# Project update: Telomere size and compaction

August 29, 2014

Kyle M. Douglass

## Introduction

Telomeres are nucleotide repeats located at the end of chromosomes. They help protect chromosomes from fusing together and play an essential role in cell division during chromosome duplication.

The primary goal of this project is to assess the size and compaction of human telomeres in the presence and absence of a number of telomere-assosciated proteins such as *SmchD1, LRif1* and *TRF2*. Little is known about SmchD1, but there is reason to believe that a SmchD1 knockdown will result in telomere decompaction Bibliography entry. TRF2 is a component of the shelterin complex and is also known to compact telomeric DNA and form t-loops Bibliography entry.

## Assessing telomere size and compaction

Telomeres in Hela L and Hela S cells were imaged using 3D astigmatic STORM microscopy. The telomere repeats were labeled using DNA-FISH with 18 bp probes linked to the fluorescent Cy5 marker. Typically, about fifteen fields of view were taken for each treatment, with ten thousand images recorded for each field of view. The Hela L cells contained telomeres between 17 kbp and 35 kbp long; the Hela S cells contained telomeres between between 8 and 17 kbp long. The southern blot for Hela L and Hela S is in Figure Illustration.

Figure A description...1: Southern blot from Hela S (short) and Hela L (long) cells demonstrating the difference in telomeric length between the two cell lines.

Fluorescent probe localizations from each field of view were recorded in a text file as x-, y-, and z-coordinates and drift-corrected coordinates. The drift-corrected coordinates are labeled as Xc, Yc, and Zc columns within the text files. Other information was recorded as well but not used in this analysis. Except when noted, the drift-corrected coordinates were used. Drift correction was performed in Nikon Elements and is thought to be an image autocorrelation-based correction.

The algorithm for processing the images is in Figure Illustration. Briefly, the drift-corrected coordinates of the localizations for each field of view are read from the corresponding text file. The localizations are grouped into clusters using the DBSCAN algorithm after an initial round of data-mining that optimizes the clustering algorithm parameters for the data. Once all the clusters are identified, the number of localizations, their center locations, the radius of gyration, second moments along the three coordinate axes, and volumes are computed for each cluster.

Figure A description...2: The algorithm for processing the localization data.

The radius of gyration is a common statistical measure of polymer size which is simply the sum of the variances of the locations along the x-, y, and z-directions of each monomeric subunit. For a homopolymer in solvent, the radius of gyration obeys a rigorously-proven scaling law that depends on the hydrophobicity of the solvent Bibliography entry. The volume is computed by finding the complex hull of each cluster and determining the volume inside this hull. This is a volume estimate biased towards larger volumes since a single outlier in the cluster can significantly increase the estimate.

## Results

Figure Illustration displays the histogram of the radius of gyration of each identified cluster in the untreated Hela L and Hela S cells. Hela L cells have an average radius of gyration of 106 nm, whereas Hela S cells have an average radius of gyration of 85 nm.

Figure A description...3: Distribution of the radius of gyration for Hela L and Hela S cells. The distribution reflects the radius of gyration from clusters of localizations from three different slides each of Hela L and Hela S cells. About 15 fields of view were taken for each slide. There were 2424 identified clusters for Hela L and 1405 clusters for Hela S.

In Figure 4 the histogram of the number of localizations per cluster is displayed. The average number of localizations in Hela L clusters is 191. In Hela S, the average number is 120. Both distributions are statistically different.

Figure A description...4: Histogram of the number of localizations per cluster for the Hela L and Hela S cells discussed above.

In Figure 5, the average radius of gyration is plotted for each experiment. The experiments are labeled at the top of the graph and divided by dashed vertical lines. In each experiment, the left-most set of blue and red bars is the control and are labeled as either pSuper or siGFP. The experiments from July 29 and July 30 are knockouts of the protein SmchD1 and the experiment from August 5 is a LRif1 knockout. In August 5, only Hela L cells were examined and the colors denote different transfections, not Hela L and Hela S.

Figure A description...5: Mean radius of gyration for all the datasets. Error bars represent the standard error of the mean.

We also explored the scaling between the radius of gyration and the number of localizations in the clusters. Scatter plots of the radius of gyration vs the number of localizations for the original Hela L and Hela S data sets are displayed in Figure 6. The data seems to follow a power-law model such as that predicted for the radius of gyration of a polymer vs. the number of monomer units. Three different types of fits of this data to a power lawwere performed. The robust fit routine matched the data best for all datasets as judged by eye and was robust against outliers. The fitted exponent in the original dataset was 0.31 +/- 0.01 for Hela L and 0.22 +/- 0.02 for Hela S. Errors on these values are 95% confidence intervals on the fit parameters. Scaling exponents for all the datasets are plotted in Figure 7.

Figure A description...6: Scatter plots of radius of gyration of a cluster vs. the number of localizations in the cluster for the original Hela L and Hela S datasets. Three different fits to a power law are displayed: Original, which includes all data points weighted equally, Fit with outliers excluded, which excludes points greater than 1.5 standard deviations from the curve, and Robust fit, which weighs points based on their distance from the curve.

Figure A description...7: Scaling exponents for the radius of gyration vs. the number of localizations. The solid black line denotes an exponent of 1/3, which denotes clusters that maintain a constant density of localizations with size. Error bars represent 95% confidence intervals on the fit parameters.

## Discussion

### Hela L cells contain larger telomeres on the average

In Figure 5 we can see that the mean radius of gyration of telomeres in Hela L cells is larger in every experiment that compared Hela L to Hela S (the original data and July 29 and 30). This is expected because the Hela L telomeres are longer than the Hela S telomeres. The average number of localizations measured from the telomeres is also larger for Hela L in all experiments.

### The radius of gyration appears to decrease slightly relative to the controls in most experiments but only for Hela L

Also in Figure 5 we can see that the mean radius of gyration tends to decrease in Hela L relative to the control for that experiment. The only case where it does not appear to is in the August 5 first transfection si1LRif1 dataset (the blue bar). A decrease is in this sense is noted when mean value is less than the control and the standard errors of the means do not overlap. Hypothesis testing to confirm that the means are different has not yet been done.

Hela S cells do not seem to respond to the SmchD1 knockdown. We cannot say how Hela S responds to the LRif1 knockdown since it was not tested.

The decrease in radius of gyration is unexpected because SmchD1 is thought to compact the telomeric DNA. Knocking it down might be thought to increase the average telomere size.

### The density of labels increases with radius of gyration, most notably in Hela S cells

A scaling exponent of 1/3 denotes a constant density of labeling with telomere radius of gyration. If the exponent is less than 1/3, it means that the density of labels increases with Rg, whereas if its greater, the density decreases with increasing Rg. This can be understood from the definition of density . If the density ρ is constant, the number of labeled molecules will grow as the cube of radius of the cluster like the volume does. If the number grows faster than the cube of the radius, the density of labels will increase with as the cluster radius increases.

The scaling exponent of 0.22 +/- 0.02 in the original Hela S data set reveals that the larger a Hela S telomere volume becomes, the larger the density of fluorescent labels is for increasing telomere size. The same is true for Hela L in the original dataset, but to a much smaller degree since its scaling exponent of 0.31 +/- 0.01 is very close to the constant density 0.33 scaling.

For the other datasets, it appears that Hela S labeling densities also increase with size based on Figure 7. Hela L remains close to the constant density scaling.

### We cannot conclude whether scaling exponents are different in the knockdown experiments

We cannot determine whether the exponents are significantly different from the controls in the knockdown experiments because fewer data points were collected than in the original dataset. This is reflected in the larger error bars for the July 29 and 30 and August 5 experiments. The error bars are the 95% confidence intervals on the fitted exponent parameter.

### Data from the TRF2 experiments were more noisy and not filtered in the same way. They are not compared here.

The data from the TRF2 experiments was not included in this report since it was much more noisy and an analysis on the filtered data was not performed at this time. The analyses were performed on the unfiltered data but the results from the same experiment repeated on two different days contradict one another.

We're currently waiting to analyze the filtered data to see if the results change.

## Conclusion

It appears that SmchD1 and LRif1 have a very slight effect in controlling the average size of the L telomeres, but the decrease in the telomeric size is unexpected. However, we have not yet tested for significance in the differences in average size relative to the controls, but will do so soon. Since both SmchD1 and LRif1 are related to telomeric packaging, it was expected that their knockdowns would result in an increase in telomere size.

We also learned that the labeling density of fluorophores on the Hela S telomeres increased with the the size of the telomere, making it difficult to assess compaction by comparing the scaling of radius of gyration with the number of detected labels. The labeling density also increases with size in Hela L, but to much less of an extent. The reason for the nonconstantlabeling density is not currently known.

Unfortunately, the TRF2 knockdown data was not conclusive due to its noisiness and because the unfiltered data, which did not include drift-correction, was analyzed. The drift-corrected data will be analyzed soon to determine whether we can see an increase in telomere size when TRF2 is absent from the shelterin complex. Additional experiments imaging more TRF2 knockdown cells are currently being performed. Should we see an increase in average telomere size in the TRF2 knockdowns, we'd have a stronger degree of confidence that we can indeed measure changes in telomeric size due to biological perturbations of essential telomere packing proteins.

Bibliography

SMCHD1: Ryu-Suke Nozawa, Koji Nagao, Ken-Taro Igami, Sachiko Shibata, Natsuko Shirai, Naohito Nozaki, Takashi Sado, Hiroshi Kimura, Chikashi Obuse, Human inactive X chromosome is compacted through aPRC2-independent SMCHD1-HBiX1 pathway, 2013

TLOOP: Ylli Doksani, John Y. Wu, Titia de Lange, and Xiaowei Zhuang, Super-Resolution Fluorescence Imaging of Telomeres Reveals TRF2-Dependent T-loop Formation, 2013

POLSCALE: Liu Hong, Jinzhi Lei, Scaling law for the radius of gyration of proteins and its dependence on hydrophobicity, 2009