Towards Understanding Functional Nuclear Alterations with Cellular Senescence

By

Audrey Sara Dalgarno

B.S., Biology, University of Wisconsin – Madison, 2017

Thesis

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Division of Biology of Medicine at Brown University

PROVIDENCE, RHODE ISLAND
May 2024

This dissertation by Audrey Sara Dalgarno is accepted in its present form by the Division of Biology and Medicine as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date _____ _____

Nicola Neretti, Ph.D.,
Advisor

Recommended to the Graduate Council

Date_____ _____

Erica Larschan, Ph.D.

Date _____ _____

John Sedivy, Ph.D.

Date _____ _____

Michelle Dawson, Ph.D.

Date _____ _____

Jian Ma, Ph.D.

Approved by the Graduate Council

Date_____ _____

Thomas Lewis, Ph.D.,
Dean of the Graduate School

Curriculum Vitae: Audrey Dalgarno

1. Name, position, department

Audrey Sara Dalgarno
PhD Candidate
Department of Molecular Biology, Cell Biology, and Biochemistry
Brown University, Providence, RI

2. Office Address

Department of Molecular Biology, Cell Biology, and Biochemistry
Laboratories for Molecular Medicine
Brown University
70 Ship Street
Providence, RI 02903
Work email: audrey_dalgarno@brown.edu

3. Education

Graduate: Brown University, Providence, RI
Expected 2024 PhD, Molecular Biology, Cell Biology, and Biochemistry
Advisor: Nicola Neretti, PhD

Undergraduate: University of Wisconsin, Madison, WI
2017 BS Biology, French Minor, Graduate with distinction

4. Professional Appointments

2017-2019 Dana-Farber Cancer Institute, Boston, MA
Research Technician, Center for Cancer Genome Discovery

5. Completed Publications and Invited Talks

My work as a Graduate Student in the Neretti Lab:

<u>Publications</u>

1) **Dalgarno, A.,** Evans, S.A., Kelsey, M., Clark, K., Sedivy, J., and Neretti, N. A kilobase resolution analysis of the senescent genome reveals novel architectural features supporting cell cycle arrest and inflammation. (In preparation).

2) Soto, C., Bryner, D., **Dalgarno, A**., Neretti, N. & Srivastava, A. TADBay: A Bayesian Construction of Topologically Associated Domains. in 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 233–240 (2022).

3) Rocha, A.*, **Dalgarno, A.*** & Neretti, N. The functional impact of nuclear reorganization in cellular senescence. Brief. Funct. Genomics 21, 24–34 (2022).

4) Soto, C.*, **Dalgarno, A.*** et al. Representation of Chromosome Conformations Using a Shape Alphabet Across Modeling Methods. in 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 151–156 (2021).

***Authors contributed equally**

Invited Lectures/Talks

1) Dalgarno, A. et al. "A Hi-C map of the senescent genome at kilobase resolution reveals novel associations between architectural features and senescence-associated gene expression", CGI Group Meeting, CEGS, 11/18/22.

2) Dalgarno, A. et al. "Shape Alphabets: A Novel Representation of Variability in 3D Chromosome Structures", NOFIC-AICS Group Meeting, 4DNucleome, 07/10/20.

Abstracts/Conferences Attended

1) Dalgarno, A. et al. "A Hi-C map of the senescent genome at kilobase resolution reveals novel associations between architectural features and senescence-associated gene expression", 4DNucleome Annual Meeting, Boston, MA, 12/9/23.

2) Dalgarno, A. et al. "A Hi-C map of the senescent genome at kilobase resolution reveals novel associations between architectural features and senescence-associated gene expression", Center for Physical Genomics, Northwestern, 4/21/23.

3) Dalgarno, A. et al. "A Hi-C map of the senescent genome at kilobase resolution reveals novel associations between architectural features and senescence-associated gene expression", MCB Data Club, Brown University, 10/21/22.

4) Dalgarno, A. et al. "A Hi-C map of the senescent genome at kilobase resolution reveals novel associations between architectural features and senescence-associated gene expression", Mechanisms of Aging, CSHL, 9/27/22 – 10/1/22.

5) Dalgarno, A. et al. "A Hi-C map of the senescent genome at kilobase resolution

reveals novel associations between architectural features and senescence-associated gene expression", MCB Retreat, Brown University, 9/2/22.

6) Dalgarno, A. et al. "Investigating Transcriptional Signatures of Cellular Senescence", MCB Data Club, Brown University, 1/14/22.

7) Dalgarno, A. et al. "A Novel Shape Alphabet to Represent the 3D Genome", CEGS Annual Meeting, 10/13/21.

8) Dalgarno, A. et al. "Investigating the role of genomic shape in cellular senescence", MCB Retreat, Brown University, 9/3/21.

9) Dalgarno, A. et al. "Characterizing the variability of chromosome 3D structures via Shape Alphabets", Epigenetics and Chromatin, CSHL, 9/15/20 – 9/18/20.

10) Dalgarno, A. et al. "Shape Alphabets: a novel representation of variability in 3D chromosome structures", MCB Retreat, Brown University, 9/4/20.

11) Attended Genome Organization and Nuclear Function, CSHL, 5/20.

My work as a Research Technician at Dana-Farber Cancer Institute

Publications

1) Chung, T. K. H. et al. Dissection of PIK3CA Aberration for Cervical Adenocarcinoma Outcomes. Cancers 13, (2021).

2) Sewastianik, T. et al. Human MYD88L265P is insufficient by itself to drive neoplastic transformation in mature mouse B cells. Blood Adv 3, 3360–3374 (2019).

Abstracts/Conferences Attended

1) Attended NEBNext Single Cell RNA Workshop, Cambridge, MA, 11/18.

2) Attended CCGD Symposium: Genomic Approaches towards Precision Cancer Medicine, Boston, MA, 3/18.

My work as an Undergraduate in the Ané lab:

Abstracts/Conferences Attended

1) Dalgarno, A. et al. "MBOA had no effect on *Rhizobium* Autoaggregation" University of Wisconsin Undergraduate Research Symposium, Madison, WI, 5/13.

6. Research Grants

NIH 2020 T32 AG
7/1/21 - 8/31/23
"Predoctoral Training in the Molecular Biology of Aging"
PI: John Sedivy
Role: Trainee

NIH 2020 T32 GM
9/1/20 – 6/31/21
"Predoctoral Training Program in Biological Data Science at Brown University"
PI's: Sohini Ramachandran, Bjorn Sandstede, Eliezer Upfal, Jean Zhijin
Role: Trainee

NIH 2019 T32 GM
9/1/19 – 8/31/20
"Training in Molecular and Cell Biology and Biochemistry"
PI: Kim Mowry
Role: Trainee

7. Service

Service to the University, Brown University

2023 New Student Mentor
2023 Admissions Committee Member
2022 Admissions Committee Member

Service to the University, University of Wisconsin

2017 Discovery Center Volunteer
2016-17 Badgers Give Back Volunteer
2016 Biocore Board of Directors

8. Honors

Academic

2023 Blavatnik Family Graduate Fellowship
2022 Poster Winner, MCB Retreat
2020 Oliver Cromwell Gorton Arnold Biological Fellowship
2017 Graduate with Distinction
2017 Biology Core Curriculum Honors
2016 Germaine Mercier Scholarship

Academic/Athletic

2017 Sue Ela Award
2017 Student Athlete of the Month
2016 IRA National Champion, Varsity Ltwt 4+
2016-17 Athletic Board Scholar
2016-17 CRCA Scholar-Athlete
2016-17 IRA All-Academic Team
2015,17 Big Ten Distinguished Scholar
2015-17 Academic All-Big Ten
2014 Chi Alpha Sigma

Honorary Societies

2017 Phi Beta Kappa

9. Teaching

Brown University

BIOL0470 (Fall 2020): Genetics
Developed Computational Biology Lab Curriculum*
Teaching Assistant

**Due to COVID-19 (novel coronavirus pandemic) this course was offered virtually. I developed new lab content consisting of computational biology modules with online tools because we were unable to conduct regular labs.*

University of Wisconsin

Biocore 382 (Spring 2017): Cellular Biology Lab
Undergraduate teaching assistant

Biocore 486 (Fall 2016): Organismal Biology Lab
Undergraduate teaching assistant

10. Advisees

Undergraduates

Kikuyo Shaw (2021-2022)
Camille Leung (2021)
Carolyn Lai (2020-2021)

11. Date

May 2024

Acknowledgements

Wow – it truly takes a village. As I stand on the other side of this lengthy and arduous degree, I am immensely grateful for the army who got me here.

First and foremost, I would like to thank my family and partner for their unwavering support through the ups and downs of my PhD. To Ethan – my partner and my rock. I have no words. Thank you for making it through with me! I'm sure you will get your own fun hat soon. To my father, David – you are my original inspiration, and I will be forever influenced by your profound love for science and all the good it can do in this world. To my mother, Brenda – thank you for being there for me, always, and for expressing your pride in my accomplishments. To my siblings Robert, Philip, and Carol – thank you for filling my breaks with fun and distractions. Especially Carol, thank you for serving as an additional lab member. It has been such a surprise we have ended up working in such parallel fields and I'm excited for you to start your own PhD journey soon! To my grandparents, Herb and Carol – it has been so much fun to share my research accomplishments with you. I hope that my research helps you to continue to age gracefully. To my new family of Dimmocks – I am grateful for your enthusiasm and support. The escapes of your house, dogs, and the beach have been so important throughout my degree. To Nessie – (although she can't read because she is a dog) thank you for standing by my side on your tiny little legs throughout the end of my degree with me. You have made my life full of adventure.

Of course, I must acknowledge my lab! Being in the Neretti lab truly extends past coworkers and into friendship. You all made it fun to come to work every day (well…most days haha). To my advisor, Nicola – I am always blown away by your creativity and passion for your research. I have never met someone so eager to discuss their research and where it can go. I am glad to have learned from you over the past years. Your expertise has transformed my thinking

from that of a rigid genomics core technician into that of a budding computational biologist, which is exactly what I wanted from this experience. Thank you. To my fellow lab members, Azu, Jeremy, Miiko, Nick, Anthony, Sam, Sam (!), Andrew, Aiya, and Fifi – thank you for filling my days and nights with hilarious conversations, unexpected youtube videos of pumas, and delicious and overly complicated meals. You are all incredibly smart and creative, and I am thankful for your input into all my projects. I will miss you all!

Thank you as well to the MCB graduate program. It is rare to find a program who cares about their students as much as you do. Thank you to Mark, Greg, and Alison for your leadership throughout my time in the program. Thank you as well to my committee – Erica, John, Michelle, Jian, and formerly Sohini. I have enjoyed learning from you during and outside of my committee meetings. Your support has been invaluable to moving my projects forward. Thank you as well to Christoph, for your patience and timeliness, and for all your help with my samples. Furthermore, I am also thankful to all the collaborators I have worked with who have extended my knowledge base beyond the scope of Brown's research.

I feel I would also be remiss not to acknowledge the unusual climate in which I completed my degree. Although the first months of my PhD proceeded 'normally', in March of my first year the world shut down due to the COVID-19 pandemic. The next years were shrouded by the virus and this was an incredibly isolating time where it could be undeniably difficult to be a graduate student. I joined the Neretti lab after a virtual rotation, worked virtually for an entire year, and still did not see the lower half of my colleagues faces for just over half a year after that. I am especially grateful for the support and flexibility of my family, my lab, and the program during this time. Furthermore, I am proud of myself, my cohort, and the MCB students for all we accomplished during this experience, which added a level of difficulty to an already demanding degree.

Table of Contents

Manuscripts and Publications Included

Chapter 1 (Introduction) includes sections or portions of sections written by Audrey Dalgarno that were published in (Rocha, Dalgarno, & Neretti, 2022). Chapter 2 contains material from an unpublished manuscript authored by Audrey Dalgarno and colleagues. Chapters 3 and 4 contain material from two published conference papers (C. Soto, Bryner, Dalgarno, Neretti, & Srivastava, 2022; C. Soto, Dalgarno, et al., 2021) on which Audrey Dalgarno is an author.

Rocha, A.*, **Dalgarno, A.***, & Neretti, N. (2022). The functional impact of nuclear reorganization in cellular senescence. *Briefings in Functional Genomics*, *21*(1), 24–34.

Soto, C., Bryner, D., **Dalgarno, A.**, Neretti, N., & Srivastava, A. (2022). TADBay: A Bayesian Construction of Topologically Associated Domains. *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 233–240.

Soto, C.*, **Dalgarno, A.***, Bryner, D., McLaughlin, B., Neretti, N., & Srivastava, A. (2021). Representation of Chromosome Conformations Using a Shape Alphabet Across Modeling Methods. *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 151–156.

***indicates co-first authorship**

List of Tables

List of Illustrations

Chapter 1: Introduction

Introduction sections 1.4.1-1.4.3. and portions of section 1.5 were published in (Rocha et al., 2022). *Audrey Dalgarno was responsible for the research, writing, and figures for these sections. All authors were responsible for editing these sections.*

1.1: What is aging?

Aging is the one thing that no person has survived. While humans have developed innovative ways to stave off diseases such as cancer and diabetes, aging is the one ailment which is considered inevitable. Why is this? What *is* aging?

Aging can be defined as, "the time-related deterioration of the physiological functions necessary for survival and fertility (Gilbert, 2000)." One of the first contexts in which aging was considered was the context of evolution (Kirkwood, 1977; Medawar, 1952; Zainabadi, 2018). The evolutionary cycle of life and natural selection involves survival, reproduction, and the subsequent passing on of hereditary information which has conferred survival advantage. There is, therefore, less evolutionary necessity to maintain organisms after they have passed their reproductive age. The negative aspects of aging we see are thus a result of the declining impact of natural selection. In this context, aging is considered unavoidable.

However, evidence emerged in the late 20th century which challenged this belief (Kaeberlein, McVey, & Guarente, 2001). Studies pointing to the similarities between mortality curves across diverse organisms suggested that while aging itself may not be regulated, the pace of aging could be influenced (Zainabadi, 2018). For example, there is variation between lifespans among related species and even within the same species, shedding light on the role of epigenetics in aging. Since these observations, the study of aging has gained significant traction, leading to an improved, but still incomplete, understanding of how and why we age. In the next section I share a broad overview of what is known about the molecular mechanisms of aging.

1.2: The Hallmarks of Aging

Following the monumental turn of the century paper, "The Hallmarks of Cancer," the aging field has established its own Hallmarks of Aging (Hanahan & Weinberg, 2000; López-Otín, Blasco, Partridge, Serrano, & Kroemer, 2013, 2023). Originally starting with nine hallmarks, this list has recently been expanded to twelve hallmarks, which can be categorized as primary, antagonistic, and integrative. The primary hallmarks are hallmarks that entail damage to the genome itself. In contrast, antagonistic hallmarks arise in response to this damage. Integrative hallmarks occur when the damage caused by the primary and antagonistic hallmarks becomes unsurmountable. In this context a hallmark is defined as: (1) being associated with age, and (2) having the capacity to accelerate or decelerate/stop/reverse aging with experimental accentuation or therapeutic intervention, respectively. The primary hallmarks include genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, and disabled macroautophagy. The antagonistic hallmarks include deregulated nutrient sensing, mitochondrial dysfunction, and cellular senescence. The integrative hallmarks include dysbiosis, chronic inflammation, altered intercellular communication, and stem cell exhaustion. Of particular interest to this thesis is the antagonistic hallmark cellular senescence, which is driven by the primary hallmark of epigenetic alterations, and in turn fuels the integrative damaging hallmarks, namely chronic inflammation.

1.3: Cellular Senescence

Cellular senescence, first characterized by Hayflick and Moorehead in 1961, is an irreversible state of proliferative arrest, which can have detrimental pro-aging effects (Di Micco, Krizhanovsky, Baker, & d'Adda di Fagagna, 2021; Hayflick & Moorhead, 1961).

1.3.1: Different Types of Cellular Senescence

Senescence can be induced in different ways, but stems from significant DNA damage, leading to cell cycle arrest (Di Micco et al., 2021). The first type of senescence discovered was replicative senescence (RS), wherein the progressive shortening of telomeres due to the serial passaging of cells leads to a DNA damage response (Hayflick & Moorhead, 1961). There is also oncogene-induced senescence (OIS), which is caused by DNA damage due to the collapse of replication forks when the cell enters an oncogene-induced hyperproliferative state (Serrano, Lin, McCurrach, Beach, & Lowe, 1997). Additionally, there is DNA-damage induced senescence (DDIS), which can be caused by various DNA-damaging agents such as chemicals (e.g. etoposide) (Leontieva & Blagosklonny, 2010). These forms of primary senescence can also induce senescence on the surrounding healthy cells through either paracrine (Acosta et al., 2013) or juxtacrine signaling (Hoare et al., 2016; Parry et al., 2018; Teo et al., 2019). Senescence is a dynamic process, as its phenotype changes significantly over time (van Deursen, 2014). The transition from early to late senescence involves changes such as lamin B1 downregulation, chromatin remodeling, and retrotransposition, among others.

1.3.2: Biomarkers of Senescence

Although the biomarkers of senescence are hotly debated (Gorgoulis et al., 2019), several have general support from the field. First, there is senescence-associated beta-galactosidase. This enzyme accumulates and increases in activity in the enlarged lysosomes of senescent cells and can demarcate them in colorimetric x-gal assays (Dimri et al., 1995). Lipofuscin, an aggregate of lipids, proteins, and oligosaccharides, also accumulates in the lysosomes of senescent cells and can therefore serve a similar role to senescence-associated beta-galactosidase (Salmonowicz & Passos, 2017). The cyclin-dependent kinase inhibitors CDKN2A (p16) and CDKN1A (p21) are also considered biomarkers for senescence (Gorgoulis et al., 2019). CDKN2A is generally thought

to be more specific of a senescence marker as compared to CDKN1A, and marks a later and more stable state of senescence (De Cecco et al., 2019; Ito, Teo, Evans, Neretti, & Sedivy, 2018). Other markers of senescence include markers of DNA damage, such as p53BP1 and gamma-H2AX (Gorgoulis et al., 2019), as well as members of the senescence-associated secretory phenotype (SASP), a secretion of numerous inflammatory factors and extracellular matrix remodeling components released by senescent cells (Coppé, Desprez, Krtolica, & Campisi, 2010). Due to the non-specific nature of these biomarkers a combined approach has been recommended (Gorgoulis et al., 2019). For example, to indicate a cell is senescent it must be (1) senescence-associated beta-galactosidase positive (2) p16 positive *and* (3) SASP positive. Other combinations have also been suggested, but generally the proposed idea is that there must be multiple sources of support. Highlighting the relevance of this issue, the NIH has recently designated significant funding in this area through the SenNet consortium (SenNet Consortium, 2022).

1.3.3: The Pro-Aging Effects of Senescence

Senescence itself can be beneficial, as it is critical for biological processes such as wound healing (Demaria et al., 2014), embryogenesis (Muñoz-Espín et al., 2013), and tumor suppression (Coppé et al., 2010) (Figure 1.1). However beneficial, senescence represents an example of hormesis, i.e. when low levels of some stressor are beneficial to an organism, but high levels are detrimental (Mattson, 2008). For example, in the short term the SASP is beneficial, as it signals to immune cells that there are damaged cells which need clearing. However, as we age, our immune system's ability to clear senescent cells is impaired and the senescent cells persist (López-Otín et al., 2013). This persistence can lead to problems such as chronic inflammation, tissue and organ dysfunction, as well as tumorigenesis (Figure 1.1). Studies using mouse models have been especially persuasive, as they have indicated that clearing p16-positive cells can extend healthspan

(i.e. the amount of time an organism lives in good health) and lifespan (Baker et al., 2016, 2011; Bussian et al., 2018; Demaria et al., 2017). Cellular senescence has many causes, including epigenetic alterations, which are explored in the next section.

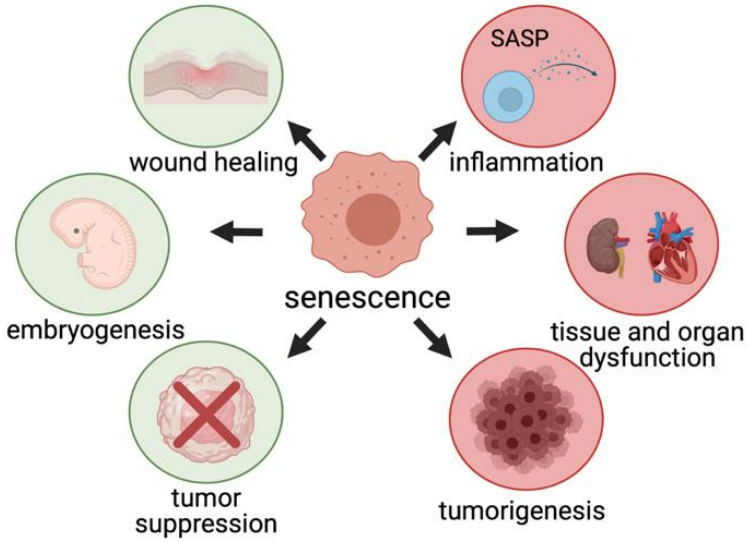


Figure 1.1. Visualization of the beneficial (green) and detrimental (red) roles of senescence.

1.4: Epigenetic Alterations with Senescence

Epigenetics, combining the Greek prefix "epi" meaning on top of, with genetics, refers to the study of changes in gene expression or phenotype that do not involve alterations in DNA sequence, but instead involve more plastic modifications to DNA or chromatin structure. There are many different levels of epigenetic organization which can influence gene expression, ranging from DNA methylation at the small-scale to genomic architecture at the large-scale. Chromatin changes significantly with senescence and these changes have important functional consequences. These changes are outlined in the following sections.

1.4.1: Global Loss and Focal Gains of Heterochromatin

Senescent cells display both a global loss and focal gains of heterochromatin. Early work on senescence supported the heterochromatin loss model of aging, which posits that senescence-

associated changes in gene expression are due to the progressive loss of heterochromatin and the consequent transcription of otherwise-silenced genes (Villeponteau, 1997). In support of this model, drug-induced demethylation impairing heterochromatin formation shortens proliferative lifespan (Gray, Jesch, & Stein, 1991), and histone deacetylase inhibitors also induce premature senescence (Ogryzko, Hirai, Russanova, Barbie, & Howard, 1996).

This global loss of heterochromatin is contrasted with local gains in heterochromatin (Figure 1.2). In RS, although the loss of methylation occurs in gene-poor late-replicating regions associated with the nuclear envelope (NE), the gains of hypermethylation occur in promoter regions, including those of genes which, when repressed, inhibit cell cycle progression. The global hypomethylation is attributed to aberrant localization of DNA methyltransferase 1 (DNMT1) (Cruickshanks et al., 2013). Studies on chromatin accessibility support these findings, as RS cells show increased accessibility in gene-poor heterochromatin and decreased accessibility at promoters and enhancers (De Cecco, Criscione, Peckham, et al., 2013). These changes also lead to the derepression of repetitive elements such as LINE1, which can contribute to chronic inflammation and shorten lifespan (De Cecco, Criscione, Peterson, et al., 2013; De Cecco et al., 2019). Overall, chromatin accessibility increases in both OIS and Notch-Induced Senescence (NIS), a form of secondary senescence, including at gene-distal sites such as enhancers and repeat regions, although largely at separate sites in the two different types of senescence. NIS also shows a decrease in accessibility at gene-distal sites (Parry et al., 2018).

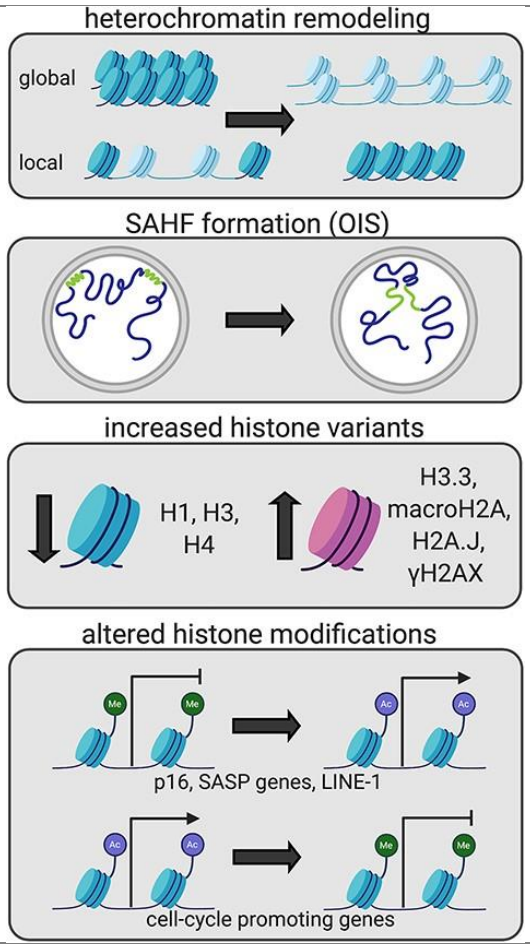


Figure 1.2. Senescence-associated chromatin changes. Globally, heterochromatin (shown in blue) is lost, but there are additional focal gains of heterochromatin. In OIS, SAHFs (shown in green) are formed. Canonical histones become less prevalent, whereas increased proportions of histone variants are observed. Modifications to histones drive a senescence-associated transcriptional program. Acetylation = Ac, Methylation = Me.

1.4.2: Senescence-Associated Heterochromatic Foci

Most prominently in OIS, senescence features the formation of senescence-associated heterochromatic foci (SAHF). SAHF are regions of compacted heterochromatin observed primarily in OIS (Figure 1.2). SAHF exhibit enrichment in H3K9me3, HP1 proteins, H4K20me3 and H3K27me3 around the SAHF periphery (Chandra et al., 2012; Narita et al., 2003; Nelson et al., 2016), all markers of heterochromatin. SAHF contribute to cell cycle arrest by forming on loci

containing genes responsive to E2F, a transcription factor associated with cell cycle progression (Narita et al., 2003).

SAHF formation is orchestrated by High-mobility group (HMG) proteins such as HMGA1 and HMGA2 (Narita et al., 2006). HMG proteins are abundant non-histone regulatory proteins that associate with chromatin and alter its structure. HMGA, HMGB and HMGN are three distinct families of HMG proteins and differ based on their DNA binding motifs (Reeves, 2010). Ectopic overexpression of HMGA2 can induce senescence and SAHF formation (Shi et al., 2017). In the context of NOTCH induced senescence, NOTCH signaling represses HMGA1, disrupting SAHF formation (Parry et al., 2018). Another HMG protein, HMGB2, does not affect SAHF formation but promotes the SASP (Aird et al., 2016). In OIS, HMGB2 localizes to SASP genes and prevents their incorporation into SAHF by fending off the spread of heterochromatin marks (Aird et al., 2016). This contrasts with findings that indicate HMGB2 does not play a similar role in RS. Although there are de novo HMGB2 peaks observed with RS, these do not include loci associated with the SASP (Zirkel et al., 2018). One plausible reason for this discrepancy is the heterogeneity of the SASP across different types of cellular senescence (Coppé et al., 2010; Kirschner, Rattanavirotkul, Quince, & Chandra, 2020).

Finally, recent work has implicated nuclear pores in the formation of SAHF, as the increase in nuclear pore density observed in OIS is required for SAHF formation. The association of the nucleoporin translocated promoter region to the nuclear pore complex was shown to be necessary for SASP activation in OIS (Boumendil, Hari, Olsen, Acosta, & Bickmore, 2019).

1.4.3: Histone Modifications and Variants

All forms of senescence display broad changes in the landscape of histone variants and post-translational modifications. These changes yield altered transcriptional programs, contributing to the senescent phenotype.

Regarding the histones themselves, the prevalence of canonical histone proteins decreases with senescence (Figure 1.2). In RS, stress from telomere shortening causes reduced expression of histones H3 and H4. The decreased quantity of histones compromises the chromatin landscape and thus amplifies local damage to a larger scale (O'sullivan, Kubicek, Schreiber, & Karlseder, 2010). Levels of H1, a linker histone, also decrease in senescent cells containing SAHF (Funayama, Saito, Tanobe, & Ishikawa, 2006).

Histone variants also become more abundant as cells progress into the senescent state (Figure 1.2). For example, increased levels of macroH2A, a family of transcription-silencing histone variants, have been found in SAHF (R. Zhang et al., 2005). In OIS, macroH2A1, a member of the macroH2A family, is redistributed by ATM, notably moving away from SASP genes and allowing their transcription (H. Chen et al., 2015). ATM also mediates the DDR associated with senescence and phosphorylates the histone variant H2AX (Burma, Chen, Murphy, Kurimasa, & Chen, 2001; Mallette, Gaumont-Leclerc, & Ferbeyre, 2007). In its phosphorylated state H2AX (γH2AX) may 'anchor' the ends of double-stranded breaks (DSBs) in close proximity, enabling repair (Bassing & Alt, 2004); this becomes more prevalent with senescence and is even used as a marker of senescence. Another histone variant, H2A.J, accumulates with DNA-damage-associated senescence. This rare variant is important to upregulating inflammatory and immune response genes, including those associated with the SASP (Contrepois et al., 2017). Lastly, the histone variant H3.3 becomes more prevalent in senescence, and its cleavage leads to cell cycle arrest through the silencing of cell cycle regulators (Duarte et al., 2014).

Modifications to histones also play significant roles in senescence by altering the transcriptional landscape (Figure 1.2). For example, H3K9ac and H4K16ac at the promoters of the SASP genes IL-8 and IL-6 increase with senescence, promoting their transcription. Sirtuin 1 (SIRT1), an NAD+-dependent protein deacetylase, normally prevents acetylation of these SASP genes, but its expression is decreased with senescence (Hayakawa et al., 2015). Additionally, The MYST-family histone acetyltransferase MOZ maintains H3K9ac and H3K27ac at several INK4A-ARF pathway inhibitors, including CDC6, EZH2 and E2F2, thus inhibiting senescence (Sheikh et al., 2015). Accordingly, inhibiting MOZ promotes senescence (Baell et al., 2018). The downregulation of EZH2, the aforementioned INK4A-ARF inhibitor, leads to DDR activation and thus senescence (Ito et al., 2018). At later stages of senescence, EZH2 downregulation is followed by H3K27me3 loss, notably at p16 and activation of the SASP. Additionally, in RS histone acetyltransferase p300 drives a senescence-associated transcriptional program because of its induction of super-enhancers enriched in several acetylation marks and H3K4me1 (Sen et al., 2019). These examples are not exhaustive, but indicate some ways in which aspects of senescent cells (SASP, p16 expression) are functionally supported by histone modifications.

1.5: Alterations to Genomic Architecture with Senescence

One of the most exciting emerging areas of epigenetic regulation in senescence is genomic architecture. Recent technological advancements have revolutionized our knowledge of this large-scale chromosome organization. For example, Hi-C, a chromosome conformation capture technique, has revealed the genome is organized hierarchically in three-dimensional (3D) space into open and active A compartments and more closed and repressed B compartments (Lieberman-Aiden et al., 2009). Within these larger compartments are subcompartments: portions of the genome several hundred kilobases in size that have unique patterns of long-range interactions and

have specific transcriptional and epigenetic associations (Rao et al., 2014). Locally, chromosomes are organized into topologically associating domains (TADs), ~1 Mb regions with elevated levels of intra-domain contacts (Dixon et al., 2012). At higher resolution, DNA is folded into loops, formed through a process mediated by cohesin called loop-extrusion. DNA loops often connect promoters and enhancers and are associated with gene activation; convergent CCCTC-binding factor (CTCF) motifs are normally found at loop anchors (Rao et al., 2014).

Senescence is an especially interesting context in which to study chromosome organization, as there are many changes to the nucleus with senescence such as nuclear enlargement and blebbing (Pathak, Soujanya, & Mishra, 2021). Emphasizing these distinctive changes, recent studies have even demonstrated that nuclear morphology can predict if cells are senescent (Belhadj, Surina, Hengstschläger, & Lomakin, 2023; Heckenbach et al., 2022).

1.5.1: Genomic Architecture in HGPS

Although not directly related to senescence, the first study on chromosome organization in an age-related context examined Hutchinson-Gilford progeria syndrome (HGPS), a disease in which patients experience accelerated aging (McCord et al., 2013). McCord *et al.* performed Hi-C on HGPS fibroblasts and identified a loss of compartments in late passage HGPS cells. Albeit the compartment loss, the study still identified an increase in compartment switching (i.e. switching from A to B or B to A) as compared with controls. Compartment switching can affect gene expression, as regions that switch compartments tend to reflect the transcriptional level of the compartment they join (Dixon et al., 2015).

1.5.2: Alterations to Long- and Short-Range Contacts in Senescence

Several studies have examined changes to long- and short-range contacts with senescence. Quantification of the long- and short-range contacts is one of the least granular analyses of Hi-C

data, but can reveal insight into how the genome is organized. The standard cutoff between long- and short-range contacts is 20 kilobases (Lieberman-Aiden et al., 2009) The first Hi-C study of RS found a decrease in long-range and an increase in short-range contacts, supporting the shrinkage of chromosomes that is seen with senescence (Criscione et al., 2016). In contrast, Zirkel et al. investigated chromosome organization in RS using three cell lines at higher resolution and observed an increase in long-range interactions (Zirkel et al., 2018). This discrepancy could be explained by this latter study using cells entering senescence (early senescence), whereas the former one used cells kept in a senescent stage for a long period of time (deep senescence). In agreement with Zirkel et al., Sati et al., who also used early senescence, found a loss of local interactions and gain of distal interactions for both types of senescence (Sati et al., 2020) (Figure 1.3).

	long-range interactions	short-range interactions	compartment switching	TAD boundaries	loop architecture
RS	↑	↓	some; BA > AB	TADs largely unchanged*	*de novo* long-range CTCF loops
OIS	↑	↓	some; BA > AB	TADs largely unchanged	*de novo* cohesin loops

* Varies based on TAD caller used

Figure 1.3. Overview of chromosome organization changes in RS and OIS. Long-range interactions increase while short-range interactions decrease . There is evidence of compartment switching with more B to A than A to B compartment switches. TADs are largely unchanged in both types of senescence, although a less commonly used method identified changes such as fusing and shifting TADs in RS. In RS, there is evidence of de novo long-range CTCF loops formed across sites occupied by HMGB2 in proliferating but not senescent cells. In OIS, de novo loops connecting enhancers and promoters were observed.

1.5.3: Alterations to A/B Compartments with Senescence

There are significant changes to the organization of A/B compartments with senescence. The first Hi-C study of RS found a subset of the genome switched compartments, driving aspects of senescence such as cell cycle arrest and upregulation of the SASP (Criscione et al., 2016). Iwasaki et al. also found that compartment switching occurs with senescence, with a higher proportion of B to A switches than A to B switches (Figure 1.3). The regions undergoing compartment switching are significantly conserved between OIS and RS and exhibit transcriptional changes: upregulation in B to A switches and downregulation in A to B switches. Iwasaki et al. also found condensin is important to maintaining senescence, as it enforces the A compartment and is implicated in B to A transitions, allowing the expression of senescence-related genes (Iwasaki et al., 2019).

Cancer and senescence exhibit similar epigenetic changes at the compartment level. With colorectal cancer, the A and B compartments lose spatial segregation, becoming more homogeneous in 3D space. Additionally, there is a novel self-interacting compartment, I, intermediate to A and B, with distinct patterns of contacts between the A and B compartments within nuclear space. In tumors, the I compartment becomes hypomethylated, resembling the B compartment, and enriched for H3K27me3. Very similar changes were observed with late passage fibroblasts, indicating that the compartmental changes are a feature of excess cell division rather than cancer and thus aid in impeding malignant progression (Johnstone et al., 2020).

While A/B compartments have been analyzed in the context of senescence, no studies to date have examined alteration to subcompartments with senescence. This is an area we will explore in this thesis.

1.5.4: Alterations to TADs with Senescence

As previously mentioned, Topologically Associating Domains have the potential to regulate genes through restraining enhancer activity. Studies have had mixed results regarding alterations to Topologically Associating Domains (TADs) with senescence (Figure 1.3). The first study to examine nuclear architecture in OIS found TAD boundaries were mainly conserved between proliferating and OIS cells (Chandra et al., 2015). In the context of RS, while Criscione et al. reported limited alterations to TAD boundaries, Zirkel et al. found there was shifting, fusing and separating of TADs with RS. Of note, these TADs were identified using a less commonly used method, which was not included in a study that evaluated and compared many TAD callers (Zufferey, Tavernari, Oricchio, & Ciriello, 2018). Olan et al. found TAD borders to be mostly conserved with OIS, but the TAD containing HMGA2 was among those most changed (Olan et al., 2020).

The differing results regarding TADs may also find its roots in the hazy definition of what a Topologically Associating Domain is. Clearly stated by Schoenfelder and Fraser:

> TADs, sub-TADs and loops are related, cohesin-dependent structures; in fact, the current nomenclature in the field is confusing and sometimes misleading. For example, 'loop domains', 'insulated neighbourhoods' and 'CTCF contact domains' all describe loop structures between CTCF-bound sites that are either identical or share extensive overlap. Adding to the nomenclature confusion, even universally accepted descriptors such as 'TAD' are sometimes replaced with much less widely used terms ('contact domain') (Schoenfelder & Fraser, 2019).

1.5.5: Alterations to Loops with Senescence

The field lacks a deep understanding of how genomic architecture is altered at the level of loops due the limiting resolutions of available Hi-C maps, however some topical analysis has been performed examining their role in senescence (Figure 1.3).

In RS, Zirkel et al. found that nuclear HMGB2 was depleted during senescence affects loop architecture (Zirkel et al., 2018). HMGB2 is located at a subset of TAD boundaries, helps insulate CTCF loops, and its depletion is sufficient to form senescence-induced CTCF clusters, a reorganization of CTCF that occurs with senescence. In accordance with HMGB2's insulating role, de novo long-range CTCF loops were observed with RS across locations where HMGB2 was formerly present in proliferating cells.

More in depth analysis has been completed in OIS, but is still limited due a Hi-C map resolution of >10 kb (Olan et al., 2020). Olan et al. found that OIS features altered enhancer-promoter interactions, indicative of loop formation, notably at the IL1 cluster, which contains important SASP and cell cycle-associated genes. OIS-associated changes in enhancer-promoter contacts support that inflammation-related genes are upregulated, and cell cycle-related genes are downregulated. Some of these changes can be attributed to transcription-dependent cohesin repositioning. These repositioned structures or 'cohesin islands' form at the 3′ ends of activated genes after cohesin is loaded on to the transcription start site and progresses because of transcription, but there is inefficient offloading and no impeding CTCF. This process creates de novo cohesin peaks and plausibly de novo loops (Busslinger et al., 2017; Olan et al., 2020).

However, as previously mentioned, we lack a complete understanding of the loop landscape of senescence. We will define this landscape in the upcoming chapters.

1.5.6: The Role of the Nuclear Lamina in Senescence-Associated Genomic Architecture

As early as 2015, with limited resolution Hi-C maps of OIS, Chandra et al. were able to identify the role of the nuclear lamina in driving senescence-associated changes in genomic architecture. This analysis indicated a sequence- and lamin-specific heterochromatic loss of local interactions; these regions were inaccessible, GC poor, enriched for H3K9me3 and associated with lamins. Although these regions lose local interactions, they also come together in space, indicative of SAHF formation (Chandra et al., 2015). This study contrasted their findings with those in HGPS to support a two-step formation of SAHF, as OIS and HGPS share similar changes in local interactions, but HGPS does not exhibit a gain of distal interactions.

Sati et al. supported this finding that SAHF formation is driven by nuclear lamina detachment through the use of polymer modeling (Sati et al., 2020). Furthermore, they found that both RS and OIS feature senescence-associated heterochromatin domains (SAHDs), areas enriched for H3K9me3, which develop into SAHF in OIS. In OIS, DNMT1 drives SAHF formation through increasing the expression of HMGA2. The architectural formation of SAHFs is crucial to senescence as they are able to alter gene expression. Although Iwasaki et al. found genes 500–700 kb from SAHF exhibit statistically significant downregulation, Sati et al. found that SAHF bring together specific loci to enable their gene expression; these include genes relating to cancer and inflammatory response, but do not include SASP genes.

A different study also used polymer modeling to recapitulate senescence-associated changes in chromosome organization (Chiang et al., 2019). A polymer model with varying heterochromatin-heterochromatin and heterochromatin-nuclear lamina interactions identified four chromatin states, including those resembling growing cells, senescent cells and progeroid cells. According to the model, the transition between proliferating and senescent cells is abrupt and is stabilized by hysteresis.

1.5.7: Genomic Shape

While the aforementioned levels of genomic architecture have been explored in depth, one area that has not been recognized is the *shape* of the genome itself, regardless of senescence context. We are limited in our analysis because most interrogations of Hi-C data involve identifying features from contact maps and comparing them, rather than looking at the structure itself. Instead of performing this disjointed process, we could instead use more appropriate shape analysis tools to determine and compare genome architecture. Using elastic shape analysis as in (Srivastava & Klassen, 2018) we can more comprehensively compare the structure of the genome itself in different regions (i.e. TADs, subcompartments, etc.), or different conditions (i.e. proliferating and senescent cells). This would also allow for systematic connections to biology, for example determining if specific structural features are more transcriptionally active, are enriched for specific epigenetic marks, etc. These ideas will be explored in more detail in the upcoming chapters of this thesis.

1.6: Aging and Senescence Interventions

As outlined above, we can appreciate that epigenetic changes, notably those at the level of genomic architecture, drive senescence, which in turn accelerates aging. This leads us to another important question I will explore in later chapters: how do we modulate aging, especially through modulating its more plastic epigenetic changes and especially those relating to senescence? There are several known ways to extend health- and lifespan, and progress is actively being made in these areas.

1.6.1: Senolytics

Senolytics are drugs that preferentially clear senescent cells (Kirkland & Tchkonia, 2020). Due to the technical challenges of completing high-throughput screens with senescent cells and

the lack of specific biomarkers, the first successful senolytics were discovered through hypothesis driven research. The Kirkland lab discovered drugs based on the observation that senescent cells resist apoptosis (Wang, 1995) including the tyrosine-kinase inhibitor Dasatinib (D), and the flavonoids Quercetin (Q) and Fistetin (F). The combination of D and Q has been shown to be a potent anti-aging combination through elimination of senescent cells when used in combination (Y. Zhu et al., 2015) and is beginning to be used in clinical trials (Kirkland & Tchkonia, 2020). Navitoclax is another frequently used senolytic, but it is less specific and can also target non-senescent cells such as immune cells (Mahfoudhi et al., 2016; Wilson et al., 2010). In single-cell experiments Navitoclax has been shown to preferentially act on cells with the phenotype of secondary senescence (Evans et al., 2023).

1.6.2: Senomorphics

Another tactic to extend lifespan is through modulating the more damaging aspects of senescence, such as the SASP. The drugs and treatments which focus on mitigating parts of the senescent phenotype, as opposed to killing them, are called senomorphics (Di Micco et al., 2021). For example, rapamycin and metformin are examples of senomorphics, as they have proven to beneficially diminish the SASP (Herranz et al., 2015; Maruthur et al., 2016; Moiseeva et al., 2013; Noren Hooten et al., 2016; Oubaha et al., 2016).

1.6.3: Epigenetic Reprogramming

Epigenetic reprogramming is one of the most revolutionary techniques developed this century. In their nobel-prize-winning paper Takahaski and Yamanaka showed that exogenously administering the 4 ‘OSKM’ factors (Oct3/4, Sox2, Klf4, and c-Myc) could transform differentiated cells into induced Pluripotent Stem Cells (iPSCs) (Takahashi & Yamanaka, 2006). In the context of aging, a full reversion to a stem cell would not be feasible, as it would disrupt

organism function and has the potential to give rise to teratomas. However, partial reprogramming, i.e. reprogramming while maintaining cell identity, can rejuvenate cells and organisms.

For example, Sarkar et al. transiently transfected old fibroblasts and endothelial cells with OSKM as well as Lin28 and Nanog (OSKMLN) modified RNA and observed evidence of rejuvenation, as transfected cells resembled younger cells according to measures such as heterochromatic marks (H3K9me3) and epigenetic age (Sarkar et al., 2020). In this study epigenetic age was measured by a variation of Steve Horvath's epigenetic clock, a model which examines methylation at sites across the genome to determine biological age (Horvath, 2013). Lin28 and Nanog were added to the modified RNA cocktail to decrease inflammation and increase reprogramming efficiency. The partial reprogramming in this study was ensured by a shorter reprogramming time course. Gill et al. conducted similar experiments and found similar results using patient derived fibroblasts and an OSKM cassette (Gill et al., 2022). Supporting the benefit of partial reprogramming, *in vivo* studies using a progeroid mouse model and cyclic reprogramming treatments with OSKM (as to avoid teratoma formation) led to a ~50% lifespan extension (Ocampo et al., 2016). *In vivo* studies have also indicated that administration of the factors without Myc (OSK), again to avoid teratoma formation, can improve age-associated defects of the eye (Lu et al., 2020). Transient reprogramming factor expression is able to decrease the aging across the body and has been demonstrated to be successful through multi-omic analyses of pancreas, liver, spleen, and blood (Chondronasiou et al., 2022).

The role of senescence in reprogramming is just now being elucidated, as there are a few studies that have examined their contribution. One study was able to partially reprogram senescent cells with OSKML, diminishing aspects of senescence such as beta-galactosidase expression. However, senescence returned after the treatment (Manukyan & Singh, 2014). Senescent cells'

ability to reprogram is also variable, as demonstrated by a study which found that depending on fibroblasts' cytokine profile cells can reprogram more or less favorably, which impacts processes such as wound healing (Mahmoudi et al., 2019). Evidence has also indicated that senescence is important to the reprogramming process (Mosteiro et al., 2016). Reprogramming-induced senescent cells (created by reprogramming-associated damage) express elements of the SASP, such as IL-6, and can thus create a microenvironment permissive to partial reprogramming. However, studies using *Drosophila* have shown that a combination of senolytics and partial reprogramming (cyclical) extends lifespan more than each intervention alone (Kaur et al., 2022), indicating senescent cells are *not* beneficial for reprogramming. Due to these ambiguities, there is a strong need for research in this area to explore and elucidate the role of senescent cells. These studies should ideally be *in vivo* studies, making use of inducible reprogramming mouse models, so we can get a clear picture of how senescent cells contribute to the overall process of reprogramming. Combining these models with single-cell techniques will be especially advantageous to examining senescent cells, which are a rare cell population composing up to ~5% of cells in aged organisms. These ideas will be explored in this thesis.

1.7: Experimental Rationale

Broadly, this thesis aims to increase our potential to extend lifespan and healthspan by expanding knowledge of the structure and function of the nuclear alterations that occur with cellular senescence.

Chapters 2 and 3 of this thesis focus on genomic architecture and its role in cellular senescence. As aforementioned, senescence is driven in part by epigenetic changes, and chromosome organization is a newly discovered and comparatively unexplored area of epigenetic organization. Chapter 2 presents an in-depth analysis of the highest resolution Hi-C maps of

replicative senescence available to date. Due to the 2-3 kb resolution of the maps, I am able to present the first description of high resolution features of the 3D genome in senescence, such as loops. Furthermore, I integrated this Hi-C map with other multi-omic data such as RNA-Seq, bisulfite sequencing, and long-read DNA sequencing to characterize the functional effects of the structural changes we see with senescence.

In chapter 3 I delve into method development and testing of methods to systematically analyze chromosome structures. Many techniques exist to compare and pick out features from Hi-C maps (i.e. large matrices) of genomes, but few exist to analyze the shape of the genome itself, and how it responds to biological changes, such as senescence. I present work on the "Chromosome Shape Alphabet" method, which picks out reoccurring shapes in the 3D genome with the aim of linking these shapes to biological features. We can think of these reoccurring shapes being analogous to features like alpha-helices and beta-sheets in 3D protein structures. I also present work on a novel TAD caller, for use with the Chromosome Shape Alphabet method covered in chapter 4, which I benchmarked compared to existing methods.

In chapter 5 I present a preliminary analysis of reprogrammed mouse liver. This work explores the epigenetic landscape of partial reprogramming with OSKM at the single cell and multi-omic (scATAC and scRNA) level, allowing us to uncover a clearer picture of the role of senescent cells in rejuvenation. Although we observe that some of the data quality is not as good as we would have hoped, we troubleshoot how we can extract the most information possible from the remaining data, and modify future experiments accordingly.

Finally, I conclude with a reflection and summation of how these works contribute to unraveling the mystery of aging and propose future directions.

Chapter 2: A Kilobase Resolution Analysis of the Senescent Genome Reveals Novel Architectural Features Supporting Cell Cycle Arrest and Inflammation

In Chapter 2 Audrey Dalgarno was responsible for all writing, figure generation, and computational analysis besides the analysis of the long-read DNA sequencing and custom LINE1 reference, which were performed by Maxfield Kelsey (Sedivy Lab, Brown University). Audrey Dalgarno also performed all wet lab experiments to prepare the quiescent samples for Hi-C and RNA-seq.

2.1: Abstract

Cellular senescence, a driver of aging, is an irreversible and pro-inflammatory state of proliferative arrest caused by progressive DNA damage. Cellular senescence is associated with many epigenetic changes, including those to genomic architecture. However, the field lacks a complete picture of these alterations due to the limited resolution of the available Hi-C maps. To better understand how changes to the genome's structure drive cellular senescence, and thus aging, we created the highest resolution Hi-C maps to date (~3kb) of proliferating, quiescent, and replicative senescent lung fibroblasts. We also included the highest resolution Hi-C maps of oncogene-induced senescence in our analysis for comparison. We confirm that with senescence we see a loss of heterochromatin, with a shift towards the A compartment and the A subcompartments with senescence. Furthermore, through establishing a novel framework to compare loops across conditions, we observe that senescence features ~6x more unique loops as compared to proliferating cells. Many of these loops form due to specific loss of methylation at their anchors. We present a custom reference genome including novel retrotransposon insertions and use this to determine that structural changes help to support the derepression of LINE1, notably the 'hotspot' we define at L1HS_14q23.2_3. Functionally, the architectural changes we observe help to support senescence, as they are associated with cell cycle arrest and inflammation, as well as loss of cellular identity and epigenetic drift.

2.2: Introduction

Cellular senescence is a deleterious state of proliferative arrest. This hallmark of aging is caused by irreparable DNA damage, stemming from sources such as telomere shortening, oncogenic stress, or irradiation (López-Otín et al., 2013). Although in principle senescence is beneficial, as it prevents the spread of DNA damage and plays important roles in embryogenesis (Muñoz-Espín et al., 2013), wound healing (Demaria et al., 2014), and tumor suppression (Coppé et al., 2010), it also has more detrimental pro-aging effects. For example, senescent cells exhibit the Senescence-Associated Secretory Phenotype (SASP), a secretion of proinflammatory factors which can lead to tumorigenesis (Coppé et al., 2010) and secondary senescent of the surrounding healthy cells (Acosta et al., 2013; Teo et al., 2019). Studies in mice have also indicated that the clearance of senescent cells can extend lifespan and healthspan (Baker et al., 2016, 2011).

At the root of this pro-aging phenotype are epigenetic alterations, which occur across every level of epigenetic organization (Rocha et al., 2022). Most recently, epigenetic changes have been explored at the level of genomic architecture. This has been facilitated by the advent of chromosome conformation capture technologies, such as Hi-C, which reveals the structure of nuclear DNA at the genome-wide scale (Lieberman-Aiden et al., 2009). This technology has shown that the genome is organized hierarchically. At lower resolution there are chromosome territories, which are composed of more open, active A compartments and more closed, repressed B compartments. These compartments are further divided into subcompartments, which have unique patterns of long range interactions and are associated with specific epigenetic marks (Rao et al., 2014). As the resolution of Hi-C data has increased, DNA loops, which connect promoters and enhancers, have also been characterized. These loops are thought to form through a process called loop extrusion, which involves CCCTC-binding factor (CTCF) and cohesin.

Senescent cells undergo coordinated changes to their nuclear structure and organization, such as increased cell size (Zhao & Darzynkiewicz, 2013) and nuclear lamina break-down (Sadaie et al., 2013; Shah et al., 2013). To date, several studies have examined nuclear architecture in the context of senescence (Chandra et al., 2012; Criscione et al., 2016; Iwasaki et al., 2019; Olan et al., 2020; Sati et al., 2020; Zirkel et al., 2018) and reported alterations to long- and short-range chromosome contacts, as well as genomic regions undergoing 'compartment switching' between the A and B compartments. These changes drive features of senescence such as chromosome shrinking and transcriptional changes to the SASP, p53 signaling, and cell cycle regulation. In oncogene-induced senescence (OIS), there is an increased prevalence of cohesin-mediated loops as compared to proliferating cells (Olan et al., 2020).

Although these studies have identified extensive changes in the genomic architecture of senescent cells, their genomic resolution was limited to >10 kb (Olan et al., 2020), hence, they could only provide limited information at the finer scale where loops can be detected. They also lack any comparison with quiescent cells, which share the crucial feature of cell cycle arrest with senescent cells, making it difficult to tease out which architectural changes are associated with senescence-specific control of the cell cycle exit. Furthermore, senescence is a dyanmic process, progressing through different stages from "early" senescence to "late" senescence. Throughout this progression cells undergo transformations, such as lamin B1 downregulation, chromatin remodeling, SASP expression, and retrotransposition (van Deursen, 2014). Here we present a high-resolution map of late senescence, which likely represents a more physiological state of senescence, as the most problematic senescence cells are those that are not cleared and remain in the body.

To gain a more comprehensive view of how genomic architecture changes with senescence we created the highest resolution (~3 kb) Hi-C contact maps to date of proliferating, quiescent, and RS fibroblasts. These maps, as well as recent advancements in Hi-C data analysis, have allowed us to gain a better understanding of how senescence is influenced by the 3D genome. With this new dataset we have been able to validate previous findings found at lower resolution, and furthermore learn about how finer resolution levels of genomic architecture, such as loops and subcompartments are altered. Here we show there is a loss of heterochromatin at the compartment and subcompartment levels with senescence. Additionally, there are many more (~6x) unique loops with RS, as compared to proliferating cells. We establish this by creating a novel method to call loops, which uniquely has the capacity to compare loops from more than two conditions. These loops are driven by senescence-associated hypomethylation, mostly from loops in the B compartment without CTCF at their bases. We find that these levels of organization all support transcriptional features of senescence such as cell-cycle arrest and inflammation, as well as the recently described (Yang et al., 2023) loss of cellular identify due to epigenetic drift with aging. Furthermore, senescence-associated architectural changes support the derepression of replicative elements, namely the retrotransposon LINE1, which leads to chronic inflammation through an interferon response (De Cecco et al., 2019). Our creation of a custom LINE1 reference genome from long-read DNA sequencing and careful analysis of total RNA-Seq data we were able to establish a LINE1 'hotspot' at L1HS_14q23.2_3, which is upregulated due to architectural changes at every level of organization. These observations provide important insight about how genomic architecture contributes to cellular senescence, and thus aging. Additionally, through our high-resolution analysis we pave the way for future mechanistic interventions which have the potential to extend human healthspan and lifespan.

2.3: Methods

2.3.1: Cell Culture

For this analysis female human diploid lung fibroblast cells (LF1) were used. Proliferating cells were passaged in regular growth media and harvested at 60 - 80% confluency. Quiescent cells were obtained by allowing cells to reach 100% confluency and subsequently incubating cells in serum free media for 1-2 days. Replicative senescent (RS) cells were passaged in regular growth media until replicative exhaustion. These cells were maintained in the senescent state for 3-4 months. Growth media consisted of HAMS F10, 15% FBS, and 1x of Penicillin Streptomycin and Glutamine. Cells were maintained at 37 degrees Celsius at 5% CO2 and 2.5% O2.

2.3.2: Hi-C and Analysis

All conditions (proliferating, quiescent, RS) were crosslinked with 1% PFA at room temperature for 10 minutes while still attached to cell culture plates. Crosslinking was quenched with glutamine. Cells were washed with PBS three times before being scraped into a microcentrifuge tube. To perform Hi-C the Arima Genomics Hi-C kit was used. For each condition three replicates were prepared. For each condition samples were pooled and run on 3 lanes of an Illumina Hi-Seq. The resulting fastqs as well as OIS data from GSE135093 were aligned against hg38 and assembled into Hi-C maps with the Juicer pipeline version 1.6 (Durand et al., 2016). Replicates were checked using HiCRep (Lin, Sanders, & Noble, 2021) at 100kb. Visualizations were created using CoolBox (Xu et al., 2021) and Juicer (Durand et al., 2016). The default cutoffs for long- and short-range contacts were used from Juicer (20kb). T-tests were performed across samples comparing the percentage of intra-chromosomal contacts across replicates.

2.3.3: RNA-Seq and Analysis

Total RNA for proliferating and RS samples (extracted using trizol) and cell pellets for quiescent cells were sent to Azenta where libraries were prepared using a stranded ribosomal depletion method. The proliferating and RS cells were split across two HiSeq lanes. The quiescent cells were run on one HiSeq lane. The resulting fastqs as well as the RNA-Seq from GSE72404 were aligned to hg38 with STAR (Dobin et al., 2013). Features were counted with HTSeq (Anders, Pyl, & Huber, 2014), and assessed for differential expression with EdgeR (Robinson, McCarthy, & Smyth, 2010). Replicates were checked using EdgeR's MDS function. GSEA was performed with fgsea (Korotkevich et al., 2016) and MSigDB's Hallmark set (Liberzon et al., 2015).

2.3.4: A/B Analysis

A/B compartments and switching were assessed with dcHiC (Chakraborty, Wang, & Ay, 2022) at 100kb. Compartment switches with an FDR <= 0.05 were considered significant. For each gene we determined what compartment they were in, or if they were in a region switching compartments. We then determined the overlap of genes moving towards the A/B compartments with RS or OIS. Quiescence was considered separately. To determine concordant differential expression within these conditions the logFC and FDR were considered. For example, moving towards the A compartment, genes in the region with a positive logFC and FDR <= 0.05 were considered. GO analysis and KEGG were performed with cluster profiler (Wu et al., 2021). Unless otherwise specified, all venn diagrams were created with eulerr (Larsson, 2021) and the overlap was determined as significant using R's fisher.test with the alternative hypothesis "greater."

2.3.5: Subcompartment Analysis

Subcompartments were called at 25kb using Calder (Y. Liu et al., 2021). More specifically, we used Calder 2, the hg38 compatible version. Regions undergoing subcompartment switching were first identified as either unique to RS or OIS. Quiescence was considered separately. The

genes within these regions were determined with bedtools intersect (Quinlan & Hall, 2010) with the default settings. Genes within the regions undergoing compartment switching were filtered based on RNA-Seq (i.e. concordant v discordant based on logFC and FDR <= 0.05). GO analysis was performed with cluster profiler (Wu et al., 2021). Terms of interest were curated from the overall GO results to created the visual diagram with the arrows. (Figure 2.3 B/C).

2.3.6: Loop Analysis

Loops were called with Juicer's CPU HiCCUPS (Durand et al., 2016) and the default parameters. To compare the proliferating and senescent genes we wrote a custom script. Quiescence was considered separately. We set a threshold of size n (which can be varied by the user) and used a window of size +/- n around the base of each loop to decide if loops were the same. If a loop fell within +/- n of both loop bases then loops were called the same (Supplementary Figure 2.5A). By varying the threshold between 10kb and 100kb we determined a window size (n) of 30kb to use (Supplementary Figure 2.5B/C). Graph visualizations were created with NetworkX (Hagberg, Schult, & Swart, n.d.). Due to the nested nature of loops we also established "lenient" (more loops altered) and "stringent" criteria (less loops altered) to determine if regions within loops were altered with senescence. For example, we consider the situation where a region within a loop shared by proliferating and senescent conditions is within a loop unique to senescence. In the lenient condition this would be called altered with senescence, whereas with the stringent criteria this would be called not altered with senescence (Supplementary Figure 2.5D). The genes within these regions were determined with bedtools intersect (Quinlan & Hall, 2010) and a cutoff of .6. Genes within the loops were filtered based on RNA-Seq (i.e. upregulated v. downregulated based on logFC and FDR <= 0.05). GO analysis and KEGG were performed with cluster profiler (Wu et al., 2021).

2.3.7: Methylation Analysis

Methylation data were acquired from GSE48580 and lifted over to hg38. This data was intersected with the loop bases using bedtools intersect (Quinlan & Hall, 2010) and counted. For the analysis included in Figure 2.5 we specifically examined the senescence-specific loops losing methylation at both anchors. Bedtools intersect was also used to determine the overlap of the senescent loops losing methylation at both anchors with A/B compartments and subcompartments with a threshold of .6. CTCF motifs locations were determined using a CTCF motif from JASPAR (Rauluseviciute et al., 2023) and the MEME suite's FIMO (Bailey et al., 2009). These were intersected with methylated loop bases using bedtools intersect and counted.

2.3.8: SASP Analysis

We curated a SASP list by combining the genes of the SASP atlas (Basisty et al., 2020) with the interferon and SASP lists from De Cecco et al. (De Cecco et al., 2019) and removing duplicates. A SASP unique to each condition (i.e. RS SASP or OIS SASP) was identified by taking the overlap of this SASP list and genes upregulated with that specific condition.

2.3.9: Long-Read DNA Sequencing

LF1 genomic DNA was extracted using the Monarch Nucleic Acid Purification kit. Nanopore sequencing libraries were prepared using the Ligation Sequencing Kit V14 (SQK-LSK114) and sequenced on a r10.4.1 flow cell using a P2-solo sequencer. Sequencing parameters were unchanged from the default settings.

2.3.10: Custom LF1 reference genome generation

A custom genome incorporating non-reference (private) retrotransposable element insertions was produced on the basis of long-read DNA sequencing data. In brief, >60X coverage worth of nanopore long sequencing reads were fed to the "TLDR" program (Ewing et al., 2020),

which outputs insertion sites and sequences, alongside various quality metrics. Extra chromosomal insertion "patches" were appended to the T2T-HS1 reference (Nurk et al., 2022), which were comprised of the full private insert sequence plus 500 base pairs of flanking sequence on either side of the insert-breakpoint. Custom gene transfer file annotations were subsequently produced to document the presence of these private insertions.

All reference and private L1HS and L1PA2 elements were inspected for the presence of intact (full length ORF1p and ORF2p) protein coding potential. Sequences were extracted from our custom reference using bedtools getfasta (Quinlan & Hall, 2010) in addition to the private insert augmented T2T-HS1 RepeatMasker gtf annotation. Open reading frames were detected using the R package "Orfik" (Tjeldnes et al., 2021) and filtered on the basis of length. Elements which passed these criteria were denoted as "intact".

Reference and private intact L1HS coordinates were lifted over to hg38 with the liftOver tool from the Broad Institute for use in Hi-C analyses which had reads aligned to this reference.

2.3.11: LINE1 Analysis

To determine the overlap of the LINE1s with A/B compartments and subcompartments we used bedtools intersect with a cutoff of 0.6 (Quinlan & Hall, 2010). To determine the proximity of loops to LINE1 elements we used bedtools closest.

2.3.12: LINE1 Expression Analysis

Sequencing reads were trimmed using fastp (S. Chen, Zhou, Chen, & Gu, 2018) and aligned to our custom LF1 genome reference with STAR (Dobin et al., 2013). Gene count estimates were obtained using featureCounts (Liao, Smyth, & Shi, 2014) and the RefSeq transcript annotation (October 2023 revision, GCF_009914755.1). Repetitive elements count estimates were produced with the Telescope tool (Bendall et al., 2019) and a custom RepeatMasker derived

annotation allowing for multi-mapping reads overlapping RTEs to be assigned to the best supported locus via an expectation maximization algorithm (see Telescope methods (Bendall et al., 2019)). Count normalization for genes and repetitive elements was performed using DESeq2's (Love, Huber, & Anders, 2014) median of ratio's method, wherein size factors were estimated by setting the "controlGenes" parameter of the "estimateSizeFactors" factors to include only RefSeq genes. Differential expression was assessed using DESeq2. Intact L1HS loci were subset by genome architectural features derived from our HIC analysis using the GenomicRanges (Lawrence et al., 2013) R package.

2.3.13: Quiescence RT-qPCR

For this analysis female human diploid lung fibroblast cells (IMR90) were used. Proliferating cells were passaged in regular growth media and harvested at 60 - 80% confluency. Contact inhibited quiescent cells were obtained by allowing cells to reach 100% confluency in regular growth media. Serum withdrawal quiescent cells were obtained by incubating cells in regular growth media for 0.01% FBS for 1-2 days. Senescent cells were obtained by treating cells with regular growth media supplemented 40 uM etoposide every other day for two weeks. Growth media consisted of DMEM, 10% FBS, and 1% of Penicillin Streptomycin and Glutamine. Cells were maintained at 37 degrees Celsius at 5% CO2 and 2.5% O2.

RNA was extracted using the QIAGEN RNeasy micro kit and reverse transcribed using the QUIAGEN QuantiTect Rev. Transcription Kit kit. A qPCR was performed on the cDNA using the Applied Biosciences *Power* SYBR Green PCR Master Mix and Applied Biosystems QuantStudio 6 Pro Real-Time PCR System machine according to manufacturer's instructions. The primers used were from (De Cecco et al., 2019). PUM1 was used as a control, as in (González-Bermúdez, Anglada, Genescà, Martín, & Terradas, 2019).

2.3.14: Cell Cycle Analysis

Cell cycle genes were retrieved from the "cell cycle" GO term (GO:0007049) from bioconductor's org.Hs.eg.db package (Carlson, 2019). We determined the coordinates of each gene and intersected them with the A/B compartments, subcompartments, and loops from the quiescent and RS conditions using bedtools intersect with a cutoff of .6 (Quinlan & Hall, 2010).

2.4: Results

2.4.1: Hi-C Quality Control

To better understand how genomic organization supports senescence we performed Hi-C on proliferating and late RS lung fibroblast (LF1) cells in triplicate. We sequenced the resulting Hi-C libraries to a depth of ~1 billion reads per sample (Figure 2.1A, Supplementary Table 2.1, Supplementary Figure 2.1). After confirming low variability between replicates (Figure 2.1B) we combined the replicate samples, resulting in the deepest available map of any type of senescence at 2.75 kb resolution. Our proliferating controls achieved a resolution of 3.35 kb.

With these maps we first examine how patterns of long- and short-range contacts are altered with RS. Past studies have had mixed results when it comes to these patterns. An earlier paper from our lab on genomic architecture in RS (Criscione et al., 2016) found a decrease in long-range interactions and an increase in short-range interactions, which is consistent with the shrinking of chromosomes that accompanies RS. However, this data conflicts with other studies which found an increase in long-range interactions and a decrease in short-range interactions (Sati et al., 2020; Zirkel et al., 2018). Initially, we hypothesized that this inconsistency may be due to some of the improvements to the in situ Hi-C protocol in Rao et al.'s seminal 2014 paper (Rao et al., 2014). However, our current high-resolution maps support our lab's original findings in that there is a decrease in long-range interactions (p-value = 0.03) and an increase in short-range

interactions (p-value = 0.09) with RS (Figure 2.1C/D). From this we can infer that this is not a technical artifact, but may be due to the stage of RS in which Hi-C was performed, as both of our lab's analyses focused on significantly later stages of RS, as mentioned in Zirkel et al. In this case the seemingly dynamic nature of genomic architecture would echo aspects of senescence that are also altered over time, such as SASP gene expression, lamin B1 downregulation, and retrotransposition (van Deursen, 2014). We also compared this finding to the most recently published and highest resolution OIS dataset from the Narita lab (> 10kb) (Olan et al., 2020) and similarly found that OIS cells feature more short-range contacts (p-value = 0.15) and less long-range contacts (p-value = 0.02) as compared to proliferating cells (Figure 2.1E/F). These trends are consistent with the shrinking of chromosomes seen with senescence.

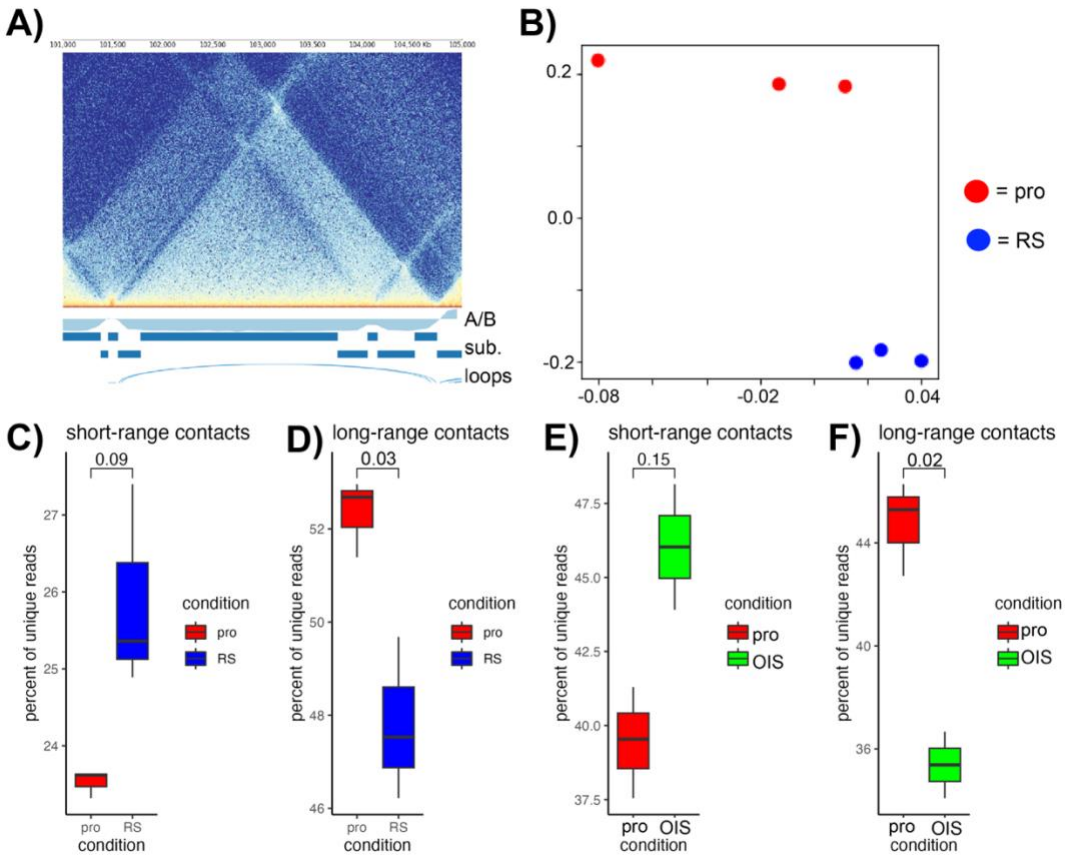


Figure 2.1. Quality Control for Hi-C Maps. A) Visualization of sample high resolution Hi-C map and annotated features. B) Multidimensional scaling plot of proliferating and RS replicates. C) Short-Range intra-chromosomal interactions for proliferating and RS cells. D) Long-Range intra-chromosomal interactions for proliferating and RS cells. E) Short-Range intra-chromosomal interactions for proliferating and OIS cells. F) Long-Range intra-chromosomal interactions for proliferating and OIS cells. Statistics represent p-value outputs from student's t-test.

2.4.2: Genomic Architecture at the Compartment and Subcompartment Levels Reflects Senescence-Associated Loss of Heterochromatin

Past studies have noted increased compartment switching from inactive B to active A compartments with senescence (Iwasaki et al., 2019). Our results support this, as with both RS and OIS we see more of the genome moving towards the A compartment as compared to towards the B compartment (Figure 2.2A). Of note, RS features even more compartment switching than OIS (chi-sq test with p value < 0.05) with ~4% of the genome moving towards the A compartment and ~1% of the genome moving towards the B compartment. This is consistent with the loss of heterochromatin model (Villeponteau, 1997), which states that with senescence there is an overall loss of heterochromatin, leading to aberrant transcription such as the upregulation of repetitive elements (De Cecco, Criscione, Peckham, et al., 2013). Our total RNA-Seq supported canonical senescence expression patterns (Supplementary Figure 2.2) and allowed us to examine gene expression associated with genomic architecture. While only a fraction of the genes undergoing compartment switching were differentially expressed, more were expressed in a concordant direction with the switch (Figure 2.2B). For example, of the 1872 genes moving towards the A compartment with RS alone, 126 are concordantly expressed (upregulated), while 64 are discordantly expressed (downregulated). There was a statistically significant overlap (fishers exact test with p-value <0.01) between the genes undergoing compartment switching with RS and OIS.

The four shared concordantly expressed genes moving towards the A compartment included ATP6V0A4, TTC9, SERPINA9, SERPINA1, while the two shared concordantly expressed genes moving towards the B compartment were TRPA1 and TRHDE-AS1. Supporting the inflammatory associations of senescence, genes moving towards the A compartment with RS only were enriched for terms such as immunoglobulin mediated immune response, B cell mediated immunity, complement activation, and regulation of transforming growth factor beta (Figure 2.2C). The KEGG pathway complement and coagulation cascade was also enriched, due to the upregulation of genes such as SERPINA5 and C7.

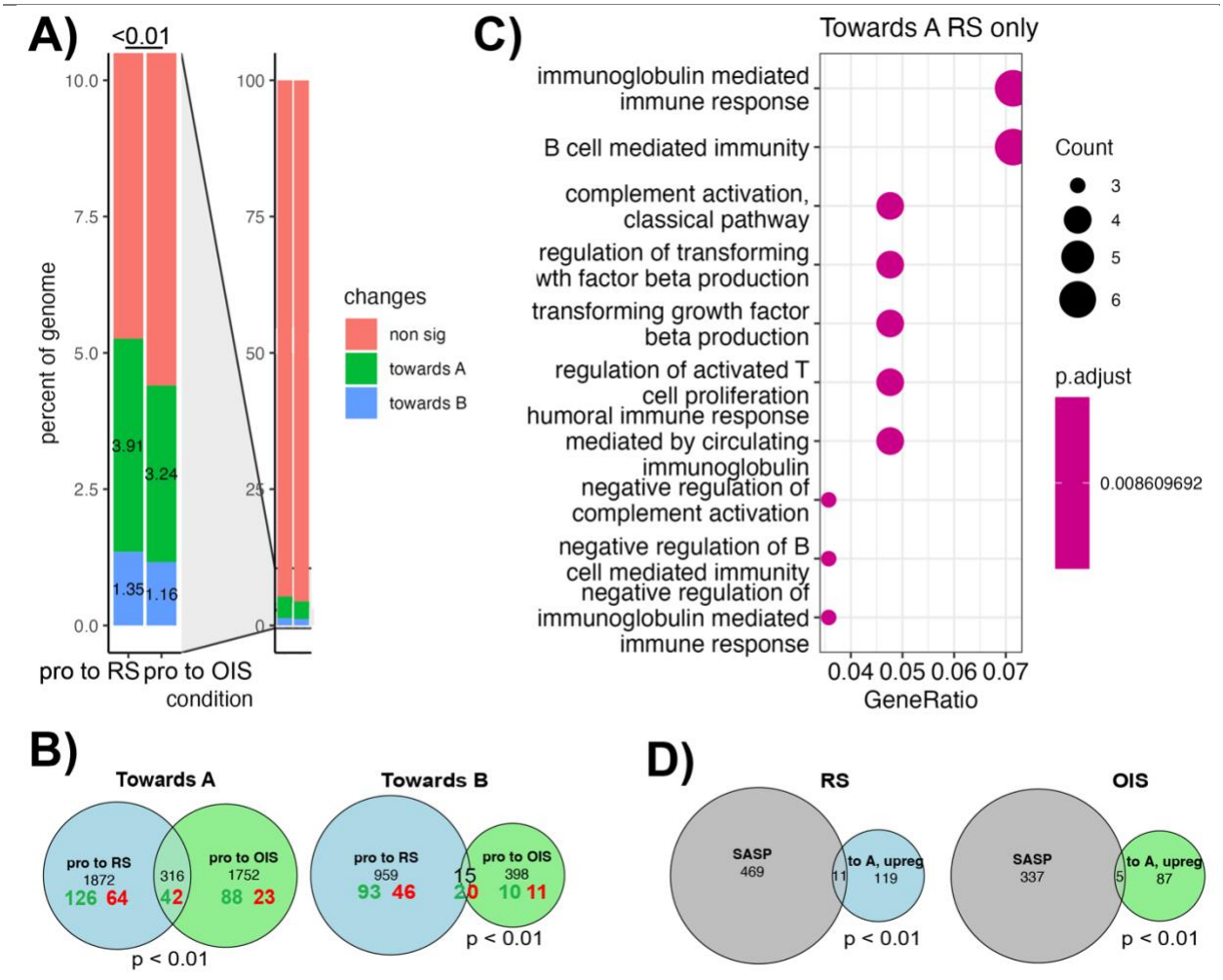


Figure 2.2. Senescence-associated loss of heterochromatin contributes to inflammation. A) Compartment switching with RS and OIS. P-value represents output of chi-sq test. B) Overlap

between total number of genes (black), genes differentially expressed in the concordant direction (green), and genes differentially expressed in the discordant direction (red) with RS and OIS. C) GO analysis of genes moving towards the A compartment with RS D) Overlap between the SASP and genes moving towards the A compartment which are upregulated in RS (left) and OIS (right). P-values for venn diagrams are from Fisher's exact test.

The four concordantly expressed genes moving towards the A compartment with both RS and OIS were enriched in the KEGG complement and coagulation cascade as well as GO terms relating to endopeptidase, peptidase, and proteolysis (Supplementary Figure 2.3A, B). Terms relating to endopeptidase and peptidase were also enriched moving towards the A compartment with OIS only (Supplementary Figure 2.3C). Examining the enrichment moving towards the B compartment, unique to OIS there were terms such as condensed chromosome (in centromeric regions) (Supplementary Figure 2.3D). These support the expansion of repetitive regions, such as centromeres, seen with senescence (Criscione et al., 2016). This expansion of repetitive regions is seen with different types of senescence, including RS, but this result indicates that it is uniquely supported by compartment switching in the context of OIS. The two shared concordant genes moving towards the B compartment with RS and OIS were enriched for terms relating to pain, temperature, and stress response (Supplementary Figure 2.3E). Additionally, these shared terms moving towards the B compartment were enriched for the KEGG pathway inflammatory mediator regulation of TRP channels (Supplementary Figure 2.3F).

Only a small portion of the SASP activation in RS and OIS can be explained by compartment switching (Figure 2.2E), but this is statistically significant (fishers exact test with p-value < 0.01). Eleven SASP genes, including KLF4 (a Yamanaka factor) (Takahashi & Yamanaka, 2006) and inflammatory genes like C7 and SERPINE1, exhibited compartment switching towards the A compartment in response to RS. KLF4 has also been implicated in inducing architectural changes to organize enhancer networks in this process (Di Giammartino et al., 2019). SERPINEA1

was also among the five SASP genes undergoing compartment switching with OIS, along with genes such as the cell-cycle-associated CDK20.

Although subcompartments were first characterized almost ten years ago (Rao et al., 2014) the field lacked tools to identify subcompartments until recently. We used the newly-developed Calder (Y. Liu et al., 2021) to identify eight different subcompartments. Subcompartments range from A.1.1., the most active, to B.2.2., the most repressed. With RS, the greatest percentage of the genome (4.86%) switched from the A.1.2. subcompartment to the A.1.1. subcompartment (Figure 2.3A). This was reflected in the number of genes undergoing subcompartment switching and being concordantly expressed (Supplementary Figure 2.4A - C). In agreement with the findings at the compartment level, we also see a loss of heterochromatin with RS, as noted by the enrichment of switching in the bottom right half of Figure 2.3A, indicating switching from the more repressed B-type subcompartments to the A-type subcompartments more enriched for transcription. Functionally, this subcompartment switching contributes to senescence characteristics such as cell cycle arrest and inflammation (Figure 2.3B, Supplementary Table 2.2). Reflecting the compartment switching, inflammatory terms are enriched switching to more active subcompartments. For example, complement-associated terms switch to more active subcompartments from the B.2.2. subcompartment. Terms associated with replication, such as replisome and DNA polymerase move towards more repressed subcompartments. Other interesting subcompartment switches include enrichment for lamin binding moving towards a more repressed subcompartment.

With OIS, the greatest percentage of genome subcompartment switching is from the B.2.2. subcompartment to the B.2.1. subcompartment (4.21%) followed by A.1.1. to A.1.2. switching (3.82%) and then B.2.1. to B.1.2. switching (3.67%) (Figure 2.3C). As in RS, this was largely

reflected in the number of genes undergoing subcompartment switching and being concordantly expressed (Supplementary Figure 2.4D - F). There was some loss of heterochromatin with OIS, but significantly more (chi-sq test with p-value < 0.05) with RS. Again, the GO terms associated with subcompartment switching with OIS supported many functional aspects of senescence (Figure 2.3D, Supplementary Table 2.3). For example, the inflammatory terms CXCR chemokine receptor binding and IL1 receptor binding switch to more active subcompartments. Cell-cycle associated terms such as mitotic cell cycle checkpoint signaling and spindle cell cycle checkpoint signaling also transition from more active subcompartments to more repressed subcompartments. Enrichment for the GO term nuclear lamina also moves to a more repressed subcompartment, as well as telomere maintenance via recombination. Subcompartment switching significantly helps to regulate the SASP as 33% and 29% of SASP genes also undergo subcompartment switching towards more active subcompartments with RS and OIS, respectively (fishers exact test with p-value < 0.01) (Figure 2.3E/F).

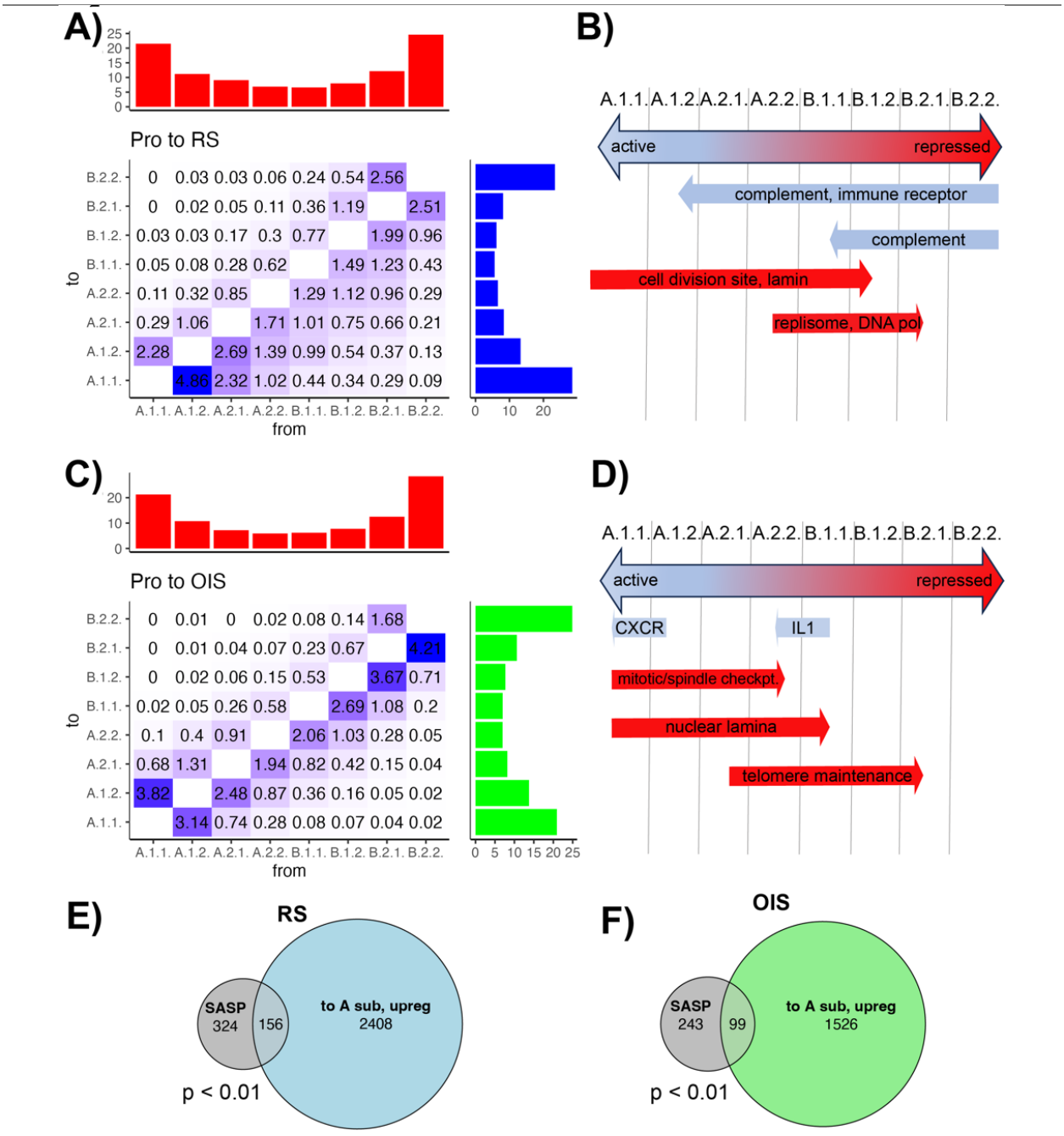


Figure 2.3. Senescence-associated subcompartment switching contributes to characteristics of senescence. A) Percent of genome switching subcompartments from proliferating (columns) to RS (rows). Marginal histograms show distribution in each subcompartment for proliferating (top, red) and RS (right, blue). B) Select enriched GO terms moving between subcompartments with RS. C) Percent of genome switching subcompartments from proliferating (columns) to OIS (row). Marginal histograms show distribution in each subcompartment for proliferating (top, red) and OIS (right, green). D) Select enriched GO terms moving between subcompartments with OIS. E) Overlap between the SASP and genes moving towards more active

subcompartments which are upregulated in RS. F) Overlap between the SASP and genes moving towards more active subcompartments which are upregulated in OIS. P-values for venn diagrams are from Fisher's exact test.

2.4.3: The Changing Loop Landscape of Replicative Senescence

The high resolution of our data also permitted us to examine loops. This is the first study to analyze loops in the context of RS. We did not examine loops in OIS, as the OIS dataset did not have a sufficiently high resolution to reliably detect loops, especially in comparison to our high-resolution datasets. From our high resolution Hi-C maps we identified 9440 and 17020 loops in proliferating and RS cells, respectively. We did not find a correlation between the resolution of the map and the number of loops called, indicating the high number of loops in senescence is a biological, rather than a technical result.

In order to effectively compare loops across conditions we developed a novel methodology to discern if loops were unique to each condition or shared. Publicly available tools such as HiCCUPSDiff (Durand et al., 2016) can compare loops across two conditions, but are limited in that they do not report common loops and that they are unable to compare more than two sets of loops. Our method uses a window of varying size around each loop base to compare loops and determine if they are the same. If both loop bases are within this window threshold of each other they are relabeled as the same loop (Supplementary Figure 2.5A). Using chromosome 21 as a representative chromosome, we varied this threshold between 10 kb and 100 kb and determined 30 kb as an appropriate cutoff, as the graph of unique and shared loops leveled off at this point (Supplementary Figure 2.5B). Graph visualizations of unique and shared loops at thresholds of 10 kb and 100 kb can be seen in Supplementary Figure 2.5C. Comparing proliferating and RS cells, we identified 1536 loops unique to proliferating cells and about 6 times more unique loops with RS cells (8930). There were 7643 common loops (Figure 2.4A).

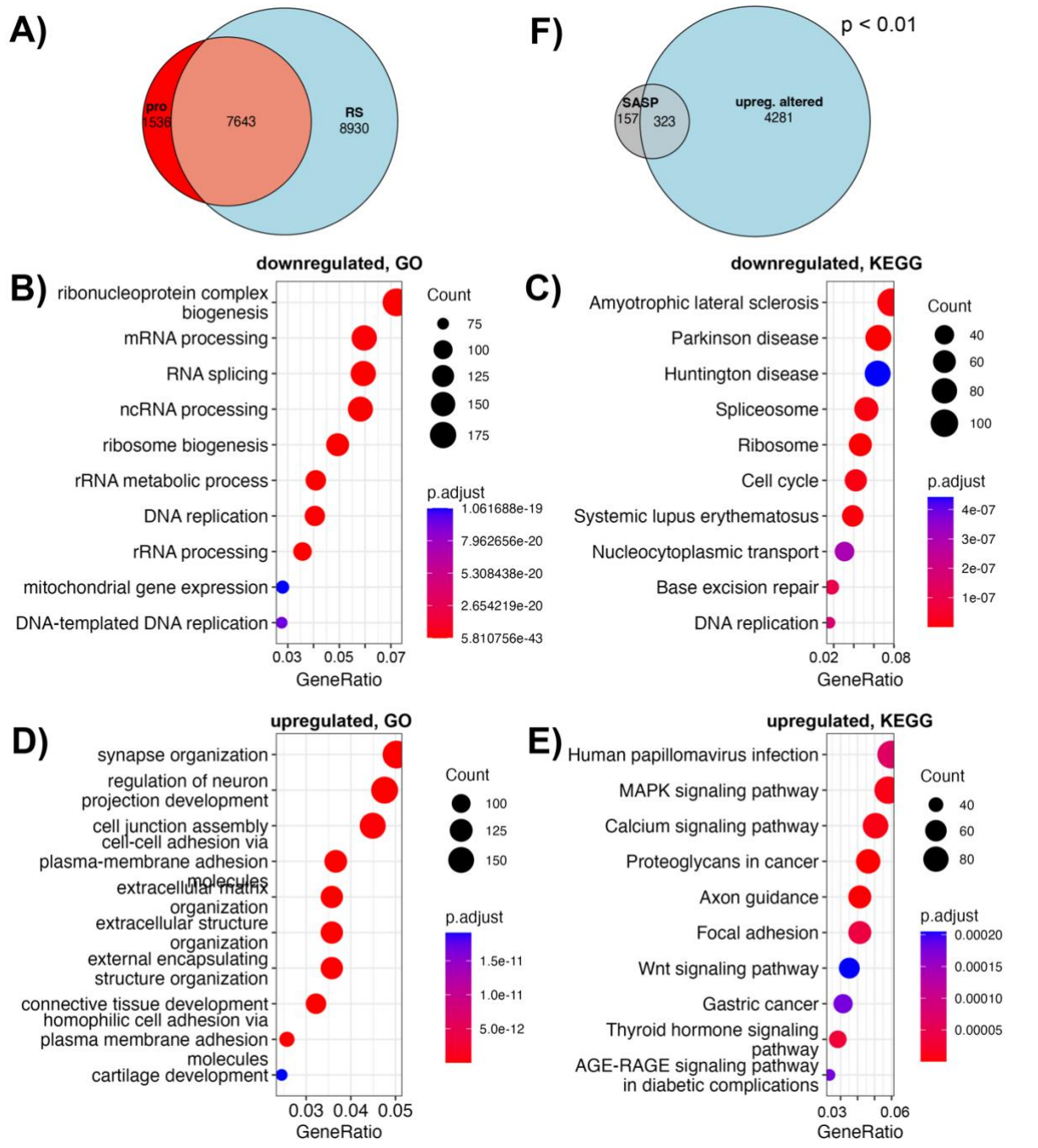


Figure 2.4. Senescence features many new loops associated with cell cycle arrest. A) Venn diagram of loops in proliferating and RS cells. B) GO analysis of genes downregulated within senescence-altered loops. C) KEGG analysis of genes downregulated within senescence-altered loops. D) GO analysis of genes upregulated within senescence-altered loops. E) KEGG analysis of genes upregulated within senescence-altered loops. F) Venn diagram of SASP genes and genes within loops altered with senescence. P-values for venn diagrams are from Fisher's exact test.

To probe the functional roles of these loops we performed GO and KEGG analysis for the upregulated and downregulated genes within these regions. Due to the complex nested nature of loops we also needed to decide if regions within a group of nested loops were affected if any of the loops within the nested group were altered with senescence. We established two criteria: (1) lenient, in which the loops are considered altered in this situation and (2) stringent, in which case the loops are not considered altered in this situation (Supplementary Figure 2.5D). For the lenient criteria of the loops altered with RS, those associated with downregulation included enrichment for terms such as DNA replication (Figure 2.4B). These genes were also enriched for downregulation of proliferation related KEGG pathways such as cell cycle and DNA replication (Figure 2.4C). Interestingly, many of the RS-loop specific genes upregulated with senescence were associated with neurons. These genes were enriched for GO terms, such as synapse organization (Figure 2.4D), and neuron-associated KEGG pathways, such as axon guidance (Figure 2.4E). The stringent criteria loops gave similar results (Supplementary Figure 2.6A-C). Similarly to with the compartment and subcompartment levels, loop alteration with senescence also significantly supports the SASP with 67% of SASP genes being affected by loops (fishers exact test with p-value < 0.01) (Figure 5F).

We also sought to determine why the senescent condition had so many more unique loops, as compared to the proliferating cells. One explanation is altered methylation. Methylation is known to be lost with senescence (Cruickshanks et al., 2013) and a loss of methylation can also contribute to loop formation (X. Zhang et al., 2020). We used a published RS methylation dataset (Cruickshanks et al., 2013), which indicated differentially methylated regions with senescence. Using this, we determined if methylation is lost or gained at 0, 1, or 2 of the loop anchors in the proliferating and RS specific loops. From this analysis we observed that methylation is specifically

lost at RS loop anchors, potentially contributing to the formation of these loops (chi-square with p-value < 0.01) (Figure 2.5A, Supplementary Figure 2.6D). We then investigated the architectural environment of these loops that are specifically losing methylation with senescence. Most of these loops (> 60%) did not have CTCF motifs at either anchor (Figure 2.5B). Although the majority of these loops were found in the B compartment, a subset underwent compartment switching towards the A compartment, as would be expected with this loss of methylation (Figure 2.5C). Reflecting the subcompartment switching trends (Figure 2.2) many loops that are hypomethylated with senescence underwent switching from more repressed to more active subcompartments (Figure 2.5D). The majority of the hypomethylated loops were in the B.2.2. subcompartment. Although only a small percentage of the genes contained within these loops were differentially expressed (~10%) this accounted for many SASP genes (25%) (fishers exact test with p-value < 0.01) (Figure 2.5E).

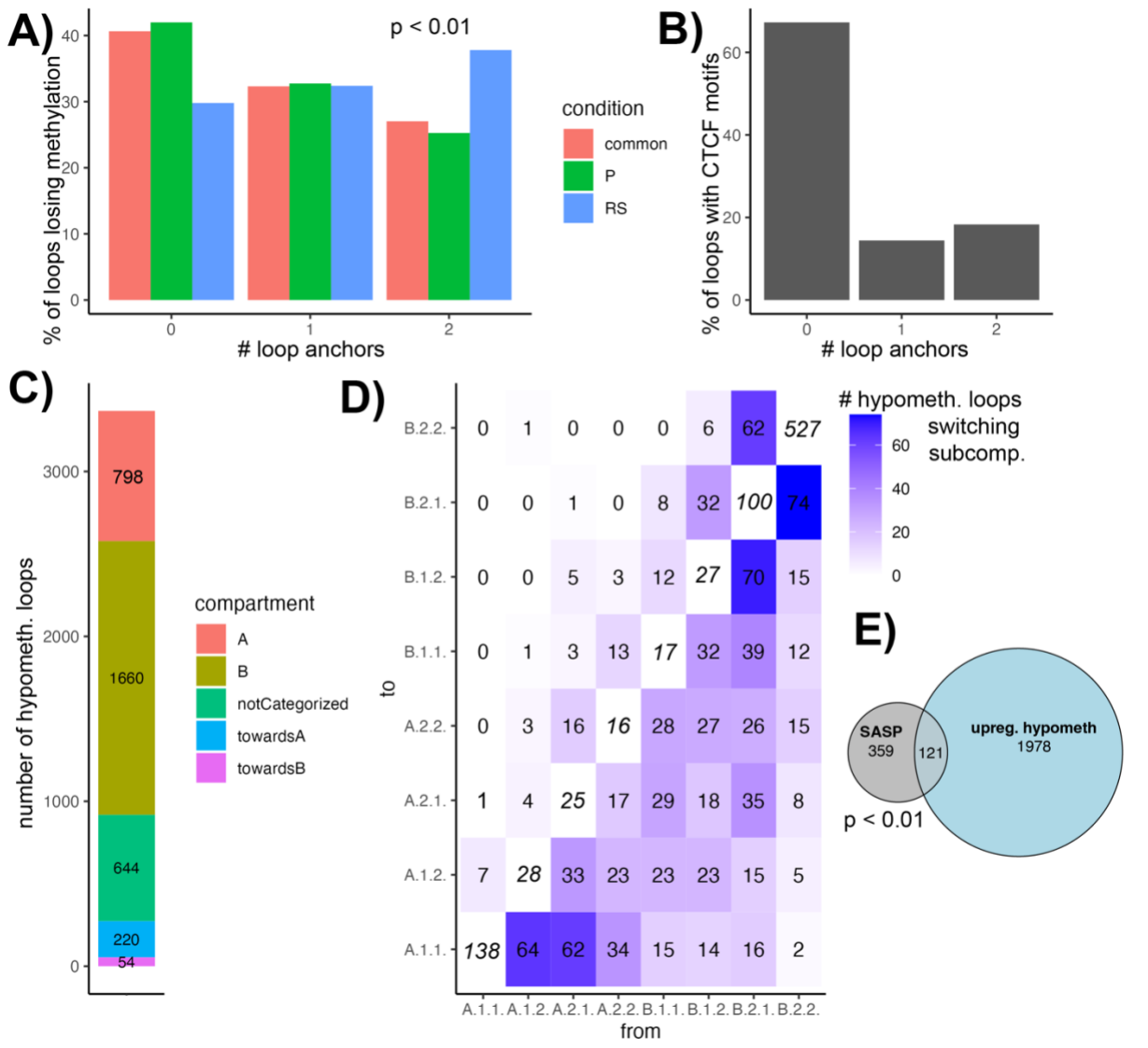


Figure 2.5. Senescence-specific loops preferentially lose methylation at loop anchors. A) Percent of loops losing methylation at 0, 1, or 2 of their anchors. P-value < 0.01 from chi-square test. B) Percent of senescence-specific hypomethylated loops with CTCF at 0, 1, or 2 of their anchors. C) Compartmentalization of senescence-specific hypomethylated loop. D) Subcompartmentalization of senescence-specific hypomethylated loop. E) Venn diagram of SASP genes and upregulated genes within senescence-specific hypomethylated loops. P-values for venn diagrams are from Fisher's exact test.

2.4.4: An Architectural Landscape favoring Retroelement De-Repression

Senescence is also associated with the upregulation of repetitive elements, which can lead to inflammation (De Cecco, Criscione, Peckham, et al., 2013; De Cecco, Criscione, Peterson, et

al., 2013; De Cecco et al., 2019). Specifically, the derepression of LINE1 retrotransposable elements is known to cause an interferon response, leading to the SASP. For this reason, we examined whether architectural changes contribute to LINE1 derepression. We used a curated list of LINE1 elements identified with long-read DNA sequencing for this analysis. Using our total RNA-Seq data we were able to determine that overall LINE1s were upregulated with RS (Figure 2.6A). We were unable to perform this analysis with the OIS data due to the sequencing configuration used by the Narita lab (Hoare et al., 2016; Olan et al., 2020). Furthermore, through careful analysis of our RNA-Seq data we were able to determine which individual LINE1 elements were upregulated. With both RS and OIS there were more LINE1 elements moving towards the A compartment (as opposed to towards the B compartment), supporting derepression (Supplementary Figure 2.7A). One LINE1 element (L1HS_14q23.2_3) was significantly upregulated moving towards the A compartment (Supplementary Figure 2.7B). There was also notable switching of active LINE1 elements between subcompartments with both RS and OIS, generally moving in a derepressed direction, as noted by enrichment in the bottom right half of the heatmap. (Figure 2.6B, Supplementary Figure 2.7C). At the level of loops, there was also evidence of derepression, as there were more loops with RS (v proliferating cells) that formed closely to active LINE1 elements (Figure 2.6C). Overall, it seems architectural changes can contribute to LINE1 upregulation, as LINE1 in more active subcompartments, moving towards more active subcompartments, or in loops were more highly expressed (Figure 2.6D). Although this result was not significant we can appreciate this trend, especially given the low read count due to the difficulties of identifying individual LINE1 elements with short-read sequencing. We also determined a LINE1 ‘hotspot,’ i.e. a specific LINE1 element (L1HS_14q23.2_3) that was significantly upregulated due to architectural changes (compartments, subcompartments, and

loops) (Figure 2.6D, 2.6E, Supplementary Figure 2.7B). This LINE1 hotspot is controlled by a switch towards the A compartment, from the B.1.2. subcompartment towards the A.1.2. subcompartment, and is within a loop that is unique to RS, supporting that in the context of senescence LINE1 activation is driven, at least in part, by architectural changes.

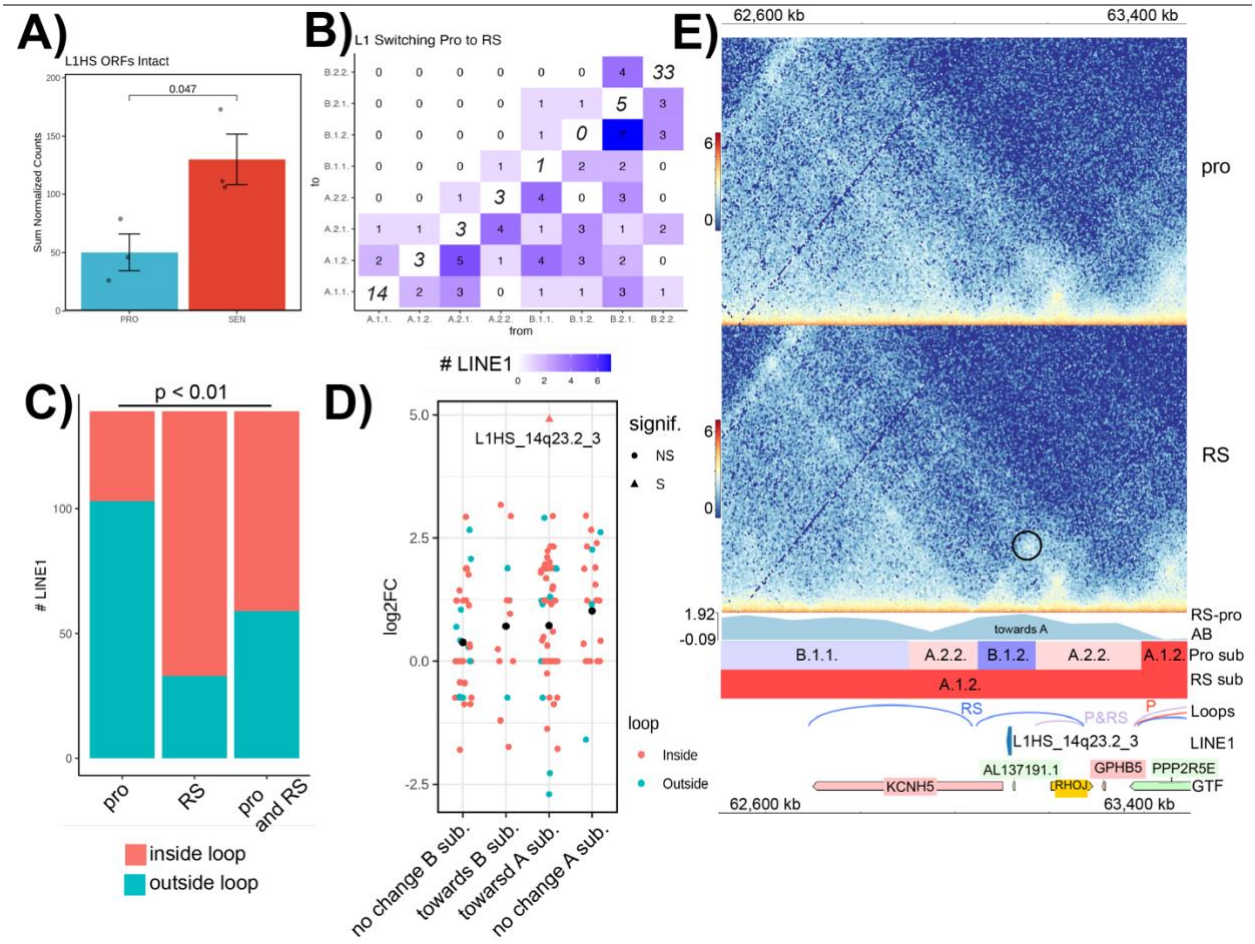


Figure 2.6. Senescence-associated architectural changes contribute to LINE1 derepression. A) Expression of LINE1 in proliferating and senescent (RS) cells. B) Subcompartmentalization of LINE1 with RS. C) Proximity of LINE1 with pro, RS, and common (pro & RS) loops. D) Expression of individual LINE1 elements categorized by subcompartment type, loop proximity, and significance level. Black dots represent means. E) Visualization of architectural landscape surrounding significantly upregulated L1HS_14q23.2_3.

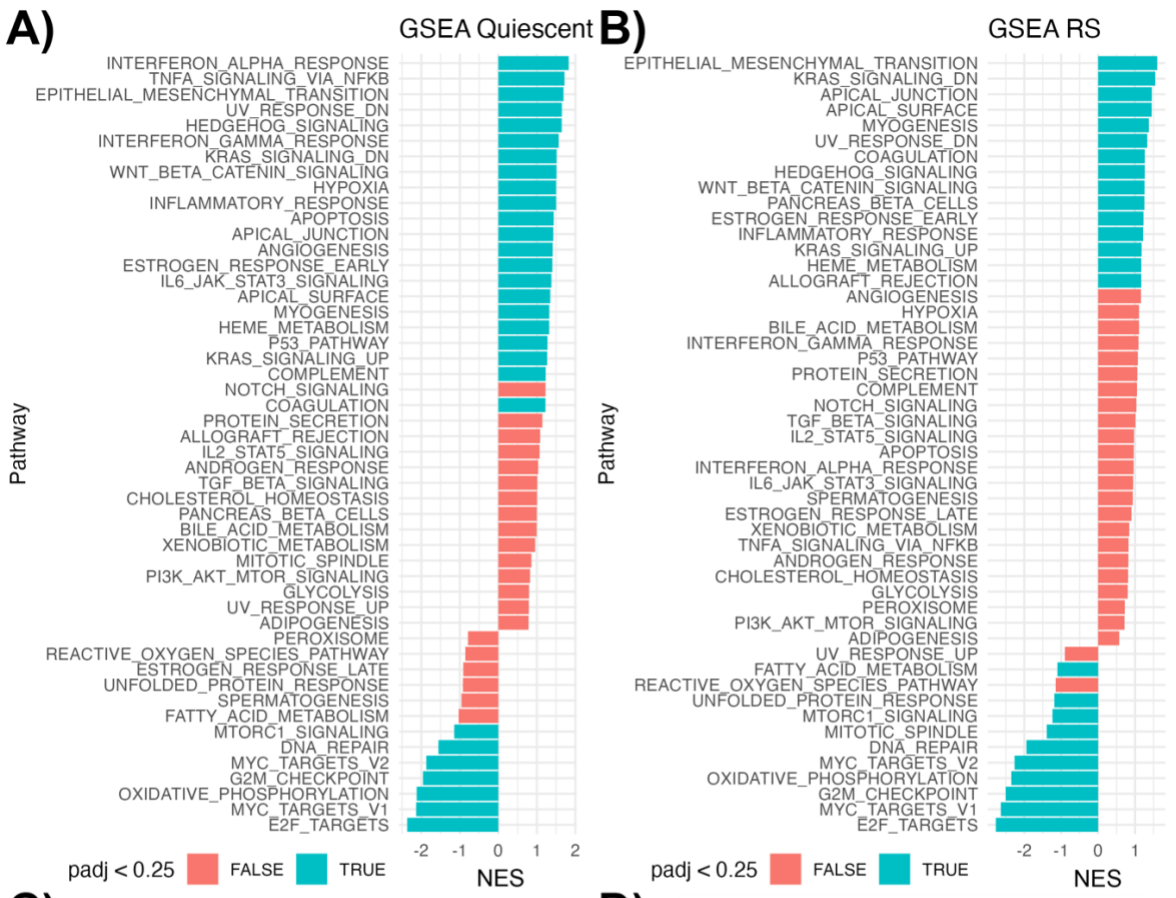
2.4.5: Complications of Using *In Vitro* Models of Quiescence

In addition to preparing high-resolution Hi-C maps of proliferating and RS cells we also created a map of quiescent LF1 cells with a resolution of 2.5 kb after combining the technical replicates, which clustered together (Supplementary Table 2.1, Supplementary Figure 2.8A). We intended to use quiescent cells as a control for senescence, as it represents a more natural physiological and temporary state of cell cycle arrest.

However, as we analyzed the quiescent data, specifically the RNA-Seq, we noticed some unexpected characteristics for quiescence. We performed Gene Set Enrichment Analysis Set Enrichment Analysis (GSEA) to get an idea of the functional effects of altered gene expression with these conditions (Figure 2.7A, B). This analysis of quiescent (as compared to proliferating) and RS (as compared to proliferating) turned up many expected terms. For example, E2F targets were downregulated in both quiescence and RS, supporting cell cycle arrest. Furthermore, p53 signaling was enriched in both quiescence and RS. However, we noticed that while many characteristic inflammatory terms (indicative of the SASP) were enriched in senescence, such as complement and TGF-beta, these were also enriched in quiescence, and arguably even more so. To further investigate this inflammatory phenotype, we also determined the overlap of the differentially expressed genes in quiescence and senescence with a curated list of SASP genes, which were overrepresented in both quiescence and senescence (fishers exact test with p-value < 0.01) (Figure 2.7C). The quiescent cells had more SASP genes upregulated than RS cells (551 v 480). As such they are poor controls for senescence, for if we used them as controls we would likely 'filter out' many of the inflammatory aspects of RS, which contribute to aging.

To further investigate if this inflammatory phenotype was unique to specific methods of quiescence induction we performed RT-qPCR on a panel of senescence markers and SASP genes. We compared proliferating cells, quiescent cells induced with contact inhibition, quiescent cells

induced with serum starvation, and senescent cells (Figure 2.7D). These results indicated that many of the SASP genes were upregulated in either or both types of quiescence as well as senescence, supporting the inflammatory phenotype of commonly used *in vitro* quiescence induction methods. Furthermore, in both contact inhibition and serum withdrawal induced quiescence we see the upregulation of genes considered senescence markers such as CDKN1A (p21) and CDKN2A (p16). We would like to draw attention to this, as quiescence is a commonly used control in the senescence field. In the future we caution that more physiological quiescence induction methods should be investigated and used so as to not confound a comparison to senescence. For this reason, we omitted quiescence as a control, and chose to analyze it separately.



	shared	unique	uncat.
A/B	1694	93	38
subcomp.	883	831	111
loops	147	1285	393

Figure 2.7. Complexities of using *in vitro* quiescence as a senescence control. A) GSEA of quiescent (as compared to proliferating) cells. B) GSEA of RS (as compared to proliferating) cells. C) Overlap between the SASP and genes upregulated with quiescence (as compared to proliferating) (left). Overlap between the SASP and genes upregulated with RS (as compared to proliferating) (right). P-values for venn diagrams are from Fisher's exact test. D) RT-qPCR of senescence marker genes and SASP genes. CI = quiescence induced with contact inhibition. Sen = senescence SW = serum withdrawal. E) Architecture surrounding 1825 cell cycle genes. Shared indicates same compartment, etc. whereas unique indicates a different compartmentalization, etc. uncat = uncategorized (gene was not clearly in a specific compartment, etc.).

Despite the inflammation, we wanted to compare the genomic structure surrounding cell cycle genes. We determined a list of cell cycle genes and investigated if the architecture of different levels of organization were shared or unique between quiescence and RS at each level (Figure 2.7E). At the level of A/B compartments, most of the cell cycle genes were within the same compartment, indicating that their architecture at this level is largely shared to repress cell cycle progression. However, at the subcompartment level about half of the architecture was shared and about half was unique. This indicates that there are differences in the way that cell cycle genes are controlled with subcompartment architecture between quiescence and senescence, but leading to the same net result of cell cycle arrest. A similar finding was true at the level of loops but to an even more extreme degree; the majority (70%) of cell cycle genes were in an altered loop landscape between quiescence and senescence, indicating their structure largely differs but attains the same cell cycle arrest.

Although it was analyzed separately, we also examined the trends with quiescence at each level of organization. The quiescent condition displays a decrease in both long- and short-range intra-chromosomal contacts (Supplementary Figure 2.8B). Because Hi-C measures contacts as either intra (within a chromosome) or inter (between chromosomes), this supports that there is an increase in inter-chromosomal contacts, and therefore the chromosomes are closer together with quiescence. Similarly to RS and OIS, quiescence also displays a shift to the A compartment (Supplementary Figure 2.9A). The only functional association with quiescence was in the genes switching towards the A compartment, and these were associated with response to lead ion (Supplementary Figure 2.9B). With quiescence the most subcompartment switching occurred from the A.1.1. subcompartment to the A.1.2. subcompartment (4.43%) (Supplementary Figure 2.10). Functionally many of these subcompartment switches support features of quiescence, such as

repression of mitotic nuclear division and G2/M transition, among others (Supplementary Table 2.4). In quiescence we identified 7062 loops. The downregulated genes they included were enriched for GO terms such as cell cycle DNA replication (Supplementary Figure 2.11A) and those upregulated were enriched for GO terms such as regulation of anatomical structure size (Supplementary Figure 2.11B).

2.5: Discussion

Genomic architecture is a burgeoning field, driving discovery in the complex interaction between the structure of the genome and gene expression. Senescence research is also an area of increasing attention, as it's clear this inflammatory phenotype is a driver of aging (Baker et al., 2016). This study represents an important advancement at the apex of these two fields, as we have created the highest resolution Hi-C map of the senescent genome (specifically RS) to date. This map has reinforced the loss of heterochromatin hypothesis of senescence, as there is compartment and subcompartment switching towards the A compartment and A-type subcompartments. These switches, along with the comparatively high number of loops specific to RS support functional aspects of senescence such as cell cycle arrest and inflammation. Many of these senescence-loops are hypomethylated with senescence, indicating a mechanism for their formation. Furthermore, we show these architectural changes support the derepression of LINE1 elements and to some degree the SASP, supporting age-associated inflammation. In this study we also present the highest resolution Hi-C map to date of quiescence.

Along with an analysis of RS, we also compared this type of senescence to OIS, using the highest resolution available Hi-C maps (>10kb) of OIS (Olan et al., 2020). Overall these types of senescence were similar, but had distinct differences. Both RS and OIS exhibit a gain of short-range and loss of long-range interactions. At the compartment level, both RS and OIS shift towards

the A compartment, but OIS to a lesser degree. Functionally this shift is associated with inflammation with RS, but not with OIS. With subcompartments there is also movement towards the A-type subcompartments, and more so with RS than OIS. There were some shared functional components at this level. Cell-cycle associated genes and lamin-associated genes were downregulated with both types of senescence. Inflammatory genes were upregulated with both types of senescence, however RS is associated with complement and OIS is associated with chemokines and interleukins. The SASP as well as LINE1 were more affected by compartment and subcompartment switching in RS, as compared to OIS.

The most striking and novel feature we saw in our results was the high number of loops with senescence, as compared to proliferating cells. This is the first time that loops have been examined in the context of RS, and thus the first time that this emergence of loops has been described. This finding is reminiscent of Olan et al.'s finding that there is an increase in *de novo* cohesin binding in OIS, driving the expression of important genes such as the inflammatory SASP genes IL1A and IL1B (Olan et al., 2020). Transcription was determined to be the driver of these cohesin loops, at it pushed the cohesin along the gene bodies, depositing it at the 3' end of the gene bodies. We were unable to examine if a similar finding occurred with our data, however, we determined a different mechanism of loop formation with RS. Many RS-specific loops lost methylation at both anchors, as compared to proliferating-specific loops and loops shared between proliferating and RS cells. Many of these loops did not have CTCF motifs at either anchor, indicating they could be cohesin loops, similar to Olan et al.'s findings.

Many of our study's findings echo recent work which revealed that there is epigenetic erosion with age (Yang et al., 2023). According to a study from the Sinclair lab, the root cause of this epigenetic drift seen with aging is the response to double-stranded breaks, leading the

epigenetic landscape to stray from its more faithful youthful form. Although the Sinclair lab's study examined aging through use of the ICE (Inducible Changes to the Epigenome) mouse model, and not senescence specifically, one may hypothesize that a similar phenomenon would occur with senescence, as it is caused by DNA damage. Reflecting this, our results show epigenetic drift at every level of organization that we examined, from loops to long- and short-range contacts. This is especially interesting at the level of loops, where our results indicating more unique loops with RS reflects the gain of promoter-enhancer contacts seen in ICE cells. Furthermore, our functional analysis of RS-specific loops revealed they lead to the upregulation of many neuron-associated genes. This supports the Sinclair lab's findings that the epigenetic drift associated with aging may lead cell types such as fibroblasts to resemble others, such as neurons.

This study is the first to establish an architectural hotspot supporting the derepression of a LINE1 element, specifically L1HS_14q23.2_3. Our establishment of this hotspot was enabled specifically by our creation of long-read DNA sequencing, high-resolution Hi-C, and total RNA-Seq for proliferating and senescent fibroblasts. This is the first documentation of architecture, specifically at the level of subcompartments and loops, contributing to the significant upregulation of a retrotransposon. This is especially impressive given that the individual LINE1 elements were identified by total RNA-Seq data created with standard short-read Illumina technology. It is possible that more of these hotspots exist, and future studies using long-read RNA-Seq will allow for further investigation in this area, cementing the link between architectural changes and LINE1 upregulation.

Our results also showed there was upregulation of TGF-beta through compartment switching towards the A compartment with RS. Many papers have established that TGF-beta signaling is an important marker of secondary senescence (Evans et al., 2023; Teo et al., 2019).

Although this has been more fully investigated in OIS and Notch-Induced Senescence (NIS) (Hoare et al., 2016; Parry et al., 2018), clustering of single-cell analysis has indicated primary and secondary populations of senescent cells with RS. This would be an important and interesting future direction for this analysis, as this high-resolution RS Hi-C map likely represents a combination of primary and secondary senescent cells and the ratio of these cell populations is not known. There are single-cell Hi-C approaches that could be used to differentiate which architectural features are attributed to primary vs. secondary senescence, but these single-cell methods are extremely limited in resolution (Nagano et al., 2013). In this case, bulk Hi-C examining each of the two populations with a co-culture or transwell system as in (Teo et al., 2019) would be advantageous and allow us to examine structural differences between primary and secondary cells.

The downregulation of lamin-associated genes seen with subcompartment switching in RS and OIS is of interest as well. A relevant study used polymer modeling to investigate architectural changes with senescence and found that varying heterochromatin and lamin interactions created a genomic architectural state resembling that of cellular senescence (Chiang et al., 2019). Other studies have shown that lamin depletion had profound effects on chromatin structure (Sadaie et al., 2013; Shah et al., 2013). This would be an interesting future direction as well, as performing Hi-C using a degron controlled system or an alternative to deplete the nuclear lamina would allow us to investigate the role senescence-associated lamin breakdown has on genome structure in senescence.

Although we present new and high-resolution datasets, this study does have limitations. With a resolution of ~3 kb, this study undoubtedly gives us the best picture of senescence compared to other published studies. However, there are other, newer tools such as micro-C

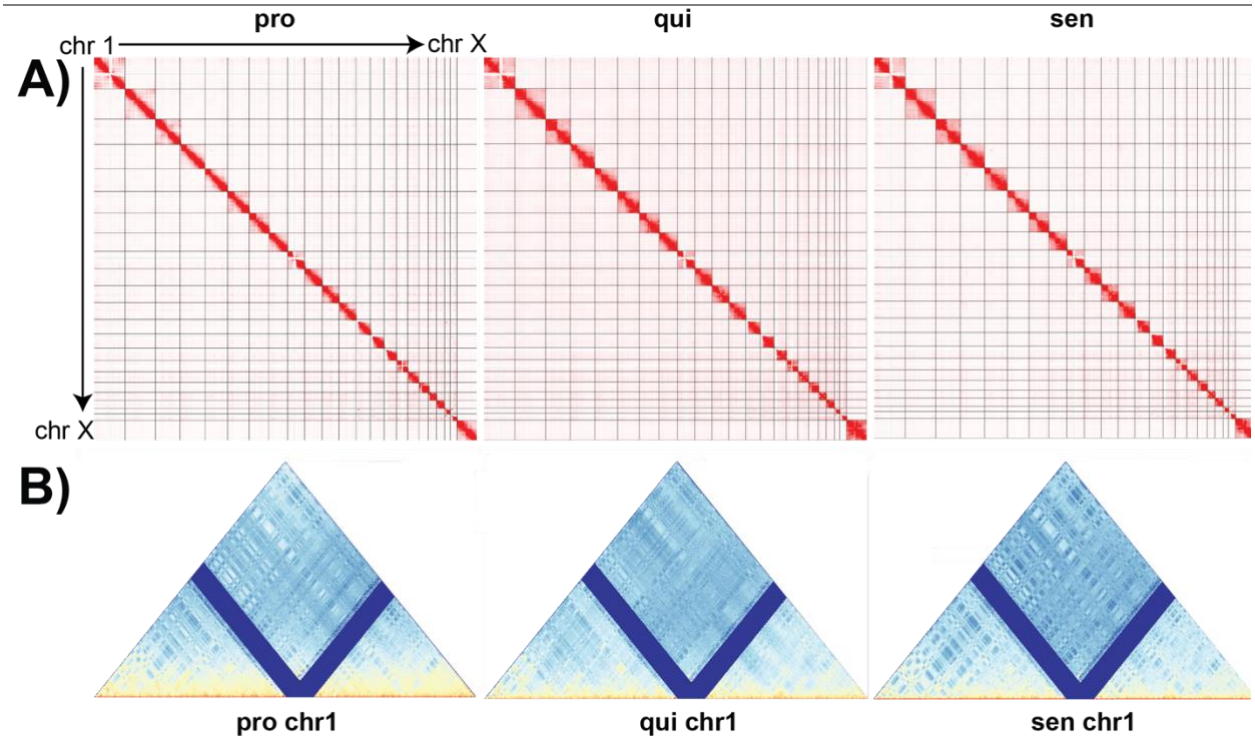
(Krietenstein et al., 2020) and capture Hi-C (Schoenfelder, Javierre, Furlan-Magaril, Wingett, & Fraser, 2018), which would permit more detailed analysis of genomic structure, especially at the level of loops. Although we were able to examine loops we were limited in our analysis by our 3 kb resolution, as the loop anchors were relatively large in size (~15-20 kb) making it difficult to identify precisely which genes were being regulated and how. In addition, we were also limited in attaining other features of interest, such as histone marks, transcription factor/protein localization (via CUT&RUN, for example), or chromatin accessibility due to the technical difficulties of working with late stage RS. Despite this, a strength of our study was our ability to examine late RS, which is technically challenging as this process takes months, and it is easy for cultures to run into technical problems during this time (contamination, excess cell death leading to sparsity, etc.). Another difficulty we encountered in this study was that our intended control, quiescence, was very inflammatory, and thus proved to be a poor control for senescence. However, our RT-qPCR analysis is instrumental, as it indicates this traditional senescence control should be used with caution.

Altogether, this study and the datasets it encompasses provide important insight into genomic architecture and its role in cellular senescence. The experiments and analysis we have performed pave the way for future, more mechanistic studies, which have the potential to extend human healthspan and lifespan.

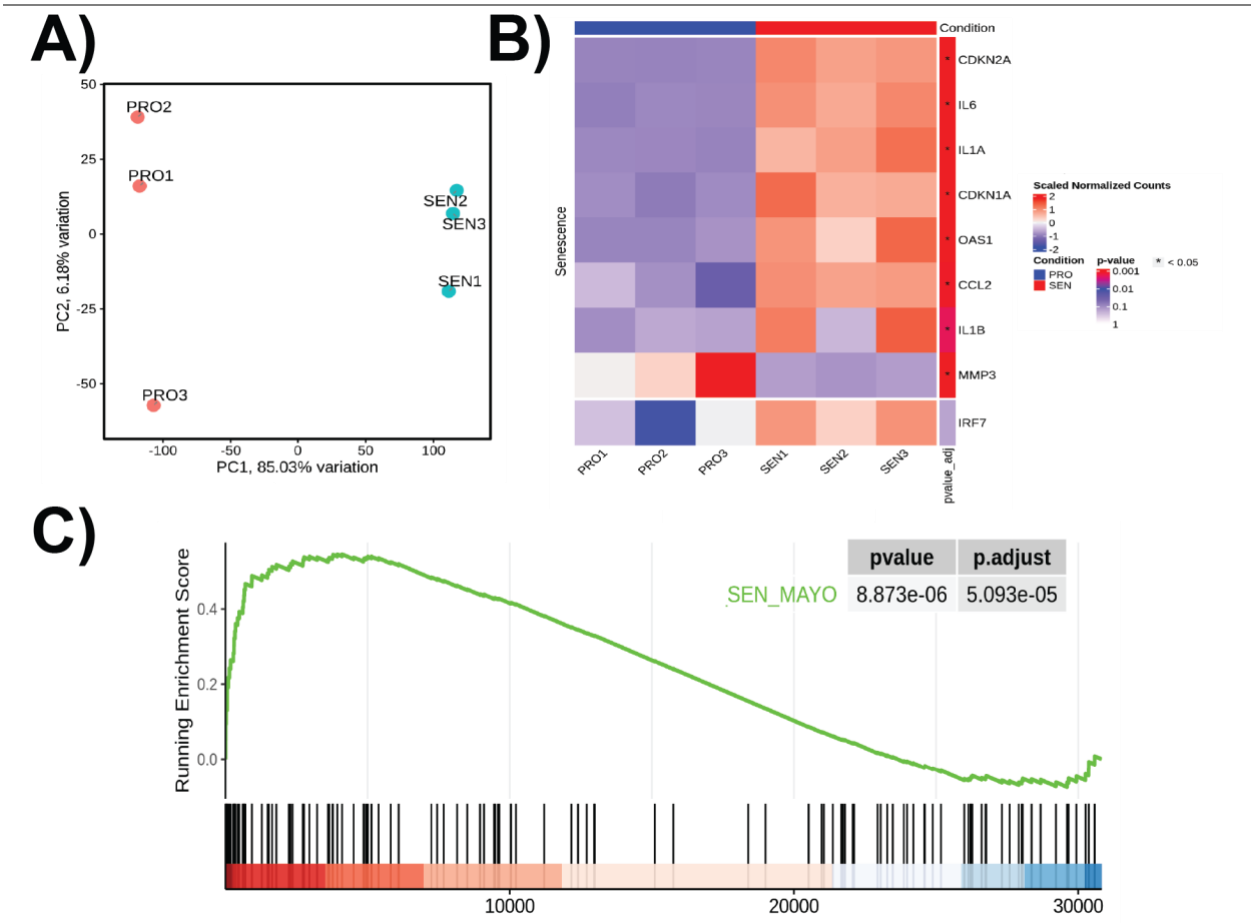
2.6: Supplementary Information

Sample Name	Replicate	Sequenced Reads	Hi-C Contacts
Proliferating	1	479,606,053	335,144,966
Proliferating	2	376,935,533	263,471,076
Proliferating	3	394,567,767	274,118,940
Quiescent	1	453,732,287	286,795,901
Quiescent	2	444,787,457	281,594,189
Quiescent	3	479,962,618	300,708,434
Senescent	1	508,313,932	316,312,022
Senescent	2	466,055,552	294,431,607
Senescent	3	449,124,662	304,263,231

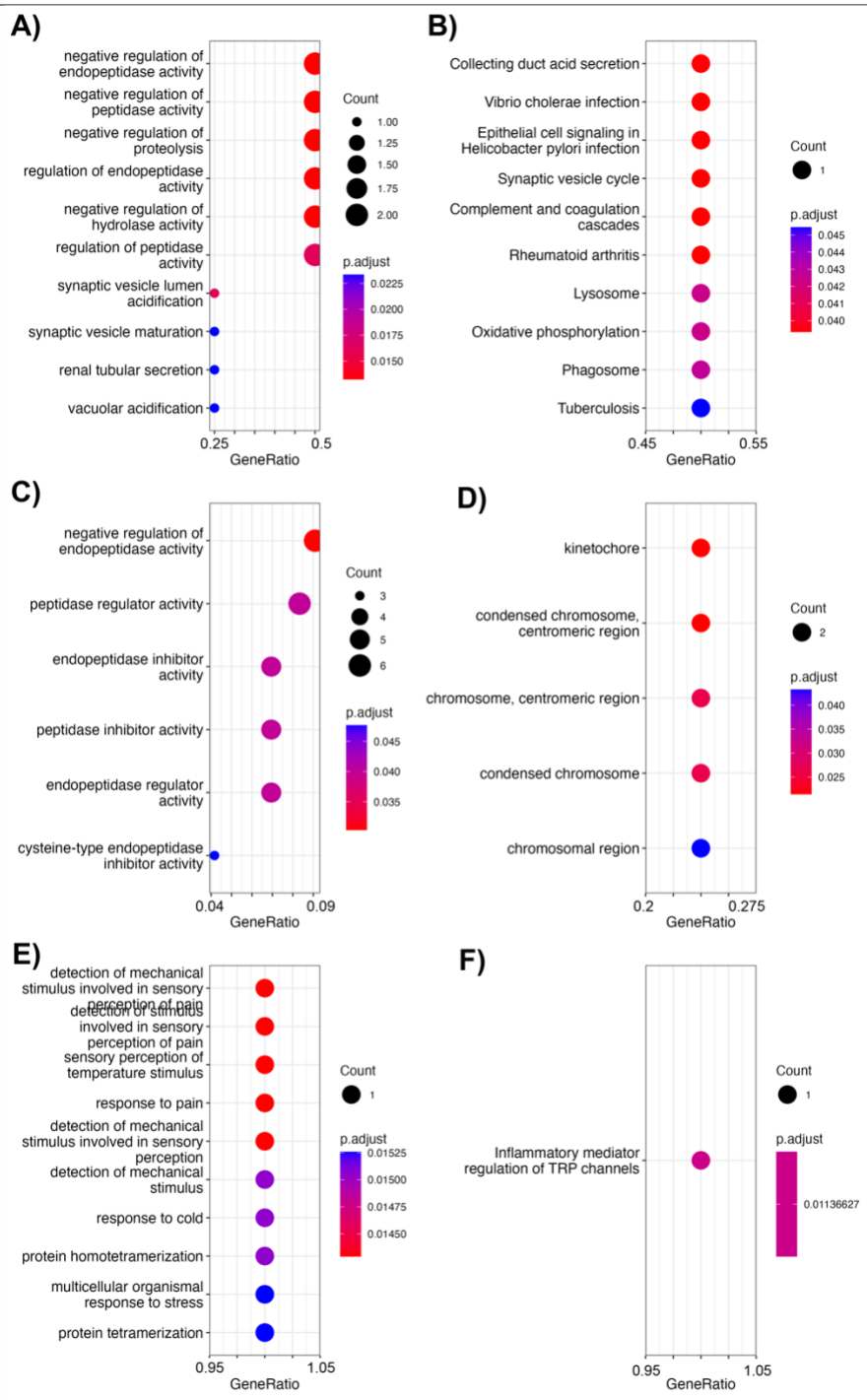
Supplementary Table 2.1. Number of sequenced reads and resulting Hi-C contacts per replicate.



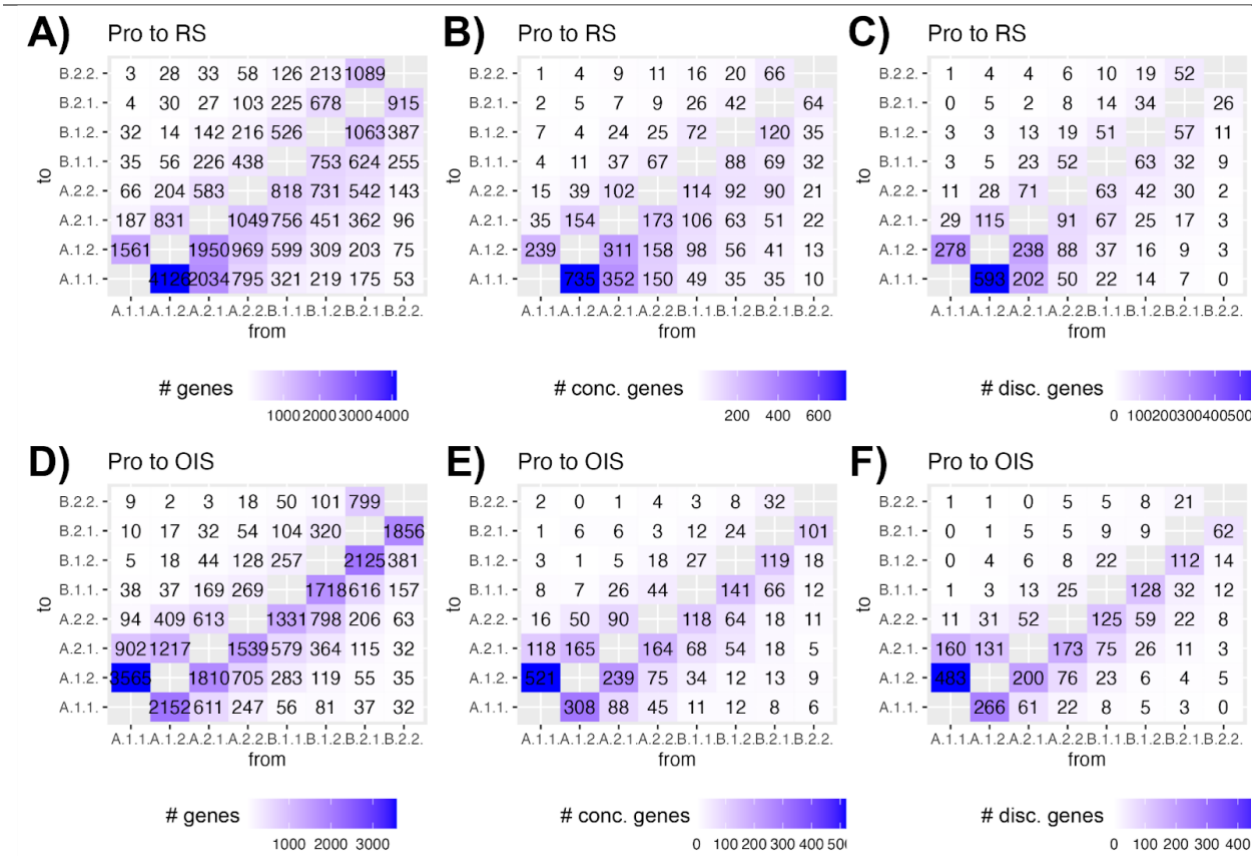
Supplementary Figure 2.1. A) Full genome visualizations of Hi-C maps of proliferating, quiescent, and senescent cells. B) Visualization of chromosome 1 Hi-C map for proliferating, quiescent, and senescent cells.



Supplementary Figure 2.2. A) PCA of RNA-Seq from proliferating and senescent replicates. B) Heatmap of canonical senescence markers and SASP genes. C) GSEA plot of SEN_MAYO genes.



Supplementary Figure 2.3. Additional functional analysis of proliferating to RS and OIS compartment switching. A) GO analysis of shared (RS and OIS) towards A compartment switching. B) KEGG analysis of shared (RS and OIS) towards A compartment switching. C) GO analysis of OIS only towards A compartment switching. D) GO analysis of OIS only towards B compartment switching. E) GO analysis of shared (RS and OIS) towards B compartment switching. F) KEGG analysis of shared (RS and OIS) towards B compartment switching.



Supplementary Figure 2.4. Genes undergoing subcompartment switching. A) Number of genes undergoing subcompartment switching with RS. B) Number of genes concordantly expressed undergoing subcompartment switching with RS. C) Number of genes discordantly expressed undergoing subcompartment switching with RS. D) Number of genes undergoing subcompartment switching with OIS. E) Number of genes concordantly expressed undergoing subcompartment switching with OIS. E) Number of genes discordantly expressed undergoing subcompartment switching with OIS.

Supplementary Table 2.2. Full list of GO terms enriched for regions undergoing subcompartment switching between proliferating and RS cells.

GO term	from	to
vacuolar membrane	A.1.2.	A.1.1.
lysosomal membrane	A.1.2.	A.1.1.
lytic vacuole membrane	A.1.2.	A.1.1.
adherens junction	A.1.2.	A.1.1.
lysosomal lumen	A.1.2.	A.1.1.
cell cortex	A.2.2.	A.1.1.

regulation of metal ion transport	B.1.1.	A.1.1.
regulation of calcineurin-NFAT signaling cascade	B.1.1.	A.1.1.
regulation of calcineurin-mediated signaling	B.1.1.	A.1.1.
calcineurin-NFAT signaling cascade	B.1.1.	A.1.1.
calcineurin-mediated signaling	B.1.1.	A.1.1.
ATPase binding	B.2.1.	A.1.1.
adipose tissue development	B.2.2.	A.1.1.
positive regulation of hippo signaling	B.2.2.	A.1.1.
membrane disassembly	B.2.2.	A.1.1.
locomotion involved in locomotor behavior	B.2.2.	A.1.1.
phosphatidylethanolamine acyl-chain remodeling	B.2.2.	A.1.1.
nuclear chromosome segregation	A.1.1.	A.1.2.
sister chromatid segregation	A.1.1.	A.1.2.
mitotic nuclear division	A.1.1.	A.1.2.
mitotic sister chromatid segregation	A.1.1.	A.1.2.
mitotic sister chromatid cohesion	A.1.1.	A.1.2.
intermediate filament	A.2.1.	A.1.2.
intermediate filament cytoskeleton organization	A.2.1.	A.1.2.
intermediate filament-based process	A.2.1.	A.1.2.
keratin filament	A.2.1.	A.1.2.
intermediate filament organization	A.2.1.	A.1.2.
atrial cardiac muscle tissue development	A.2.2.	A.1.2.
thyroid hormone metabolic process	A.2.2.	A.1.2.
cardiac pacemaker cell differentiation	A.2.2.	A.1.2.
cardiac muscle cell fate commitment	A.2.2.	A.1.2.
sinoatrial node development	A.2.2.	A.1.2.
muscle cell differentiation	B.1.1.	A.1.2.
postsynaptic membrane	B.1.1.	A.1.2.
homotypic cell-cell adhesion	B.1.1.	A.1.2.
heterophilic cell- cell adhesion via plasma membrane cell adhesion molecules	B.1.1.	A.1.2.
lateral plasma membrane	B.1.1.	A.1.2.
transmembrane transporter binding	B.1.2.	A.1.2.
actin filament	B.2.1.	A.1.2.
3 prime,5 prime cyclic-AMP phosphodiesterase activity	B.2.1.	A.1.2.
intraciliary transport particle	B.2.1.	A.1.2.
intraciliary transport particle	B.2.1.	A.1.2.
filamentous actin	B.2.1.	A.1.2.
complement receptor mediated signaling pathway	B.2.2.	A.1.2.

complement receptor activity	B.2.2.	A.1.2.
immune receptor activity	B.2.2.	A.1.2.
G protein-coupled peptide receptor activity	B.2.2.	A.1.2.
peptide receptor activity	B.2.2.	A.1.2.
nucleotide-excision repair	A.1.2.	A.2.1.
nuclear DNA-directed RNA polymerase complex	A.1.2.	A.2.1.
transcription factor TFII core complex	A.1.2.	A.2.1.
transcription factor TFIlH holo complex	A.1.2.	A.2.1.
carboxy-terminal domain protein kinase complex	A.1.2.	A.2.1.
regulation of neuron projection development	A.2.2.	A.2.1.
cell-substrate adhesion	A.2.2.	A.2.1.
cell-matrix adhesion	A.2.2.	A.2.1.
vacuole organization	A.2.2.	A.2.1.
negative regulation of actin filament bundle assembly	A.2.2.	A.2.1.
alpha-actinin binding	B.1.1.	A.2.1.
nuclear membrane	B.1.2.	A.2.1.
basal plasma membrane	B.1.2.	A.2.1.
basal part of cell	B.1.2.	A.2.1.
main axon	B.1.2.	A.2.1.
connexin complex	B.1.2.	A.2.1.
regulation of monoatomic ion transmembrane transport	B.2.1.	A.2.1.
cation channel complex	B.2.1.	A.2.1.
ion channel complex	B.2.1.	A.2.1.
transmembrane transporter complex	B.2.1.	A.2.1.
voltage-gated calcium channel complex	B.2.1.	A.2.1.
integral component of endoplasmic reticulum membrane	B.2.2.	A.2.1.
intrinsic component of endoplasmic reticulum membrane	B.2.2.	A.2.1.
main axon	B.2.2.	A.2.1.
neuromuscular junction	B.2.2.	A.2.1.
guanyl-nucleotide exchange factor activity	A.1.2.	A.2.2.
GTPase regulator activity	A.1.2.	A.2.2.
GTPase regulator activity	A.1.2.	A.2.2.
catalytic activity, acting on RNA	A.2.1.	A.2.2.
single-stranded DNA binding	A.2.1.	A.2.2.
regulation of membrane potential	B.1.1.	A.2.2.
membrane repolarization	B.1.1.	A.2.2.
membrane depolarization during cardiac muscle cell action potential	B.1.1.	A.2.2.
membrane depolarization during action potential	B.1.1.	A.2.2.
regulation of membrane repolarization	B.1.1.	A.2.2.

postsynaptic specialization organization	B.1.2.	A.2.2.
amyloid-beta binding	B.1.2.	A.2.2.
regulation of NMDA receptor activity	B.1.2.	A.2.2.
synapse organization	B.2.1.	A.2.2.
calcium channel complex	B.2.1.	A.2.2.
voltage-gated calcium channel complex	B.2.1.	A.2.2.
postsynaptic specialization organization	B.2.1.	A.2.2.
receptor localization to synapse	B.2.1.	A.2.2.
dendrite cytoplasm	B.2.2.	A.2.2.
maturation of LSU-rRNA from tricistronic rRNA transcript (SSU-rRNA, 5.8S rRNA,LSU-rRNA)	A.1.1.	B.1.1.
pentose-phosphate shunt	A.1.1.	B.1.1.
glycerol metabolic process	A.1.1.	B.1.1.
dolichol-linked oligosaccharide biosynthetic process	A.1.1.	B.1.1.
oligosaccharide-lipid intermediate biosynthetic process	A.1.1.	B.1.1.
DNA replication initiation	A.1.2.	B.1.1.
peptidase activator activity involved in apoptotic process	A.1.2.	B.1.1.
peptidase activator activity	A.1.2.	B.1.1.
peptidase regulator activity	A.1.2.	B.1.1.
dipeptidyl-peptidase activity	A.1.2.	B.1.1.
tRNA metabolic process	A.2.2.	B.1.1.
tRNA processing	A.2.2.	B.1.1.
chromosome, centromeric region	A.2.2.	B.1.1.
tRNA modification	A.2.2.	B.1.1.
mitochondrial respiratory chain complex assembly	A.2.2.	B.1.1.
positive regulation of protein kinase activity	B.1.2.	B.1.1.
regulation of protein serine/threonine kinase activity	B.1.2.	B.1.1.
regulation of MAP kinase activity	B.1.2.	B.1.1.
positive regulation of protein serine/threonine kinase activity	B.1.2.	B.1.1.
positive regulation of MAP kinase activity	B.1.2.	B.1.1.
ABC-type transporter activity	B.2.2.	B.1.1.
ATPase-coupled transmembrane transporter activity	B.2.2.	B.1.1.
lipid transporter activity	B.2.2.	B.1.1.
primary active transmembrane transporter activity	B.2.2.	B.1.1.
complement receptor activity	B.2.2.	B.1.1.
cleavage furrow	A.1.1.	B.1.2.
cell division site	A.1.1.	B.1.2.
lamin binding	A.1.1.	B.1.2.
acylglycerol lipase activity	A.1.1.	B.1.2.

nucleoside diphosphate phosphatase activity	A.1.1.	B.1.2.
microtubule anchoring at centrosome	A.1.2.	B.1.2.
regulation of DNA-directed DNA polymerase activity	A.1.2.	B.1.2.
positive regulation of DNA-directed DNA polymerase-activity	A.1.2.	B.1.2.
inner cell mass cell proliferation	A.1.2.	B.1.2.
microtuble anchoring at microtubule organizing center	A.1.2.	B.1.2.
ATP hydrolysis activity	A.2.1.	B.1.2.
regulation of sister chromatid cohesion	A.2.2.	B.1.2.
tRNA metabolic process	B.1.1.	B.1.2.
DNA duplex unwinding	B.1.1.	B.1.2.
DNA geometric change	B.1.1.	B.1.2.
DNA conformation change	B.1.1.	B.1.2.
postsynaptic specialization	B.2.1.	B.1.2.
synaptic membrane	B.2.1.	B.1.2.
postsynaptic density	B.2.1.	B.1.2.
asymmetric synapse	B.2.1.	B.1.2.
postsynaptic membrane	B.2.1.	B.1.2.
outflow tract septum morphogenesis	B.2.2.	B.1.2.
cardiac septum morphogenesis	B.2.2.	B.1.2.
serotonin uptake	B.2.2.	B.1.2.
polyamine transmembrane transport	B.2.2.	B.1.2.
histamine transport	B.2.2.	B.1.2.
mitochondrial protein processing	A.1.1.	B.2.1.
proton-transporting two-sector ATPase complex assembly	A.1.1.	B.2.1.
double-strand break repair via nonhomologous end joining	A.1.1.	B.2.1.
protein processing	A.1.1.	B.2.1.
double-strand break repair	A.1.1.	B.2.1.
protein polyubiquitination	A.1.2.	B.2.1.
double-strand break repair	A.1.2.	B.2.1.
transcription-coupled • nucleotide-excision repair	A.1.2.	B.2.1.
single strand break repair	A.1.2.	B.2.1.
double-strand break repair o via break-induced replication	A.1.2.	B.2.1.
CGMP binding	A.2.1.	B.2.1.
oxidoreductase activity, acting on the CH-NH2 group of donors, oxygen as acceptor	A.2.1.	B.2.1.
oxidoreductase activity, acting on the CH-NH2 group of donors	A.2.1.	B.2.1.
CoA hydrolase activity	A.2.1.	B.2.1.
mitogen-activated protein kinase binding	A.2.1.	B.2.1.
nuclear replisome	A.2.2.	B.2.1.

replisome	A.2.2.	B.2.1.
DNA polymerase complex	A.2.2.	B.2.1.
nuclear replication fork	A.2.2.	B.2.1.
acetylcholine receptor regulator activity	A.2.2.	B.2.1.
synapse assembly	B.2.2.	B.2.1.
cardiac muscle tissue development	B.2.2.	B.2.1.
striated muscle tissue development	B.2.2.	B.2.1.
kidney development	B.2.2.	B.2.1.
regulation of synapse assembly	B.2.2.	B.2.1.
mitochondrial electron transport. NADH to ubiquinone	A.1.1.	B.2.2.
NADH dehydrogenase complex assembly	A.1.1.	B.2.2.
mitochondrial respiratory chain complex I assembly	A.1.1.	B.2.2.
proton motive force-driven mitochondrial ATP synthesis	A.1.1.	B.2.2.
proton motive force-driven ATP synthesis	A.1.1.	B.2.2.
luteinizing hormone secretion	A.1.2.	B.2.2.
gonadotropin secretion	A.1.2.	B.2.2.
asymmetric cell division	A.1.2.	B.2.2.
thyroid hormone generation	A.1.2.	B.2.2.
positive regulation of steroid biosynthetic process	A.1.2.	B.2.2.
apical part of cell	A.2.2.	B.2.2.
positive regulation of DNA-templated DNA replication	B.1.1.	B.2.2.
positive regulation of DNA replication	B.1.1.	B.2.2.
regulation of DNA-templated DNA replication	B.1.1.	B.2.2.
regionalization	B.1.2.	B.2.2.
atrioventricular canal development	B.1.2.	B.2.2.
cardiac chamber formation	B.1.2.	B.2.2.
pericardium development	B.1.2.	B.2.2.
organ induction	B.1.2.	B.2.2.
post-mRNA release spliceosomal complex	B.2.1.	B.2.2.

Supplementary Table 2.3. Full list of GO terms enriched for regions undergoing subcompartment switching between proliferating and OIS cells.

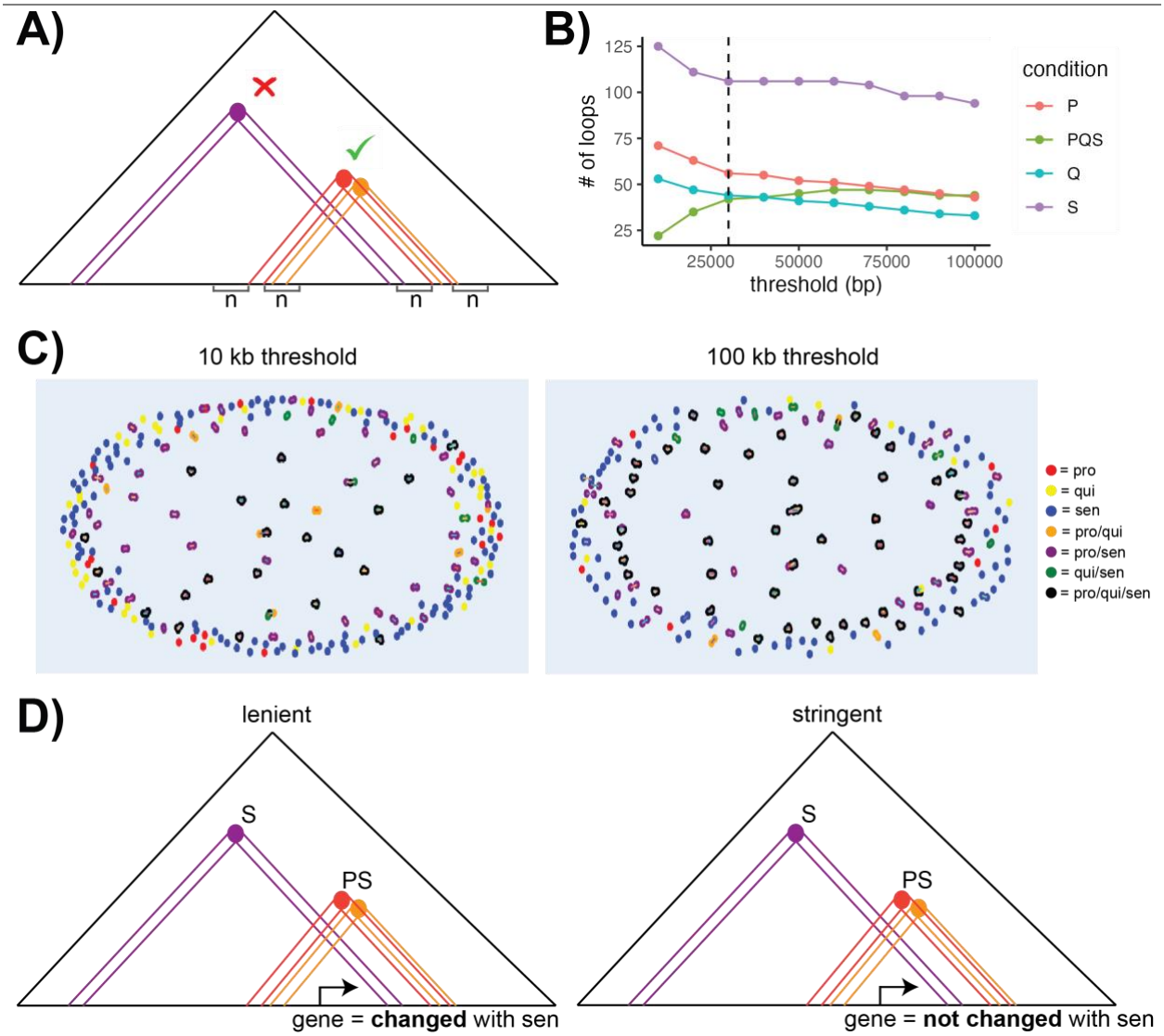
GO term	from	to
chemokine activity	A.1.2.	A.1.1.
CXCR chemokine receptor binding	A.1.2.	A.1.1.
IMP biosynthetic process	B.1.2.	A.1.1.
AMP biosynthetic process	B.1.2.	A.1.1.
GMP biosynthetic process	B.1.2.	A.1.1.
IMP metabolic process	B.1.2.	A.1.1.
oxidative demethylation	B.1.2.	A.1.1.
potassium:proton antiporter activity	B.2.1.	A.1.1.
sodium:proton antiporter activity	B.2.1.	A.1.1.
solute:potassium antiporter activity	B.2.1.	A.1.1.
basic amino acid transmembrane transporter activity	B.2.1.	A.1.1.
opsonin binding	B.2.1.	A.1.1.
nuclear cyclin-dependent protein kinase holoenzyme complex	B.2.2.	A.1.1.
transcription factor TFIIH core complex	B.2.2.	A.1.1.
transcription factor TFIIH holo complex	B.2.2.	A.1.1.
carboxy-terminal domain protein kinase complex	B.2.2.	A.1.1.
extracellular matrix	A.1.1.	A.1.2.
extracellular structure organization	A.1.1.	A.1.2.
external encapsulating structure organization	A.1.1.	A.1.2.
mitotic cell cycle phase transition	A.1.1.	A.1.2.
cell-substrate adhesion	A.1.1.	A.1.2.
oxidoreductase activity, acting on the CH-CH group of donors	A.2.2.	A.1.2.
contractile actin filament bundle assembly	B.2.2.	A.1.2.
stress fiber assembly	B.2.2.	A.1.2.
actin cytoskeleton reorganization	B.2.2.	A.1.2.
actin filament bundle assembly	B.2.2.	A.1.2.
actin filament bundle organization	B.2.2.	A.1.2.
endoplasmic reticulum lumen	A.1.1.	A.2.1.
cell projection membrane	A.1.1.	A.2.1.
basement membrane	A.1.1.	A.2.1.
DNA-templated DNA replication	A.1.2.	A.2.1.
preribosome	A.2.2.	A.2.1.
intermediate filament organization	B.1.2.	A.2.1.
keratinization	B.1.2.	A.2.1.
intermediate filament cytoskeleton organization	B.1.2.	A.2.1.
intermediate filament-based process	B.1.2.	A.2.1.

keratin filament	B.1.2.	A.2.1.
active monoatomic ion transmembrane transporter activity	B.2.1.	A.2.1.
active transmembrane transporter activity	B.2.1.	A.2.1.
ATPase-coupled monoatomic cation transmembrane transporter activity	B.2.1.	A.2.1.
ATPase-coupled transmembrane transporter activity	B.2.1.	A.2.1.
proton transmembrane transporter activity	B.2.1.	A.2.1.
sulfur compound metabolic process	B.2.2.	A.2.1.
adenosine metabolic process	B.2.2.	A.2.1.
atrial cardiac muscle cell membrane repolarization	B.2.2.	A.2.1.
G protein-coupled adenosine receptor signaling pathway	B.2.2.	A.2.1.
heparan sulfate proteoglycan biosynthetic process, enzymatic modification	B.2.2.	A.2.1.
mitotic cell cycle checkpoint signaling	A.1.1.	A.2.2.
cell cycle checkpoint signaling	A.1.1.	A.2.2.
mitotic spindle assembly checkpoint signaling	A.1.1.	A.2.2.
spindle assembly checkpoint signaling	A.1.1.	A.2.2.
mitotic spindle checkpoint signaling	A.1.1.	A.2.2.
actin binding	A.1.2.	A.2.2.
nuclear receptor activity	A.1.2.	A.2.2.
ligand-activated transcription factor activity	A.1.2.	A.2.2.
coreceptor activity	A.1.2.	A.2.2.
positive regulation of cell projection organization	A.2.1.	A.2.2.
catalytic activity, acting on a glycoprotein	A.2.1.	A.2.2.
interleukin-1 receptor binding	B.1.1.	A.2.2.
phosphorus-oxygen lyase activity	B.1.1.	A.2.2.
cytoskeletal anchor activity	B.1.1.	A.2.2.
leading edge membrane	B.2.1.	A.2.2.
cell projection membrane	B.2.1.	A.2.2.
cell leading edge	B.2.1.	A.2.2.
transport vesicle	B.2.1.	A.2.2.
ruffle membrane	B.2.1.	A.2.2.
negative regulation of cell adhesion	A.1.1.	B.1.1.
X chromsome	A.1.1.	B.1.1.
semaphorin receptor complex	A.1.1.	B.1.1.
nuclear lamina	A.1.1.	B.1.1.
COP9 signalosome	A.1.1.	B.1.1.
aldehyde-lyase activity	A.1.2.	B.1.1.
mannosyltranferse activity	A.1.2.	B.1.1.

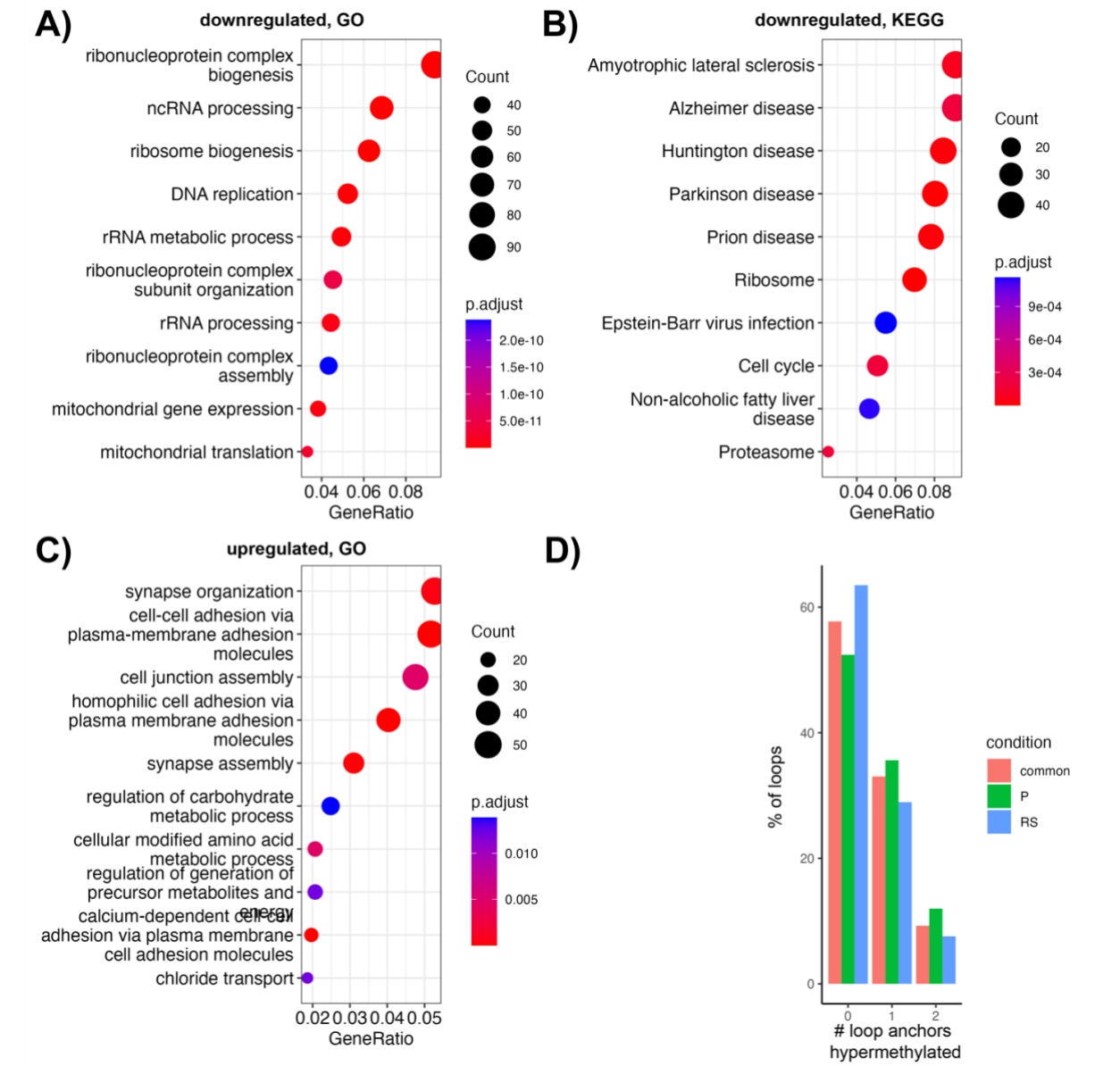
carbon-carbon lyase activity	A.1.2.	B.1.1.
pyridoxal phosphate binding	A.1.2.	B.1.1.
pericentriolar material	A.1.2.	B.1.1.
microtubule	A.2.1.	B.1.1.
spindle microtubule	A.2.1.	B.1.1.
centriole	A.2.1.	B.1.1.
mitotic spindle	A.2.1.	B.1.1.
constitutive heterochromatin formation	A.2.2.	B.1.1.
chromatin-protein adaptor activity	A.2.2.	B.1.1.
T-tubule	B.1.2.	B.1.1.
transmembrane receptor protein kinase activity	B.2.1.	B.1.1.
U2-type prespliceosome assembly	A.1.1.	B.1.2.
negative regulation of axon extensive involved in axon guidance	A.1.1.	B.1.2.
ventricular system development	A.1.1.	B.1.2.
regulation of axon extension involved in axon guidance	A.1.1.	B.1.2.
axon extension involved in axon guidance	A.1.1.	B.1.2.
positive regulation of Ras protein signal transduction	A.1.2.	B.1.2.
positive regulation of small GTPase mediated signal transduction	A.1.2.	B.1.2.
fibroblast growth factor receptor signaling pathway	A.1.2.	B.1.2.
cellular response to fibroblast growth factor stimulus	A.1.2.	B.1.2.
response to fibroblast growth factor	A.1.2.	B.1.2.
single fertilization	A.2.1.	B.1.2.
fertilization	A.2.1.	B.1.2.
negative regulation of humoral response mediated by circulating imm unoglobulin	A.2.1.	B.1.2.
sequestering of extracellular ligand from receptor	A.2.1.	B.1.2.
regulation of memory T cell differentiation	A.2.1.	B.1.2.
regulation of viral entry into host cell	A.2.2.	B.1.2.
modulation by symbiont of entry into host	A.2.2.	B.1.2.
regulation of biological process involved in symbiotic interaction	A.2.2.	B.1.2.
regulation of viral life cycle	A.2.2.	B.1.2.
positive regulation of autophagy	A.2.2.	B.1.2.
regulation of synapse organization	B.1.1.	B.1.2.
receptor localization to synapse	B.1.1.	B.1.2.
regulation of behavior	B.1.1.	B.1.2.
protein localization to synapse	B.1.1.	B.1.2.
neurotransmitter-gated ion channel clustering	B.1.1.	B.1.2.
heart contraction	B.2.1.	B.1.2.
heart process	B.2.1.	B.1.2.

MAP kinase kinase activity	B.2.1.	B.1.2.
G1 to G0 transition	A.1.1.	B.2.1.
neural precursor cell proliferation	A.1.1.	B.2.1.
peripheral nervouse system axon regeneration	A.1.2.	B.2.1.
bone trabecula formation	A.1.2.	B.2.1.
ether lipid biosynthetic process	A.1.2.	B.2.1.
deoxyribonucleoside metabolic process	A.1.2.	B.2.1.
glycerol ether biosynthetic process	A.1.2.	B.2.1.
regulation of fat cell proliferation	A.2.1.	B.2.1.
fat cell proliferation	A.2.1.	B.2.1.
telomere maintenance via recombination	A.2.1.	B.2.1.
regulation of respiratory system process	A.2.1.	B.2.1.
mitochondrion distribution	A.2.1.	B.2.1.
ketone body metabolic process	A.2.2.	B.2.1.
regulation of secondary metabolic process	A.2.2.	B.2.1.
short-chain fatty acid metabolic process	A.2.2.	B.2.1.
regulation of melanin biosynthetic process	A.2.2.	B.2.1.
regulation of secondary metabolite biosynthetic process	A.2.2.	B.2.1.
regulation of cGMP-mediated signaling	B.1.1.	B.2.1.
cGMP-mediated signaling	B.1.1.	B.2.1.
energy homeostatsis	B.1.1.	B.2.1.
cGMP binding	B.1.1.	B.2.1.
3 prime 5 prime cyclic-GMP phosphodiesterase activity	B.1.1.	B.2.1.
laminin binding	B.1.2.	B.2.1.
neurotransmitter transport	B.2.2.	B.2.1.
detection of external stimulus	B.2.2.	B.2.1.
detection of abiotic stimulus	B.2.2.	B.2.1.
polyamine transport	B.2.2.	B.2.1.
detection of temperature stimulus	B.2.2.	B.2.1.
positive regulation of GTPase activity	A.1.1.	B.2.2.
regulation of small GTPase mediated signal transduction	A.1.1.	B.2.2.
regulation of GTPase activity	A.1.1.	B.2.2.
small GTPase mediated signal transduction	A.1.1.	B.2.2.
TRAIL-activated apoptotic signaling pathway	A.2.1.	B.2.2.
positive regulation of extrinsic apoptotic signaling pathway via death domain receptors	A.2.1.	B.2.2.
negative regulation of metallopeptidase activity	A.2.1.	B.2.2.
regulation of metalloendopeptidase acitvity	A.2.1.	B.2.2.
regulation of membrane protein extodomain proteolysis	A.2.1.	B.2.2.

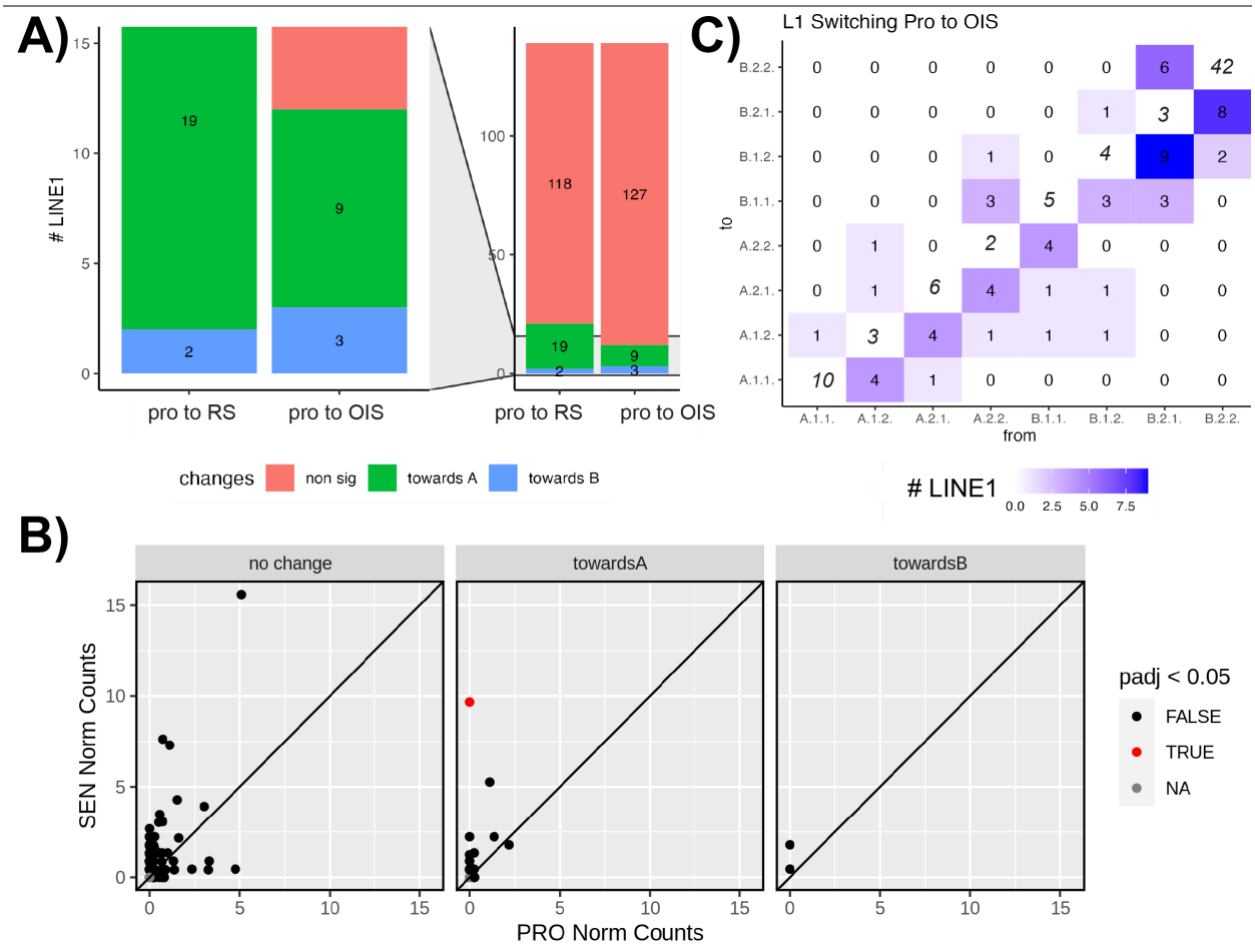
proteasome assembly	A.2.2.	B.2.2.
positive regulation of cytosolic calcium ion concentration involved in phospholipase C-activating G protein-coupled signaling pathway	A.2.2.	B.2.2.
positive regulation of Rho protein signal transduction	A.2.2.	B.2.2.
glycogen biosynthetic process	A.2.2.	B.2.2.
glucan biosynthetic process	A.2.2.	B.2.2.
protein oxidation	B.1.1.	B.2.2.
resolution of meiotic recombination intermediates	B.1.1.	B.2.2.
regulation of single stranded viral RNA replication via double strande d intermediate	B.1.1.	B.2.2.
single stranded viral RNA replication via double stranded intermediat e	B.1.1.	B.2.2.
viral RNA genome replication	B.1.1.	B.2.2.
sarcoplasm	B.1.2.	B.2.2.
S-methyltransferase activity	B.1.2.	B.2.2.
dystrophin-associated glycoprotein complex	B.1.2.	B.2.2.
protein phosphatase type 1 complex	B.1.2.	B.2.2.
glycoprotein complex	B.1.2.	B.2.2.
neuron recognition	B.2.1.	B.2.2.
neuron projection terminus	B.2.1.	B.2.2.
costamere	B.2.1.	B.2.2.



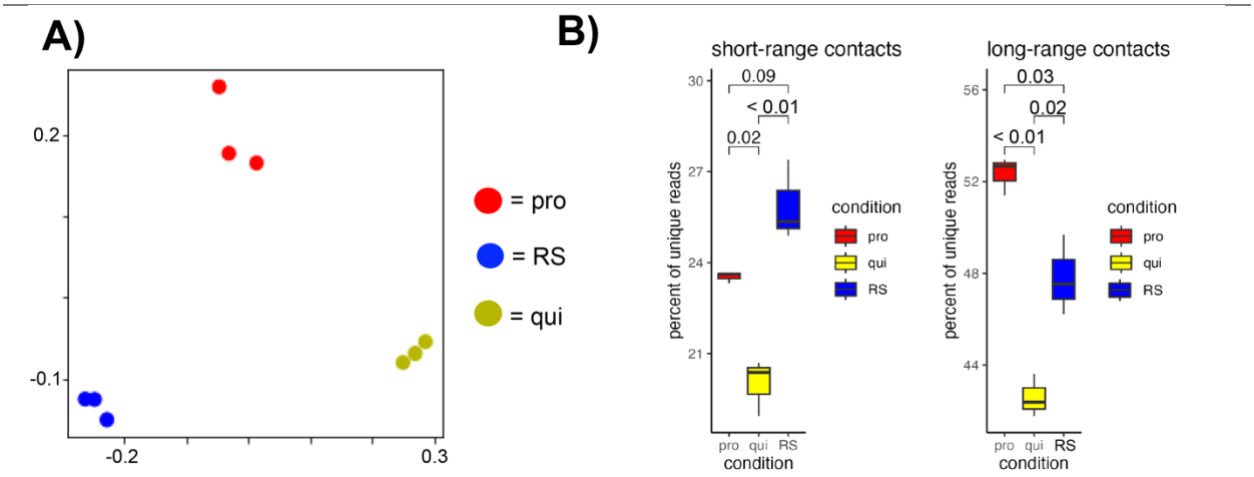
Supplementary Figure 2.5. Methodological details of loop analysis. A) Visualization of custom loop comparison script. Considering a loop (red) we compare it to other loops (purple, orange). We establish a window of +/- n around the loop base. If another loop falls within this window at both bases (orange) we call it the same, and if not it is a different loop (purple). B) Thresholding analysis of chromosome 21 loops. P = pro, Q = qui, S = sen, PQS = pro/qui/sen. Dotted line is at 30,000 bp. C) Graph visualization of thresholding analysis from B. D) Visualization of lenient (more loops altered) and stringent (less loops altered) criteria for loop analysis.



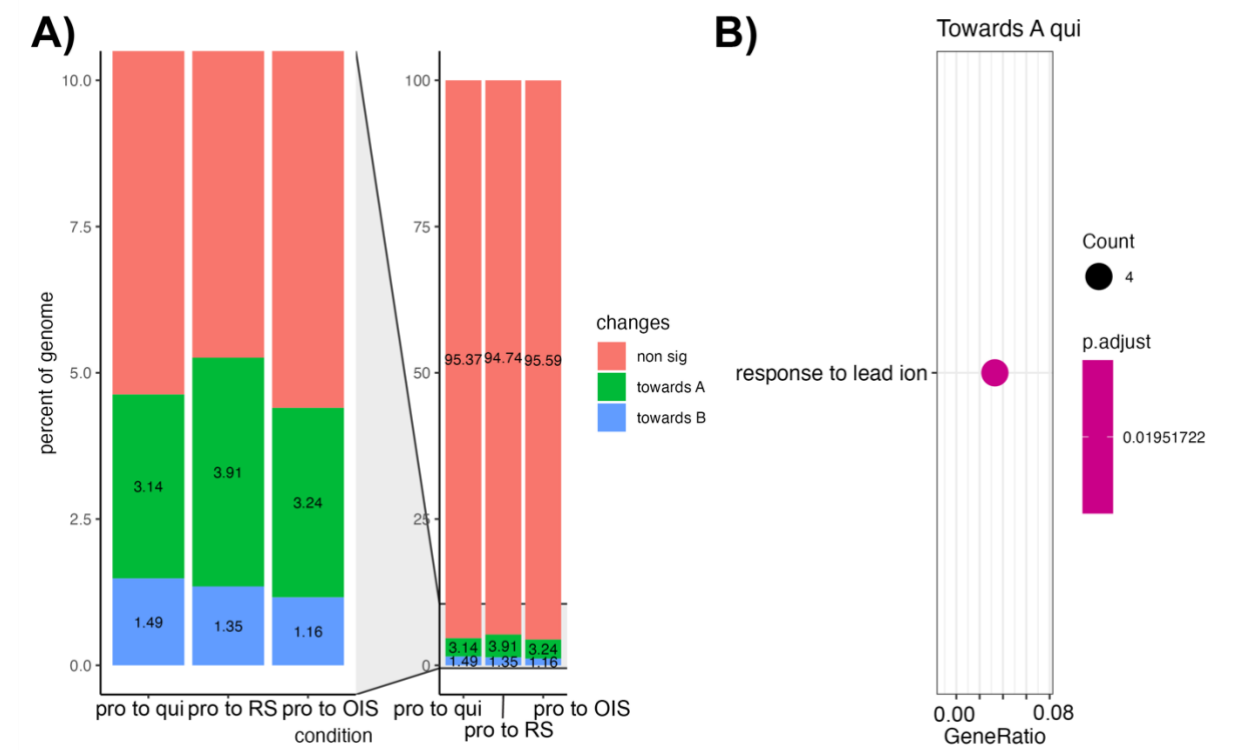
Supplementary Figure 2.6. Additional functional analysis of loops (stringent criteria). A) GO analysis of genes downregulated within senescence-altered loops. B) KEGG analysis of genes downregulated within senescence-altered loops. C) GO analysis of genes upregulated within senescence-altered loops. D) Percent of loops with methylation gained at 0/1/2 loop anchors.



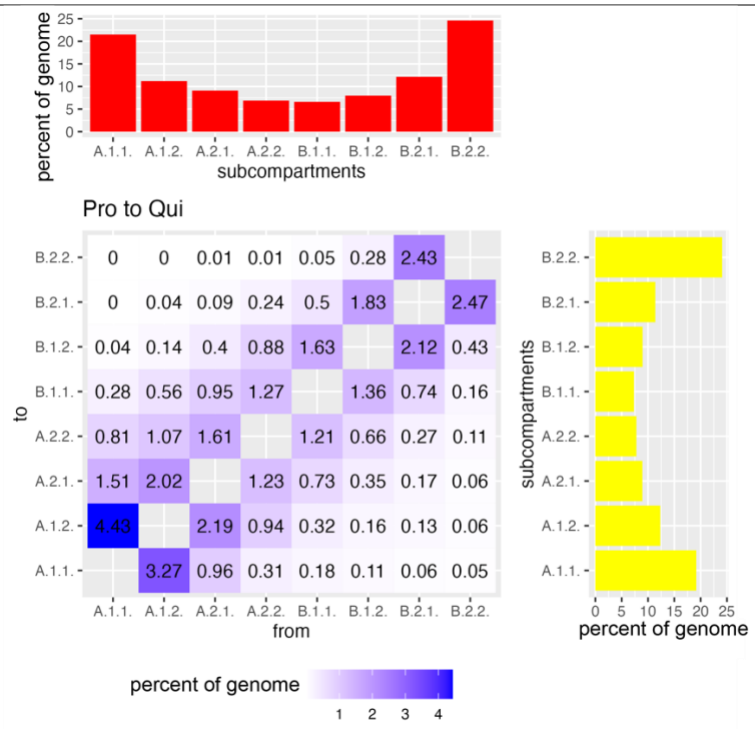
Supplementary Figure 2.7. A) Compartment switching of LINE1 elements with RS and OIS. B) Gene expression analysis of LINE1 elements switching compartments with RS. C) Subcompartment switching of LINE1 with OIS.



Supplementary Figure 2.8. Quality Control for Hi-C Maps with quiescence. A) Multidimensional scaling plot of proliferating, quiescent, and RS replicates. B) Short-Range and long-range intra-chromosomal interactions for proliferating, quiescent, and RS cells. Statistics represent p-value outputs from student's t-test.



Supplementary Figure 2.9. Functional analysis of compartment switching with quiescence. A) Compartment switching with quiescence, RS, and OIS. B) GO analysis of genes moving towards the A compartment with quiescence.



Supplementary Figure 2.10. Subcompartment switching with quiescence. Percent of genome switching subcompartments from proliferating (columns) to quiescence (rows). Marginal histograms show distribution in each subcompartment for proliferating (top, red) and quiescence (right, yellow).

Supplementary Table 2.4. Full list of GO terms enriched for regions undergoing subcompartment switching between proliferating and quiescent cells.

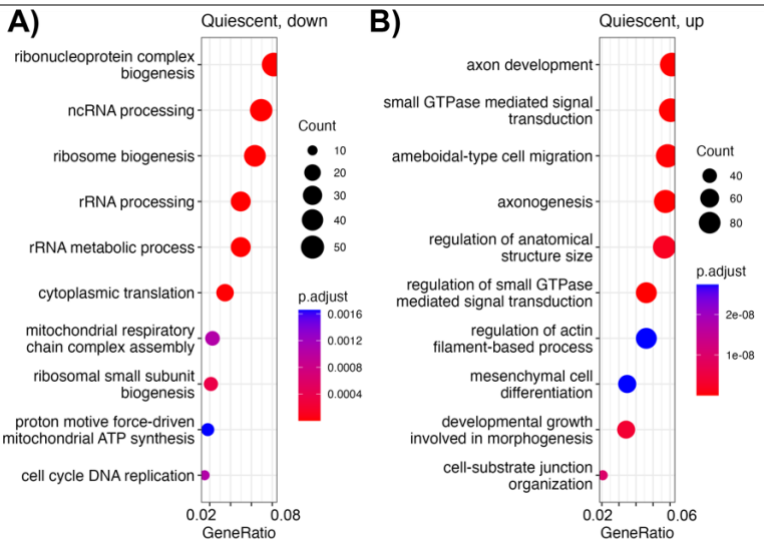
GO terms	from	to
epithelial cell proliferation	A.1.2.	A.1.1.
regulation of epithelial cell proliferation	A.1.2.	A.1.1.
cell-cell adhesion via plasma-membrane adhesion molecules	A.1.2.	A.1.1.
homophilic cell adhesion via plasma membrane adhesion molecules	A.1.2.	A.1.1.
embryonic skeletal system morphogenesis	A.1.2.	A.1.1.
Golgi apparatus subcompartment	A.2.1.	A.1.1.
transcription repressor complex	A.2.1.	A.1.1.
lysosomal lumen	A.2.1.	A.1.1.
inflammasome complex	A.2.1.	A.1.1.
negative chemotaxis	A.2.2.	A.1.1.
embryonic pattern specification	A.2.2.	A.1.1.
blastoderm segmentation	A.2.2.	A.1.1.
histone deacetylase activity	B.1.2.	A.1.1.
protein lysine deacetylase activity	B.1.2.	A.1.1.
regulation of nervous system process	B.2.1.	A.1.1.
midgut development	B.2.1.	A.1.1.
enteric nervous system development	B.2.1.	A.1.1.
regulation of cell maturation	B.2.1.	A.1.1.
response to pain	B.2.1.	A.1.1.
immunological synapse	B.2.2.	A.1.1.
ribonucleoprotein complex biogenesis	A.1.1.	A.1.2.
mitotic cell phase transition	A.1.1.	A.1.2.
ncRNA processing	A.1.1.	A.1.2.
ribosome biogenesis	A.1.1.	A.1.2.
rRNA metabolic process	A.1.1.	A.1.2.
small GTPase mediated signal transduction	A.2.1.	A.1.2.
positive regulation of cell projection organization	A.2.1.	A.1.2.
collagen catabolic process	A.2.1.	A.1.2.
extracellular matrix disassembly	A.2.1.	A.1.2.
telencephalon cell migration	A.2.1.	A.1.2.
connective tissue development	A.2.2.	A.1.2.
central nervous system neuron development	A.2.2.	A.1.2.
regulation of chondrocyte differentiation	A.2.2.	A.1.2.
positive regulation of chondrocyte differentiation	A.2.2.	A.1.2.
pvramidal neuron development	A.2.2.	A.1.2.
cytoplasmic ribonucleoprotein granule	B.1.1.	A.1.2.

ribonucleoprotein granule	B.1.1.	A.1.2.
regulation of canonical Wnt signaling pathway	B.1.2.	A.1.2.
canonical Wnt signaling pathway	B.1.2.	A.1.2.
regulation of Wnt signaling pathway	B.1.2.	A.1.2.
small GTPase binding	B.1.2.	A.1.2.
positive regulation of canonical Wnt signaling pathway	B.1.2.	A.1.2.
3 prime 5 prime-cyclic-AMP phosphodiesterase activity	B.2.1.	A.1.2.
3 prime,5 prime-cyclic-nucleotide_ phosphodiesterase activity	B.2.1.	A.1.2.
cyclic-nucleotide phosphodiesterase activity	B.2.1.	A.1.2.
response to temperature stimulus	B.2.2.	A.1.2.
calcium ion import across plasma membrane	B.2.2.	A.1.2.
calcium ion import into cytosol	B.2.2.	A.1.2.
vesicle-mediated transport • between endosomal compartments	B.2.2.	A.1.2.
nucleotide biosynthetic process	A.1.1.	A.2.1.
nucleoside phosphate biosynthetic process	A.1.1.	A.2.1.
purine ribonucleotide biosynthetic process	A.1.1.	A.2.1.
ribonucleotide biosynthetic process	A.1.1.	A.2.1.
ribose phosphate o biosynthetic process	A.1.1.	A.2.1.
mitochondrial tricarboxvlic acid cycle enzyme complex	A.1.2.	A.2.1.
transcription factor TFIIH core complex	A.1.2.	A.2.1.
outer kinetochore	A.1.2.	A.2.1.
transcription factor TFIIH holo complex	A.1.2.	A.2.1.
tricarboxlic acid cycle enzyme complex	A.1.2.	A.2.1.
17-beta-hydroxysteroid dehydrogenase (NAD+) activity	A.2.2.	A.2.1.
estradiol 17-beta-dehydrogenase activity	A.2.2.	A.2.1.
polyketide metabolic process	A.2.2.	A.2.1.
aminoglycoside antibiotic metabolic process	A.2.2.	A.2.1.
doxorubicin metabolic process	A.2.2.	A.2.1.
negative regulation of neuron projection development	B.1.2.	A.2.1.
negative regulation of cell projection organization	B.1.2.	A.2.1.
neuron projection extension	B.1.2.	A.2.1.
developmental cell growth	B.1.2.	A.2.1.
developmental growth o involved in morphogenesis	B.1.2.	A.2.1.
glucose transmembrane transport	B.2.1.	A.2.1.
hexose transmembrane transport	B.2.1.	A.2.1.
monosaccharide transmembrane transport	B.2.1.	A.2.1.
carbohydrate transmembrane transport	B.2.1.	A.2.1.
carbohydrate transport	B.2.1.	A.2.1.
response to tumor cell	B.2.2.	A.2.1.

organelle fission	A.1.1.	A.2.2.
nuclear division	A.1.1.	A.2.2.
regulation of cell cycle phase transition	A.1.1.	A.2.2.
mitotic nuclear division	A.1.1.	A.2.2.
postsynaptic density	B.2.1.	A.2.2.
Wnt signaling pathway	B.2.1.	A.2.2.
cell-cell signaling by wnt	B.2.1.	A.2.2.
positive regulation of autophagy	B.2.1.	A.2.2.
glial cell-derived neurotrophic factor receptor signaling pathway	B.2.1.	A.2.2.
SH3 domain binding	B.2.2.	A.2.2.
protein kinase activator activity	B.2.2.	A.2.2.
growth factor receptor binding	B.2.2.	A.2.2.
kinase activator activity	B.2.2.	A.2.2.
neuromuscular junction	B.2.2.	A.2.2.
amino acid metabolic process	A.1.1.	B.1.1.
amino acid activation	A.1.1.	B.1.1.
tRNA aminoacylation	A.1.1.	B.1.1.
intermediate filament organization	A.2.1.	B.1.1.
intermediate filament cytoskeleton organization	A.2.1.	B.1.1.
intermediate filament-based process	A.2.1.	B.1.1.
ribonucleoprotein complex assembly	A.2.1.	B.1.1.
spliceosomal snRNP assembly	A.2.1.	B.1.1.
synaptic membrane	B.1.2.	B.1.1.
neuronal cell body	B.1.2.	B.1.1.
presynaptic membrane	B.1.2.	B.1.1.
neuronal cell body membrane	B.1.2.	B.1.1.
cell body membrane	B.1.2.	B.1.1.
positive regulation of MAPK cascade	B.2.1.	B.1.1.
cellular response to tumor necrosis factor	B.2.1.	B.1.1.
response to tumor necrosis factor	B.2.1.	B.1.1.
eosinophil chemotaxis	B.2.1.	B.1.1.
eosinophil migration	B.2.1.	B.1.1.
phosphatidylethanolamine acyl-chain remodeling	B.2.2.	B.1.1.
pulmonary valve morphogenesis	B.2.2.	B.1.1.
pulmonary valve development	B.2.2.	B.1.1.
phosphatidylethanolamine metabolic process	B.2.2.	B.1.1.
aortic valve morphogenesis	B.2.2.	B.1.1.
regulation of dorsal identity	A.1.1.	B.1.2.
bone trabecula formation	A.1.1.	B.1.2.

mesonephric tubule formation	A.1.1.	B.1.2.
osteoclast proliferation	A.1.1.	B.1.2.
sequestering of extracellular ligand from receptor	A.1.1.	B.1.2.
NAD+ binding	A.1.2.	B.1.2.
G protein-coupled chemoattractant receptor activity	A.1.2.	B.1.2.
chemokine receptor activity	A.1.2.	B.1.2.
protein kinase A regulatory subunit binding	A.1.2.	B.1.2.
tRNA methyltransferase activity	A.1.2.	B.1.2.
single-stranded break repair	A.2.1.	B.1.2.
ovarian follicle development	A.2.2.	B.1.2.
female gonad development	A.2.2.	B.1.2.
protein targeting to mitochondrion	A.2.2.	B.1.2.
mitochondrial protein processing	A.2.2.	B.1.2.
signal peptide processing	A.2.2.	B.1.2.
axonogenesis	B.2.1.	B.1.2.
positive regulation of nervous system development	B.2.1.	B.1.2.
olfactory bulb interneuron differentiation	B.2.1.	B.1.2.
olfactory bulb development	B.2.1.	B.1.2.
olfactory lobe development	B.2.1.	B.1.2.
mammary gland epithelium development	B.2.2.	B.1.2.
mammary gland development	B.2.2.	B.1.2.
ERBB4 signaling pathway	B.2.2.	B.1.2.
olfactory bulb interneuron differentiation	B.2.2.	B.1.2.
complement receptor mediated signaling pathway	B.2.2.	B.1.2.
intrinsic apoptotic signaling pathway in response to oxidative stress	A.1.1.	B.2.1.
cell death in response to oxidative stress	A.1.1.	B.2.1.
G2/M transition of mitotic cell cycle	A.1.1.	B.2.1.
neural precursor cell proliferation	A.1.1.	B.2.1.
cell cycle G2/M phase transition	A.1.1.	B.2.1.
methionine biosynthetic process	A.2.1.	B.2.1.
S-adenosylmethionine metabolic process	A.2.1.	B.2.1.
methionine metabolic process	A.2.1.	B.2.1.
sulfur amino acid biosynthetic process	A.2.1.	B.2.1.
aspartate family amino acid biosynthetic process	A.2.1.	B.2.1.
gamma-catenin binding	A.2.2.	B.2.1.
GTPase activating protein binding	A.2.2.	B.2.1.
ankyrin binding	A.2.2.	B.2.1.
beta-catenin binding	A.2.2.	B.2.1.
protein tyrosine kinase binding	A.2.2.	B.2.1.

dendritic cell antigen, processing and presentation	B.1.1.	B.2.1.
MHC class II protein complex assembly	B.1.1.	B.2.1.
peptide antigen assembly with MHC class II protein complex	B.1.1.	B.2.1.
MHC protein complex assembly	B.1.1.	B.2.1.
peptide antigen assembly. with MHC protein complex	B.1.1.	B.2.1.
cell-cell adhesion via plasma-membrane adhesion molecules	B.2.2.	B.2.1.
homophilic cell adhesion via plasma membrane adhesion molecules	B.2.2.	B.2.1.
cell junction assembly	B.2.2.	B.2.1.
synapse assembly	B.2.2.	B.2.1.
calcium-dependent cell- cell adhesion via plasma membrane cell adhesion molecules	B.2.2.	B.2.1.
folic acid-containing compound metabolic process	B.1.2.	B.2.2.
pteridine-containing compound metabolic process	B.1.2.	B.2.2.
cellular modified amino acid metabolic process	B.1.2.	B.2.2.
mitochondrial translation	B.1.2.	B.2.2.
mitochondrial gene expression	B.1.2.	B.2.2.
carbohydrate:monoatomic cation symporter activity	B.2.1.	B.2.2.
glucose transmembrane transporter activity	B.2.1.	B.2.2.
hexose transmembrane transporter activity	B.2.1.	B.2.2.
monosaccharide transmembrane transporter activity	B.2.1.	B.2.2.
sugar transmembrane transporter activity	B.2.1.	B.2.2.



Supplementary Figure 2.11. Functional analysis of quiescence associated loops. A) GO analysis of genes downregulated with quiescence loops. B) GO analysis of genes upregulated with quiescence loops.

Chapter 3: Writing an Alphabet to Read the Senescent Genome

In Chapter 3 Audrey Dalgarno was responsible for all preliminary CSA analysis (GM12878, WTC11, LF1 proliferating and senescent). Audrey Dalgarno was also responsible for the selection of modeling method and TAD caller, as well as the generation of the full genome Chromosome Shape Alphabet from the IGM structures.

3.1: Abstract

Despite the enormous structural variability exhibited in 3D chromosomal conformations at a global scale, there is a significant commonality of structures visible at smaller, local levels. We hypothesize that chromosomal conformations are representable as concatenations of a handful of prototypical shapelets, termed Chromosome Shape Letters. This is akin to expressing complicated sentences in a language using only a small set of letters. Our goal is to organize the vast variability of 3D chromosomal conformation by constructing a set of predominant shape letters, termed a shape alphabet, using statistical shape analysis of curvelets taken from training conformations. Furthermore, we aim to use this analysis to determine the functional effects of these prototypical shape letters, leading to an efficient understanding of the biological role of genomic shape, and how it is altered with different cell lines and conditions, such as senescence. This chapter utilizes conformations generated from Integrative Genome Modeling (IGM) to develop a shape alphabet as follows: it first segments 3D conformations into curvelets according to their Topologically Associated Domains. It then clusters these segments, estimates mean shapes, and refines and reorders these shapes into a Chromosome Shape Alphabet. The chapter demonstrates effectiveness of this construction by successfully representing independent test conformations taken from IGM and other methods such as SIMBA3D, both symbolically and structurally, using the constructed alphabet. We also show preliminary analysis indicating shape letters may have biological roles, and change with conditions such as cellular senescence.

3.2: Introduction

Genomic architecture is one of the most rapidly evolving modern biology fields, largely due to recent sequencing (Lieberman-Aiden et al., 2009) and microscopy (Beliveau et al., 2017; Mateo, Sinnott-Armstrong, & Boettiger, 2021) innovations. From these technologies we have begun to see how the genome is carefully organized in 3D to regulate functions such as gene expression. However informative these assays are, it can be difficult to link structure to function, due to the potentially vast scale of these datasets, and the necessity to parse through cell-to-cell structural variability. Furthermore, the structure-function links that we know of, such as A and B compartments, are derived from Hi-C data (Lieberman-Aiden et al., 2009), rather than genomic structure itself. Even in the context of imaging, to connect structure and function imaging coordinates are usually transformed back into contact matrix-like structures for analysis (Mateo et al., 2021). We lack a set of tools to parse through genome structure and shape itself to identify critical features and connect them to biological function.

To fill this void, we worked with collaborators in the Srivastava Lab at Florida State University to create the Chromosome Shape Alphabet (CSA) method, which identifies reoccurring shapes in the 3D genome (C. Soto, Bryner, Neretti, & Srivastava, 2021). We can think of these reoccurring shapes as being analogous to functional secondary structures in proteins such as alpha helices and beta-sheets. The CSA method involves several steps (Figure 3.1). First, an ensemble of chromosomes modeled to resemble natural cell-to-cell heterogeneity are segmented according to a segmentation of interest, such as TAD segmentation. Next, the distance between each of these TAD curves is computed using a specialized shape metric invariant to factors such as rotation and scaling. The resulting matrix is then clustered, and a median shape is then computed for each cluster. This is repeated with chromosomes and conditions of interest and ordered by complexity

to create a CSA composed of Chromosome Shape Letters (CSL). These CSLs can be used to describe local variability across a chromosome, connect structure to biological features such as transcriptional activity and epigenomic marks.

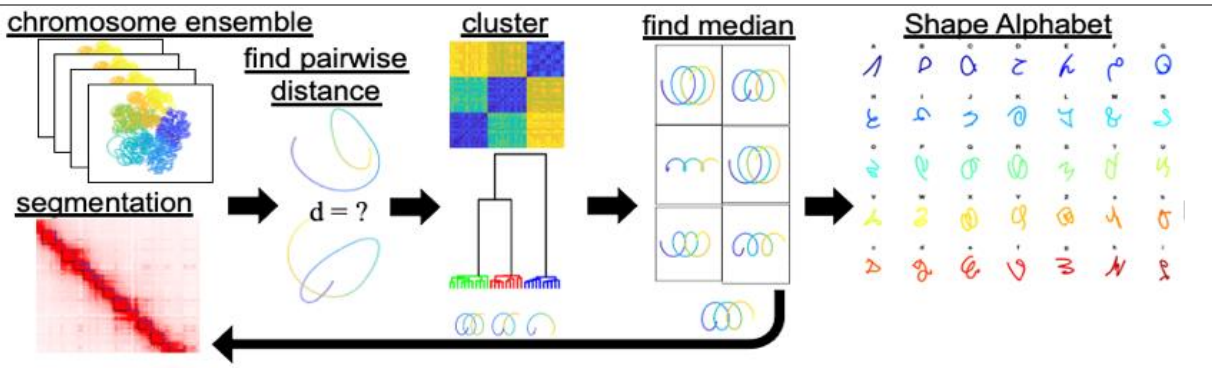


Figure 3.1. Overview of Chromosome Shape Alphabet (CSA) method. The chromosome ensemble is segmented into curves (by TADs, for example). A pairwise distance matrix is prepared by determining the distance (d) between each curve using a specialized distance metric. This matrix is then clustered, a median shape is determined, and this process is repeated to incorporate curves from multiple chromosomes, conditions, etc. The results of this process is a CSA of reoccurring shapes found across the genome.

The original publication of the CSA method was impactful, but contained no explicit connection to biology. My interest in this project was to elucidate the functional role of these reoccurring shapes in the genome, especially within the context of aging. For example, are certain CSL found more often in the A/B compartments? Do they undergo compartment switching between conditions? Are some CSL more 'active' as shown through associations with transcription and epigenetic marks? How does the distribution of CSL change with senescence? Additionally, through starting this type of analysis we discovered it was necessary to improve several aspects of this method, due to both computational and practical limitations. Preliminary biological CSL associations as well as improvements to the CSA methods are outlined in this thesis chapter.

3.3: Methods

3.3.1: Elastic Shape Analysis

In the CSA method we compare and cluster chromosomal segments or curvelets in terms of their shapes. The *shape* of a curve is an intrinsic attribute that is invariant to transformations. such as rotation, reflection, scale, location, and parameterization. To analyze shapes of these curves, we utilize *elastic shape analysis* (Srivastava & Klassen, 2018). This approach is summarized next.

Let $\mathcal{F}$ be the set of all absolutely continuous functions from [0,1] to $\mathbb{R}^3$. Each element $f \in \mathcal{F}$ is a parameterized space curve. This framework utilizes the notion of a *square-root velocity function* (SRVF) to compute shape differences and shape summaries. For a curve $f \in \mathcal{F}$ its SRVF is defined by a map $q : [0,1] \rightarrow \mathbb{R}^3$ according to the formula $q(t) = \dot{f}(t)/\sqrt{|\dot{f}(t)|}$. Let Γ be the set of all re-parameterization functions and $\mathbb{O}$ be the set of all 3D rotations. For any $\gamma \in \Gamma$ and $O \in \mathbb{O}$, the curve $O(f \circ \gamma)$ represent a rotated and re-parameterized version of f while retaining the same shape as f. The SRVF of this new curve is given by $O(q \circ \gamma)\sqrt{\dot{\gamma}}$. This sets up the definition of the elastic shape metric.

Given any two curves $f_1, f_2 \in \mathcal{F}$, and their SRVFs q_1 and q_2, the elastic shape distance between them is:

$$d_s(f_1, f_2) = \inf_{O \in \mathbb{O}, \gamma \in \Gamma} \cos^{-1}\left[\left\langle \frac{q_1}{\|q_1\|}, O \cdot \left(\frac{q_2}{\|q_2\|} \circ \gamma\right)\sqrt{\dot{\gamma}}\right\rangle\right]$$

where $\|\cdot\|$ represents the $\mathbb{L}^2$ norm of a curve. For more details on this construction, we refer the reader to (Srivastava & Klassen, 2018).

3.3.2: Clustering

We will use the pairwise shape metric d_s to cluster chromosome segments generated earlier. In any metric space, there are several algorithms for clustering points in that space. We have explored several approaches, including Bayesian clustering outlined in (Z. Zhang, Pati, & Srivastava, 2015). However, it requires several tuning parameters that significantly influence the final results. Another approach is hierarchical clustering in MATLAB. We found both of these methods to produce similar results, with hierarchical clustering being 200 times faster, so we use it from now onwards. We perform the clustering using built-in MATLAB functions 'linkage' and 'dendrogram'. For the IGM CSA we set the number of clusters as approximately 70% of the number of TADs per chromosome as determined empirically in (C. Soto, Bryner, et al., 2021).

3.3.3: Karcher Median

To determine a median curve, we find the curve f , which minimizes the sum of shape distances between curves in a cluster of *n* parameterized curves $\{f_i\}$ (C. Soto, Bryner, et al., 2021; Srivastava & Klassen, 2018)

$$\widehat{m} = \operatorname{argmin}_{f \in \mathcal{F}} \left(\sum_{i=1}^{n} d_s \left(f, f_i \right) \right).$$

3.3.4: Alphabet Refinement

Using hierarchical clustering and computation of cluster means, we can generate thousands of candidates for shape letters. This is too large to form an alphabet, so we need further refinement into a smaller set of distinct shapes. We accomplish this by performing one more round of clustering and averaging. We denote the elements of this refined set as $\Delta = \{\delta i\}$; this set is called the CSA, and its elements are called CSLs.

It is convenient to order the selected shape letters according to their shape complexity, starting from the simplest shapes. Similar to (C. Soto, Bryner, et al., 2021) we use the *total absolute*

curvature (TAC) (Willmore, 1971), defined as $\int_0^1 \frac{\|f'(t)\times f''(t)\|}{\|f'(t)\|^3}dt$, to order these letters. To develop a simple mneumonic system, we assign a letter symbol to each of these 52 shapes, with some examples listed in Figure 3.2.

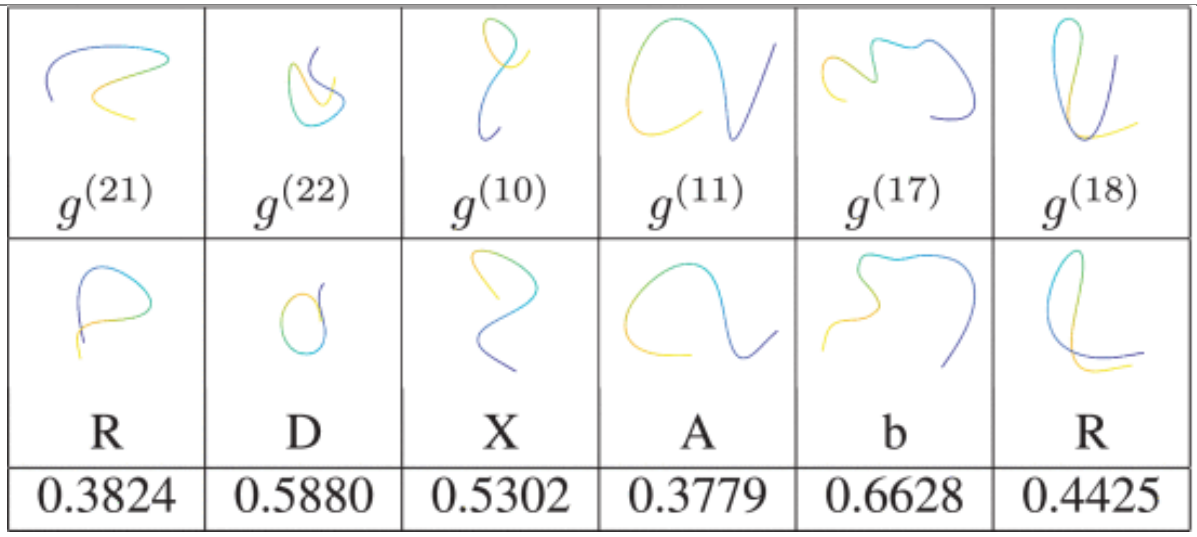


Figure 3.2. A TAD segment from a chromosome conformation denoted $g^{(I)}$ (top). The most similar CSL to the above segment measured with elastic shape distance (middle). The elastic shape distance between the chromosome segment and the CSL (bottom).

3.3.5: Representation by Letter Sequences

Suppose we have a test 3D conformation $g \in \mathcal{F}$ with TAD based segments $g^{(1)}, g^{(2)}, \ldots g^{(K)}$. We first construct a sequence of CSLs which most closely resemble the segments in terms of shape distance by $\hat{\delta}_k = \arg\min_{\delta\in\Delta} d_S(g^{(k)}, \delta)$. Figure 3.2 displays some segments of a curve with their corresponding most similar CSLs. We can use this sequence of representative shapes to generate a string of letters that represent the original conformation g symbolically.

In the case of TAD segmentations which do not cover the full chromosome, that is, between $g^{(i)}$ and $g^{(i+1)}$, before $g^{(1)}$, and after $g^{(K)}$ there may be additional data points. For reconstruction analysis we require a segmentation corresponding to each coordinate of the contact matrix, so we fill the gaps by using the ground truth. Further, since our CSA has 52 elements we can symbolically

represent each element using the upper and lower cases of the English alphabet. We set $\delta_1 - \delta_{26} \rightarrow A - Z$ and $\delta_{27} - \delta_{52} \rightarrow a - z$ as shown in Figure 3.2.

3.3.6: CSA Senescence Analysis

The software SIMBA3D (Rosenthal et al., 2019) was used with different random initializations to create an ensemble of chromosome structures with heterogeneity. We created 10 initializations of chromosomes 1, 3, and 5 using SIMBA3D from Hi-C data at 50 kb resolution extracted using Juicer's 'dump' tool (Durand et al., 2016) on data from the cells lines GM12878 (Rao et al., 2014), WTC11 (Dekker Lab, 4DNucleome Portal)., and LF1 proliferating and senescent cells (see Methods sections 2.3.1-2). Hi-C maps were segmented into TADs using 3DNetMod (Norton et al., 2018). Clustering, median determination, and CSA creation were completed as in Methods sections 3.3.1-5. The curves in the shape alphabet are ordered by radius of curvature as in (Mjaavatten, 2020). Chromosome motifs were created using ggseqlogo (Wagih, 2017). A and B compartments were determined as in (Miura, Poonperm, Takahashi, & Hiratani, 2018). In brief, Juicer's eigenvector (Durand et al., 2016) was used to determine the first principal component of the Pearson's correlation matrix, then assigned to the A/B compartment based on gene density.

3.3.7: IGM Analysis

The software Integrative Genome Modeling (IGM) (Boninsegna et al., 2022) was used by Lorenzo Boninsegna (Alber Lab) to create an ensemble of structures of the cell line GM12878 at 200 kb (Rao et al., 2014). These structures were graciously gifted to us. We used 100 initializations of the 'a' homolog for chromosomes 1 – 22 and X to create a CSA. The TAD caller TopDom was used on the 200 kb matrices. TADs between 800 kb and 3 Mb were retained, to avoid size bias. As there were many gaps, TADs were used from calling with TopDom at 100 kb. For comparison

we use H9 (human stem cell) data from the original CSA publication (C. Soto, Bryner, et al., 2021). For this data SIMBA3D (Rosenthal et al., 2019) was used to create 60 chromosome structures (by changing the initialization) from portions of chromosome 1 and 2 at 50 kb. 10-20 of these 1000 x 1000 matrices were used for training and the rest for testing. For the H9 data an insulation score method was used to call TADs (Schwarzer et al., 2017).

3.3.8: Recursive Fréchet Mean (RecFME)

For these clustered chromosome segments, the next step is to define a representative shape for each cluster. For this, we utilize the notion of a Fréchet mean of shapes. The traditional approach for estimating these means defines an objective function involving given shapes and uses a gradient- based approach to minimize that function. This approach is computationally expensive, as each iterative update requires computing geodesic or optimal deformations between each given shape and the current mean estimate. In the following, we describe a novel, more efficient algorithm for estimating a mean shape.

The main idea is to start with an initial guess and update the estimate using one given curve at a time. The update uses the notion of a *geodesic*, or the optimal deformation, between any two shapes (Figure 3.3). A geodesic is the manifold equivalent of a straight line. For any two curves f_1 and f_2 with corresponding SRVFs q_1 and q_2, the geodesic path between them is defined to be $\alpha_\tau(q_1, q_2) = \frac{1}{\sin(\theta)}(\sin((1-\tau)\theta)q_1 + \sin(\tau\theta)\tilde{q}_2)$ where $\theta = \cos^{-1}(\langle q_1, \tilde{q}_2 \rangle)$, $\tau \in [0,1]$, and $\tilde{q}_2$ is the SRVF of the second curve after optimal rotation and re-parameterization. Here $\tau \in [0, 1]$ is a scalar parameter indexing the geodesic path. In this setting, one can easily obtain a weighted mean of the shapes f_1 and f_2, with the weights $1-\tau$ and τ respectively, as simply α_τ.

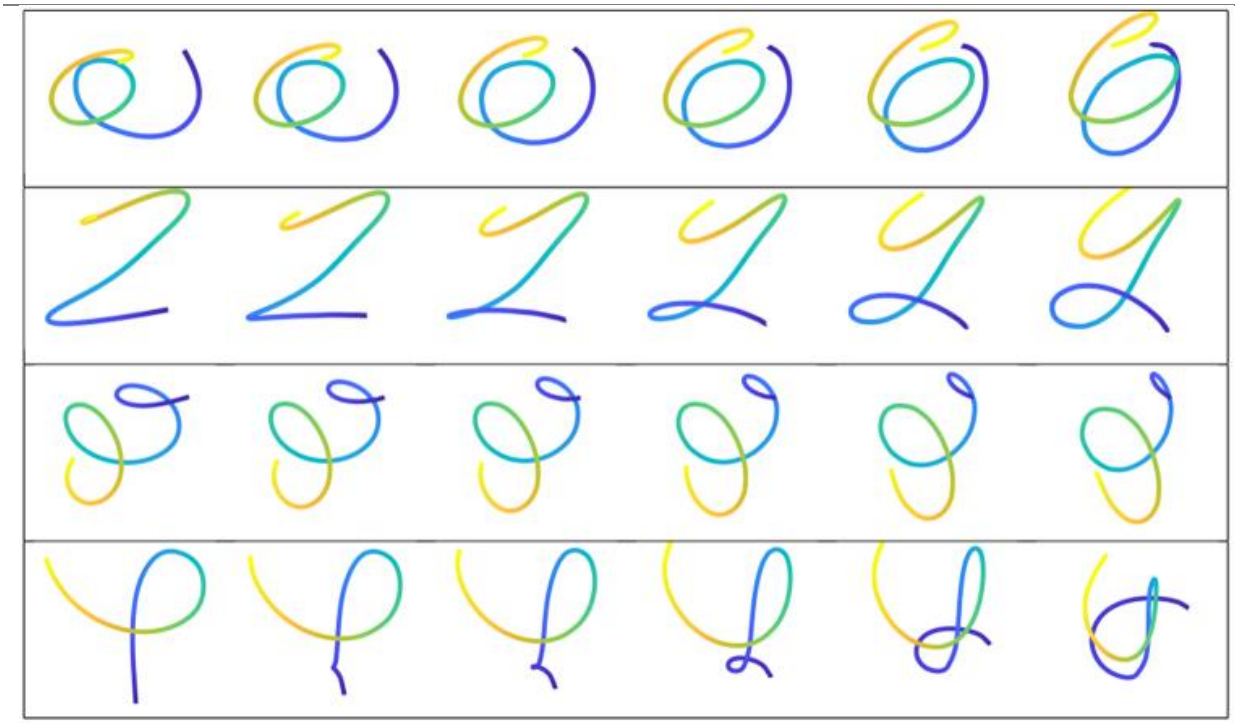


Figure 3.3. Examples of geodesic deformations between the left most and rightmost curves in each row. From Soto et al. (C. Soto, Bryner, et al., 2021).

Returning to the problem of finding a mean of n curves, we start with an initial estimate and update it using one given curve at a time. We compute a weighted mean of the current mean estimate and the next given curve in each step. Specifically, we start with the first two curves f_1, f_2, and compute their mean shape with equal weights (1/2) and (1/2). We call the result $\mu^{(2)}$. Next, we compute the weighted mean of $\mu^{(2)}$ with the next curve f_3, with the weights (2/3) and (1/3) respectively. Call the result $\mu^{(3)}$. Next, we compute the weighted mean of $\mu^{(3)}$ with the curve f_4, with the weights (3/4) and (1/4) respectively. And so on. This approach performs a single pass on the data and computes $n - 1$ geodesics in the process. One can modify this algorithm to improve the efficiency further. As illustrated in Fig. 3.4, one can use a hierarchical approach – group the curves in some fashion, compute means of within the groups at one level, and then further compute

the mean of those at the next level. We will call this approach the Recursive Fréchet Mean Estimator or RecFME.

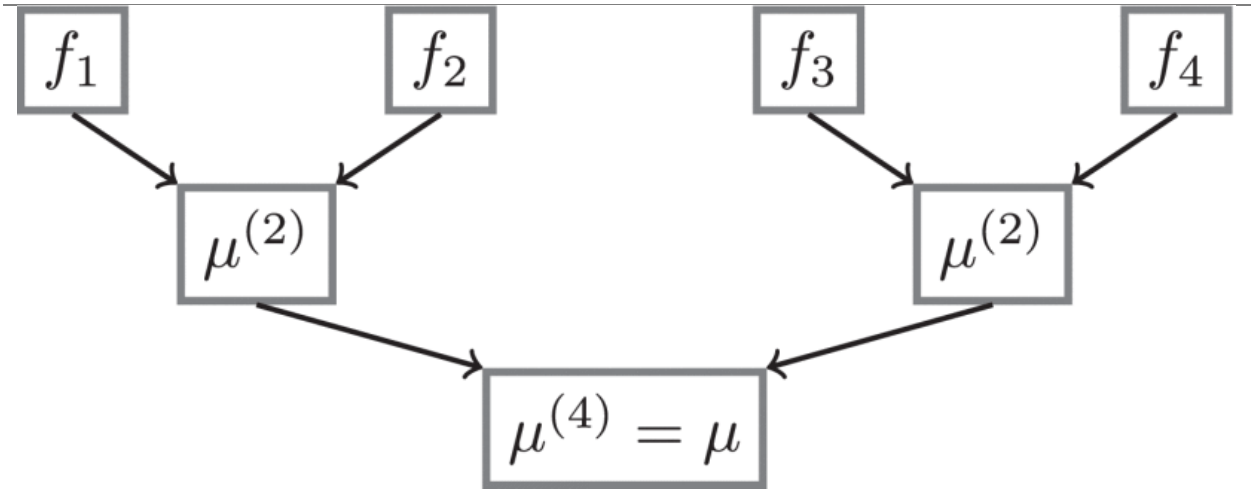


Figure 3.4. An example of the recursive Fréchet mean algorithm for n = 5. The data itself are the leaves at the top. The superscript is for keeping track of the weight for each new element. The algorithm recursively iterates top-down, with the final estimate being the bottom-most element.

3.2.9: Full 3D Reconstruction

Given a sequence of segments and the corresponding CSLs, as described in the previous section, we want to use these CSLs to reconstruct a full 3D conformation that approximates the original structure. This requires solving for optimal translation, orientation (rotation and reflection), and scale for each shape letter to best match the corresponding segments. We formulate an objective function that is based on a penalized negative log-likelihood as in (Rosenthal et al., 2019) and devise an optimization scheme to solve for the free parameters as outlined in (C. Soto, Dalgarno, et al., 2021).

In the proof-of-concept test we generate one solution from SIMBA3D (which we then divide into 12 equally sized substructures) for each chromosome 1 – 19 in the mouse embryonic stem cell (mESC) dataset given in the paper. We randomize each substructure's relative scales,

orientations, and positions and run the structure alignment algorithm to recover the original SIMBA3D structure approximately. We generate ten such solutions from an initial randomized configuration for each chromosome and record the objective function's computation time and final value (final energy).

3.4: Results

3.4.1: Examining Senescence-Associated Architectural Changes with the CSA Method

While we have the tools to comprehensively assess the changes in Hi-C maps across different conditions, such as proliferating and senescent cells (Chapter 2), we lack concrete tools to compare how the *shape* of the genome itself is changing. For this reason, we were interested in using the CSA method to examine how the shape of the genome changes with senescence in a methodical way. This CSA method is appropriate to use in this situation, as it allows us to more easily pick out trends in how genome shape is altered, and furthermore facilitates the connection of these alterations to biological features, such as transcriptional activity.

We ran the CSA method on a combination of data from GM12878, WTC11, and LF1 proliferating and senescent cells using the original published CSA method (C. Soto, Bryner, et al., 2021). This resulted in a CSA of 35 letters, ranging from A – I (Figure 3.5). With this CSA we were able to perform basic analysis. For example, do certain CSL occur more frequently? The most frequently occurring CSL across this data was 'g', followed by CSL 'J', 'Z', and 'D'. There was not a clear correlation between CSL complexity and frequency, however many of the more commonly occurring CSL represented loop multiples, such as D, J, and g (single loop, double loop, triple loop, etc.).

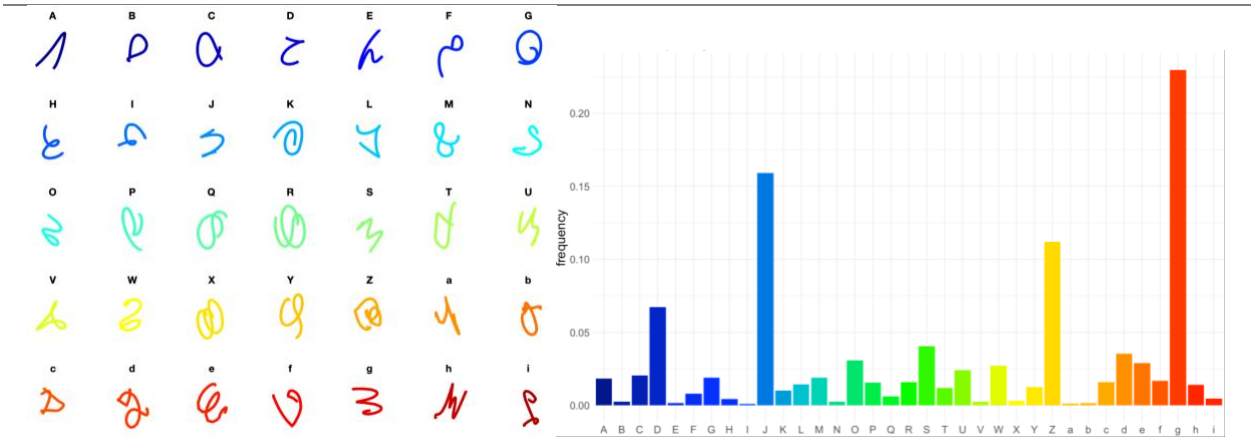


Figure 3.5. Shape alphabet ranked by complexity created from GM12878, WTC11, LF1 proliferating, and LF1 senescent data (right). Corresponding frequency of each CSL (left)

We can also use these CSL to represent the variability seen across a population of cells, as single cell studies have indicated that there is significant structural heterogeneity (Bintu et al., 2018; Nagano et al., 2013, 2017; Nir et al., 2018). One effective way to do this is a sequence logo, which is generally used to represent DNA motifs, such as those targeted by transcription factors. In Figure 3.6 we display a sequence logo motif for chromosome 1 of LF1 proliferating and senescent cells. From these motifs we can see that CSL are affected by cell line/condition. For example, LF1 proliferating cells have more shapes of intermediate complexity, whereas LF1 senescent cell CSL are more polarized, as we see more simple shapes and complex shapes.

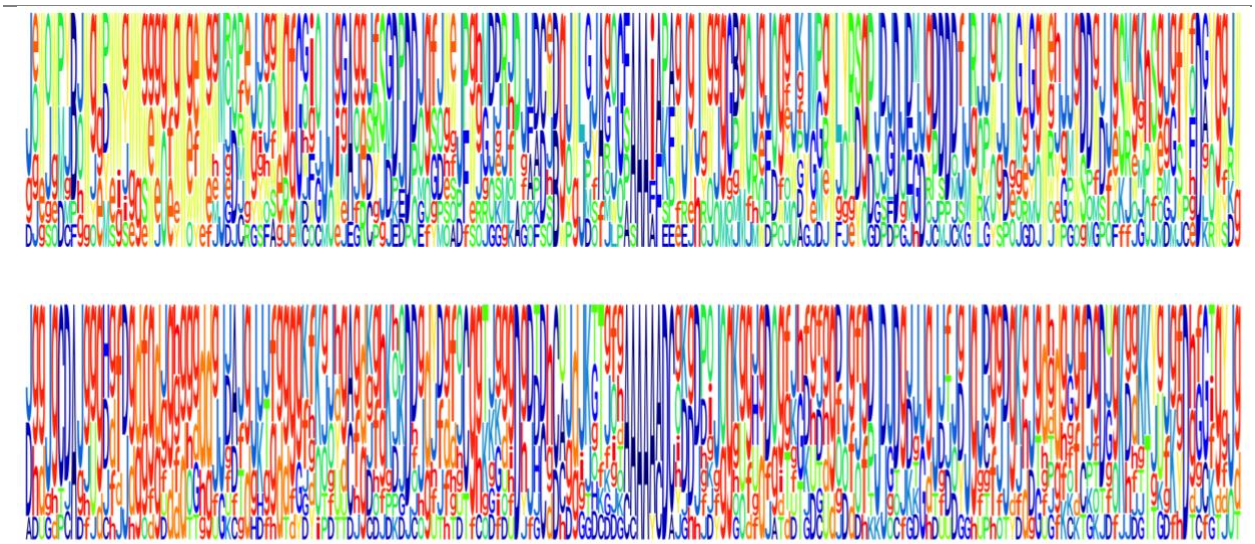


Figure 3.6. Chromosome motifs for LF1 chr 1 of proliferating cells (top) and senescent cells (bottom. Each column represents a TAD and the height of each CSL represents the probability of it occurring at a given position.

CSL can also be intersected with A/B compartments to give us more biological information about these stereotypical shapes. For example, are certain CSL more commonly found in the A v. the B compartment? Does this change across conditions or cell lines? From Figure 3.7 we can see that cell lines and conditions can affect the frequency of Shape Letters occurring in the A or B compartment. For example, CSL 'A' is mostly in the B compartment with GM12878, mostly in the A compartment with WTC11, and partially in the A and partially in the B compartment in LF1. Furthermore, specific CSL change their distribution significantly between proliferating cells. For example, CSL 'C' is found more frequently in the A compartment with senescence.

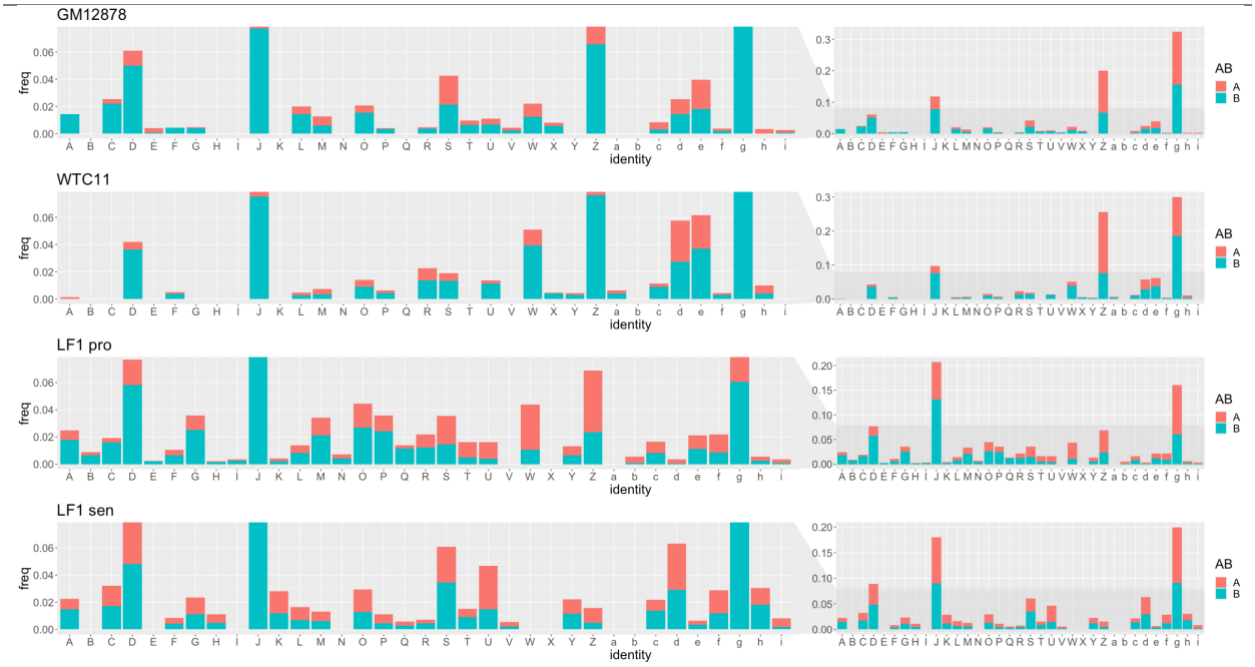


Figure 3.7. Frequency of Shape Letters occurring in A/B compartments for GM12878, WTC11, LF1 proliferating and LF1 senescent cells. Left panels are a closer view of the right panels.

3.4.2: Extending and Improving the CSA Method

In a previous study, and in the previous sections, we introduced the concept of a Chromosome Shape Alphabet (CSA). This concept involves constructing a set of recurrent local structures, called Chromosome Shape Letters (CSL), that can potentially be used to represent and reconstruct complete 3D conformations. This representation is akin to expressing complex words, sentences, and texts in a language using only a handful of letters that form an alphabet. Throughout our preliminary analysis we discovered shortcomings of this method, such as computational and practical limitations and considerations. We explore these in this section.

For example, in our initial development (C. Soto, Bryner, et al., 2021), we used SIMBA3D (Structural Inference via Multiscale Bayesian Approach), a method originally designed for generating chromosomes structures from innately sparse single-cell Hi-C data by using bulk Hi-C data (Rosenthal et al., 2019). SIMBA3D can also generate consensus chromosome structures to

describe bulk Hi-C matrices, which we used for alphabet previously. To capture structural heterogeneity, i.e., to reach variable genomic shape from cell to cell (Rowley & Corces, 2018), we used random initializations for estimation in SIMBA3D.

However, there are better, more biologically plausible ways to model structural heterogeneity. For example, a recent method called Integrative Genome Modeling (IGM) (Boninsegna et al., 2022) uses structure-based deconvolution to optimize a population of distinct full-genome chromosome structures from bulk Hi- C data, as in (Hua et al., 2018). IGM is thus very appropriate for use in structural analysis. Furthermore, IGM increases model accuracy by including multimodal data integration from sources such as DamID and SPRITE. Due to the paucity of ground truth full-genome structures, these methods are regarded as some of the best models for analyzing genomic shape. Using conformations resulting from IGM is critical in establishing the universality of CSA as a concept, and we pursue that goal in this analysis. We use 200 kb IGM models of the cell line GM12878 (B-Lymphocytes). Fig. 3.8 displays some sample conformations from IGM with the GM12878 cell line in the top row and from SIMBA3D with the H9 cell line in the bottom row. Even in this small sample, we see larger structural variability amongst the IGM conformations when compared to the SIMBA3D samples.

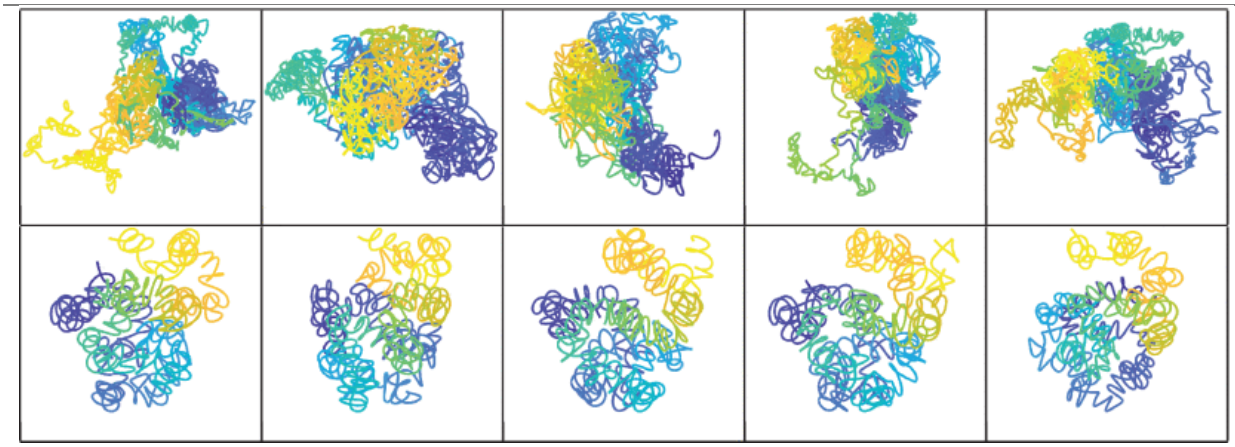


Figure 3.8. Five conformations each from the first chromosome from cell line GM12878 generated with IGM (top). Five conformations from a portion of the first chromosome from cell line H9 generated with SIMBA3D (bottom).

As mentioned earlier, the past work on CSA focused entirely on one method, namely SIMBA3D, and one small dataset due to computational bottlenecks. Here we demonstrate and validate this concept using multiple methods. We learn a shape alphabet using one method – IGM – and use it to represent conformations from another method – SIMBA3D, thus highlighting the universality of these concepts and constructions. Furthermore, we utilize advanced numerical techniques from shape clustering and shape analysis to facilitate the processing of many conformations. The main contributions of this work are:

1) Develop a chromosome shape alphabet for 3D conformations obtained from IGM.
2) Utilize advanced techniques from shape clustering and shape averaging to handle extensive training data. Specifically, we (1) use a novel iterative approach to estimate shape means; and (2) perform a two-step hierarchical clustering method to effectively cluster and represent large sets of shapes with a post-refinement step.
3) Demonstrate representations of test conformations using generated shape letters, when the test and training conformations may come from different methods altogether.

The hierarchical nature of the 3D genome gives us many options for segmenting a genome into its basic structural units. For our analysis, we chose to segment chromosome structures using Topologically Associating Domains (TADs). TADs are approximately 1 Mb self-interacting regions that are believed to influence the regulatory landscape since promoters and enhancers function within them (Rowley & Corces, 2018).

To select from the many available tools to define TADs, we consider a recent study that compares different TAD callers according to consistency across resolutions and other criteria, such

as alignment with biological features enriched at TAD boundaries (e.g., CTCF binding) (Zufferey et al., 2018). A high-performing TAD caller is TopDom that finds TAD boundaries by identifying local minima in contact frequencies in a window of a specified size around each bin across the genome (Shin et al., 2016). For the IGM conformations, we use TopDom for segmentation, while for the H9 dataset, we use an insulation score (IS) method (Schwarzer et al., 2017). Fig. 3.9 displays the TopDom TAD segmentation for a section of GM12878 chromosome 2. Note that TopDom and the TAD filtering can produce TADs with gaps along a chromosome and that requires careful processing.

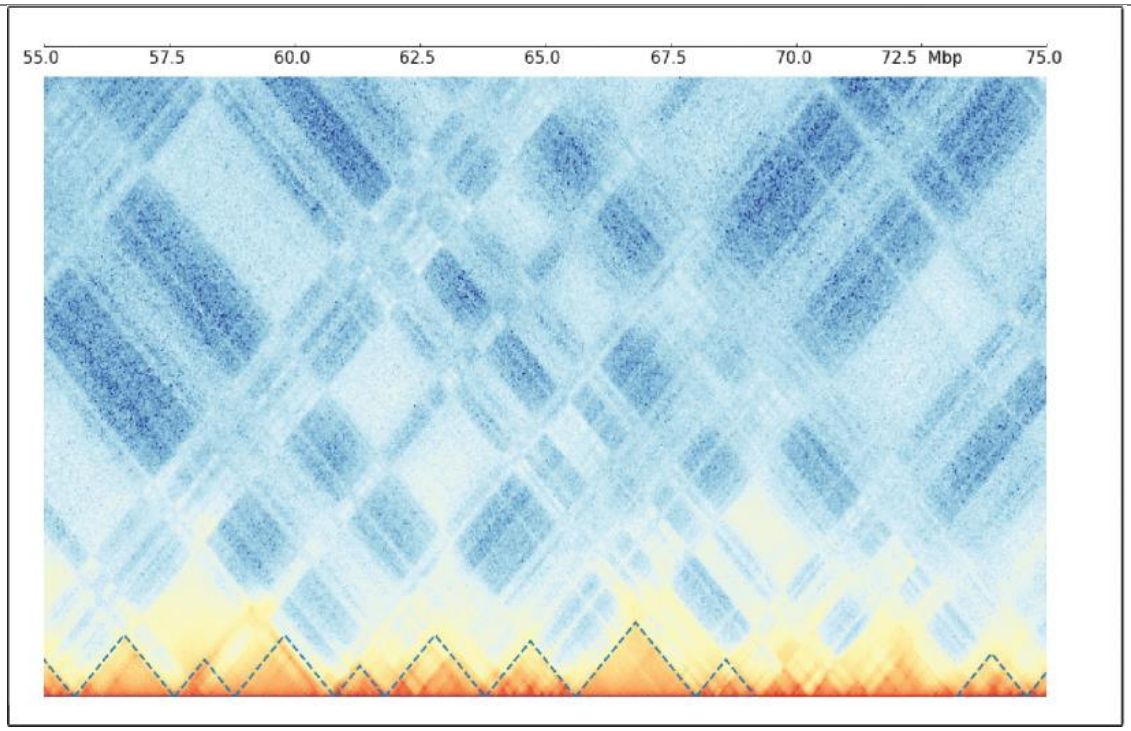


Figure 3.9. TAD segmentation created using TopDom (GM12878 chromosome 2, 55 – 75 Mb).

We segment each chromosome according to the TAD segmentation. The next step is to compare and cluster these chromosomal segments or curvelets in terms of their shapes using the CSA method, as in sections 3.2.1-5 with the following updates.

To more efficiently be able to generate cluster means we introduce the concept of a Recursive Fréchet Mean (RecFME) (Methods section 3.3.9). In short, because we can hierarchically and strategically group and compute curve means with this method it costs an order of magnitude less. Because the RecFME algorithm produces estimates very similar to the gradient descent approach we thus choose to use the RecFME for computing means of the clusters and treat them as candidates for CSLs (Figure 3.10).

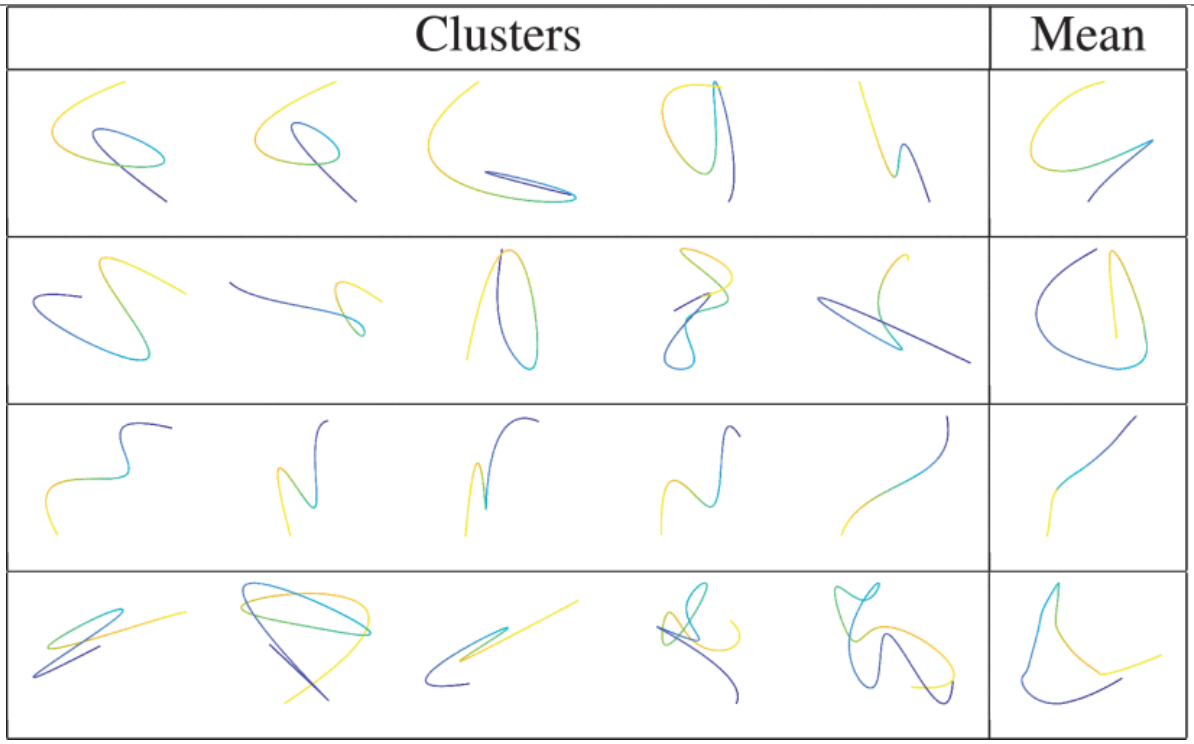


Figure 3.10. Each row displays members of clusters from hierarchical clustering with their respective mean rightmost.

Next, because we aimed to create a CSA from many (100) initializations of the full genome (23 chromosomes) we needed to implement new methods for combining shape alphabets. For example, we generate approximately 1200 shape letters from the full genome. Instead of 'pruning' alphabets as in (C. Soto, Bryner, et al., 2021) we instead refine the alphabet into a smaller set of distinct shapes by performing one more round of clustering and averaging from the chromosome

level CSA. Fig. 3.11 displays an example of this refinement showing 11 members a cluster and their weighted mean in the bottom right. We denote the elements of this refined set as the CSA, and its elements are called CSLs.

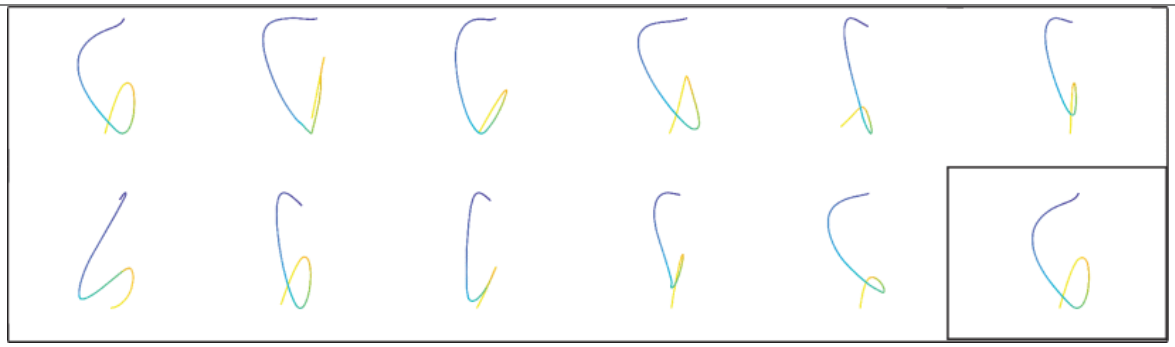


Figure 3.11. A set of candidate shapes and the refined CSL boxed in the bottom right.

After ordering these letters by total absolute curvature (TAC) we achieve an alphabet of 52 letters (Figure 3.12). This is the first demonstration of the CSA at large scale, as we have created an alphabet to represent many (100) initializations of the full genome. The computation of this CSA was facilitated by our critical computational advancements, such as the recFME.

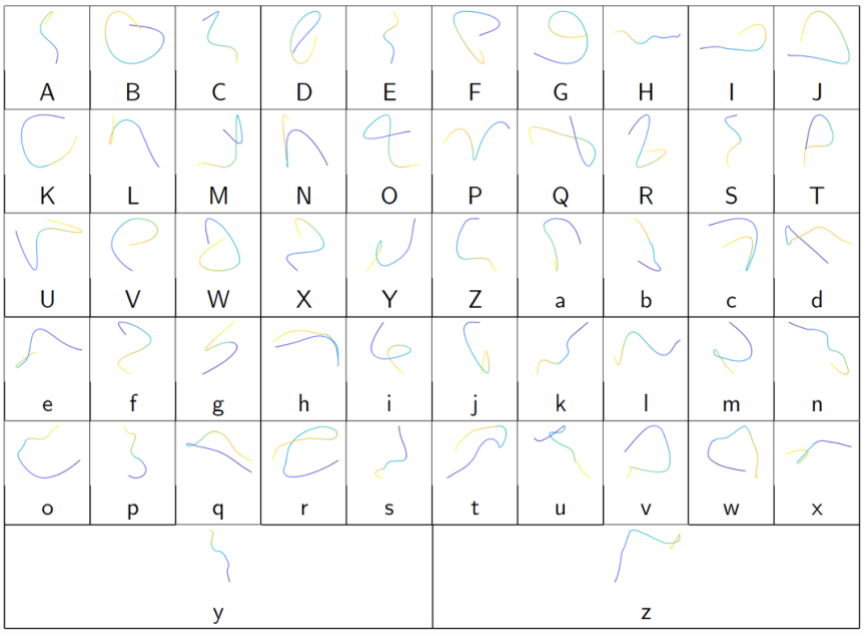


Figure 3.12. Full, sorted CSA created from full genome of IGM GM12878 data.

As in (C. Soto, Bryner, et al., 2021) we test this 52 letter CSA by reconstructing chromosomes from the CSL and determining the energies from the reconstructions. In Soto et al. we presented 3 different reconstruction algorithms to evaluate the CSA. Here we present a novel, improved algorithm. Fig. 3.13 shows the results of a proof-of-concept test of the structure alignment algorithm demonstrating the methods efficacy. The left panel shows the computation times of each solution on a log-log scale – SIMBA3D in red and structure alignment in blue – along with the fitted linear regression line for both methods. The structural-alignment algorithm is about an order of magnitude more efficient than SIMBA3D. Finally, as shown in the right panel, the final energies of the structure alignment algorithm are close to SIMBA3D, proving that the algorithm is not only efficient but well recovers the underlying structure.

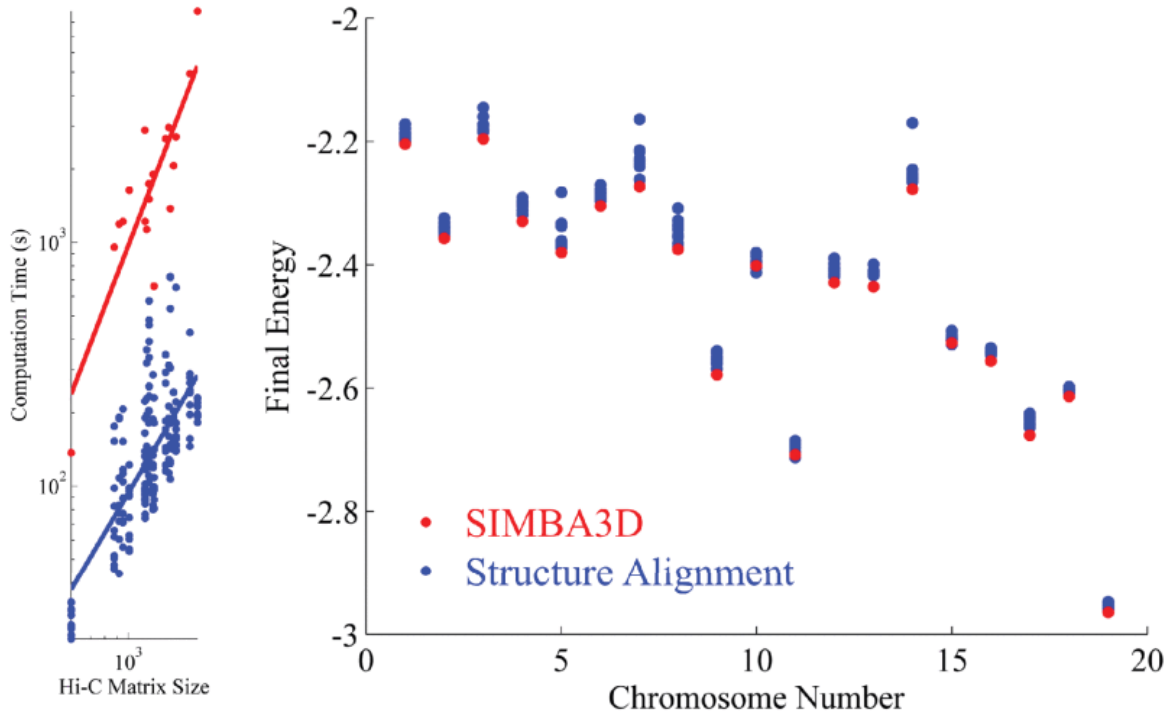


Figure 3.13. Comparison of reconstruction algorithm performance to that of SIMBA3D. The left panel shows a plot of computation time versus Hi-C matrix size on a log-log scale. Red data points represent SIMBA3D runs, and blue data points represent instances of the structure

alignment algorithm. The right panel shows the final energy of the SIMBA3D solutions (red) and the structure alignment solutions (blue) for each chromosome 1-19 of the mESC dataset. The data points on both the left and right plots result from the same solution set.

To evaluate these new methodological advancements, we first consider a dataset of the H9 human stem cells with conformations generated using SIMBA3D. Using the Insulation Score TADs described earlier, we segment these conformations and represent these segments using the CSA constructed from the IGM data. That is, this dataset did not influence the construction of the CSA. We first associate the closest shape letter to each of the segments, and then using these letter sequences, we then reconstruct the 3D structure of these test conformations.

In the top row of Fig 3.14, we display five conformations from a chromosome region of this dataset; in the second row, we display the reconstructions of each of these conformations using the CSA. In addition to the conformations and reconstructions, we display their reconstruction energies also. Remarkably, even though this alphabet was constructed from IGM data, these shape letters can reconstruct the SIMBDA3D conformations reasonably well. One can see that the energies of the reconstructed structures are similar to the original curves, despite being constrained to the shape letters only.

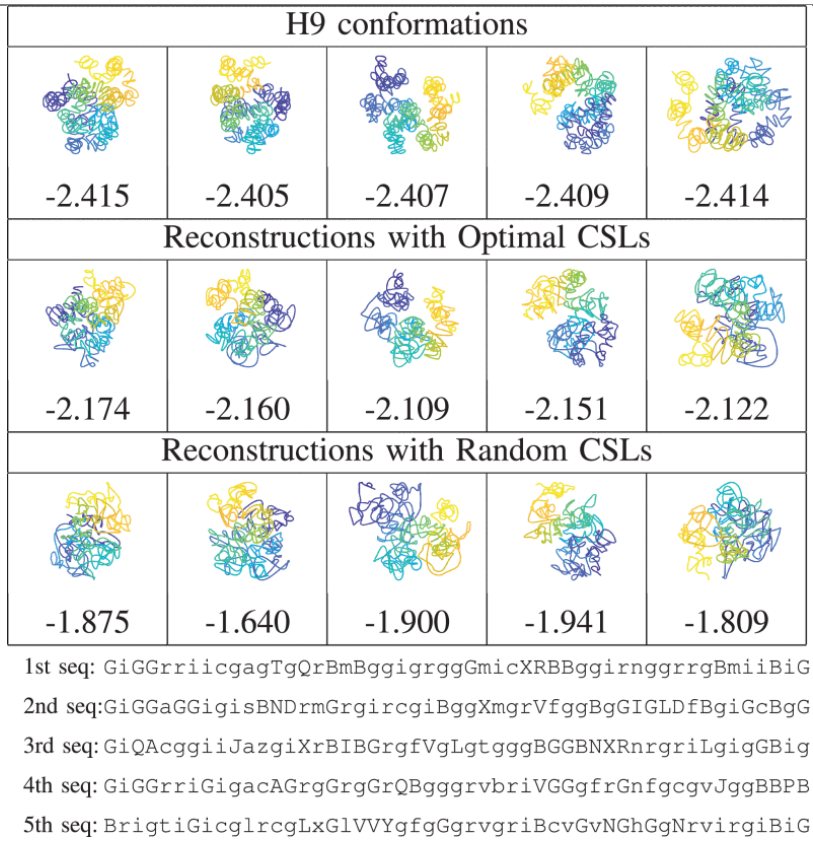


Figure 3.14. Top row: Five conformations from the H9 database. Second row: A corresponding reconstructions of the above conformation from the CSA and the structure alignment algorithm. Third row: A reconstructions of the above conformation with a random sequence of shape letters. Each conformation has its corresponding energy below it. Bottom: The five CSL symbol sequences corresponding to SIMBA3D conformations segmented using TADs of the first row.

To examine the significance of choosing the closest shape letter, we created "reconstructions" with randomly selected shape letters and displayed them in the last row. We see that the reconstructions using random CSLs have much higher energy than those with optimally chosen CSLs. This implies that appropriate CSLs indeed hold crucial structural information about segment shapes. The ground truth conformations naturally have the lowest energy as these are estimated as entirely unconstrained curves in SIMBA3D. This also validates a vital hypothesis that despite vast variability in the chromosomal structures globally, due to differences in estimation techniques, data resolutions, biological variability, etc., the local structures show remarkable consistency and patterns.

Lastly, we display the symbolic CSA sequence representations for some SIMBA3D conformations in Fig 3.14. All the five conformations result from the same contact matrix and therefore share structural similarities.

Next, we analyze the IGM data in the same way as the last section. We remind the reader that IGM conformations are estimated as an ensemble, using objective functions different from what we use here for reconstruction. Therefore, a comparison of the energies of the originals and the reconstructions is not meaningful. Fig. 3.15 shows some IGM conformations in the top row with corresponding reconstructions in the bottom row. We generate several reconstructions (since it is a gradient descent method) for each conformation and select one with minimal reconstruction energy.

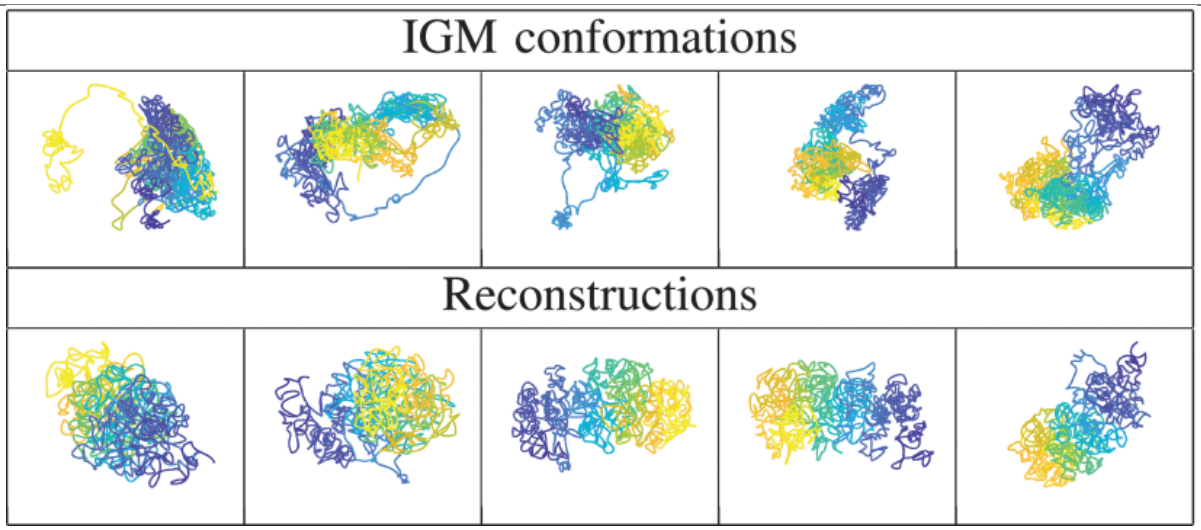
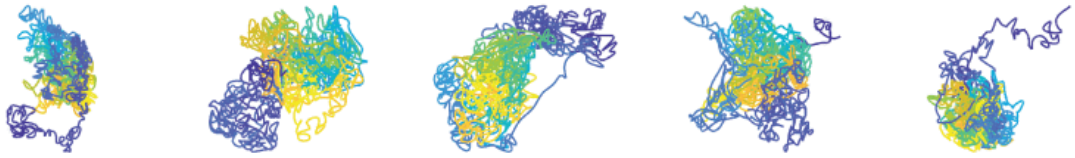


Figure 3.15. Top row: Five conformations from the IGM database. Bottom row: A corresponding reconstructions of the above conformation from the CSA and the structure alignment algorithm.

In this figure, one can see that the reconstructions differ significantly in shape from the IGM conformations. Once again, this can be attributed to having the reconstruction cost function being different from the original criterion in the IGM method. One can address this issue by choosing the IGM energy function for reconstruction, although we have not done that. A strength

of shape-letter representation is that one can always represent a TAD-segmented conformation with a sequence of CSLs. In Fig 3.16 we display some IGM conformations, all from the same chromosome, with their respective symbolic CSA sequences. Since these conformations are all from the same chromosome, they have the same TADs and thus have gaps in the same locations denoted by dots in the sequences. Because of the IGM data heterogeneity, these sequences show more variability than the SIMBA3D sequences.



1st sequence: lil.m.jGVrWcWJ.LJMGN.f.PQl.jjVc.W.xw.YbevCxAcJ

2nd sequence: arX.k.hFVaZfVV.mVrsa.A.WgN.BSTv.J.JL.MKmgcYpAH

3rd sequence: LBz.N.RtlLXxAr.SWiAW.A.aOG.lrIQ.W.Iv.rBWvfcwNw

4th sequence: baI.c.mrVnhbQH.gXjuL.l.Lgc.ecGa.B.zW.HDaZbiNLe

5th sequence: XrC.u.ixZxlWbB.VGlyO.H.JgG.jWAf.Q.QV.TrtNCHrAp

Figure 3.16. Top row: Five conformations from the IGM database. Bottom row: A corresponding reconstructions of the above conformation from the CSA and the structure alignment algorithm.

3.5: Discussion

In this chapter we present a preliminary analysis of different cell lines and conditions (proliferating cells, senescent cells) using the CSA method. We show that CSL have different frequencies, and those with 'loop' structures occur most frequently. Furthermore, we show CSL may have ties to biological features, as different cell lines and conditions have different distributions of CSL at each position in a chromosome, and in each compartment.

Additionally, we present a significant expansion of the CSA method that allows us to process more data more efficiently. We then demonstrate the use of these shape letters in successfully representing, both symbolically and structurally, independent test conformations

taken the same method (IGM) and another method (SIMBA3D). The success of this representation underscores the universal nature of this construction. It emphasizes that the conformations exhibit common patterns at the local, segment level, and one can easily represent them using a few shape letters. These representations via shape letters can be further used to characterize and analyze chromosome populations in a fully statistical manner.

In the future we seek to create more meaningful connections between the shape of the genome and functional biology. As we explored this method and the structures, we became aware that our ability to do this is complicated by the effect that modeling method, resolution, etc. have on CSA. For example, in Figure 3.8 we can clearly see the confounding effects modeling method can have. Although these structures are created from different cell lines, we see that the types of structures they create are distinct. While IGM creates more heterogenous structures, SIMBA3D's are more uniform, and feature distinct looping structures that are not present in the IGM structures. While we have demonstrated that CSA from one method can be used to effectively reconstruct structures from other methods (Figure 3.14) it is difficult to interpret these results in terms of biology, as we lack a 'ground truth' of genomic structures and their heterogeneity across a population.

One potential solution to this problem is working with imaging data. Novel methods such as DNA-MERFISH, which can image and trace the genome across thousands of cells, achieve more accurate structures and depictions of population level heterogeneity than those modeled from Hi-C structures (Su, Zheng, Kinrot, Bintu, & Zhuang, 2020). Although we sought to work with this data, its limited resolution (3 Mb) was not sufficient for connecting structural features to biology. Furthermore, the structures were highly heterogenous and and we did not see an prototypical structural features, as in the modeling data (Figure 3.17).

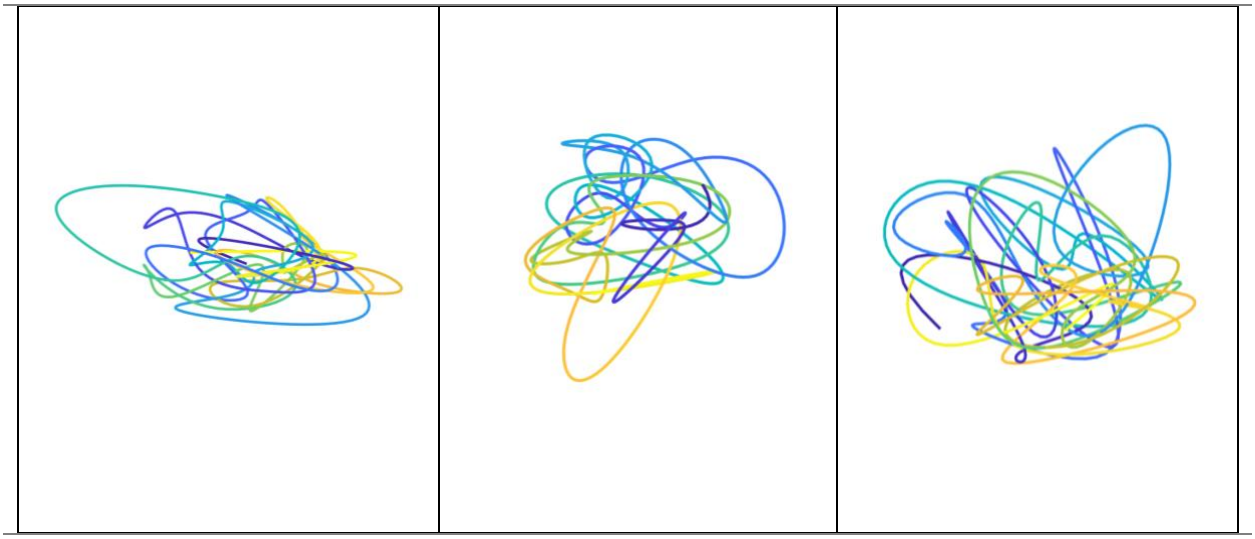


Figure 3.17. Chromosome 1 from n = 3 different cells from Su et al. full genome imaging experiement.

As the field of imaging progresses and we are able to image more of the genome at higher resolution we are confident that the CSA method will become increasingly relevant to connect structure to function. The principles of this method can also be used in other ways, for example, we could use a sliding window to move across chromosomes and compare different structural features, such as curvature, and in this way address structural features, or use the shape distance metric to simply cluster the cells.

Chapter 4: Benchmarking a Novel TAD Caller for use with the Chromosome Shape Alphabet method

In Chapter 4 Audrey Dalgarno was responsible for the benchmarking of TADBay, a novel TAD caller.

4.1: Abstract

Advances in imaging and sequencing techniques provide increasingly informative but complex structural data for understanding the 3D genome. Segmenting 3D conformations of chromosomes into their topologically associated domains (TADs) – structurally homogeneous parts of a chromosome – is an essential tool in shape analysis, such as that presented in chapter 3, which can be used to explore the structural changes that occur with senescence. TAD segmentation allows one to divide complex chromosomal structures into smaller, simpler geometries and facilitates statistical analysis of their shapes. There are several TAD segmentation procedures in the current literature, but they often lack consistency and interpretability and are not ideal for use with the Chromosome Shape Alphabet methodology proposed in chapter 3. In this chapter we assess a novel algorithm for determining TADs, called TADBay, directly from contact matrices, which avoids the difficult step of estimating 3D conformations. While consistent with some current top-performing existing callers, it has an added ability to provide a hierarchical organization of TADs and subTADs. To validate TADBay, in this chapter we utilize simulated contact matrices with ground truth TADs which can be used to benchmark TAD callers. We demonstrate the strengths of our caller through both simulated data and the IMR90 real dataset.

4.2: Introduction

Recent advancements in genomic imaging and data preprocessing techniques have led to more informative and complex data for inferring 3D genome (Dekker, Rippe, Dekker, & Kleckner, 2002), (Lieberman-Aiden et al., 2009). These advancements have enabled research on analyzing structural properties of 3D chromosome conformations, such as differentiating A and B

compartments (Lieberman-Aiden et al., 2009), estimating 3D conformations from contact data (Hua et al., 2018), (Rosenthal et al., 2019), forming topologically associated domains (Dixon et al., 2012), conformation loops (Rao et al., 2014), and so on. These techniques can be applied to studying structures for individual cells or ensemble data for multiple cells.

Since 3D structures of the whole genome and even individual chromosomes are quite complex, one often takes a divide-and-conquer approach to analyze chromosomal morphometry. One can divide or partition chromosomal conformations into structurally homogeneous parts, called Topologically Associated Domains (TADs), and analyze these smaller structures individually or jointly across cells and chromosomes. TADs are fundamental unit structures of a 3D chromosome and are routinely used in studying their relationship to gene expression (Acemel, Maeso, & Gómez-Skarmeta, 2017), (Krefting, Andrade-Navarro, & Ibn-Salem, 2018), their shape across populations (C. J. Soto, Zhao, Klein, Gilbert, & Srivastava, 2021), (W. Liu, Srivastava, & Zhang, 2011), their shape commonality across modalities, in shape alphabet learning (C. Soto, Dalgarno, et al., 2021), and many more. Thus, there is a critical need to perform reliable and interpretable TAD segmentations in chromosome data to be able to systematically investigate structural changes and with conditions such as senescence.

TAD segmentation can be performed using the Hi-C contact matrix data, the estimated 3D conformations, or a combination of the two. It seems more natural to work with the raw data, i.e., the contact matrices, since the estimation of 3D conformations is problematic and error-prone. However, contact matrices, representing a measure of interactions, are noisy and sparse. Thus algorithms need to account for noise and sparsity by adding prior terms in the objective functions. Different choices of objective functions and prior knowledge in different methods lead to inconsistencies across methods for TAD segmentations (Zufferey et al., 2018). For instance, some

methods allow adjacent TADs to overlap and intersect, some have adjacent TADs only sharing the boundaries, and some have TADs covering the entire contact map. Further, the methods differ in terms of the optimization techniques and the number of parameters. These differences lead to fundamental structural variability in what TAD callers produce and reproducibility issues. Some familiar TAD callers have been reviewed and compared as in (Zufferey et al., 2018).

In this chapter, we benchmark TADBay, a TAD caller that works directly with the contact matrix. It produces TAD segmentations in Hi-C data which:

1) has one parameter, theta, which is tunable to find TADs of varying sizes;

2) determines a TAD hierarchy of TADs and subTADs;

3) and covers the entire contact map, including the endpoints.

In the following, we demonstrate our caller's strengths and compare TADBay to other state-of-the-art TAD callers. We compare TADBay to other callers using existing metrics and visually display TADs with apparent differences.

4.3: Methods

4.3.1: TADBay

TADs were called using a novel TAD caller, TADBay. In short, this method uses Bayesian sequential clustering, forms a composite energy function with different terms, and solves an optimization through stochastic simulated annealing to call TADs as in (C. Soto et al., 2022).

4.3.2: Simulated Data

Simulated contact matrices were generated as in (C. Soto et al., 2022) for comparing TADBay to other callers.

4.3.3: Determining concordance between TAD callers

Two methods were used to compare the overlap between TAD callers. In the case of the simulated data, the boundaries were simply compared for overlap (i.e. we determined if the boundaries between each TAD caller were the same). For comparison to the Hi-C data, we used the measure of concordance (MoC) metric introduced in (Pfitzner, Leibbrandt, & Powers, 2009) and adapted for comparisons of TADs as in (Zufferey et al., 2018). The MoC measures agreement of sets of TADs, taking into account TAD size, and returns a value between 0 and 1 where 1 is complete agreement.

4.3.4: Hi-C Data

IMR90 Hi-C data at 5 kb and 50 kb from (Rao et al., 2014) was used to benchmark the TAD caller. Data was extracted from .hic files using juicertools 'dump' command (Durand et al., 2016) Chromosomes 2, 12, and 21 were used, as they represent a range of chromosome sizes.

4.3.5: TAD Callers

In addition to our novel TAD caller TADBay we used TopDom (Shin et al., 2016) with a window size of 5 or 10, and HiCSeg (Lévy-Leduc, Delattre, Mary-Huard, & Robin, 2014) with the "G" (gaussian) distribution and the "D" block diagonal model. For the maximal number of change points we determined this based on the number of TADs expected for the matrix size if TADs were ~1 Mb. These were compared to the ground truth in simulated data, and to a uniform distribution with the IMR90 data.

4.3.6: CTCF Data

The CTCF narrow peak ChIP-Seq track was retrieved from the ENCODE portal (Sloan et al., 2016) (https://www.encodeproject.org/) with the following identifier: ENCSR000EFI. Data was lifted from hg19 to hg38 using UCSC's liftover tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver).

4.4: Results

To contextualize the performance of our TAD caller, we compare our results to that of other TAD callers. We reference a sizeable comparative study (Zufferey et al., 2018), which uses several measures to compare numerous TAD callers and concludes that HiCSeg (Lévy-Leduc et al., 2014) and TopDom (Shin et al., 2016) are top perfomers.

We compare our TAD caller to HiCSeg, TopDom, and a naïve uniform segmentation on simulated data and real data (Hi-C) examples. For each comparison, we compute the measure of concordance and the distance from the segmentation sites to CTCF for the real data examples.

4.4.1: Simulated Data

For the initial testing of our TAD caller, we turn to simulated data. Simulated data is helpful in this situation because we have a ground truth solution in contrast to experimental data.

We first create a simple example of simulated Hi-C data with precise, distinguishable TADs (Figure 4.1). Figure 4.1 also displays a Venn diagram comparing the computed boundaries called by our TAD caller, HiCSeg, and TopDom with the ground truth solution. We see that the calls are almost identical, the only difference being that TopDom, with a window size of 5 calls an additional eight boundaries.

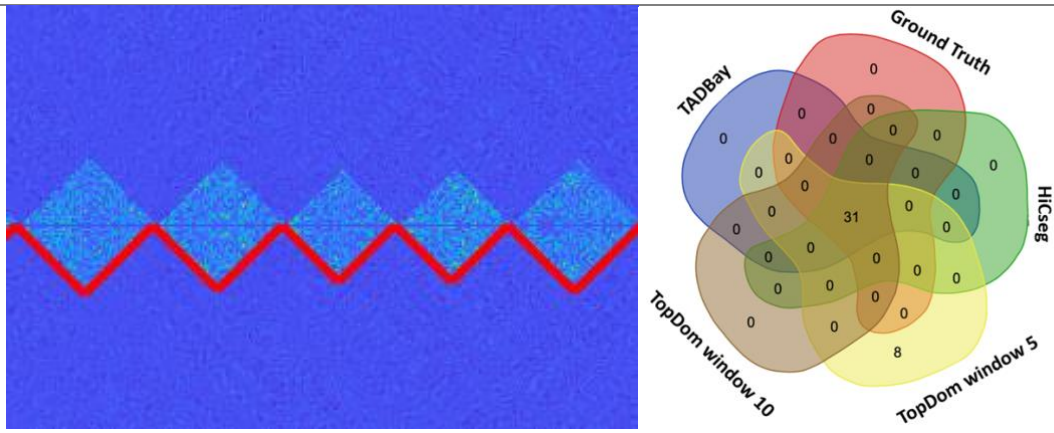


Figure 4.1. Left: Simulated data with a simple structure. Right: Overlap of TAD boundaries for simpler simulated data.

Figure 4.2 displays a portion of the simpler simulated data example with our caller's TADs in red and TopDom at a window size of 5 in black. We see some agreement in this portion of the matrix, but TopDom calls extra boundaries. Visually we see no cause for these different boundaries. We do not display HiCSeg as it entirely agrees with our caller and the ground truth.

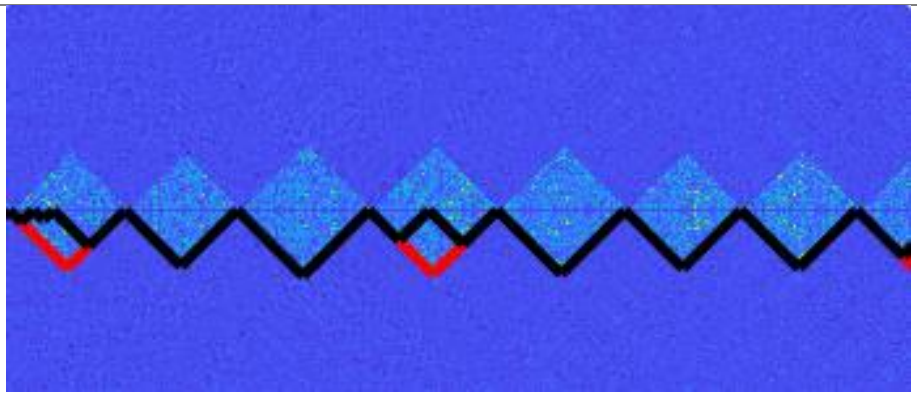


Figure 4.2. The simpler simulated data with our TADBay in red and TopDom With a window size of 5 in black.

We next compare performance under the more complex simulated data (Figure 4.3). Figure 4.3 also displays the results via a Venn diagram. We again see that our caller, HiCSeg, TopDom, and the ground truth solution were almost identical, with the only difference being that HiCSeg misses two boundaries.

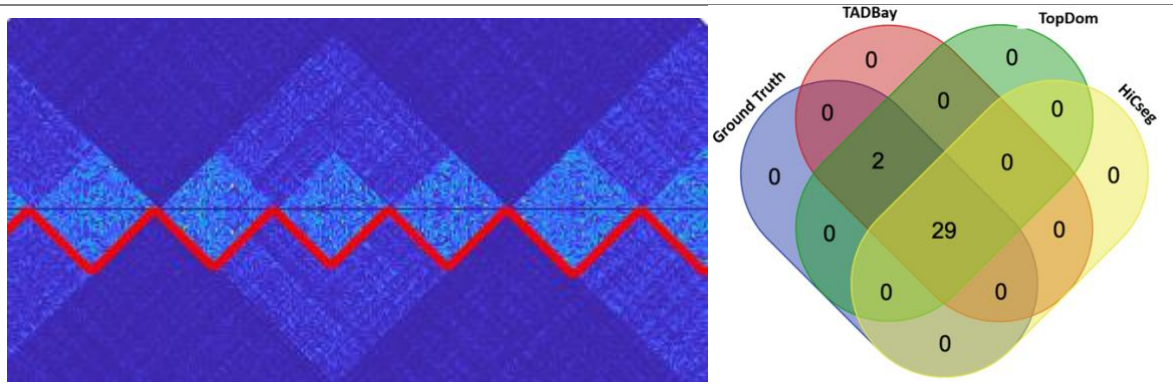


Figure 4.3. Left: Simulated data with a complex structure. Right: Overlap of TAD boundaries for complex simulated data.

Further, the Figure 4.4 displays the more complex simulated data example with our TAD caller in red and HiCSeg in yellow. We see that HiCSeg misses two TADs but captures a larger superTAD structure. While there is usefulness in capturing the larger structures, in this case, HiCSeg does not consistently capture them.

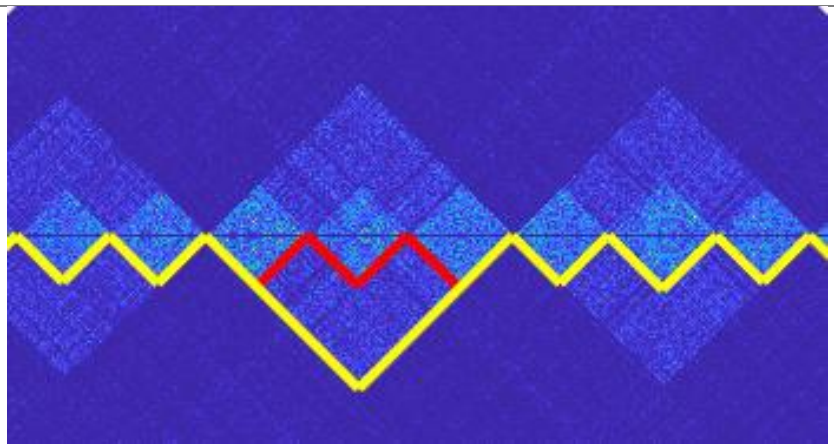


Figure 4.4. Complex simulated data with our TADs in red and HiCSeg in yellow.

In both cases of simulated data, our caller detects the ground truth precisely. Nevertheless, a top TAD caller failed to capture all the TADs in each case. In the complex simulated example, HiCSeg misses a couple of boundaries but captures a superTAD instead. A more problematic result is in the simpler simulated data example where TopDom, with a window size of 5, finds extra boundaries that do not appear to have any structural relevance.

4.4.2: Real Data

We also test our TAD caller on real experimental data – high-resolution lung fibroblast (IMR90) Hi-C data from (Rao et al., 2014)**.** We use data from several chromosomes to explore the perfomance of our TAD caller across the hierarchically organized 3D genome. In addition to comparing TADBay to HiCseg and TopDom, we also compare it to a naive uniform segmentation.

For the uniform segmentation, we set equally spaced boundaries as TADs where the number of TADs is comparable to the other TAD callers. Further, for the uniform segmentation, we constrain the TADs between 500 and 1000KB on their size.

In addition to computing a pairwise measure of concordance (MoC) for each caller we also consider another metric for evaluation. Literature supports that biological features such as CCCTC-binding factor (CTCF) are enriched at TAD boundaries (Rowley & Corces, 2018; Zufferey et al., 2018). That is, we expect

to find a CTCF site at, or near, each TAD boundary so we can thus use this to measure performance. We therefore compute, for each caller, the distance from each respective TAD boundary to the nearest CTCF site.

We first compare performance of all callers and the uniform segmentation on chromosomes 2, 12, 21 at 50KB resolution. We select chromosomes 2, 12, and 21 as instances of large, medium, and small chromosomes, respectively. For each caller we report the top quartile and average of distances from each TAD to the nearest CTCF site for chromosome 2 at 50 kb resolution (Figure 4.5). Table 4.1 displays the average and median distances for chromosomes 12 and 21. At this resolution and considering the top quartile of distances, our caller performed similarly to HiCSeg and TopDom. Further, TADBay found TADs closer to CTCF boundaries when compared to the uniform caller.

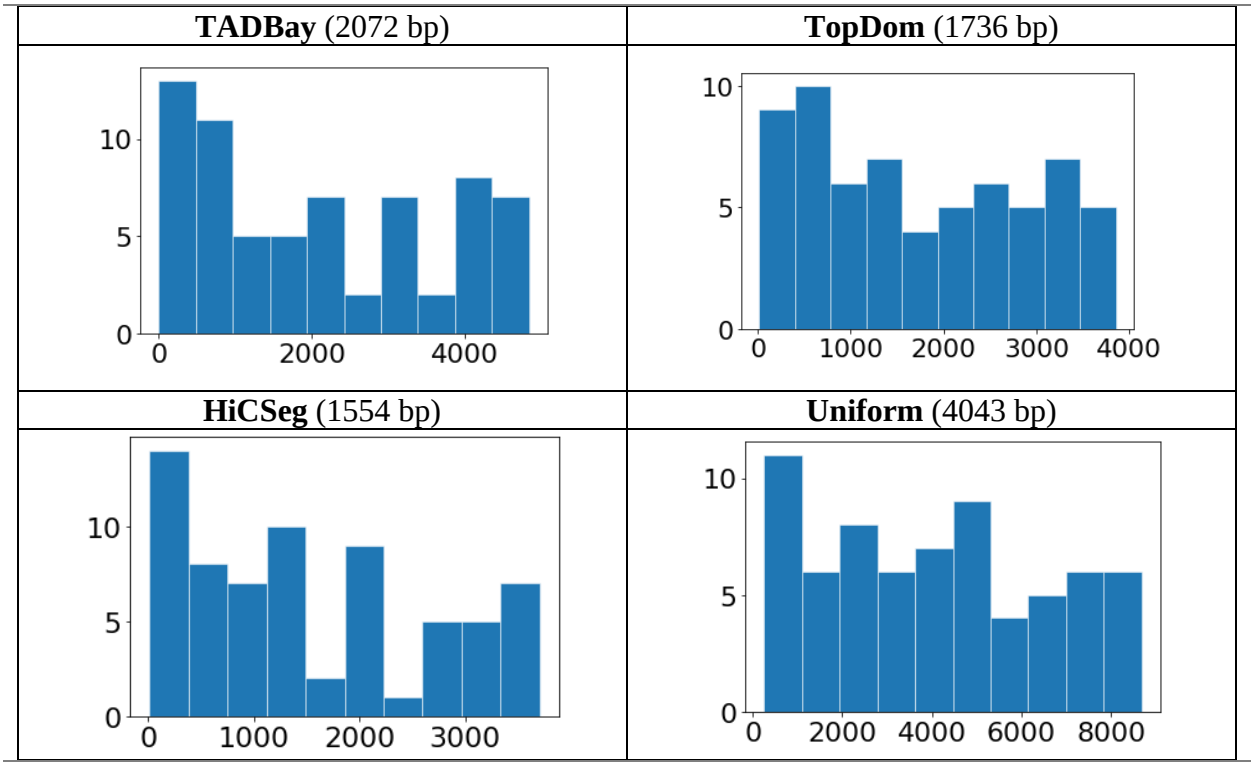


Figure 4.5. Histograms of the top quartile of distances between CTCF motifs and HiCseg, TopDom, our caller, and a uniform segmentation of chromosome 2 at 50KB resolution. Means shown in parentheses.

Table 4.1. Average and (median) distances from CTCF sites on chromosomes 12 and 21.

Chromosome	TADBay	TopDom	HiCSeg	Uniform
12	2513 (2553)	2087 (1912)	1741 (1670)	4302 (4723)
21	5226 (4271)	5374 (3594)	5007 (4271)	11409 (10380)

Table 4.2 displays the MoC comparison. At the 50KB resolution for each chromosome, our caller is more similar to HiCseg and TopDom than to the uniform segmentation. While this is not in and of itself evidence of finding TADs, being more similar to well-established TAD callers than a naive method in conjunction with the CTCF site distances is evidence of truly finding TADs. Further, according to the MoC, TADBay is similar to HiCSeg, which in almost all cases had a minimal average and median distance to CTCF sites.

Table 4.2. Measure of Concordance (MoC) comparison between HiCSeg, TopDom, our caller TADBay, and a uniform segmentation of chromosome 2, 12, and 21 at 50kb resolution.

Chromosome 2

	TADB.	TopD.	HiCS.	Unif
TADB.	1	0.72	0.75	0.61
TopD.	-	1	0.79	0.54
HiCS.	-	-	1	0.58
Unif.	-	-	-	1

Chromosome 12

	TADB.	TopD.	HiCS.	Unif
TADB.	1	0.58	0.64	0.58
TopD.	-	1	0.62	0.53
HiCS.	-	-	1	0.53
Unif.	-	-	-	1

Chromosome 21

	TADB.	TopD.	HiCS.	Unif
TADB.	1	0.63	0.60	0.60
TopD.	-	1	0.74	0.51
HiCS.	-	-	1	0.46
Unif.	-	-	-	1

Average

	TADB.	TopD.	HiCS.	Unif
TADB.	1	0.64	0.66	0.59
TopD.	-	1	0.71	0.53
HiCS.	-	-	1	0.52
Unif.	-	-	-	1

We also perform a similar analysis on the p arm of chromosome 2 at 5KB resolution. Due to the hierarchical nature of TADs and the observed presence of structures such as subTADs, we run the callers to yield TAD-like structures at 500KB and 1MB. Table 4.3 presents the pairwise MoC comparison of the callers. Figure 4.6 displays the histograms and averages of the top quartile of distances from the CTCF sites. Considering these distances, we outperform TopDom and the Uniform segmentation. HiCSeg performed very well at this resolution. Our caller is again more similar to the HiCseg caller than the uniform segmentation at both resolutions. Also, the TopDom caller yielded significantly less control over TAD size, as the only parameter the user controls is the window size. The TADs called by TopDom are much smaller, resulting in poor MoC values with other callers. This is a critical point to consider when selecting a TAD caller. Additionally, our caller gives flexibility in the choice of its hyperparameters.

Table 4.3. Measure of Concordance (MoC) comparison between HiCSeg, TopDom, our caller TADBay, and a uniform segmentation of chromosome 2's p arm at 5 kb resolution.

500 kb

	TADB.	TopD.	HiCS.	Unif
TADB.	1	0.21	0.53	0.50
TopD.	-	1	0.33	0.30
HiCS.	-	-	1	0.52
Unif.	-	-	-	1

1 Mb

	TADB.	TopD.	HiCS.	Unif
TADB.	1	0.17	0.59	0.50
TopD.	-	1	0.24	0.22
HiCS.	-	-	1	0.49
Unif.	-	-	-	1

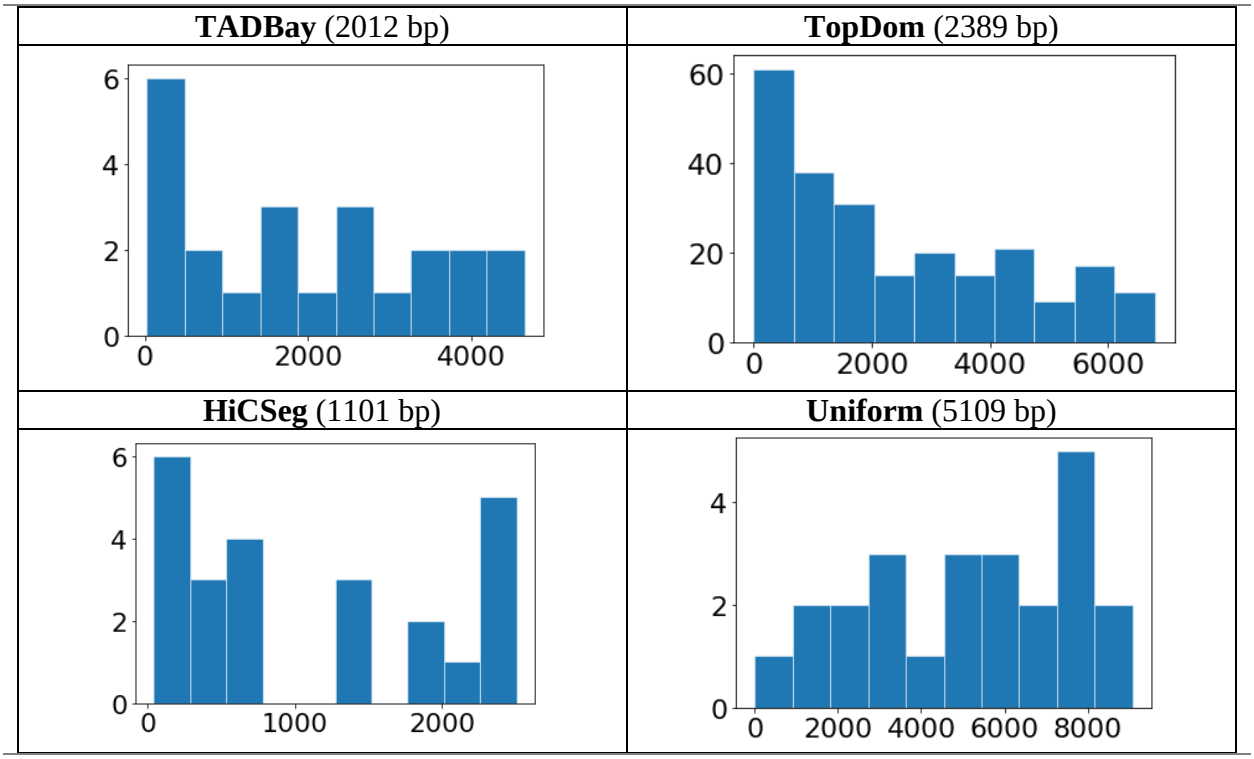


Figure 4.6. Average distances (top quartile) between CTCF motifs and HiCseg, TopDom, our caller, and a random segmentation of chromosome 2's p arm at 5KB resolution.

From these examples, our caller performs similarly on experimental data to HiCseg and TopDom, the state-of-the-art TAD callers (Zufferey et al., 2018). Further, HiCSeg tends to perform the best under our evaluation metric, and our caller was consistently most similar to HiCSeg of the compared callers. At the 5KB resolution, TADBay outperforms the TopDom caller since it cannot

control the size of TADs very well, while we have the flexibility to do so. We thus see that TADBay is both a consistent and adaptive TAD caller.

4.4.3: Hierarchical Structure

As previously mentioned, TADBay has two tuning parameters, a constant multiplier in the initial θ and the expected number of TADs. Through exploratory analysis, however, we noticed that the multiplier in theta has the most influence on the resulting number of TADs. The multiplier in θ seemingly determines the approximate size of the TADs and number of TADs. The larger the theta multiplier, the more TADs we find and hence the smaller the average TAD. We leverage this inverse relationship at each application.

Figure 4.7 displays the effect of the theta multiplier tuning parameter on three independent runs of the algorithm. The resulting TADs from a theta multiplier of 1.6 are displayed in red, 10 in green, and 30 in pink. As theta increases, the TADs seemingly divide into subTADs mainly preserving the boundaries with the occasional slight deviation. This hierarchical structure is not enforced into the algorithm but rather an artifact. That is, one can tune theta to fit their needs of finding TADs or subTADs of the contact map. Many other TAD callers lack the flexibility of finding TADs of varying size, while this is a strength of TADBay.

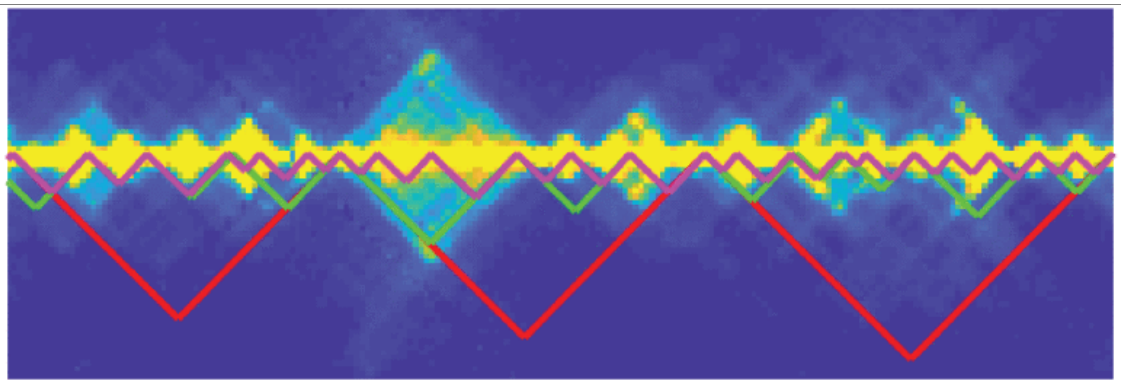


Figure 4.7. A portion of chromosome 21 at 50KB from the IMR90 database. Three sets of TADs are plotted with a θ multiplier of 1.6 in red, 10 in green, and 30 in pink.

4.5: Discussion

This analysis presented in this chapter benchmarks the novel algorithm, TADBay, for determining topologically associated domains. We indicate that TADBay performs similarly to the state-of-the-art TAD callers on both simulated and real data. TADBay performed flawlessly on our simulated data and reasonably well on the experimental data.

TADs are difficult to detect consistently and correctly, most apparent in that TAD callers rarely agree with each other completely. Thus there are many opportunities for future research. It would be interesting to consider the 3-dimensional conformation of the chromosome instead of the contact matrix. However, to our knowledge, there is no rich dataset containing contact matrices and actual conformations. One could, of course, estimate the conformation from the contact matrix; however, this may lead to a correlation between estimation techniques and the resulting TADs.

In the context of this thesis, we now have a novel TAD caller, TADBay, which is able to create chromosome segmentations effectively, tunably, and specifically compatibly with the Chromosome Shape Alphabet (CSA) method outlined in Chapter 2. To reiterate, the CSA method can be used to determine prototypical shapes in 3D structure and this can help us to connect structure to biological function, which is critical, especially in situations where nuclear structure is significantly altered, such as with cellular senescence.

Chapter 5: Uncovering the Role of Senescence in Epigenetic Reprogramming

In Chapter 5 Audrey Dalgarno was responsible for the performing the wet lab experiments (10x multiome) and the computational analysis, which she performed with the help of Andrew Nunez. Mousework and epigenetic reprogramming was performed by members of Vittoro Sebastiano's Lab at Stanford.

5.1: Abstract

To slow the pace of aging, a portion of recent aging research has focused on aging interventions. Many of these interventions have focused on clearing (senolytics) and modulating (senomorphics) senescent cells, which accumulate in the body with age and become damaging, contributing to chronic inflammation. Another emerging rejuvenation method is epigenetic reprogramming. Partial reprogramming through expression of the Yamanaka Factors (OSKM, and variations) can render signatures of aging (epigenetic clock, inflammation, age-associated visual impairment, etc.) more youthful. However, as we are early in our understanding of how this process works, many aspects of this mechanism are poorly understood. In this chapter we focus on outlining the role of senescent cells in murine hepatic rejuvenation. In our analysis we approach reprogramming from a multiomic standpoint at the single cell level, examining the transcriptome (scRNA-seq) and chromatin landscape (scATAC-seq). Although we show data supporting that the RNA associated with this multiomic analysis is of prohibitively poor quality, we begin an exploration of the changes to chromatin accessibility that occur with epigenetic reprogramming. We determine motifs critical for reprogramming at early and late stages, which are distinct from each other. The results encompassed in this chapter represent an early stage of this project's analysis, which will be an ongoing effort in the Neretti lab.

5.2: Introduction

As the field of aging research has developed, one emphasis has been on developing ways to improve healthspan and lifespan, as aging is the primary risk factor for chronic diseases such as cancer and diabetes (López-Otín et al., 2013). These interventions have included everything from senolytics (Kirkland & Tchkonia, 2020) to caloric restriction (McCay, Crowell, & Maynard, 1935), and now epigenetic reprogramming.

Epigenetic reprogramming entails the reversion of cells into a pluripotent state (Takahashi & Yamanaka, 2006). This concept was first introduced in by Takahashi and Yamanaka, who used the four transcription factors OCT4, SOX2, KLF4 and MYC (OSKM) to transform cells into induced pluripotent stem cells (iPSCs). Alternative factors such as IL6 (L. Li et al., 2023; Mosteiro et al., 2016; H. Zhu et al., 2022) and TGF-beta (W. Li, Wei, & Ding, 2016) have also been implicated as reprograming agents. Although revolutionary, the full de-differentiation of cells would not be feasible in the context of aging, and this process carries risk, such as the generation of teratomas. Instead, researchers in the field of aging have turned to partial reprogramming, through techniques such as cyclical reprogramming (Ocampo et al., 2016), or simply incomplete reprogramming timecourses (Sarkar et al., 2020).

Studies on partial reprorgramming have definitively shown that administration of OSKM, and variants, can induce rejuventation in different organisms (mouse, human cells) and tissues (retina, spleen, liver, blood, pancreas, etc.), as it can revert measures of aging such as biological age, as measured by epigenetic clock (Chondronasiou et al., 2022; Lu et al., 2020; Mosteiro et al., 2016; Ocampo et al., 2016; Sarkar et al., 2020). However, due to the novelty of this field many aspects of epigenetic reprogramming are not well understood. For example, the role of senescent cells has not been clearly elucidated. It is known that senescent cells beneficially contribute to

reprogramming, as they secrete factors such as IL6, which aids with de-differentiation (Mosteiro et al., 2016). However, *Drosophila* studies have indicated that a combination of senolytics and partial reprogramming (cyclical) extends lifespan more than each intervention alone (Kaur et al., 2022), indicating that they conversely may not help the reprogramming process.

To reconcile these findings, determine the role of senescent cells in age-associated rejuvenation, and more concretely determine the epigenetic landscape of epigenetic reprogramming we completed a multiomic (RNA-seq, ATAC-seq) analysis of mouse liver at the single-cell level. Furthermore, we explored epigenetic reprogramming at two different time points – immediately after partial reprogramming, and 4 weeks after reprogramming. While the single cell transcriptomic assay was of poor quality, we have begun to explore the chromatin accessibility landscape of epigenetic reprogramming. We determine the early and late stages of reprogramming are controlled by Signal Transducers and Activators of Transcription (STAT) family and Krüppel-Like Factors (KLF), respectively. Lastly, we conclude with an outline of potential future directions which can be explored with this dataset.

5.3: Methods

5.3.1: Mousework

Fresh frozen mouse liver were prepared from young (3 month) and old (12 month) i4F (Mosteiro et al., 2016) mice. I4F mice, or "inducible four factors" mice express OCT4, SOX2, KLF4, and cMYC under doxycycline (dox) control. There were 4 conditions of old mice – mice treated with and without dox for 7 days at two different time points: either harvested immediately after the 7 days of dox treatment (early reprogramming), or 4 weeks after dox treatment (late reprorgramming). All mice included in this analysis were female, with each sample coming from

a different mouse (biological replicates, n = 2). Mice were perfused with PBS and pieces (~1/6) of their liver was then harvested and fresh frozen in liquid nitrogen.

5.3.2: scMultiome experiments

10x's multiome kit (scRNA and scATAC) was used for this analysis. Nuclei were extracted as in 10x's "Nuclei Isolation from Complex Tissues for Single Cell Multiome ATAC + Gene Expression Sequencing" protocol with a 3 minute lysis time, homogenization performed with a disposable pellet pestle and cordless motor, and a 1.8 M sucrose gradient replacing the FACS sorting. Nulcei quality was examined after extraction using tryphan blue and determined to be sufficient for each sample. We targeted 10k nuclei per sample for the multiome experiment. The protocol was completed as in 10x's "Chromium Next GEM Single Cell Multiome ATAC + Gene Expression User Guide" manual. Sequencing was completed on an Illumina NovaSeq as outlined in the user guide.

5.3.3: Computational Analysis

The sequenced samples were demultiplexed and aligned to mm10 with cellranger arc version 2.0.1. The filtered feature bc matrix and atac fragment files were analyzed using Seurat (Hao et al., 2023) and Signac (Stuart, Srivastava, Madad, Lareau, & Satija, 2021), largely according to Signac's "Joint RNA and ATAC analysis: 10x multiomic" vignette, as well as other Seurat and Signac vignettes.

5.4: Results

There is little known about the role of senescent cells in the context of epigenetic reprogramming, an accelerating area of aging research. We sought to clearly elucidate the role of senescent cells, and additionally define the epigenetic landscape of murine hepatic reprogramming through use of the 10x multiome kit, which assays gene expression (scRNA) and chromatin

accessibility (scATAC) at the single-cell level (Figure 5.1). We examined two different time points of reprogramming – immediately after reprogramming (D0 after 7 days of Dox treatment, "early reprogramming") and 4 weeks after reprogramming ("late reprogramming"). For the analysis contained in this chapter we discuss the older timepoints only, as the young timepoint has yet to be assayed with the 10x multiome kit and sequenced. The one missing early reprogramming sample will also need repeating, due to insufficient nuclei yield.

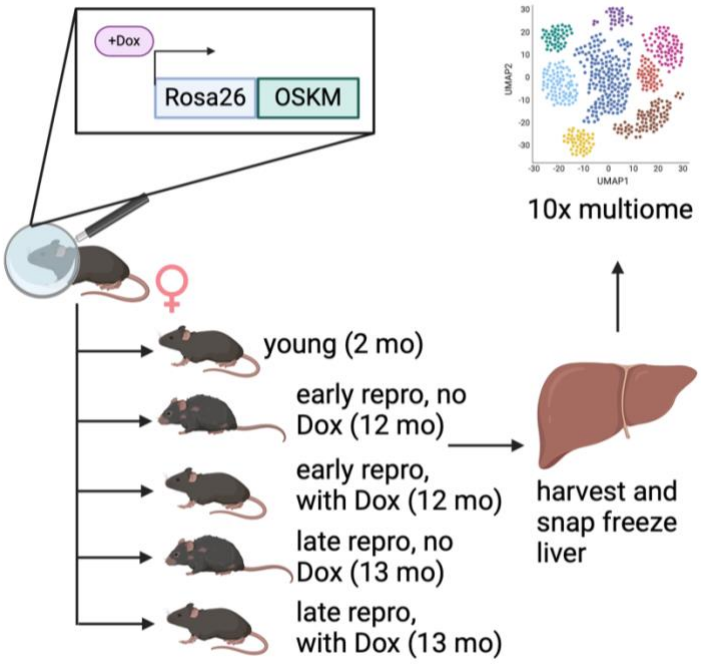


Figure 5.1. Overall methods. Liver was extracted and snap frozen from n = 2 female mice at "early" (D0 after 7 days of Dox treatment) and "late" reprogramming timepoints (4 weeks after reprogramming. 10x multiome was performed on the resulting samples.

After aligning the data to mm10 we were able to examine general summary statistics about the quality of the data. While some samples looked better than others, it appeared that the ATAC data was of much higher quality than the RNA (Table 5.1). After filtering using the default parameters we generally kept ~30% of the cells. Samples that were overloaded significantly (138, 162) had the lowest numbers of cells retained after filtering.

Table 5.1. Sample quality control, pre- and post- standard filtering.

Sample Name	Sample Type	n Cells	ATAC median	RNA median	n Cells post-filter default	% cells retained
138	4 week + dox	18,852	3,912	116	1551	8
152	4 week - dox	10,246	8,516	480	3042	30
153	4 week - dox	9,412	7,440	408	2021	21
162	7 days + dox	20,000	5,746	170	907	5
164	7 days + dox	9,747	11,414	718	4959	51
165	7 days - dox	14,931	8,817	856	8759	59
166	7 days - dox	11,897	6,199	299	2153	18

After merging the seven samples together we were able to generate a Uniform Manifold Approximation and Projection (UMAP), pictured below. As this is a multiomic assay, we can view the RNA and ATAC separately (Figure 5.2A/B), or jointly (Figure 5.2C). Generally, the samples clustered by sample type, rather than cell type, indicating a significant, and likely problematic, amount of heterogeneity.

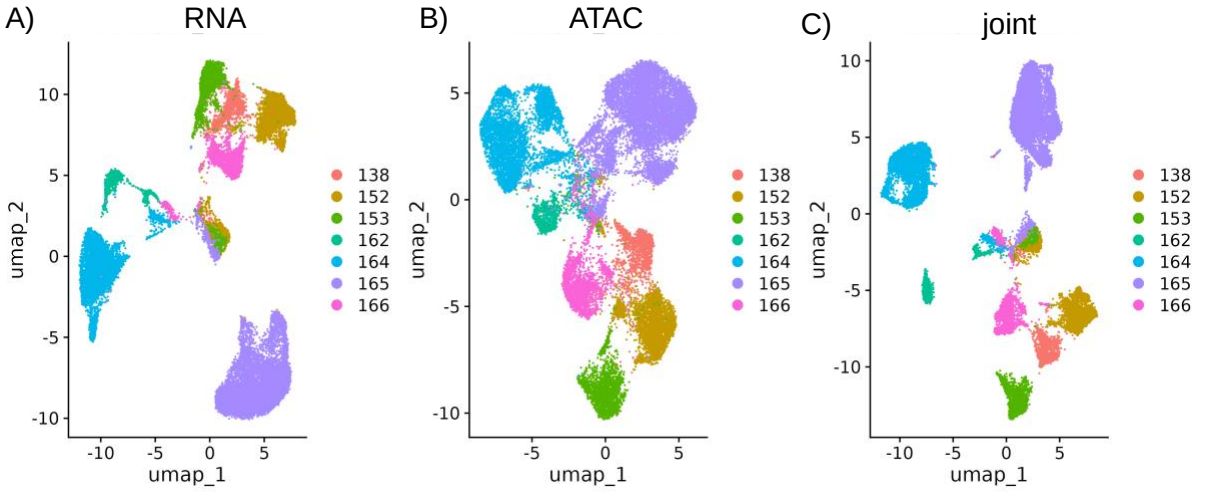


Figure 5.2. Uniform Manifold Approximation and Projection (UMAP) plots for multiome data. A) RNA only UMAP B) ATAC only UMAP C) Combined RNA and ATAC UMAP. Colors indicate sample names (see Table 5.1).

Before proceeding with our analysis, we examined some important aspects of the gene expression data. First, we looked at the expression of several critical senescence markers.

Unfortunately, the key senescence marker Cdkn2a (p16) was not captured in any cells and Cdkn1a (p21), another marker, was captured in very few (Figure 5.3).

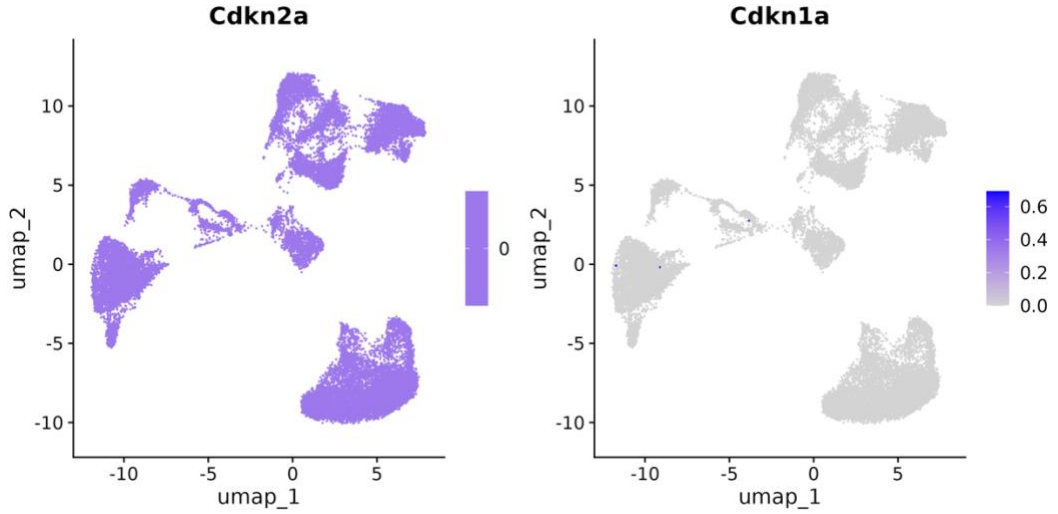


Figure 5.3. Normalized expression for senescence markers Cdkn2a (p16) (left) and Cdkn1a (p21) (right).

Furthermore, we tried our best at performing cell typing with this data. First, we used Seurat's standard clustering resolution (0.8), which identified nineteen different clusters (Figure 5.4). In this case, many different samples were classified as entirely one cluster, which we know not to be true, as each sample should contain multiple cell types. Additionally, in the samples that were divided into multiple clusters, representing different cell types, these were not shared across samples, accounting for the high number of clusters. This is also problematic, as we would expect to see shared cell types across the samples. We also tried using a lower clustering resolution to identify fewer clusters (Figure 5.4), as we know from reference datasets there should relatively few (Tabula Muris Consortium, 2020). In this case we saw that each sample clustered individually, and we did not see the heterogeneity we would expect from having different cell types in each sample. Next, we performed cell typing using the Tabula Muris Senis (Tabula Muris Consortium, 2020) data with Seurat's reference mapping. Although we were able to map the predicted IDs,

again we saw entire sample populations mapped to specific cell types (Figure 5.4), which we know to be inaccurate as again, each sample should contain multiple cell types.

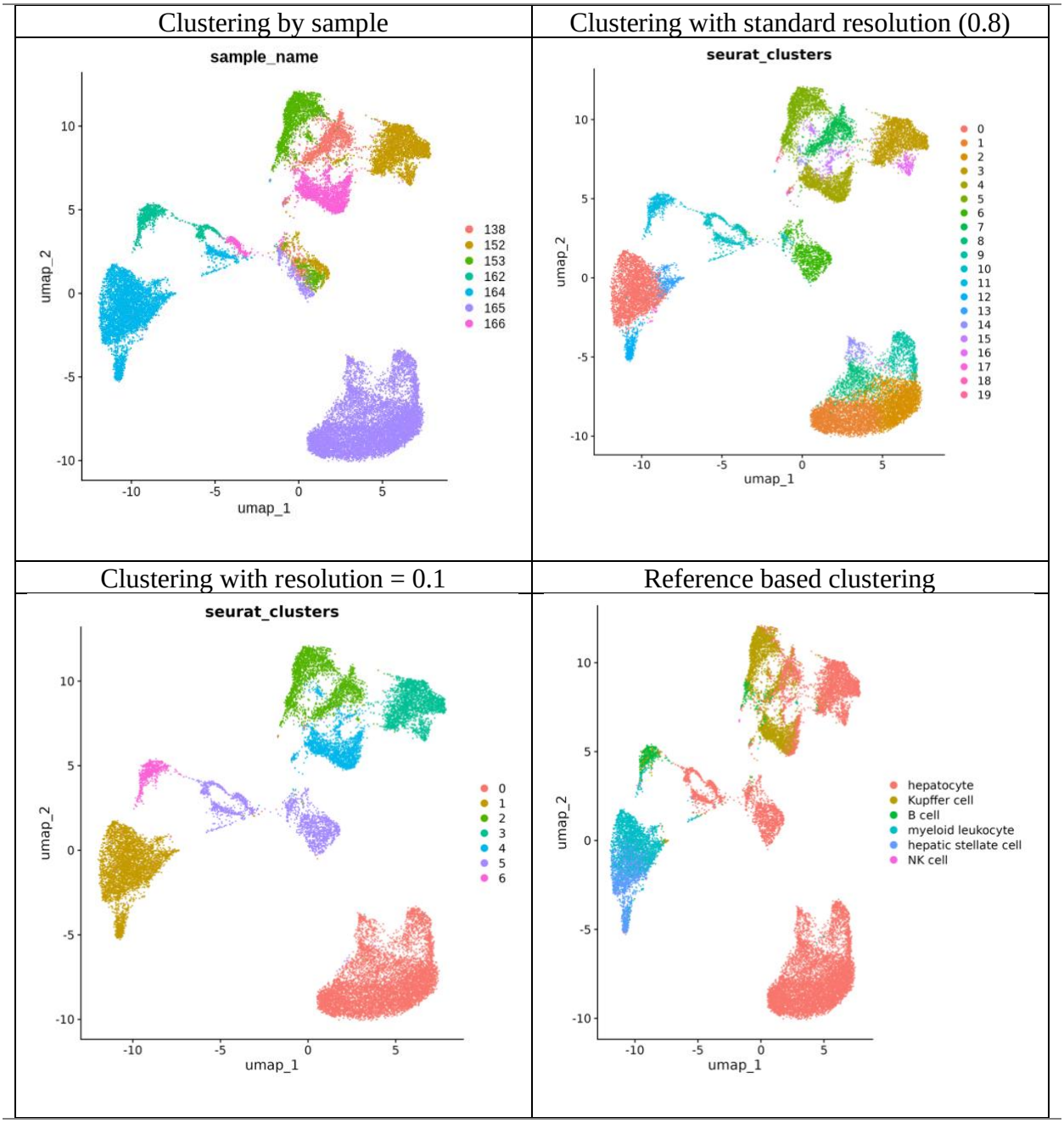


Figure 5.4. (top left) UMAP colored by sample name. (top right) UMAP colored by standard Seurat clustering. (bottom left) UMAP colored by Seurat clustering with low resolution. (bottom right) UMAP colored by Seurat's reference-based cell typing.

Taken together, these data indicate that the RNA was of insufficient quality for our analysis. We proceeded to look at the ATAC data alone, to see if we could elucidate more biological information from this dataset. By filtering using just the ATAC data we were able to retain a significantly higher percentage of the cells, supporting our conclusions that the RNA associated with this assay were of poor quality (Table 5.2). A UMAP representation of this data is represented in Figure 5.5.

Table 5.2. Sample quality control, pre- and post-filtering using the ATAC data only.

Sample Name	Sample Type	n Cells	ATAC median	RNA median	n Cells post-filter ATAC	% cells retained
138	4 week + dox	18,852	3,912	116	18262	97
152	4 week - dox	10,246	8,516	480	10020	98
153	4 week - dox	9,412	7,440	408	8952	95
162	7 days + dox	20,000	5,746	170	19954	99
164	7 days + dox	9,747	11,414	718	9517	98
165	7 days - dox	14,931	8,817	856	14702	98
166	7 days - dox	11,897	6,199	299	11586	97

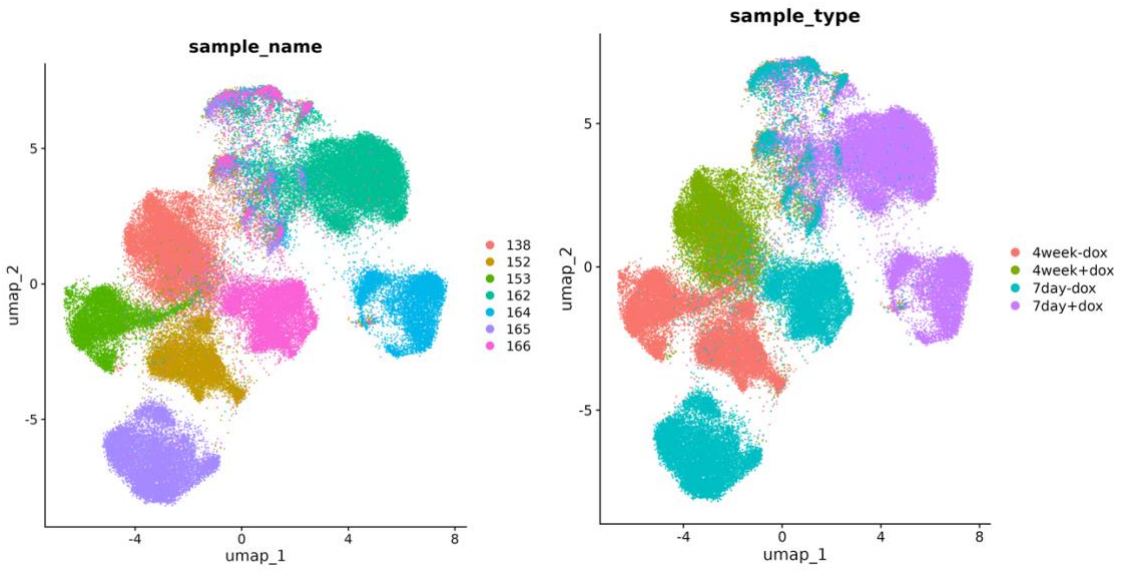


Figure 5.5. UMAP of ATAC-seq data grouped by sample (left) and condition (right)

Although generally the data clusters less distinctly by sample type when we examined the ATAC data alone (Figure 5.5) as comparted with both modalities (Figure 5.2) there is still

heterogeneity between the samples, which can potentially be problematic for downstream processes such as cell typing. We explore some of the sources of this variation by delving deeper into the reduced dimension components for this dataset. For example, normally in scATAC-seq the first reduced dimensionality component is omitted, as it is strongly correlated with sequencing depth. In our scATAC-seq data we see there is a significant depth correlation in the first four reduced dimension components, especially component three, which is almost as correlatedas the first dimension (Figure 5.6). If we omit these from visualization, we see the data becomes more homogenous, clustering less significantly by sample type, and potentially giving us the ability to derive more biological differences from the variance between samples. This will undoubtedly be important in future analysis and can also be explored in a more in depth and systematic way.

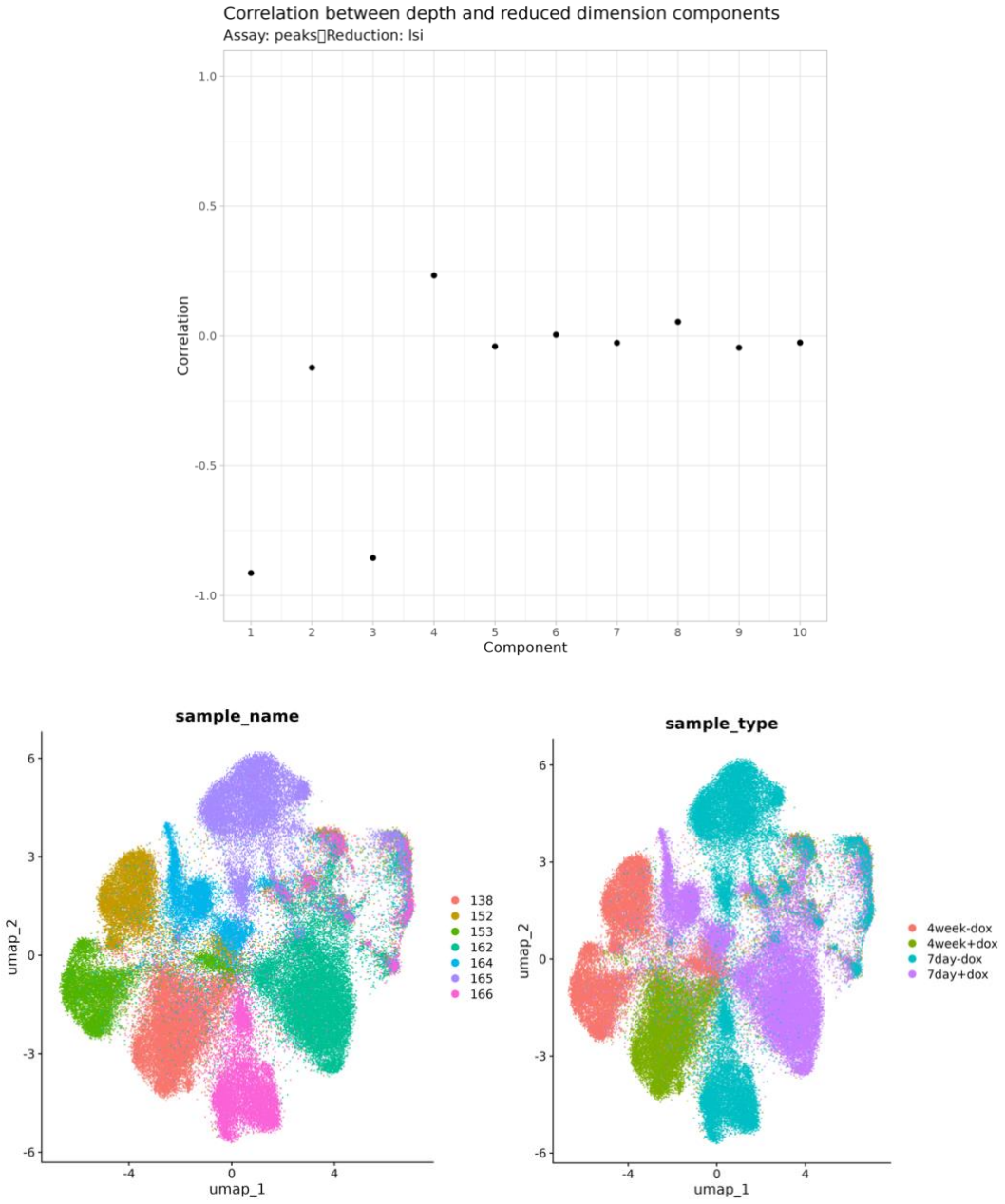


Figure 5.6. (top) Correlation between sequencing depth and reduced dimension components. (bottom left) UMAP of ATAC-seq data grouped by sample. (bottom right) UMAP of ATAC-seq grouped by condition.

For the early (7 day) and late (4 week) reprogramming timepoints we also investigated the differences in peaks (i.e. open chromatin) between the samples with and without dox treatment. By intersecting these differing areas of chromatin accessibility with a motif database (JASPAR)

we were able to determine some of the factors responsible for epigenetic reprogramming at these different timepoints (Figure 5.7). From this data we observe Signal Transducers and Activators of Transcription (STAT) family transcription factors are largely responsible for early reprogramming, whereas Krüppel-Like Factors (KLF) support late-stage reprogramming.

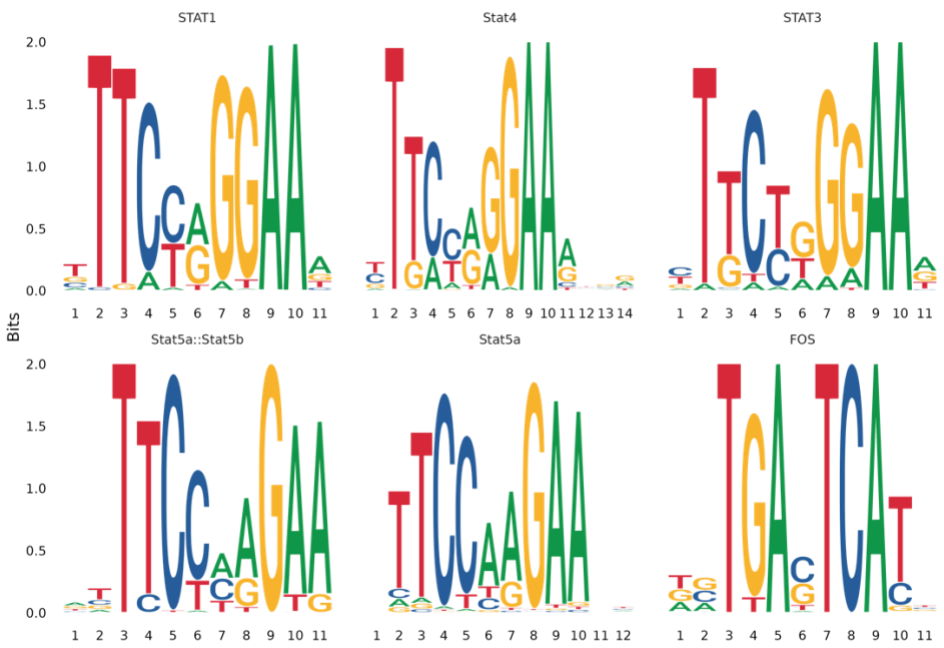
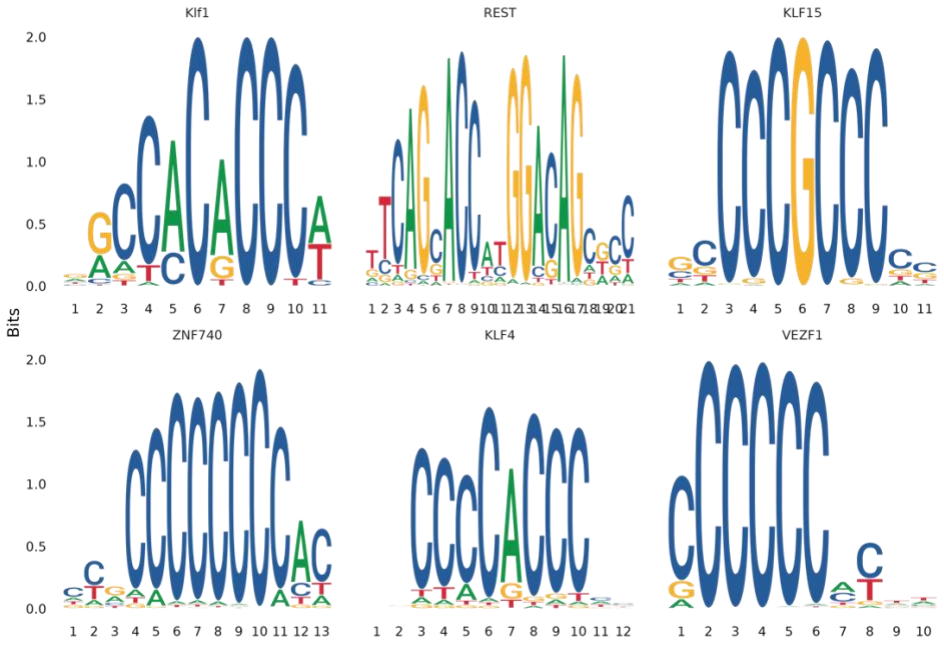
early repro	
late repro	

Figure 5.7. Top n = 6 differentially enriched motifs in accessible chromatin at the early (top) and late (bottom) reprogramming time points.

5.5: Discussion

In this chapter we used a multiomic approach to determine the role senescent cells play in epigenetic reprogramming. Although we were able to establish distinct motifs from the early and late stages of reprogramming from the ATAC data, indicating the key regulators of this process at different stages, there were significant complications in completing this analysis due to data quality. In light of the poor scRNA-Seq caliber and time restraints, as this data was received only two months before the completion of this thesis, this discussion will focus on alternative approaches and suggestions for (1) salvaging the existing data and (2) further analysis.

The gene expression data from the samples presented in this chapter prevented us from performing critical pieces of this analysis, such as accurate cell typing. Furthermore, as key transcripts such as Cdkn2a (p16) are already sparse in single-cell data, our poor gene expression data quality is likely to blame for their lack of detection in our data, which prevents us from detecting senescent cells accurately. In contrast, the scATAC-seq data from this multiomic analysis was of much higher quality. While traditional filtering using both modalities retained ~30% of cells, we were able to retain almost all the cells sequenced by only filtering using the ATAC-seq data.

There are several options to determine transcriptomic signatures from chromatin accessibility data, which could be done in future analyses, allowing for the completion of cell typing, among other aspects. Firstly, a gene activity matrix can be created from the peaks called by the scATAC-seq. With this gene activity matrix one can then examine the expression of cell type marker genes in different clusters and perform manual cell typing. We can also incorporate

scRNA-seq data from an assay completed on the same tissue or cell type (for example the Tabula Muris Senis (Tabula Muris Consortium, 2020), in our case mouse liver. Seurat's functions to find transfer anchors and integrate the datasets can then be used to join the reference scRNA-seq dataset with our experimental scATAC-seq data.

With the new, integrated multiomic analysis we would have many options available at our disposal. The cells can now be typed, and we can determine senescent cells based off of markers (p16, p21, SASP genes, lamin b1, etc.) or the recently published signatures of senescence such as SenSig (Cherry et al., 2023) or SenMayo (Saul et al., 2022). Once identified, we can determine the effect of reprogramming on senescent cells, asking questions such as – does reprogramming reduce the number of senescent cells? Are these cells less inflammatory with reprogramming? Is this more notable in early or late reprogramming? The scATAC-seq will also allow us to identify key regulators of SASP genes and senescence pathways through motif analysis, and how these differ with reprogramming.

Aside from senescent cells, we can determine how reprogramming effects cell populations generally. For example, does reprogramming decrease the proportion of specific cell types? Additionally, what genes and pathways become altered with reprogramming? What factors regulates these pathways? The possibilities are endless and will be explored in future analysis by members of the Neretti lab.

The data analysis presented in this chapter is also enlightening for future multiome experiment planning. The first and most obvious issue is the scRNA-seq data quality. While we are still investigating the source of this problem, our leading hypothesis is that the perfusion of the mice with PBS damaged the RNA present in the liver. While troubleshooting this experiment we performed the 10x multiome on mouse livers flash frozen without perfusion by our lab, as well as

members of the Jain lab at Brown University. Generally, we saw that these yielded ATAC-seq samples lower in concentration, while the RNA-seq samples were comparatively higher, and there was more cDNA in the proper size range. Furthermore, while the non-perfused samples homogenized nicely, it was very difficult to breakdown the perfused livers provided to us by the Sebastiano lab. In this case it seems that the perfusion, or something that happened in the harvesting process, damaged the RNA and altered the tissue structure, while stabilizing, or even damaging and the chromatin. In the future we will compare perfused and non-perfused samples from the Sebastiano lab to explore this hypothesis.

An additional takeaway from this experiment is to target fewer than 10k cells. In our initial troubleshooting we established that the data yield from targeting 5k cells was insufficient, so we instead targeted the highest suggested range by 10x (10k). However, due to the inaccuracy of manual cell counting, and undoubtedly automatic cell counting, it is easy to underestimate the number of cells and incidentally overload the 10x machine, as in samples 138 and 162. For future experiments it would be ideal to target fewer cells, such as 7k-8k nuclei, to avoid this problem.

Although the analysis presented in this chapter is not as fruitful as we would have hoped at this stage, it has incredible potential and has taught us a great deal about how best to perform multiomic analysis. With additional future experiments and computational analysis this datatset, the first of its kind to allow in depth exploration of the epigenetic landscape of senescent cells in reprogramming, will propel the rejuvenation field forward.

Chapter 6: Conclusion

This thesis presents an ensemble of innovative research which sheds light on how changes within the nucleus functionally support senescence.

6.1: High-Resolution Hi-C has expanded our understanding of senescence-associated nuclear architecture

Most chapters of this body of work focus on alterations to nuclear architecture. In chapter 2 I present work indicating that with senescence we see a loss of heterochromatin, as noted by a shift towards the A compartment and A-type subcompartments. Functionally, this supports features of senescence such as cell cycle arrest and the damaging SASP. Furthermore, this work is the first to explore genome structure at the level of loops, enabled by the high-resolution of our Hi-C maps. We define a new method to call loops, which allows us to see the formation of many novel loops with senescence. These loops play functional roles as they support an erosion of cellular identity, cell cycle arrest and SASP gene expression. These loops' formation is facilitated by a specific loss of methylation at their anchors, and they mainly lack CTCF motifs and form within the B compartment and B-type subcompartments. The architectural changes that occur with senescence also support the derepression of repetitive elements, namely LINE1, and we establish the first LINE1 "hotspot" where changes to genome structure at every level of organization help drive retroelement expression.

6.2: The Chromosome Shape Alphabet method has the potential to systematically connect structure to function across conditions and cell lines

In chapter 3 we improve upon a published method (C. Soto, Bryner, et al., 2021), which uses elastic shape analysis and clustering to identify prototypical shapes in the 3D genome. By implementing changes, such as the recursive Fréchet mean, and secondary clustering of the Chromosome Shape Alphabets created at the level of chromosomes, we are now able to process

entire genomes and therefore can apply this method to explore more biological problems. In this chapter we also discuss the complications of selecting the modeling method and segmentation strategy. These are open questions, which do not yet have definitive answers. Regarding modeling, we instead suggest the use of imaging, as it presents a more natural way to observe cell-to-cell heterogeneity. Although full genome imaging at the scale needed to complete this analysis does not yet exist, the leaps and bounds being made in the field of spatial omics will likely rise to this challenge in the upcoming years. Additionally, regarding segmentation, we have completed all analysis at the level of TADs. Although in chapter 3 we explain our methods for selecting TopDom as our TAD caller of choice, in chapter 4 we present and benchmark a novel TAD caller TADBay, which would likely be more advantageous for this type of analysis. We show that TADBay performs comparatively to other state-of-the-art callers as measured by measure of concordance analysis and proximity to CTCF. Furthermore, it can be tuned to generate hierarchical TAD calls, giving flexibility. However, the choice of TADs as a segmentation is imperfect within itself – many recent studies have challenged the concept of TADs at the single cell level and instead imply that they are an effect of averaging at the population level (Bintu et al., 2018; Mateo et al., 2019). As one of the goals of the Chromosome Shape Alphabet is to look at single-cell heterogeneity, a universal segmentation could complicate this. Regardless of these complications, we believe that this Chromosome Shape Alphabet method has utility, such as in the preliminary analysis we present highlighting the structural differences between proliferating and senescent cells.

6.3: Epigenetic Reprogramming and complications of multi-modal assays

Chapter 5 contains the results from an early investigation of the landscape of epigenetic reprogramming as assayed with the 10x multiome kit. In this chapter we explore the data quality, finding that unfortunately the scRNA-seq data created in this assay was not of sufficient condition

for our analysis. This result was unexpected, and we theorize that this could be due to the perfusion performed on the mice, as this was the only experimental difference between the livers we troubleshooted and the livers we received from our collaborators. Anecdotally, we have heard that the data quality from these multiome kits is significantly worse than with each modality alone. While the RNA quality was poor, there is hope that we can use the scATAC-seq associated with this assay to infer RNA and move on to elucidate the role of senescent cells in reprogramming from there. From this project we also learned about optimizing some technical aspects of the 10x multiome assay, such as to target 7000 to 8000 cells, rather than the upper limit of 10000.

6.4: Technical challenges of working with Hi-C data

Chromosome conformation capture techniques, specifically Hi-C, are the workhorse of this thesis. However informative these data are, there are also significant limitations to working with them. First, with Hi-C we sacrifice depth for breadth. While we are able to survey the spatial organization of the entire genome, which is incredibly powerful, we do not get sufficient resolution to be able to look at more granular details. In chapter 2, for example, we are unable to determine specific transcription factors regulating loops, or what promoters they control, from the Hi-C data alone. In chapter 3, we see there are limitations to working with Hi-C data as well. For example, Hi-C conveys the structure of the genome indirectly, by showing spatial proximity genome-wide. While we can model chromosome structures and genome structures from this data, it is highly dependent on which modeling method is used as well as other factors such as Hi-C data resolution. This is true of cell-to-cell heterogeneity as well; we can model this heterogeneity using methods such as deconvolution (Hua et al., 2018), but it is difficult to evaluate their accuracy due to a lack of ground truth structures. While some of these aspects can be improved by using newer methods

such as micro-C (Krietenstein et al., 2020) and capture Hi-C (Schoenfelder et al., 2018), a complementary and sometimes advantageous approach is to use imaging.

6.5: Spatial multiomics: a new player in genomic architecture

Advances from the sequencing-based side of genomic architecture research have been equally matched, if not more so, by novel imaging techniques. These techniques include volumetric approaches, such as the super-resolution OligoSTORM (Beliveau et al., 2017), which uses renewable, customizable oligopaint probes to investigate genomic architecture. Methods such as OligoSTORM reveal informative characteristics such as sphericity, volume, and radius of gyration. We can also observe aspects such as high-resolution colocalization with this technique. While informative, OligoSTORM has its limitations. For example, it can be extremely time consuming and expensive. For this reason, it is currently only possible to survey portions of the genome with this method; the largest OligoSTORM experiment to date surveyed almost the entirety of Chromosome 9 (Horrell, 2022).

Another type of imaging which can significantly inform our knowledge of genomic architecture is 'ball-and-stick' imaging, which collects points of data along DNA structures instead of giving volumetric information. These points can be visualized as dots connected by lines, thus giving them the alias ball-and-stick. These types of imaging include methods such as Optical Reconstruction of Chromatin Architecture (ORCA) (Mateo et al., 2019) and Multiplexed Imaging of Nucleome Architectures (MINA) (M. Liu et al., 2021), as well as DNA-Multiplexed Error-Robust Fluorescence In Situ Hybridization (DNA-MERFISH) (Su et al., 2020). Although it cannot capture the same resolution as super-resolution methods such as OligoSTORM, ball-and-stick methods have several advantages. For example, they can capture much larger scales of imaging. The Zhuang lab has demonstrated that DNA-MERFISH can investigate the structure of the entire

genome at low resolution (Su et al., 2020). Furthermore, these imaging techniques can be more easily combined with other modalities such as transcriptomic and proteomic information. For these reasons these ball-and-stick-type methods are likely more advantageous for our analysis.

The most powerful and accurate discoveries likely lie at the apex of chromosome conformation capture and imaging techniques. In the context of this thesis, to advance our knowledge of the functional role of senescence-associated conformational changes, future analyses should include results from both sequencing- and microscopy-based interrogations.

6.6: Mechanistic Interventions

This thesis contains many exciting ideas about how the genome's conformation is altered with senescence. For example, in chapter 2 we observe that many of the loops formed with senescence are associated with specific hypomethylation. However, what would be more critical is to test and directly support this link. Due to the advent of CRISPR/Cas9 (Doudna & Charpentier, 2014) we have many options for doing so. Using the aforementioned example, we could use technologies such as CRISPR-driven methylation (Katayama, Shiraishi, Gorai, & Andou, 2021; Vojta et al., 2016) to explore this correlation. For example, we could direct methylation of loops which have lost methylation with senescence and determine if this prevents their formation, and how this effects senescence (Figure 6.1). To assay this robustly, we could trace the loops using imaging techniques to determine loop formation at higher resolution. The Chromosome Shape Alphabet method, or an adaptation using the distance metric, could be used for this analysis. We could cluster the shapes to determine if the structure is informative of the transcriptional status, thus informing the structure-function relationship.

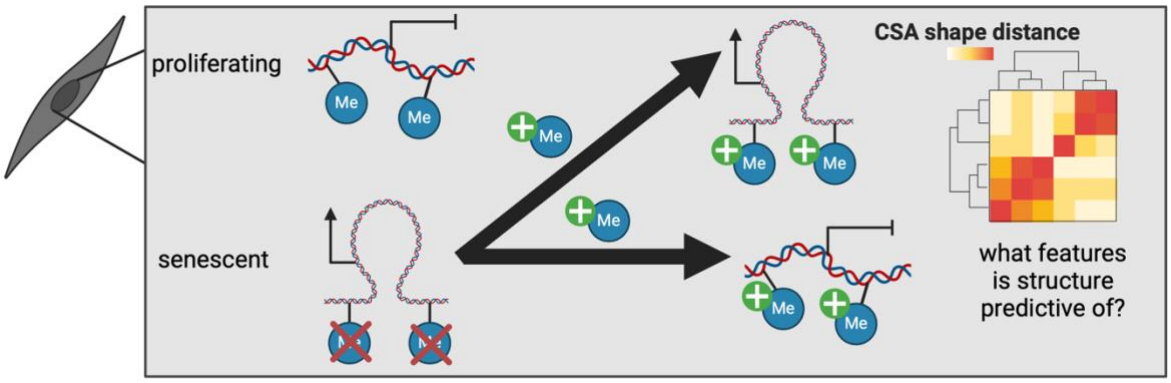


Figure 6.1. Methodological overview of proposed mechanistic interventions (effect of targeted methylation on loops).

Another intervention of interest to more concretely link structure to function would be senomorphics. This class of drugs which modulates the damaging inflammatory phenotype can more specifically target genes, and unlike senolytics, leaves cells alive so we can study them. An exciting future direction would be to use a combination of sequencing- and microscopy-based techniques to investigate the effect of senomorphics on some of the 67% of loops that form around SASP genes. As this is a large percentage of the SASP genes, and as senomorphics modulate the SASP, there are undoubtedly senomorphics which would target regions where specific SASP genes are modulated by loops. We could trace and interrogate these loops, again using a combination of imaging and high-resolution chromatin conformation capture techniques. If we see that there is a structural change to these with the senomorphic it would indicate a stronger structure-function link.

6.7: Complications of Working with Senescence

As mentioned in chapter 2, there are many complications to working with senescent cells. First and foremost, senescent cells can be difficult to culture. This varies based on the model used (DDIS, OIS, RS, etc.), but over time senescent cells tend to die off, and especially in the context of replicative senescence and this can make garnering sufficient cell numbers for experiments

difficult. For sequencing-based assays such as Hi-C which require large sums of cells this can make experiments technically difficult and large numbers of cells need to be cultured and maintained in the same plates for long periods, increasing the chance of contamination. In this case the imaging-based assays may be technically easier to perform, but there are still significant challenges. For example, many imaging applications require high cell numbers for statistical power, and in recent years, machine learning algorithms. It is difficult to achieve high confluence with senescent cells, even when re-seeded in high numbers onto coverslips.

Senescent cells are also quite fragile. When culturing them we have noted that they often do not respond well to processes such as scraping, seeding, and plate transfers. We have also noticed in single-cell assays that the data gleaned from senescent cells is of comparatively poor quality and requires careful, customized filtering (Evans et al., 2023). This is likely because their fragility is incompatible with the microfluidics contained within the chromium devices. In this case, we must take our single-cell observations with a grain of salt. Although the scRNA-seq data quality presented in chapter 5 was poor, it is also possible we did not see p16 transcripts because the senescent cells did not survive transport throughout the chromium's microfluidics system. A possible solution in future analyses would be to use split-pooling based method such as SPLiT-seq (Rosenberg et al., 2018), which may circumvent the problems with the fragility of senescent cells.

A last, broad complication of working with senescent cells is that we generally lack a perception of their role *in vivo*. Some key studies, such as those using Darren Baker's INK-ATTAC mouse, have indicated that clearing p16-positive cells extends lifespan (Baker et al., 2016). However, because the biomarkers of senescence are unclear, it is somewhat difficult to interpret what this finding means for senescent cells in general, as opposed to p16-positive cells. Groups such as the NIH's SenNet Consortium (SenNet Consortium, 2022) are a step in the right direction

to solve this problem. This consortium is using the most cutting-edge techniques and technologies, from the sequencing and microscopy sides, to elucidate how senescent cells contribute to aging. The SenNet group will create atlases of aging animals and in this way, allow an integrated, multiomic approach to more comprehensively determine the role of senescent cells, propelling the field forward.

6.8: Bringing it all together: a unified future direction

At the intersection of the areas of this thesis (genomic architecture, senescence, shape analysis tools, epigenetic reprogramming) lies a natural, impactful continuation and future direction for this work. What is the role of genomic architecture in epigenetic reprogramming? While movement has been made to determine how epigenetic reprogramming rejuvenates organisms, little, if anything, is known about how structure supports this anti-aging process. Are there key architectural changes that are driving the transcriptomic changes seen? And in particular, what happens to senescent cells with this treatment and how is their architecture altered? As recommended in earlier sections of this chapter, the most advantageous choice would be multifaceted and multimodal (Figure 6.2). First, I would perform a broad scale assay to assess the general architectural landscape of reprogramming. In this context I would use micro-C (Krietenstein et al., 2020), as it provides higher resolution data than Hi-C. Combining this with bulk RNA-seq, I would determine regions of interest to epigenetic reprogramming, where both the architecture and gene expression was altered significantly. Next, I would take a multiomic and spatial imaging approach. I would use a ball-and-stick chromatin tracing method and combine this with imaging to assess the transcriptome and potentially proteins of interest. These areas can even be perturbed, through the use of CLOuD9 (Morgan et al., 2017), or other CRISPR-based techniques to determine whether structure is a key regulator of these rejuvenating changes.

Additionally, we can use the distance metric from the Chromosome Shape Alphabet method to cluster the traced structures and determine what biological features they are predictive of (condition, transcriptional activity, etc.). Together these assays will provide a clear picture of how epigenetic reprogramming is supported, and potentially driven, by changes to the genome's structure.

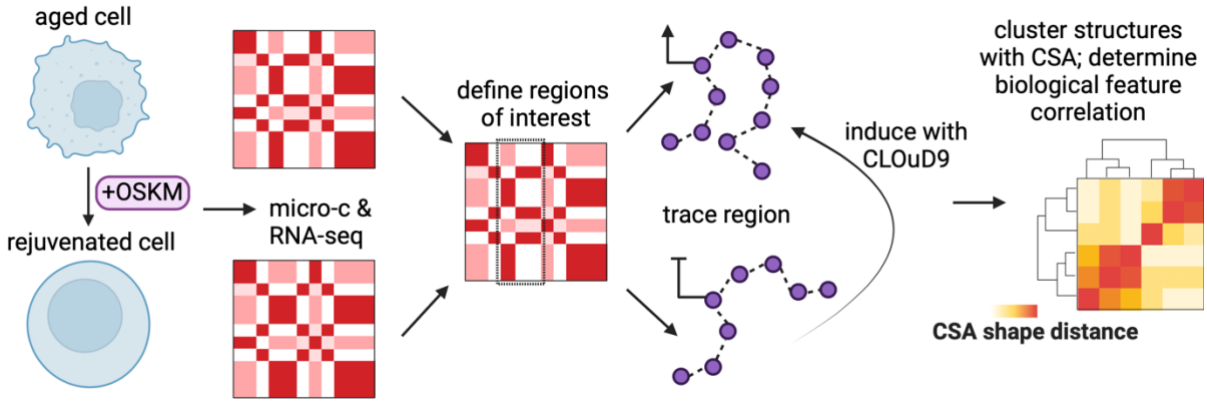


Figure 6.2. Methodological overview of unified future direction (investigating architectural features supporting reprogramming).

6.9: On Aging

The year 2024 is a fascinating time to be working in the field of aging. We are truly in a revolutionary time in the field and are making strides towards living longer, healthier lives. Senescence research is proving crucial in these advancements, and there is great hope for senescence interventions' impact (i.e. senolytics, senomorphics, etc.). This thesis represents a piece of the body of discoveries that brings us closer to modulating aging by outlining key observations about the architectural regulation of senescence and methodological innovations to analyze these conformational changes. It also presents an exciting new experimental design and dataset to examine the role of senescence in epigenetic reprogramming. Altogether these novel

ideas shape a critical niche in the aging field and bring us closer *towards understanding functional nuclear alterations with cellular senescence*.

References

Acemel, R. D., Maeso, I., & Gómez-Skarmeta, J. L. (2017). Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. *Wiley Interdisciplinary Reviews. Developmental Biology*, *6*(3), e265. doi:10.1002/wdev.265

Acosta, J. C., Banito, A., Wuestefeld, T., Georgilis, A., Janich, P., Morton, J. P., … Gil, J. (2013). A complex secretory program orchestrated by the inflammasome controls paracrine senescence. *Nature Cell Biology*, *15*(8), 978–990. doi:10.1038/ncb2784

Aird, K. M., Iwasaki, O., Kossenkov, A. V., Tanizawa, H., Fatkhutdinov, N., Bitler, B. G., … Zhang, R. (2016). HMGB2 orchestrates the chromatin landscape of senescence-associated secretory phenotype gene loci. *The Journal of Cell Biology*, *215*(3), 325–334. doi:10.1083/jcb.201608026

Anders, S., Pyl, P. T., & Huber, W. (2014). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* , *31*(2), 166–169. doi:10.1093/bioinformatics/btu638

Baell, J. B., Leaver, D. J., Hermans, S. J., Kelly, G. L., Brennan, M. S., Downer, N. L., … Thomas, T. (2018). Inhibitors of histone acetyltransferases KAT6A/B induce senescence and arrest tumour growth. *Nature*, *560*(7717), 253–257. doi:10.1038/s41586-018-0387-5

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., … Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, *37*(Web Server issue), W202-8. doi:10.1093/nar/gkp335

Baker, D. J., Childs, B. G., Durik, M., Wijers, M. E., Sieben, C. J., Zhong, J., … van Deursen, J. M. (2016). Naturally occurring p16(Ink4a)-positive cells shorten healthy lifespan. *Nature*, *530*(7589), 184–189. doi:10.1038/nature16932

Baker, D. J., Wijshake, T., Tchkonia, T., LeBrasseur, N. K., Childs, B. G., van de Sluis, B., … van Deursen, J. M. (2011). Clearance of p16Ink4a-positive senescent cells delays ageing-associated disorders. *Nature, 479*(7372), 232–236. doi:10.1038/nature10600

Basisty, N., Kale, A., Jeon, O. H., Kuehnemann, C., Payne, T., Rao, C., … Schilling, B. (2020). A proteomic atlas of senescence-associated secretomes for aging biomarker development. *PLoS Biology, 18*(1), e3000599. doi:10.1371/journal.pbio.3000599

Bassing, C. H., & Alt, F. W. (2004). H2AX may function as an anchor to hold broken chromosomal DNA ends in close proximity. *Cell Cycle (Georgetown, Tex.), 3*(2), 149–153. doi:10.4161/cc.3.2.689

Belhadj, J., Surina, S., Hengstschläger, M., & Lomakin, A. J. (2023). Form follows function: Nuclear morphology as a quantifiable predictor of cellular senescence. *Aging Cell*, e14012. doi:10.1111/acel.14012

Beliveau, B. J., Boettiger, A. N., Nir, G., Bintu, B., Yin, P., Zhuang, X., & Wu, C.-T. (2017). In Situ Super-Resolution Imaging of Genomic DNA with OligoSTORM and OligoDNA-PAINT. *Methods in Molecular Biology , 1663*, 231–252. doi:10.1007/978-1-4939-7265-4_19

Bendall, M. L., de Mulder, M., Iñiguez, L. P., Lecanda-Sánchez, A., Pérez-Losada, M., Ostrowski, M. A., … Nixon, D. F. (2019). Telescope: Characterization of the retrotranscriptome by accurate estimation of transposable element expression. *PLoS Computational Biology*, *15*(9), e1006453. doi:10.1371/journal.pcbi.1006453

Bintu, B., Mateo, L. J., Su, J.-H., Sinnott-Armstrong, N. A., Parker, M., Kinrot, S., … Zhuang, X. (2018). Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science*, *362*(6413). doi:10.1126/science.aau1783

Boninsegna, L., Yildirim, A., Polles, G., Zhan, Y., Quinodoz, S. A., Finn, E. H., … Alber, F. (2022). Integrative genome modeling platform reveals essentiality of rare contact events in 3D genome organizations. *Nature Methods*, *19*(8), 938–949. doi:10.1038/s41592-022-01527-x

Boumendil, C., Hari, P., Olsen, K. C. F., Acosta, J. C., & Bickmore, W. A. (2019). Nuclear pore density controls heterochromatin reorganization during senescence. *Genes & Development*, *33*(3–4), 144–149. doi:10.1101/gad.321117.118

Burma, S., Chen, B. P., Murphy, M., Kurimasa, A., & Chen, D. J. (2001). ATM phosphorylates histone H2AX in response to DNA double-strand breaks. *The Journal of Biological Chemistry*, *276*(45), 42462–42467. doi:10.1074/jbc.C100466200

Bussian, T. J., Aziz, A., Meyer, C. F., Swenson, B. L., van Deursen, J. M., & Baker, D. J. (2018). Clearance of senescent glial cells prevents tau-dependent pathology and cognitive decline. *Nature*, *562*(7728), 578–582. doi:10.1038/s41586-018-0543-y

Busslinger, G. A., Stocsits, R. R., van der Lelij, P., Axelsson, E., Tedeschi, A., Galjart, N., & Peters, J.-M. (2017). Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*, *544*(7651), 503–507. doi:10.1038/nature22063

Carlson, M. (2019). org.Hs.eg.db: Genome wide annotation for Human. Retrieved from Bioconductor website: https://www.bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html

Chakraborty, A., Wang, J. G., & Ay, F. (2022). dcHiC detects differential compartments across multiple Hi-C datasets. *Nature Communications*, *13*(1), 1–21. doi:10.1038/s41467-022-34626-6

Chandra, T., Ewels, P. A., Schoenfelder, S., Furlan-Magaril, M., Wingett, S. W., Kirschner, K., … Reik, W. (2015). Global reorganization of the nuclear landscape in senescent cells. *Cell Reports*, *10*(4), 471–483. doi:10.1016/j.celrep.2014.12.055

Chandra, T., Kirschner, K., Thuret, J.-Y., Pope, B. D., Ryba, T., Newman, S., … Narita, M. (2012). Independence of repressive histone marks and chromatin compaction during senescent heterochromatic layer formation. *Molecular Cell*, *47*(2), 203–214. doi:10.1016/j.molcel.2012.06.010

Chen, H., Ruiz, P. D., McKimpson, W. M., Novikov, L., Kitsis, R. N., & Gamble, M. J. (2015). MacroH2A1 and ATM Play Opposing Roles in Paracrine Senescence and the Senescence-Associated Secretory Phenotype. *Molecular Cell*, *59*(5), 719–731. doi:10.1016/j.molcel.2015.07.011

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics (Oxford, England)*, *34*(17), i884–i890. doi:10.1093/bioinformatics/bty560

Cherry, C., Andorko, J. I., Krishnan, K., Mejías, J. C., Nguyen, H. H., Stivers, K. B., … Elisseeff, J. H. (2023). Transfer learning in a biomaterial fibrosis model identifies in vivo senescence heterogeneity and contributions to vascularization and matrix production across species and diverse pathologies. *GeroScience*, *45*(4), 2559–2587. doi:10.1007/s11357-023-00785-7

Chiang, M., Michieletto, D., Brackley, C. A., Rattanavirotkul, N., Mohammed, H., Marenduzzo, D., & Chandra, T. (2019). Polymer Modeling Predicts Chromosome Reorganization in Senescence. *Cell Reports*, *28*(12), 3212-3223.e6. doi:10.1016/j.celrep.2019.08.045

Chondronasiou, D., Gill, D., Mosteiro, L., Urdinguio, R. G., Berenguer-Llergo, A., Aguilera, M., … Serrano, M. (2022). Multi-omic rejuvenation of naturally aged tissues by a single cycle of transient reprogramming. *Aging Cell*, *21*(3), e13578. doi:10.1111/acel.13578

Contrepois, K., Coudereau, C., Benayoun, B. A., Schuler, N., Roux, P.-F., Bischof, O., … Mann, C. (2017). Histone variant H2A.J accumulates in senescent cells and promotes inflammatory gene expression. *Nature Communications*, *8*, 14995. doi:10.1038/ncomms14995

Coppé, J.-P., Desprez, P.-Y., Krtolica, A., & Campisi, J. (2010). The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annual Review of Pathology*, *5*, 99–118. doi:10.1146/annurev-pathol-121808-102144

Criscione, S. W., De Cecco, M., Siranosian, B., Zhang, Y., Kreiling, J. A., Sedivy, J. M., & Neretti, N. (2016). Reorganization of chromosome architecture in replicative cellular senescence. *Science Advances*, *2*(2), e1500882. doi:10.1126/sciadv.1500882

Cruickshanks, H. A., McBryan, T., Nelson, D. M., Vanderkraats, N. D., Shah, P. P., van Tuyn, J., … Adams, P. D. (2013). Senescent cells harbour features of the cancer epigenome. *Nature Cell Biology*, *15*(12), 1495–1506. doi:10.1038/ncb2879

De Cecco, M., Criscione, S. W., Peckham, E. J., Hillenmeyer, S., Hamm, E. A., Manivannan, J., … Sedivy, J. M. (2013). Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell*, *12*(2), 247–256. doi:10.1111/acel.12047

De Cecco, M., Criscione, S. W., Peterson, A. L., Neretti, N., Sedivy, J. M., & Kreiling, J. A. (2013). Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. *Aging*, *5*(12), 867–883. doi:10.18632/aging.100621

De Cecco, M., Ito, T., Petrashen, A. P., Elias, A. E., Skvir, N. J., Criscione, S. W., … Sedivy, J. M. (2019). L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature*, *566*(7742), 73–78. doi:10.1038/s41586-018-0784-9

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Capturing chromosome conformation. *Science (New York, N.Y.)*, *295*(5558), 1306–1311. doi:10.1126/science.1067799

Demaria, M., Ohtani, N., Youssef, S. A., Rodier, F., Toussaint, W., Mitchell, J. R., … Campisi, J. (2014). An essential role for senescent cells in optimal wound healing through secretion of PDGF-AA. *Developmental Cell*, *31*(6), 722–733. doi:10.1016/j.devcel.2014.11.012

Demaria, M., O'Leary, M. N., Chang, J., Shao, L., Liu, S., Alimirah, F., … Campisi, J. (2017). Cellular senescence promotes adverse effects of chemotherapy and cancer relapse. *Cancer Discovery*, *7*(2), 165–176. doi:10.1158/2159-8290.CD-16-0241

Di Giammartino, D. C., Kloetgen, A., Polyzos, A., Liu, Y., Kim, D., Murphy, D., … Apostolou, E. (2019). KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. *Nature Cell Biology*, *21*(10), 1179–1190. doi:10.1038/s41556-019-0390-6

Di Micco, R., Krizhanovsky, V., Baker, D., & d'Adda di Fagagna, F. (2021). Cellular senescence in ageing: from mechanisms to therapeutic opportunities. *Nature Reviews. Molecular Cell Biology*, *22*(2), 75–95. doi:10.1038/s41580-020-00314-w

Dimri, G. P., Lee, X., Basile, G., Acosta, M., Scott, G., Roskelley, C., … Pereira-Smith, O. (1995). A biomarker that identifies senescent human cells in culture and in aging skin in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(20), 9363–9367. doi:10.1073/pnas.92.20.9363

Dixon, J. R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J. E., Lee, A. Y., … Ren, B. (2015). Chromatin architecture reorganization during stem cell differentiation. *Nature*, *518*(7539), 331–336. doi:10.1038/nature14222

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., … Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–380. doi:10.1038/nature11082

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* , *29*(1), 15–21. doi:10.1093/bioinformatics/bts635

Doudna, J. A., & Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)*, *346*(6213), 1258096. doi:10.1126/science.1258096

Duarte, L. F., Young, A. R. J., Wang, Z., Wu, H.-A., Panda, T., Kou, Y., … Bernstein, E. (2014). Histone H3.3 and its proteolytically processed form drive a cellular senescence programme. *Nature Communications*, *5*, 5210. doi:10.1038/ncomms6210

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems*, *3*(1), 95–98. doi:10.1016/j.cels.2016.07.002

Evans, S. A., Teo, Y. V., Hinthorn, S. J., Clark, K., Takahiro, I., Sedivy, J. M., & Neretti, N. (2023). Single-Cell Transcriptomics Reveals Global Markers of Transcriptional Diversity across Different Forms of Cellular Senescence. *AgingBio*.

Ewing, A. D., Smits, N., Sanchez-Luque, F. J., Faivre, J., Brennan, P. M., Richardson, S. R., … Faulkner, G. J. (2020). Nanopore Sequencing Enables Comprehensive Transposable

Element Epigenomic Profiling. *Molecular Cell*, *80*(5), 915-928.e5. doi:10.1016/j.molcel.2020.10.024

Funayama, R., Saito, M., Tanobe, H., & Ishikawa, F. (2006). Loss of linker histone H1 in cellular senescence. *The Journal of Cell Biology*, *175*(6), 869–880. doi:10.1083/jcb.200604005

Gilbert, S. F. (2000). *Aging: The Biology of Senescence*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK10041/

Gill, D., Parry, A., Santos, F., Okkenhaug, H., Todd, C. D., Hernando-Herraez, I., … Reik, W. (2022). Multi-omic rejuvenation of human cells by maturation phase transient reprogramming. *ELife*, *11*. doi:10.7554/eLife.71624

González-Bermúdez, L., Anglada, T., Genescà, A., Martín, M., & Terradas, M. (2019). Identification of reference genes for RT-qPCR data normalisation in aging studies. *Scientific Reports*, *9*(1), 13970. doi:10.1038/s41598-019-50035-0

Gorgoulis, V., Adams, P. D., Alimonti, A., Bennett, D. C., Bischof, O., Bishop, C., … Demaria, M. (2019). Cellular Senescence: Defining a Path Forward. *Cell*, *179*(4), 813–827. doi:10.1016/j.cell.2019.10.005

Gray, M. D., Jesch, S. A., & Stein, G. H. (1991). 5-Azacytidine-induced demethylation of DNA to senescent level does not block proliferation of human fibroblasts. *Journal of Cellular Physiology*, *149*(3), 477–484. doi:10.1002/jcp.1041490317

Hagberg, A. A., Schult, D. A., & Swart, P. J. (n.d.). Aug. 2008. Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference*.

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, *100*(1), 57–70. doi:10.1016/s0092-8674(00)81683-9

Hao, Y., Stuart, T., Kowalski, M. H., Choudhary, S., Hoffman, P., Hartman, A., … Satija, R. (2023). Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology*. doi:10.1038/s41587-023-01767-y

Hayakawa, T., Iwai, M., Aoki, S., Takimoto, K., Maruyama, M., Maruyama, W., & Motoyama, N. (2015). SIRT1 suppresses the senescence-associated secretory phenotype through epigenetic gene regulation. *PloS One*, *10*(1), e0116480. doi:10.1371/journal.pone.0116480

Hayflick, L., & Moorhead, P. S. (1961). The serial cultivation of human diploid cell strains. *Experimental Cell Research*, *25*(3), 585–621. doi:10.1016/0014-4827(61)90192-6

Heckenbach, I., Mkrtchyan, G. V., Ezra, M. B., Bakula, D., Madsen, J. S., Nielsen, M. H., … Scheibye-Knudsen, M. (2022). Nuclear morphology is a deep learning biomarker of cellular senescence. *Nature Aging*, *2*(8), 742–755. doi:10.1038/s43587-022-00263-3

Herranz, N., Gallage, S., Mellone, M., Wuestefeld, T., Klotz, S., Hanley, C. J., … Gil, J. (2015). mTOR regulates MAPKAPK2 translation to control the senescence-associated secretory phenotype. *Nature Cell Biology*, *17*(9), 1205–1217. doi:10.1038/ncb3225

Hoare, M., Ito, Y., Kang, T.-W., Weekes, M. P., Matheson, N. J., Patten, D. A., … Narita, M. (2016). NOTCH1 mediates a switch between two distinct secretomes during senescence. *Nature Cell Biology*, *18*(9), 979–992. doi:10.1038/ncb3397

Horrell, J. C. (2022). *Observing the natural and heteromorphic organization of the human genome in nanoscale resolution*. Brown University.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*(10), R115. doi:10.1186/gb-2013-14-10-r115

Hua, N., Tjong, H., Shin, H., Gong, K., Zhou, X. J., & Alber, F. (2018). Producing genome structure populations with the dynamic and automated PGS software. *Nature Protocols*, *13*(5), 915–926. doi:10.1038/nprot.2018.008

Ito, T., Teo, Y. V., Evans, S. A., Neretti, N., & Sedivy, J. M. (2018). Regulation of cellular senescence by polycomb chromatin modifiers through distinct DNA damage- and histone methylation-dependent pathways. *Cell Reports*, *22*(13), 3480–3492. doi:10.1016/j.celrep.2018.03.002

Iwasaki, O., Tanizawa, H., Kim, K.-D., Kossenkov, A., Nacarelli, T., Tashiro, S., … Noma, K.-I. (2019). Involvement of condensin in cellular senescence through gene regulation and compartmental reorganization. *Nature Communications*, *10*(1), 5688. doi:10.1038/s41467-019-13604-5

Johnstone, S. E., Reyes, A., Qi, Y., Adriaens, C., Hegazi, E., Pelka, K., … Bernstein, B. E. (2020). Large-scale topological changes restrain malignant progression in colorectal cancer. *Cell*, *182*(6), 1474-1489.e23. doi:10.1016/j.cell.2020.07.030

Kaeberlein, M., McVey, M., & Guarente, L. (2001). Using yeast to discover the fountain of youth. *Science of Aging Knowledge Environment: SAGE KE*, *2001*(1), e1. doi:10.1126/sageke.2001.1.pe1

Katayama, S., Shiraishi, K., Gorai, N., & Andou, M. (2021). A CRISPR/Cas9-based method for targeted DNA methylation enables cancer initiation in B lymphocytes. *Advanced Genetics (Hoboken, N.J.)*, *2*(1), e10040. doi:10.1002/ggn2.10040

Kaur, P., Otgonbaatar, A., Ramamoorthy, A., Chua, E. H. Z., Harmston, N., Gruber, J., & Tolwinski, N. S. (2022). Combining stem cell rejuvenation and senescence targeting to synergistically extend lifespan. *Aging*, *14*(20), 8270–8291. doi:10.18632/aging.204347

Kirkland, J. L., & Tchkonia, T. (2020). Senolytic drugs: from discovery to translation. *Journal of Internal Medicine*, *288*(5), 518–536. doi:10.1111/joim.13141

Kirkwood, T. B. L. (1977). The evolution of aging. *Nature*, (270), 301–304.

Kirschner, K., Rattanavirotkul, N., Quince, M. F., & Chandra, T. (2020). Functional heterogeneity in senescence. *Biochemical Society Transactions*, *48*(3), 765–773. doi:10.1042/BST20190109

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M. N., & Sergushichev, A. (2016). Fast gene set enrichment analysis. doi:10.1101/060012

Krefting, J., Andrade-Navarro, M. A., & Ibn-Salem, J. (2018). Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC Biology*, *16*(1), 87. doi:10.1186/s12915-018-0556-x

Krietenstein, N., Abraham, S., Venev, S. V., Abdennur, N., Gibcus, J., Hsieh, T.-H. S., … Rando, O. J. (2020). Ultrastructural Details of Mammalian Chromosome Architecture. *Molecular Cell*, *78*(3), 554-565.e7. doi:10.1016/j.molcel.2020.03.003

Larsson, J. (2021). *eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses*. Retrieved from https://CRAN.R-project.org/package=eulerr

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., … Carey, V. J. (2013). Software for computing and annotating genomic ranges. *PLoS Computational Biology*, *9*(8), e1003118. doi:10.1371/journal.pcbi.1003118

Leontieva, O. V., & Blagosklonny, M. V. (2010). DNA damaging agents and p53 do not cause senescence in quiescent cells, while consecutive re-activation of mTOR is associated with conversion to senescence. *Aging*, *2*(12), 924–935. doi:10.18632/aging.100265

Lévy-Leduc, C., Delattre, M., Mary-Huard, T., & Robin, S. (2014). Two-dimensional segmentation for analyzing Hi-C data. *Bioinformatics (Oxford, England)*, *30*(17), i386-92. doi:10.1093/bioinformatics/btu443

Li, L., Cui, L., Lin, P., Liu, Z., Bao, S., Ma, X., … Hui, L. (2023). Kupffer-cell-derived IL-6 is repurposed for hepatocyte dedifferentiation via activating progenitor genes from injury-specific enhancers. *Cell Stem Cell*, *30*(3), 283-299.e9. doi:10.1016/j.stem.2023.01.009

Li, W., Wei, W., & Ding, S. (2016). TGF-β signaling in stem cell regulation. *Methods in Molecular Biology (Clifton, N.J.)*, *1344*, 137–145. doi:10.1007/978-1-4939-2966-5_8

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, *30*(7), 923–930. doi:10.1093/bioinformatics/btt656

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Systems*, *1*(6), 417–425. doi:10.1016/j.cels.2015.12.004

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., … Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, *326*(5950), 289–293. doi:10.1126/science.1181369

Lin, D., Sanders, J., & Noble, W. S. (2021). HiCRep.py: fast comparison of Hi-C contact matrices in Python. *Bioinformatics (Oxford, England)*, *37*(18), 2996–2997. doi:10.1093/bioinformatics/btab097

Liu, M., Yang, B., Hu, M., Radda, J. S. D., Chen, Y., Jin, S., … Wang, S. (2021). Chromatin tracing and multiplexed imaging of nucleome architectures (MINA) and RNAs in single mammalian cells and tissue. *Nature Protocols*, *16*(5), 2667–2697. doi:10.1038/s41596-021-00518-0

Liu, W., Srivastava, A., & Zhang, J. (2011). A mathematical framework for protein structure comparison. *PLoS Computational Biology*, *7*(2), e1001075. doi:10.1371/journal.pcbi.1001075

Liu, Y., Nanni, L., Sungalee, S., Zufferey, M., Tavernari, D., Mina, M., … Ciriello, G. (2021). Systematic inference and comparison of multi-scale chromatin sub-compartments connects spatial organization to cell phenotypes. *Nature Communications*, *12*(1), 1–11. doi:10.1038/s41467-021-22666-3

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell*, *153*(6), 1194–1217. doi:10.1016/j.cell.2013.05.039

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2023). Hallmarks of aging: An expanding universe. *Cell*, *186*(2), 243–278. doi:10.1016/j.cell.2022.11.001

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. doi:10.1186/s13059-014-0550-8

Lu, Y., Brommer, B., Tian, X., Krishnan, A., Meer, M., Wang, C., … Sinclair, D. A. (2020). Reprogramming to recover youthful epigenetic information and restore vision. *Nature, 588*(7836), 124–129. doi:10.1038/s41586-020-2975-4

Mahfoudhi, E., Lordier, L., Marty, C., Pan, J., Roy, A., Roy, L., … Plo, I. (2016). P53 activation inhibits all types of hematopoietic progenitors and all stages of megakaryopoiesis. *Oncotarget, 7*(22), 31980–31992. doi:10.18632/oncotarget.7881

Mahmoudi, S., Mancini, E., Xu, L., Moore, A., Jahanbani, F., Hebestreit, K., … Brunet, A. (2019). Heterogeneity in old fibroblasts is linked to variability in reprogramming and wound healing. *Nature, 574*(7779), 553–558. doi:10.1038/s41586-019-1658-5

Mallette, F. A., Gaumont-Leclerc, M.-F., & Ferbeyre, G. (2007). The DNA damage signaling pathway is a critical mediator of oncogene-induced senescence. *Genes & Development, 21*(1), 43–48. doi:10.1101/gad.1487307

Manukyan, M., & Singh, P. B. (2014). Epigenome rejuvenation: HP1β mobility as a measure of pluripotent and senescent chromatin ground states. *Scientific Reports, 4*(1), 4789. doi:10.1038/srep04789

Maruthur, N. M., Tseng, E., Hutfless, S., Wilson, L. M., Suarez-Cuervo, C., Berger, Z., … Bolen, S. (2016). Diabetes medications as monotherapy or metformin-based combination therapy for type 2 diabetes. *Annals of Internal Medicine, 164*(11), 740. doi:10.7326/m15-2650

Mateo, L. J., Murphy, S. E., Hafner, A., Cinquini, I. S., Walker, C. A., & Boettiger, A. N. (2019). Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature, 568*(7750), 49–54. doi:10.1038/s41586-019-1035-4

Mateo, L. J., Sinnott-Armstrong, N., & Boettiger, A. N. (2021). Tracing DNA paths and RNA profiles in cultured cells and tissues with ORCA. *Nature Protocols*, *16*(3), 1647–1713. doi:10.1038/s41596-020-00478-x

Mattson, M. P. (2008). Hormesis defined. *Ageing Research Reviews*, *7*(1), 1–7. doi:10.1016/j.arr.2007.08.007

McCay, C. M., Crowell, M. F., & Maynard, L. A. (1935). The effect of retarded growth upon the length of life span and upon the ultimate body size. *The Journal of Nutrition*, *10*(1), 63–79. doi:10.1093/jn/10.1.63

McCord, R. P., Nazario-Toole, A., Zhang, H., Chines, P. S., Zhan, Y., Erdos, M. R., … Cao, K. (2013). Correlated alterations in genome organization, histone methylation, and DNA-lamin A/C interactions in Hutchinson-Gilford progeria syndrome. *Genome Research*, *23*(2), 260–269. doi:10.1101/gr.138032.112

Medawar, P. B. (1952). An Unsolved Problem of Biology. *H. K. Lewis, London*.

Miura, H., Poonperm, R., Takahashi, S., & Hiratani, I. (2018). Practical analysis of Hi-C data: Generating A/B compartment profiles. *Methods in Molecular Biology (Clifton, N.J.)*, *1861*, 221–245. doi:10.1007/978-1-4939-8766-5_16

Moiseeva, O., Deschênes-Simard, X., St-Germain, E., Igelmann, S., Huot, G., Cadar, A. E., … Ferbeyre, G. (2013). Metformin inhibits the senescence-associated secretory phenotype by interfering with IKK/NF-κB activation. *Aging Cell*, *12*(3), 489–498. doi:10.1111/acel.12075

Morgan, S. L., Mariano, N. C., Bermudez, A., Arruda, N. L., Wu, F., Luo, Y., … Wang, K. C. (2017). Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nature Communications*, *8*(1), 15993. doi:10.1038/ncomms15993

Mosteiro, L., Pantoja, C., Alcazar, N., Marión, R. M., Chondronasiou, D., Rovira, M., … Serrano, M. (2016). Tissue damage and senescence provide critical signals for cellular reprogramming in vivo. *Science (New York, N.Y.), 354*(6315), aaf4445. doi:10.1126/science.aaf4445

Muñoz-Espín, D., Cañamero, M., Maraver, A., Gómez-López, G., Contreras, J., Murillo-Cuesta, S., … Serrano, M. (2013). Programmed cell senescence during mammalian embryonic development. *Cell, 155*(5), 1104–1118. doi:10.1016/j.cell.2013.10.019

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., … Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature, 502*(7469), 59–64. doi:10.1038/nature12593

Nagano, T., Lubling, Y., Várnai, C., Dudley, C., Leung, W., Baran, Y., … Tanay, A. (2017). Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature, 547*(7661), 61–67. doi:10.1038/nature23001

Narita, M., Narita, M., Krizhanovsky, V., Nuñez, S., Chicas, A., Hearn, S. A., … Lowe, S. W. (2006). A novel role for high-mobility group a proteins in cellular senescence and heterochromatin formation. *Cell, 126*(3), 503–514. doi:10.1016/j.cell.2006.05.052

Narita, M., Nũnez, S., Heard, E., Narita, M., Lin, A. W., Hearn, S. A., … Lowe, S. W. (2003). Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell, 113*(6), 703–716. doi:10.1016/s0092-8674(03)00401-x

Nelson, D. M., Jaber-Hijazi, F., Cole, J. J., Robertson, N. A., Pawlikowski, J. S., Norris, K. T., … Adams, P. D. (2016). Mapping H4K20me3 onto the chromatin landscape of senescent cells indicates a function in control of cell senescence and tumor suppression through

preservation of genetic and epigenetic stability. *Genome Biology*, *17*(1). doi:10.1186/s13059-016-1017-x

Nir, G., Farabella, I., Pérez Estrada, C., Ebeling, C. G., Beliveau, B. J., Sasaki, H. M., … Wu, C.-T. (2018). Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genetics*, *14*(12), e1007872. doi:10.1371/journal.pgen.1007872

Noren Hooten, N., Martin-Montalvo, A., Dluzen, D. F., Zhang, Y., Bernier, M., Zonderman, A. B., … Evans, M. K. (2016). Metformin-mediated increase in DICER1 regulates microRNA expression and cellular senescence. *Aging Cell*, *15*(3), 572–581. doi:10.1111/acel.12469

Norton, H. K., Emerson, D. J., Huang, H., Kim, J., Titus, K. R., Gu, S., … Phillips-Cremins, J. E. (2018). Detecting hierarchical genome folding with network modularity. *Nature Methods*, *15*(2), 119–122. doi:10.1038/nmeth.4560

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., … Phillippy, A. M. (2022). The complete sequence of a human genome. *Science*, *376*(6588), 44–53. doi:10.1126/science.abj6987

Ocampo, A., Reddy, P., Martinez-Redondo, P., Platero-Luengo, A., Hatanaka, F., Hishida, T., … Izpisua Belmonte, J. C. (2016). [Review of *In Vivo Amelioration of Age-Associated Hallmarks by Partial Reprogramming*]. *Cell*, *167*(7), 1719-1733.e12. doi:10.1016/j.cell.2016.11.052

Ogryzko, V. V., Hirai, T. H., Russanova, V. R., Barbie, D. A., & Howard, B. H. (1996). Human fibroblast commitment to a senescence-like state in response to histone deacetylase

inhibitors is cell cycle dependent. *Molecular and Cellular Biology*, *16*(9), 5210–5218. doi:10.1128/MCB.16.9.5210

Olan, I., Parry, A. J., Schoenfelder, S., Narita, M., Ito, Y., Chan, A. S. L., … Narita, M. (2020). Transcription-dependent cohesin repositioning rewires chromatin loops in cellular senescence. *Nature Communications*, *11*(1), 6049. doi:10.1038/s41467-020-19878-4

O'sullivan, R. J., Kubicek, S., Schreiber, S. L., & Karlseder, J. (2010). Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres. *Nature Structural & Molecular Biology*, *17*(10), 1218. Retrieved from https://idp.nature.com/authorize/casa?redirect_uri=https://www.nature.com/articles/nsmb.1897.pdf%3Forigin%3Dppub&casa_token=OPnTBEvHtqsAAAAA:-gXbq4jQeljqXvJyNZnN_YN5NNEwuLQpp5PlPbG9d5AeEyGUzBh2leItI9uruR08K1IaoDpb3DSyOjeR

Oubaha, M., Miloudi, K., Dejda, A., Guber, V., Mawambo, G., Germain, M.-A., … Sapieha, P. (2016). Senescence-associated secretory phenotype contributes to pathological angiogenesis in retinopathy. *Science Translational Medicine*, *8*(362), 362ra144. doi:10.1126/scitranslmed.aaf9440

Parry, A. J., Hoare, M., Bihary, D., Hänsel-Hertsch, R., Smith, S., Tomimatsu, K., … Narita, M. (2018). NOTCH-mediated non-cell autonomous regulation of chromatin structure during senescence. *Nature Communications*, *9*(1), 1840. doi:10.1038/s41467-018-04283-9

Pathak, R. U., Soujanya, M., & Mishra, R. K. (2021). Deterioration of nuclear morphology and architecture: A hallmark of senescence and aging. *Ageing Research Reviews*, *67*(101264), 101264. doi:10.1016/j.arr.2021.101264

Pfitzner, D., Leibbrandt, R., & Powers, D. (2009). Characterization and evaluation of similarity measures for pairs of clusterings. *Knowledge and Information Systems*, *19*(3), 361–394. doi:10.1007/s10115-008-0150-6

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* , *26*(6), 841–842. doi:10.1093/bioinformatics/btq033

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., … Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, *159*(7), 1665–1680. doi:10.1016/j.cell.2014.11.021

Rauluseviciute, I., Riudavets-Puig, R., Blanc-Mathieu, R., Castro-Mondragon, J. A., Ferenc, K., Kumar, V., … Mathelier, A. (2023). JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*. doi:10.1093/nar/gkad1059

Reeves, R. (2010). Nuclear functions of the HMG proteins. *Biochimica et Biophysica Acta*, *1799*(1–2), 3–14. doi:10.1016/j.bbagrm.2009.09.001

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* , *26*(1), 139–140. doi:10.1093/bioinformatics/btp616

Rocha, A., Dalgarno, A., & Neretti, N. (2022). The functional impact of nuclear reorganization in cellular senescence. *Briefings in Functional Genomics*. doi:10.1093/bfgp/elab012

Rosenberg, A. B., Roco, C. M., Muscat, R. A., Kuchina, A., Sample, P., Yao, Z., … Seelig, G. (2018). Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science (New York, N.Y.)*, *360*(6385), 176–182. doi:10.1126/science.aam8999

Rosenthal, M., Bryner, D., Huffer, F., Evans, S., Srivastava, A., & Neretti, N. (2019). Bayesian Estimation of Three-Dimensional Chromosomal Structure from Single-Cell Hi-C Data. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, *26*(11), 1191–1202. doi:10.1089/cmb.2019.0100

Rowley, M. J., & Corces, V. G. (2018). Organizational principles of 3D genome architecture. *Nature Reviews. Genetics*, *19*(12), 789–800. doi:10.1038/s41576-018-0060-8

Sadaie, M., Salama, R., Carroll, T., Tomimatsu, K., Chandra, T., Young, A. R. J., … Narita, M. (2013). Redistribution of the Lamin B1 genomic binding profile affects rearrangement of heterochromatic domains and SAHF formation during senescence. *Genes & Development*, *27*(16), 1800–1808. doi:10.1101/gad.217281.113

Salmonowicz, H., & Passos, J. F. (2017). Detecting senescence: a new method for an old pigment. *Aging Cell*, *16*(3), 432–434. doi:10.1111/acel.12580

Sarkar, T. J., Quarta, M., Mukherjee, S., Colville, A., Paine, P., Doan, L., … Sebastiano, V. (2020). Transient non-integrative expression of nuclear reprogramming factors promotes multifaceted amelioration of aging in human cells. *Nature Communications*, *11*(1), 1545. doi:10.1038/s41467-020-15174-3

Sati, S., Bonev, B., Szabo, Q., Jost, D., Bensadoun, P., Serra, F., … Cavalli, G. (2020). 4D Genome Rewiring during Oncogene-Induced and Replicative Senescence. *Molecular Cell*, *78*(3), 522-538.e9. doi:10.1016/j.molcel.2020.03.007

Saul, D., Kosinsky, R. L., Atkinson, E. J., Doolittle, M. L., Zhang, X., LeBrasseur, N. K., … Khosla, S. (2022). A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues. *Nature Communications*, *13*(1), 4827. doi:10.1038/s41467-022-32552-1

Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nature Reviews. Genetics*, *20*(8), 437–455. doi:10.1038/s41576-019-0128-0

Schoenfelder, S., Javierre, B.-M., Furlan-Magaril, M., Wingett, S. W., & Fraser, P. (2018). Promoter Capture Hi-C: High-resolution, Genome-wide Profiling of Promoter Interactions. *Journal of Visualized Experiments: JoVE*, (136). doi:10.3791/57320

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., … Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature*, *551*(7678), 51–56. doi:10.1038/nature24281

Sen, P., Lan, Y., Li, C. Y., Sidoli, S., Donahue, G., Dou, Z., … Berger, S. L. (2019). Histone acetyltransferase p300 induces DE Novo super-enhancers to drive cellular senescence. *Molecular Cell*, *73*(4), 684-698.e8. doi:10.1016/j.molcel.2019.01.021

SenNet Consortium. (2022). NIH SenNet Consortium to map senescent cells throughout the human lifespan to understand physiological health. *Nature Aging*, *2*(12), 1090–1100. doi:10.1038/s43587-022-00326-5

Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D., & Lowe, S. W. (1997). Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell*, *88*(5), 593–602. doi:10.1016/s0092-8674(00)81902-9

Shah, P. P., Donahue, G., Otte, G. L., Capell, B. C., Nelson, D. M., Cao, K., … Berger, S. L. (2013). Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. *Genes & Development*, *27*(16), 1787–1799. doi:10.1101/gad.223834.113

Sheikh, B. N., Phipson, B., El-Saafin, F., Vanyai, H. K., Downer, N. L., Bird, M. J., … Thomas, T. (2015). MOZ (MYST3, KAT6A) inhibits senescence via the INK4A-ARF pathway. *Oncogene*, *34*(47), 5807–5820. doi:10.1038/onc.2015.33

Shi, X., Tian, B., Ma, C., Liu, L., Zhang, N., Na, Y., … Qiao, Y. (2017). GSK3β activity is essential for senescence-associated heterochromatin foci (SAHF) formation induced by HMGA2 in WI38 cells. *American Journal of Translational Research*, *9*(1), 167–174. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28123643

Shin, H., Shi, Y., Dai, C., Tjong, H., Gong, K., Alber, F., & Zhou, X. J. (2016). TopDom: an efficient and deterministic method for identifying topological domains in genomes. *Nucleic Acids Research*, *44*(7), e70. doi:10.1093/nar/gkv1505

Sloan, C. A., Chan, E. T., Davidson, J. M., Malladi, V. S., Strattan, J. S., Hitz, B. C., … Cherry, J. M. (2016). ENCODE data at the ENCODE portal. *Nucleic Acids Research*, *44*(D1), D726-32. doi:10.1093/nar/gkv1160

Soto, C., Bryner, D., Dalgarno, A., Neretti, N., & Srivastava, A. (2022, December). TADBay: A Bayesian Construction of Topologically Associated Domains. *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 233–240. doi:10.1109/BIBM55620.2022.9995248

Soto, C., Bryner, D., Neretti, N., & Srivastava, A. (2021). Toward a Three-Dimensional Chromosome Shape Alphabet. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*. doi:10.1089/cmb.2020.0383

Soto, C., Dalgarno, A., Bryner, D., McLaughlin, B., Neretti, N., & Srivastava, A. (2021, December). Representation of Chromosome Conformations Using a Shape Alphabet

Across Modeling Methods. *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 151–156. doi:10.1109/BIBM52615.2021.9669716

Soto, C. J., Zhao, P. A., Klein, K. N., Gilbert, D. M., & Srivastava, A. (2021). Statistical comparisons of chromosomal shape populations. *Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, *2021*, 788–791. doi:10.1109/isbi48211.2021.9433812

Srivastava, A., & Klassen, E. P. (2018). *Functional and shape data analysis*. New York, NY: Springer.

Stuart, T., Srivastava, A., Madad, S., Lareau, C. A., & Satija, R. (2021). Single-cell chromatin state analysis with Signac. *Nature Methods*, *18*(11), 1333–1341. doi:10.1038/s41592-021-01282-5

Su, J.-H., Zheng, P., Kinrot, S. S., Bintu, B., & Zhuang, X. (2020). Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell*, *182*(6), 1641-1659.e26. doi:10.1016/j.cell.2020.07.032

Tabula Muris Consortium. (2020). A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature*, *583*(7817), 590–595. doi:10.1038/s41586-020-2496-1

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663–676. doi:10.1016/j.cell.2006.07.024

Teo, Y. V., Rattanavirotkul, N., Olova, N., Salzano, A., Quintanilla, A., Tarrats, N., … Chandra, T. (2019). Notch signaling mediates secondary senescence. *Cell Reports*, *27*(4), 997-1007.e5. doi:10.1016/j.celrep.2019.03.104

Tjeldnes, H., Labun, K., Torres Cleuren, Y., Chyżyńska, K., Świrski, M., & Valen, E. (2021). ORFik: a comprehensive R toolkit for the analysis of translation. *BMC Bioinformatics, 22*(1), 336. doi:10.1186/s12859-021-04254-w

van Deursen, J. M. (2014). The role of senescent cells in ageing. *Nature, 509*(7501), 439–446. doi:10.1038/nature13193

Villeponteau, B. (1997). The heterochromatin loss model of aging. *Experimental Gerontology, 32*(4–5), 383–394. doi:10.1016/s0531-5565(96)00155-6

Vojta, A., Dobrinić, P., Tadić, V., Bočkor, L., Korać, P., Julg, B., … Zoldoš, V. (2016). Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Research, 44*(12), 5615–5628. doi:10.1093/nar/gkw159

Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics (Oxford, England), 33*(22), 3645–3647. doi:10.1093/bioinformatics/btx469

Wang, E. (1995). Senescent human fibroblasts resist programmed cell death, and failure to suppress bcl2 is involved. *Cancer Research, 55*(11), 2284–2292. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/7757977

Willmore, T. J. (1971). Tight immersions and total absolute curvature. *Bulletin of the London Mathematical Society, 3*(2), 129–151. doi:10.1112/blms/3.2.129

Wilson, W. H., O'Connor, O. A., Czuczman, M. S., LaCasce, A. S., Gerecitano, J. F., Leonard, J. P., … Humerickhouse, R. A. (2010). Navitoclax, a targeted high-affinity inhibitor of BCL-2, in lymphoid malignancies: a phase 1 dose-escalation study of safety, pharmacokinetics, pharmacodynamics, and antitumour activity. *The Lancet Oncology, 11*(12), 1149–1159. doi:10.1016/S1470-2045(10)70261-8

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., … Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation Journal: The Public Sector Innovation Journal, 2*(3), 100141. doi:10.1016/j.xinn.2021.100141

Xu, W., Zhong, Q., Lin, D., Zuo, Y., Dai, J., Li, G., & Cao, G. (2021). CoolBox: a flexible toolkit for visual analysis of genomics data. *BMC Bioinformatics*, *22*(1), 489. doi:10.1186/s12859-021-04408-w

Yang, J.-H., Hayano, M., Griffin, P. T., Amorim, J. A., Bonkowski, M. S., Apostolides, J. K., … Sinclair, D. A. (2023). Loss of epigenetic information as a cause of mammalian aging. *Cell, 186*(2), 305-326.e27. doi:10.1016/j.cell.2022.12.027

Zainabadi, K. (2018). A brief history of modern aging research. *Experimental Gerontology*, *104*, 35–42. doi:10.1016/j.exger.2018.01.018

Zhang, R., Poustovoitov, M. V., Ye, X., Santos, H. A., Chen, W., Daganzo, S. M., … Adams, P. D. (2005). Formation of MacroH2A-containing senescence-associated heterochromatin foci and senescence driven by ASF1a and HIRA. *Developmental Cell, 8*(1), 19–30. doi:10.1016/j.devcel.2004.10.019

Zhang, X., Jeong, M., Huang, X., Wang, X. Q., Wang, X., Zhou, W., … Goodell, M. A. (2020). Large DNA Methylation Nadirs Anchor Chromatin Loops Maintaining Hematopoietic Stem Cell Identity. *Molecular Cell, 78*(3), 506-521.e6. doi:10.1016/j.molcel.2020.04.018

Zhang, Z., Pati, D., & Srivastava, A. (2015). Bayesian clustering of shapes of curves. *Journal of Statistical Planning and Inference*, *166*, 171–186. doi:10.1016/j.jspi.2015.04.007

Zhao, H., & Darzynkiewicz, Z. (2013). Biomarkers of Cell Senescence Assessed by Imaging Cytometry. In L. Galluzzi, I. Vitale, O. Kepp, & G. Kroemer (Eds.), *Cell Senescence: Methods and Protocols* (pp. 83–92). doi:10.1007/978-1-62703-239-1_5

Zhu, H., Liu, X., Ding, Y., Tan, K., Ni, W., Ouyang, W., … Ding, Z. (2022). IL-6 coaxes cellular dedifferentiation as a pro-regenerative intermediate that contributes to pericardial ADSC-induced cardiac repair. *Stem Cell Research & Therapy*, *13*(1), 44. doi:10.1186/s13287-021-02675-1

Zhu, Y., Tchkonia, T., Pirtskhalava, T., Gower, A. C., Ding, H., Giorgadze, N., … Kirkland, J. L. (2015). The Achilles' heel of senescent cells: from transcriptome to senolytic drugs. *Aging Cell*, *14*(4), 644–658. doi:10.1111/acel.12344

Zirkel, A., Nikolic, M., Sofiadis, K., Mallm, J.-P., Brackley, C. A., Gothe, H., … Papantonis, A. (2018). HMGB2 Loss upon Senescence Entry Disrupts Genomic Organization and Induces CTCF Clustering across Cell Types. *Molecular Cell*, *70*(4), 730-744.e6. doi:10.1016/j.molcel.2018.03.030

Zufferey, M., Tavernari, D., Oricchio, E., & Ciriello, G. (2018). Comparison of computational methods for the identification of topologically associating domains. *Genome Biology*, *19*(1), 217. doi:10.1186/s13059-018-1596-9