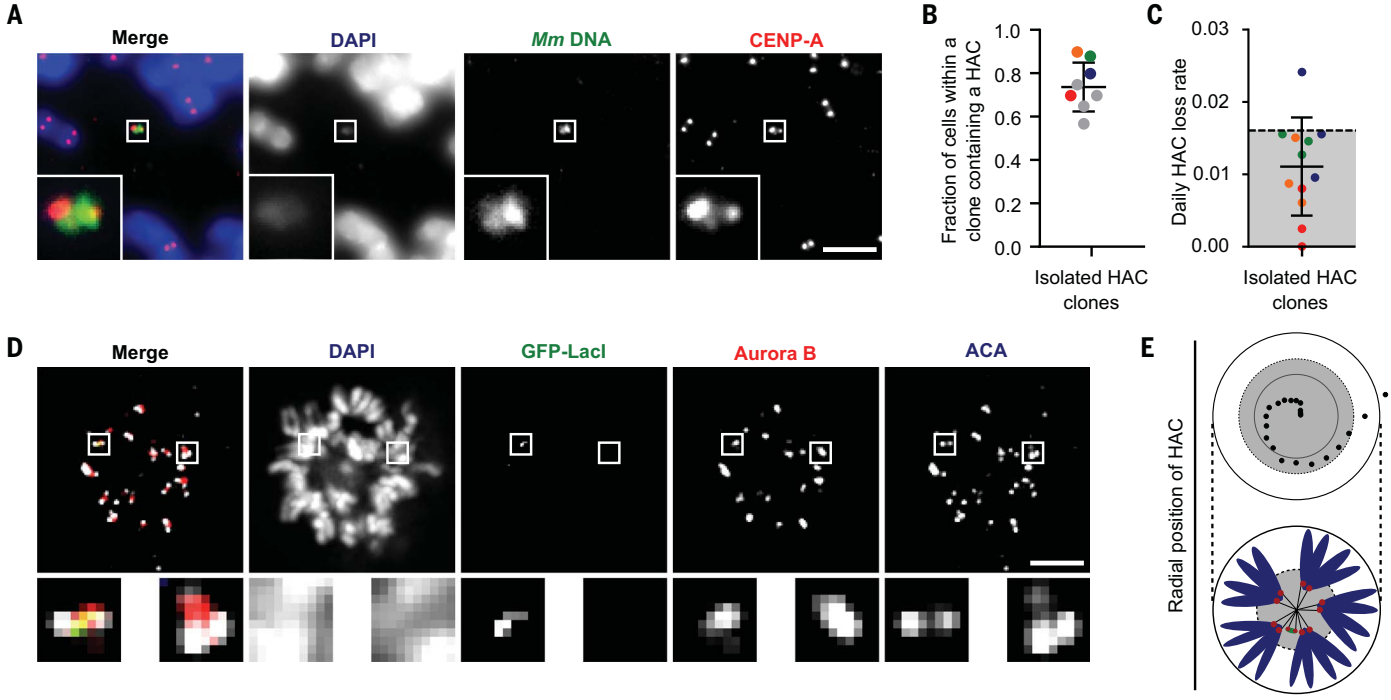

Fig. 2. YAC-*Mm*-4q21^{LacO}–based HACs are inherited as autonomous chromosomes with functional kinetochores and robust CPC recruitment. (**A**) Representative image of a single-copy HAC that has been isolated in a monoclonal cell line. (Inset) 5x magnification. Bar, 5 μm. (**B**) Quantification of fraction of spreads with a HAC in monoclonal cell lines. The mean (± SD) is shown. Colors indicate individual HAC lines that were assessed in HAC retention assays. Red, orange, green, and blue correspond to HAC clones 1, 2, 3, and 5, respectively. See also table S1. (**C**) Quantification of HAC loss rate after culturing without selection for 30 days. The mean (± SD) is shown. Experiments are color coded to correspond to the clones shown in (B). Gray shading indicates the range of loss rates for prior generations of HACs (*12*, *16*, *29*). (**D**) Representative images of HACs synchronized in mitosis showing Aurora B and ACA. The image shows eight 0.2-μm z-projected stacks (see also fig. S5 for centromere delineation in the z-dimension). (Inset) 5x magnification; bar, 5 μm. (**E**) The radial position of HACs was measured relative to endogenous centromeres. The position of 20 HACs, each endogenous centromere and the center of DNA mass was measured. The distance between HAC or endogenous centromere and the center of DNA mass was calculated. The distance of each HAC from the center was normalized based on the total length across (i.e., the diameter) of mitotic chromosomes. The inner black circle represents the mean radial position of endogenous centromeres and the dotted line represents one standard deviation from the mean. An illustration is shown below the graph.

To further examine our initial hypothesis that a larger initial HAC construct would confer full centromere function, we assessed the mitotic recruitment of a key component of the inner centromeric error correction mechanism, the Aurora B kinase (*18*). The inner centromere includes chromatin marked with H3^{T3phos} and H2A^{T121phos} modifications for recruiting the chromosome passenger complex (CPC), containing Aurora B (*30*–*32*). Inner centromeric chromatin spans several Mb on natural chromosomes (*33*). The inner centromere in metazoans is not a single thread of chromatin but rather thought to be a densely packed region that spans a linear distance of 500 to 1000 nm between sister centromeres (*34*, *35*). Given the high-fidelity transmission of our HACs (Fig. 2C), we predicted that they are sufficiently large to generate a robust inner centromere that recruits Aurora B. In order to have the necessary dispersion of chromosomes in the mitotic cell for robust detection of the HAC through expression of GFP-LacI, we induced the formation of monopolar spindles and assessed both the kinetochore forming part of the paired sister HACs [with anticentromere antibodies (ACA)] and the inner centromere (with antibodies to Aurora B) (Fig. 2, D and E, and fig. S5). Aurora B was found in 33/35 HACs. Thus, our findings suggest that the increased size of YAC-*Mm*-4q21^{LacO} relative to prior HAC constructs permits a robust inner centromere without the need to undergo large-scale multimerization. This is critical as it endows YAC-*Mm*-4q21^{LacO} HACs with the ability to segregate in mitosis at high fidelity alongside natural counterparts.

Single-copy HACs

The small physical size of HACs formed from YAC-*Mm*-4q21^{LacO} (Fig. 1) raised the possibility that they can form without any multimerization at all. To test this notion we sought a cytological approach that reports on copy number without the deformations that happen naturally when chromosomes are attached to and stretched by the spindle or otherwise confounded by mitotic chromosome condensation. We found that nuclear envelope lysis during isolation of nuclei releases the small HACs formed with YAC-*Mm*-4q21^{LacO} that are subsequently efficiently separated from the rest of the genome through centrifugation (Fig. 3A and fig. S6). We harvested the top gradient fractions in the 10% sucrose layer (i.e., above the visible cell debris) and determined the location of CENP-A and the LacO sequences (Fig. 3B). We anticipated a single CENP-A focus on interphase HACs, even after replication, because sister centromeres are not separated into distinct foci on natural chromosomes until just prior to nuclear envelope breakdown near mitotic onset (*36*). LacO arrays, on the other hand, when present on repeated HAC constructs do not coalesce into a single focus (*16*). HACs were readily identified in these fractions, permitting us to visualize an individualized and functional metazoan chromosome in its decondensed, interphase form (Fig. 3B). In most HACs, CENP-A and LacO staining each produced a single focus (Fig. 3B). We did not observe any HACs with a single CENP-A focus and more than one LacO focus. A small number of HACs harbored two CENP-A foci, consistent with them coming from cells that were in late G2 or early mitosis (i.e., prior to sister chromatid separation) at

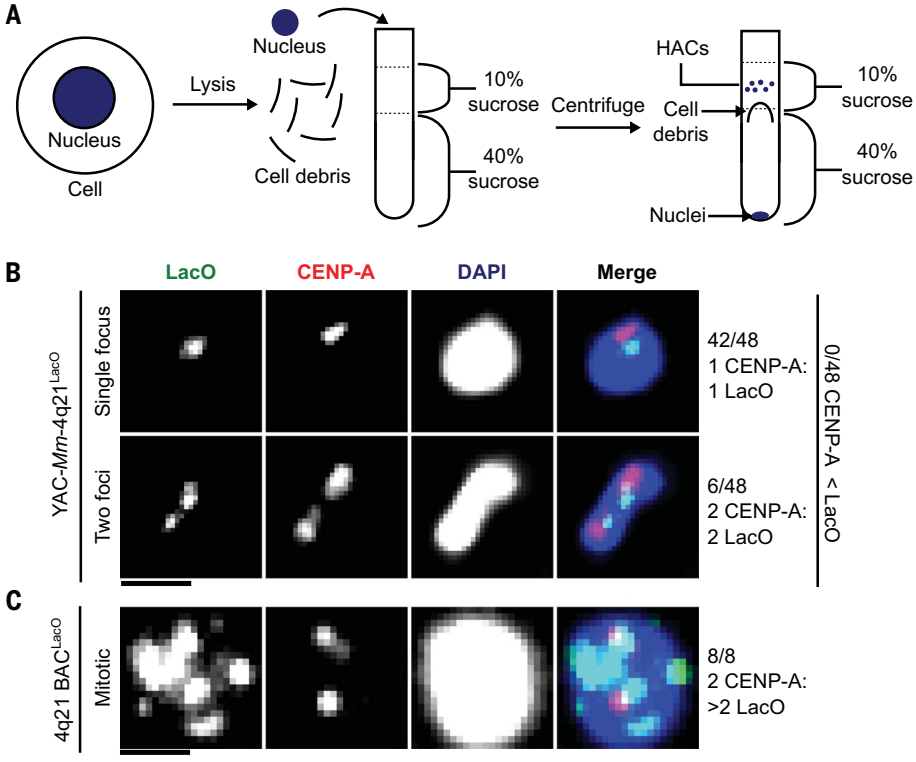

Fig. 3. YAC-*Mm*-4q21^{LacO} HACs are functional as single-copy DNA. (**A**) Schematic of approach used to enrich single-copy HACs. (**B**) Representative image of HACs isolated by sucrose gradient with either a single or two foci of LacO. The proportion of HACs with a single or two foci is noted. HACs with two LacO foci also had two CENP-A foci, suggesting that they are mitotic. Bar, 1 μm. (**C**) Representative image of a multimerized HAC [Clone 27 from (*16*)] from mitotic chromosome spreads. Bar, 1 μm.

the time of isolation. Each of these also had precisely two LacO foci (Fig. 3B). Unlike the single-copy YAC-*Mm*-4q21^{LacO} HACs, prior generations of HACs are large multimers that do not separate from endogenous chromosomes during nuclei isolation (*16*). In these HACs, CENP-A and LacO arrays are readily visualized on mitotic HACs in chromosome spreads (Fig. 3C). For the prior generation of HACs, the paired, replicated centromeres are visible as "double dots" of CENP-A whereas the LacO arrays are visible as numerous foci (Fig. 3C). Taken together with the earlier detection from uniformly small-sized HACs from populations of cells with nascent YAC-*Mm*-4q21^{LacO} HACs (Fig. 1, B to D), our interphase HAC experiments (Fig. 3B) indicate that the new HACs are formed and maintained without multimerization.

We next assessed the size and topology of functional YAC-*Mm*-4q21^{LacO}-based HACs (Fig. 4). This is important because earlier generations of HACs typically formed in a manner accompanied by large-scale DNA sequence multimerization and even acquisition of portions (>100 kb) of host cell chromosomal DNA (*16*, *37*, *38*). YAC-*Mm*-4q21^{LacO} contains a single *Fse*I site (fig. S1), and we found that two isolated HAC-containing cell lines required *Fse*I digestion to enter a pulse-field gel (Fig. 4A and fig. S7). This is consistent with well-established topological trapping of circular chromosomes prior to digestion (*39*). The mobility of the linearized HAC suggests that it has not lost or gained large fragments of DNA (Fig. 4A and fig. S7). We compared this to a circular, multimerized BAC-based HAC (*16*) that has one *Fse*I site per repeating ~200-kb monomer (Fig. 4A and fig. S7). Along with our cytological data indicating that the HACs are single-copy (Fig. 3), their behavior on pulse-field gels (Fig. 4A and fig. S7) support the notion that they function and are inherited through cell divisions with the same single-copy circular nature in which they were initially constructed in yeast.

To cytologically examine YAC-*Mm*-4q21^{LacO} as it exists in yeast and then after it forms a functional HAC in human cells, we employed a well-established chromatin stretching approach (*34*, *40*) (Fig. 4, B to H and fig. S8). This permits us to monitor the 182-kb region of 4q21 present on the HACs in comparison to the initial YAC construct or the endogenous locus on chromosome 4 (Fig. 4, B to H). The degree of stretching we achieved is about threefold, making the circle fold back upon itself (Figs. 4, B to D, and fig. S8). Stretching maintains large blocks (roughly 40 kb) of chromatin linked by highly stretched regions with little or no detectable FISH signal (Fig. 4, C, D, and G, and fig. S8). The overall length of the HAC in human cells was greater than the YAC in yeast but smaller than native 4q21 (Fig. 4E). The number of 4q21 foci produced by stretching of the YAC in yeast or at the native locus and HAC in mammalian cells is similar (Fig. 4F). We conclude that the similarity of YACs and HACs reflects the absence of rampant multimerization or other substantial molecular changes (Fig. 4, C to F, and figs. S1 and S4), with relatively subtle changes in length and foci number likely reflecting differences in chromatin properties between budding yeast and human cells (*41*, *42*).

The number of 4q21 foci observed on the HAC varied greatly, revealing the possible existence of two HAC populations (Fig. 4F). We reasoned that the HAC would be visualized as a single chromatid early in the cell cycle (G1 and early S phase) but would be visualized as paired chromatids later in the cell cycle (late S, G2, and M). On the other hand, the native 4q21 locus would be visualized in stretching experiments as a single chromatid, even late in the cell cycle, because each natural chromosome arm location would make its own chromatin fiber. To test this notion, we enriched cells in mitosis, prior to sister chromatid separation, and found that the average number of 4q21 foci in the HAC increased from 4.9 ± 2.9 to 8.2 ± 2.1 (Fig. 4H) relative to those from asynchronous cell populations (Fig. 4F) whereas the endogenous 4q21 locus has a similar number of foci after stretching in both instances (Fig. 4, F and H). We note that the *M. mycoides* chromatin appears to be relatively resistant to stretching as there is a similar number of foci and length (Fig. 4, C and D and fig. S8) despite its threefold greater length relative to 4q21 on YAC-*Mm*-4q21^{LacO}. A likely explanation is that the AT-rich *M. mycoides* sequence has denser chromatin relative to the human 4q21 sequence, which could be rendered by histone modifications and/or AT-hook protein-mediated compaction. Indeed, in addition to CENP-A–containing chromatin, a portion of the HAC DNA contains chromatin harboring H3K9me3 (fig. S6), suggesting that it harbors the heterochromatin typically found in pericentromeric regions and offering a means through which differences in chromatin stretching could be achieved (*43*, *44*). HAC stretching supports the notion of a chromatinized circular topology supporting propagation and inheritance in dividing cells.

Discussion

De novo HACs, similar to the ones we exhibit in this paper, are the only viable platform to generate a new mammalian chromosome where the entire DNA sequence can be designed in the lab. This presents an extremely wide horizon of possibilities for downstream biological and applied uses, and we report a system to create HACs that faithfully exist in their functional

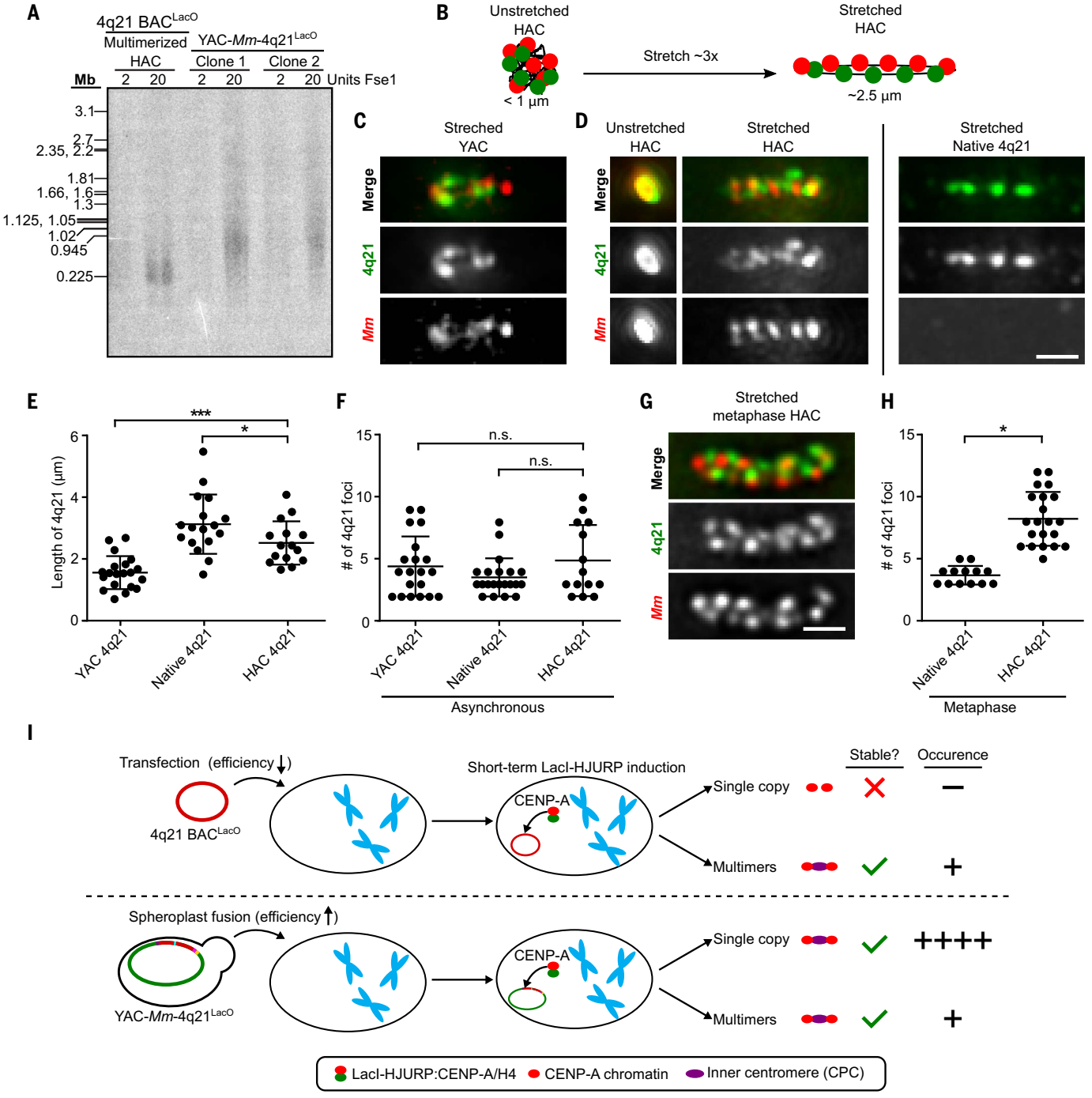

Fig. 4. YAC-*Mm*-4q21^{LacO}–based HACs are intact 760-kb circles with similar chromatin- stretching properties as natural chromosomes. (**A**) Southern blot analysis of the indicated HAC lines using a LacO probe. The multimerized HAC line was diluted 20-fold in HT1080 cells to reduce signal relative to the YAC-*Mm*-4q21^{LacO}-based HAC so that they could be in a similar range of detection. (**B**) Schematic showing extent of stretching HACs in our experiments (**D**) to (**H**), with indicated regions detected by FISH. The number of foci shown is in the range predicted by prior stretching experiments with natural chromosomes, with actual outcomes measured in (D) to (H) and fig. S8. (**C**) Representative images of a stretched YAC with both the 4q21 and *M. mycoides* sequence labeled through FISH compared with endogenous 4q21 in asynchronous cells. Bar, 1 µm. (D) Representative images of an unstretched and stretched HAC with both the 4q21 and *M. mycoides* sequence labeled through FISH compared with endogenous 4q21 in asynchronous cells. Bar, 1 µm. (**E**) Quantification of the length of 4q21 FISH in the YAC, HAC, and endogenous chromosome after stretching chromatin in asynchronous cells. The mean (± SD) is shown. $P < 0.001$ and $P < 0.05$ based on an unpaired, two-tailed *t* test. (**F**) Quantification of the number of foci from 4q21 FISH in the YAC, HAC and in the endogenous chromosome after stretching chromatin in asynchronous cells. The mean (± SD) is shown. *P* values > 0.05 are based on an unpaired, two-tailed *t* test and are marked as not significant (n.s.). (**G**) Representative images of a stretched HAC with both the 4q21 and *M. mycoides* sequence labeled through FISH after enriching for cells in metaphase. Bar, 1 µm. (H) Quantification of the number of foci from 4q21 FISH in the HAC and the endogenous chromosome after stretching chromatin and enriching for cells in metaphase. The mean (± SD) is shown. $P < 0.0001$ based on an unpaired, two-tailed *t* test. (**I**) Model illustrating how construct size influences HAC formation outcomes. Single-copy BAC-based HACs have never been isolated and are shown as too small to harbor a functional inner centromere.

form as a single copy. The advances are centered on the portion of the chromosome that controls segregation of the HAC at cell division, the centromere. Single-copy HAC formation requires establishment of a high local density of CENP-A nucleosomes that can self-propagate alongside natural centromeres (*16*), but this is not sufficient for centromere function. Rather, an entirely different chromatin domain, the inner centromere, must function in mitotic quality control and control of sister chromatid cohesion. The HACs formed from the YAC-*Mm*-4q21^{LacO} are large enough to harbor robust CENP-A arrays and inner centromeric chromatin (Fig. 2, Fig. 4I, and fig. S9).

The overhauled HAC cell delivery approach through spheroplast fusion was necessary because the initial construct is too large to reliably purify. It further has the benefit of being far more efficient than HAC construct transfections. YACs from spheroplasts will be packaged into chromatin, which may also contribute to the high rate of HAC formation upon introduction into mammalian cells. The overall high efficiency of HAC delivery and formation upon moving to spheroplast fusion–based delivery has important practical implications for the development and testing of specific features of HACs. In prior generations of HACs, rigorous testing of a modest number of constructs or cell lines required the isolation and subsequent cytological assessment of hundreds of cell lines that are cloned weeks after initial HAC construct transfection (*16*, *27*), because the initial selected cell populations are so sparse and HAC formation is so inefficient. With the YAC-*Mm*-4q21^{LacO} platform, we can measure high HAC formation efficiency in rapidly generated cell populations, obviating the need to isolate clonal lines. To quantify the proportion of HACs with the replicates shown in Fig. 1C, we would have needed to generate a minimum of hundreds of cell lines with our prior approach (*16*). One can easily envision a multitude of HAC features—including their genetic cargo and recipient cell types—that are attractive to test in the future. There are many potential opportunities for further optimization: some common to prior generations of HACs and some that arise because of the single-copy nature of YAC-*Mm*-4q21^{LacO}. For instance, gene silencing on HACs is a known challenge (*45*) that confounds expression of selection markers, reporters, and eventually other useful genetic payloads. The instability of all generations of HACs, including YAC-*Mm*-4q21^{LacO} HACs (Fig. 2C), relative to natural chromosomes could be affected by several modifiable factors. These include, but are not limited to, HAC topology (i.e., circular versus linear), size of pericentromeric heterochromatin (fig. S6), and the choice of stuffer DNA. For the latter, we note that although the extreme AT-richness in the stuffer *M. mycoides* DNA present in YAC-*Mm*-4q21^{LacO} may positively contribute to forming a functional inner centromere (Fig. 2), it results in barriers to routine high-throughput sequence-based approaches (fig. S4) that could otherwise be employed to map distinct chromatin features. YAC-*Mm*-4q21^{LacO} represents an attractive platform for HAC debugging—using synthetic budding yeast chromosomes as a model (*46*)—and for gaining new insights into mammalian chromosome biology while doing so. The central innovations with YAC-*Mm*-4q21^{LacO} are that it avoids the rampant multimerization that has complicated past HAC efforts and that it permits rapid assessment of HAC function (Fig. 1). The latter is bound to accelerate the pace of future HAC debugging.

Artificial chromosomes, more broadly, could deliver useful cargos for biomedical and industrial applications in a variety of eukaryotes. Because both the CENP-A–based epigenetic centromere specification mechanism and the inner centromeric dimensions and molecular constituency (e.g., sister centromere cohesion components and CPC) are common to diverse eukaryotic species, including in many agriculturally important plants, we envision that YAC-*Mm*-4q21^{LacO}-based artificial chromosomes will be readily modified and extended into many useful biological systems. For advancing the understanding of chromosomes, HACs can be used as testing grounds for designing what nature has evolved to ensure the stability of our genome, as well as designing the functional features that govern chromosome "outputs" in gene expression programs and epigenetic regulation. In summary, the advancements made in this study to HAC design, delivery, formation, and function will expedite both discovery-based and applied genome science.

REFERENCES AND NOTES

1. D. T. Burke, G. F. Carle, M. V. Olson, *Science* **236**, 806–812 (1987).
2. A. W. Murray, J. W. Szostak, *Nature* **305**, 189–193 (1983).
3. C. N. Traver, S. Klapholz, R. W. Hyman, R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 5898–5902 (1989).
4. E. D. Green, M. V. Olson, *Science* **250**, 94–98 (1990).
5. D. G. Gibson *et al*., *Science* **319**, 1215–1220 (2008).
6. D. G. Gibson *et al*., *Science* **329**, 52–56 (2010).
7. N. Annaluru *et al*., *Science* **344**, 55–58 (2014).
8. J. D. Boeke *et al*., *Science* **353**, 126–127 (2016).
9. R. K. Dawe, *Exp. Cell Res.* **390**, 111951 (2020).
10. E. J. Barrey, P. Heun, *Prog. Mol. Subcell. Biol.* **56**, 193–212 (2017).
11. J. J. Harrington, G. Van Bokkelen, R. W. Mays, K. Gustashaw, H. F. Willard, *Nat. Genet.* **15**, 345–355 (1997).
12. M. Ikeno *et al*., *Nat. Biotechnol.* **16**, 431–439 (1998).
13. K. Kixmoeller, P. K. Allu, B. E. Black, *Open Biol.* **10**, 200051 (2020).
14. K. E. Hayden *et al*., *Mol. Cell. Biol.* **33**, 763–772 (2013).
15. N. Kouprina *et al*., *ACS Synth. Biol.* **1**, 590–601 (2012).
16. G. A. Logsdon *et al*., *Cell* **178**, 624–639.e19 (2019).
17. J. Basu, H. F. Willard, *Trends Mol. Med.* **11**, 251–258 (2005).
18. M. Carmena, M. Wheelock, H. Funabiki, W. C. Earnshaw, *Nat. Rev. Mol. Cell Biol.* **13**, 789–803 (2012).
19. N. Altemose *et al*., *Science* **376**, eabl4178 (2022).
20. N. Altemose *et al*., *Nat. Methods* **19**, 711–723 (2022).
21. D. Hasson *et al*., *Nat. Struct. Mol. Biol.* **20**, 687–695 (2013).
22. G. A. Logsdon *et al*., *Nature* **593**, 101–107 (2021).
23. D. M. Brown *et al*., *Nucleic Acids Res.* **45**, e50 (2017).
24. N. Kouprina, V. Larionov, *Chromosoma* **125**, 621–632 (2016).
25. J. Song, F. Dong, J. W. Lilly, R. M. Stupar, J. Jiang, *Genome* **44**, 463–469 (2001).
26. T. Okada *et al*., *Cell* **131**, 1287–1300 (2007).
27. J. Ohzeki *et al*., *EMBO J.* **31**, 2391–2402 (2012).
28. K. Jülicher *et al*., *Genomics* **43**, 95–98 (1997).
29. T. A. Ebersole *et al*., *Hum. Mol. Genet.* **9**, 1623–1631 (2000).
30. A. E. Kelly *et al*., *Science* **330**, 235–239 (2010).
31. F. Wang *et al*., *Science* **330**, 231–235 (2010).
32. Y. Yamagishi, T. Honda, Y. Tanno, Y. Watanabe, *Science* **330**, 239–243 (2010).
33. E. A. Bassett *et al*., *J. Cell Biol.* **190**, 177–185 (2010).
34. M. D. Blower, B. A. Sullivan, G. H. Karpen, *Dev. Cell* **2**, 319–330 (2002).
35. S. A. Ribeiro *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10484–10489 (2010).
36. P. K. Allu *et al*., *Curr. Biol.* **29**, 2625–2639.e5 (2019).
37. I. Erliandri *et al*., *Nucleic Acids Res.* **42**, 11502–11516 (2014).
38. E. Pesenti *et al*., *ACS Synth. Biol.* **9**, 3267–3287 (2020).
39. V. N. Noskov *et al*., *Nat. Protoc.* **6**, 89–96 (2011).
40. E. Kyriacou, P. Heun, *Epigenetics Chromatin* **11**, 68 (2018).
41. Z. Y. Tan *et al*., *eLife* **12**, RP87672 (2023).
42. S. Cai, D. Böck, M. Pilhofer, L. Gan, *Mol. Biol. Cell* **29**, 2450–2457 (2018).
43. A. H. F. M. Peters *et al*., *Mol. Cell* **12**, 1577–1589 (2003).
44. J. C. Rice *et al*., *Mol. Cell* **12**, 1591–1598 (2003).
45. N. C. O. Lee *et al*., *Cell. Mol. Life Sci.* **70**, 3723–3737 (2013).
46. Y. Zhao *et al*., *Cell* **186**, 5220–5236.e16 (2023).

ACKNOWLEDGMENTS

We thank our UPenn colleagues M. Lampson (for comments on the manuscript) and M. Gerace (for assistance with preparing reagents). We also thank E. Makeyev (King's College) for reagents. **Funding:** This work was funded by the following: National Institutes of Health grant GM130302 (to B.E.B.) National Institutes of Health grant HG012445 (to B.E.B. and J.I.G.) National Institutes of Health grant CA261198 (to K.K.) National Institutes of Health grant GM007229 (to G.J.B.) **Author contributions:** C.W.G., D.M.B., J.I.G., and B.E.B. conceived the project. C.W.G., G.J.B., E.M., D.M.B., G.Y., K.K., J.N.G., and P.K. performed experiments. C.W.G., G.J.B., E.M., G.Y., J.N.G., J.L.E., C.L.D., P.H., and B.E.B. analyzed data. C.W.G., D.M.B., and G.A.L. generated reagents. C.W.G. and B.E.B. wrote the paper. All authors edited the paper. B.E.B. supervised the project. **Competing interests:** C.W.G., G.J.B., D.M.B., J.I.G., and B.E.B. are inventors on a patent application submitted by UPenn related to this work. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials or have been uploaded to SRA (PRJNA985068). The materials used in this study are available from commercial sources or from the corresponding authors on reasonable request.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adf2572
Materials and Methods
Figs. S1 to S10
Table S1
References (*47*–*55*)
MDAR Reproducibility Checklist

Submitted 21 June 2023; accepted 23 January 2024
10.1126/science.adj3566