Age-associated cryptic transcription in mammalian stem cells is linked to permissive chromatin at cryptic promoters

Brenna McCauley
Baylor College of Medicine
Luyang Sun
Baylor College of Medicine
Ruofan Yu
Baylor College of Medicine https://orcid.org/0000-0002-5281-2508
Dena Leeman
Stanford University
Yun Huang
Institute of Bioscience and Technology, Texas A&M University
Ashley Webb
Brown University
Weiwei Dang (weiwei.dang@bcm.edu)
Baylor College of Medicine https://orcid.org/0000-0002-6931-4636

Article

Keywords: age-associated cryptic transcription, mammalian stem cells, permissive chromatin, cryptic promoters

Posted Date: October 5th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-82156/v1

Version of Record: A version of this preprint was published at Nature Aging on August 2nd, 2021. See the published version at https://doi.org/10.1038/s43587-021-00091-x.

Age-associated cryptic transcription in mammalian stem cells is linked to permissive chromatin at cryptic promoters

Authors

Brenna S. McCauley[1#], Luyang Sun[1#], Ruofan Yu[1],Dena Leeman[2, 3], Yun Huang[4], Ashley E. Webb[5], Weiwei Dang[1*]

[#]Equal contribution

[*]Correspondence should be addressed to Weiwei.Dang@bcm.edu

Affiliations

[1]Huffington Center on Aging, Baylor College of Medicine, Houston, TX 77030, USA; [2]Department of Genetics, Stanford University, Stanford, CA, 94305 USA; [3] Department of Discovery Immunology, Genentech, Inc. 1 DNA Way, South San Francisco, CA, 94080, USA [4]Center for Epigenetics & Disease Prevention, Institute of Bioscience and Technology, Texas A&M University, Houston, TX 77030, USA; [5]Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, RI, USA

Abstract

Suppressing spurious cryptic transcription by a repressive intragenic chromatin state featuring trimethylated lysine 36 on histone H3 (H3K36me3) and DNA methylation is critical for maintaining self-renewal capacity in mouse embryonic stem cells. In yeast and nematodes, such cryptic transcription is elevated with age, and reducing the levels of age-associated cryptic transcription extends yeast lifespan. Whether cryptic transcription is also increased during mammalian aging is unknown. We show for the first time an age-associated elevation in cryptic transcription in several stem cell populations, including murine hematopoietic stem cells (mHSCs) and neural stem cells (NSCs) and human mesenchymal stem cells (hMSCs). Using DECAP-seq, we mapped and quantified age-associated cryptic transcription in hMSCs aged *in vitro*. Regions with significant age-associated cryptic transcription have a unique chromatin signature: decreased H3K36me3 and increased H3K4me1, H3K4me3, and H3K27ac with age. Furthermore, genomic regions undergoing such age-dependent chromatin changes resemble known promoter sequences and are bound by the promoter-associated protein TBP even in young cells. Hence, the more permissive chromatin state at intragenic cryptic promoters likely underlies the increase of cryptic transcription in aged mammalian stem cells.

Introduction

Aging, the degeneration of an organism over time, is a universal phenomenon. Although the specific physical and phenotypic manifestations of aging vary between different organisms, the underlying molecular and cellular mechanisms that drive this process are broadly conserved among eukaryotes[1]. One such hallmark, epigenetic alteration, ultimately drives the opening of chromatin with age, which limits lifespan [2]. A more open chromatin structure can also have a profound effect on the regulation of transcription, and, indeed, there are numerous examples of age-associated transcriptional dysregulation that can impair cellular function and, in mammals, even induce a partial loss of cell identity[3–12]. Recent work in yeast and worms has also revealed that a more open chromatin state, characterized by the loss of trimethylated lysine 36 of histone H3 (H3K36me3) within gene bodies during aging, causes an increase in cryptic transcription[12]. As reduction of H3K36me3 levels is a common age-associated phenotype in eukaryotes[13–15], increased cryptic transcription may play an underappreciated role in the biology of aging.

Cryptic transcription is the aberrant initiation of transcription from non-promoter regions, and often occurs at sites within the gene body. This phenomenon has been extensively studied in yeast (reviewed in ref. [16]) and more recently in mammals[17–19]. Importantly, cryptic transcription is thought to result from the chromatin state changes that arise when RNA polymerase II transits the gene body. During this process, chromatin is opened by the removal of histones and altered by deposition of specific histone modifications that are conferred co-transcriptionally (reviewed in ref. [20]). Following transcription, a closed chromatin state is restored by recruitment of the FACT complex, which promotes the re-incorporation of nucleosomes in yeast and mammals[21,22]; the histone deacetylase complex Rpd3S in yeast[23]; and the histone demethylase Kdm5b[19] and the DNA methyltransferase Dnmt3b[18] in mice. Loss of any of these proteins, which are recruited to

actively transcribed gene bodies by H3K36me3, results in increased cryptic transcription (reviewed in ref. [24]).

Evidence from both yeast and worms suggests that increased cryptic transcription has a negative impact on lifespan. Perturbations that decrease H3K36me3 increase cryptic transcription and decrease lifespan in yeast, while those that increase H3K36me3 have the opposite effect on both lifespan and cryptic transcription[12]. Similarly, in worms, knocking down *met-1*, which is thought to trimethylate H3K36, shortens lifespan, while RNAi against *jmjd-2*, an H3K36 demethylase, extends it[13,25], though the impact of these treatments on cryptic transcription has not been assessed. Nevertheless, an increase in cryptic transcription, likely driven by chromatin structure dysregulation as the result of reduced levels of H3K36me3, is a conserved feature of aging in yeast and worms.

Given the broad conservation of molecular mechanisms that drive aging, asked whether cryptic transcription increases with age in mammals and if chromatin dysregulation contributes to this process. In particular, we focused on stem cells, which may play an outsized role in the aging process (reviewed in ref. [26]). Analysis of transcriptome data, mapping of transcription start sites, and assessment of several transcription-associated histone modifications suggest that cryptic transcription increases with age in mammals, and that this is associated with an opening of the chromatin, just as seen in yeast and worms.

Results

Mammalian stem cell aging models.

In this study, we focused on aging in two models: murine hematopoietic stem cells (mHSCs) and human mesenchymal stem cells (hMSCs). Transcriptional and chromatin-level changes were previously characterized in highly purified HSCs isolated from young (4mo) and aged (24mo) mice[7]; we reanalyzed these data to look for evidence of intragenic cryptic transcription. We further used culture expanded hMSCs isolated from cord blood as an *in vitro* model of aging in a human stem cell population. Culture expansion is a commonly used model of MSC aging as it recapitulates many of the phenotypes observed in MSCs isolated from aged individuals[27]. We confirmed that culture expanded hMSCs have a decrease in differentiation potential; an increase in the proportion of senescent cells, as measured by β-galactosidase staining; and decreased growth rate (Figure S1). Interestingly, the reduction in differentiation potential occurs before decreased growth rate and increased β-galactosidase activity; for this study, we chose to expand the cells to the point that the mock-aged cells had decreased differentiation potential, but had not yet slowed growth or increased cellular senescence.

Analysis of RNA-seq data indicates an increase of intragenic cryptic transcription with age.

Intragenic cryptic transcription is the initiation of transcription from a non-promoter region within a gene body[22,28]. This phenomenon should be reflected in RNA-seq data as an increase in mapped reads downstream of the cryptic initiation site relative to upstream thereof. As such sites can occur throughout a gene, we developed a method to detect cryptic transcription in RNA-seq samples by examining exon-based expression levels within transcripts. Simply, if

downstream exons have a higher transcripts per kilobase million (TPM) than the first (or second) exon of a transcript, as reflected by the ratio of exon-based TPMs, this indicates that cryptic transcription is occurring in the cognate transcript. When comparing between samples, an elevation in cryptic transcription can be inferred from an increased average exon-based TPM ratio in one sample vs. the other. We tested our method using RNA-seq data from murine embryonic stem cells (mESCs) that lack Setd2, which are known to have elevated cryptic transcription[18]. We detected a significant increase in the average exon-based TPM ratio in the Setd2 knockout sample vs. the control mESCs (Figure S1), indicating that our method can detect increased cryptic transcription from RNA-seq data. Analysis of mESCs undergoing Setd2 knockdown[29] or knockout[30] and Setd2 knockout murine oocytes[31] confirmed an elevation of cryptic transcription in these systems (Figure S1).

We applied this method to RNA-seq datasets from aged mHSCs[7] and culture-expanded hMSCs. The average exon-based TPM ratio is increased in mHSCs isolated from old mice relative to young, indicating that cryptic transcription increases with age (Figure 1A). These ratios are shown separately for young and old samples in Figure S1. A similar phenomenon is observed in hMSCs (Figure 1B and Figure S1). There is also a trend of transcripts with higher expression having greater increases in exon-based TPM ratios with age in both mHSCs and hMSCs (Figure S1), suggesting that highly expressed genes are more susceptible to age-associated increases in cryptic transcription. Additionally, the exon-based TPM ratio increases deeper into the transcript, which is consistent with sustained transcription from cryptic sites and expected in a cumulative measurement from RNA-seq. This is confirmed by the number of exons that have exon-based TPM ratios greater than 2 increasing as exon number increases

(Figure S1). Thus, the RNA-seq data from both mHSCs and hMSCs indicates that, on a global scale, cryptic transcription is elevated with age in mammalian stem cells.

We next identified transcripts that have an age-associated increase in cryptic transcription, *i.e.*, transcripts for which the exon-based TPM ratio is highest in the old vs. young samples. Because the data are not normally distributed (Figure S1), we ranked the log2-transformed data and determined the right-tailed inflection point to identify transcripts with a significantly high increase in exon-based TPM ratio with age[32,33] (Figure 1C). All transcripts to the right of this cut-off (199 in mHSCs and 304 in hMSCs) were considered to exhibit an age-associated elevation of cryptic transcription. Heatmaps of the exon-based TPM ratios confirm an increase in the old samples (Figure 1D), and a scatterplot of the old ratio vs. the young ratio on a per-transcript basis shows that the identified transcripts all lie above the line y=x (Figure S1). As expected, metagene plots of the RNA-seq read distribution in these samples show that read density increases towards the 3' end of the transcripts vs. the 5' end (Figure 1E). Two examples of cryptic transcripts identified by this analysis, *Car11* in mHSCs and *IAH1* in hMSCs, are shown in Figure 1F. The transcripts identified by our analysis are longer than average (Figure S1), suggesting that, along with higher expression levels, gene length is associated with elevated cryptic transcription with age.

As cryptic transcription increases with age in both mHSCs and hMSCs, we asked whether this also occurs in other adult stem cells. Neural stem cells (NSCs) are maintained in a resting state, but are activated under certain conditions[34]. We examined RNA-seq data from quiescent and activated NSCs (qNSCs and aNSCs, respectively) freshly isolated from the subventricular zone of young and old mice for evidence of an age-associated increase in cryptic transcription. Principle component analysis (PCA) showed that most gene expression changes

correlate with activation state and a smaller fraction are associated with sex, while age has less of an effect (Figure 2A). However, in both males and females, aNSCs show an increased exon-based TPM ratio with age, though qNSCs do not (Figure 2B). In males, this increase is only evident in the last exon, while in females, there is a gradual increase throughout the transcript. When divided into expression quartiles, in aNSCs isolated from female mice, only transcripts in the fourth quartile show elevated cryptic transcription with age, suggesting a strong link between expression level and age-increased cryptic transcription. However, no such trend was observed in aNSCs from male mice (Figure S2). We identified 237 transcripts in males and 266 in females that exhibit increased cryptic transcription with age, confirmed by heatmaps of the young and old ratios vs. the first exon (Figure 2C) and metagene plots of RNA-seq signal (Figure 2D); as in mHSCs and hMSCs, these transcripts tend to be longer (Figure S2). Thus, increased cryptic transcription is also a hallmark of NSC aging, but only in activated NSCs.

We additionally looked for signs of an age-associated increase in cryptic transcription in publicly available aging and senescence RNA-seq datasets (E-GEOD-59966; E-GEOD-46486; GSE53330; E-MTAB-4879; and refs. [35–44]). In 21 out of 25 datasets spanning a range of tissues, we detected an elevation of cryptic transcription with age (Figure S2), similar to what was observed in the stem cells. Overall, this suggests that increased cryptic transcription is a hallmark of aging in many mammalian tissues. When we applied this method to samples from Rett and Werner syndromes[45,46], both premature aging models, no increase in cryptic transcription was detected (Figure S2), indicating that these models do not recapitulate this feature of aging. Nevertheless, taken together, our data indicate that elevated cryptic transcription is a hallmark of mammalian aging.

Global analysis of the 5' ends of transcripts identifies sites of cryptic transcription initiation.

Although examination of RNA-seq data revealed that cryptic transcription increases with age, this analysis cannot identify the sites from which these cryptic transcripts initiate (cryptic transcription start sites, cTSSes). For this, we used a recently-developed protocol called DECAP-seq that allows for the specific sequencing of the 5' ends of transcripts[18]. We were only able to obtain sufficient RNA to perform this analysis on hMSCs, which can be expanded to large numbers during culture. A modified peak-calling algorithm was used to identify DECAP-seq peaks in young and old hMSC samples. Most peaks cluster around annotated RefSeq TSSes (referred to as "endogenous TSSes"), as expected (Figure 3A). Although the DECAP-seq signal is reduced downstream of endogenous TSSes, it persists throughout the gene body in both young and old samples, suggesting that intragenic cryptic transcription occurs in both young and old hMSCs. DECAP-seq peaks were considered to indicate cryptic transcription if they were greater than 2000bp distant from an endogenous TSS, a stringent cut-off that ensures no signal from annotated promoters is mis-called as cryptic and prevents chromatin signatures associated with such promoters from confounding assessment of chromatin state at cryptic promoters; subsequent analysis considers these peaks only. A comparison of these peaks revealed that 58% of the total peaks were found in both samples, while nearly 12% were unique to the young sample and almost 30% were unique to the old sample (Figure 3B), which suggests that cryptic transcription increases with age.

We further developed a differential peak calling pipeline to identify non-endogenous TSS-associated DECAP-seq peaks with more reads in the old sample vs. the young. While only 127 peaks have significantly higher signal in the young sample, there are 1375 peaks with higher

signal in the old sample (Figure 3C); these peaks are found in both young and old samples. That the old sample has more higher peaks than the young sample suggests that cryptic transcription increases with age in hMSCs, consistent with our other analyses. A metagene plot clearly shows that these sites have higher DECAP-seq signal in the old vs. young samples (Figure 3D), while those identified as having more signal in the young show the converse (Figure S3). DECAP-seq signal along the whole gene is shown for both *GAPDH* and *CAPNS1*, two genes with 3 identified cTSSes between them. An enrichment of DECAP-seq signal is seen around the endogenous TSS of each gene, and several peaks can be found within the gene bodies; the intragenic peaks are clearly higher in the old samples vs. the young (Figure 3E). We consider the 1375 loci identified by this analysis to be *bona fide* sites of age-associated cryptic transcription.

If the age-increased DECAP-seq peaks are sites where cryptic transcription increases with age, this should be reflected in the RNA-seq data. As expected, there is an increase in the ratio of RNA-seq reads mapping immediately up- and downstream from these sites (Figure 3F), indicating that the increased transcription identified by DECAP-seq can be independently detected. When considering the 158 cTSSes located within introns (Figure 3G), this trend is more obvious. The transcripts determined to have increased cryptic transcription with age by DECAP-seq are longer (Figure 3H) and tend to be more highly expressed (Figure 3I). There is a slight trend of higher gene expression with age in genes with an age-associated increase in cryptic transcription, though those genes with decreased cryptic transcription with age have no clear trend in expression changes (Figure S3). We also searched for enriched transcription factor binding sites (TFBSs) in the vicinity of the age-associated cTSSes. HOMER[47] analysis identified enrichment of five motifs within 200bp of these DECAP-seq peaks (Figure 3J and Figure S3).

Changes in the abundance or binding of transcription factors that recognize these sites with age could contribute to increased Pol II recruitment to the cTSSes and therefore cryptic transcription.

Transcription-related chromatin states are largely maintained during aging.

Transcription is fundamentally regulated at the level of chromatin. It is well known that chromatin becomes more open with age in many systems (reviewed in ref. [48]), which could contribute to increased cryptic transcription by promoting intragenic Pol II entry. We therefore performed ChIP-seq to characterize how several transcription-associated histone modifications (H3K4me1, H3K4me3, H3K27ac, and H3K36me3) and heterochromatin-associated histone marks (H3K9me3 and H3K27me3) change with age in hMSCs. Additionally, we performed a ChromHMM analysis[49] to better understand how the chromatin state changes during aging. A 10 state model in ChromHMM identified four categories of chromatin states, based on known functions of particular histone modifications[50]: heterochromatin (states 1-3), enhancers (states 4-6), actively transcribed genes (states 7-9), and chromatin with no enriched histone modifications (state 10) (Figure 4A). These designations are consistent with the genome-level distribution of states and their enrichment around TSSes and transcription end sites (Figure S4).

Globally, heterochromatin decreases with age in hMSCs (Figure 4B), consistent with numerous other reports (reviewed in ref. [51]). However, there are only small changes in the fraction of the genome associated with enhancer- and transcription-related chromatin states (Figure 4B and Figure S4). Most of these changes occur between similar states, *e.g.*, state 7 (enriched for H3K36me3, H3K4me1, and H3K27ac) changing to state 8 (H3K36me3 only) by losing enrichment of several histone modifications (Figure 4C and Figure S4). Heterochromatin loss can be attributed largely to the conversion of states 1-3 to state 10, unmodified chromatin,

rather than to a gain of canonically euchromatic histone modifications in previously heterochromatic regions. Overall, the most drastic changes in chromatin state during aging are a loss of constitutive heterochromatin, a gain of facultative heterochromatin, and a decrease in the overlap between H3K4me1 and H3K27ac with H3K36me3.

We also examined how chromatin state changes with age in particular genomic regions (Figure 4D and Figure S4). State 1 is most enriched at intergenic and lamin-associated domains (LADs), but shows a drastic reduction in the enrichment with age, though in some of these regions, there is an increase in state 3 enrichment (Figure S4). However, the particular LADs that lose state 1 (H3K9me3) are not the loci that gain state 3 (H3K27me) (Figure 4E and F). Within CpG islands and genic regions, there is a strong trend of increasing enrichment of state 3, and a weaker trend of state 7 enrichment. In contrast, only small changes are seen in the enrichment of states 6 and 8, while states 5 and 9 are generally depleted with age. The increase in state 3 and decrease in state 9, particularly in promoter regions, suggests an overall decrease in transcription with age, consistent with the observation that TPMs of expressed genes tend to decrease with age (not shown). Overall, this analysis suggests that normally heterochromatic regions of the genome become more open with age, while genic regions become both more open and more closed at different loci.

There are subtle age-associated changes in histone modification enrichment in expressed genes. As seen in Figure 4G, H3K4me3 distribution does not change with age, but its enrichment around the promoter increases. While enrichment levels of H3K4me1 and H3K27ac do not appreciably change, their distribution becomes wider with age. This is reflected in Figure 4C as the fraction of state 8 that changes to state 7 with age and Figure 4D as the increase of state 7 in genic regions. Unlike the increased prevalence of promoter-associated histone modifications at

actively expressed genes during aging, H3K36me3 is depleted within gene bodies (Figure 4G). This is not a complete loss of H3K36me3, but a reduction in its enrichment. As H3K36me3 promotes the restoration of chromatin structure following Pol II transit[23], it is possible that its decreased enrichment has effects on the chromatin that are undetectable by ChromHMM analysis. Thus, although chromatin structure at actively expressed genes is largely preserved with aging, the chromatin at these loci may be more open in old hMSCs.

The gain of an active promoter-like chromatin structure correlates with age-associated cryptic transcription.

Reduced H3K36me3 is associated with increased cryptic transcription and H3K4me3 (ref. [19]). A chromatin environment enriched for H3K4me3 and depleted of H3K36me3 is reminiscent of the state at active endogenous promoters. We hypothesized that the chromatin structure around age-associated cTSSes might take on an active promoter-like state with age. Indeed, H3K4me3 is specifically enriched near age-increased cTSSes in old, but not young, hMSCs (Figure 5A), while no such increase in H3K4me3 is observed at cTSSes with decreased expression with age (Figure S5). The pattern of H3K4me3 enrichment around the age-associated cTSSes is even comparable to what is seen at endogenous TSSes (Figure S5). A similar trend is also seen for H3K4me1 and H3K27ac (Figure S5). While H3K36me3 persists in the vicinity of age-increased cTSSes, it is decreased in the old samples compared to the young (Figure 5A and Figure S5). Although H3K36me3 levels are reduced with age at age-associated cTSSes, the levels are still much higher than observed at endogenous TSSes, which likely contributes to the relatively low levels of transcription observed at these sites. Thus, in hMSCs the chromatin state

around age-increased cTSSes becomes more active promoter-like with age, with a reduction in H3K36me3 and an accumulation of H3K4me3.

This observation prompted us to determine whether age-associated cTSSes have other characteristics of promoters. We therefore performed ChIP-seq for TBP and compared its enrichment at endogenous TSSes and age-increased cTSSes. As expected, over 85% of the TBP peaks were located within 2000bp of an annotated promoter (Figure S5). Surprisingly, TBP is bound near a subset of age-associated cTSSes to a similar extent in both old and young hMSCs (Figure 5B). Independent clustering of the TBP ChIP-seq signal further showed that the distribution of TBP enrichment is similar at endogenous and cryptic TSSes, which suggests that at least part of the transcriptional machinery recognizes select age-associated cTSSes in a similar manner to which it recognizes endogenous TSSes. The pre-initiation complex may be actively recruited to age-increased cTSSes and, as the local chromatin state becomes more open and permissive with age, transcription initiation may increase at these loci.

If age-associated cTSSes gain an active promoter-like chromatin structure in aged hMSCs, it is possible that this chromatin signature is associated with additional age-increased cTSSes that were not identified by DECAP-seq. We therefore searched for sites within gene bodies where H3K4me3 is gained and H3K36me3 enrichment decreases with age (Figure 5C). The focus was on H3K4me3 rather than H3K4me1 or H3K27ac as its distribution is known in mouse HSC aging[7], allowing us to extend our model to this system. Using this method, we identified an additional 2118 putative age-associated cTSSes in hMSCs, exemplified by the site found in *ATOH8* (Figure 5D). The putative age-increased cTSS is enriched for H3K4me3 in both young and old cells, but there is a local depletion of H3K36me3 with age, as seen by a broadening of the valley in the signal track for that histone modification. The increase of

DECAP-seq reads in this region in the old vs. young sample suggests an age-associated increase in cryptic transcription, though it does not rise to the level of significance required by our differential peak calling algorithm.

To confirm that the sites identified in this analysis are associated with elevated cryptic transcription, we analyzed RNA-seq and DECAP-seq data surrounding these loci. There is a significant increase in DECAP-seq signal in old hMSCs vs. young associated with these sites (Figure S5). Likewise, the ratio of RNA-seq signal in the exon downstream of the predicted cTSS vs. the first exon of the transcript is elevated in old vs. young hMSCs (Figure 5E). Additionally, the putative age-increased cTSSes have higher scores in a promoter 2.0 analysis[52] than random sequence, indicating that they have promoter-like characteristics (Figure 5F). We performed a similar analysis using ChIP-seq and RNA-seq data from aging mHSCs[7] and identified 510 putative age-associated cTSSes. As in hMSCs, intragenic sites that lost H3K36me3 and gained H3K4me3 with age have an age-associated elevation of RNA-seq reads downstream of the predicted cTSS (Figure 5G). Likewise, these loci also have high promoter prediction scores (Figure 5H). Taken together, these analyses suggest that the sites identified by our chromatin state analysis are *bona fide* age-increased cTSSes.

Discussion

While it is well established that the chromatin landscape changes with age, the impact of these changes on cells and organisms have not been exhaustively characterized. Here we show for the first time that intragenic cryptic transcription is elevated with age in mammals and link this increase to an altered chromatin state that is more permissive for transcription activation. Fundamentally, cryptic promoters are sites from which RNA polymerase II can aberrantly

initiate transcription. These sites have sequence features similar to those in endogenous TSSes (Figure S3; Figure 5); such promoter-like sequences may be the result of random chance during evolution. Nevertheless, these cTSSes can be bound by TBP in both young and old cells (Figure 5), which suggests that they are functioning promoters that are routinely silenced, likely due to a repressive chromatin state maintained downstream of H3K36me3[17–19] (Figure 6). Indeed, this histone modification is reduced within gene bodies with age, and its reduction is especially pronounced around age-increased cTSSes (Figures 4 and 5). A concomitant increase in H3K4me1, H3K4me3, and H3K27ac in these regions with age results in a chromatin state similar to what is seen at active endogenous promoters (Figure 5). Thus, as the intragenic chromatin state becomes more permissive for transcription initiation during aging, overall levels of cryptic transcription increase.

This increase in cryptic transcription occurs in a wide range of tissues during aging: in addition to hMSCs, mHSCs, and NSCs, evidence of this phenomenon is found in skin, bone, liver, and various regions of the brain (Figures 1 and 2; Figures S1 and S2). Interestingly, in mouse aNSCs, the age-associated increase in cryptic transcription has distinct characteristics in males and females (Figure 2). Indeed, in the human dermal fibroblasts dataset analyzed in Figure S2 (ref. [39]), when samples are segregated by sex, only males show an elevation of cryptic transcription with age (not shown), further suggesting sex-specific differences in this aspect of aging. As cryptic transcription increases in a broad spectrum of tissues with age, it may contribute to aging pathologies throughout the body. While the direct phenotypic consequences of elevated cryptic transcription are uncharacterized in adult tissues, loss of Kdm5b, which increases cryptic transcription, impairs self-renewal of embryonic stem cells[19]. This suggests that elevated cryptic transcription may also be detrimental in adult stem cells and aging tissues in

mammals, as has been found in yeast[12]. As cryptic transcription increases with age throughout the body, and is epigenetically encoded, interventions that target this phenomenon may be approachable pro-longevity treatments.

Acknowledgements

We would like to thank Dr. Rui Chen and the Human Genome Sequence Center at Baylor College of Medicine for performing the Illumina sequencing reported here. This work was funded by NIH grants R01AG052507 to WD and R01AG053268 to AW; R01HL134780 and R01HL146852 to YH; CPRIT award R1306, to WD; and a Ted Nash Long Life Foundation research grant to WD. BSM was supported by T32AG000183.

Author Contributions

Conceptualization, WD, BSM, and LS; Methodology, WD, BSM, LS, and YH; Investigation, BSM, LS, RY, DL, YH, AW, and WD; Writing: Original Draft, BSM, LY and WD; Writing: Review & Editing, BSM, LS, RY, DL, YH, AW, MK and WD; Funding Acquisition, WD, YH, and AW; Supervision, WD.

Competing Interests

The authors declare no competing interests.

Methods

hMSC cell culture and growth curve

Human cord blood mesenchymal stem cells (hMSCs) were purchased from ATCC (PCS-500-010, lot #63216949). Cells were grown in low glucose DMEM (Life Technologies 11885-084) supplemented with 10% FBS (Life Technologies 16000-044, lot #1314735) and 1× penicillin/streptomycin (Life Technologies 15140-122) at 37ºC with 5% CO_2 and 3% O_2. Medium was replaced every 4 days in the absence of passaging. Cells were grown to ~70% confluence and split 1:4 as needed. Cell density was below the concentration needed for reliable detection with a hemocytometer (10^5 cells/mL), so the growth curve was estimated as 2 population doublings per passage.

hMSC senescence-associated β-galactosidase assay

Approximately 5×10^4 hMSCs were plated in one well of a 6 well plate and incubated at 37ºC with 5% CO_2 and 3% O_2 for 4-6 hours. Staining was performed as described in ref. [53].

hMSC differentiation assays

Approximately 10^5 hMSCs were plated in one well of a 12 well plate and grown to full confluence under conditions described above. For adipogenic differentiation, growth medium was replaced with the Adipocyte Differentiation Tool (ATCC, PCS500050), and cells were cultured for 14 days in this medium; medium was refreshed every 4 days. For osteogenic differentiation, growth medium was replaced with the Osteocyte Differentiation Tool (ATCC, PCS500052) and cells were cultured for 21 days; medium was refreshed every 4 days.

After 14 days, adipogenic differentiation was assessed by oil red O staining for lipid droplet accumulation. Briefly, cells were fixed in 3.7% formaldehyde for 1 hour, permeabilized with 60% isopropanol for 5 minutes, and incubated in oil red O stain (1.8mg/mL in 60% isopropanol) for 30 minutes. Osteogenic differentiation was assessed after 21 days of differentiation by staining for calcium deposits with Alizarin Red S. Briefly, cells were fiexed in 3.7% formaldehyde for 10 minutes and incubated with Alizarin Red S solution (2% solution, Lifeline Cell Technology CM-0058) for 10 minutes.

hMSC RNA-seq library preparation and sequencing

Total RNA was extracted from PD12 and PD32 hMSCs using the RNEasy Mini Kit (Qiagen 74104) following the manufacturer's protocol, including the optional DNase treatment. 100ng of total RNA was depleted of rRNA (NEB, E6310) following the manufacturer's protocol; this rRNA-depleted RNA was used as the template for a stranded RNA-seq library using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, E7760) following the manufacturer's protocol. Resulting libraries were submitted to the Human Genome Sequencing Center at Baylor College of Medicine for sequencing on the Illumina HiSeq 2500 sequencing platform. All sequencing data have been deposited in the GEO database at NCBI (#GSE156409).

hMSC ChIP-seq library preparation and sequencing

Approximately 10^6 PD12 and PD32 hMSCs were fixed in 1% formaldehyde at room temperature for 10 minutes. Following 3 washes in cold PBS, cell pellets were flash frozen on dry ice and store at -80ºC. Samples were thawed on ice and resuspended in 100µL RIPA with 10µM PFMS and 1× Halt Protease Inhibitor Cocktail (Thermo Fisher PI78439). Samples were

sonicated to release chromatin with an average fragment length of 400bp in an EpiShear Multi-Sample sonicator (Active Motif 53062) using the following parameters: 5 minutes sonication, 30 second ON/OFF cycles, 50% amplitude at 4ºC.

For each ChIP, 10µL of clarified lysate was added to 90µL of HBSS and used as input into the True MicroChIP Kit (Diagenode, C01010130); manufacturer's instructions were followed beyond the fixation/sonication part of the protocol, using 2µg of each antibody. Antibodies used: histone H3: Active Motif #61475, lot #17316003 and Millipore #05-928, lot #2884434; H3K4me1: AbCam #ab8895, lot #GR1278894; H3K4me3: Diagenode #C15410030, lot #002; H3K9me3: Active Motif #39765, lot #16513004; H3K27ac: Active Motif #39133, lot #01613007; H3K27me3: Active Motif #39155, lot #23813016; H3K36me3: Active Motif #61101, lot #32412003; and TBP: Cell Signaling Technologies #440595, lot #1. 2µL of clarified lysate was used as input and was treated as a sample starting from the reverse crosslinking step in the protocol. The ChIPed DNA was cleaned and concentrated using the Qiaquick PCR Purification Kit (Qiagen, 28104), following the manufacturer's protocol with the following exception: samples were eluted 3× in 40µL of EB and eluates were combined. Sequencing libraries were prepared using the MicroPlex Library Preparation Kit v2 (Diagenode, C05010014) following manufacturer's instructions. Libraries were submitted to the Human Genome Sequencing Center at Baylor College of Medicine for sequencing on the Illumina HiSeq 2500 sequencing platform. All sequencing data have been deposited in the GEO database at NCBI (#GSE156409).

hMSC DECAP-seq library preparation and sequencing

DECAP-seq was performed as described in ref. [18] with some modifications. The starting material was poly-A+ RNA was isolated from 100μg of total RNA using the Dynabeads mRNA Purification Kit (Life Technologies, 61006). 5' phosphate groups were removed using CIP (NEB, M0290) rather than AP, and Cap-Clip acid pyrophosphatase (CellScript, C-CC15011H) was used in place of RppH to de-cap the mRNAs. Size selection of products was performed after 5' adapter ligation, rather than after 1st strand synthesis to select against cDNAs generated solely from binding of the RT primer to the 5' adapter. As in ref. [18], a negative control library was generated by omitting the de-capping step. Libraries were submitted to the Human Genome Sequencing Center at Baylor College of Medicine for sequencing on the Illumina HiSeq 2500 sequencing platform. All sequencing data have been deposited in the GEO database at NCBI (#GSE156409).

Murine NSC isolation

Activated and quiescent NSCs (aNSC and qNSC, respectively) were freshly isolated from the subventricular zone of adult (~7 month) and aged (19-21 month) hGFAP-GFP transgenic mice according to established methods[54,55]. NSCs were isolated by FACS using the surface markers prominin-1 (CD133) and the EGF receptor, and GFAP-GFP fluorescence. For sorting, cells were stained with 1:300 EGF-Alexa 647 (Molecular probes, E-35351), and 1:400 Prominin-1-biotin (eBioscience, 13-1331-80). Dead cells were excluded using propidium iodide and cells were sorted on a BD FACS Aria into PBS. aNSCs and qNSCs can be distinguished by the surface expression of EGFR. Using this scheme, aNSC are prominin-1+; GFP+; EGFR+ and

qNSC are prominin-1+; GFP+; EGFR-. We performed this experiment in biological triplicate, generating each RNA-seq library from a single animal, using ~400 cells per library.

Murine NSC RNA-seq library preparation and sequencing

RNA-seq libraries were generated by GENEWIZ LLC. using the SMART-Seq v4 Ultra Low Input RNA Kit for cDNA synthesis and the Nextera XT DNA Library Preparation Kit. Libraries were sequenced using a 2x125 paired end configuration on an Illumina HiSeq2500. All sequencing data have been deposited in the GEO database at NCBI (#GSE156409).

Data quality control, read trimming, and read mapping

Raw reads were trimmed to remove sequencing adaptors and low quality reads using Trim Galore version 0.4.4 with default parameters (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). RNA-seq data was mapped to the genome using HISAT2 version 2.1.0 (ref. [56]). DECAP-seq data was mapped using STAR version 2.7 (ref. [57]). ChIP-seq data was mapped using bowtie2 version 2.2.4 (ref. [58]). Human genome assembly hg19 and mouse genome assembly mm10 were used as reference genomes in the mapping steps. Reads mapped to the ENCODE blacklist regions[59,60] were filtered. Multi-aligned reads were removed by a homemade script.

RNA-seq data analysis and global cryptic transcription event assessment

After mapping, transcript-level read count, and subsequent FPKM and TPM measurements, were performed using salmon version 1.0.0 (ref. [61]) with following parameters: *-l A –validateMappings*.

An exon-level approach was used to assess cryptic transcription. The exon-level read count was performed using featureCounts version 1.5.3 (ref. [62]) with the following parameters: *-t exon -g transcript_id -f -O -T 8 -p -C*. To eliminate the bias caused by multiple isoforms, cryptic transcript identification was performed on the major transcript of each gene, defined as the transcript with the highest TPM as determined in the transcript-level analysis. Major transcripts with length >3 kb and TPM ≥1 were used for further analysis.

The length-normalized TPM of each exon was calculated using a homemade script, and the expression ratio between the first or second exon and the remaining exons was calculated using the following formula, hereafter called relative expression value (E_i):

$$E_i = \frac{TPM_i/L_i}{TPM_b/L_b}$$

Where *i* is the *i* th exon of a transcript; *L* is the length of the exon; and *b* = 1 or 2, depending on whether the length-normalized expression of the first or second exon was used as the baseline value.

Fold change of relative expression (ratio of ratios) were calculated using the following formula:

$$FC_i = \frac{E_{iO}}{E_{iY}}$$

Where *i* is the *i* th exon of a transcript, *O* is the data of the old sample, and *Y* is the data of the young sample.

Global changes in cryptic transcription with age in paired old and young samples were assessed by calculating the log2-transformed FC_i for the second, third, fourth, and last exons of all major transcripts relative to their first and second exons. A log2-transformed FC_i value of >0 indicates increased cryptic transcription with age. Changes in cryptic transcription were also

assessed after transcripts were divided into quantiles based on their TPM and the same analysis was performed independently for each group. For hMSCs, mHSCs, and NSCs, we additionally compared the averaged E_i values (exons 2 through the second to last exon) in old vs. young cells. A two-tailed Wilcoxon signed-rank test was used to determine differences in cryptic transcription between the old and young samples vs. the null hypothesis that the average log2-transformed FC_i was the same in both samples. To identify specific transcripts that have increased cryptic transcription with age, we considered the exon with the maximum FC_i relative to the first exon for each major transcript; the relative expression value E_i of this exon vs. the first exon is termed the CT score of the transcript. As these values do not have a normal distribution, even after log2-transfromation, the ratio of ratios were ranked, and transcripts with significantly large ratio of ratios were identified using the rank ordering algorithm in ROSE[32,33]. Transcripts with a putative increase in cryptic transcription with age were further filtered by three criteria: 1) all of the exons downstream of the identified exon have a higher relative expression value in the old sample than in the young; 2) transcripts in which the first exon of a transcript has the highest FC_i were excluded; and 3) the absolute expression of the second exon in the old sample must be larger than or similar to the expression in the young sample to ensure that transcripts with few reads mapping to the 5' end of the transcript did not dominate the analysis.

DECAP-seq data analysis

Following initial read mapping, strand-specific reads were separated and analyzed independently in order to increase sensitivity. A metagene plot was constructed to confirm that the majority of reads mapped near annotated TSSes. Cryptic TSSes (cTSSes) were identified in

10bp bins across the genome by comparing sample data to the negative control. Read count assessment was performed using the csaw package version 1.20 (ref. [63]) and the count data was normalized using the TMM method in edgeR[64]. Each bin was tested for significant differences between young or old sample and negative control using a Poisson test with the null hypothesis that for each bin, the signal in the sample is equal to the signal in the negative control.

$$p = Pois(x; \lambda)$$

Where p is the output p-value, x is the number of reads in a 10 bp bin of the sample data (young or old sample) and λ is the number of reads in the same region in the negative control data.

Neighboring bins were then merged into 100 bp windows, with the most significant bin representing the window. Adjustment for multiple comparisons were perform by calculating FDR using the Benjamini & Hochberg method[65]. cTSS peaks were defined as windows with FDR < 0.05 and fold change > 1.2 in the young or old sample vs. the negative control. Windows that overlapped with annotated TSSes and 3' UTRs were filtered out. A common cTSS was defined as a peak identified in both young and old datasets; otherwise, the cTSS was considered unique to the young or old sample.

Age-associated cTSS peaks are a subset of cTSS peaks in old sample. Age-associated cTSS peaks were defined as windows with significantly more reads than negative control (FDR < 0.05) and significantly more reads than the young sample (FDR < 0.05 and fold change > 1.2). Significance was determined using a Poisson test using signals in the young sample as λ. The null hypothesis is that for each bin, signal abundance in the old sample is the same to that in the young sample. Merge of neighboring bins and multiple comparison correction were performed as described above.

Analysis of age-increased cryptic transcription

The population distribution of length and expression levels of transcripts with age-increased cryptic transcription was compared to the major transcripts of expressed genes, as defined above. Significance was determined using a two-tailed Wilcoxon rank sum test vs. the null hypothesis that they are equal.

Motif features within 200bp of age-increased cTSSes were identified by HOMER findMotif.pl with default parameters[47].

To assess whether transcripts that have increased cryptic transcription are associated with a change in gene expression, FPKM fold changes (old/young) of all the major transcripts were calculated and ordered by rank, from negative fold changes to positive fold changes. The distribution of the fold change rank of transcripts with age-increased cTSSes were compared to transcripts with age-decreased cTSSes, using the distribution of the rank list of all the genes as control.

RNA-seq data was used to validate the age-increased cTSSes identified in the DECAP-seq analysis. RNA-seq read coverage 100 bp up- and downstream of age-associated cTSSes was counted using deeptools version 3.2.0 (ref. [66]) with the computeMatrix function and a 10bp bin size. The sum of read counts for each bin was calculated in the young and the old sample and its ratio was computed. A two-sided Wilcoxon rank sum test vs. the null hypothesis that the ratio was equal in young and old samples was used to determine significance.

Heatmaps and metagenes plot were made using deeptools version 3.2.0 (ref. [66])using plotHeatmap and computeMatrix function, respectively.

ChIP-seq data analysis

After mapping, duplicated reads were removed using Picard (http://broadinstitute.github.io/picard). Peak calling was performed using MUSIC[67]. Differentially bound regions were identified by the csaw package[63] with default parameters, except that H3K9me3, H3K36me3 and H3K27me3 were performed using the get_motimal_broad_ERs model and peaks were called for the remaining datasets using get_optimal_punctate_ERs model. Differentially bound regions were identified by comparing the changes between young and old data using the csaw package[63] with H3 total and input as internal control for histone markers and TBP, respectively.

Heatmaps of read abundance of histone markers in genic regions and around TSSes/cTSSes were generated by deeptools version 3.2.0 (ref. [66]).

TBP ChIP-seq quality was assessed by determining the percentage of peaks located on promoter regions, defined as 1 kb up- and downstream from annotated TSSes.

TBP read abundance within 500bp of age-increased cTSSes or endogenous TSSes was calculated in 10bp bins using cTSSes or TSSes as reference and processed TBP bigwig files as input; regions without TBP signal were excluded. Three clusters were made using the k-means method based on the pattern of TBP read coverage around cTSSes or TSSes.

H3K36me3 coverage around age-increased cTSSes and endogenous TSSes was calculated using deeptools computeMatrix function[66] in 10 bp bins. Endogenous TSSes of major transcripts were used in this analysis (defined above). The sum of read depth in this region was calculated for cTSSes and TSSes in both young and old samples, and changes were assessed using a two-sided Wilcoxon rank sum test vs. the null hypothesis that the normalized read depth is equal in the young and old samples.

Chromatin state analysis

Chromatin states were identified using a 10 state model in chromHMM[49] with pre-defined peaks as input. In addition to pre-defined genomic regions in the package, we also assessed the enrichment of chromatin states 1kb up- and downstream of TSSes and transcription end sites (TESes), using genomic coordinates from hg19.

The genomic positions of LADs were downloaded from USCS genomic browser using human assembly hg19. Heatmaps of the LADs regions were made with the deeptools plotHeatmap function[66] using histone modification signal tracks as sample and H3 total signal tracks as control. k-means clustering was used to separate the LADs into two groups based on the signal enrichment of H3K9me3. Heatmaps of H3K9me3 and H3K27me3 were made for both young and old data with the LADs pileup in the same order.

cTSS region prediction from histone modification patterns

Putative age-associated cTSS regions in hMSCs and mHSCs were predicted by changes in H3K4me3 and H3K36me3 enrichment. Non-promoter genic regions, defined as gene body regions at least 1kb away from any endogenous TSSes, were considered putative age-increased cTSS regions if they: 1) contained significantly more H3K4me3 reads in the old sample and the regions did not overlap with H3K36me3 peak in the old sample than the young; or 2) contained significantly fewer H3K36me3 reads in the old sample than the young and the region was within 2kb of a H3K4me3 peak in the old sample. A Poisson test was used to determine whether reads were increased (for H3K4me3) or decreased (for H3K36me3) in the old sample relative to the young. Signal abundance the young sample was used as λ. The null hypothesis is that for each peak region, signal abundance in the old sample is the same as in the young.

Analysis of putative age-increased cTSS regions

Promoter characteristics of putative age-associated cTSS regions were analyzed by the promoter prediction algorithm Promoter – 2.0 (ref. [52]). The promoter score was defined as the number of sequences containing promoter features identified by the algorithm from 100 2kb sequences. In each iteration, 100 randomly selected promoter sequences, 100 randomly selected putative age-increased cryptic promoters and 100 randomly selected genomic sequences were analyzed by the algorithm separately. This process was repeated 100 times to get a distribution. The sample sequences analyzed each time were distinct and samples were randomly selected without replacement. Promoter sequences were defined as 1kb up- and downstream of endogenous TSSes. Putative age-increased cryptic promoters were defined as 1kb up- and downstream of the midpoint of the identified age-increased cTSS regions. Random genomic sequences were 2kb sequences that randomly extracted from all autosomes.

Validation of increased cryptic transcription with age from putative age-increased cTSSes was assessed by analyzing RNA-seq and DECAP-seq data. The number of RNA-seq reads were compared between the exon downstream of the identified region and the first exon of the transcript. Read counts were normalized by total mapped read counts and exon length. A two-sided Wilcoxon rank sum test was used to determine if this ratio was different between the samples. Validation of these sites in hMSCs were also conducted by comparing the DECAP-seq signal within 1kb of the midpoint of the identified region in young and old samples. A two-sided Wilcoxon rank sum test was used to determine significance with the null hypothesis that the normalized DECAP-seq signal is equal in the young and old samples.

References

1. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194 (2013).
2. Booth, L. N. & Brunet, A. The Aging Epigenome. *Mol. Cell* **62**, 728–744 (2016).
3. Sousa-Victor, P. *et al.* Geriatric muscle stem cells switch reversible quiescence into senescence. *Nature* **506**, 316–321 (2014).
4. Rowland, J. *et al.* Uncovering genetic mechanisms of kidney aging through transcriptomics, genomics, and epigenomics. *Kidney Int.* **95**, 624–635 (2019).
5. White, R. R. *et al.* Comprehensive transcriptional landscape of aging mouse liver. *BMC Genomics* **16**, 1–15 (2015).
6. Lai, R. W. *et al.* Multi-level remodeling of transcriptional landscapes in aging and longevity. *BMB Rep.* **52**, 86–108 (2019).
7. Sun, D. *et al.* Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell* **14**, 673–688 (2014).
8. Enge, M. *et al.* Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. *Cell* **171**, 321-330.e14 (2017).
9. Laun, P. *et al.* A comparison of the aging and apoptotic transcriptome of Saccharomyces cerevisiae. *FEMS Yeast Res.* **5**, 1261–1272 (2005).
10. Girardot, F., Lasbleiz, C., Monnier, V. & Tricoire, H. Specific age related signatures in Drosophila body parts transcriptome. *BMC Genomics* **7**, 1–16 (2006).
11. Zhan, M. *et al.* Temporal and spatial transcriptional profiles of aging in Drosophila melanogaster. *Genome Res.* **17**, 1236–1243 (2007).
12. Sen, P. *et al.* H3K36 methylation promotes longevity by enhancing transcriptional fidelity.

Genes Dev. **29**, 1362–1376 (2015).

13. Pu, M. *et al.* Trimethylation of Lys36 on H3 restricts gene expression change during aging and impacts life span. *Genes Dev.* **29**, 718–731 (2015).
14. Wang, C. M., Tsai, S. N., Yew, T. W., Kwan, Y. W. & Ngai, S. M. Identification of histone methylation multiplicities patterns in the brain of senescence-accelerated prone mouse 8. *Biogerontology* **11**, 87–102 (2010).
15. Wood, J. G. *et al.* Chromatin remodeling in the aging genome of Drosophila. *Aging Cell* **9**, 971–978 (2010).
16. Hennig, B. P. & Fischer, T. Chromatin and cryptic transcription. *Transcription* **4**, 97–101 (2013).
17. Carvalho, S. *et al.* Histone methyltransferase SETD2 coordinates FACT recruitment with nucleosome dynamics during transcription. *Nucleic Acids Res.* **41**, 2881–2893 (2013).
18. Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72–77 (2017).
19. Xie, L. *et al.* KDM5B regulates embryonic stem cell self-renewal and represses cryptic intragenic transcription. *EMBO J.* **30**, 1473–1484 (2011).
20. Venkatesh, S. & Workman, J. L. Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* **16**, 178–189 (2015).
21. Belotserkovskaya, R. *et al.* FACT facilitates transcription-dependent nucleosome alteration. *Science (80-.).* **301**, 1090–1093 (2003).
22. Kaplan, C. D., Laprade, L. & Winston, F. Transcription elongation factors repress transcription initiation from cryptic sites. *Science (80-.).* **301**, 1096–1099 (2003).
23. Carrozza, M. J. *et al.* Histone H3 methylation by Set2 directs deacetylation of coding

regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**, 581–592 (2005).

24. McDaniel, S. L. & Strahl, B. D. Shaping the cellular landscape with Set2/SETD2 methylation. *Cell. Mol. Life Sci.* **74**, 3317–3334 (2017).

25. Ni, Z., Ebata, A., Alipanahiramandi, E. & Lee, S. S. Two SET domain containing genes link epigenetic changes and aging in Caenorhabditis elegans. *Aging Cell* **11**, 315–325 (2012).

26. Goodell, M. A. & Rando, T. A. Stem cells and healthy aging. *Science (80-.).* **350**, 1199–1204 (2015).

27. Wagner, W. *et al.* Aging and replicative senescence have related effects on human stem and progenitor cells. *PLoS One* **4**, (2009).

28. Smolle, M. & Workman, J. L. Transcription-associated histone modifications and cryptic transcription. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1829**, 84–97 (2013).

29. Ferrari, K. J. *et al.* Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. *Mol. Cell* **53**, 49–62 (2014).

30. Zhang, Y. *et al.* H3K36 histone methyltransferase Setd2 is required for murine embryonic stem cell differentiation toward endoderm. *Cell Rep.* **8**, 1989–2002 (2014).

31. Xu, Q. *et al.* SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. *Nat. Genet.* **51**, 844–856 (2019).

32. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).

33. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).

34. Urbán, N., Blomfield, I. M. & Guillemot, F. Quiescence of Adult Mammalian Neural

Stem Cells: A Highly Regulated Rest. *Neuron* **104**, 834–848 (2019).

35. Adelman, E. R. *et al.* Aging Human Hematopoietic Stem Cells Manifest Profound Epigenetic Reprogramming of Enhancers That May Predispose to Leukemia. *Cancer Discov.* **9**, 1080–1101 (2019).

36. Boisvert, M. M., Erikson, G. A., Shokhirev, M. N. & Allen, N. J. The Aging Astrocyte Transcriptome from Multiple Regions of the Mouse Brain. *Cell Rep.* **22**, 269–285 (2018).

37. Clarke, L. E. *et al.* Normal aging induces A1-like astrocyte reactivity. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E1896–E1905 (2018).

38. Fleischer, J. G. *et al.* Predicting age from the transcriptome of human dermal fibroblasts. *Genome Biol.* **19**, 1–8 (2018).

39. Kaisers, W. *et al.* Age, gender and UV-exposition related effects on gene expression in in vivo aged short term cultivated human dermal fibroblasts. *PLoS One* **12**, 1–21 (2017).

40. MacRae, S. L. *et al.* DNA repair in species with extreme lifespan differences. *Aging (Albany. NY).* **7**, 1171–1184 (2015).

41. Marthandan, S. *et al.* Similarities in Gene Expression Profiles during In Vitro Aging of Primary Human Embryonic Lung and Foreskin Fibroblasts. *Biomed Res. Int.* **2015**, (2015).

42. Marthandan, S. *et al.* Conserved senescence associated genes and pathways in primary human fibroblasts detected by RNA-seq. *PLoS One* **11**, 1–31 (2016).

43. Rai, T. S. *et al.* HIRA orchestrates a dynamic chromatin landscape in senescence and is required for suppression of Neoplasia. *Genes Dev.* **28**, 2712–2725 (2014).

44. Stilling, R. M. *et al.* De-regulation of gene expression and alternative splicing affects distinct cellular pathways in the aging hippocampus. *Front. Cell. Neurosci.* **8**, 1–15

(2014).

45. Johnson, B. S. *et al.* Biotin tagging of MeCP2 in mice reveals contextual insights into the Rett syndrome transcriptome. *Nat. Med.* **23**, 1203–1214 (2017).

46. Zhang, W. *et al.* A Werner syndrome stem cell model unveils heterochromatin alterations as a driver of human aging. *Science (80-.).* **348**, 1160–1163 (2015).

47. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**, 576–589 (2010).

48. Feser, J. & Tyler, J. Chromatin structure as a mediator of aging. *FEBS Lett.* **585**, 2041–2048 (2011).

49. Ernst, J. & Kellis, M. ChromHMM: Automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).

50. Kimura, H. Histone modifications for human epigenome analysis. *J. Hum. Genet.* **58**, 439–445 (2013).

51. Tsurumi, A. & Li, W. X. Global heterochromatin loss: A unifying theory of aging? *Epigenetics* **7**, 680–688 (2012).

52. Knudsen, S. Promoter2.0: For the recognition of PolII promoter sequences. *Bioinformatics* **15**, 356–361 (1999).

53. Dimri, G. P. *et al.* A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 9363–9367 (1995).

54. Codega, P. *et al.* Prospective Identification and Purification of Quiescent Adult Neural Stem Cells from Their In Vivo Niche. *Neuron* **82**, 545–559 (2014).

55. Leeman, D. S. *et al.* Lysosome activation clears aggregates and enhances quiescent neural

stem cell activation during aging. *Science (80-.).* **359**, 1277–1283 (2018).

56. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

57. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

58. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

59. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 1–5 (2019).

60. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

61. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).

62. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

63. Lun, A. T. L. & Smyth, G. K. Csaw: A Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.* **44**, e45 (2015).

64. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, (2010).

65. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. . Ser. B (Methodol.)* **57**, 289–300 (2016).

66. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

67. Harmanci, A., Rozowsky, J. & Gerstein, M. MUSIC: identification of enriched regions in ChIP-Seq experiments using a mappability-corrected multiscale signal processing framework. *Genome Biol.* **15**, 474 (2014).

Figure Legends

Figure 1. Analysis of RNA-seq in mHSCs and hMSCs suggests an age-associated increase in CT in mammalian stem cells. A and B) Boxplots of the log2-transformed ratio of reads mapping to the second, third, fourth, and last exon vs. reads mapping to the first exon (dark orange on the left) or vs. reads mapping to the second exon (light orange on the right) in old vs. young samples (ratio in old divided by ratio in young, or ratio of ratios), as indicated, in mHSCs (A) and hMSCs (B). All expressed transcripts were included in the analysis for mHSC (n=10068) and hMSCs (n=9230). A two-sided Wilcoxon signed-rank test was used to determine significance with the null hypothesis that the calculated log2 ratios are equal to 0. C) Ranked plot of CT scores for mHSCs (top) and hMSCs (bottom). The red lines show inflection points where the slope of the curve has a tangent of 1; transcripts with significantly increased CT scores, *i.e.*, those showing evidence of an age-associated increase in cryptic transcription, are located to the right of second inflection point. D) Heatmaps of the CT scores of transcripts with age-increased cryptic transcription shows an increase in old vs. young cells in both mHSCs (left) and hMSCs (right). E) Metagene plot of RNA-seq read density in old (red) and young (blue) samples. mHSCs are on top; hMSCs on the bottom. F) Examples of age-associated cryptic transcription in mHSCs (left, *Car11* gene) and hMSCs (right, *IAH1* gene). The panels show mapped reads from young samples (top, blue), old samples (middle, red), and the difference (old-young, red depicts regions where reads are higher in old samples, blue in young; bottom panel).

Figure 2. Analysis of RNA-seq in NSCs and other mammalian tissues suggests a widespread increase in CT during mammalian aging. A) PCA analysis of transcriptomes in

murine NSCs shows that the primary differences in transcription are based on sex and status (activated vs. quiescent). B) Boxplots of the log2-transformed ratio of reads mapping to the second, third, fourth, and last exon vs. reads mapping to the first exon (dark orange on the left) or vs. reads mapping to the second exon (light orange on the right) in old vs. young samples (ratio in old divided by ratio in young, or ratio of ratios), as indicated, in activated (top) or quiescent (bottom) NSCs in females (left) and males (right). All expressed transcripts were included in the analysis for active female NSC (n=6110), quiescent female NSC (n=4985), active male NSC (n=4654), and quiescent male NSC (n=2956). A two-sided Wilcoxon signed-rank test was used to determine significance with the null hypothesis that the calculated log2 ratios are equal to 0. C) Heatmaps of the CT score of cryptic transcripts shows an increase in old vs. young cells in aNSCs from both female (left) and male (right) mice. D) Metagene plot of RNA-seq read density in old (red) and young (blue) aNSC samples. Cells isolated from females are on the top; those from males on the bottom.

Figure 3. Sequencing the 5' ends of capped RNA shows increased CT in a human stem cell aging model. A) Metagene plot showing the distribution of normalized DECAP-seq reads along expressed genes. Young sample in blue; old shown in red. B) Venn diagram showing the overlap of DECAP-seq peaks (cryptic TSSes) in the young and old samples. C) Bar chart indicating the number of DECAP-seq peaks with significantly increased reads in old (red bar, age-inc. cTSS) and young (blue bar, age-dec. cTSS) samples. The DECAP-seq peaks in the first group represent sites at which cryptic transcription increases with age, referred to as age-associated cTSSes. D) Read pile up of DECAP-seq signal surrounding the age-associated cTSSes; young sample is on the left, and old on the right. E) Examples of age-associated CT in the *GAPDH* and *CAPAS1*

genes. The panels show mapped DECAP-seq reads from young samples (top, blue) and old samples (bottom, red). F and G) Log2 transformed old to young ratio of averaged normalized RNA-seq reads mapping to the 200bp interval surrounding age-associated cTSSes. All such sites are shown in (F); only those sites located in introns are shown in (G). H) Histogram and boxplot of length of genes that undergo age-associated cryptic transcription (n=961, blue) and all genes (n=14062, red). P-values were calculated by a two-sided Wilcoxon signed-rank test with the null hypothesis that the two groups have the same average length. I) Boxplot of the log2-transformed FPKM of genes that undergo age-associated CT (n=836) compared to all expressed genes (n=13148). Genes with FPKM < 1 were filtered. J) HOMER analysis of the DNA sequence ±200bp from the identified age-associated cTSSes shows enrichment for specific motifs. The *de novo* identified motifs with significant p-values are shown, along with the closest matching known motif for each.

Figure 4. Chromatin states associated with transcription are largely preserved during aging. A) Chromatin states identified by ChromHMM using a 10 state model. B) Proportion of the genome covered by ChromHMM states 1-9 in young (top) and old (bottom) samples. C) Stacked bar graph showing how chromatin states 1-9 change with age, with each column representing the totality of the indicated state in the young sample and the shaded regions indicating its state in the old sample. D) Proportional dot plot showing the change in chromatin state with age, segregated by genomic region. Each column represents the total genomic space of the indicated region, with the size of the dot for each state indicating the age averaged proportion of that region that is covered by that state; the color of the dot represents its change with age in that region. E) Heatmaps depicting the changes in repressive histone modifications H3K9me3

and H3K27me3 in LADs with age. For both modifications, the maps on the left shows the young sample. F) Metagene plot showing the same LAD regions as in (E). G) Read pile ups showing the enrichment of the indicated histone modifications along expressed genes (FPKM≥1). For each modification, the map on the left shows the young sample.

Figure 5. Chromatin near cTSSes takes on promoter-like characteristics with age. A) Read pile ups showing enrichment of H3K36me3 (left) and H3K4me3 (right) in old and young hMSCs surrounding the 1375 age-increased cTSSes identified by DECAP-seq. Loci are clustered by H3K4me3 enrichment pattern. B) Metagene plots showing TBP enrichment patterns of endogenous TSSes (n=10215, top) and the age-increased cTSSes (n=134, bottom) previously identified that are associated with TBP peaks. cTSSes without TBP signal were excluded from analysis. C) Depiction of the predicted change in chromatin state our algorithm uses to find additional putative cTSSes, based on the change in chromatin state observed at cTSSes in panel (A). H3K4me3 enrichment is represented in gold and H3K36me3 in green. D) An example of a putative age-increased cTSS identified by our chromatin state algorithm. The predicted cTSS in *ATOH8* is highlighted in light blue. H3K4me3, H3K36me3, and DECAP-seq signal tracks are shown, with the signal for the young sample in blue and the old in red. E) Boxplot showing the log-normalized ratio of ratios of RNA-seq reads mapping to the exon downstream of the putative age-increased cTSSes identified by chromatin state vs. the first exon of the transcript in old vs. young hMSCs. Predicted age-increased cTSSes found within expressed major transcripts (n = 1056, teal) and randomly selected genic regions (n=1000, blue) were analyzed. F) Boxplot showing promoter prediction scores for endogenous TSSes (n=100, red), putative age-increased cTSSes (n=100, teal), and random DNA sequences (n=100, blue) in hMSCs. G and H) as (E) and

(F), respectively, except these depict putative age-increased cTSSes identified in mHSCs (n=264). P-values for both hMSCs and mHSCs boxplots were calculated by a two-sided Wilcoxon rank-sum test against the null hypothesis that the RNA-seq ratios or promoter scores are equal in the two groups, as appropriate.

Figure 6. Model of the mechanisms driving elevated cryptic transcription in mammalian aging.

In young cells, H3K36me3 (red peaks) within the gene body of actively transcribed genes serves as a scaffold for the chromatin modifiers KDM5B and DNMT3B (top panel). These enzymes remove intragenic H3K4me3 (blue peaks) and confer DNA methylation (solid black circles) at CpG residues, respectively, generating a repressive chromatin environment that maintains cTSSes in an inactive state[18,19]. During aging, H3K36me3 levels are reduced within gene bodies, which inhibits the recruitment of KDM5B and DNMT3B to these regions, resulting in accumulation of H3K4me3 and reduced DNA methylation (bottom panel). This results in a chromatin state that is more permissive for transcription initiation at cTSSes (red arrow), causing an elevation of cryptic transcription with age.

Figures

Fig. 1

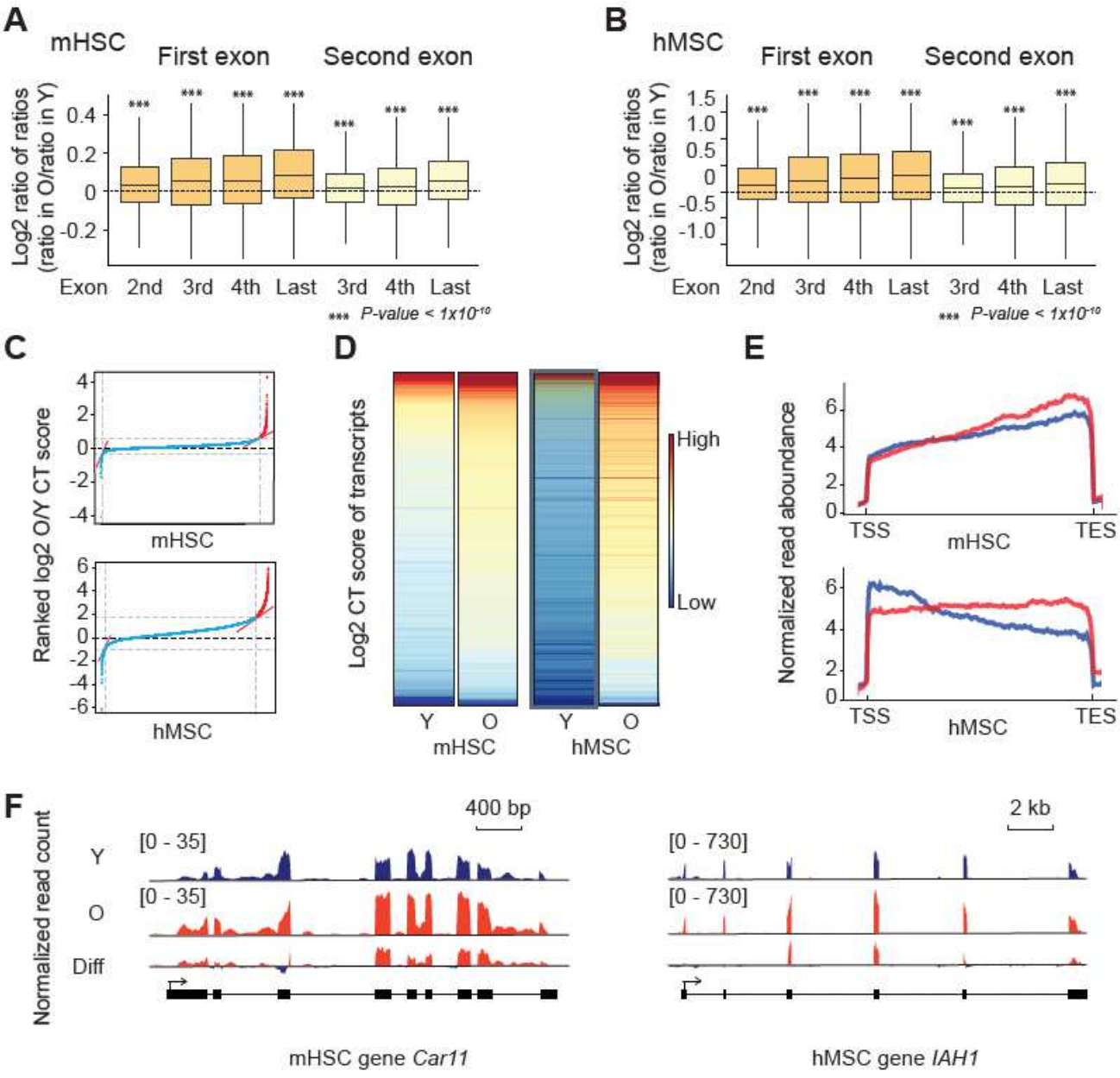

Figure 1

Analysis of RNA-seq in mHSCs and hMSCs suggests an age-associated increase in CT in mammalian stem cells. A and B) Boxplots of the log2-transformed ratio of reads mapping to the second, third, fourth, and last exon vs. reads mapping to the first exon (dark orange on the left) or vs. reads mapping to the second exon (light orange on the right) in old vs. young samples (ratio in old divided by ratio in young, or ratio of ratios), as indicated, in mHSCs (A) and hMSCs (B). All expressed transcripts were included in the analysis for mHSC (n=10068) and hMSCs (n=9230). A two-sided Wilcoxon signed-rank test was used to

determine significance with the null hypothesis that the calculated log2 ratios are equal to 0. C) Ranked plot of CT scores for mHSCs (top) and hMSCs (bottom). The red lines show inflection points where the slope of the curve has a tangent of 1; transcripts with significantly increased CT scores, i.e., those showing evidence of an age-associated increase in cryptic transcription, are located to the right of second inflection point. D) Heatmaps of the CT scores of transcripts with age-increased cryptic transcription shows an increase in old vs. young cells in both mHSCs (left) and hMSCs (right). E) Metagene plot of RNA-seq read density in old (red) and young (blue) samples. mHSCs are on top; hMSCs on the bottom. F) Examples of age-associated cryptic transcription in mHSCs (left, Car11 gene) and hMSCs (right, IAH1 gene). The panels show mapped reads from young samples (top, blue), old samples (middle, red), and the difference (old-young, red depicts regions where reads are higher in old samples, blue in young; bottom panel).

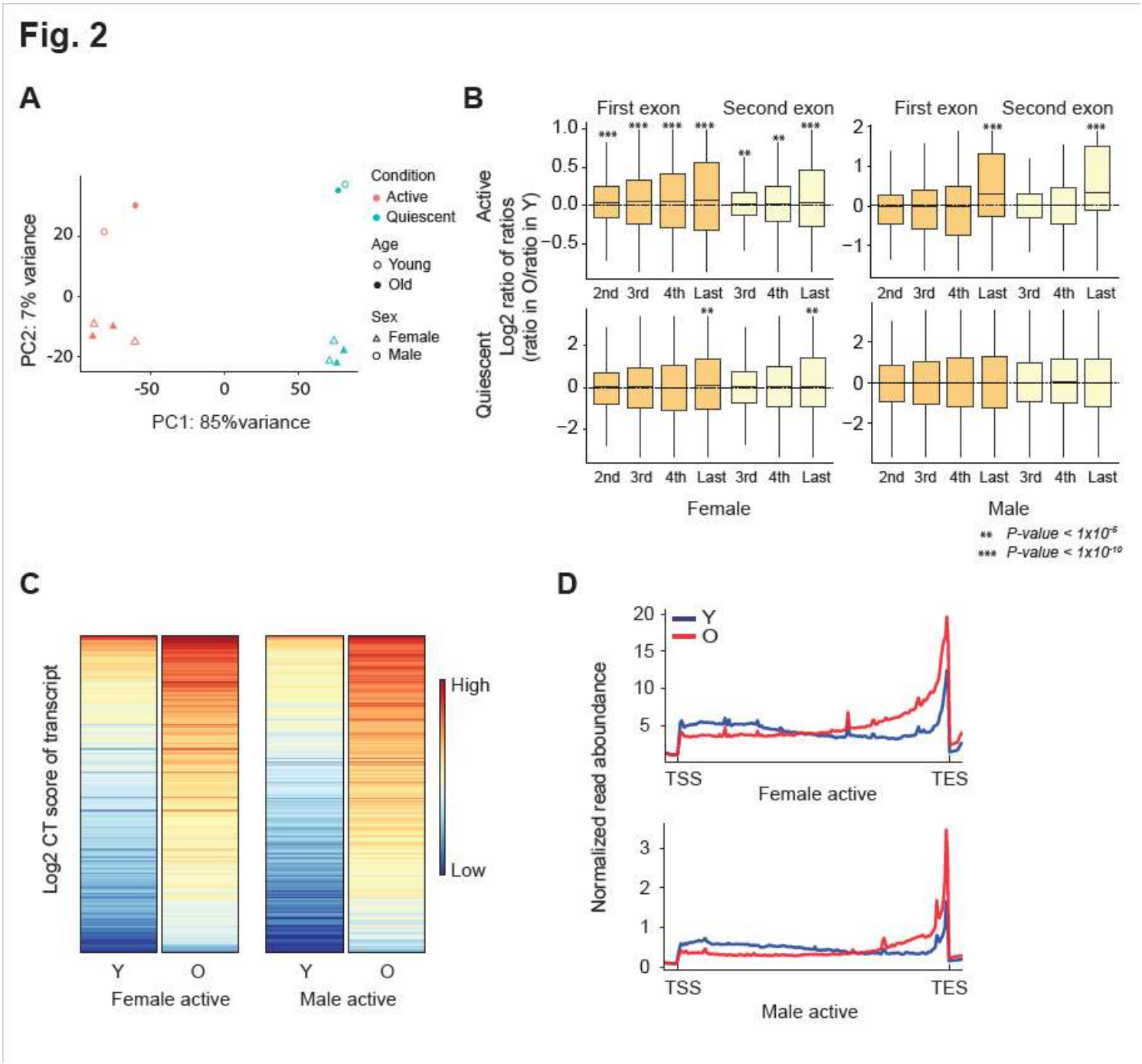

Figure 2

Analysis of RNA-seq in NSCs and other mammalian tissues suggests a widespread increase in CT during mammalian aging. A) PCA analysis of transcriptomes in murine NSCs shows that the primary differences in transcription are based on sex and status (activated vs. quiescent). B) Boxplots of the log2-transformed ratio of reads mapping to the second, third, fourth, and last exon vs. reads mapping to the first exon (dark orange on the left) or vs. reads mapping to the second exon (light orange on the right) in old vs. young samples (ratio in old divided by ratio in young, or ratio of ratios), as indicated, in activated (top) or quiescent (bottom) NSCs in females (left) and males (right). All expressed transcripts were included in the analysis for active female NSC (n=6110), quiescent female NSC (n=4985), active male NSC (n=4654), and quiescent male NSC (n=2956). A two-sided Wilcoxon signed-rank test was used to determine significance with the null hypothesis that the calculated log2 ratios are equal to 0. C) Heatmaps of the CT score of cryptic transcripts shows an increase in old vs. young cells in aNSCs from both female (left) and male (right) mice. D) Metagene plot of RNA-seq read density in old (red) and young (blue) aNSC samples. Cells isolated from females are on the top; those from males on the bottom.

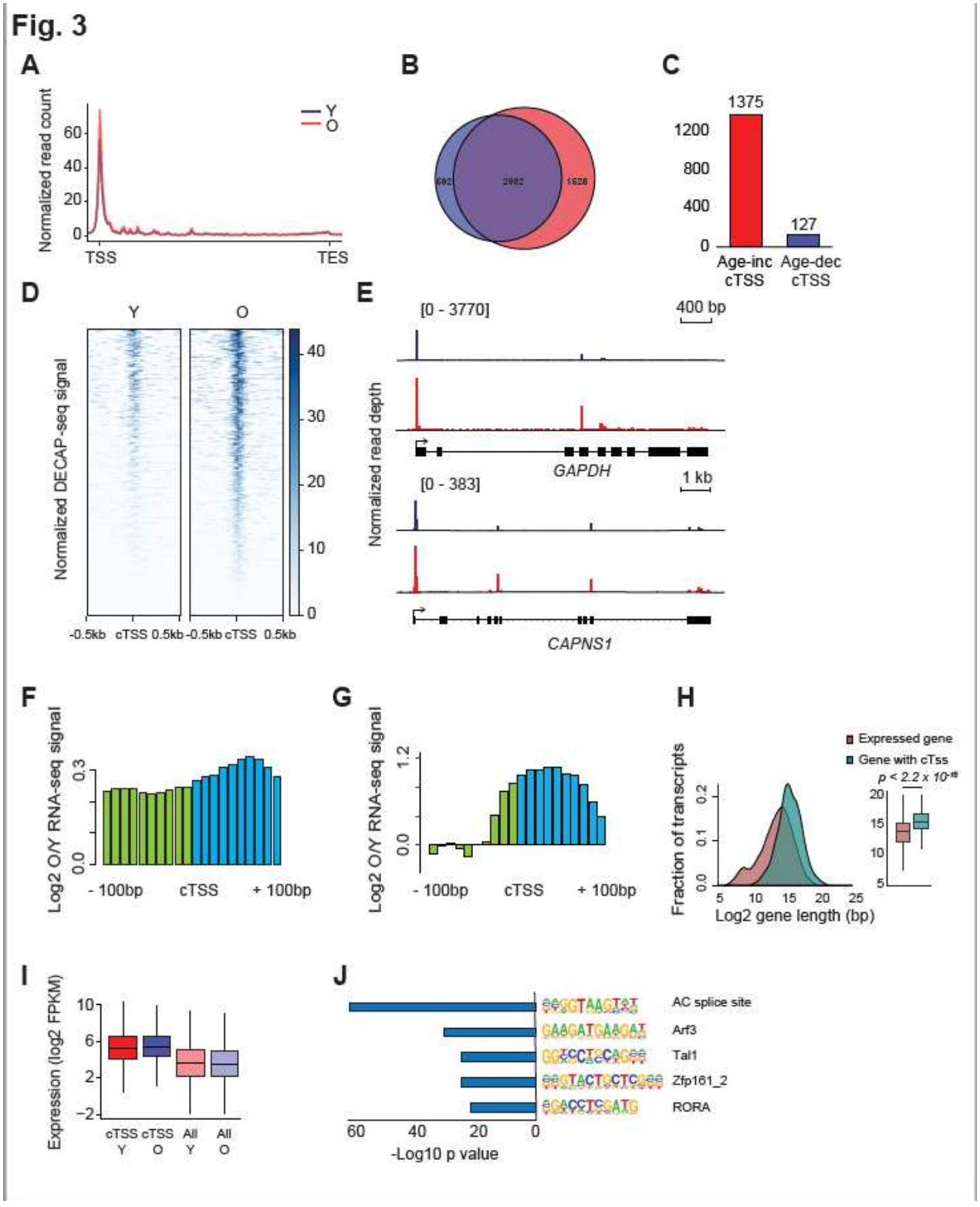

Figure 3

Sequencing the 5' ends of capped RNA shows increased CT in a human stem cell aging model. A) Metagene plot showing the distribution of normalized DECAP-seq reads along expressed genes. Young sample in blue; old shown in red. B) Venn diagram showing the overlap of DECAP-seq peaks (cryptic TSSes) in the young and old samples. C) Bar chart indicating the number of DECAP-seq peaks with significantly increased reads in old (red bar, age-inc. cTSS) and young (blue bar, age-dec. cTSS) samples.

The DECAP-seq peaks in the first group represent sites at which cryptic transcription increases with age, referred to as age-associated cTSSes. D) Read pile up of DECAP-seq signal surrounding the age-associated cTSSes; young sample is on the left, and old on the right. E) Examples of age-associated CT in the GAPDH and CAPAS1 genes. The panels show mapped DECAP-seq reads from young samples (top, blue) and old samples (bottom, red). F and G) Log2 transformed old to young ratio of averaged normalized RNA-seq reads mapping to the 200bp interval surrounding age-associated cTSSes. All such sites are shown in (F); only those sites located in introns are shown in (G). H) Histogram and boxplot of length of genes that undergo age-associated cryptic transcription (n=961, blue) and all genes (n=14062, red). P-values were calculated by a two-sided Wilcoxon signed-rank test with the null hypothesis that the two groups have the same average length. I) Boxplot of the log2-transformed FPKM of genes that undergo age-associated CT (n=836) compared to all expressed genes (n=13148). Genes with FPKM < 1 were filtered. J) HOMER analysis of the DNA sequence ±200bp from the identified age-associated cTSSes shows enrichment for specific motifs. The de novo identified motifs with significant p-values are shown, along with the closest matching known motif for each.

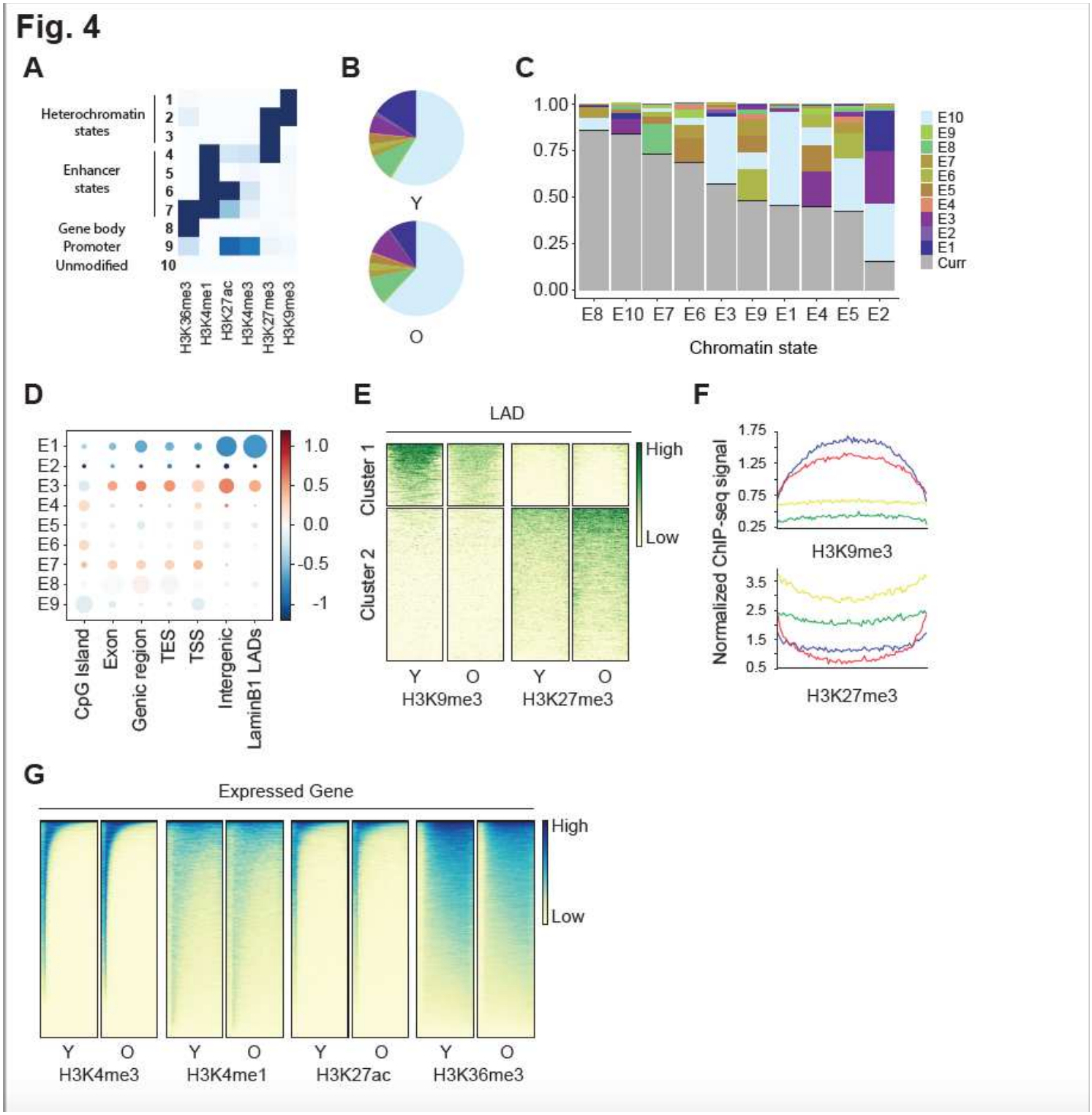

Figure 4

Chromatin states associated with transcription are largely preserved during aging. A) Chromatin states identified by ChromHMM using a 10 state model. B) Proportion of the genome covered by ChromHMM states 1-9 in young (top) and old (bottom) samples. C) Stacked bar graph showing how chromatin states 1-9 change with age, with each column representing the totality of the indicated state in the young sample and the shaded regions indicating its state in the old sample. D) Proportional dot plot showing the change in chromatin state with age, segregated by genomic region. Each column represents the total

genomic space of the indicated region, with the size of the dot for each state indicating the age averaged proportion of that region that is covered by that state; the color of the dot represents its change with age in that region. E) Heatmaps depicting the changes in repressive histone modifications H3K9me3 and H3K27me3 in LADs with age. For both modifications, the maps on the left shows the young sample. F) Metagene plot showing the same LAD regions as in (E). G) Read pile ups showing the enrichment of the indicated histone modifications along expressed genes (FPKM≥1). For each modification, the map on the left shows the young sample.

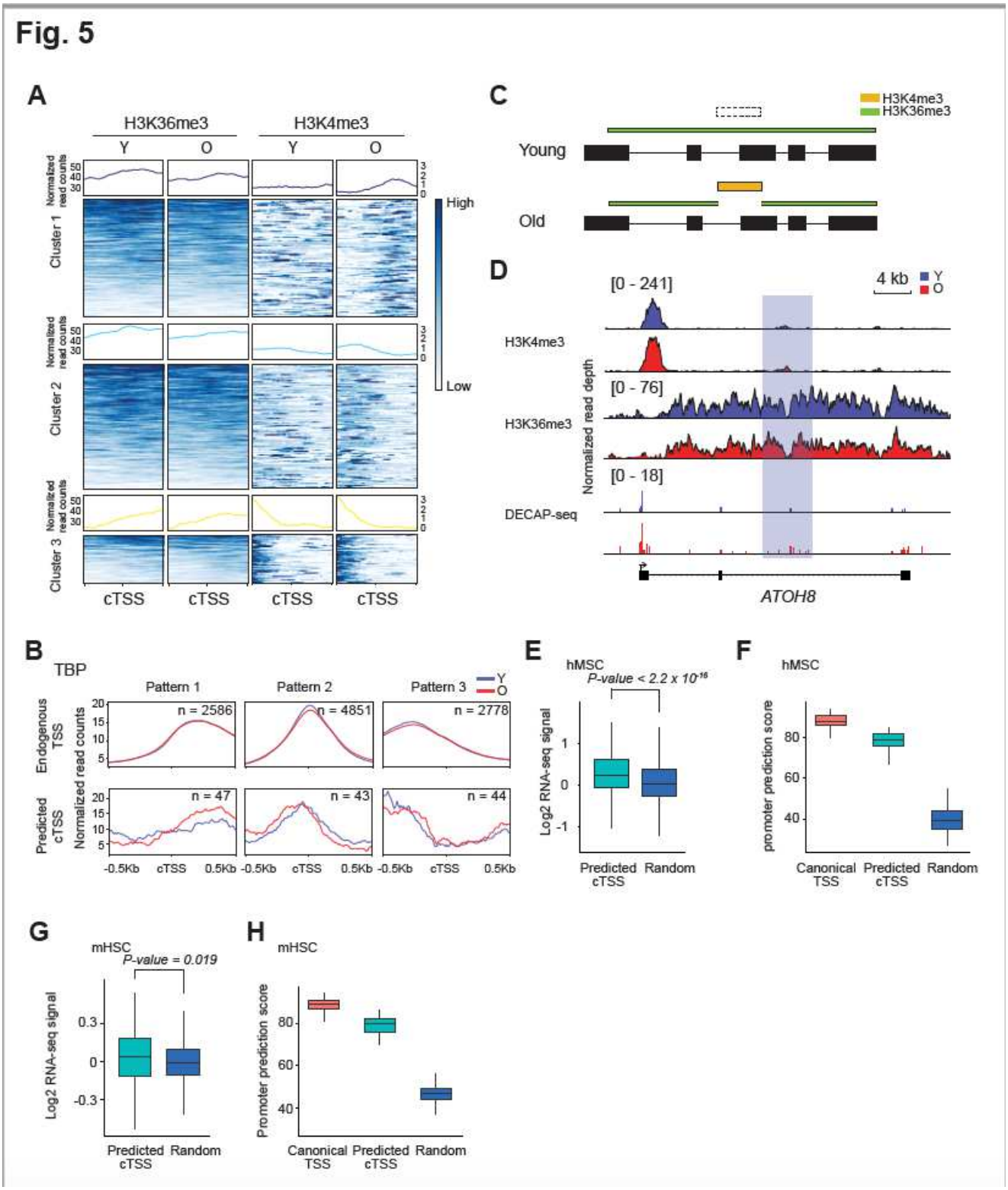

Figure 5

Chromatin near cTSSes takes on promoter-like characteristics with age. A) Read pile ups showing enrichment of H3K36me3 (left) and H3K4me3 (right) in old and young hMSCs surrounding the 1375 age-increased cTSSes identified by DECAP-seq. Loci are clustered by H3K4me3 enrichment pattern. B) Metagene plots showing TBP enrichment patterns of endogenous TSSes (n=10215, top) and the age-increased cTSSes (n=134, bottom) previously identified that are associated with TBP peaks. cTSSes without TBP signal were excluded from analysis. C) Depiction of the predicted change in chromatin state our algorithm uses to find additional putative cTSSes, based on the change in chromatin state observed at cTSSes in panel (A). H3K4me3 enrichment is represented in gold and H3K36me3 in green. D) An example of a putative age-increased cTSS identified by our chromatin state algorithm. The predicted cTSS in ATOH8 is highlighted in light blue. H3K4me3, H3K36me3, and DECAP-seq signal tracks are shown, with the signal for the young sample in blue and the old in red. E) Boxplot showing the log-normalized ratio of ratios of RNA-seq reads mapping to the exon downstream of the putative age-increased cTSSes identified by chromatin state vs. the first exon of the transcript in old vs. young hMSCs. Predicted age-increased cTSSes found within expressed major transcripts (n = 1056, teal) and randomly selected genic regions (n=1000, blue) were analyzed. F) Boxplot showing promoter prediction scores for endogenous TSSes (n=100, red), putative age-increased cTSSes (n=100, teal), and random DNA sequences (n=100, blue) in hMSCs. G and H) as (E) and (F), respectively, except these depict putative age-increased cTSSes identified in mHSCs (n=264). P-values for both hMSCs and mHSCs boxplots were calculated by a two-sided Wilcoxon rank-sum test against the null hypothesis that the RNA-seq ratios or promoter scores are equal in the two groups, as appropriate.

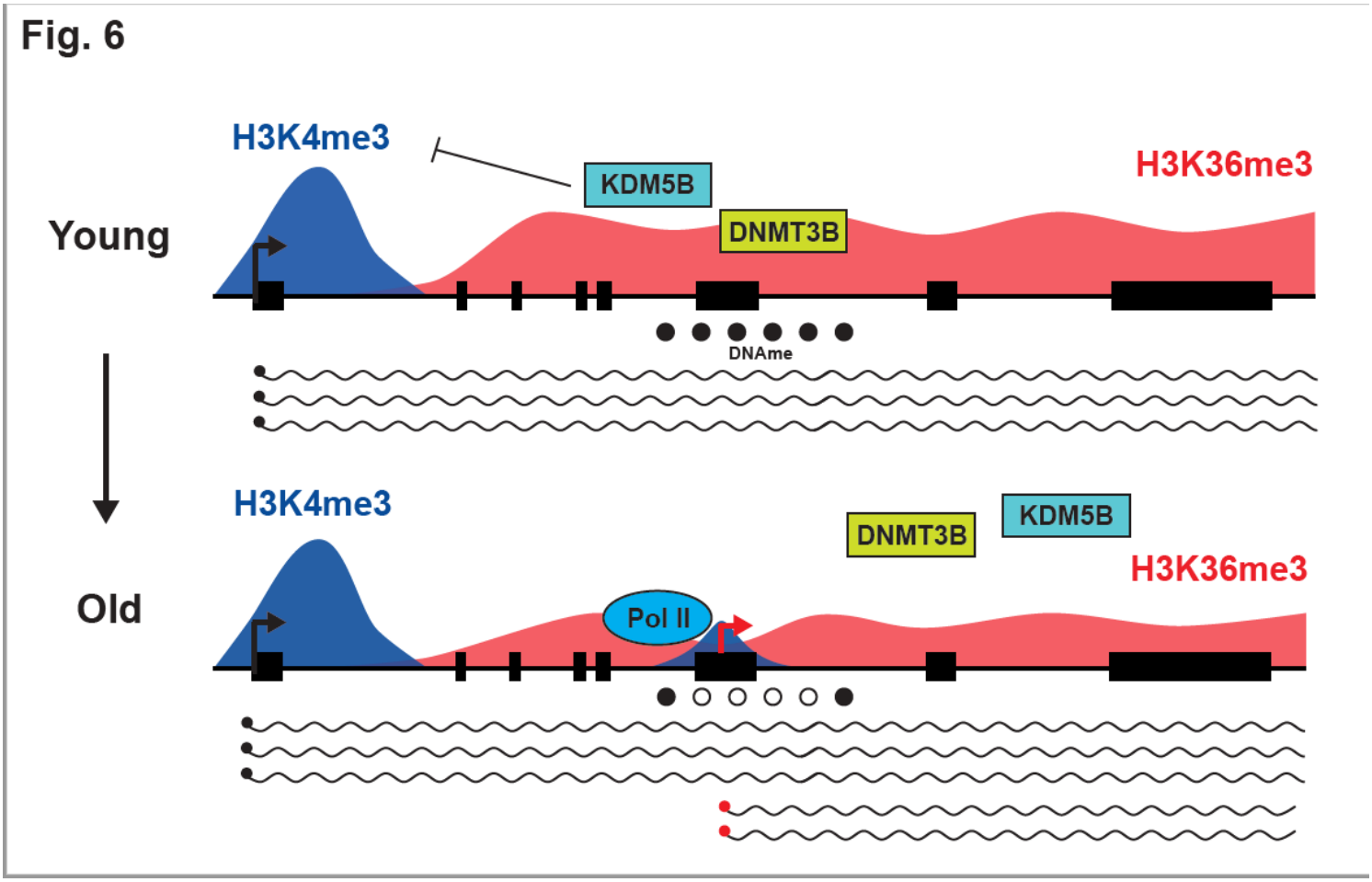

Figure 6

Model of the mechanisms driving elevated cryptic transcription in mammalian aging. In young cells, H3K36me3 (red peaks) within the gene body of actively transcribed genes serves as a scaffold for the chromatin modifiers KDM5B and DNMT3B (top panel). These enzymes remove intragenic H3K4me3 (blue peaks) and confer DNA methylation (solid black circles) at CpG residues, respectively, generating a repressive chromatin environment that maintains cTSSes in an inactive state18,19. During aging, H3K36me3 levels are reduced within gene bodies, which inhibits the recruitment of KDM5B and DNMT3B to these regions, resulting in accumulation of H3K4me3 and reduced DNA methylation (bottom panel). This results in a chromatin state that is more permissive for transcription initiation at cTSSes (red arrow), causing an elevation of cryptic transcription with age.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SIFigures.pdf
- SupplementalFigureLegends.pdf