Correlated alterations in genome organization, histone methylation, and DNA-lamin A/C interactions in Hutchinson-Gilford progeria syndrome

Rachel Patton McCord¹, Ashley Nazario-Toole³, Haoyue Zhang³, Peter S. Chines², Ye Zhan¹, Michael R. Erdoes², Francis S. Collins², Job Dekker¹, and Kan Cao³,⁴

¹Program in Systems Biology, Department of Biochemistry and Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA
²Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892-8004, USA
³Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA

⁴Corresponding author:
Kan Cao, Ph.D.
Department of Cell Biology and Molecular Genetics
2114 Bioscience Research Building
University of Maryland
College Park, MD 20742
Phone: 301-405-3016
Fax: 301-314-1248
e-mail: kcao@umd.edu

Running title: Epigenomic alterations in progeria cells
Methods

Chromatin Immunoprecipitation (ChIP)

Primary fibroblasts cells were grown until confluency for a week in 145 mm dishes. Protein complexes were then crosslinked by addition of formaldehyde to the culture medium to a final concentration of 1% for 10 minutes. The reaction was stopped by addition of glycine (final concentration 125µM). Fixed cells were rinsed, scraped in PBS, pelleted, and flash frozen in liquid nitrogen. Cell pellets were thawed on ice and resuspended in Lysis buffer 1 (50 mM HEPES-KOH, pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1x protease inhibitors) and rocked for 10 min at 4°C. Cells were spun down (2000 rpm for 2 min at 4°C), resuspended in Lysis buffer 2 (10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1x protease inhibitors), and pelleted by centrifugation at 4°C. Pellets were resuspended in 3 mL Lysis buffer 3 (10 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine, 1x protease inhibitors), and sonicated on ice for 12 times 20 seconds at 30% amplitude with a microtip attached to a Branson 450 sonifier. Immunoprecipitations (IPs) were performed overnight with 15-25 µg of antibody coupled to Dynal Protein G Magnetic Beads (Invitrogen). Beads were then washed with 5 times in RIPA buffer (50 mM HEPES-KOH, pH 7.55, 500 mM LiCl, 1 mM EDTA, 1.0% NP-40, 0.7% Na-Deoxycholate, 1x protease inhibitors) and once in ice-cold PBS. For ChIP-seq assays, DNA was eluted in Elution buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1.0% SDS) and prepared for sequencing as recommended by Illumina.

ChIP-seq data mapping, processing, and datasets comparison
Supplementary Information

McCord et al., 2012

ChIP’ed DNA fragments were sequenced by the Illumina Genome Analyzer 1G and aligned to the hg18 genome assembly using ELAND. As an internal control, total genomic DNA isolated from the same extracts was sequenced and processed in parallel (“Input”). At least a total of 30 million uniquely aligned reads were generated for each H3K27me3 ChIP and Input library and at least 10 million reads for each Lamin A/C ChIP and Input library (Table S2). We removed reads that mapped to satellite repeat regions as defined in hg18. To avoid PCR amplification artifacts, we applied a stringent filter to remove duplicated reads by allowing only one mapped read at any position in the genome in one direction. As a result, on average, about 5-10% of the total aligned reads were removed, and the remaining uniquely aligned and uniquely positioned reads were used for the comparisons between HGPS and normal samples.

Before comparing changes in H3K27me3 between HGPS and normal samples, we used a variety of filtering steps recommended by previous literature (Kharchenko et al. 2008; Zhang et al. 2008; Ho et al. 2011) to carefully control for possible Input biases. First, to adjust for read depth among different experiments and samples, we multiplied counts of Input reads in all windows by the ratio of the sum of all ChIP read counts to the sum of all Input reads, and then all read counts were reported as reads per 10 million ChIP reads. The total read counts in these calculations were found by taking the sum of reads in every 1 kb bin, excluding any bins that contained significantly enriched “peaks” of reads compared to the background average reads per kb (Poisson p<10^{-5}). Next, 1 kb bins with high Input variability between replicates (variance > 20) and 10 kb bins with Input read counts much higher than ChIP read counts (Zscore > 5) were filtered out to avoid results that would be strongly affected by Input outliers. Finally, the log2 ratio of the ChIP and Input read counts was calculated for each 1 kb bin. If the
input signal of a given bin was low, the average number of reads per kb over a broader window (maximum of 5 kb, 10 kb, or the whole genome average) was used in the denominator of the log ratio for that bin to prevent artificial inflation of the ratio due to poor Input coverage. The 1 kb bin log ratios were then binned and smoothed in 25 kb bins with a 5 kb smoothing step across the genome.

To determine changes of H3K27me3 in HGPS cells, we subtracted Father or Age Control ratio from the HGPS ratio for each 100kb window \[\log_2(\text{ChIP} / \text{Input})_{\text{HGPS}} - \log_2(\text{ChIP} / \text{Input})_{\text{Normal}}\]. In comparisons of H3K27me3 between HGPS and control samples, we required that at least one of the samples had a positive log ratio of IP to Input signal in order to consider a “change” in signal to be reliable.

Lamin A/C ChIP-Seq data was analyzed as above, with a few alterations. The initial filtering and read count normalization was performed in 5 kb bins to account for the somewhat lower read depth of these datasets. The final log ratios were then binned at 100 kb, because lamin associated domains have previously been shown to occur over broad genomic regions.

**H3K27me3 Comparisons with Gene Density and CpG Islands**

Gene density was defined as the number of genes in the hg18 Known Canonical gene list per megabase for each 25 kb bin. CpG island locations were obtained from the UCSC Genome Browser. All comparisons of H3K27me3 levels at high and low gene density or CGI and non-CGI locations were visualized using violin plots and quantified using a KS test. The violin plots display the median, 25%, and 75% quartiles as well as the shape of the H3K27me3 distribution for sets defined by gene density or CGI.
locations. The differences between the distributions were quantified using a KS test. The significance of the KS test statistic was then evaluated by calculating the same KS test statistic for 1000 circular permutations of the data.

**Gene expression analysis**

Gene expression was analyzed using Affymetrix GeneChip Human Genome U133 Plus 2.0 Array in duplicates. Total RNA preparation and microarray processing were performed as described (Cao et al. 2011a). To compare changes in gene expression, we calculated the log2 ratios of RMA-normalized expression between HGPS and Father and HGPS and Control. We considered the set of genes with greater than 4 fold changes in expression for further analyses. For the comparisons between H3K27me3 and gene expression, the average H3K27me3 log ratio of the 25 kb bin overlapping the gene of interest was used.

**Immunoprecipitation**

For chromatin-lamin A/C interaction studies, normal fibroblast cells were formaldehyde crosslinked and flash frozen as described above in ChIP section. Each sample was sonicated 8 times 30 seconds at 30% amplitude. Immunoprecipitations were performed overnight in the sonicated nuclear extracts with 15µg antibody coupled to Dynal beads (Invitrogen). The following antibodies were used in the immunoprecipitation experiments: H3K27me3 (Cao et al. 2011b)Cruz®, sc2027). Immunoprecipitates were washed and eluted in 100µl western sample buffer (BioRad). Proteins were separated by SDS-PAGE gels and analyzed by Western blotting assays.
Quantitative RT-PCR

Quantitative RT-PCR (qRT-PCR) was performed to measure the expression levels of EZH2, SALL1, PTGS2, PDE3A, and β-actin. All reactions were carried out at least in triplicate on CFX96 Real-Time PCR System (Bio-Rad) using SYBR Green mix (Qiagen) according to the manufacturer’s instructions. Reaction conditions were: 2 min at 50°C, 1 cycle; 10 min at 95°C, 1 cycle; 15 s at 95°C, 15 s at 58°C, 30 s at 72°C, 40 cycles. Primer sequences are described in supplementary table S3. β-actin primers were obtained from Ambion (Applied Biosystems, Inc.). To normalize for mRNA input in each reaction, the relative expression values for each gene were calculated by normalizing its mRNA level to the endogenous β-actin RNA level in each sample. Expression values for each gene were expressed as the mean of +/- the standard deviation of the experimental triplicates.

Hi-C library preparation

20 million cells from an HGPS cell line (HGADFN167) at two increasing passages (p17 and 19), as well as from two normal fibroblast cell lines at similar passages (HGFDFN168-p18 and AG08470-p20) were crosslinked in 1% formaldehyde. HGFDFN168-p18 is the father of the HGPS patient HGADFN167, and AG08470 is an age matched, unrelated child. Hi-C was performed essentially as described previously (Lieberman-Aiden et al. 2009). Cells were lysed, and chromatin was digested with HindIII. Digested ends were filled in with biotinylated dCTP and then ligated for 4 hours at 16°C. After reversing the formaldehyde crosslinks by incubation at 65°C with Proteinase K overnight and removing unligated biotinylated ends with T4 DNA polymerase, the DNA was fragmented by Covaris sonication to an average size of 200
bp and then the ideal size for Illumina sequencing (100-300 bp) was selected by Ampure fractionation. The DNA ends were repaired and ‘A’-tailed and then biotinylated junctions were pulled down using MyOne streptavidin beads. Illumina paired end adapters were ligated onto the DNA ends and then the fragments were PCR amplified for the minimum number of cycles necessary to generate 10 nM final DNA concentration.

**Hi-C data processing**

Samples were sequenced on an Illumina GAII instrument using the Paired End 75 bp module. Sequencing reads from the Hi-C experiment were mapped to the hg18 genome using Bowtie2 using the “very-sensitive” settings in an iterative procedure as follows: first, the 5’ 25 bp of each sequence was mapped, and then any reads that were unmapped or not mapped uniquely were extended to 30 bp, then 35 bp, etc. until the maximum length of the sequence was reached. This procedure aids in mapping sequences that read through a ligation junction near their 3’ end and whose full length sequence would thus be unmappable. Aligned reads were assigned to restriction fragments and filtered to discard duplicate read pairs (PCR over-amplification products) and molecules for which both ends map to the same restriction fragment. Restriction fragments shorter than 100 bp or longer than 100kb as well as those with the top 0.5% of read counts were removed. After these filtering steps, 10-20 million valid interaction pairs were obtained for each sample. Reads were assigned to genomic bins of 200 kb, according to the center of their corresponding restriction fragment. The binned interaction maps were then corrected for systematic biases by equalizing the total coverage (1D sum across the matrix) of every bin in the genome using 50 iterations.
of a normalization procedure previously described (Imakaev et al. 2012; Zhang et al. 2012). The final data was then smoothed with a 1 Mb bin size and 200 kb step size.

**Hi-C data analysis and comparison to other datasets**

Open and closed chromatin compartments were identified as previously described (Lieberman-Aiden et al. 2009). Briefly, the expected number of Hi-C reads between bins separated by each genomic distance was calculated using a loess-smoothed average over the dataset. The log ratio of observed Hi-C reads to this expected value was then calculated. The Pearson correlation between the patterns of chromosomal interactions at each pair of bins was then calculated, and this correlation matrix was used to perform Principal Components Analysis. The eigenvector of the first principal component was then plotted as the compartment assignment, with positive values corresponding to regions of high gene density (“compartment A” or “open chromatin”) and negative values corresponding to regions of low gene density (“compartment B” or “closed chromatin”). The gene density was determined by calculating the number of genes in each bin according to the UCSC Known Canonical table of human genes.

We next assessed the H3K27me3 changes and lamin A/C binding changes within regions of the genome that change from A to B or from B to A compartment between control and patient samples. “Compartment change” was defined as both a change in the sign of the eigenvector and a change of at least 0.03 in the value of the eigenvector. The distributions of H3K27me3 and lamin A/C changes from Father to HGPS were found for 200 kb windows centered on bins that changed from A to B or B to A. These distributions of changes were then compared visually in a violin plot and statistically using a KS test. The significance of the KS test statistic was then evaluated by
calculating the same KS test statistic for 1000 circular permutations of the data. The changes in H3K27me3 and lamin A/C at sites of compartment change were also visualized using an average signature plot. For this analysis, all compartment change bins were aligned and then the average H3K27me3 log ratios or lamin A/C binding log ratios for changes between control and patient were calculated at these bins as well as at each 1 Mb bin over 4 Mb upstream and downstream of these bins.

**Electron Microscopy**

Primary fibroblasts were seeded onto 35mm glass bottom dishes. The fixation includes three steps: the first step was carried out in 2% glutaraldehyde in 0.1M cacodylate buffer for 48 hours, followed by wash in 0.1M cacodylate buffer (3 X 5 min). Second, post-fix was performed in OsO₄, 1% in 0.1M cacodylate buffer for an hour, followed by wash in double distilled water (3 X 10 min). Third, en bloc fix was done in Uranyl Acetate, 2% aqueous for one hour. After the completion of fixation, the specimen were dehydrated in Ethanol gradients: 35% for 10min, 50% for 10min, 70% for 10min, for 95% for 10min, and 100% for 3 X 5 min. Infiltration was performed in EtOH: EPON Resin mixture at 1:1 ratio for one hour, then at 1:2 ratio for the second hour, at 1:3 ratio for the third hour, and in EPON Resin 100% for the fourth hour. Specimen were then embed in fresh EPON Resin and cure at 60°C. The ultramicrotomy and post-staining were done in Uranyl Acetate, 2% aqueous for 5 min and Lead citrate, 0.2-0.4% aqueous for 1.5 min. Microscopic analysis was performed on Zeiss EM10 CA.
Supplementary Figure Legends:

Figure S1: An overview of ChIP-seq analysis.

Figure S2: Reproducibility of ChIP-seq biological replicates. Normalized read counts (read counts per 10 million in dataset) in 25 kb windows are plotted for each dataset.  
(A) Father replicate 2 vs. replicate 1.  (B) HGPS replicate 2 vs. replicate 1.

Figure S3: Percent of H3K27me3 in the Input chromatin samples of HGPS fibroblasts vs. Father control fibroblasts.  (A) Western blot analysis of ChIP Input fractions with anti-Lamin A/C, anti-histone H3, and anti-H3K27me3 (Millipore).  (B) Quantification of levels of H3K27me3 in HGPS vs. Father control cells.  Input fractionations from 3 separate ChIP experiments were analyzed.  H3K27me3 levels in each sample were normalized to H3 histone levels, and HGPS was then normalized to Father (p= 0.09).

Figure S4: Relationship between Father H3K27me3 and gene density and CpG island locations.  (A) The distribution of the average 25 kb bin log ratios (IP/Input) of H3K27me3 from the two Father replicate datasets are plotted on the y-axis comparing low and high gene density (left) and CGI vs. non-CGI bins (right).  H3K27me3 is more often enriched in low gene density bins and non-CGI bins.  (B) Same as (A), except only those bins with a positive log ratio of H3K27me3 IP/Input are included in the distribution.  Among regions that have some H3K27me3 enrichment, the overall enrichment levels tend to be higher at CpG islands than non-CpG islands.  (C) The same analyses are repeated using normalized read counts (reads/10 million in dataset) instead of log(IP/Input).  (D) Normalized read counts (read counts per 10 million in
Supplementary Information

Pr Cul et a., 2012

Dataset) in 25 kb windows are plotted for H3K27me3 ChIP in Father skin fibroblasts (Father) vs. in IMR90 lung fibroblasts (R=0.22). Weak to moderate correlations were observed (R=0.52 for Age Control, data not shown). The lack of full correlation may arise from the differences in the tissues from which these fibroblasts were isolated.

Figure S5: Decreases in H3K27me3 for Age Control as compared with HGPS tend to occur in gene poor regions. Analyses in (A) and (B) were performed in the same way as for Figure 1 (C) and (D).

Figure S6: Changes between Father and HGPS H3K27me3 levels are plotted for non-CGI regions and CGI regions. H3K27me3 tends to decrease in non-CGI regions than at CGI regions.

Figure S7: Quantitative RT-PCR analysis indicated a significant decrease of the transcription of EZH2 in 4 HGPS fibroblast cell lines (HGADFN167-p13, HGADFN169-p13, HGADFN155-p15, and HGADFN164-p16, HGPS1, 2, 3, and 4, respectively), as compared to three passage-matched normal control lines (HGFDFN168-p16, HGFDFN090-p14, and AG08470-p15). Error bars show the standard deviation of 3 replicates.

Figure S8: The distributions of lamin association signal in HGPS fibroblasts are shown for regions that were previously classified as LADs or non-LADs in Tig3 human lung fibroblasts. Analysis performed as in Figure 3B.
Figure S9: (A) Scatterplot analysis comparing the H3K27me3 changes and Lamin A/C changes between Father and HGPS at 100 kb resolution. R = 0.28. (B) The significance of the relationship between H3K27me3 and Lamin changes is shown by comparing distributions of H3K27me3 changes at sites where Lamin A/C interactions increase (left) or decrease (right) in HGPS vs. Father.

Figure S10: Quantification of nuclear blebbing in HGPS-p17 and HGPS-p19 cells. Mean Negative Curvature, a quantitative measure of nuclear blebbing as described in a previous study (Driscoll et al. 2012), was used to quantify the nuclear abnormalities in HGSP p17 and p19 samples. We found there is a slight but insignificant increase of nuclear abnormality in passage 19 HGPS cells compared to passage 17 HGPS cells.

Figure S11: (A) Corrected Hi-C heatmaps showing interactions between all bins along chromosome 7. Data is binned at 1 Mb with a 200 kb step size. White = weakest interaction; red = moderate interaction; black = highest interaction. Grey lines indicate bins that were filtered out due to lack of sequence mappability or other filters (see Methods). Distinct patterns of long distance interaction are visible for Father-p18 and HGPS-p17, but only local interactions (along black diagonal) are clearly evident in HGPS-p17. (B) The same heatmap as in (A) but for the Age Control fibroblast cell line. (C) The Pearson correlation heatmap for the Age Control cell line, calculated as for Figure 4A. Compartment structure is apparent in the plaid pattern of this control.

Figure S12: Histogram of changes in eigenvector values (HGPS vs. normal). Normal refers to the consistent regions in both Father and Age Control samples. Eigenvector
values change more in HGPS-p19 relative to normal than in HGPS-p17 relative to normal.

Figure S13: Violin plots of H3K27me3 change (left) and lamin A/C change (right) from Father to HGPS in regions that change compartments from open to closed (AtoB) or closed to open (BtoA) in HGPS-p17 vs. Normal. Analysis performed as in Figure 5. Normal refers to the consistent regions in both Father and Age Control samples.

Figure S14: Average signature plots comparing compartment changes from normal to HGPS cells with H3K27me3 changes (left) or Lamin A/C binding changes (right) for HGPS-p17 (upper) or HGPS-p19 (bottom) cells. The average H3K27me3 or lamin log ratios (HGPS/Father) were calculated at each 1 Mb bin that experienced a compartment change as well as at its +/- 4 bins around this region to create an average signature.

Table S1: Reads counts summary for ChIP-seq and Hi-C analyses.

Table S2: A list of genes with correlated expression and H3K27me3 changes in between both Father and Age Control fibroblasts and HGPS fibroblasts.

Table S3: Primer sequences for RT-PCR analyses.
References


