Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements .

Marco De Cecco,[1] Steven W. Criscione,[1] Edward J. Peckham,[1] Sara Hillenmeyer,[1] Eliza A. Hamm,[1] Jayameenakshi Manivannan[1], Abigail L. Peterson[1], Jill A. Kreiling,[1] Nicola Neretti,[1] John M. Sedivy[1,2]

Supplemental Experimental procedures, Figures and Tables

Supplemental Experimental procedures

Preparation of FAIRE DNA

For the FAIRE-chip experiments, cells were grown from one thaw and propagated using one batch of serum, but were passaged and harvested on 3 separate occasions. Each harvest was then processed separately by FAIRE, and subsequently processed individually on the tiled arrays. Hence, the FAIRE-chip dataset was acquired from 3 independent replicates. At a later time, cells were grown again for the dot blot experiments, which required a larger amount of material. A new thaw was made, and the previous procedure was followed again, but scaled up. The FAIRE samples were processed as 3 independent replicates, and were treated as such throughout the dot blot experiments. For the FAIRE-seq experiments, equal parts of the 3 samples from the dot blot experiments were mixed and sequenced in a single lane as a pool. The same pools were also used for the qPCR validations shown in Fig. S9-S11. In addition, we ran several primer pairs on the 3 replicates individually, and in no case did we see significant differences between the individual replicates.

To prepare FAIRE DNA, cells were crosslinked by adding 37% formaldehyde directly to the culture medium to a final concentration of 1% and incubating at room temperature for 4 min on a rotating platform. The formaldehyde was quenched by adding glycine to a final concentration of 125 mM, and incubation was continued for 10 min at room temperature. The dishes were washed twice with 5 ml chilled PBS, and harvested by scraping into 1.5 ml PBS containing 1 mM PMSF. All subsequent steps were performed on ice. Cells were collected by centrifugation and resuspended in FAIRE lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA). Cell suspensions were sonicated in an ice water bath using a Fischer Scientific Sonic Dismembrator (model 500). Each sample received 5 sessions of 2 min in length (1 sec on and 1 sec off at 25% intensity), followed by 2 min of rest on ice. In later experiments a Bioruptor UCD-200 (Diagenode) instrument was used, set to pulse on high (30 sec followed by 30 sec rest) for a total time of 15 min. Samples were cleared of debris by centrifugation in a microfuge at top speed (5 min, 4°C). The sonication was verified by running DNA samples (decrosslinked and purified) on agarose gels to ensure fragment sizes in the 100-500 bp range. Cleared samples were extracted with an equal volume of phenol, chloroform, isoamyl alcohol (25:24:1, Sigma) by vortexing at high speed for 30 sec. Phases were separated by centrifugation in a microfuge at top speed (1 min, 4°C). The top aqueous layer was removed and extracted again with an equal volume of phenol, chloroform, isoamyl alcohol. The first organic phase was back extracted with an equal volume of 10 mM Tris-HCl pH 7.5, 1 mM EDTA (TE). The two aqueous phases were combined, and extracted one more time with an equal volume of phenol, chloroform, isoamyl alcohol. DNA was precipitated by addition of 2 volumes of 95% ethanol, 0.1 volumes of 3 M sodium acetate, and glycogen to a final concentration of 20 µg/ml. Pellets were resuspended in 50 µl of 10 mM Tris-HCl pH 7.5, and incubated at 37^{o}C for 2 hr with 0.2 mg/ml of RNase A and 50 U/ml RNase T1 (Fermentas). Proteinase K (Invitrogen) was added to a concentration of 0.2 mg/ml and incubation was continued overnight at 65^{o}C. Finally, the samples were extracted with an equal volume of phenol, chloroform, isoamyl alcohol, precipitated with ethanol as described above, and resuspended in a small volume of TE. Two types of "input" controls were used. In the first, samples were processed as above, but the formaldehyde crosslinking step was omitted. In the second, cells were crosslinked, but the crosslinks were reversed (by incubation at 65°C in the presence of proteinase K) before the phenol, chloroform, isoamyl alcohol extraction step. These two methods yielded very similar

results.

FAIRE-seq

DNA prepared by the FAIRE method was sequenced on an Illumina GAIIx instrument, and later on an Illumina HighSeq 2000 instrument by the Brown Genomics Core. The read lengths were 42 nucleotides on the GAIIx, and 50 nucleotides on the HighSeq 2000. In both cases single-end reads were used. Further data on the number of reads and their mapability are provided in Table S2. 100 ng of FAIRE-extracted (or input) DNA per sample was end repaired with the End-It DNA end repair kit (Epicentre), according to the provided protocol. DNA was purified using Agencourt® AMPure XP paramagnetic beads (Beckman Coulter), and eluted in molecular grade water. dATP was added to the DNA ends (Kozarewa & Turner, 2011), and after another DNA purification, pre-annealed adapters were ligated to each sample (Quail *et al.*, 2008). 10 cycles PCR amplification were performed with Phusion High-Fidelity DNA Polymerase (New England Biolabs), and the libraries were agarose gel purified in the range of 200-500 base pairs. The sequencing data were uploaded to the Galaxy platform (https://main.g2.bx.psu.edu/) (Giardine *et al.*, 2005; Blankenberg *et al.*, 2010; Goecks *et al.*, 2010), mapped, filtered and normalized to the number of reads in each sample. These uniquely mapping reads were processed for peaks using the Model-based Analysis of ChIP-Seq (MACS) algorithm (Zhang *et al.*, 2008), and the resulting Browser Extendible Data (BED) files were evaluated for coverage, peaks size, and distribution. Both the BED files and the sorted mapped reads were processed by the gene scoring analysis of the EpiChIP software (Hebenstreit *et al.*, 2011), using a window of 1.5 kb upstream and downstream of each transcriptional start site (TSS) in the human genome (hg18 or hg19) annotation files. Finally, a correlation analysis was performed with the gene scoring results and the expression values from the Genome U133 Plus 2.0 microarray dataset calculated with Probe Logarithimic Intensity Error (PLIER) software (Affymetrix).

Computation of FAIRE enrichment in genomic features

For Fig. 1B, K-means clustering was performed on the mean interpolated FAIRE enrichment signals of the 19,524 selected RefSeq genes (Fig 1A). We varied the number of clusters from 2 to 10, and independently the window from 2 to 10 kbp, in various combinations. No features that were considered significant were detected using larger windows or higher cluster numbers. However, we reproducibly detected peaks centered on the TSS using 2 clusters and a window of 3 kbp. The variability at the TSS level (taking into account all TSS) can be seen as the shaded area in Fig. 1A. Fig. 1B shows the average enrichment for the class of TSS with signal. The other class was essentially flat and hence not informative, and thus was not shown in the figure.

For Fig. 1C, we used 36,589 enhancer regions that were predicted in HeLa cells by Heintzman *et al.* (2009). They performed ChIP-chip throughout the entire human genome, mapping HeLa enrichment patterns of H3K4me1 and H3K4me3. They binned the ChIP-chip data into 100 bp bins, averaging probes that fell into the same bin. Using a sliding window of H3K4Me1 and H3K4Me3 features, they scanned a training set of enhancer patterns defined previously by p300 binding sites, keeping only those windows that correlated most with the training sets and had significant enrichment of chromatin modifications.

For FAIRE-seq data, the FAIRE and input samples were first aligned to the unmasked reference human genome (build hg18 or hg19) using the Bowtie short read aligner (version 0.12.7) (Langmead *et al.*, 2009), requiring reads to map uniquely to the genome. Late and early replicating coordinate BED files were generated using published datasets of replication timing in

human cells (Hansen *et al.*, 2010). Late and early replicating regions were determined by calculating the log2 ratio of late/early from the normalized data. Late replication regions were defined as > 1.5 log2 ratio late/early. Early replicating regions were defined as < 1.5 log2 ratio late/early. The H3K9me3 and H3K4me3 genomic feature BED files were generated from broadPeak files describing regions of ChIP-seq enrichment for normal human lung fibroblasts (NHLF) (Ernst *et al.*, 2011) which were obtained from ENCODE. Regions that were defined as active promoters and strong enhancers were also obtained from the UCSC genome browser tracks generated by Ernst *et al.* (2001) and converted to a BED file format. BEDTools was used to compute the coverage, namely, the number of reads in a sample BAM alignment file overlapping with the features specified in a chosen BED coordinates file, using the *coverageBed* tool (Quinlan & Hall, 2010). The read counts occupying the features in the BED coordinates file were normalized to the total reads that mapped uniquely to the genome for each individual sample. FAIRE-seq enrichment was examined at early-replicating, late-replicating, H3K9me3, and H3K4me3 genomic features by computing the log2 (FAIRE/input) fold change for senescent and early passage cells for all features. Kernel smoothing density estimation (Bowman & Azzalini, 1997) was performed on the computed log2 fold changes to generate an estimate of the probability density function using Matlab (ver. R2012a) for early and late replicating tracks. P-values were computed by a two-tailed t-test assuming unequal variance.

We estimated the probability that the observed enrichments in early-replicating, late-replicating, H3K9me3, and H3K4me3 genomic features could have arisen by chance by performing a bootstrap randomization of BED file genomic features. A random BED file of genomic features was generated from the regions annotated in a given BED file, such that regions with lengths specified by the BED file were randomly shuffled throughout the genome. Chromosomes were selected randomly with a probability p = (chromosome length / total length of all chromosomes), and a coordinate start position within the chosen chromosome was selected randomly, excluding positions within genome assembly gaps. The coordinate BED files were shuffled in this manner for 2000 iterations. BEDTools was used to compute the coverage in the shuffled coordinate BED file, using the *coverageBed* tool (Quinlan & Hall, 2010). The read counts were normalized to the total reads that mapped uniquely to the genome for each individual sample. The log2 (FAIRE/input) values for senescent and early passage cells were computed for the randomized coordinates and the original coordinate BED file. To compare early-replicating, late-replicating, H3K9me3, and H3K4me3 regions versus random features we estimated a p-value by calculating sequential two-tailed t-tests assuming equal variance between the coordinate BED file log2 fold changes and a subset of the log2 fold changes for the random features with equal sample size for 1000 iterations. The p-value was then estimated by the mean of the 1000 independent t-tests.

Computation of repetitive element enrichment

The software pipeline described by Day *et al.* (Day *et al.*, 2010) was used to examine differences in repetitive element enrichment in FAIRE-seq datasets. Source code and repetitive enrichment estimator software were downloaded from the support website cited in the publication (http://compbio.med.harvard.edu/repeats/). A current human repetitive element assembly was generated from available source code and repeat annotation files downloaded from Repbase (http://www.girinst.org/repbase/; EMBL format, April 18, 2012 release) and RepeatMasker (http://www.repeatmasker.org/; hg19, Feb. 2009, Repeat Library 20120124 release) (Jurka *et al.*, 2005). The Day *et al.* software constructs repetitive element pseudogenomes from canonical

(Repbase annotation) and genomic (RepeatMasker annotation) instances of repeat element sequences. To generate our current human repetitive element assembly we used the *instance-only* option, which builds repetive element pseudogenomes only from the genomic instances (RepeatMasker annotation) of repeat elements. Standard Burrows-Wheeler Aligner (BWA, version 0.4.6) commands were used to index the resultant human repeat element assemblies (Li & Durbin, 2009). Alignments to the human repetitive element assemblies were performed using the BWA aligner. The *readmap* tool in the Day *et al.* software was used to compute read count tables, which are coverage estimates for the repeat elements. The read counts were normalized to the total number of mapping reads (including reads mapping to multiple locations) in order to account for differences in read counts between experimental samples. Fold changes were computed for the normalized read counts by calculating the Log2 (FAIRE/input) for early passage and senescent cell samples. A direct comparison of FAIRE signals between early and senescent samples was performed by computing Log2 (senescent FAIRE/early FAIRE). The repetitive element class and family annotations was used to group repeat elements into larger phylogenetic categories (Alu, L1, SVA and Satellites).

Dot blotting and hybridization

DNA preparations (FAIRE and input) from three independent experiments of early passage and senescent cells were deposited in serial 2-fold dilutions (40 ng – 2.5 ng per spot) onto a positively-charged nylon membrane (BioRad) under alkaline conditions using a 96 well vacuum manifold. DNA was additionally fixed by baking at 80^{o}C for 2 hr. Membranes were washed in 0.5 x SSC, 0.5% SDS (1 hr at 65^{o}C), and prehybridized (2 hr at 65^{o}C) and hybridized (24 hr at 65^{o}C) under standard aqueous conditions (Maniatis *et al.*, 1989) (6x SSPE, 5x Denhardt's solution, 0.5% SDS, 100 µg/ml denatured herring sperm carrier DNA). The hybridization solution also included 5% dextran sulfate (Fisher Scientific) and ^{32}P-dCTP labeled probe. Probe fragments for Alu (285 bp) and L1 (383 bp) were PCR amplified from human total genomic DNA using primers described in Table S1, purified by agarose gel electrophoresis, and labeled using the random octadeoyribonucleotide kit (NEB). Unincorporated nucleotides were removed using an Illustra microspin G-25 gel filtration column (GE Healthcare). Two washes for in 2x SSC, 1% SDS at 65^{o}C for 10 min. were followed by a stringent wash at in 0.1x SSC, 0.1 % SDS at 60°C for 30 min. The latter was repeated until the average background signal on the membrane fell below 200 cpm using a handheld Geiger counter. The membranes were exposed to storage phosphor screens and imaged on a Typhoon 9410 (GE Healthcare). Several identical membranes were prepared. After hybridization with one probe, the membranes were stripped and hybridized with the other probe.

The amounts of DNA deposited on the membranes were initially quantified using the Nano-Drop instrument (Thermo Scientific). To further improve accuracy, these values were corrected using careful qPCR measurements of the DNA samples, using 5 primer sets (TNF1, ETF3, Z281, CCD1, C191), which were previously determined to be in FAIRE non-enriched regions of the genome in both early passage and senescent cells. The best fit line generated from a dilution series performed with each primer pair was then used to quantify and fine tune the effective DNA concentrations that were deposited on the membrane. The measurements made by qPCR were in very close agreement with independent measurements made using the Quant-it PicoGreen dsDNA Assay Kit (Invitrogen).

The signal intensity of each spot on the membranes was measured as the average pixel intensity of a circle with a fixed area (0.204 cm^2) centered on the spot. Background was

calculated by measuring 20 circles of equal area around the periphery of the membrane. The background-subtracted signal of each spot was plotted against its qPCR corrected DNA concentrations, a best-fit line was generated for each sample, and the equation of this line was used to calculate the DNA-normalized signal of that sample. Values from the 3 independent experiments were averaged, and the FAIRE signals were normalized to their inputs. Finally, input-normalized senescent FAIRE signals were normalized to early passage FAIRE signals.

Design of PCR primers

See Table S1 for a listing of all primers. Primers used in Fig. 3B (dot blotting) were based on consensus sequences to Alu elements (Batzer *et al.*, 1996; Weisenberger *et al.*, 2005). Primers for L1 were designed to the 3' end of ORF2, and were biased to preferentially amplify elements of the primate-specific L1PA and human-specific L1H subfamilies. These L1 primer sequences were analyzed using the UCSC genome browser *in silico* PCR tool, which identified 1383 genomic L1 elements as the most likely products of amplification. The primary subfamilies that were amplified were L1HS, L1PA2, L1PA3, and L1PA4, with 134, 395, 632, and 160 genomic positions, respectively (accounting for 96% of L1 target sites identified). In Fig. 3C, primers to detect Alu RNA were described by Marullo *et al.* (Marullo *et al.*, 2010). Primers to detect L1 RNA (ORF2) were described by Coufal *et al.* (Coufal *et al.*, 2009), and are also biased for the detection of L1PA and L1H subfamilies.

In order to detect individual repetitive elements, the most recent RepeatMasker table (hg19Patch5) containing annotated masked regions was downloaded (version 3.3.0, RMLib: 20110920). The table was subsequently filtered to return the family of repeats desired. For L1 and AluYb9, the full list of annotated elements was then intersected with the coordinates of late-replicating regions, to specifically identify elements in those domains. Complete inclusion within a domain was required. For L1 elements, in addition a >5 kb overlap was imposed. The genomic coordinates of the resulting intervals were then used in qPCR primer design, and primer pairs were sought to contain at least 2 mismatches to all non-target sites, which were evaluated by BLAST searches against the human reference genome assembly as well as the nucleotide database (nr, nt and RefSeq). All primer pairs were then tested for their amplification efficiency (E) using three 10-fold serial dilutions of genomic DNA. Only primer pairs above a specified efficiency threshold were kept for further analysis: L1, E >0.95; Alu, E >0.95, hSATII, E >0.92. Primer pairs that failed these criteria or produced primer dimers were discarded. Next, the highly efficient primer pairs were compared for their Ct values against two different known single copy control amplicons, and only pairs with a $\Delta\Delta$Ct of <0.9 were retained.

Chromatin immunoprecipitation

All procedures followed the protocols in the Magna ChIP kit (Millipore). Briefly, $2x10^6$ cells were crosslinked in their culture dishes with 1% formaldehyde (10 min, room temperature), quenched with glycine, washed twice with ice-cold PBS containing protease inhibitors, and finally scraped into a microfuge tube. Cell pellets were resuspended in 100 µl lysis buffer, and sonicated with a Bioruptor UCD-200 instrument (Diagenode), set to pulse on high (30 sec followed by 30 sec rest) for a total time of 10 min. The extracts were centrifuged in a microfuge (top speed, 5 min, 4°C) to remove debris, supernatants were placed in fresh tubes, and diluted 10-fold with ChIP dilution buffer containing protease inhibitors. A small aliquot (~ 1%) of each chromatin preparation was saved as input for further quantification, and the remainder was divided into two 500 µl immunoprecipitation reactions. Antibodies (mH2A, 3 µg, provided by

Peter Adams; histone H4, 10 µg, clone 62-141-13, Millipore) were added, followed by 20 µl of magnetic protein A beads, and the reactions were incubated overnight at 4°C with constant rotation. The immunocomplexes on the beads were carefully washed as described in the Magna ChIP instructions (once each with low salt wash buffer, high salt wash buffer, LiCl wash buffer, and twice with TE buffer). The captured DNA was eluted and decrosslinked by incubation in the presence of proteinase K for 2 hr at 65°C, followed immediately by 10 min at 95°C. The samples were subsequently cleaned up by phenol-chloroform extraction, precipitated with ethanol, and resuspended in 16 µl of TE. The yield was quantified using a Qubit 2.0 dsDNA HS assay kit (Invitrogen), and normalized either to starting cell number or input DNA.

Electron microscopy

Cells were grown in 10 cm dishes as indicated. Early passage and senescent cells were harvested by scraping, washed in 0.1 M sodium cacodylate pH 7.4, and fixed with 2.5% glutaraldehyde in 0.1 M sodium cacodylate for 4 hr at room temperature. Cells were post fixed with 1.5% potassium ferrocyanide and 1% osmium tetroxide in 0.1M sodium cacodylate, followed by 1% aqueous thiocarbohydrazide, then 2% aqueous osmium tetroxide. Samples were dehydrated in ice cold, graded ethanol series (70%, 90%, 100%), then ice cold anhydrous acetone. Infiltration of specimens was achieved using Durcupan ACM resin (Electron Microscopy Sciences) with decreasing proportions of acetone, and finally Durcupan alone. Cells were embedded in Durcupan and polymerized at 60°C for 48 hrs. Cells were en bloc stained with 1% aqueous uranyl acetate, followed by Walton's lead aspartate. Cell blocks were thin sectioned at 80 nm on a Reichert Ultracut Ultramicrotome, placed on copper grids and viewed on a Philips 410 transmission electron microscope equipped with 1 megapixel Advantage HR CCD camera (Advanced Microscopy Techniques). Images were taken at a magnification of 54,800x. Heterochromatin content of the region peripheral to the nuclear envelope was determined by analyzing the area within 200 nm of the membrane. This was done by drawing a series of ROI (regions of interest, which were rectangles), that on one side abutted (but did not include) the inner membrane of the nuclear envelope, and on the other side projected by 200 nm into the nucleoplasm. The pixels in these ROI were the analyzed for their intensity using ImageJ open source software (http://rsbweb.nih.gov/ij/). A threshold was applied, such that all pixels darker than this threshold were considered heterochromatin, and all pixels lighter were considered not to be heterochromatin. The threshold was established by averaging a number of ROI drawn randomly in the lightest regions of the nucleoplasm, similarly averaging a number of ROI drawn randomly in the darkest regions of the nucleoplasm, and taking the threshold as the midpoint. We also arbitrarily varied the threshold (both up and down), and this treatment did not change the results of the analysis. Significance was determined using Student's t test.

Fluorescence in situ hybridization

Cells were grown on coverslips, fixed with 4% paraformaldehyde for 20 min at room temperature, washed 3x in phosphate buffered saline (PBS), and stored in 70% ethanol at 4°C until used. Cells were rehydrated in PBS, denatured in 80% formamide, 1x sodium citrate buffer (SSC; 15 mM sodium citrate, 150 mM NaCl, pH 7.0), for 5 min at 80°C, immediately dehydrated in a graded ethanol series (70%, 90%, 100%), and air dried. Human centromeric FISH probe (to the Sat.2 repeat) was obtained from Exiqon, and used according to the supplier's instructions. After overnight hybridization the cover slips were washed twice with 15% formamide, 1x SSC, and once with 1x SSC for 15 min at room temperature. Coverslips were mounted using Prolong

Gold with DAPI (Invitrogen). Cells were visualized with a Zeiss LSM 710 confocal laser scanning microscope equipped with a 63x Plan-Apochromat objective (NA 1.4) and a 34-channel QUASAR detector. A Z-series was collected through each nucleus and three dimensional reconstructions were performed using ImageJ software.

References for supplemental Materials and Methods

Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM, Schmid CW, Zietkiewicz E, Zuckerkandl E (1996) Standardized nomenclature for Alu repeats. *J. Mol. Evol.* **42**, 3-6.

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* **Chapter 19**, 1-21.

Bowman AW, Azzalini A (1997). *Applied smoothing techniques for data analysis*. New York: Oxford University Press.

Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, Morell M, O'Shea KS, Moran JV, Gage FH (2009) L1 retrotransposition in human neural progenitor cells. *Nature* **460**, 1127-1131.

Day DS, Luquette LJ, Park PJ, Kharchenko PV (2010) Estimating enrichment of repetitive elements from high-throughput sequence data. *Genome Biol.* **11**, R69.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* **15**, 1451-1455.

Goecks J, Nekrutenko A, Taylor J (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**, R86.

Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, Dorschner MO, Gartler SM, Stamatoyannopoulos JA (2010) Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 139-144.

Hebenstreit D, Gu M, Haider S, Turner DJ, Lio P, Teichmann SA (2011) EpiChIP: gene-by-gene quantification of epigenetic modification levels. *Nucleic Acids Res.* **39**, e27.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462-467.

Kozarewa I, Turner DJ (2011) Amplification-free library preparation for paired-end Illumina sequencing. *Methods Mol. Biol.* **733**, 257-266.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760.

Maniatis T, Fritsch EF, Sambrook J (1989). *Molecular Cloning: A Laboratory Manual (2nd edition)*: Cold Spring Harbor Laboratory Press.

Marullo M, Zuccato C, Mariotti C, Lahiri N, Tabrizi SJ, Di Donato S, Cattaneo E (2010) Expressed Alu repeats as a novel, reliable tool for normalization of real-time quantitative RT-PCR data. *Genome Biol.* **11**, R9.

Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow H, Turner DJ (2008) A large genome center's improvements to the Illumina sequencing system. *Nat. Methods* **5**, 1005-1010.

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842.

Weisenberger DJ, Campan M, Long TI, Kim M, Woods C, Fiala E, Ehrlich M, Laird PW (2005) Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res.* **33**, 6823-6836.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137.

Supplemental Figure 1

SA-β-Gal activity

Early passage Senescent

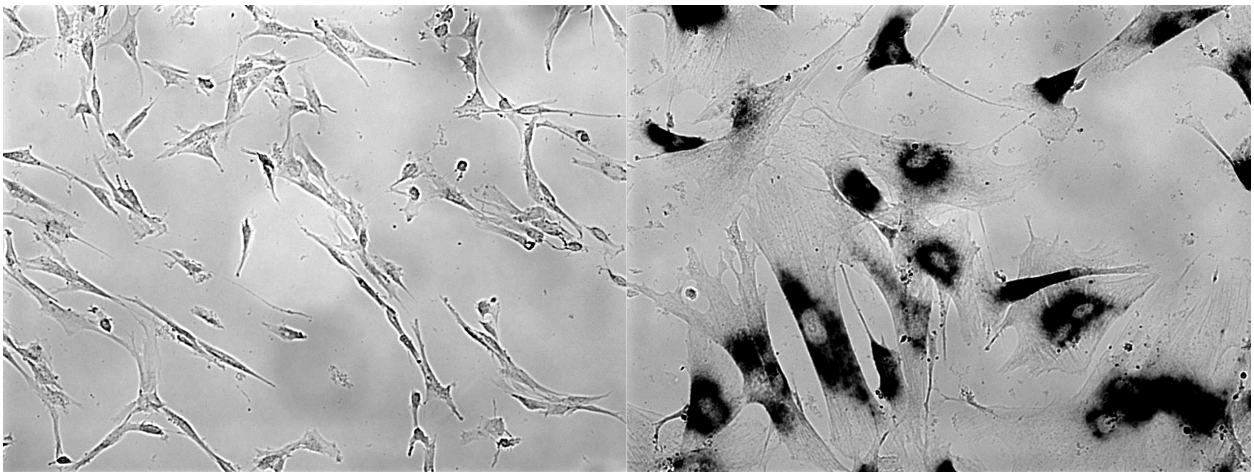

Fig. S1. Staining of early passage and senescent cells for senescence-associated beta galactosidase activity (SA-β-Gal). Cultures were propagated and stained for SA-β-Gal activity as indicated in Supplemental Materials, Materials and Methods. The culture pictured here on right was stained 4 weeks after the number of cells ceased to increase. Counting over 200 cells showed that >98.5% of the cells stained strongly for SA-β-Gal.

Supplemental Figure 2

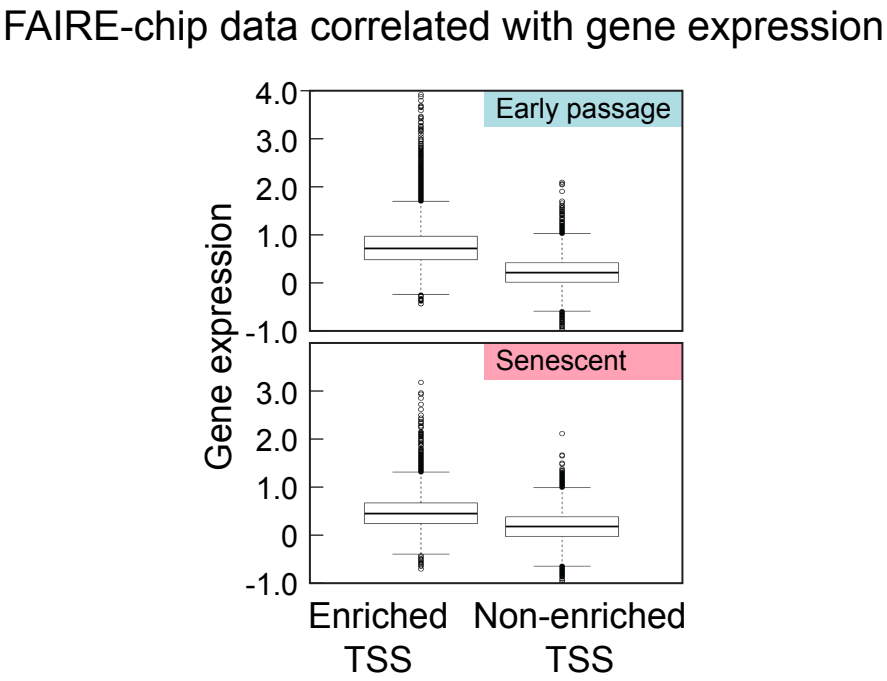

Fig. S2. FAIRE-enriched genes are expressed at higher levels than FAIRE-non-enriched genes. mRNA expression in early passage and senescent cells was assessed using Affymetrix U133 Plus 2.0 microarrays. An expression value for each gene was calculated as the median value of all the probesets that mapped to that gene. Of the 19,524 genes used in the clustering analysis (Fig. 1, panel B), 18,549 mapped to corresponding transcripts in the gene expression dataset, and 18,264 had unambiguous TSS. The expression values of these 18,264 genes that fell into FAIRE-enriched and FAIRE-non-enriched clusters were then plotted for early passage cells (upper panel) and senescent cells (lower panel). The central box delineates the 25th and 75th percentiles, and the thin lines represent the 5th and 95th percentiles. 80% of genes in the FAIRE-enriched cluster in early passage cells had a mean log2 expression value greater than the median expression value of all the genes in the FAIRE-non-enriched cluster. 75% of genes in the FAIRE-enriched cluster in senescent cells showed the same relationship. FAIRE-enriched genes had a higher overall median expression value in early passage cells relative to senescent cells. Y-axis: log2 gcRMA expression values.

Supplemental Figure 3

FAIRE enrichment of active promoters and strong enhancers defined in NHLF

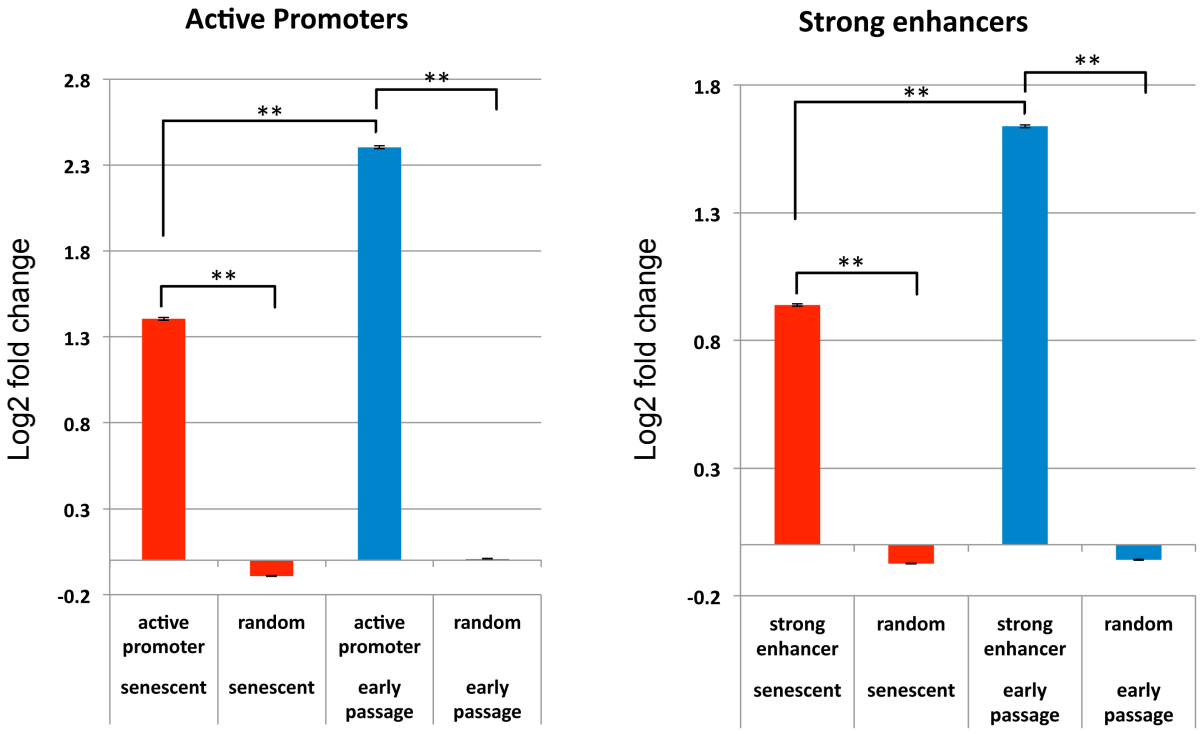

Fig. S3. Genome-wide FAIRE enrichment of regions representing active promoters and strong enhancers in early passage and senescent cells. Genomic feature coverage was computed for reads mapping to unique locations. As in Fig. 2B-D, FAIRE-seq and input sample read counts were normalized to total unique mapping reads, used to calculate log2 fold changes for senescent and early passage cells, and are represented in bar graph format. The active promoter and strong enhancer genomic feature BED files were generated from broadPeak files determined by Ernst *et al.* (2011) for normal human lung fibroblasts (NHLF), and were obtained from the ENCODE data repository. Statistical validation was performed using a bootstrap randomization of the coordinate files. Log2 (FAIRE/input) values of the original and randomized datasets were computed, and a two-tailed t-test was applied assuming unequal variances. **The results were significant in all cases ($p < 0.01$).

Supplemental Figure 4

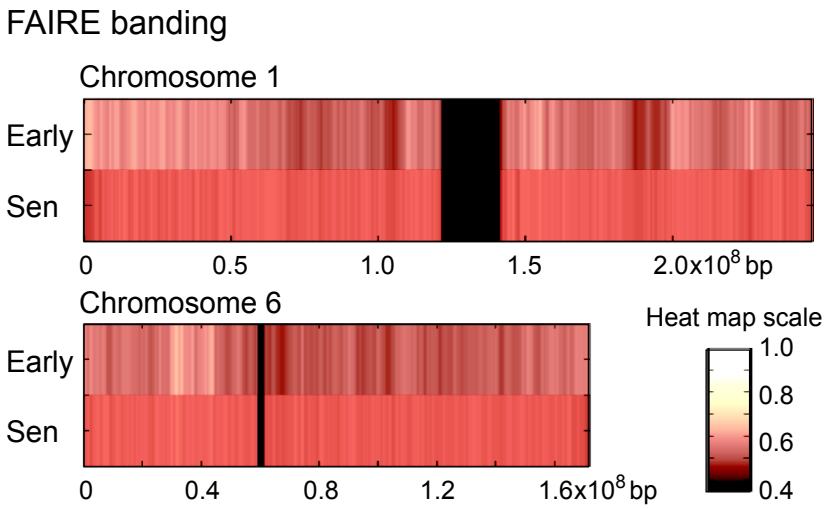

Fig. S4. The global FAIRE signal pattern becomes more homogeneous in senescent cells. FAIRE-seq enrichment data are shown represented as a heat map. Two chromosomes are shown here, and this pattern was consistent across the whole genome. Enrichment was calculated as the mean FAIRE signal of all probes within a sliding window of 1 Mbp at 100,000 bp intervals. The higher the enrichment signal, the brighter the image. Centromeres are shown as black regions and telomeres are omitted. Each chromosome shows data for both early passage (top) and senescent cells (bottom).

Supplemental Figure 5

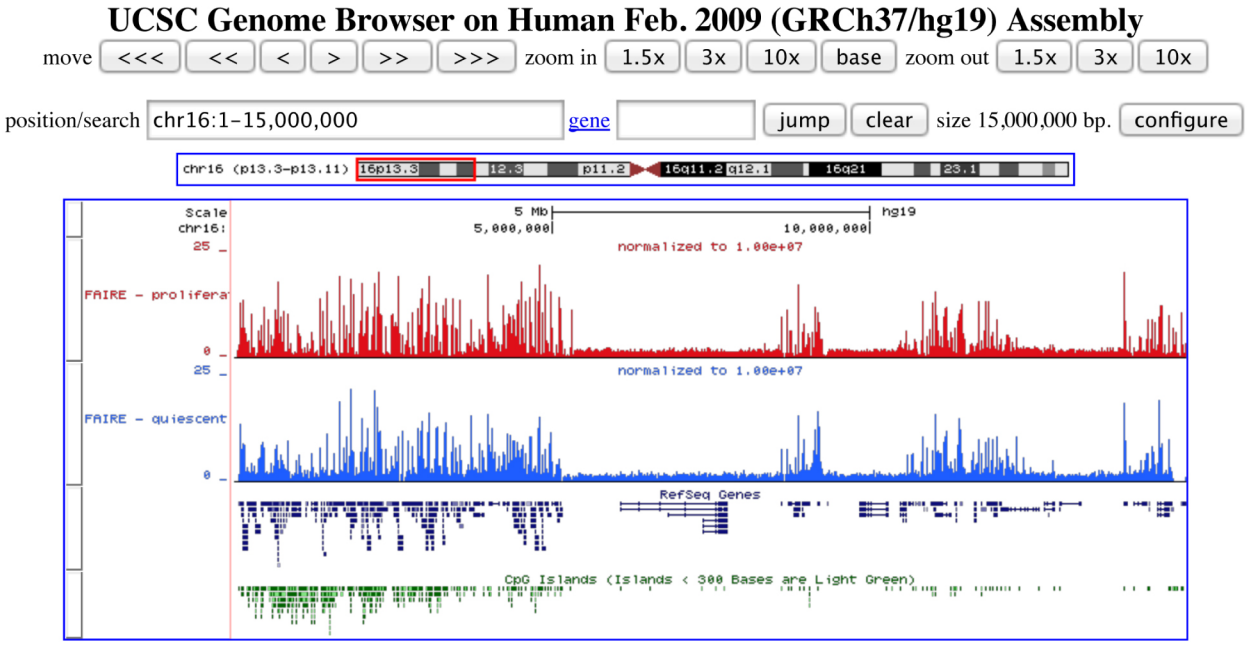

Fig. S5. Growing and quiescent cells have very similar FAIRE profiles. The chromosomal distribution of FAIRE-seq enrichment was compared between early passage proliferating and quiescent cells, and is shown as a genome browser view for illustration. Cells were made quiescent by serum deprivation. A parallel culture that was allowed to proliferate in full medium was used as the control. The same 15 Mbp region at the left end of chromosome 16 as is shown in Fig. 2A is shown here. Note the overall high similarity of the profiles, and in particular the absence of the relative increase of FAIRE signal in late replicating regions that was observed in senescent cells (Fig. 2A). Hence, this feature is not a mere consequence of cell cycle withdrawal.

Supplemental Figure 6

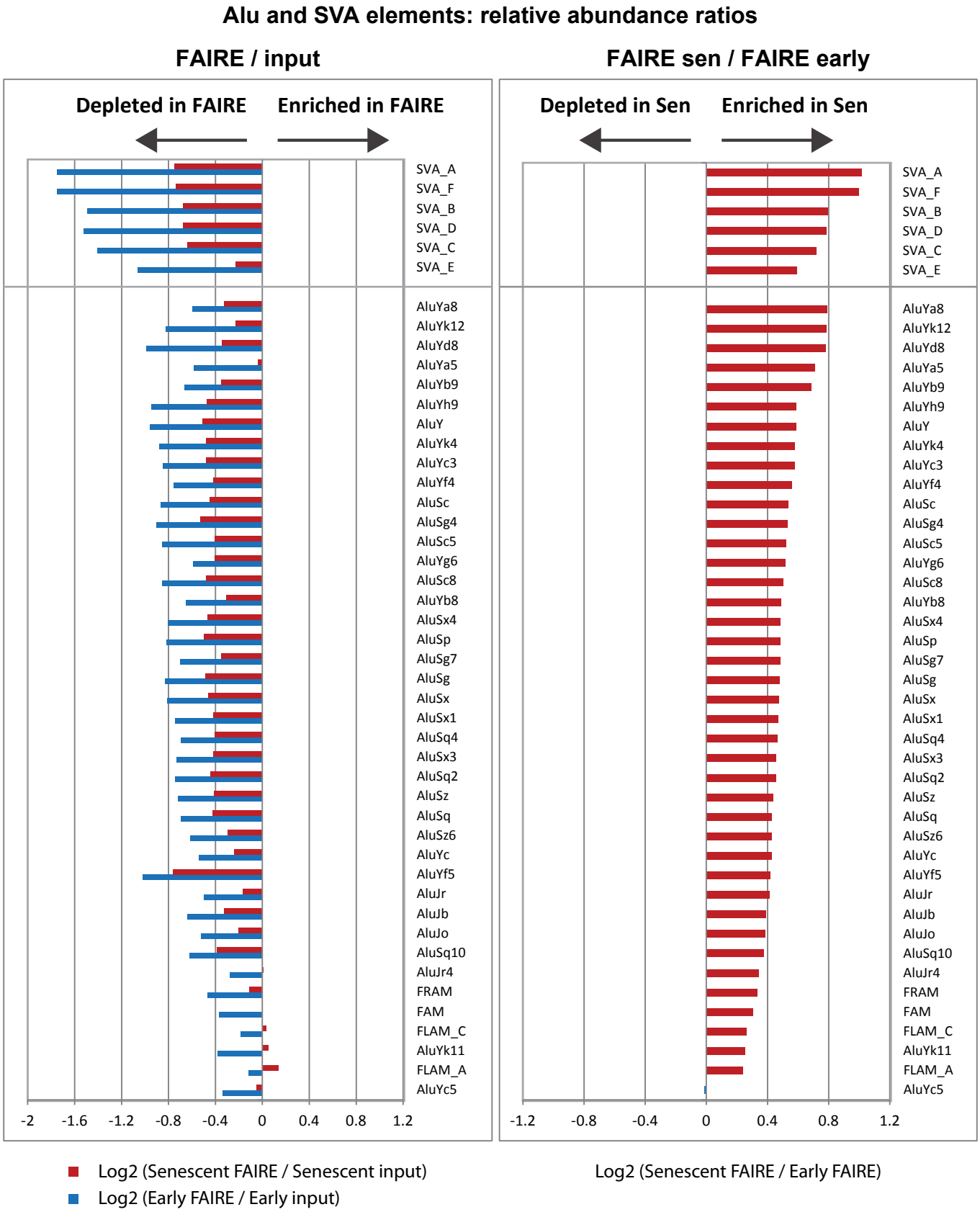

Fig. S6. See next page.

Fig. S6. Relative abundance of Alu and SVA repetitive element subfamilies. The relative abundance of RepeatMasker annotated repetitive elements was computed in the FAIRE-seq datasets using the software pipeline of Day *et al.* (*18*). Read counts were normalized to the total number of mapping reads, and fold changes were calculated as the Log2 ratios. For further details see Materials and Methods. The graphs show each repetitive element subfamily as a horizontal bar, and the log2 fold changes are shown along the X axis. The left panel shows FAIRE/input, with the Log2 (senescent FAIRE/senescent input) shown as red bars, and the Log2 (early FAIRE/early input) as blue bars. Note that elements depleted in FAIRE relative to input have negative log2 values, and hence project to the left of the Y axis. The right panel shows the FAIRE/FAIRE comparison, with positive Log2 (senescent FAIRE/early FAIRE) values shown as red bars, and negative Log2 (senescent FAIRE/early FAIRE) values shown as blue bars. Note that elements enriched in senescent FAIRE relative to early FAIRE (as is the case with all but one of the Alu subfamilies) have positive log2 values, and hence project to the right of the Y axis. The SVA elements are shown as a separate group on the top of the graph. The Alu elements are listed in descending order in the right panel, and then in the same order in the left panel. The RepeatMasker names of the subfamilies appear on the left side of each panel.

Supplemental Figure 7

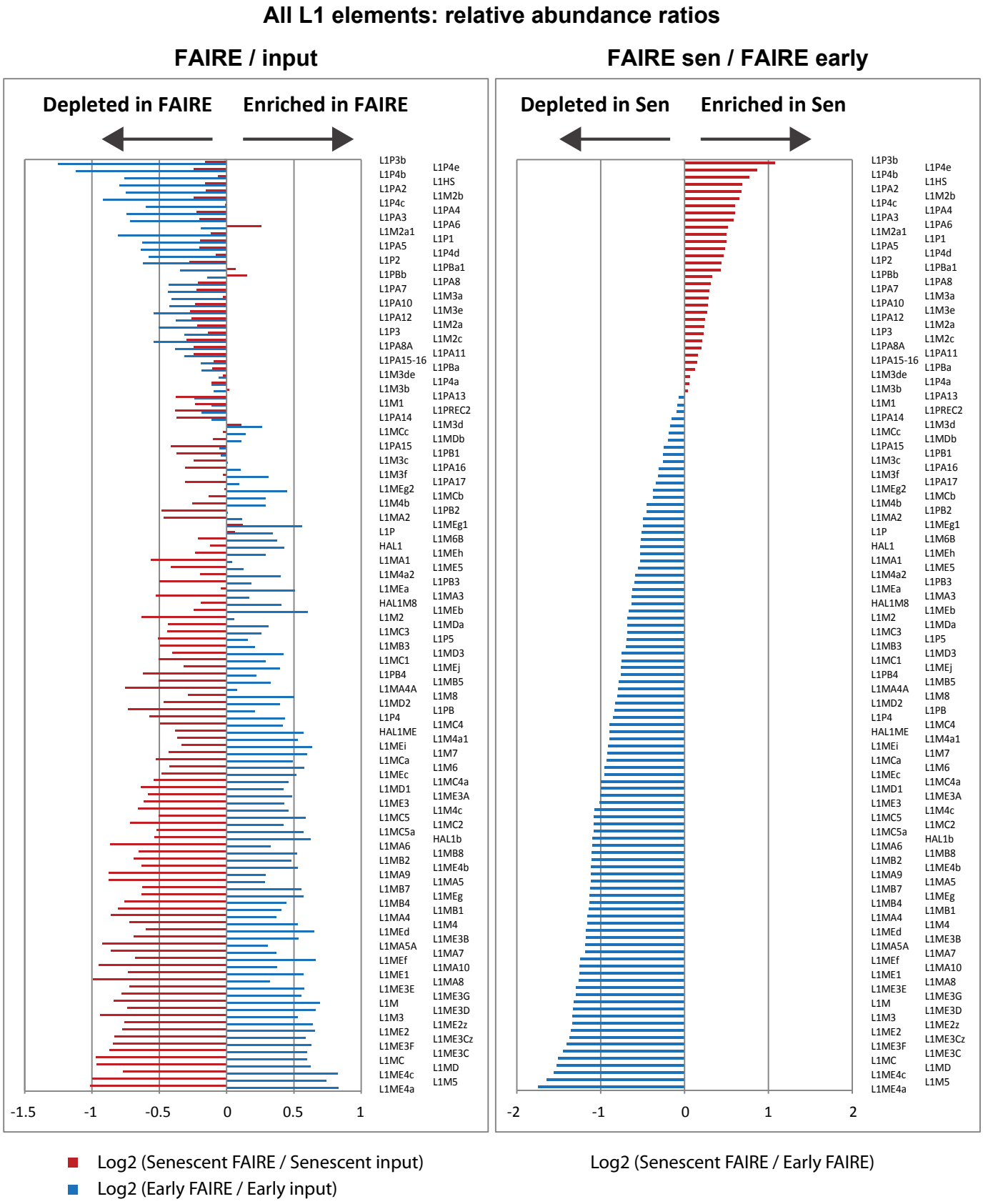

Fig. S7. Relative abundance of L1 repetitive element subfamilies. All the L1 subfamilies annotated in RepeatMasker are shown here. The graphs follow the same conventions and annotations as in Fig. S6.

Supplemental Figure 8

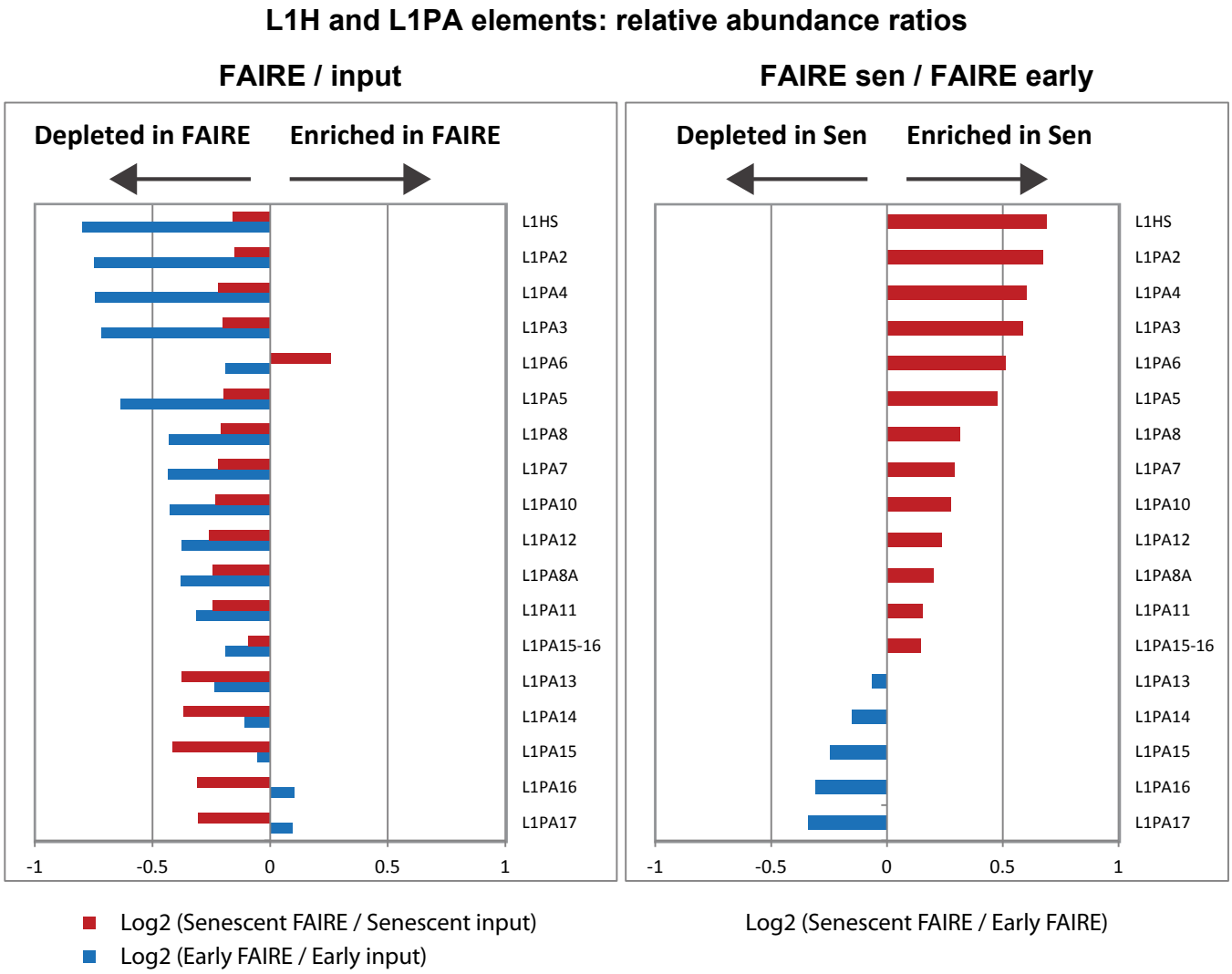

Fig. S8. Relative abundance of human-specific L1H and primate-specific L1PA subfamilies of L1 repetitive elements. The graphs follow the same conventions and annotations as in Fig. S6.

Supplemental Figure 9

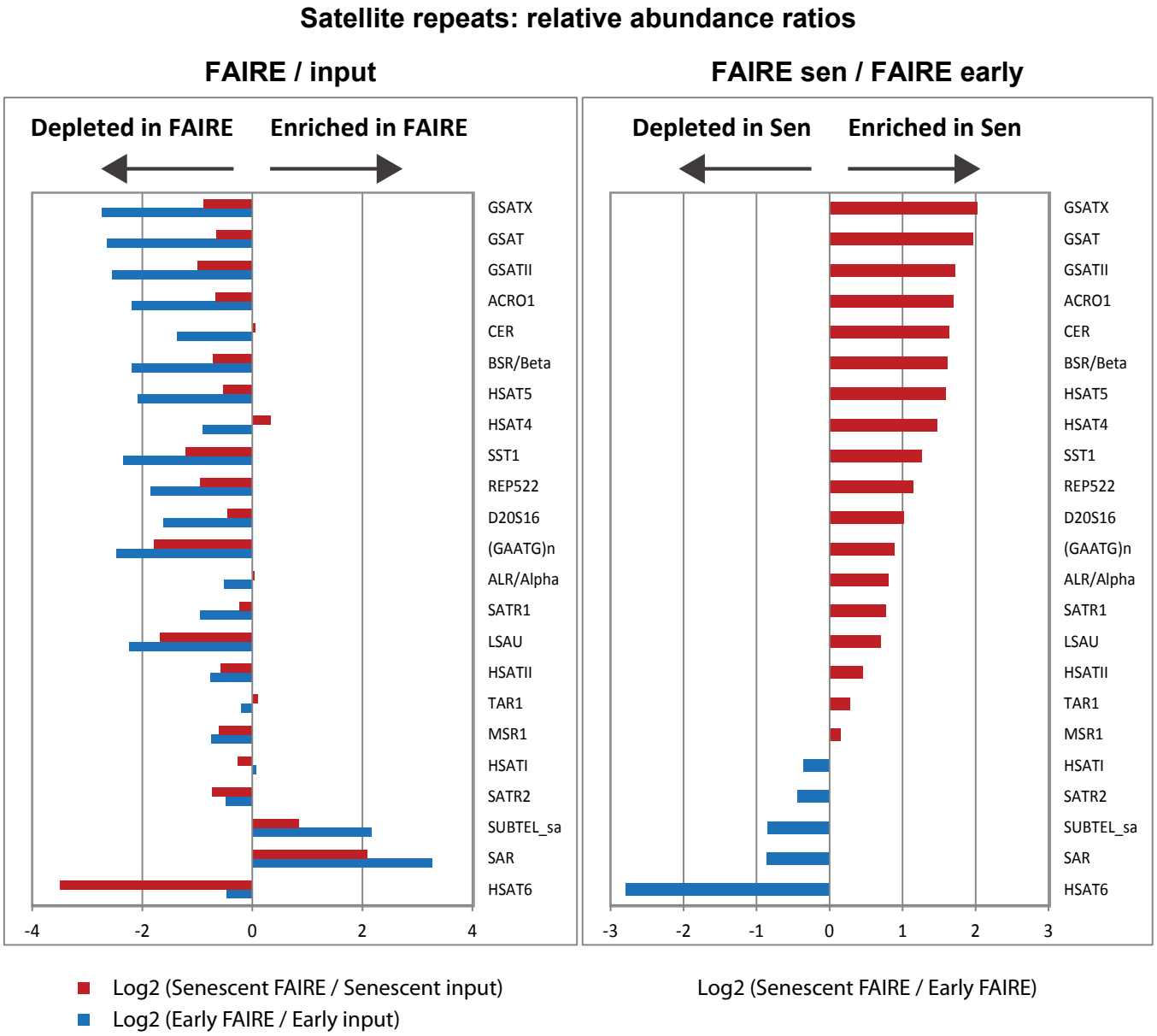

Fig. S9. Relative abundance of satellite repeat subfamilies. All the satellite subfamilies annotated in RepeatMasker are shown here. The graphs follow the same conventions and annotations as in Fig. S6.

Supplemental Figure 10

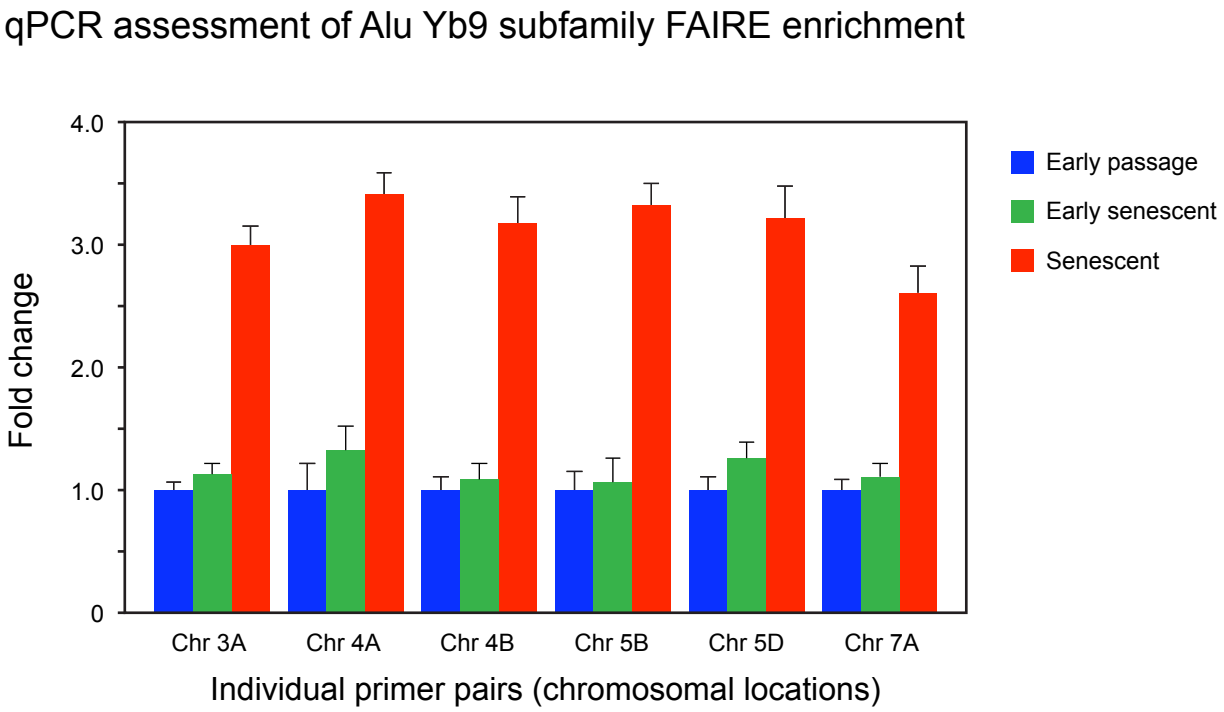

Fig. S10. qPCR quantification of FAIRE enrichment of Alu elements in senescent cells. Primer pairs were designed to detect individual members of the evolutionarily recent Yb9 subfamily of Alu repeats. Annotated sequences corresponding to this subfamily were downloaded from RepeatMasker, and those that localized to late replicating regions were used for subsequent primer design (see Materials and Methods for details). Primer pairs were subsequently empirically tested for performance: absence of primer dimers, efficiency of amplification >0.95, and kinetics of amplification that matched ($\Delta\Delta$Ct of <0.9) known single copy amplicons. Out of some 30 primer pairs tested, 6 passed these quality control criteria and are shown here. FAIRE DNA was prepared from early passage, early senescent and late senescent cells, the yields were quantified using a Qubit 2.0 dsDNA HS assay kit (Invitrogen), and 1 ng was used per qPCR reaction. qPCR was performed using the SYBR Green system (Applied Biosystems) on the ABI 7900 Fast Sequence Detection instrument. Data were normalized to controls of amplicons previously shown not to be differentially enriched in FAIRE DNA, and additionally normalized to the early passage signal for each primer pair. Experiments were performed on three separate occasions, and the data are shown as means with standard deviations (The p values for the differences between senescent and other samples were <0.01 in all cases, and were determined by the t-test. Differences between early passage and early senescent samples were not significant). Early senescent samples were harvested 2-3 weeks after proliferation ceased, and senescent cells were harvested 6-8 weeks after proliferation ceased.

Supplemental Figure 11

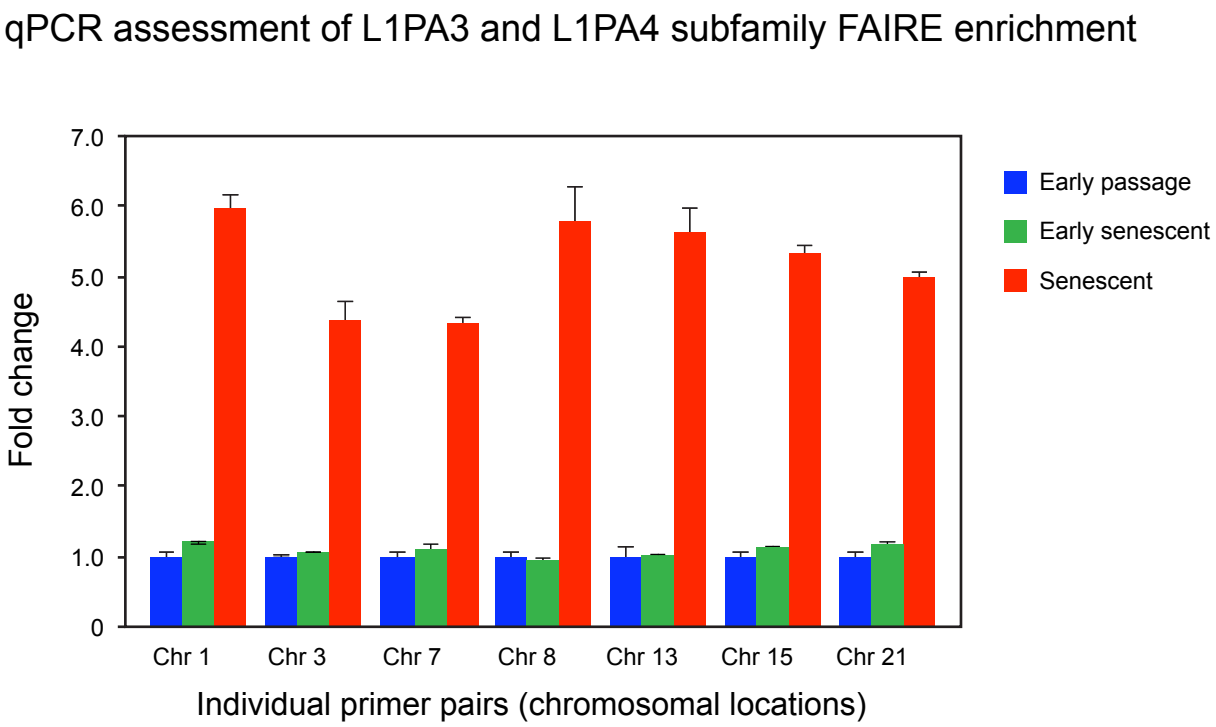

Fig. S11. qPCR quantification of FAIRE enrichment of L1 elements in senescent cells. Primer pairs were designed to detect individual members of the evolutionarily recent PA3 and PA4 subfamilies of L1 repeats. Primers were designed and reactions were performed as indicated for Alu repeats in Fig. S10. The same DNA samples were used as in Fig. S10. Experiments were performed on three separate occasions, and the data are shown as means with standard deviations. The p values for the differences between senescent and other samples were <0.01 in all cases, and were determined by the t-test. Differences between early passage and early senescent samples were not significant.

Supplemental Figure 12

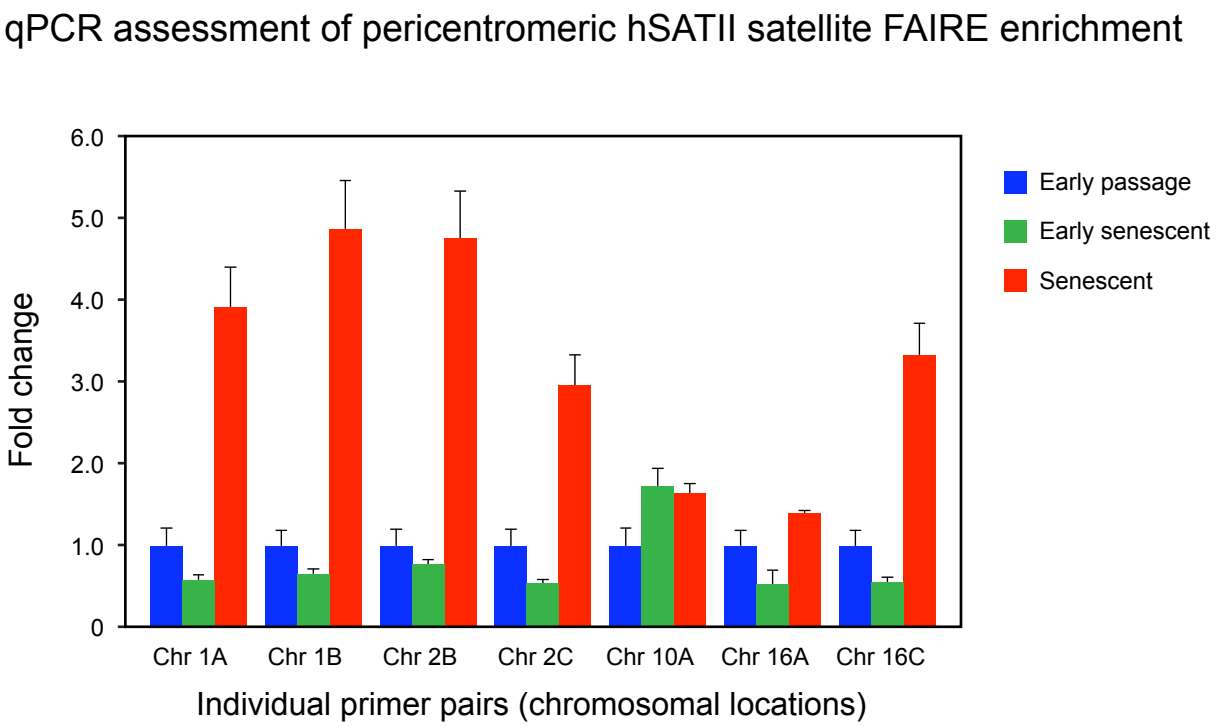

Fig. S12. qPCR quantification of FAIRE enrichment of satellite repeats in senescent cells. Primer pairs were designed to detect individual elements of the hSATII family of satellite repeats. Annotated sequences corresponding to this family were downloaded from RepeatMasker, and those that localized to pericentromeric regions were used for subsequent primer design (see Fig. S10 for details of design and testing). Primer pairs were designed to detect individual members of the hSATII family of satellite repeats on different chromosomes. Reactions were performed as indicated for Alu repeats in Fig. S10. The same DNA samples were used as in Fig. S10. Experiments were performed on three separate occasions, and the data are shown as means with standard deviations. The p values for the differences between senescent and early passage cells were <0.01 for samples 1A, 1B, 2B, 2C and 16C, <0.05 for sample 10A, and not significant for sample 16A. The somewhat reduced levels seen in early senescent versus early passage cells were significant ($p<0.05$) for samples 1A, 1B, 2C and 16C. p values were determined by the t-test.

Supplemental Figure 13

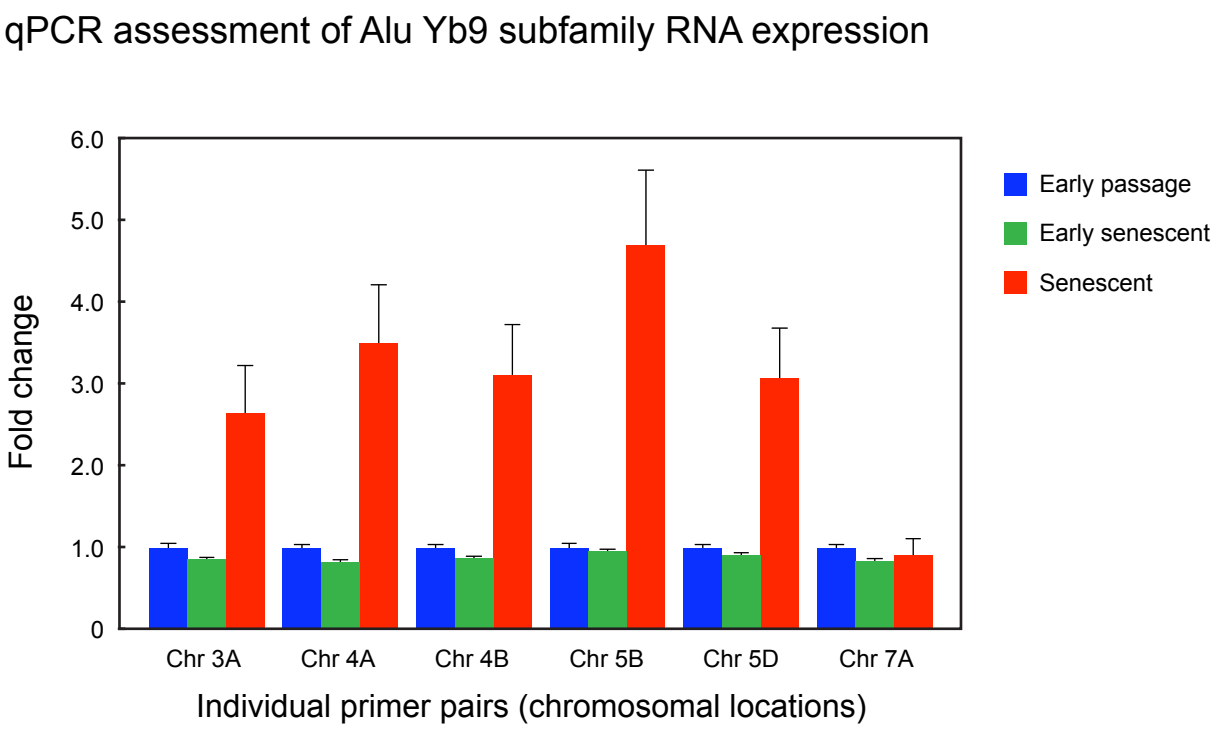

Fig. S13. qRT-PCR quantification of RNA expressed from Alu elements in senescent cells. The same primer pairs that were designed in Fig. S10 to detect individual members of the evolutionarily recent Yb9 subfamily of Alu repeats were used in qRT-PCR assays here. Total RNA from early passage, early senescent and late senescent cells was exhaustively depleted of DNA and reverse transcribed into cDNA. Effectiveness of the DNase digestion was assessed using controls that omitted reverse transcriptase. qRT-PCR was performed using the SYBR Green system (Applied Biosystems) on the ABI 7900 Fast Sequence Detection instrument. Data were normalized to GAPDH controls, and additionally normalized to the early passage signal for each primer pair. Experiments were performed on three separate occasions, and the data are shown as means with standard deviations. The p values for the differences between senescent and other samples were <0.01 in all cases, except for 7A, where none of the differences were significant. Differences between early passage and early senescent samples were not significant in any case. p values were determined by the t-test. Early senescent samples were harvested 2-3 weeks after proliferation ceased, and senescent cells were harvested 6-8 weeks after proliferation ceased. RNA was prepared from the same batches of cells as were used to prepare FAIRE DNA for Fig. S10, S11 and S12.

Supplemental Figure 14

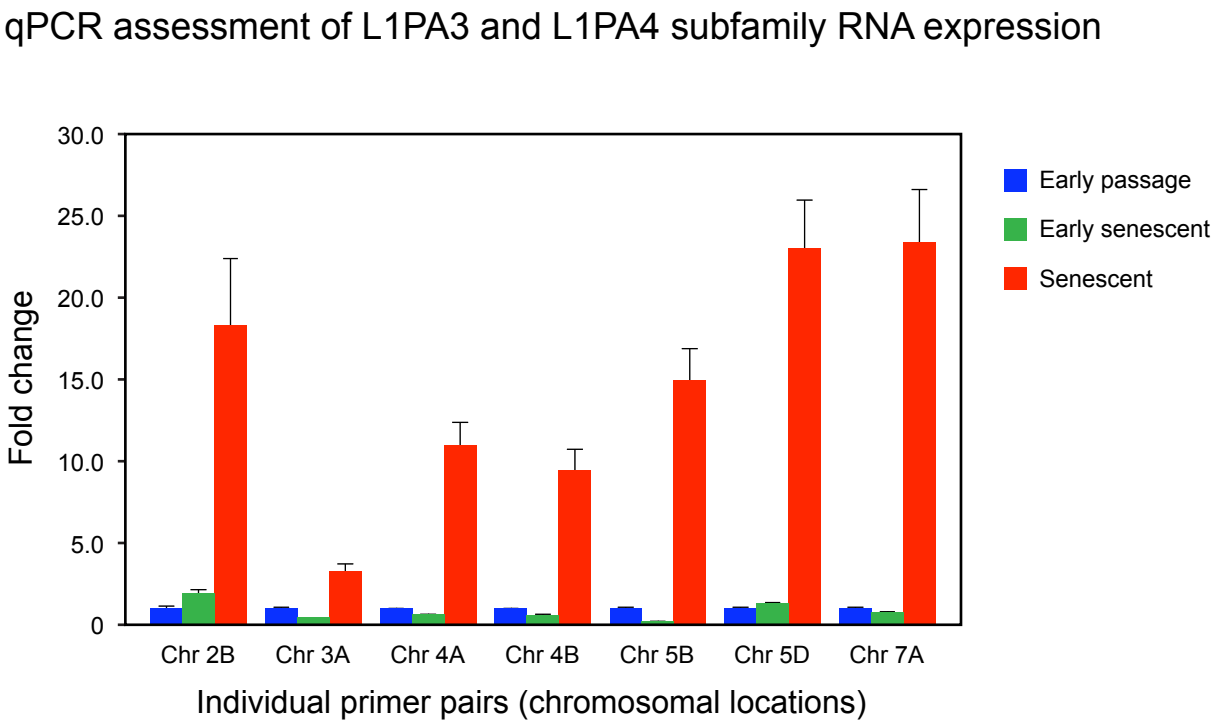

Fig. S14. qRT-PCR quantification of RNA expressed from L1 elements in senescent cells. The same primer pairs that were designed in Fig. S11 to detect individual members of the evolutionarily recent PA3 and PA4 subfamilies of L1 repeats were used in qRT-PCR assays here. RNA was prepared and reactions were performed as indicated for Alu repeats in Fig. S13 The same RNA samples were used as in Fig. S13. Experiments were performed on three separate occasions, and the data are shown as means with standard deviations. The p values for the differences between senescent and other samples were <0.01 in all cases, and were determined by the t-test. Differences between early passage and early senescent samples were not significant, except for 3A (<0.05) and 5B (<0.01).

Supplemental Figure 15

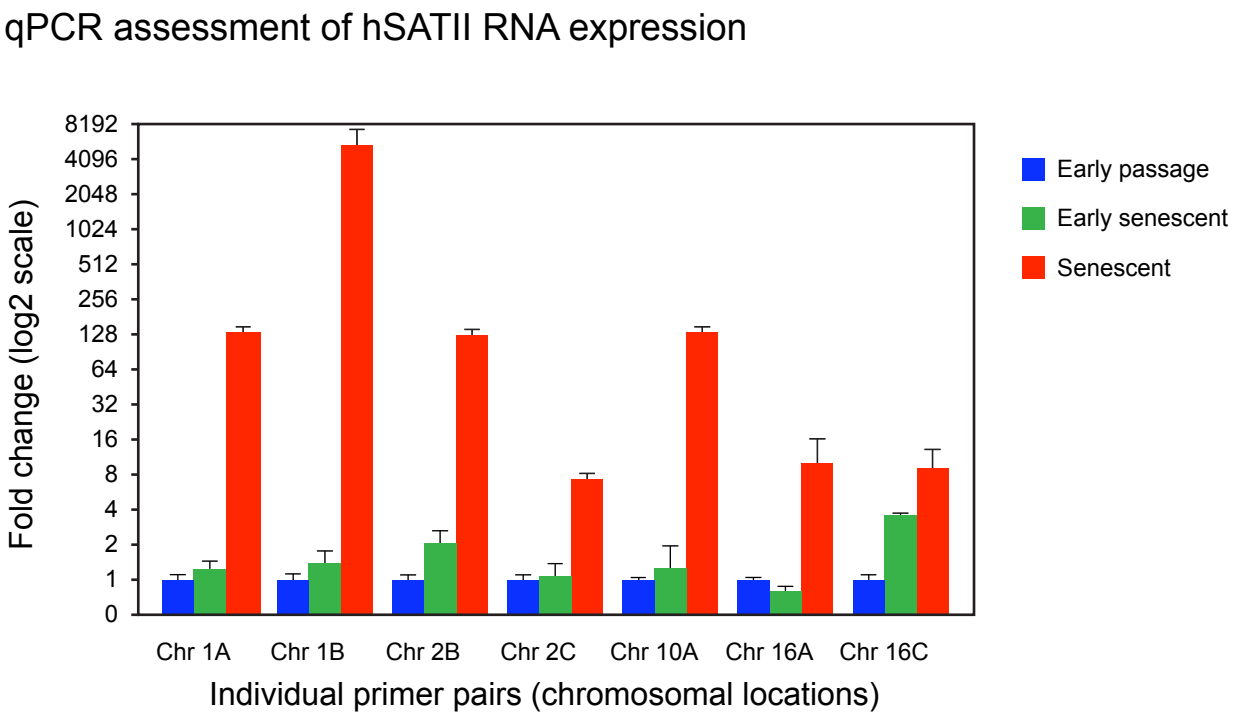

Fig. S15. qRT-PCR quantification of RNA expressed from hSATII elements in senescent cells. The same primer pairs that were designed in Fig. S12 to detect individual members of the hSATII family of satellite repeats were used in qRT-PCR assays here. RNA was prepared and reactions were performed as indicated for Alu repeats in Fig. S13. The same RNA samples were used as in Fig. S13. Experiments were performed on three separate occasions, and the data are shown as means with standard deviations. The p values for the differences between senescent and other samples were <0.01 in all cases, and were determined by the t-test. Differences between early passage and early senescent samples were not significant, except for 2B (<0.05) and 16C (<0.01).

Table S1: List of PCR primers.

Dot blotting of FAIRE DNA (Fig. 3B).

Primer name	Sequence
FAIREALUF	CTCACGCCTGTAATCCCAGC
FAIREALUR	TTGAGACGGAGTCTCGCTC
FAIREL1F	TCGCCAAGTCAATCCTAAGC
FAIREL1R	AATGCCTAGGTTTTCTTCTAGGG

Detection of Alu and L1 RNA (Fig. 3C)

Primer name	Sequence
RNAALUF	GAGGCTGAGGCAGGAGAATCG
RNAALUR	TGTCGCCCAGGCTGGAGTG
RNAL1F	CAAACACCGCATATTCTCACTCA
RNAL1R	CTTCCTGTGTCCATGTGATCTCA

Multiplex qPCR for L1 copy number (Fig. 3D)

Primer name	Sequence
ORF2 probe	AGGTGGGAATTGAAC-VIC
ORF2F	CAAACACCGCATATTCTCACTCA
ORF2R	CTTCCTGTGTCCATGTGATCTCA
5'UTR probe	TCCCAGCACGCAGC-6FAM
5'UTRF	ACAGCTTTGAAGAGAGCAGTGGTT
5'UTRR	AGTCTGCCCGTTCTCAGATCT
5S probe	AGGGTCGGGCCTGG-6FAM
5SF	CTCGTCTGATCTCGGAAGCTAAG
5SR	GCGGTCTCCCATCCAAGTAC

qPCR detection of AluYb9 (Fig. S9, S12)

Primer name	Sequence
ALYB9C3AF	GATGGAGTCTCGCTCTGTCG
ALYB9C3AR	GAATGGTGTGAACCCGGGAA
ALYB9C4AF	AGATCATGCCACTGCACTCC
ALYB9C4AR	TGCAGTCTCACTCTGTTGCC
ALYB9C4BF	AAATTAGTGGGGCGTGGTGG
ALYB9C4BR	AGGTTCACGCCATTCTCCTG
ALYB9C5BF	GCTTCCTGGGTTCACACCAT
ALYB9C5BR	AGTCCCAGCTACTAGGGAGG
ALYB9C5DF	CACTGCGCCTGGCTAATTTT
ALYB9C5DR	AGATCGAGACCATCCTGGCT
ALYB9C7AF	TCCCCAGTAGCTGGGACTAC

ALYB9C7AR	GGCTAACAAGGTGAAACCCCT

qPCR detection of L1PA3 and L1PA4 (Fig. S10, S13)

Primer name	Sequence
MDL1C1A	AAGGAATCCTGCCTCCCTCT
MDL1C1B	TTGTTTGTCTTTGCCCTGCC
MDL1C3A	TCCCAGCATGAGTGACACAG
MDL1C3B	CTGGCCCTAGTGGGATGAAC
MDL1C7A	GGCTAGTGACGGTGAGCATT
MDL1C7B	GGGCGAAGGACAAGAACAGA
MDL1C8A	CTTCCAGAGGAACGAGGCAG
MDL1C8B	GCAGAACAGTGGTGGCTGTA
MDL1C13A	GCTCTCCACCAAGCAGACTT
MDL1C13B	TTCGGAATGGGTGTGGAGTG
MDL1C15A	TGCCAGAAAGTAGGTGCAGG
MDL1C15B	TGCTTCGGCTTGTGTATGGT
MDL1C21A	CAAGGGGTCAGAAGGTTCCC
MDL1C21B	GGTGGGAGTGACCCGATTTT

qPCR detection of hSATII (Fig. S11, S14)

Primer name	Sequence
C1AHSAT2F	TGATTCCATTTGGGTACAATAGATG
C1AHSAT2R	AGTGGACTGCAATAGAAACATCA
C1BHSAT2F	AATTCAAGTCCCTTCATTGATTCC
C1BHSAT2R	CGAATGCAGTCATCATACAATGG
C2BHSAT2F	CGAGTCCATTAGAGGATTCCATTT
C2BHSAT2R	GCAGTCATCATCGAATGGTATCAAA
C2CHSAT2F	TCGATGTTGATTCCATTAGTTTCCA
C2CHSAT2R	AAATGTGATCTTCATTGAATGGACT
C10AHSAT2F	TGATTCCGTTCGATTCCATTCT
C10AHSAT2R	CGGACGAAATCATCAAGTGGG
C16AHSAT2F	CCCACTCGATGATGATTTCGT
C16AHSAT2R	TGGGATCGATCAGACTCTTCAT
C16CHSAT2F	CCTTTCGAGTCCGTTCGATGA
C16CHSAT2R	TCGAAAGGAATCATCTTCAAGTGG

Table S2: Alignment statistics of Illumina sequencing data.

Sample Name	reads mapping uniquely	% reads mapping uniquely	reads mapping to multiple locations	%reads mapping to multiple locations	reads not mapping	% reads not mapping	total sample reads
senescent FAIRE (GAIIx)	15819416	60.38	6993306	26.69	3387523	12.93	26200245
input senescent (GAIIx)	12953744	49.93	6332340	24.41	6657313	25.66	25943397
early passage FAIRE (GAIIx)	23212656	74.18	6500429	20.77	1579359	5.05	31292444
input early passage (GAIIx)	12060195	58.67	5839253	28.41	2657559	12.93	20557007
quiescent FAIRE (HiSeq2000)	23871291	75.39	5949139	18.79	1841979	5.82	31662409