PI: **Collins, Kathleen**	Title: Human genetic supplementation without donor DNA or a DNA break	
Received: 09/06/2019	Opportunity: RFA-RM-19-005 Clinical Trial:Optional	Council: 05/2020
Competition ID: FORMS-E	FOA Title: NIH Directors Pioneer Award Program (DP1 Clinical Trial Optional)	
1DP1HL156819-01	Dual: RM,OD	Accession Number: 4346124
IPF: 577502	Organization: UNIVERSITY OF CALIFORNIA BERKELEY	
Former Number: 1DP1OD029727-01	Department: Molecular and Cell Biology	
IRG/SRG: ZRG1 BCMB-N (50)R	AIDS: N	Expedited: N
Subtotal Direct Costs (excludes consortium F&A) Year 1: 3,500,000	Animals: N Humans: N Clinical Trial: N Current HS Code: 10 HESC: N HFT: N	New Investigator: N Early Stage Investigator: N
Senior/Key Personnel:	*Organization:*	*Role Category:*
Kathleen Collins	The Regents of the University of California	PD/PI

Reference Letters

[redacted]

OMB Number: 4040-0001
Expiration Date: 10/31/2019

APPLICATION FOR FEDERAL ASSISTANCE
SF 424 (R&R)

3. DATE RECEIVED BY STATE	**State Application Identifier**

1. TYPE OF SUBMISSION*

❍ Pre-application ● Application ❍ Changed/Corrected Application

4.a. Federal Identifier

b. Agency Routing Number
6 MCB; 2 CB

2. DATE SUBMITTED | **Application Identifier**

c. Previous Grants.gov Tracking Number

5. APPLICANT INFORMATION

Organizational DUNS*: [redacted]

Legal Name*: The Regents of the University of California
Department: Sponsored Projects Office
Division:
Street1*: [redacted]
Street2:
City*:
County:
State*:
Province:
Country*:
ZIP / Postal Code*:

Person to be contacted on matters involving this application
Prefix: First Name*: Margaret Middle Name: Last Name*: Nguyen Suffix:
Position/Title: Contract and Grant Officer
Street1*: [redacted]
Street2:
City*:
County:
State*:
Province:
Country*:
ZIP / Postal Code*:
Phone Number*: [redacted] Fax Number: [redacted] Email: [redacted]

6. EMPLOYER IDENTIFICATION NUMBER *(EIN) or (TIN)** [redacted]

7. TYPE OF APPLICANT* H: Public/State Controlled Institution of Higher Education

Other (Specify):
Small Business Organization Type ❍ Women Owned ❍ Socially and Economically Disadvantaged

8. TYPE OF APPLICATION*

● New ❍ Resubmission
❍ Renewal ❍ Continuation ❍ Revision

If Revision, mark appropriate box(es).
❍ A. Increase Award ❍ B. Decrease Award ❍ C. Increase Duration
❍ D. Decrease Duration ❍ E. Other *(specify)* :

Is this application being submitted to other agencies?* ❍Yes ●No What other Agencies?

9. NAME OF FEDERAL AGENCY*
National Institutes of Health

10. CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER
TITLE:

11. DESCRIPTIVE TITLE OF APPLICANT'S PROJECT*
Human genetic supplementation without donor DNA or a DNA break

12. PROPOSED PROJECT
Start Date*: 09/30/2020
Ending Date*: 07/31/2025

13. CONGRESSIONAL DISTRICTS OF APPLICANT
CA-013

SF 424 (R&R) APPLICATION FOR FEDERAL ASSISTANCE

14. PROJECT DIRECTOR/PRINCIPAL INVESTIGATOR CONTACT INFORMATION

Prefix: First Name*: Kathleen Middle Name: Last Name*: Collins Suffix:
Position/Title: Professor
Organization Name*: The Regents of the University of California
Department: [redacted]
Division:
Street1*: [redacted]
Street2: [redacted]
City*: [redacted]
County: [redacted]
State*: [redacted]
Province:
Country*: [redacted]
ZIP / Postal Code*: [redacted]
Phone Number*: [redacted] Fax Number: Email*: [redacted]

15. ESTIMATED PROJECT FUNDING

a. Total Federal Funds Requested*	$3,500,000.00
b. Total Non-Federal Funds*	$0.00
c. Total Federal & Non-Federal Funds*	$3,500,000.00
d. Estimated Program Income*	$0.00

16.IS APPLICATION SUBJECT TO REVIEW BY STATE EXECUTIVE ORDER 12372 PROCESS?*

a. YES ○ THIS PREAPPLICATION/APPLICATION WAS MADE AVAILABLE TO THE STATE EXECUTIVE ORDER 12372 PROCESS FOR REVIEW ON:
DATE:

b. NO ● PROGRAM IS NOT COVERED BY E.O. 12372; OR
○ PROGRAM HAS NOT BEEN SELECTED BY STATE FOR REVIEW

17. By signing this application, I certify (1) to the statements contained in the list of certifications* and (2) that the statements herein are true, complete and accurate to the best of my knowledge. I also provide the required assurances * and agree to comply with any resulting terms if I accept an award. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties. (U.S. Code, Title 18, Section 1001)

● I agree*

** The list of certifications and assurances, or an Internet site where you may obtain this list, is contained in the announcement or agency specific instructions.*

18. SFLLL or OTHER EXPLANATORY DOCUMENTATION File Name:

19. AUTHORIZED REPRESENTATIVE

Prefix: First Name*: Margaret Middle Name: Last Name*: Nguyen Suffix:
Position/Title*: Contract and Grant Officer
Organization Name*: The Regents of the University of California
Department: [redacted]
Division:
Street1*: [redacted]
Street2: [redacted]
City*: [redacted]
County: [redacted]
State*: [redacted]
Province:
Country*: [redacted]
ZIP / Postal Code*: [redacted]
Phone Number*: [redacted] Fax Number: [redacted] Email*: [redacted]

Signature of Authorized Representative*
Margaret Nguyen

Date Signed*
09/06/2019

20. PRE-APPLICATION File Name:

21. COVER LETTER ATTACHMENT File Name:

424 R&R and PHS-398 Specific
Table Of Contents

OMB Number: 4040-0010
Expiration Date: 10/31/2019

Project/Performance Site Location(s)

Project/Performance Site Primary Location

❍ I am submitting an application as an individual, and not on behalf of a company, state, local or tribal government, academia, or other type of organization.

Organization Name: The Regents of the University of California
Duns Number: [redacted]
Street1*: [redacted]
Street2: [redacted]
City*: [redacted]
County: [redacted]
State*: [redacted]
Province:
Country*: [redacted]
Zip / Postal Code*: [redacted]
Project/Performance Site Congressional District*: CA-013

Additional Location(s) File Name:

Funding Opportunity Number: RFA-RM-19-005. Received Date: 2019-09-06T13:01:39.000-04:00

OMB Number: 4040-0001
Expiration Date: 10/31/2019

RESEARCH & RELATED Other Project Information

1. Are Human Subjects Involved?* ❍ **Yes** ● **No**

1.a. If YES to Human Subjects

Is the Project Exempt from Federal regulations? ❍ Yes ❍ No

If YES, check appropriate exemption number: _ 1 _ 2 _ 3 _ 4 _ 5 _ 6 _ 7 _ 8

If NO, is the IRB review Pending? ❍ **Yes** ❍ **No**

IRB Approval Date:

Human Subject Assurance Number

2. Are Vertebrate Animals Used?* ❍ **Yes** ● **No**

2.a. If YES to Vertebrate Animals

Is the IACUC review Pending? ❍ Yes ❍ No

IACUC Approval Date:

Animal Welfare Assurance Number

3. Is proprietary/privileged information included in the application?* ❍ **Yes** ● **No**

4.a. Does this project have an actual or potential impact - positive or negative - on the environment?* ❍ **Yes** ● **No**

4.b. If yes, please explain:

4.c. If this project has an actual or potential impact on the environment, has an exemption been authorized or an environmental assessment (EA) or environmental impact statement (EIS) been performed? ❍ Yes ❍ No

4.d. If yes, please explain:

5. Is the research performance site designated, or eligible to be designated, as a historic place?* ❍ **Yes** ● **No**

5.a. If yes, please explain:

6. Does this project involve activities outside the United States or partnership with international collaborators?* ❍ **Yes** ● **No**

6.a. If yes, identify countries:

6.b. Optional Explanation:

	Filename
7. Project Summary/Abstract*	CollinsAbstract.pdf
8. Project Narrative*	CollinsNarrative.pdf
9. Bibliography & References Cited	
10.Facilities & Other Resources	D_Facilities.pdf
11.Equipment	

Funding Opportunity Number: RFA-RM-19-005. Received Date: 2019-09-06T13:01:39.000-04:00

ABSTRACT

Human genome engineering has widely anticipated promise as a healthcare strategy, but current technologies are unlikely to provide the safe, efficient, and broadly useful implementation of transgene introduction essential to complete the next big leap forward for gene therapy. CRISPR-based approaches for transgene integration have major impediments, including the need for donor DNA delivery, the propensity of that DNA to undergo non-specific integration, and the low efficiency of repair by homologous recombination relative to sloppy rejoining of the broken DNA ends. Also severely limiting is the fact that slowly proliferating cells are rarely in a cell cycle phase favorable for homologous recombination, and just the presence of a DNA break can be toxic. The alternative approach of adeno-associated virus introduction of a transgene also has limitations, among others including the small transgene size permitted by the virus capsid and the challenges of engineering virus uptake into different cell types. It remains an unmet need to have a non-mutagenic, non-toxic approach for gene introduction to the human genome. Therapy for many loss-of-function pathologies hinges on this missing technology. Also, only transgene introduction offers the opportunity for non-native control of protein expression, isoform selectivity, and myriad other clinically useful outcomes.

Starkly missing from current efforts to develop transgene introduction techniques is an approach exploiting the gene insertion strategy widespread endogenously across eukaryotes: cDNA synthesis. The ancestral, evolutionarily persistent type of eukaryotic LINE/non-LTR retroelement integrates by nick-primed reverse transcription that is rigorous both it its sequence specificity of target site selection and in its specificity for use of an RNA transcript with the retroelement 3' UTR as template. The biochemical activities required for target site selection, introduction of precisely positioned nick, and cDNA synthesis are carried out by a single protein. Any RNA sequence flanked by 5' and 3' regions of the retroelement genome should assemble with a favorably modified retroelement protein, and this RNP would then seek its native insertion site. Because several LINE/non-LTR retroelement families target highly conserved, repetitive sequences invariant across multicellular eukaryotes, there is no need to re-engineer DNA site-specificity of these retroelement proteins, although that may become of interest to undertake. The simple architecture of the non-LTR retroelements begs to be exploited for developing an approach to human genome supplementation with genes of therapeutic impact. The novelty of this approach demands continuous innovation and obliges high risk of failure to reach the goal of delivering an engineered RNP capable of transgene introduction into human cells. Success of this strategy would usher in a new modality of therapeutic treatment for loss-of-function diseases.

NARRATIVE

Human genome engineering has widely anticipated promise as a healthcare strategy, but technologies in development are unlikely to safely or efficiently introduce a gene into the human genome. Starkly missing from transgene introduction strategies is an approach based on what is already quite successful across eukaryotic evolution: gene insertion by synthesis of an RNA-templated complementary DNA (cDNA). The evolutionary success of RNA template-specific, DNA site-specific LINE/non-LTR retroelement insertion begs to be engineered to develop a transgene introduction technology for human gene therapy.

FACILITIES AND OTHER RESOURCES

Scientific Environment
UC Berkeley is a widely cross-disciplinary and enthusiastically collaborative hub for basic research with relevance to the improvement of human health. Collins lab research benefits from the breadth and depth of expertise on campus and nearby through mechanisms including super-group meetings, Bay Area RNA clubs, and direct collaborations with colleagues including our published collaborations with campus labs in the Departments of Statistics, Physics, and our own MCB. There are almost daily seminars, frequent Symposia, annual retreats, and numerous relevant international meetings.

Laboratory
The Collins lab has ~1,800 square feet of space on the fifth floor of Barker Hall on the UC Berkeley campus. This space includes 14 benches with desks, including common benches for gel electrophoresis and other equipment. Additional rooms within the Collins lab space include a computer & lunch room, a room devoted to radioactive work, and a tissue culture room with several incubators, a microscope, and a 6' hood with a biolistic transformation appartaus. We have an in-lab liquid nitrogen and cryotank, chromatography systems, a nanodrop and plate reader, gel driers and speed vac, multiple hybridization ovens, PCR machines and other molecular biology tools, wavelenth-specific crosslinking light sources, and a water distillation system. Adjacent areas have LiCOR and Typhoon imaging systems and a digital imaging system for gel documentation. Controlled temperature rooms nearby are equiped with large-scale cell growth shakers. On neighboring floors we use electroporation and nucleofection machines, dry ice, a tape station, cell disruption equipment, and other tools.

Computer
The laboratory includes a computer room with two networked Macs and one PC that are common use, printers, and a gel scanner. Mobile laptops tethered to desks are used wirelessly throughout the lab. Journal articles are available online by campus library subscriptions and web-based document delivery.

Office
An office directly adjoins the lab. The office contains an additional Mac, printer, and scanner.

Animal Use
The campus Northwest Animal Facility provides consultation for antibody production and ordering. They approve the use of commercial antibody vendors to ensure appropriate standards of animal care.

Other Facilities
We are users of campus mass spectrometry facilities, a facility for fluorescence microscopy, a flow cytometry facility, and an Illumina and PacBio NGS sequencing facility. In the same building we use a resources from the departmental tissue culture facility, low-throughput DNA sequencing service, quantitative real-time PCR machines, large-scale bacterial and eukaryotic cell culture facilities, carpentry and equipment repair shops, a self-service stockroom, and myriad other resources.

OMB Number: 4040-0001
Expiration Date: 10/31/2019

RESEARCH & RELATED Senior/Key Person Profile (Expanded)

PROFILE - Project Director/Principal Investigator

Prefix: First Name*: Kathleen Middle Name Last Name*: Collins Suffix:

Position/Title*: Professor
Organization Name*: The Regents of the University of California
Department: [redacted]
Division:
Street1*: [redacted]
Street2: [redacted]
City*: [redacted]
County: [redacted]
State*: [redacted]
Province:
Country*: [redacted]
Zip / Postal Code*: [redacted]

Phone Number*: [redacted] Fax Number:

E-Mail*: [redacted]

Credential, e.g., agency login: [redacted]

Project Role*: PD/PI Other Project Role Category:

Degree Type: PHD,MS,BS Degree Year: 1992,1987,1987

Attach Biographical Sketch*: File Name: D_Bio_Collins.pdf

Attach Current & Pending Support: File Name: Collins_Other_Support_revised.pdf

BIOGRAPHICAL SKETCH

NAME: Collins, Kathleen

eRA COMMONS USER NAME: [redacted]

POSITION TITLE: Professor of Biochemistry, Biophysics & Structural Biology

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE	Completion Date	FIELD OF STUDY
Yale University, CT	B.S./M.S.	05/1987	Molecular Biophysics & Biochemistry
Massachusetts Institute of Technology, MA	Ph.D.	05/1992	Biology
Cold Spring Harbor Laboratory, NY	Postdoctoral	05/1995	Biology

A. Personal Statement

My laboratory's primary accomplishments have been towards understanding mechanisms and regulations of telomerase, a eukaryotic RNP reverse transcriptase specialized for telomeric repeat synthesis. We were the first to reconstitute recombinant telomerase that could be assayed directly by primer extension (Collins and Gandhi 1998), to demonstrate that telomerase RNA (TER) motifs control specific features of the repeat-synthesis catalytic cycle (Licht and Collins 1999; Miller and Collins 2002; Lai, Miller and Collins 2003), to produce and characterize an autonomously functional RNA binding domain of telomerase reverse transcriptase (TERT) (Lai, Mitchell and Collins 2001; O'Connor, Lai and Collins 2005), and to demonstrate hierarchical steps of telomerase protein and RNA folding required for assembly of active RNP (Prathapam et al. 2005; O'Connor and Collins 2006; Stone et al. 2007). For 20 years we have been field-leading pioneers in discovering and defining endogenously assembled telomerase holoenzymes (starting with Mitchell et al. 1999b for human & Witkin and Collins 2004 for *Tetrahymena*). With several collaborators we contributed atomic-resolution structures of telomerase proteins, RNA motifs, and RNP and protein-DNA complexes (for example, Zeng et al. 2011; Singh et al. 2012; Jiang et al. 2015; Wan et al., 2015), with biochemical assays to elucidate domain structure/function (for example, Min and Collins 2010; Wu and Collins 2014). With EM collaborators, we determined ciliate and human telomerase holoenzyme structures (Jiang et al. 2013; Hong et al. 2013; Jiang et al. 2015; Nguyen et al. 2018).

In our human telomerase studies we discovered the presence and the biological significance of the vertebrate telomerase H/ACA RNA motif for telomerase RNA precursor processing and RNP assembly (Mitchell, Chen and Collins 1999), made the first link between telomerase deficiency and human disease (Mitchell, Wood and Collins 1999), and established disease-linked loss-of-function mechanisms (Fu and Collins, 2003; Wong and Collins 2006; Errington et al. 2008; Robart and Collins 2010; Egan and Collins 2012; Vogan et al. 2016; Nguyen et al. 2018). We elucidated surprising principles of telomerase mechanism by finessing telomerase reconstitution strategies (Mitchell and Collins, 2000; Robart and Collins 2011; Wu and Collins 2014; Wu et al. 2015; Wu, Tam and Collins 2017). We are defining the biochemical and genetic requirements for telomerase action at telomeres in human stem and tumor cell lines using genome engineering (Sexton et al. 2014; Vogan et al. 2016; Wu, Tam and Collins 2017).

Other pioneering Collins lab accomplishments relate to our discoveries of non-coding RNAs and ncRNA RNP biogenesis and function. We discovered strand-specific loading of Piwi proteins with endogenous double-stranded RNA (Lee and Collins 2006), stress-induced eukaryotic tRNA anticodon loop cleavage (Lee and Collins 2005; Andersen and Collins 2012), and unprecedented RNA binding activity for an interferon-induced anti-viral protein (Katibah et al. 2013; Katibah et al. 2014). We developed robust methods for entirely RNA-based purifications to homogeneity of endogenously assembled RNPs, including the first recombinant use of the phage PP7 coat protein hairpin now a widely used tool (Hogg and Collins 2007a, 2007b, 2008). Our studies of eukaryotic RNA-dependent RNA polymerase complexes and endogenous RNA silencing pathways provided remarkable new paradigms for diverse cellular specificities of transcript recognition and processing (Lee and Collins 2007; Lee, Talsky and Collins 2009; Couvillion et al. 2009; Talsky and Collins 2010; Couvillion, Sachidanandam and Collins 2010; Couvillion et al. 2012; Talsky and Collins 2012; Farley and Collins 2017).

B. Positions and Honors

Positions and Employment

1990 (June/July)	Visiting Scientist, Institute of Physical and Chemical Research (RIKEN), Japan
1992-1995	Postdoctoral Fellow, Cold Spring Harbor Laboratory
1995-2001	Assistant Professor, University of California at Berkeley
2001-2004	Associate Professor, University of California at Berkeley
2004-present	Professor, University of California at Berkeley
2008	Visiting Scientist, New England Biolabs, Inc.
2008-2009	Visiting Scientist, Tetragenetics, Inc.

Other Experience and Professional Memberships

1989-present	Member ASCB, ASBMB, ASM, RNA Society
1992	Teaching Assistant, ASCB workshop “Protein and Peptide Purification”
1997	Author of “Working with Proteins” educational video, Cogito Learning Media, Inc.
1999-2017	Member *ad hoc* NIH study sections BIO2 (1999*)*, CDF2 (2001), MGA (2006-2007, 2017), Special Emphasis Panel (2012)
2001-2015	Review Panel (2001-2006), Executive Policy Committee (2005-2009), and Fellowships Coordinator (2007-2015) for the UC Cancer Research Coordinating Committee
2003-2006	Monitoring Editor, *The Journal of Cell Biology*
2003-2006	Scientific Advisory Board, *Tetrahymena* genome projects
2007-2011	Project Lead for *Tetrahymena* micronuclear genome pilot sequencing through the JGI (Joint Genome Institute) Community Sequencing Program
2007-2017	Chair and Member, NIH Study Section MGA (Molecular Genetics A)
2007-2013	Program Director, NIH T32 GM07232 (MCB department predoctoral training grant)
2008-2009	Visiting Scientist, New England Biolabs and Tetragenetics (May-December)
2010	Co-organizer, CBD Division Symposium “Non-coding RNA Frontiers”
2011	Study section chair panel, NIH New Faculty Investigator Conference, CSHL
2011-2012	Board of Directors, RNA Society
2011-present	Associate Editor, *RNA*
2011	Editor, Methods in Cell Biology volume 109 "*Tetrahymena thermophila*"
2012-2016	Chair and Member, Publications Committee, ASBMB
2013	Chair, NIH Common Fund Study Section “Extracellular RNA Biogenesis, Biodistribution, Uptake, and Effector Function”
2014	Chair, NIH Common Fund Study Section “Defining A Comprehensive Reference Profile of Circulating Human Extracellular RNA”
2014-2015	Executive Committee, Miller Institute for Basic Research, UC Berkeley
2014	Decennial Review Committee, University of Texas, Austin Biosciences
2015-2017	Board of Reviewing Editors, *eLife*
2016	Member *ad hoc*, National Advisory General Medical Sciences Council
2017-present	Steering Committee and Faculty Lead, Science@Cal
2017-2019	Chair and Vice-Chair, Gordon Conference “Nucleic Acids”
2018	Co-organizer, CNIO Meeting "Molecular, Cellular and Organismal Drivers of Aging"
2019	Decennial Review Committee, UMass Medical School Dept Biochem & Mol Pharmacol

Honors

1992-1995	Life Sciences Research Foundation Postdoctoral Fellowship
1998-2001	Burroughs Wellcome Fund New Investigator Award in the Pharmacological Sciences
2002	American Society for Cell Biology/Promega Early Career Life Scientist Award
2010-2011	Miller Professor, University of California at Berkeley
2015-present	Walter and Ruth Schubert Family Chair

C. Contribution to Science

1. ***Telomerase deficiency imposes human disease.*** Collins lab discoveries established the unanticipated, essential role of human telomerase activation within a single human lifespan. We discovered that the X-linked form of human dyskeratosis congenita (DC), caused by dyskerin amino acid substitutions, arises from a partial telomerase deficiency that imposes premature bone marrow failure in the second decade of life. From studies of the fundamental properties of human telomerase, we bootstrapped into the realization that dyskerin is associated with a motif in human telomerase RNA (hTR) shared with the H/ACA-motif small nucleolar and small Cajal body RNAs required for maturation of ribosomal and spliceosomal RNAs. We used primary fibroblasts, lymphocytes, and lymphoblasts from patients with DC and family-matched controls, obtained from repositories and clinician collaborations, to show that DC is a telomerase deficiency. Building from our insights, we and others showed that other telomerase holoenzyme subunit mutations cause autosomal dominant and recessive forms of DC and other tissue renewal failures including aplastic anemia and pulmonary fibrosis. Based on our pioneering connection of human telomerase deficiency to disease, telomere length has become an established clinical diagnostic and determining factor for a tailored clinical therapy that promotes patient survival.

Beyond these initial insights, we persevered to define the mechanisms by which telomerase subunit mutations in dyskerin, hTR, and telomerase reverse transcriptase (TERT) impose a spectrum of disease phenotypes. Our studies of disease mechanism are unparalleled in the field. We demonstrated that rescuing telomerase deficiency rescues all molecular and cellular phenotypes of DC patient cells. We solved the mystery of how gene mutations affecting H/ACA RNP proteins impose telomerase-specific defects in human diseases. The hTR precursor is an independent transcript with an exonuclease-accessible 3' end, whereas other H/ACA RNAs are processed from exonuclease-resistant intron lariats. Therefore, hTR biogenesis is differentially sensitized to compromise of the protein scaffold for initial H/ACA RNP assembly, which we were the first to characterize biochemically. Our insights about H/ACA RNP assembly mechanism explain the tandem-hairpin architecture of all eukaryotic small nucleolar and Cajal body H/ACA RNAs and rationale for telomerase-specific phenotypes of H/ACA protein mutations.

a. Mitchell, J. R., Wood, E. and Collins, K. A novel telomerase component is defective in the human disease dyskeratosis congenita. ***Nature*** 402: 551-5 (1999).
b. Wong JMY, Collins K. Telomerase RNA deficiency limits telomere length maintenance in X-linked dyskeratosis congenita. ***Genes Dev*** 20: 2848-58 (2006).
c. Egan ED, Collins K. An enhanced H/ACA RNP assembly mechanism for human telomerase RNA. ***MCB*** 32: 2428-39 (2012).
d. Vogan JM, Zhang X, Youmans DT, Regalado SG, Johnson JZ, Hockemeyer D, Collins K. Minimized human telomerase maintains telomeres and resolves endogenous roles of H/ACA proteins, TCAB1, and Cajal bodies. ***eLife*** 5: e18221 1-21 (2016).

2. ***Telomerase holoenzymes.*** We approached the need to understand telomerase enzyme biogenesis and function in cellular context by identifying and characterizing *Tetrahymena* and human telomerase holoenzyme proteins. Our *Tetrahymena* telomerase holoenzyme purification, subunit identification, genetic characterization, and reconstitution from individual subunits provided long-awaited opportunities to define enzyme structure and function at coordinated cellular and structural levels. From our combination of approaches, a general subunit architecture of telomerase holoenzyme emerged: the RNP catalytic core (p65-TER-TERT) has a connector hub (p50) coordinating two different RPA-like heterotrimer sub-complexes. One of these sub-complexes, TEB, combines telomerase-specific large-subunit Teb1 with two smaller subunits shared with *Tetrahymena* RPA. We demonstrated that the TEB sub-complex mediates telomerase recruitment to telomeres as well as other functions. The other RPA-like holoenzyme-associated sub-complex, CST (p75-p45-p19), is a conserved negative regulator of telomerase and activator of complementary-strand synthesis by Polymerase α – Primase.

This body of work and parallel studies characterizing human telomerase holoenzyme are described in publications from the lab spanning nearly two decades. As a result of this persistent effort, we have at last accomplished structural shapshots of telomerase holoenzymes' organization at the subunit level, approaching atomic resolution. Structural characterization of a multitude of holoenyzme complexes was done as close collaboration, for *Tetrahymena* telomerase with previous Collins lab sabbatical visitor

Professor Juli Feigon and her UCLA neighbor Professor Hong Zhou and for human telomerase in collaboration with Professor Eva Nogales at Berkeley. These structures yielded breakthrough revelations about composition, domain architecture, and function of telomerase RNA (TER) and TERT.

a. Jiang J, Miracco EJ, Hong K, Eckert B, Chan H, Cash DD, Min B, Zhou ZH*, Collins K*, Feigon J* (*co-corresponding). The architecture of *Tetrahymena* telomerase holoenzyme. ***Nature*** 496: 187-92 (2013).
b. Hong K, Upton H, Miracco EJ, Jiang J, Zhou ZH, Feigon J*, Collins K* (*co-corresponding). *Tetrahymena* telomerase holoenzyme assembly, activation, and inhibition by domains of the p50 central hub. ***MCB*** 33: 3962-71 (2013).
c. Jiang J, Chan H, Cash DD, Miracco EJ, Loo RRO, Upton HE, Cascio D, Johnson RO, Collins K, Loo JA, Zhou ZH, Feigon J. Structure of *Tetrahymena* telomerase reveals previously unknown subunits, functions, and interactions. ***Science*** 350: aab4070 1-10 (2015).
d. Nguyen THD, Tam J, Wu RA, Greber BJ, Toso D, *Nogales E, *Collins K. Cryo-EM structure of substrate-bound human telomerase holoenzyme. ***Nature*** 557:190-5 (2018). (*co-corresponding).

3. ***Principles of RNA and protein co-folding****.* In our studies of telomerase RNP biogenesis in human cells and *Tetrahymena*, we expanded the knowledge of fundamental principles for cellular RNP biogenesis. We showed, first for the *Tetrahymena* telomerase RNP catalytic core, that active telomerase RNP folds in a hierarchical cascade of induced fit. One *Tetrahymena* telomerase holoenzyme protein that we identified, p65, is genetically essential for accumulation of TERT and TER. Using bacterially expressed p65 and the first functional purified TERT RNA-binding domain polypeptides, we reconstituted p65-dependent TERT RNP assembly. In collaboration we used a single-molecule FRET approach to resolve individual steps of p65-mediated TERT RNP assembly. Direct observations in real time established a hierarchical RNP assembly pathway: p65-TER interaction induces a TER structural rearrangement that in turn creates the binding surface for TERT, illuminating the first function for a telomerase holoenzyme protein other than TERT. Hierarchical folding principles also explain the function of the most conserved element of TER: a unique pseudoknot that must be delayed in its formation until the template loop threads around the entire circumference of TERT.

a. Prathapam R, Witkin KL, O'Connor CM, Collins K. A telomerase holoenzyme protein enhances telomerase RNA assembly with telomerase reverse transcriptase. ***NSMB*** 12: 252-7 (2005).
b. O'Connor CM, Collins K. A novel RNA binding domain in *Tetrahymena* telomerase p65 initiates hierarchical assembly of telomerase holoenzyme. ***MCB*** 26: 2029-36 (2006).
c. Stone MS, Mihalusova M, O'Connor CM, Prathapam R, Collins K, Zhuang X. Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. ***Nature*** 446: 458-61 (2007).
d. Singh M, Wang Z, Koo BK, Patel A, Cascio D, Collins K, Feigon J. Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. ***Mol Cell*** 47: 16-26 (2012).

4. ***Unique nucleic acid recognition properties critical for telomeric repeat synthesis****.* The Collins lab changed the paradigm for understanding telomerase by demonstrating that the RNA subunit has both template-related and template-independent functions. We showed that beyond providing an internal template, TER is a required co-factor for TERT activity. Using circularly permuted variants of TER with or without deletions, we showed that TERT could copy an exchangeable oligonucleotide template. However, *trans*-template copying required the presence of the template-less remainder of TER! This work convincingly resolved TER function as a template from its other functions and led us and others to define several independently folded TER motifs with specific roles in enzyme activity and regulation.

We then defined dynamic TERT and hTR contacts across the complex catalytic cycle of repeat synthesis, first using cellular RNP assembly to ensure physiological specificity then reconstituting this specificity in vitro. Our studies revealed a complex RNP domain architecture that places and holds template-primer duplex and single-stranded product within the RNP. We made the discovery that TERT ring is sufficient to bind DNA substrate but can only elongate it by addition of a single repeat. The TEN domain confers repeat synthesis processivity by recapturing the very short template-product duplex formed at the template 3' end. By pressing the technical envelope of *trans*-active domain reconstitution, we changed the dogma for telomerase enzyme specialization with the discovery of new specificities of template-paired and single-stranded DNA handling.

a. Lai CK, Miller MC, Collins K. Roles for RNA in telomerase nucleotide and repeat addition processivity. ***Mol Cell*** 11: 1673-83 (2003).
b. Robart AR, Collins K. Networked human telomerase domain interactions capture the template-paired substrate and single-stranded product for processive elongation. ***Mol Cell*** 42: 308-18 (2011).
c. Wu RA, Collins K. Human telomerase specialization for repeat synthesis by unique handling of primer-template duplex. ***EMBO J*** 8: 921-35 (2014).
d. Wu RA, Tam J, Collins K. DNA-binding determinants and cellular thresholds for human telomerase repeat addition processivity. ***EMBO J*** 36: 1908-27 (2017).

5. ***Telomerase action at telomeres.*** The Collins lab has made critical contributions to understanding the still-murky requirements for telomerase action at telomeres and regulation underlying telomere length homeostasis. For *Tetrahymena* telomerase, using panels of genome-engineered cell lines expressing biochemically characterized selective loss-of-function alleles, we showed that direct telomerase-DNA interaction is necessary for physical recruitment of telomerase to telomeres. We characterized genetic requirements for cell-cycle regulated telomere recruitment of telomerase holoenzyme subunits, G-strand synthesis by telomerase, and complementary C-strand synthesis.

In human cells we defined the specificity of telomerase interaction with the telomere shelterin protein TPP1, then we demonstrated that this interaction is genetically essential for telomerase action at telomeres. Our studies of human telomerase subunits, TPP1, and TPP1-interacting proteins tease apart multi-layered positive and negative regulations of human telomerase action at telomeres. Genome engineering of human cancer cell lines in the Collins lab paralleled by genome engineering of human pluripotent stem cells in the nearby lab of our collaborator Dirk Hockemeyer allows us to address whether layers of telomerase regulation have similar or different mechanisms in human stem versus cancer cells.

a. Sexton AN, Youmans DT, Collins K. Specificity requirements for human telomere protein interaction with telomerase holoenzyme. ***JBC*** 287: 34455-64 (2012).
b. Upton HE, Hong K, Collins K. Direct single-stranded DNA binding by Teb1 mediates the recruitment of *Tetrahymena* telomerase to telomeres. ***MCB*** 34: 4200-12 (2014).
c. Sexton AN, Regalado SG, Lai CS, Cost GJ, O'Neil CM, Urnov FD, Gregory PD, Jaenisch R, Collins K*, Hockemeyer D* (*co-corresponding). Genetic and molecular identification of three human TPP1 functions in telomerase action: Recruitment, activation, and homeostasis set-point regulation. ***Genes Dev*** 28: 1885-99 (2014).
d. Vogan JM, Zhang X, Youmans DT, Regalado SG, Johnson JZ, Hockemeyer D, Collins K. Minimized human telomerase maintains telomeres and resolves endogenous roles of H/ACA proteins, TCAB1, and Cajal bodies. ***eLife*** 5: e18221 1-21 (2016).

Complete List of Published Work in MyBibliography:
http://www.ncbi.nlm.nih.gov/myncbi/browse/collection/48105283/?sort=date&direction=ascending

D. Research Support

ACTIVE
R35 GM 130315-01 (Collins) [redacted]
Mechanism and Regulation of Telomerases

Bakar Fellows Program (Collins) [redacted]
Reverse Transcriptase Technologies

RECENTLY COMPLETED
R01 GM 054198-21 (Collins) [redacted]
Structure and Function of Telomerase

R01 HL 079585-12 (Collins) [redacted]
Biogenesis and Regulation of Human Telomerase

PHS 398 Cover Page Supplement

OMB Number: 0925-0001

Expiration Date: 03/31/2020

1. Vertebrate Animals Section

Are vertebrate animals euthanized? ❍ Yes ❍ No

If "Yes" to euthanasia

Is the method consistent with American Veterinary Medical Association (AVMA) guidelines?

❍ Yes ❍ No

If "No" to AVMA guidelines, describe method and provide scientific justification

2. *Program Income Section

*Is program income anticipated during the periods for which the grant support is requested?

❍ Yes ● No

If you checked "yes" above (indicating that program income is anticipated), then use the format below to reflect the amount and source(s). Otherwise, leave this section blank.

*Budget Period	*Anticipated Amount ($)	*Source(s)

PHS 398 Cover Page Supplement

3. Human Embryonic Stem Cells Section

*Does the proposed project involve human embryonic stem cells? ❍ Yes ● No

If the proposed project involves human embryonic stem cells, list below the registration number of the specific cell line(s) from the following list: http://grants.nih.gov/stem_cells/registry/current.htm. Or, if a specific stem cell line cannot be referenced at this time, check the box indicating that one from the registry will be used:

❑ Specific stem cell line cannot be referenced at this time. One from the registry will be used.

Cell Line(s) (Example: 0004):

4. Inventions and Patents Section (Renewal applications)

*Inventions and Patents: ❍ Yes ❍ No

If the answer is "Yes" then please answer the following:

*Previously Reported: ❍ Yes ❍ No

5. Change of Investigator/Change of Institution Section

❑ Change of Project Director/Principal Investigator

Name of former Project Director/Principal Investigator

Prefix:

*First Name:

Middle Name:

*Last Name:

Suffix:

❑ Change of Grantee Institution

*Name of former institution:

PHS 398 Research Plan

OMB Number: 0925-0001
Expiration Date: 03/31/2020

Introduction	
1. Introduction to Application (for Resubmission and Revision applications)	
Research Plan Section	
2. Specific Aims	
3. Research Strategy*	Essay.pdf
4. Progress Report Publication List	
Other Research Plan Section	
5. Vertebrate Animals	
6. Select Agent Research	
7. Multiple PD/PI Leadership Plan	
8. Consortium/Contractual Arrangements	
9. Letters of Support	
10. Resource Sharing Plan(s)	
11. Authentication of Key Biological and/or Chemical Resources	D_KeyResources.pdf
Appendix	
12. Appendix	

Tracking Number: GRANT12932613
Funding Opportunity Number: RFA-RM-19-005. Received Date: 2019-09-06T13:01:39.000-04:00

Human genetic supplementation without donor DNA or a DNA break (6 MCB; 2 CB)

PROJECT DESCRIPTION

A. Significance and Strategy

Personalized healthcare, including proactive safeguarding of genetic and environmental challenges to human heath, is vital to the future of a diverse society. Affordable genome sequencing and molecular diagnostics have dramatically outpaced the ability of clinicians to use the information for preventive medicine or therapy. Based on model organism precedents, there is an assumption that human cells can be “rescued” using homologous recombination to integrate a complementing transgene. Academic research has benefited enormously from innovations that increase the efficiency of this process by introduction of a recombination-stimulating DNA break. However, there are limits to the success of this approach in mammalian cells, because homologous recombination is suppressed in non-dividing cells and is suppressed even in rapidly dividing cells for most of the cell cycle. Also, a generally applicable method for donor DNA delivery is lacking, and should that emerge, our large genome size inevitably reduces the specificity and efficiency of donor DNA use for break repair. The alternative approach of adeno-associated virus introduction of a transgene also has limitations, among others including the small transgene size permitted by the virus capsid.

It remains an unmet need to have a non-mutagenic, non-toxic, general approach for gene introduction to the human genome. Safe, stable, efficient genetic supplementation of the genome of human somatic cells would fulfill a necessary transformative leap forward for personalized medicine. Therapy for many loss-of-function pathologies hinges on this missing technology. Transgene introduction also offers broad opportunities for non-native control of protein expression, isoform selectivity, and myriad other clinically useful outcomes. I propose to innovate a strategy for supplying additional genetic material to human cells, and subsequently human individuals, that will bypass the need for production of an unrestrained, damage-signaling DNA break and also bypass the challenge and hazard of introducing donor DNA. To do this I will exploit eukaryotic evolution, which has selected for this strategy in the perpetuation of LINE/non-LTR retrotransposons (LRTs).

B. Opportunity

Starkly missing from current efforts to develop transgene introduction techniques for human cells is an approach exploiting the gene-insertion mechanism endogenous to eukaryotes. A retroelement that invaded the genome of a relatively early eukaryote co-evolved and diversified to sustain itself as a mobile passenger in its host organisms ever since. Most LRT lineages are site-specific in their genome occupancy, targeting themselves to a repeated sequence. Occasionally in evolution LRTs devolve to semi-random insertion, one example of which is human LINE-1. LRTs with this loss of specificity appear not to have the evolutionary durability of site-specific LRTs. However, even these more recent LRT lineages offer the potential to be harnessed for human genome supplementation, if constraints are added to direct them to a specific safe-harbor insertion site.

Precision insertion of site-specific LRTs is ongoing in diverse organisms including birds, fish, and insects. Some of these LRT families may be re-tasked to human genome engineering without alteration of their target site specificity. For example, the families of LRTs that have a DNA target site corresponding to conserved sequence in 18S or 28S rRNA will find the same site, literally unchanged or minimally changed, in the human genome. Furthermore, because the site-specific LRTs always target a repetitive sequence (e.g. the rRNA precursor gene; also 5S rRNA, an snRNA, and satellite repeats), disruption of a few target sites by cDNA insertion is not deleterious. Indeed, flies from some *Drosophila* species can have more than 50% of rRNA genes with a retroelement insertion. I envision exploiting the ability of LRT protein to insert cDNA at a specific target site as a new method of human gene therapy. I want to use RNP assembled from modified LRT protein and a template RNA encoding the transgene, with engineered template 5’ and 3’ handles that promote efficient insertion of full-length cDNA.

C. Research Plan

Engineering LRTs to accomplish human gene therapy will involve many high-risk stages of investigation including (a) discovery of active elements among the many copies of many LRTs in sequenced genomes, (b) establishing, enhancing, or redirecting specificity for a safe-haven DNA target site with sequence conservation across human genetic diversity, (c) engineering, producing, and delivering an RNP with modified LRT protein and transgene template at sufficient concentration for efficient gene insertion, and (d) establishing the range of cells and tissues that can use RNP for transgene insertion, including both mitotic and post-mitotic cell types, and the range of transgene lengths and sequences that can be added to the genome. Only some of the complexities and challenges demanding innovation can be described in this proposal.

Unrelenting innovation and the expectation of failure for many approaches will be required to reach the goal of human genome supplementation mediated by RNPs containing a modified LRT protein and transgene-encoding RNA template. The project will benefit from evolving the experimental plan as the research unfolds. Collaborators with a diversity of expertise will be recruited as the project progresses, as will lab members with appropriate interests and training.

Finding active LRTs:

Bioinformatic and computational approaches will be applied to mine LRT sequences from eukaryotic genomes and repeat databases and to interrogate them for evidence of recent mobility. Also, comparison of sequences from independent organisms of the same or closely related species could uncover mobile elements; this approach was successful in demonstrating the mobility of human LINE-1. RNA sequences can also be mined to detect LRT transcripts.

The ancestral, site-specific LRT architecture has a single ORF flanked by 5' and 3' UTRs (**FIGURE 1**). These will be the initial LRTs evaluated for activity. At least three ancestral-type LRT proteins with a 28S rRNA target site have been shown to retain activities necessary for the nick-primed cDNA synthesis mechanism of LRT insertion (**FIGURE 2**), based on biochemical assays of target site binding, nicking, and nick-primed cDNA synthesis (*Bombyx mori*, silkmoth) or [redacted]. Alignments will be used to screen for additional LRT proteins for consideration. Beyond primary sequence, functionally critical amino acids can be evaluated by modeling LRT protein sequence into 3D using crystal structures of evolutionarily related reverse transcriptases (RTs), DNA binding proteins, and nucleases. To evaluate LRT function, [redacted]

FIGURE 1. The ancestral-type LRT retroelement. The 5' UTR contains an RZ and likely an IRES to support ORF translation; in at least some LRT families, the 5' region also contains a high-affinity binding site for the retroelement protein. The 3' UTR binds the retroelement protein and determines the template site of cDNA synthesis initiation. The single ORF encodes a protein with multiple DNA binding domains (ZF, Myb), an RNA binding surface (RBD), reverse transcriptase motifs (RT), a restriction-enzyme-like endonuclease domain (EN) and other conserved modules of unknown function (? and ZK, zinc knuckle). Some LRTs without target-site specificity lack sequence-specific DNA binding domains and C-terminal EN, and instead have a different N-terminal endonuclease domain. The two-ORF LRTs encode an extra protein thought to package the retroelement RNA for nuclear import after translation.

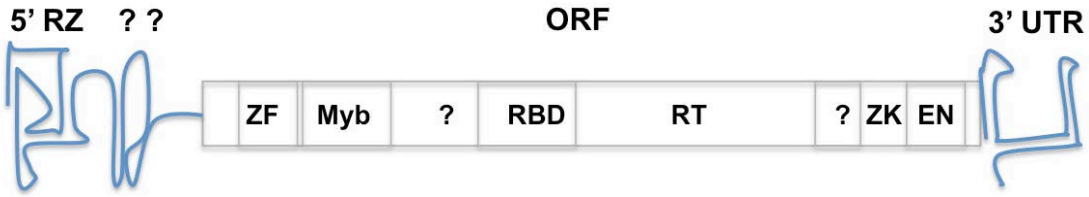
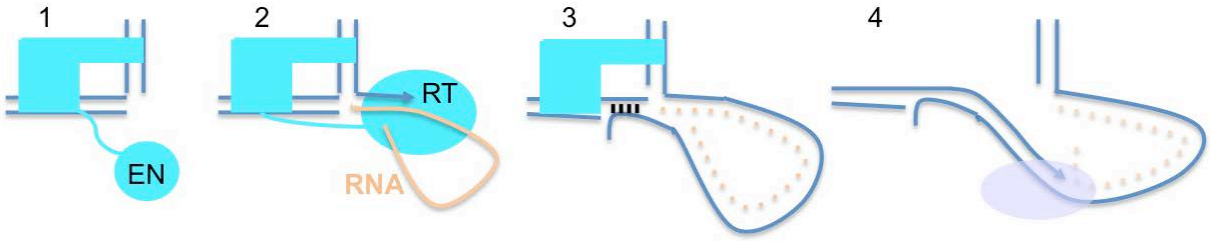


FIGURE 2. LRT insertion. Protein domains (light blue) bind to DNA sequences surrounding the site of cDNA insertion, likely distorting it for access. Target site nicking (1) provides the primer for cDNA synthesis initiation (2), the latter either without or with a few nt of primer-template base-pairing depending on the LRT family. After cDNA synthesis, the other target-site strand is nicked (3), either without an offset from the first nick (as illustrated) or in some LRT families with an offset that results in target site duplication or deletion. Second-strand synthesis (4) may be mediated by the RT and/or a cellular polymerase (purple). If the cDNA 3' end can base-pair with the target site as illustrated in (3), the LRT junction with the target site is precise.



Transgene template design:

It is essential to this project to design a framework for RNAs that will be specifically and efficiently used for cDNA synthesis from the initiating target site nick. Only a limited amount is known about key features of an RNA template for any LRT, and not even guesses have been made to anticipate what protects a template 5' end to allow synthesis of full-length cDNA. Ancestral-type LRT families have a variable-length 5' UTR and a relatively short 3' UTR (~150-250 nt). LRT RNA is typically transcribed as part of a host transcript, and that transcript is processed into the template for reverse transcription by a self-cleaving ribozyme (RZ) in the 5' UTR. Depending on the LRT family, the RNA 5' end is either precisely at the target site junction or upstream in the target site sequence. Some LRTs appear to use a mechanism other than an RZ to specify the template 5' end, based on observed failures of RNA to self-cleave. Also some more recently evolved LRTs have gained an internal promoter. The 3' processing of template is uncertain, in particular for the ancestral-type site-specific LRTs embedded within a host RNA, but the 3' UTR sequences copied into cDNA are well defined; typically the junction of target site and cDNA 5' end is precise.

The 5' and 3' ends of a RNA template for transgene insertion will be engineered based on understanding the folding, protein interaction(s), and functions of endogenous LRT RNA template UTRs. The 3' UTR will be streamlined to the minimum length and optimal folding that directs cDNA synthesis initiation first using a biochemical assay with purified components. I envision that many therapeutically valuable transgenes will include cassettes for expression of a human protein comprised of a promoter, an ORF, and a poly-adenylation signal; this vision will be built out by recruiting appropriate expertise for consultation about what sequences will or won't be suitable for therapeutic applications. To accelerate transgene testing, cassettes will be integrated at LRT target sites using CRISPR/Cas9 technology and evaluated for protein expression level and for persistence of both the transgene and its expression. [redacted]

I am keen to innovate biochemical and cellular methods to select and iteratively refine RNA handles that support LRT protein binding, cDNA synthesis initiation, and other critical RNA roles. Some of these handles could be entirely exogenous in design, if ideas for that emerge from initial insights about RNA template requirements. Of note, I hope to provide cells with the actual RNA template for transgene synthesis, not DNA encoding the template that would have to be transcribed; therefore a range of RNA modifications can be sampled, exploiting knowledge from dramatically successful uses and ongoing trials of RNA therapeutics for human disease. I believe that my group and collaborators can nucleate an enabling expansion of RNA-based therapy by developing new strategies for production of several thousand nt RNAs for clinical use.

Protein (re)design:
It will be broadly useful for genome engineering to gain insight about how the ancestral-type LRT proteins accomplish such high specificity for one particular target sequence in a eukaryotic genome. How did different lineages of LRT change this site specificity without changing protein domain architecture? This knowledge will become critical for my goal if endogenous LRT target sites aren't suitable loci for sustained transgene expression, and if a strategy of fusing additional DNA binding determinants to an LRT protein doesn't become a logical approach.

The LRT protein possesses all of the activities necessary for transgene introduction (**FIGURE 2**): target site nicking by EN domain generates a 3'OH used by the RT domain to initiate cDNA synthesis from the RNA 3' UTR, then after cDNA synthesis, nicking of the other target site strand occurs to prime second-strand DNA synthesis. It is unknown how the DNA binding and EN domains of an LRT protein coordinate to determine the site of the nicks. Likewise, it is unknown how EN and RT domains coordinate. The structural principles of this coordination could involve the regions of the LRT protein that do not yet have assigned function. I will initiate several approaches for understanding and redesigning LRT protein structure/function; it will have some similarity to the other type of eukaryotic RNP RT, telomerase, which was the ancestor or spawn of an LRT protein and has been a main subject of my research program for the past 27 years.

Validate RNP delivery as a strategy for therapeutic applications:
At no stage of the transgene insertion process should there be a DNA break (**FIGURE 2**). The retroelement wouldn't persist if it welcomed competition from DNA break repair. Current models suggest that LRT protein remains bound to DNA flanking the nicks, which would inhibit end-resection. Nonetheless, I will evaluate whether RNPs assembled from LRT protein and template RNA activate DNA damage signaling as the first step in monitoring potential toxicity. Transgene insertion efficiency, specificity, and lack of toxicity will be evaluated in parallel as the RNP engineering and delivery approaches evolve in their designs.

EVIDENCE FOR INNOVATIVENESS
I'm passionate in the pursuit of finding insights that push a field forward, bridge fields, or generate new fields, examples of which are described below. I'm driven by the thrill of integrating diverse sources of information and experimental approaches to reveal new biology and improve clinical care. I readily recruit collaborators based on my track record of valuing their contributions.

As a single answer for three seemingly different puzzles about human telomerase we were struggling with in the mid/late 1990s, I hit upon the idea that human telomerase RNA might contain the same motif then just discovered in yeast small nucleolar RNAs responsible for guiding rRNA pseudouridine modifications. My graduate student and I demonstrated that this motif could account for the mysterious 3' end processing specificity of human telomerase RNA and the nucleolar fractionation of some active telomerase. In response to manuscript rejections stating that telomerase RNA couldn't have a this motif because it didn't assemble with the protein that catalyzes pseudouridine modification, we cloned the human homolog of that enzyme and showed it to be an integral telomerase subunit. Moreover, we showed that human disease mutations affecting that subunit compromise telomerase function but not rRNA modification. For many years, our compounding evidence that telomerase deficiency rather than ribosome deficiency was the explanation for disease was ignored in favor of well-published but poorly performed studies supporting the opposite. Twenty years later, after many additional contributions, our telomerase deficiency explanation is accepted, the range of telomerase subunit mutations and "telomeropathy" diseases has grown enormously, and telomere length is used as a diagnostic and a guidance for therapy, because transplants in patients with short telomeres need to use a different protocol.

Another graduate student in my lab discovered stress-induced tRNA anticodon loop cleavage in *Tetrahymena* (Lee and Collins, 2005), and subsequently other types of tRNA fragments; the

generation of these fragments was then confirmed in budding yeast by Roy Parker's lab, later extended to human cells, and now is touted as biomarker frontier and potential basis for RNA therapeutics. Yet other graduate students discovered new types of piRNA that were unprecedented at the time but have since been found in other organisms (for example Lee and Collins, 2006; Couvillion et al., 2009; Couvillion et al., 2012). One my lab's graduate students introduced the PP7 RNA hairpin tag and modified PP7 coat protein for endogenously assembled RNP purification (2x Hogg and Collins, 2007), which was then exploited by Rob Singer's lab for RNA imaging and is now widely used; along with human telomerase expression plasmids, the PP7 coat protein plasmids we generated are top burners among AddGene orders. As one more example of innovation, my group and our collaborators have contributed new bioinformatics approaches and new enzymes for RNA-seq of non-coding RNAs with base modifications that preclude their analysis by standard protocols (Couvillion et al., 2012; Katibah et al., 2014)

HOW THE PLANNED PROJECT DIFFERS FROM MY PREVIOUS RESEARCH

The Collins lab has contributed to and launched several fields of research, but none of them involve new technology for transgene addition to human cells. Our longest-running research direction has sought to understand the biochemical underpinnings of how telomerase accomplishes processive synthesis of telomeric repeats and how telomerase activity is restricted to and regulated at telomeres. To get answers to these questions, the lab's path detoured through establishing principles of co-folded RNP biogenesis, defining new roles for non-coding RNA, developing new approaches for RNA-based purification of biologically assembled RNPs, participating in and leading *Tetrahymena thermophila* genome sequencing projects, setting new standards for human cell genome engineering, and other unanticipated but relevant quests.

Spurred on by these forays, we studied myriad non-coding RNAs and their RNPs as well as endogenous RNA silencing pathway machinery including a new type of RNA-dependent RNA polymerase activity coupled to a Dicer enzyme. These studies provide a solid foundation of RNA, polymerase, and genome engineering expertise but do not overlap the proposed research. Also we have contributed and still are developing new RNA-seq approaches using diverse types of RT including LRT proteins, but only the polymerase core of these and only for the purpose of unbiased copying of pools of RNA into cDNA libraries for sequencing.

SUITABILITY FOR THE NIH COMMON FUND PIONEER AWARD PROGRAM

This project is not suitable for funding as an R01, or any NIH mechanism other than the Pioneer Award, due to very high risk, lack of preliminary data, inability to forecast a several-year experimental plan, and need for recruitment of enabling collaborators as the project progresses rather than at the start. Also it is a new direction from all of my lab's previous work.

The constant innovation that will be required to make this project fulfill its potential is hallmark of a Pioneer Award, as is the new combination of research fields that the project will integrate. I suspect that my letters of support will attest to my ability to bridge fields and interlace experimental approaches, and also my persistence in bringing a field from skepticism about my ideas to complete conversion. I believe that my obsession, energy, ability to motivate, and proven skill at uncovering clues to new mysteries and to fill knowledge gaps and pass roadblocks make me an excellent candidate for the Pioneer Award.

STATEMENT OF RESEACH EFFORT COMMITMENT

If chosen to receive the NIH Common Fund Pioneer Award, I will commit more than 6 person-months (>51%) of my research effort to the projects supported by the Pioneer Award. Because the Pioneer Award will fund the most important research that I can envision to contribute, I will resign my current NIGMS R35 on telomerase. My group and I are eager to master a new RNP RT with strong promise to be a human therapeutic.

PHS Human Subjects and Clinical Trials Information

OMB Number: 0925-0001 and 0925-0002

Expiration Date: 03/31/2020

Are Human Subjects Involved	❍ Yes ● No
Is the Project Exempt from Federal regulations?	❍ Yes ❍ No
Exemption Number	❑ 1 ❑ 2 ❑ 3 ❑ 4 ❑ 5 ❑ 6 ❑ 7 ❑ 8
Does the proposed research involve human specimens and/or data	❍ Yes ● No
Other Requested information	

Authentication of Key Resources

We authenticate human cell lines by short tandem repeat (STR) profiling service from the MCB Department Tissue Culture Facility housed in the same building. Parental and derived cell lines are frozen in multiple aliquots, and new frozen aliquots are thawed every few months to replace continuous cultures to reduce heterogeneity. Plasmids are sequenced by the MCB Department DNA Sanger sequencing facility, also housed in the same building. When reasonable, antibodies are tested for specificity by generating knockout cell lines and/or cell lines with epitope-tagged proteins expressed in replacement of untagged protein.