SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Dipartimento di / Department of

Scuola di medicine

Dottorato di Ricerca in / PhD program Dottorato in Medicina Molecolare e Traslazionale
Ciclo / Cycle XXXVI

Curriculum in (se presente / if it is)

Chromatin conformation dynamics in cellular senescence

Cognome / Surname Di Patrizio Soldateschi Nome / Name Emanuele

Matricola / Registration number 875632

Tutore / Tutor: Lanzuolo Chiara

Cotutore / Co-tutor:
(se presente / if there is one)

Supervisor:
(se presente / if there is one)

Coordinatore / Coordinator: Prof. Francesco Mantegazza

ANNO ACCADEMICO / ACADEMIC YEAR 2023/2024

Index

Chapter 1 GENERAL INTRODUCTION

1.1 The 3D chromatin structure

All the highly regulated steps that allow the formation of a fully functional multicellular organism are driven by a macromolecule inside the nucleus of the cells, the deoxyribonucleic acid (DNA): a unique macromolecule present in all the organism's cells that, producing specific set of proteins, controls the diverse identities and tissues[1,2]. Because DNA translation into RNA has a degree of stochasticity[3,4], differences in the genic expression are due to the effect of all the epigenetic factors that directly or indirectly interact with DNA thus controlling the tridimensional shape of the genome inside the nucleus. This organization forms the chromatin that, passing by multiple levels of folding, allows almost two meters of human DNA to fit inside the nucleus with a cell-specific shape that finely regulates the gene transcription[2].

1.2 From histones to chromosome territories

At the lowest level of chromatin organization, the DNA is wrapped around the histones, composed by four subunits, H2A, H2B, H3, and H4[5]. One or more of these subunits can undergo a posttranslational modification (PTM). Among PTM there are histone modifications, also called histone mark (HM), which influences the DNA transcriptional activity[6]. HM can involve distinct residues of the histone subunits, leaving a specific covalent modification, such as methylation, acetylation, phosphorylation, ubiquitylation and others[6]. Combination of distinct PTM can give rise to diverse transcriptional output; for example, the trimethylation on the lysine K4 of the histone H3 (H3K4me3) is associated with active transcription[7], while the H3K27me3 is associated with transcriptional silencing[7]. Interestingly, co-presence of H3K4me3 and the trimethylation on the lysine K27 of the histone H3 (H3K27me3) on the same genomic regions is specific of the so called bivalent genes, transcriptionally inactive but ready to a fast switch of state[7].

Other HMs are more stable, as the trimethylation on the lysine K9 of the histone H3 (H3K9me3), marker of constitutive, transcriptionally silent, heterochromatin. However, in certain condition, as during cell differentiation, H3K9me3 can acquire a more dynamic state playing a fundamental role in the gene expression[8]. The chromatin organization level on top of histone marks are chromatin loops. These structures derive from the extrusion of a single chromatin filament through the action of the cohesin (a ring-shaped protein complex)[9] and put distal chromatin regions in physical proximity with the effect of increased DNA/DNA interactions[10].

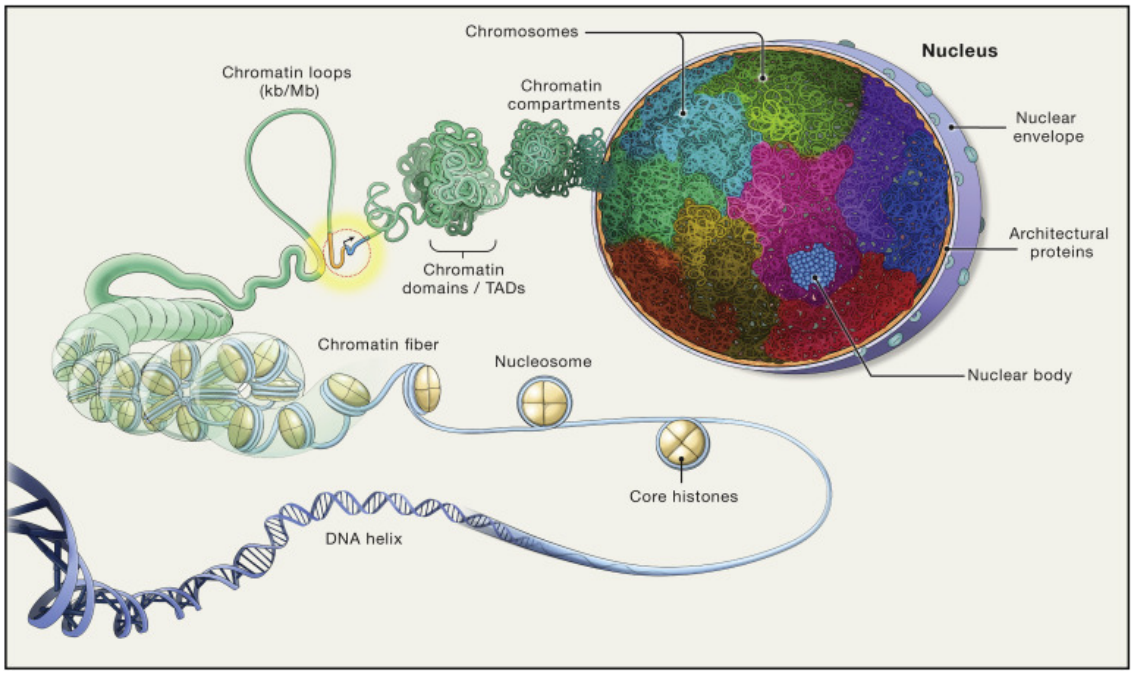


Figure 1. Cartoon of chromatin organization from DNA to chromosome territories Image taken from The Self-Organizing Genome: Principles of Genome Architecture and Function.[2]

These dynamics are fundamental in numerous nuclear processes such as enhancer-promoter interactions, gene expression, DNA repair, and DNA replication[11]. From a mechanical point of view, once cohesin starts to extrude the loops, it continuously slides through chromatin in a bidirectional process until it finds two convergent N-terminal domain of a CCCTC-binding factor (CTCF)[12]. At a higher hierarchy, larger genomic regions with a high probability of forming loops within are called topologically associated domains (TADs)[10,13], described for the first time with an innovative, sequencing based technology, the HiC[13], and later

on with the GAM (Genome Architecture Mapping)[14]. Both HiC and GAM are techniques developed to measure the physical distance between two or more chromatin segments. TAD borders correspond to the asymmetrical and convergent CTCF binding sites clusters[15], suggesting that the regions of interaction are the same for all cells, although degrees of interactions are cell dependent.

At a larger scale in the chromatin organization, there are genomic portions where chromatin segregates with lower or higher interaction frequencies, the A and B chromatin compartments (**Figure 1**). Whereas A compartments show lower interactions and correspond to the gene dense regions and transcriptionally active genome, B compartments display a higher interaction interactions frequency which signify more densely packet regions, yet poor in gene density and scarcely transcribed[13]. The genome shows a high level of plasticity in switching between compartments, in a study of chromatin conformation during differentiation from human embryonic stem cells (ESCs) toward four different human ESC-derived lineages, mesenchymal stem cells, neural progenitors, mesendoderm and trophoblast-like cells, shows that 36% of the genome have a modified compartments in at least one of the cell types[16]. These observation were corroborated by another work, in which it has been reported that during the differentiation from bone marrow mesenchymal stem cells to adipocytes the degree of genomic regions that change compartments is up to ~ 20%[17]. In particular, what undergo the most profound changes in cell differentiation are TADs boundaries. In fact, only 11-14% remain stable on multiple differentiated cell types[18]. In particular, few of them appear to be highly conserved across different tissue and also in cancer cells[19], suggesting a pivotal role for these specific TADs.

Finally, the highest level of the chromatin organization is the chromosome positioning in the nucleus[2]. The chromosomes are not randomly entangled; they occupy a discrete portion of the tridimensional nuclear space, called chromosomal territories[20,21]. They also remain conserved along similar species and were confirmed through FISH and chromosome painting[20].

The entire chromatin structure is highly conserved in cell types and is inherited through cell cycle[22]. Nonetheless, it shows a high level of dynamicity with chromatin remodelers that modify the HM upon cell exit from quiescence[23,24]. An example of this dynamicity is represented by DNA loops, dynamically created and persistent for ~ 10-30 minutes in the nuclei before return to the un-looped state[25,26].

1.3 The role of the nuclear envelope in epigenome regulation

The nuclear envelope (NE) is composed of a double layer membrane and beneath is present the Nuclear Lamina (NL). The NL is a protein meshwork composed of a type V filament, mainly A and B lamins, that shows a variable degree of thickness along the nuclear envelope[27] . Lamins in mammals are created by three lamin genes, *LMNA*, *LMNB1* and *LMNB2*, which can be spliced in no less than nine isoforms[28,29]. Whereas the B types appear to be constitutively expressed, the A type is missing in some cell types[30]. When present, A type lamins play a key role in cell differentiation. This role is also due to the interaction with several transcriptional factors, including the Polycomb group of proteins (PcG), transcriptional repressors that control the transcription of developmental and differentiation genes[31,32]. Beside the well documented role of lamin A, the ratio between A and B lamin amounts orchestrates cell differentiation[33]. Lamins distribution are cell-type specific and both types were found as detergent-soluble pools in the NL[34,35], then it has been demonstrated that the four main isoforms A, C, B1, and B2 form similar yet independent meshworks[36]. The NL have many roles in the cell response mediated by many binding proteins[37]. One of the major biological function of lamin proteins is their mechanotransduction role, that consists of transferring mechanical information from the cytoplasm to the genome, to control multiple functions such as cell differentiation and motility[38]. There are complexes that lie in the nuclear envelope passing through the double layer of the nucleus physically linking the cytoskeleton proteins and the nuclear

lamina. These complexes, called Linker of Nucleoskeleton and Cytoskeleton (LINC), are also needed to provide spatial and structural integrity of the nucleus[39,40].

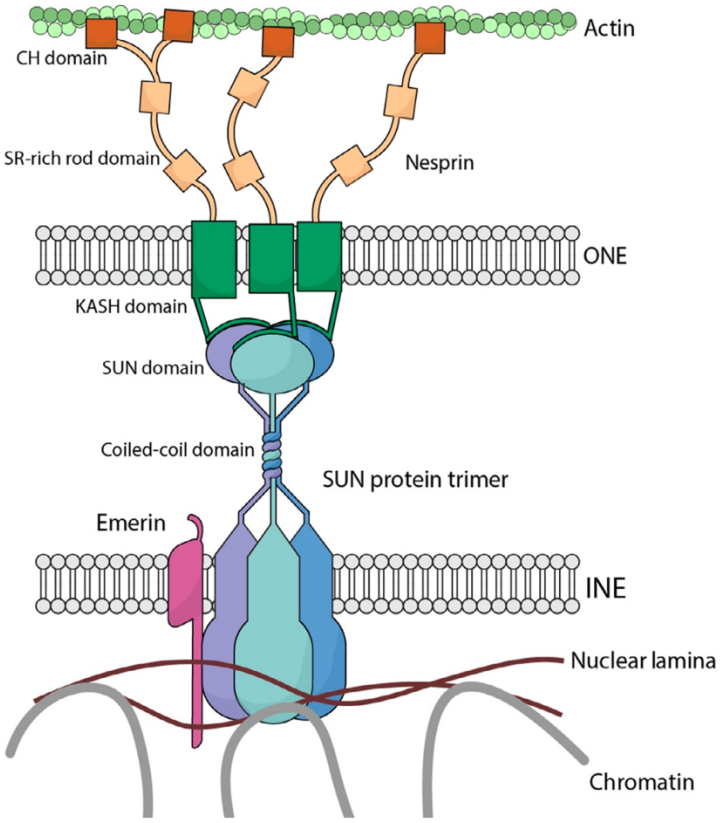


Figure 2. schematic representation of LINC complex and its main components. Image taken from Cellular Biomechanics in Skeletal Muscle Regeneration[41]

The LINC complexes are composed by the Sad1p, UNC-84 (SUN) proteins (SUN1 and SUN2) in the perinuclear space of the NE. SUN proteins can be bound to the Klarsicht/ANC-1/Syne Homology (KASH) creating various isoforms. The KASH proteins have a domain which lies on the NE, which bind nesprin proteins 1 to 4 (spectrin-repeats proteins) and ultimately bind microtubules or F-actin in the outer NE[42](**Figure 2**). In the inner NE the SUN proteins interacts with the lamina[43], that in turn directly interacts with 40% of inactive and gene-poor chromatin of sizes that may vary from 0.1–10 Mb in size[44]. These portions of chromatin, called lamina-associated domain (LAD), are enriched in repressive epigenetic markers like H3K9me2, H3K9me3 and H3K27me3[44,45]. The presence of LADs lead to the radial organization of the chromosome territories while

keeping gene-dense regions towards the nuclear centre[46,47], as recently showed by GPseq, a technique that, by slow diffusion of restriction enzyme, is able fractionate DNA by its proximity with the nuclear periphery.

Proteins that lie in the inner nuclear membrane as lamin B receptor (LBR), polypeptide 2 (LAP2), emerin, and LEM domain-containing protein 3 (LEMD3 or MAN1) interact with heterochromatin which is in close contact with the lamina, with functional consequences(**Figure 3**).

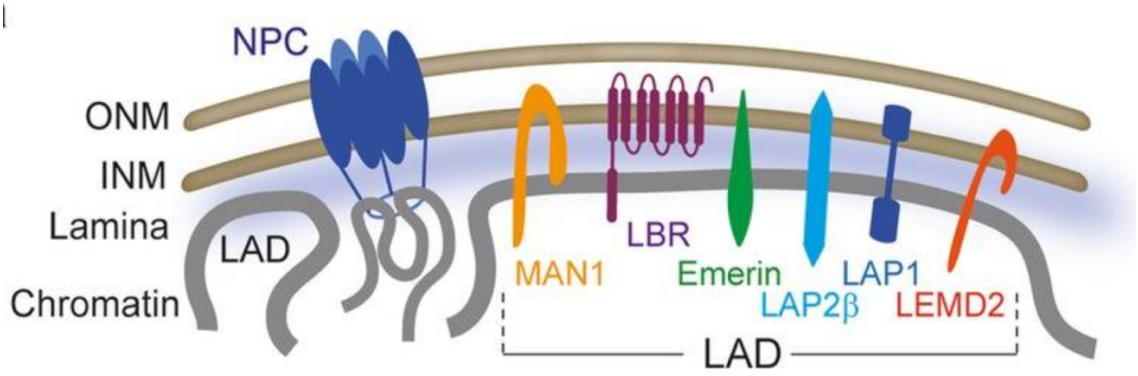


Figure 3. Cartoon representation of NE proteins interacting with chromatin. Lamina-associated domains: peripheral matters and internal affairs[48]

As an example, the direct binding during differentiation of emerin to muscle specific genes, maintains muscle progenitor identity[49]. Some of the lamina-associated domains are stable in cells (defined as constitutive LADs) while some of them change during lineage commitment and differentiation (classified as facultative LADs)[48,50] modifying the epigenetic repressive HM in these regions and subsequently the transcriptional level in the genes. LADs are dynamically formed during the cell cycle, with the contact frequency of LADs with the NL appear to be guided by H3K9me2 quantities and the H3K9 methyltransferase G9a and GLP1[51]. In addition, a subsequent work demonstrate that whereas in the early G1 phases the contacts with the NL are wide-spread through all the genome, they become progressively more defined increasing size along time[52]. During differentiation, some LADs can be formed de novo, and others increase or decrease their extension. However, during differentiation, when the association with the lamina is not accompanied by a HM remodeling, the genes

that are associated escape to the repressive LAD environment. The final LAD position after completing the differentiation appear to be pre-patterned by the HM positions[53]. This could be associated with only the facultative LADs, which most probably would not have time to remodel the HM in the differentiation process as it happens after a prolonged period of time[54]. In fact, also if the detachment happens by forced mechanical stress the chromatin remodeling happens in a matter of days by remodeling the H3K9me2/3 in the LADS by switching to H3K27me3 as a compensatory mechanism[55]. Notably, the constitutive LADs show ~ 200 genes that escape the characteristic repressive environment of LADs and are active and transcribing. These genes shows enrichment in open chromatin HM as H3K4me1/3 and H3K27ac which are not affected by the constant positioning on the lamina[56]. Also TADs, which tend to group in cliques, interact with the nuclear lamina. In a model of adipogenic differentiation, these interactions get stronger and shape radial reposition of the cliques, redefining the heterochromatin at the nuclear periphery[57].

1.4 Laminopathies

Laminopathies are a group of diseases that are caused by a mutation in the lamin genes. Over 600 known point mutations on the LMNA/C gene can cause laminopathies. These mutations occur along the LMNA gene with several “hot-spot” regions associated with 15 distinct diseases[37]. The phenotype macro-areas of these diseases are lipodystrophies, neuropathies, bone disorders, muscular dystrophies as the Emery Dreifuss, cardiomyopathies and premature aging diseases as atypical Werner, Hutchinson- Gilford Progeria Syndrome (HGPS or progeria) and dermopathies[58]. Also, if mutated, the LMNB1 and LMNB2 can lead to laminopathies[59]. However, there are fewer reported laminopathies caused by LMNB genes compared to lamin A. As described in the work of Camilla Evangelisti et al., it is probably related to the fact that lamin B is required in embryonic and fetal development, thus mutations are most likely lethal[59]. One of the most notable laminopathies is HGPS, as the affected patients show in some

tissues an accelerated physiological ageing. This disease has caught the scientists' attention because it can be used on top of the understanding of the pathology to learn mechanisms related to physiological aging[60].

1.5 Hutchinson- Gilford Progeria Syndrome (HGPS)

The HGPS is a segmental autosomal dominant ultra-rare genetic disease with an incidence of 1:18 million and patients affected usually do not live longer than 13-14 years old on average[60]. The symptoms manifest early age, usually already within the first year of life. It shows phenotypes as sclerotic skin modifications, reduced body fat, loss of vascular smooth muscle, severe atherosclerosis, ubiquitous vascular stiffening and arterial occlusive diseases, protracted prothrombin times, conductive hearing loss, alopecia, growth impairment with bone abnormalities, and abnormal dentition[60]. The most common cause of death in progeria is the heart attack or stroke generated from the vascular diseases. Notably, HGPS patients do not face a decline in functions in many organs, such as kidneys and gastrointestinal tract, or the neurocognitive capacity remains unaffected[60].

1.6 Progerin and its effects on nuclear structure

Progeria is caused by a single point mutation on the G608G site , typically caused by the mutation (GGC>GGT) on the exon 11 at position 1824 in the LMNA gene[61]. This mutation creates a cryptic splice site resulting in an mRNA without 150nt at the 3′ end. When translated, the lamin A precursor polypeptide lacks the second endoproteolytic cleavage site for ZMPSTE24, needed to eliminate 50 residues in the C-terminal tail of the protein. This leads to a persistent CaaX farnesylation motif and increases abnormally the farnesylated precursor[62], the progerin, that does not correctly integrate into the nuclear lamin mesh. The lamin isoforms are also crucial for regulating the correct distribution and regulation of the nuclear

pore complexes in cells. Progerin accumulation triggers an abnormal nuclear envelope morphology that leads to the formation of protruding structures where the nuclear lamin is thinner, the nuclear blebs[62,63]. The nuclear blebs are depleted in nuclear pores[64]. Besides their role in molecules transportation across the NE, nuclear pores interact with chromatin[65], regulating chromatin silencing and contributing to long-range interactions among Polycomb targets[66].

These blebs have been suggested to have a role in the severity of the disease through the overaccumulation of Sun1[67]. In fact, in the study of Chai et al. by reducing the SUN1 quantity in a floxed conditional LMNA mutant of mice, the progression of dilated cardiomyopathy is repressed[68]. Interestingly, nuclear blebs contain more active chromatin inside, creating a strongly deregulated epigenetic environment[69]. Recently it has been reported that micronuclei formation, which are created by the mis-segregation of an entire chromosome portions during replication, also disrupts the epigenetic environment of the nucleus[70–72]. Furthermore, because the heterochromatin is more rigid than the open and active euchromatin, by increasing the euchromatin/ heterochromatin ratio, the nucleus stiffness decrease and result in creating blebs independently of the lamins[73]. Taken all together these results show the tight connection and importance of chromatin architecture and nuclear shape to the homeostasis of the nucleus. For example, nuclear blebbing can be caused initially by mechanical forces. However, when the LADs are detached and the histone modifications are remodeled in late passages, the modified stiffness can increase the severity of nuclear blebbing. Progerin overaccumulation changes the structure of the nuclear lamina by creating lamin microaggregates, this affects the nuclear lamin ultrastructure and alters the force propagation in the nucleus[74], changing in the mechanical properties of the nuclei. Although this chain of event does not increase the nuclear fragility[75], the cells can be more sensitive to mechanical strain[76].

1.7 HGPS dysfunction of biological pathways

HGPS affect tissues expressing lamin A, which are mainly cells derived from mesenchymal lineage[77]. Some conditions are common to all the affected tissues, as heterochromatin remodeling, alteration of the mechanical properties and mechanotransduction, increased cellular senescence and apoptosis. An important deregulated pathway that affect HGPS patients is caused by the mammalian target of rapamycin (mTOR) activity reductions. mTOR is divided in two different complexes mTORC1 and mTORC2 which are involved in a plethora of functions as it regulates cell proliferation, autophagy, apoptosis and nutrient sensing[78]. In progeria, as mTOR signaling activity drops, the autophagy dramatically increases[79]. In fact, inhibition of the mTOR signaling pathway significantly increase the HGPS mouse lifespan[80].

Another important aspect of progerin accumulation in the nucleus is the oxidative stress, in part caused by mitochondrial dysfunctions in morphology and functions[81,82]. Another key player of oxidative stress is the transcriptional nuclear factor erythroid 2-related factor 2 (NRF2). Progeria shows intense oxidative stress because NRF2 is sequestered away in the nuclear periphery by progerin its transcriptional targets[83]. NRF2 target knockdown, in human fibroblasts cell lines, triggers senescence phenocopying the progeric phenotype[83]. Sirtuins are a family of histone deacetylases involved in various roles such as DNA repair, resistance to DNA damage and chromatin conformation[84]. The function of sirtuins has been linked to Lamin A: in mouse embryonic fibroblasts (MEF), the antioxidant resveratrol compound increases the binding of SIRT1 with A-type lamins stabilizing its deacetylase activity and resulting in an alleviation of progeroid features and retardation of cell death[85]. On the other hand, association of SIRT1 with progerin accelerates SIRT1 decline accelerating the aging in adult stem cells.

It has been demonstrated that the Ig-fold motif located in the C-terminus of the lamins directly binds the proliferating cell nuclear antigen (PCNA) and, by direct consequence, influence the DNA replication efficiency[86]. Another emerging role of the lamins, is the promotion of DNA damage repair by base excision repair (BER) during oxidative stress. In particular, if the lamin A/C is absent, the

oxidative stress leaves a specific fingerprint absent in the wild type[87]. Instead, lamin B does not appear to have a direct role in DNA damage response. However, it does interact with the transcription factors of genes implicated in DNA damage and repair as breast cancer gene 1 (BRCA1), RAD51, and the small ubiquitin like modifier 1 (SUMO1)[88]. Another role of the nuclear lamina meshwork is that it interacts with histones with its immunoglobulin-like domain[89]. During DNA damage, lamin A interacts with γH2AX, the phosphorylated form of the histone H2AX, this interaction is fundamental for the stability in the 3D space of the DNA repair foci[90]. All of these impaired mechanisms in DNA damage repair would lead to the supposition that HGPS patients are have an increased susceptibility in cancer development. Conversely, the patients appear to have a protection mechanism[91]. Cancer escaping mechanism rely on BRD4, a transcriptional activator which, in HGPS cells, abnormally spreads on chromatin. The abnormal dissemination of BRD4, mediated by progerin, regulates specific genes which hinder the de-differentiation process and the subsequent carcinogenic reprogramming[92,93].

1.8 Epigenetic alterations in HGPS

The chromatin remodeling in HGPS starts at the lamina-associated domains level. In these domains there are histone marks associated with heterochromatin mostly H3K9me2 or me3 and H3K27me3 at the borders (silenced, gene-poor and inactive)[44]. It has been reported that progerin increases during cell passages, which leads to a loss of heterochromatin governed by the histone mark H3K9me3, and by contrast, an increase of H4K20me3 in cells[94], a repressive HM of promoters and transposons[95]. The methyltransferase SUV39H1, responsible for the trimethylation of H3K9me3 shows a decrease in late passages, and H3K27me3 displays an aberrant deposition patterns being absent on chromosome X[94,96].

Another essential complex for heterochromatin maintenance is the nucleosome-remodeling and deacetylase (NuRD) complex[97]. In progeric cells the necessary

subunits that bind histones, RB Binding Protein 4 and 7 (RBBP4, RBBP7), Metastasis Associated 1 Family Member 3 (MTA3) and Histone Deacetylase 1 (HDAC1) are downregulated, resulting in a depletion of focal H3K9me3[98]. However, studies performed in early passages of progeric fibroblasts it has been reported that in presence of progerin one of the first alteration is LAD relocation from nuclear periphery towards the nuclear center[54]. This change is followed a low spreading and decrease of H3K27me3 quantity[54]. This chromatin relocation reflects in a subset of bivalent genes, more susceptible to variations of PcG bodies position which increase their expression[54]. On the contrary, H3K9me3 do not show any modification in the early passages[54]. Nonetheless, in late passages, these genomic relocations and epigenetic alterations result in H3K9me3 depletion and are subjected to a nuclear re-compartmentalization of Polycomb[54]. This remodeling is followed by impaired chromatin interactions seen with Hi-C analysis[96].

Beside its direct role on chromatin, progerin expression interferes with many nuclear factors that control the epigenetic state of chromatin. One of them is the one now called ubiquitin conjugating enzyme E2I (UBE2I), the specific E2-conjugating enzyme necessary for the SUMOylation, necessary for the establishment of H3K9me3[99]. The reduction of SUMOylation has significant consequences on cell identity and cell-specific gene expression at least in cell lines as mouse embryonic fibroblasts (MEFs) and embryonic stem cells (ESCs)[91].

A growing quantity of evidences suggests that Polycomb group of proteins (PcG) play important roles in several tissue strongly affected in HGPS: aortic tissue, muscle, and skin homeostasis[100–102]. As the chromatin conformation is cell-type specific, chromatin remodeling due to progerin presence in different cell types would be most probably cell-type specific. However, no studies specifically addressed this question.

1.9 Tissues alterations in HGPS

Progerin has an impact on various tissues, affecting homeostasis and maintenance. In addition as pre-laminA accumulation leads to cell toxicity, different cell lines or mouse models are used as models to understand a specific type of overaccumulation and mechanisms[103–105]. In osteogenesis, the Wnt/β-catenin signaling promotes differentiation, whereas in HGPS, an impairment in the import of β-catenin cause defective osteogenesis by decreasing the mineralization[106,107]. In addition, increased pre-lamin A levels in mesenchymal stem cells cause an increase of osteogenesis-related proteins in the secretome, thus promoting osteoblast differentiation[108]. Progeria patient also shows impairment in the correct regulation of some central genes that oppose chondrocyte differentiation, thus increasing the mesenchymal stem cell commitment[109,110]. In skin, a fundamental process for the maintenance of the tissue is the stem cell asymmetric division, which create one cell that replenish the stem pool and the other committed to differentiate[111]. The asymmetric cell division is regulated by signaling pathways as Hedgehog, Notch, and Wingless (Wnt)[112]. The latter, aberrantly activated by progerin accumulation, leads to an increased symmetric cell division. This causes an abnormal differentiation in keratinocytes resulting in an unbalance in skin homeostasis[113].

In endothelial cells, progerin accumulation causes cell senescence and, thereby, cell loss. This results in reduced contractility, excessive buildup of extracellular matrix and calcification of the middle layer of the aortic wall[114]. In the vascular smooth muscles, important for vascular contraction and for the maintenance of the correct blood pressure, prelamin A presence hinders the capacity of endothelial cells to adapt to mechanical stress, resulting in cell loss and senescence[77]. In a elegant model of HGPS mice that express progerin only in endothelial cells, it has been shown that lack of mechanical stress adaptation is accompanied by alterations in the production of nitric oxide[115], which is important for cell adhesion and inflammatory response[116]. Importantly, recent studies have shown that suppression of progerin and restoration of Lamin A levels in endothelial cells by a Cre recombinase in the mice model, increased the lifespan of these mice[117]. In addition, they also show that the sooner progerin is depleted,

the higher is the lifespan of the mice. Similar results were reported in the HGPS mouse model expressing progerin only in cardiomyocytes[117].

In cardiac tissue the effects of progerin expression have been studied for a long time. The Zmpste24$^{-/-}$mouse model, a premature aged mouse mouse that lack the enzyme for endoproteolytic cleavage of the CaaX farnesylation motif, develops cardiac symptoms similar to HGPS. Furthermore, it shows reduced amplitude of the intracellular Ca2+ fluxes, which is central in the contractility regulation[118]. With the advent of cell reprogramming, executed also in progeria cells[119], several studies started also in human cells. Human cardiomyocytes obtained from differentiating iPSCs, report morphological and biochemical alterations in mitochondria[120]. In addition, they curiously found a reduced level of intracellular ROS which usually increase in other tissues[81]. As cardiac cells are characterized by absence of proliferation in adult, in both progeria and physiological aging there are not great differences in single nucleotide variants (SNV). This suggest that the increased genomic instability in progeric cells do not affect the DNA repair machinery which remain fully functioning to maintain genomic integrity[121]. Nonetheless, at the proteomic level the alterations present in the hearth of progeric mouse model *Lmna^{G609G}*, a knock-in strain model with the same genetic mutation as a progeric patient[122], are quite different from the physiologically aged hearts. On the other hand both, physiological and accelerated aging present a proteome which relate to increased levels of cardiometabolic diseases[123].

In adipose tissue, the role of progerin has been deeply investigated, due to the strong lipodystrophic phenotype observed in progeric patients. The Janus kinase/signal transducer and activator of transcription (JAK/STAT) signaling, an inflammatory pathway related also in senescence, play and important role in the development of lipodystrophy in HGPS. In fact, treating in vitro cells with baricitinib (Bar), a JAK/STAT inhibitor, ameliorated adipogenesis[124]. At the chromatin level, the differentiation from adipose stem cells show a tight relation with the lamina[53,125]. In stem cells, cliques of non-contiguous TADs naturally interact inside the nuclei. During adipogenic differentiation, these cliques display

an increased size and interaction NL, stabilizing the heterochromatin domains and reinforcing the silencing of developmental genes[57]. The role of of lamin A in adipocyte differentiation was described in other laminopathies that affect adipose tissues as partial lipodystrophy with mandibuloacral dysplasia (MAD) and Dunningan-type familial partial lipodystrophy (DFPLD)[126]. Interestingly, the alteration of one tissue can play a role in the dysfunction of adjacent tissues: a recent study described senescence-associated secretory phenotype produced by endothelial cells that promotes adipocyte senescence[127].

1.10 DNA damage in HGPS

Several biological pathways involved in physiological cell senescence are also affected in progeria: telomere shortening, DNA damage and impairment of a correct DNA repair response, nuclear blebbing, LAD detachment followed by post translational modifications of heterochromatin[81,128]. Many factors cause DNA damage accumulation in progeria causing the replicative stress. In addition, these factor are tightly related to the nuclear lamina[129]. The first one, is the RAD51 recruitment operated by lamin A in the process of homologous recombination. The RAD51 function is to protect DNA in the stalled replicative forks (RF); loss of lamin A cause a recruitment defect that increase the nucleolytic degradation of stalled RFs, causing genomic instability and increasing DNA damage[130]. When double strand breaks (DSB) occur in heterochromatin, the portion of the genome containing the DSB relocalizes to the periphery of chromocenters, with a mechanism that involve actin and myosin binding to lamin A[131]. Here damaged DNA recruits Rad51, that start the DNA repair. In presence of progerin the crosstalk between myosin and lamin A is affected[132], causing an impairment in the double strand break repair method[133].

Both HGPS and ZMPSTE24$^{-/-}$ cells have abnormally high levels of the Xeroderma Pigmentosum Group A (XPA) protein, which is involved in nucleotide excision repair, at the sites of DNA damage[134]. The DNA damage response (DDR) is also affected in HGPS by the 53BP1, a lamin A interacting protein, which is a key

mediator in multiple DDR processes in the double strand break repair. In the absence of lamin A, dysfunction of 53BP1 causes genomic instability and inhibits DNA repair[135]. Heterochromatin protein 1 (HP1) is a family of proteins mainly known to form heterochromatin via the recruitment of SUV39H1[136]. HP1α can interact with lamins, having a role in chromatin remodeling and DNA repair. In progeroid mice, HP1α increase in the nuclear matrix, while its phosphorylation capacity drop. This results in a delay in the formation of the DNA-repairing foci mediated by the γ-H2AX which end up in increasing the number of DNA breaks and defective DNA repair, thus causing senescence[137].

Progeria cells also show telomere shortening, a process tightly linked with DNA damage[138]. Telomeres, the end segments of the linear chromosomes, increase their length by the enzyme telomerase. This enzyme activity in cells is highest only in the germline and absent in most of the cells after birth[139]. As a result, telomeres shorten progressively with age, contributing to aging[139]. Telomeres have a specific architecture consisting of two parts: a repeated TTAGGG region that is several Kb long, and a single-stranded 3' tail called the G-overhang[139]. To protect the telomeres from falsely undergo to a double strand DNA break repairment, the ending portions of the chromosome are protected with a particular structure which gave the cap structure, or the T-loop[139]. These structure are mediated by the shelterin complex which is a complex comprised of various sub-units as telomere repeat-binding factor (TRF1, TRF2), TERF1-interacting nuclear factor 2 (TIN2) and protection of telomeres protein 1 (POT1)[139]. In human cells, telomeres are bound to the lamin A by the TTAGGG repeat[140]. As a consequence, several proteins of the shelterin complex interact with lamin A. When the lamin A is perturbed, fewer protective structures are formed and, consequently, telomere are rapidly lost[141]. In fact, telomere shortening is variously reported in HGPS as a major driver of senescence[142,143]. Telomeres also interact with lamina-associated polypeptide 2 (LAP2α), a key interactor of lamin A. When progerin is present, it disrupts this interaction, altering telomeres correct positioning[144]. This probably has an effect on the telomeres protective complex shelterin as it also directly interact with LAP2α[145]. When the telomeric

DDR remain active for prolonged period of time, the cells enter cellular senescence[146]. The response of this kind of DNA damage generates non-coding RNA (ncRNA) which, which after a processing step mediated by the endoribonucleases named DROSHA and DICER, mediate the formation of the DDR foci[147]. In fact, by inhibiting the telomeric DNA damage response, the HGPS senescent phenotype and proliferation ability ameliorate[148].

1.11 The epigenetics of cellular senescence

Cellular senescence is a stage induced by a wide range of intrinsic and/or extrinsic cell vexations[149]. These inducers are correlated with processes such as ageing, cancer, tumor suppression/ promotion, tissue repair, and diabetes[150]. Cellular senescence nowadays can be divided in 3 macro categories, physiological, stress induced and replicative. The first one is a mechanism used by organisms to a plethora of physiological functions: embryonic development, wound healing, tissue remodeling, tumor suppression and brain homeostasis [151,152]. The second is a response to stresses to impede further damage often called stress induced premature senescence (SIPS)[153]. In particular in the SIPS macro category include the more known oncogene induced senescence (OIS)[154]. The third one, the replicative senescence (RS) is due to the telomere shortening called 'Hayflick's limit'[155].

Despite the initial theory that assumes cellular senescence as a static process, this is now considered as a highly dynamic multi step process[156]. The early senescent state is delineated by morphology changes, as flattened and enlarged morphology with organelle abnormalities, chromatin remodeling, and, consequently, metabolic reprogramming, which could end up in production of a specific secretome signature called senescence-associated secretory phenotype (SASP)[157]. The effect of SASP can be autocrine to auto reinforce senescence[158,159], and paracrine to both induce senescence in neighbors cells and signal the macrophages to remove senescent cells[160]. SASP can vary across cell types and stages[161]. In addition, SASP components can be categorized in

four distinct categories: metabolic processes, extracellular matrix/cytoskeleton/cell junctions, ox-redox factors, and regulators of gene expression[162]. The initiation of senescence is characterized by the arrest of the cell-cycle mediated by p53/p21^{CIP1} and p16^{INK4a}/Rb[163]. Both of them subsequently acts as barriers to re-enter the cell cycle. In fact, depletion of both factors can re-establish the cell proliferation[164].

Afifi et al. recently showed that cells undergo irreversible senescence after losing cyclin-dependent kinase 2 (CDK2) activity, which is caused by c-MYC (MYC) degradation. If MYC increases inside the cells, the cells re-enter in the cell cycle[165]. This c-MYC degradation appear to have an effect of blocking the cell cycle and let the cells enter in the senescent state stronger than the very well-known p53/p21^{CIP1} and p16^{INK4a}/Rb, as they can induce reversible arrest. Despite the fact that senescence is generally thought to be irreversible and the work of Afifi et al. strongly sustain this theory, this is still under debate as there are multiple evidences that cells can escape senescence and re-enter in the cell cycle[119,164,166,167], in particular in cancer cells[168,169]. An important consideration to take into account is that more the cells remain in senescent state, more they diverge from the genome cell-specific program, thus increasing their transcriptional heterogeneity, thus changing the cell identity[170]. During cellular senescence, chromatin is greatly remodeled with different players. At the molecular level, common triggers of cellular senescence during ageing are: DNA damage accumulation, critically short telomeres, mitochondrial defects, ROS and nutrient imbalance[150,171].

The DNA damage is known to be particularly effective in induce senescence, in particular the persistence of double strand break (DBS) not readily resolved by the cells triggers cellular senescence[149]. In particular, telomeres, when not accurately protected are considered by the cell a persistent DSB. When telomeres become critically short they do not form anymore the protective structure called cap. Then, they attract a DNA damage response (DDR) which is quite similar to normal DSB[172,173].The oncogene induced senescence (OIS) cited above, display senescence-associated heterochromatin foci also known as

SAHF. These bodies are enriched in non-overlapping concentric layers of H3K9me3 and H3K27me3[174], kept together by HP1 proteins[175]. In addition, DNMT1, a DNA methyltransferase, is an essential factor for SAHF formation[176]. Both SAHF and DDR are implicated in the OIS as they can coexist in the same cell, however, they are in each other proximity and do not colocalize. This proximity is necessary to restrain the DDR signaling[177]. Epigenetic alterations are different in replicative or not replicative senescence. In the first one, depletion of lamin B1 from the nuclear envelope accompanies cellular senescence[178]. This result in the partial loss of insulation and decrease in genome compartmentalization at TAD borders, leading to the broadly accepted knowledge that during ageing a generalized heterochromatin depletion is accompanied by LAD detachment[171,179]. On the other hand, the senescence imposed by the oncogene suppression increase compartmentalization and create senescence associated heterochromatin foci[176]. However, heterochromatin studies were mainly conducted in cultured cells and few datasets are available in the animal model. A systematic genome wide approach, done at different stages of lifespan, in distinct tissues could clarify the common mechanisms of heterochromatin alterations in senescent cells.

1.12 Cellular senescence in ageing

Aging, is loosely defined as a time-dependent deterioration of functionality that affects almost every organism[180]. Ageing is due to a plethora of cellular and molecular events, e.g. stem cell exhaustion, telomere shortening, DNA damage and epigenetic alterations. These factors are considered hallmarks of ageing are tightly intertwined together[180]. Cellular senescence is also considered a hallmark of aging[180] and the number of senescent cells increases during lifespan. As the number of senescent cells increases, the amount of senescence-associated secretory phenotype (SASP)(described in the paragraph 1.10) increase in the extracellular matrix contributing to create the low-level chronic inflammation microenvironment in ageing, named inflammaging. The inflammaging is a

complex status of the aged tissues in which the innate immune system, in particular the macrophages, are kept constantly activated[181]. The accumulation of senescent cells in tissues during aging leads to continuous release into the bloodstream of molecules, such as recruitment factors, damage-associated molecular patterns (DAMPs), and cytokines as IL-1, IL-6, interferon-gamma (IFN-γ), tumor factor necrosis TNF-α, cytokine antagonists[182,183]. Because SASP attracts macrophages that in homeostatic condition would remove senescent cells, one hypothesis is that the accumulation of senescent cells can be triggered by a decrease in the efficiency of the immune system, that contribute to the SASP production[184]. Intriguingly, it has been shown that brief exposures to SASP microenvironment can increase the expression of stem cell markers and augment regenerative capacity[151]. On the other hand, long exposures promote a cell-intrinsic senescence[151]. During physiological ageing, SASP composition and release quantity appear to be related to both tissue and age[185].

1.13 Muscle aging

During the functional decline due to ageing process, one of the most relevant and affected tissue is the muscle. Sarcopenia, the age-related loss of muscle mass and strength, is a major driver of decline in older adults[186]. Sarcopenic muscles are characterized by increased heterogeneity in muscle fiber size, increases ROS production, attenuated stress response capacity, and increased intramuscular connective and adipose tissue called marbling[187]. During the passage from old to geriatric stages muscle stem cells (MuSCs, also called satellite cells) are subjected to a conversion in a pre-senescent state characterized by hallmarks of cellular senescence as $p16^{ink4a}$, and gene sets regulated by Polycomb complexes as Deleted In Esophageal Cancer 1 Dec1 (BHLHB2), Cyclin Dependent Kinase Inhibitor 1A,2B,1C (p21, p15, p57)[188]. Sarcopenia and senesce have a tight relation. In particular, SASP production parallelly increase with p38 mitogen-activated protein kinase (MAPK) and the transcription faction NF-κB p38MAPK/NF-κB[189]. Thus, SASP components impact the correct differentiation

of myoblast and increase their aptoptosis[190]. On the other hand, the elevated p38/MAPK concomitant with the JAK/STAT signaling hinders the asymmetric division, which lead to regenerative impairment and pool exhaustion[191,192]. In another study, a dysfunction of the Notch-p53 signaling pathway has been found in aged muscle stem cells; this signalling impairment causes what it has been called a mitotic catastrophe, which affects the regenerative and self-renewal potential of the cells[193].

Also the correct regulation of lamin A/C is fundamental for a proper muscle maintenance and regeneration[194]. In the mouse model of Emery Dreifuss Muscular Dystrophy (EDMD), lack of lamin A leads to a redistribution of the H3K27me3 repressive Polycomb chromatin mark on non-target genes. This unbalance redistribution de-represses other genes as the cell cycle inhibitors ($p16^{ink4a}$), determining loss of cell identity and the appearance of senescent traits[195]. Genetic ablation of Cdkn2a locus, coding for $p16^{ink4a}$, reactivating the cell cycle, ameliorates dystrophic and cardiac phenotype[195,196]. Changes the PRC2–LaminA/C interplay have been described also in human cardiomyocytes derived from iPSC carrying a mutation in the LMNA gene (K219T). Lamin A dependent alteration of PRC2 positioning on chromatin determine an increase of transcriptional repression on a sodium channel gene that result in action potential changes affecting the contractility of the heart[197]. Premature muscle senescence has been described also in other muscular dystrophy, as in the case of Duchenne Muscular Dystrophy (DMD). Here the permanent state of inflammation dramatically increases the senescence levels of muscles which contribute to progression and exacerbation of the DMD symptoms[198]. Treatments with a senolytic drug can inhibit the progression of the disease in a rat model for DMD[198].

1.14 Scope of the thesis

In this work, we aim to investigate the role of the nuclear lamins on chromatin conformation dynamics in health and disease. We set up a new and improved version of the SAMMY-seq technology to analyze both euchromatin and heterochromatin.

We used this technology to dissect the chromatin architecture in distinct cell types and to highlight new molecular pathways involved in cellular senescence. Finally, we highlight differences in pathological aging conditions.

The thesis is divided into the following chapters:

In Chapter 2 we illustrate the novel SAMMY-seq technology.

In Chapter 3 we describe the different impacts of lamin A and B Knock Out (KO) on chromatin solubility. We use cultured fibroblasts with KO for the lamin A/C gene or lamin B1 and B2 genes. Here we present data analysis on the chromatin dynamics due to the lack of lamins in a controlled environment.

In Chapter 4 we analyze the impact of progerin on chromatin in 3 mesenchymal derived cell types: skin fibroblasts, endothelial cells and muscle satellite cells.

In Chapter 5 we present our data on muscular ageing. Using a cohort of mice representative of postnatal, adult, old and geriatric stages of lifespan, we deeply analyzed the chromatin dysregulation in muscular stem cells.

1.15 References

1. Sul, J.-Y. *et al.* Transcriptome transfer produces a predictable cellular phenotype. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7624–7629 (2009).
2. Misteli, T. The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* **183**, 28–45 (2020).

3. Raj, A. & van Oudenaarden, A. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **135**, 216–226 (2008).
4. Battich, N., Stoeger, T. & Pelkmans, L. Control of Transcript Variability in Single Mammalian Cells. *Cell* **163**, 1596–1610 (2015).
5. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
6. Onufriev, A. V. & Schiessel, H. The nucleosome: from structure to function through physics. *Current Opinion in Structural Biology* **56**, 119–130 (2019).
7. Millán-Zambrano, G., Burton, A., Bannister, A. J. & Schneider, R. Histone post-translational modifications — cause and consequence of genome function. *Nat Rev Genet* **23**, 563–580 (2022).
8. Montavon, T. *et al.* Complete loss of H3K9 methylation dissolves mouse heterochromatin organization. *Nat Commun* **12**, 4359 (2021).
9. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338–1345 (2019).
10. Hansen, A. S. CTCF as a boundary factor for cohesin-mediated loop extrusion: evidence for a multi-step mechanism. *Nucleus* **11**, 132–148 (2020).
11. Piazza, A. *et al.* Cohesin regulates homology search during recombinational DNA repair. *Nat Cell Biol* **23**, 1176–1186 (2021).
12. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U.S.A.* **112**, (2015).
13. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).
14. Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).
15. Nanni, L., Ceri, S. & Logie, C. Spatial patterns of CTCF sites define the anatomy of TADs and their boundaries. *Genome Biol* **21**, 197 (2020).

16. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
17. Teng, Z. *et al.* Deciphering the chromatin spatial organization landscapes during BMMSC differentiation. *Journal of Genetics and Genomics* **50**, 264–275 (2023).
18. Winick-Ng, W. *et al.* Cell-type specialization is encoded by specific chromatin topologies. *Nature* **599**, 684–691 (2021).
19. Smith, E. M., Lajoie, B. R., Jain, G. & Dekker, J. Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus. *The American Journal of Human Genetics* **98**, 185–201 (2016).
20. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2**, 292–301 (2001).
21. Tavares-Cadete, F., Norouzi, D., Dekker, B., Liu, Y. & Dekker, J. Multi-contact 3C reveals that the human genome during interphase is largely not entangled. *Nat Struct Mol Biol* **27**, 1105–1114 (2020).
22. Shan, C., Fang, Y. & Jia, S. Leaving histone unturned for epigenetic inheritance. *The FEBS Journal* **290**, 310–320 (2023).
23. Machado, L. *et al.* In Situ Fixation Redefines Quiescence and Early Activation of Skeletal Muscle Stem Cells. *Cell Rep* **21**, 1982–1993 (2017).
24. Kang, H. *et al.* Dynamic regulation of histone modifications and long-range chromosomal interactions during postmitotic transcriptional reactivation. *Genes Dev.* **34**, 913–930 (2020).
25. Gabriele, M. *et al.* Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* **376**, 496–501 (2022).
26. Mach, P. *et al.* Cohesin and CTCF control the dynamics of chromosome folding. *Nat Genet* **54**, 1907–1918 (2022).
27. Turgay, Y. *et al.* The molecular architecture of lamins in somatic cells. *Nature* **543**, 261–264 (2017).
28. Dittmer, T. A. & Misteli, T. The lamin protein family. *Genome Biol* **12**, 222 (2011).

29. DeBoy, E. *et al.* Identification of novel RNA isoforms of *LMNA*. *Nucleus* **8**, 573–582 (2017).

30. Marullo, F. *et al.* Nucleoplasmic Lamin A/C and Polycomb group of proteins: An evolutionarily conserved interplay. *Nucleus* **7**, 103–111 (2016).

31. Collas, P., Lund, E. G. & Oldenburg, A. R. Closing the (nuclear) envelope on the genome: How nuclear lamins interact with promoters and modulate gene expression: Prospects & Overviews. *BioEssays* **36**, 75–83 (2014).

32. Cesarini, E. *et al.* Lamin A/C sustains PcG protein architecture, maintaining transcriptional repression at target genes. *Journal of Cell Biology* **211**, 533–551 (2015).

33. Swift, J. *et al.* Nuclear Lamin-A Scales with Tissue Stiffness and Enhances Matrix-Directed Differentiation. *Science* **341**, 1240104 (2013).

34. Kolb, T., Maaß, K., Hergt, M., Aebi, U. & Herrmann, H. Lamin A and lamin C form homodimers and coexist in higher complex forms both in the nucleoplasmic fraction and in the lamina of cultured human cells. *Nucleus* **2**, 425–433 (2011).

35. Hozák, P., Marie-Josée Sasseville, A., Raymond, Y. & Cook, P. R. Lamin proteins form an internal nucleoskeleton as well as a peripheral lamina in human cells. *Journal of Cell Science* **108**, 635–644 (1995).

36. Shimi, T. *et al.* Structural organization of nuclear lamins A, C, B1, and B2 revealed by superresolution microscopy. *MBoC* **26**, 4075–4086 (2015).

37. Tatli, M. & Medalia, O. Insight into the functional organization of nuclear lamins in health and disease. *Current Opinion in Cell Biology* **54**, 72–79 (2018).

38. Raman, N., Imran, S. A. M., Ahmad Amin Noordin, K. B., Zaman, W. S. W. K. & Nordin, F. Mechanotransduction in Mesenchymal Stem Cells (MSCs) Differentiation: A Review. *IJMS* **23**, 4580 (2022).

39. Stewart, C. L., Roux, K. J. & Burke, B. Blurring the Boundary: The Nuclear Envelope Extends Its Reach. *Science* **318**, 1408–1412 (2007).

40. Maniotis, A. J., Chen, C. S. & Ingber, D. E. Demonstration of mechanical connections between integrins, cytoskeletal filaments, and nucleoplasm that stabilize nuclear structure. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 849–854 (1997).

41. Li, E. W., McKee-Muir, O. C. & Gilbert, P. M. Cellular Biomechanics in Skeletal Muscle Regeneration. in *Current Topics in Developmental Biology* vol. 126 125–176 (Elsevier, 2018).
42. Cruz, V. E., Esra Demircioglu, F. & Schwartz, T. U. Structural Analysis of Different LINC Complexes Reveals Distinct Binding Modes. *Journal of Molecular Biology* **432**, 6028–6041 (2020).
43. de Leeuw, R., Gruenbaum, Y. & Medalia, O. Nuclear Lamins: Thin Filaments with Major Functions. *Trends in Cell Biology* **28**, 34–45 (2018).
44. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
45. Wu, F. & Yao, J. Identifying Novel Transcriptional and Epigenetic Features of Nuclear Lamina-associated Genes. *Sci Rep* **7**, 100 (2017).
46. Girelli, G. *et al.* GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nat Biotechnol* **38**, 1184–1193 (2020).
47. van Steensel, B. & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780–791 (2017).
48. Briand, N. & Collas, P. Lamina-associated domains: peripheral matters and internal affairs. *Genome Biol* **21**, 85 (2020).
49. Demmerle, J., Koch, A. J. & Holaska, J. M. Emerin and histone deacetylase 3 (HDAC3) cooperatively regulate expression and nuclear positions of MyoD, Myf5, and Pax7 genes during myogenesis. *Chromosome Res* **21**, 765–779 (2013).
50. Meuleman, W. *et al.* Constitutive nuclear lamina–genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* **23**, 270–280 (2013).
51. Kind, J. *et al.* Single-Cell Dynamics of Genome-Nuclear Lamina Interactions. *Cell* **153**, 178–192 (2013).
52. van Schaik, T., Vos, M., Peric-Hupkes, D., HN Celie, P. & van Steensel, B. Cell cycle dynamics of lamina-associated DNA. *EMBO Reports* **21**, e50636 (2020).

53. Madsen-Østerbye, J., Abdelhalim, M., Baudement, M.-O. & Collas, P. Local euchromatin enrichment in lamina-associated domains anticipates their repositioning in the adipogenic lineage. *Genome Biol* **23**, 91 (2022).

54. Sebestyén, E. *et al.* SAMMY-seq reveals early alteration of heterochromatin and deregulation of bivalent genes in Hutchinson-Gilford Progeria Syndrome. *Nat Commun* **11**, 6274 (2020).

55. Le, H. Q. *et al.* Mechanical regulation of transcription controls Polycomb-mediated gene silencing during lineage commitment. *Nat Cell Biol* **18**, 864–875 (2016).

56. Madsen-Østerbye, J., Abdelhalim, M., Pickering, S. H. & Collas, P. Gene Regulatory Interactions at Lamina-Associated Domains. *Genes* **14**, 334 (2023).

57. Paulsen, J. *et al.* Long-range interactions between topologically associating domains shape the four-dimensional genome during differentiation. *Nat Genet* **51**, 835–843 (2019).

58. Shin, J.-Y. & Worman, H. J. Molecular Pathology of Laminopathies. *Annu. Rev. Pathol. Mech. Dis.* **17**, 159–180 (2022).

59. Evangelisti, C. *et al.* The wide and growing range of lamin B-related diseases: from laminopathies to cancer. *Cell. Mol. Life Sci.* **79**, 126 (2022).

60. Swahari, V. & Nakamura, A. Speeding up the clock: The past, present and future of progeria. *Develop. Growth Differ.* **58**, 116–130 (2016).

61. De Sandre-Giovannoli, A. *et al.* Lamin A Truncation in Hutchinson-Gilford Progeria. *Science* **300**, 2055–2055 (2003).

62. Glynn, M. W. & Glover, T. W. Incomplete processing of mutant lamin A in Hutchinson–Gilford progeria leads to nuclear abnormalities, which are reversed by farnesyltransferase inhibition. *Human Molecular Genetics* **14**, 2959–2969 (2005).

63. Funkhouser, C. M. *et al.* Mechanical model of blebbing in nuclear lamin meshworks. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 3248–3253 (2013).

64. Kittisopikul, M. *et al.* Computational analyses reveal spatial relationships between nuclear pore complexes and specific lamins. *Journal of Cell Biology* **220**, e202007082 (2021).

65. Pascual-Garcia, P. & Capelson, M. Nuclear pores in genome architecture and enhancer function. *Current Opinion in Cell Biology* **58**, 126–133 (2019).

66. Gozalo, A. *et al.* Core Components of the Nuclear Pore Bind Distinct States of Chromatin and Contribute to Polycomb Repression. *Molecular Cell* **77**, 67-81.e7 (2020).

67. Chen, C.-Y. *et al.* Accumulation of the Inner Nuclear Envelope Protein Sun1 Is Pathogenic in Progeric and Dystrophic Laminopathies. *Cell* **149**, 565–577 (2012).

68. Chai, R. J. *et al.* Disrupting the LINC complex by AAV mediated gene transduction prevents progression of Lamin induced cardiomyopathy. *Nat Commun* **12**, 4722 (2021).

69. Shimi, T. *et al.* The A- and B-type nuclear lamin networks: microdomains involved in chromatin organization and transcription. *Genes Dev.* **22**, 3409–3421 (2008).

70. Agustinus, A. S. *et al.* Epigenetic dysregulation from chromosomal transit in micronuclei. *Nature* **619**, 176–183 (2023).

71. Papathanasiou, S. *et al.* Heritable transcriptional defects from aberrations of nuclear architecture. *Nature* **619**, 184–192 (2023).

72. Trivedi, P., Steele, C. D., Au, F. K. C., Alexandrov, L. B. & Cleveland, D. W. Mitotic tethering enables inheritance of shattered micronuclear chromosomes. *Nature* **618**, 1049–1056 (2023).

73. Stephens, A. D. *et al.* Chromatin histone modifications and rigidity affect nuclear morphology independent of lamins. *MBoC* **29**, 220–233 (2018).

74. Danielsson, B. E. *et al.* Lamin microaggregates lead to altered mechanotransmission in progerin-expressing cells. *Nucleus* **11**, 194–204 (2020).

75. Dahl, K. N. *et al.* Distinct structural and mechanical properties of the nuclear lamina in Hutchinson–Gilford progeria syndrome. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10271–10276 (2006).

76. Verstraeten, V. L. R. M., Ji, J. Y., Cummings, K. S., Lee, R. T. & Lammerding, J. Increased mechanosensitivity and nuclear stiffness in Hutchinson–Gilford progeria cells: effects of farnesyltransferase inhibitors. *Aging Cell* **7**, 383–393 (2008).

77. Zhang, J. *et al.* A Human iPSC Model of Hutchinson Gilford Progeria Reveals Vascular Smooth Muscle and Mesenchymal Stem Cell Defects. *Cell Stem Cell* **8**, 31–45 (2011).

78. Chiarini, F. *et al.* The Cutting Edge: The Role of mTOR Signaling in Laminopathies. *IJMS* **20**, 847 (2019).

79. Cenni, V. *et al.* Lamin A involvement in ageing processes. *Ageing Research Reviews* **62**, 101073 (2020).

80. Cabral, W. A. *et al.* Genetic reduction of mTOR extends lifespan in a mouse model of Hutchinson-Gilford Progeria syndrome. *Aging Cell* **20**, e13457 (2021).

81. Cisneros, B., García-Aguirre, I., De Ita, M., Arrieta-Cruz, I. & Rosas-Vargas, H. Hutchinson-Gilford Progeria Syndrome: Cellular Mechanisms and Therapeutic Perspectives. *Archives of Medical Research* **54**, 102837 (2023).

82. Miwa, S., Kashyap, S., Chini, E. & von Zglinicki, T. Mitochondrial dysfunction in cell senescence and aging. *Journal of Clinical Investigation* **132**, e158447 (2022).

83. Kubben, N. *et al.* Repression of the Antioxidant NRF2 Pathway in Premature Aging. *Cell* **165**, 1361–1374 (2016).

84. Wu, Q.-J. *et al.* The sirtuin family in health and disease. *Sig Transduct Target Ther* **7**, 402 (2022).

85. Liu, B. *et al.* Resveratrol Rescues SIRT1-Dependent Adult Stem Cell Decline and Alleviates Progeroid Features in Laminopathy-Based Progeria. *Cell Metabolism* **16**, 738–750 (2012).

86. Shumaker, D. K. *et al.* The highly conserved nuclear lamin Ig-fold binds to PCNA: its role in DNA replication. *The Journal of Cell Biology* **181**, 269–280 (2008).

87. Maynard, S. *et al.* Lamin A/C promotes DNA base excision repair. *Nucleic Acids Research* gkz912 (2019) doi:10.1093/nar/gkz912.

88. Butin-Israeli, V. *et al.* Role of Lamin B1 in Chromatin Instability. *Molecular and Cellular Biology* **35**, 884–898 (2015).

89. Stierlé, V. *et al.* The Carboxyl-Terminal Region Common to Lamins A and C Contains a DNA Binding Domain. *Biochemistry* **42**, 4819–4828 (2003).

90. Mahen, R. *et al.* A-Type Lamins Maintain the Positional Stability of DNA Damage Repair Foci in Mammalian Nuclei. *PLoS ONE* **8**, e61893 (2013).

91. Cossec, J.-C. *et al.* SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States. *Cell Stem Cell* **23**, 742-757.e8 (2018).

92. de la Rosa, J. *et al.* Prelamin A causes progeria through cell-extrinsic mechanisms and prevents cancer invasion. *Nat Commun* **4**, 2268 (2013).

93. Fernandez, P. *et al.* Transformation Resistance in a Premature Aging Disorder Identifies a Tumor-Protective Function of BRD4. *Cell Reports* **9**, 248–260 (2014).

94. Shumaker, D. K. *et al.* Mutant nuclear lamin A leads to progressive alterations of epigenetic control in premature aging. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 8703–8708 (2006).

95. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**, 897–903 (2008).

96. McCord, R. P. *et al.* Correlated alterations in genome organization, histone methylation, and DNA–lamin A/C interactions in Hutchinson-Gilford progeria syndrome. *Genome Res.* **23**, 260–269 (2013).

97. Millard, C. J. *et al.* The structure of the core NuRD repression complex provides insights into its interaction with chromatin. *eLife* **5**, e13941 (2016).

98. Pegoraro, G. *et al.* Ageing-related chromatin defects through loss of the NURD complex. *Nat Cell Biol* **11**, 1261–1267 (2009).

99. Kelley, J. B. *et al.* The Defective Nuclear Lamina in Hutchinson-Gilford Progeria Syndrome Disrupts the Nucleocytoplasmic Ran Gradient and Inhibits Nuclear Localization of Ubc9. *Molecular and Cellular Biology* **31**, 3378–3395 (2011).

100. Woodhouse, S., Pugazhendhi, D., Brien, P. & Pell, J. M. Ezh2 maintains a key phase of muscle satellite cell expansion but does not regulate terminal differentiation. *Journal of Cell Science* **126**, 565–579 (2013).

101. Li, R. *et al.* EZH2 inhibits autophagic cell death of aortic vascular smooth muscle cells to affect aortic dissection. *Cell Death Dis* **9**, 180 (2018).

102. Cohen, I. *et al.* PRC1 Fine-tunes Gene Repression and Activation to Safeguard Skin Development and Stem Cell Specification. *Cell Stem Cell* **22**, 726-739.e7 (2018).

103. Cenni, V. *et al.* Rapamycin treatment of Mandibuloacral Dysplasia cells rescues localization of chromatin-associated proteins and cell cycle dynamics. *Aging* **6**, 755–769 (2014).

104. Capanni, C. *et al.* Altered pre-lamin A processing is a common mechanism leading to lipodystrophy. *Human Molecular Genetics* **14**, 1489–1502 (2005).

105. Benedicto, I., Chen, X., Bergo, M. O. & Andrés, V. Progeria: a perspective on potential drug targets and treatment strategies. *Expert Opinion on Therapeutic Targets* **26**, 393–399 (2022).

106. Saidak, Z. *et al.* Wnt/β-Catenin Signaling Mediates Osteoblast Differentiation Triggered by Peptide-induced α5β1 Integrin Priming in Mesenchymal Skeletal Cells. *Journal of Biological Chemistry* **290**, 6903–6912 (2015).

107. Choi, J. Y. *et al.* Diminished Canonical β-Catenin Signaling During Osteoblast Differentiation Contributes to Osteopenia in Progeria. *J Bone Miner Res* **33**, 2059–2070 (2018).

108. Infante, A. & Rodríguez, C. I. Secretome analysis of in vitro aged human mesenchymal stem cells reveals IGFBP7 as a putative factor for promoting osteogenesis. *Sci Rep* **8**, 4632 (2018).

109. Mateos, J. *et al.* Lamin A deregulation in human mesenchymal stem cells promotes an impairment in their chondrogenic potential and imbalance in their response to oxidative stress. *Stem Cell Research* **11**, 1137–1148 (2013).

110. Lemire, J. M. *et al.* Aggrecan expression is substantially and abnormally upregulated in Hutchinson–Gilford Progeria Syndrome dermal fibroblasts. *Mechanisms of Ageing and Development* **127**, 660–669 (2006).

111. Venkei, Z. G. & Yamashita, Y. M. Emerging mechanisms of asymmetric stem cell division. *Journal of Cell Biology* **217**, 3785–3795 (2018).

112. Bohl, B., Jabali, A., Ladewig, J. & Koch, P. Asymmetric Notch activity by differential inheritance of lysosomes in human neural stem cells. *Sci. Adv.* **8**, eabl5792 (2022).

113. Sola-Carvajal, A. *et al.* Accumulation of Progerin Affects the Symmetry of Cell Division and Is Associated with Impaired Wnt Signaling and the Mislocalization of Nuclear Envelope Proteins. *Journal of Investigative Dermatology* **139**, 2272-2280.e12 (2019).

114. Maier, J. A. *et al.* Aging and Vascular Disease: A Multidisciplinary Overview. *JCM* **12**, 5512 (2023).

115. Osmanagic-Myers, S. *et al.* Endothelial progerin expression causes cardiovascular pathology through an impaired mechanoresponse. *Journal of Clinical Investigation* **129**, 531–545 (2018).

116. Gao, F. *et al.* Reduction of Endothelial Nitric Oxide Increases the Adhesiveness of Constitutive Endothelial Membrane ICAM-1 through Src-Mediated Phosphorylation. *Front. Physiol.* **8**, 1124 (2018).

117. Sánchez-López, A. *et al.* Cardiovascular Progerin Suppression and Lamin A Restoration Rescue Hutchinson-Gilford Progeria Syndrome. *Circulation* **144**, 1777–1794 (2021).

118. Rivera-Torres, J. *et al.* Cardiac electrical defects in progeroid mice and Hutchinson–Gilford progeria syndrome patients with nuclear lamina alterations. *Proc. Natl. Acad. Sci. U.S.A.* **113**, (2016).

119. Liu, G.-H. *et al.* Recapitulation of premature ageing with iPSCs from Hutchinson–Gilford progeria syndrome. *Nature* **472**, 221–225 (2011).

120. Monnerat, G. *et al.* Modelling premature cardiac aging with induced pluripotent stem cells from a hutchinson-gilford Progeria Syndrome patient. *Front. Physiol.* **13**, 1007418 (2022).

121. De Majo, F. *et al.* Genomic instability in the naturally and prematurely aged myocardium. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2022974118 (2021).

122. Osorio, F. G. *et al.* Splicing-Directed Therapy in a New Mouse Model of Human Accelerated Aging. *Sci. Transl. Med.* **3**, (2011).

123. Fanjul, V. *et al.* Identification of common cardiometabolic alterations and deregulated pathways in mouse and pig models of aging. *Aging Cell* **19**, e13203 (2020).

124. Hartinger, R., Lederer, E.-M., Schena, E., Lattanzi, G. & Djabali, K. Impact of Combined Baricitinib and FTI Treatment on Adipogenesis in Hutchinson–Gilford Progeria Syndrome and Other Lipodystrophic Laminopathies. *Cells* **12**, 1350 (2023).

125. Siersbæk, R. *et al.* Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte Differentiation. *Molecular Cell* **66**, 420-435.e5 (2017).

126. Charó, N. L., Rodríguez Ceschan, M. I., Galigniana, N. M., Toneatto, J. & Piwien-Pilipuk, G. Organization of nuclear architecture during adipocyte differentiation. *Nucleus* **7**, 249–269 (2016).

127. Barinda, A. J. *et al.* Endothelial progeria induces adipose tissue senescence and impairs insulin sensitivity through senescence associated secretory phenotype. *Nat Commun* **11**, 481 (2020).

128. Martins, F., Sousa, J., Pereira, C. D., Da Cruz E Silva, O. A. B. & Rebelo, S. Nuclear envelope dysfunction and its contribution to the aging process. *Aging Cell* **19**, e13143 (2020).

129. Willaume, S. *et al.* A Link between Replicative Stress, Lamin Proteins, and Inflammation. *Genes* **12**, 552 (2021).

130. Graziano, S. *et al.* Lamin A/C recruits ssDNA protective proteins RPA and RAD51 to stalled replication forks to maintain fork stability. *Journal of Biological Chemistry* **297**, 101301 (2021).

131. Merigliano, C. & Chiolo, I. Multi-scale dynamics of heterochromatin repair. *Current Opinion in Genetics & Development* **71**, 206–215 (2021).

132. Simon, D. N., Zastrow, M. S. & Wilson, K. L. Direct actin binding to A- and B-type lamin tails and actin filament bundling by the lamin A tail. *Nucleus* **1**, 264–272 (2010).

133. Komari, C. J. *et al.* Alteration of genetic recombination and double-strand break repair in human cells by progerin expression. *DNA Repair* **96**, 102975 (2020).

134. Liu, Y. *et al.* Involvement of xeroderma pigmentosum group A (XPA) in progeria arising from defective maturation of prelamin A. *FASEB j.* **22**, 603–611 (2008).

135. Gibbs-Seymour, I., Markiewicz, E., Bekker-Jensen, S., Mailand, N. & Hutchison, C. J. Lamin A/C-dependent interaction with 53BP1 promotes cellular responses to DNA damage. *Aging Cell* **14**, 162–169 (2015).

136. Schoelz, J. M. & Riddle, N. C. Functions of HP1 proteins in transcriptional regulation. *Epigenetics & Chromatin* **15**, 14 (2022).

137. Liu, J. *et al.* HP1α mediates defective heterochromatin repair and accelerates senescence in *Zmpste24* -deficient cells. *Cell Cycle* **13**, 1237–1247 (2014).

138. Batista, N. J. *et al.* The Molecular and Cellular Basis of Hutchinson–Gilford Progeria Syndrome and Potential Treatments. *Genes* **14**, 602 (2023).

139. Lim, C. J. & Cech, T. R. Shaping human telomeres: from shelterin and CST complexes to telomeric chromatin organization. *Nat Rev Mol Cell Biol* **22**, 283–298 (2021).

140. Sobecki, M. *et al.* MadID, a Versatile Approach to Map Protein-DNA Interactions, Highlights Telomere-Nuclear Envelope Contact Sites in Human Cells. *Cell Reports* **25**, 2891-2903.e5 (2018).

141. Wood, A. M. *et al.* TRF2 and lamin A/C interact to facilitate the functional organization of chromosome ends. *Nat Commun* **5**, 5467 (2014).

142. Huang, S., Risques, R. A., Martin, G. M., Rabinovitch, P. S. & Oshima, J. Accelerated telomere shortening and replicative senescence in human

fibroblasts overexpressing mutant and wild-type lamin A. *Experimental Cell Research* **314**, 82–91 (2008).

143. Decker, M. L., Chavez, E., Vulto, I. & Lansdorp, P. M. Telomere length in Hutchinson-Gilford Progeria Syndrome. *Mechanisms of Ageing and Development* **130**, 377–383 (2009).

144. Chojnowski, A. *et al.* Progerin reduces LAP2α-telomere association in Hutchinson-Gilford progeria. *eLife* **4**, e07759 (2015).

145. Dreesen, O. Towards delineating the chain of events that cause premature senescence in the accelerated aging syndrome Hutchinson–Gilford progeria (HGPS). *Biochemical Society Transactions* **48**, 981–991 (2020).

146. Fumagalli, M. *et al.* Telomeric DNA damage is irreparable and causes persistent DNA-damage-response activation. *Nat Cell Biol* **14**, 355–365 (2012).

147. Francia, S., Cabrini, M., Matti, V., Oldani, A. & d'Adda Di Fagagna, F. DICER, DROSHA and DNA damage-response RNAs are necessary for the secondary recruitment of DNA damage response factors. *Journal of Cell Science* jcs.182188 (2016) doi:10.1242/jcs.182188.

148. Aguado, J. *et al.* Inhibition of DNA damage response at telomeres improves the detrimental phenotypes of Hutchinson–Gilford Progeria Syndrome. *Nat Commun* **10**, 4990 (2019).

149. Di Micco, R., Krizhanovsky, V., Baker, D. & d'Adda di Fagagna, F. Cellular senescence in ageing: from mechanisms to therapeutic opportunities. *Nat Rev Mol Cell Biol* **22**, 75–95 (2021).

150. Gorgoulis, V. *et al.* Cellular Senescence: Defining a Path Forward. *Cell* **179**, 813–827 (2019).

151. Ritschka, B. *et al.* The senescence-associated secretory phenotype induces cellular plasticity and tissue regeneration. *Genes Dev.* **31**, 172–183 (2017).

152. Mosteiro, L., Pantoja, C., de Martino, A. & Serrano, M. Senescence promotes in vivo reprogramming through p16 $^{INK\ 4a}$ and IL-6. *Aging Cell* **17**, e12711 (2018).

153. Bielak-Zmijewska, A., Mosieniak, G. & Sikora, E. Is DNA damage indispensable for stress-induced senescence? *Mechanisms of Ageing and Development* **170**, 13–21 (2018).

154. Lozano-Torres, B. *et al.* The chemistry of senescence. *Nat Rev Chem* **3**, 426–441 (2019).

155. Hayflick, L. & Moorhead, P. S. The serial cultivation of human diploid cell strains. *Experimental Cell Research* **25**, 585–621 (1961).

156. Ge, M.-X. *et al.* Multiple time-series expression trajectories imply dynamic functional changes during cellular senescence. *Computational and Structural Biotechnology Journal* **20**, 4131–4137 (2022).

157. Huang, W., Hickson, L. J., Eirin, A., Kirkland, J. L. & Lerman, L. O. Cellular senescence: the good, the bad and the unknown. *Nat Rev Nephrol* **18**, 611–627 (2022).

158. Martien, S. *et al.* Cellular senescence involves an intracrine prostaglandin E2 pathway in human fibroblasts. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1831**, 1217–1227 (2013).

159. Hsieh, H.-H., Chen, Y.-C., Jhan, J.-R. & Lin, J.-J. Serine protease inhibitor SerpinB2 binds and stabilizes p21 in senescent cells. *Journal of Cell Science* jcs.204974 (2017) doi:10.1242/jcs.204974.

160. Ohtani, N. The roles and mechanisms of senescence-associated secretory phenotype (SASP): can it be controlled by senolysis? *Inflamm Regener* **42**, 11 (2022).

161. Basisty, N. *et al.* A proteomic atlas of senescence-associated secretomes for aging biomarker development. *PLoS Biol* **18**, e3000599 (2020).

162. Özcan, S. *et al.* Unbiased analysis of senescence associated secretory phenotype (SASP) to identify common components following different genotoxic stresses. *Aging* **8**, 1316–1329 (2016).

163. Rovillain, E., Mansfield, L., Lord, C. J., Ashworth, A. & Jat, P. S. An RNA interference screen for identifying downstream effectors of the p53 and pRB tumour suppressor pathways involved in senescence. *BMC Genomics* **12**, 355 (2011).

164. Beausejour, C. M. Reversal of human cellular senescence: roles of the p53 and p16 pathways. *The EMBO Journal* **22**, 4212–4222 (2003).

165. Afifi, M. M. *et al.* Irreversible cell cycle exit associated with senescence is mediated by constitutive MYC degradation. *Cell Reports* **42**, 113079 (2023).

166. Cosgrove, B. D. *et al.* Rejuvenation of the muscle stem cell population restores strength to injured aged muscles. *Nat Med* **20**, 255–264 (2014).

167. De Blander, H., Morel, A.-P., Senaratne, A. P., Ouzounova, M. & Puisieux, A. Cellular Plasticity: A Route to Senescence Exit and Tumorigenesis. *Cancers* **13**, 4561 (2021).

168. Milanovic, M. *et al.* Senescence-associated reprogramming promotes cancer stemness. *Nature* **553**, 96–100 (2018).

169. Yu, Y. *et al.* Targeting the Senescence-Overriding Cooperative Activity of Structurally Unrelated H3K9 Demethylases in Melanoma. *Cancer Cell* **33**, 322-336.e8 (2018).

170. Hernandez-Segura, A. *et al.* Unmasking Transcriptional Heterogeneity in Senescent Cells. *Current Biology* **27**, 2652-2660.e4 (2017).

171. Zhu, X. *et al.* Inflammation, epigenetics, and metabolism converge to cell senescence and ageing: the regulation and intervention. *Sig Transduct Target Ther* **6**, 245 (2021).

172. Hemann, M. T., Strong, M. A., Hao, L.-Y. & Greider, C. W. The Shortest Telomere, Not Average Telomere Length, Is Critical for Cell Viability and Chromosome Stability. *Cell* **107**, 67–77 (2001).

173. Fagagna, F. d'Adda D. *et al.* A DNA damage checkpoint response in telomere-initiated senescence. *Nature* **426**, 194–198 (2003).

174. Chandra, T. *et al.* Independence of Repressive Histone Marks and Chromatin Compaction during Senescent Heterochromatic Layer Formation. *Molecular Cell* **47**, 203–214 (2012).

175. Zhang, R., Chen, W. & Adams, P. D. Molecular Dissection of Formation of Senescence-Associated Heterochromatin Foci. *Molecular and Cellular Biology* **27**, 2343–2358 (2007).

176. Sati, S. *et al.* 4D Genome Rewiring during Oncogene-Induced and Replicative Senescence. *Molecular Cell* **78**, 522-538.e9 (2020).

177. Di Micco, R. *et al.* Interplay between oncogene-induced DNA damage response and heterochromatin in senescence and cancer. *Nat Cell Biol* **13**, 292–302 (2011).

178. Freund, A., Laberge, R.-M., Demaria, M. & Campisi, J. Lamin B1 loss is a senescence-associated biomarker. *MBoC* **23**, 2066–2075 (2012).

179. Lee, J.-H., Kim, E. W., Croteau, D. L. & Bohr, V. A. Heterochromatin: an epigenetic point of view in aging. *Exp Mol Med* **52**, 1466–1474 (2020).

180. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: An expanding universe. *Cell* **186**, 243–278 (2023).

181. Franceschi, C. *et al.* Inflamm-aging: An Evolutionary Perspective on Immunosenescence. *Annals of the New York Academy of Sciences* **908**, 244–254 (2006).

182. Bruunsgaard, H., Skinhøj, P., Pedersen, A. N., Schroll, M. & Pedersen, B. K. Ageing, tumour necrosis factor-alpha (TNF- α) and atherosclerosis. *Clinical and Experimental Immunology* **121**, 255–260 (2008).

183. Walker, K. A., Basisty, N., Wilson, D. M. & Ferrucci, L. Connecting aging biology and inflammation in the omics era. *Journal of Clinical Investigation* **132**, e158448 (2022).

184. Mylonas, A. & O'Loghlen, A. Cellular Senescence and Ageing: Mechanisms and Interventions. *Front. Aging* **3**, 866718 (2022).

185. Schafer, M. J. *et al.* The senescence-associated secretome as an indicator of age and medical risk. *JCI Insight* **5**, e133668 (2020).

186. Cesari, M., Landi, F., Vellas, B., Bernabei, R. & Marzetti, E. Sarcopenia and Physical Frailty: Two Sides of the Same Coin. *Front. Aging Neurosci.* **6**, (2014).

187. Woo, J. Sarcopenia. *Clinics in Geriatric Medicine* **33**, 305–314 (2017).

188. Sousa-Victor, P., Perdiguero, E. & Muñoz-Cánoves, P. Geroconversion of aged muscle stem cells under regenerative pressure. *Cell Cycle* **13**, 3183–3190 (2014).

189. He, Y. *et al.* Cellular Senescence in Sarcopenia: Possible Mechanisms and Therapeutic Potential. *Front. Cell Dev. Biol.* **9**, 793088 (2022).

190. Moustogiannis, A. *et al.* The Effects of Muscle Cell Aging on Myogenesis. *IJMS* **22**, 3721 (2021).

191. Bernet, J. D. *et al.* p38 MAPK signaling underlies a cell-autonomous loss of stem cell self-renewal in skeletal muscle of aged mice. *Nat Med* **20**, 265–271 (2014).

192. Price, F. D. *et al.* Inhibition of JAK-STAT signaling stimulates adult satellite cell function. *Nat Med* **20**, 1174–1181 (2014).

193. Liu, L. *et al.* Impaired Notch Signaling Leads to a Decrease in p53 Activity and Mitotic Catastrophe in Aged Muscle Stem Cells. *Cell Stem Cell* **23**, 544-556.e4 (2018).

194. Sullivan, T. *et al.* Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy. *J Cell Biol* **147**, 913–920 (1999).

195. Bianchi, A. *et al.* Dysfunctional polycomb transcriptional repression contributes to lamin A/C–dependent muscular dystrophy. *Journal of Clinical Investigation* **130**, 2408–2421 (2020).

196. Pegoli, G. *et al.* Role of Cdkn2a in the Emery–Dreifuss Muscular Dystrophy Cardiac Phenotype. *Biomolecules* **11**, 538 (2021).

197. Salvarani, N. *et al.* The K219T-Lamin mutation induces conduction defects through epigenetic inhibition of SCN5A in human cardiac laminopathy. *Nat Commun* **10**, 2267 (2019).

198. Sugihara, H. *et al.* Cellular senescence-mediated exacerbation of Duchenne muscular dystrophy. *Sci Rep* **10**, 16385 (2020).

Chapter 2 Biochemical properties of chromatin domains define genome compartmentalization.

Federica Lucini[1,8,†], Cristiano Petrini[2,†], Elisa Salviato[2,†], Koustav Pal[2,†,] Valentina Rosti[1,3], Francesca Gorini[1], Philina Santarelli[1], Roberto Quadri[1], Emanuele Di Patrizio Soldateschi[1,3], Giovanni Lembo[2], Ilario Tagliaferri[2], Eva Pinatel[3], Endre Sebestyén[2], Luca Rotta[4], Francesco Gentile[5], Valentina Vaira[5,6], Chiara Lanzuolo[1,3,§,*], Francesco Ferrari[2,7,§,*]

[1] INGM, Istituto Nazionale di Genetica Molecolare "Romeo ed Enrica Invernizzi", Milan, 20122, Italy

[2] IFOM-ETS, The AIRC Institute of Molecular Oncology, Milan, 20139, Italy

[3] ITB-CNR, Institute of Biomedical Technologies, National Research Council, Segrate, 20054, Italy

[4] IEO, European Institute of Oncology IRCCS, Milan, 20141, Italy.

[5] Fondazione IRCCS Ca' Granda-Ospedale Maggiore Policlinico, Milan, 20122, Italy

[6] Department of Pathophysiology and Transplantation, University of Milan, Milan, 20122, Italy

[7] IGM-CNR, Institute of Molecular Genetics "Luigi Luca Cavalli-Sforza", National Research Council, Pavia, 27100, Italy

[8] Present Address: Federica Lucini, IFOM-ETS, The AIRC Institute of Molecular Oncology, Milan, 20139, Italy

† Joint first authors

§ These authors jointly supervised this work

* To whom correspondence should be addressed. Email: chiara.lanzuolo@cnr.it; francesco.ferrari@cnr.it

2.1 ABSTRACT

Chromatin three-dimensional (3D) organization inside the cell nucleus results in the separation of active and inactive chromatin compartments. These are driven by the local concentration of genomic domains with similar transcriptional activity and epigenetic features, as well as the local enrichment of their associated proteins leading to the formation of distinct subnuclear structures. The reference genomics techniques to map chromatin 3D architecture are based on the quantification of contact frequencies between genomic loci, that doesn't' detect chromatin compartments relocation across subnuclear regions. Chromatin domains spatially located in the same 3D nuclear neighbourhood are exposed to the same biochemical milieu, thus sharing similar solubility. Based on this rationale, we developed a new experimental technique, the 4fSAMMY-seq, paired with tailored bioinformatic data analysis methods to map chromatin accessibility, starting from the biochemical properties of genomic domains and using as little as 10,000 cells. We can extract information on the linear segmentation of euchromatic and heterochromatic genomic regions, as well as on their 3D segregation in active and inactive chromatin compartments. With 4fSAMMY-seq we can recapitulate in a single experiment the characteristic properties of the distinct chromatin states, including the highly dynamic Polycomb compartment.

2.2 INTRODUCTION

In the interphase cell nucleus, chromatin is organized in a three-dimensional (3D) architecture reflecting the epigenetic and transcriptional regulation of the genome (1). Alterations of chromatin 3D architecture have been identified in multiple human diseases (2-4). On a large-scale, euchromatic (active) and heterochromatic (inactive) domains tend to co-localize with domains of the same type, thus constituting spatially separated compartments that can be mapped by high-throughput genome-wide chromosome conformation capture (Hi-C) (5). The active and inactive compartments are conventionally named "A" and "B"

compartments, enriched in transcribed or repressed genomic regions, respectively. The Nuclear Lamina (NL) contributes to this compartment separation by directly binding specific heterochromatic genomic regions called Lamina Associated Domains (LADs) (6), belonging to the "B" compartment. Over the years, improvements in the resolution of experimental data (7) and in the computational analyses (8) allowed identifying also sub-compartments, whereby a finer grain segmentation of "A" and "B" compartments can be associated to specific combinations of chromatin marks, thus achieving a more precise link between the 3D chromatin organization and epigenetic regulation. Additional methodological improvements in this field focused on achieving higher resolution in mapping contacts between genomic loci, e.g. with Micro-C (9,10), or a reduced number of starting cells, e.g. with Low-C (11). Although undoubtedly powerful, Hi-C is based on a multistep protocol including chemical modifications and PCR amplification that can contribute technical biases masking or underestimating small chromatin remodelling dynamics. Moreover, by measuring the pairwise contact frequency of genomic loci, it does not give information on whether a specific contact is occurring in a particular subnuclear region. To this concern, multiple pieces of evidence in literature highlighted that the detachment of chromatin from lamina causes a major reorganization of chromatin 3D compartmentalization that is visible by imaging techniques, but does not result in evident changes in Hi-C compartments (12). In a different model with triple knock-out of lamina proteins, also the Topologically Associating Domains (TADs) detected by Hi-C have been shown to be largely unchanged (13). Indeed, as long as the local contact pairs are preserved, the chromosome conformation capture-based techniques may not be able to detect the change of location of a domain across subnuclear regions. Alternative methods to overcome some technical limitations of ligation by proximity were proposed with GAM (14) and SPRITE (15) techniques. Nevertheless, these approaches, that require highly specialized instruments and expertise, are also based on the detection of contact frequencies between genomic loci.

In this context, we recently presented SAMMY-seq, that is based instead on the biochemical fractionation of chromatin, followed by sequencing of the individual fractions to map the constitutive heterochromatin domains. We already reported that SAMMY-seq is able to detect early changes in heterochromatin solubility when LAD association to the nuclear lamina is compromised due to a mutated form of Lamin A (16). Notably, in the same early passage cellular model, Hi-C could not detect 3D chromatin architecture changes (17). More recently, in a completely different model of mechanical stress induced by cell motility, where the deformation of cell nucleus is expected to trigger a dissociation of heterochromatic LADs from the lamina (18), SAMMY-seq could detect reorganization dynamics involving heterochromatin domains (19).

Building on this earlier work, here we present a novel experimental protocol (4f-SAMMY-seq) and dedicated data analyses algorithms to map the position of both open and closed chromatin regions along the genome, in addition to their 3D spatial segregation in distinct chromatin compartments. This provides crucial practical advantages over state-of-the-art techniques for mapping compartments in terms of costs, versatility and scalability. Indeed, the novel 4f-SAMMY-seq works on as little as 10,000 (10k) cells, it requires only a few hours of bench work and a limited sequencing depth, as low as 25 million reads per chromatin fractions. These advantages open unprecedented possibilities for characterizing chromatin 3D compartmentalization and scaling up its analysis in a variety of experimental settings.

2.3 MATERIAL AND METHODS

Cell cultures

Human primary dermal fibroblast cell line C004 (foreskin fibroblast strain #2294, from a 4-year-old donor) was a generous gift from the Laboratory of Molecular and Cell Biology, Istituto Dermopatico dell'Immacolata (IDI)-IRCCS (Rome, Italy). Human primary dermal fibroblast cell lines C001 (foreskin fibroblast AG08498, from a 1-year-old donor) and C002 (foreskin fibroblast AG07095, from a 2-year-old donor) were obtained from the Coriell Institute.

All cell lines were cultured at 37 °C, 5% CO_2 in DMEM high glucose supplemented with GlutaMAX (Gibco, 10566-016), with further addition of 15% (v/v) FBS (Gibco, 10270106), 100 U/mL penicillin G and 100 µg/mL Streptomycin Sulphate.

Chromatin fractionation

For 4f-SAMMY-seq, 3 million fibroblasts were detached from the culture plate by 3 min incubation in Trypsin-EDTA solution at 37°C, 5% CO2. After two washes in cold PBS, the cells were resuspended in 600 µL of CSK-Triton buffer (10 mM PIPES pH 6.8, 100 mM NaCl, 1 mM EGTA, 300 mM Sucrose, 3 mM MgCl2, 1mM PMSF, 1 mM DTT, 0.5% Triton X-100, with protease inhibitors). After 10 min incubation on a wheel at 4°C, soluble proteins and the cytoskeletal structure were separated from the nuclei by centrifugation at 900g for 3 min at 4°C; the supernatant was labelled as S1 fraction. The pellet was then washed with an additional volume of CSK-Triton buffer, resuspended in 100 µL of CSK buffer (10 mM PIPES pH 6.8, 100 mM NaCl, 1 mM EGTA, 300 mM Sucrose, 3 mM MgCl2, 1mM PMSF, with protease inhibitors) and incubated for 60 min at 37°C with 25 U of RNase–free DNase I (Invitrogen, AM2222). To stop DNA digestion, ammonium sulphate was added in the CSK buffer to a final concentration of 250 mM. After 5 min incubation on ice, the sample was pelleted at 900g for 3 min at 4°C; the supernatant, containing digested chromatin fragments, was labelled as S2 fraction. Afterwards, the pellet was washed with 200 µL of CSK buffer and pelleted at 3000g for 3 min at 4°C, then resuspended in 100 µL of CSK-NaCl

buffer (CSK buffer with NaCl final concentration increased to 2 M) and incubated 10 min on a wheel at 4°C. At the end of the incubation, the sample was centrifuged at 2300g for 3 min at 4°C and the supernatant was labelled as S3 fraction. Finally, after two washes in 200 µL of CSK-NaCl buffer followed by centrifugation at 3000g for 3 min at 4°C, the pellet was solubilized in 100 µL of 8 M urea; the final suspension was labelled as S4 fraction.

For 50k and 10k 4f-SAMMY-seq, 50,000 and 10,000 fibroblasts were used, respectively. They were processed following the 4f-SAMMY-seq protocol described above but using 2U of RNase–free DNase I (Invitrogen, AM2222).

For 10kh-SAMMY-seq, 10,000 fibroblasts were processed following the 4f-SAMMY-seq protocol described above but halving the volumes in each step and using 12.5U of RNase–free DNase I (Invitrogen, AM2222); in this way we achieved a 150-fold increase in the ratio of enzyme units per starting number of cells, while maintaining the DNase concentration unaltered.

DNA sonication and sequencing

Fractions S2, S3 and S4 were diluted in TE (10 mM TrisHCl pH 7.5, 1mM EDTA pH 8.0) to a final volume of 200 μl and then incubated 90 min at 37°C with 6 µL of RNase cocktail (Ambion, AM2286), followed by 150 minutes at 55°C with Proteinase K (Invitrogen, AM2548) to a final concentration of 0.2 μg/μL. Next, DNA was isolated through phenol:chloroform:isoamyl alcohol (Sigma, 77617) extraction, precipitated in 70% ethanol, 0.3M sodium acetate and 20μg glycogen overnight at -20°C or 1 hour in dry ice and resuspended in nuclease-free water. S2 from 4f-SAMMY-seq was additionally purified using PCR DNA Purification Kit (Qiagen, 28106) and DNA fragments in this fraction were separated using AMPure XP paramagnetic beads (Beckman Coulter, A63880) to obtain S2S (< 300 bp) and S2L (> 300 bp) fractions. Beads were added to the S2 fraction in a 0.95x (v/v) ratio to bind fragments larger than 300bp. Magnetic separation of beads from supernatant allowed the physical separation of larger fragments (on the beads) from shorter ones (in the supernatant). Larger fragments bound on beads were then washed in 85% ethanol, resuspended in water and magnetically

separated from the beads (S2L fraction). Shorter fragments in the supernatant of the first step were bound to beads by adding a further 0.85x (v/v) beads ratio to the suspension; after washing in 85% ethanol and resuspension in water, they were also detached from beads (S2S fraction). Separation of S2S and S2L from S2 fraction of 10kh-SAMMY-seq S2 was also tested (C004_r1); since the enrichment profile of S2S and S2L was identical, this passage was later avoided (C002_r1, C004_r2). After DNA isolation, S2 (from 10kh-SAMMY-seq), S2L (from 4f-SAMMY-seq), S3 and S4 (from both 10kh- and 4f-SAMMY-seq) fractions were transferred to screw cap microTUBEs (Covaris, 004078) and sonicated in a Covaris M220 focused-ultrasonicator to obtain a smear of DNA fragments peaking at 200 bp (settings: water bath 20°C, peak power 30.0, duty factor 20.0, cycles/burst 50; duration: 125 sec for S2 and S2L, 175 sec for S3 and S4). DNA in the fractions was then quantified using Qubit dsDNA HS Assay Kit (Invitrogen, Q32854) and a Qubit 4.0 fluorometer; quality control was performed by run on an Agilent 2100 Bioanalyzer System using the High Sensitivity DNA Kit (Agilent, 5067-4626). Libraries were then created from each fraction using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645L) and the Unique Dual Index NEBNext Multiplex Oligos for Illumina (NEB, E6440S); final qualitative and quantitative controls were performed through an Agilent 2100 Bioanalyzer System and a Qubit 4.0 fluorometer. Libraries with distinct adapter indexes were multiplexed and, after cluster generation on FlowCell, sequenced for 50 bases in paired-end mode on an Illumina NovaSeq 6000 instrument at the IEO Genomic Unit in Milan or for 100 bases in single-end mode on an Illumina NextSeq 2000 instrument at Ospedale Policlinico in Milan. A sequencing depth of at least 24.9 million raw sequencing reads was obtained for each sample.

Protein analysis

Chromatin fractions were quantified using Qubit Protein Assay Kit (Invitrogen, Q33212) and a Qubit 4.0 fluorometer. Equal protein amounts from each fraction were run on 4-12% bis-tris plus acrylamide gels (Invitrogen, NW04122) and then immunoblotted. Anti-tubulin alpha (Sigma T5168, mouse 1:2000), anti-H3

(Abcam ab1791, rabbit 1:4000 or Invitrogen #MA3-049, mouse 1:1000), anti-Lamin A/C (Santa Cruz sc-6215, goat 1:1000) and anti-Lamin B (Santa Cruz sc-6216, goat 1:2000) were diluted in 5% (w/v) milk in PBST (0.1% Tween in PBS) and used as primary antibodies. Secondary HRP-conjugated anti-mouse (Sigma, A9044), anti-rabbit (Sigma, A9169) and anti-goat (Sigma, A5420) antibodies were then revealed through SuperSignal West Dura chemiluminescence kit (Thermo Scientific, 34076) and signals acquired in an iBright FL1500 Imaging System.

Public datasets

We used publicly available ChIP-seq, SAMMY-seq 3f and RNA-seq datasets from our previously published article on SAMMY-seq (GEO dataset GSE118633). We collected publicly available ChIP-seq and DNase-seq datasets from the following sources: Lamin A/C ChIP-seq from (17) (SRA datasets IDs SRR605493, SRR605494, SRR605495 and SRR605496), Lamin B1 ChIP-seq from (20), DNAse-seq from ENCODE dermal fibroblasts (ENCODE data file https://www.encodeproject.org/files/ENCFF261OJD/), H3K27ac, H3K36me3, H3K4me1, H3K4me3, H3K27me3 and H3K9me3 from Roadmap Epigenomics (foreskin fibroblast samples: E055 and E056 samples)(21). We also downloaded the chromatin states defined by Roadmap Epigenomics using the chromHMM 15 states model for the same samples, as well as the RNA-seq based gene expression profiles (RPKM). We downloaded the publicly available Hi-C dataset (dilution protocol, HFF-hTERT, HindIII) from the 4DN portal (22)(sample ID 4DNFIMDOXUT8: raw multiple-resolution contact matrix).

High-throughput sequencing reads alignment and filtering

High-throughput sequencing reads were trimmed using Trimmomatic (v0.39) (23) using the following parameters for SAMMY-seq and ChIP-seq: 2 for seed_mismatch, 30 for palindrome_threshold, 10 for simple_threshold, 3 for leading, 3 for trailing and 4:15 for sliding window. The sequence minimum length threshold of 35 was applied to all datasets. We used the Trimmomatic “TruSeq3-

SE.fa" (for single-end) and "TruSeq3-PE-2.fa" (for paired-end) as clip files. After trimming, reads were aligned using BWA (v0.7.17-r1188) (24) setting –k parameter as 2 and using as reference genome the UCSC hg38 genome (only canonical chromosomes were used) and the output saved in BAM file format. We uniformly used and aligned only a single read per DNA fragment for both single-end and paired-end sequencing samples. The PCR duplicates were marked with Picard (v2.22; https://github.com/broadinstitute/picard) MarkDuplicates option, then filtered using Samtools (v1.9) (25). In addition, we filtered all the reads with mapping quality lower than 1. Each sequencing lane was analysed separately and then merged at the end of the process. For C002-rep1 and C004-rep2 two sequencing runs from the same library were produced and merged after alignment and filtering.

The coverage estimation (as reported in Supplementary Figure 1e) has been performed using the Samtools (v1.16.1) "coverage" command.

High-throughput sequencing reads analyses

To compute reads distribution profile (genomic tracks), we used Deeptools (v3.4.3) (26) bamCoverage function. For these analyses the genome was binned at 50bp, the reads extended up to 250 bp and RPKM normalization method was used. We considered a genome size of 2701495761 bp (value suggested in the Deeptools manual https://deeptools.readthedocs.io/en/latest/content/feature/effectiveGenomeSize.html) and we excluded regions known to be problematic in terms of sequencing reads coverage using the blacklist from the ENCODE portal (https://www.encodeproject.org/files/ENCFF356LFX/).

To compute the genomic tracks for ChIP-seq IP over INPUT enrichment profiles (log2 normalized ratio) or for relative comparisons (relative enrichment, i.e. log2 normalized ratio) between SAMMY-seq fractions, we used the SPP R package (v1.16.0) (27) and R statistical environment (v3.5.2). The reads were imported from the BAM files using the "read.bam.tags" function, then filtered using "remove.local.tag.anomalies" and finally the comparisons were performed using

the function “get.smoothed.enrichment.mle” setting “tag.shift = 0” and “background.density.scaling = TRUE”. The resulting enrichment signal corresponds to a log2 normalized ratio between the pair of sequencing samples. To compute correlations between genomic tracks, we used R (v3.5.2) base function "cor" with “method = Spearman”. The genomic tracks were imported in R using the rtracklayer (v1.42.2) (28) library. Then the files were binned at 50kb using the function “tileGenome” and the correlation was computed per chromosome. The correlation values obtained for each chromosome were then summarized in one value describing the genome-wide sample correlations through a weighted mean, where the weight of each chromosome corresponds to its length.

To compute the average profile over a set of genomic regions of interest (i.e. the meta-profile analyses), we used DeepTools (v3.4.3) “computeMatrix” command, using as regions of interest the protein coding genes annotated by the Ensembl database (genome version GRCh38.p13) that are not overlapping to each other considering an extended window of 5 Kb (+/-2.5Kb around the gene) and that fall in a length range between 2Kb and 20Kb. The binning value was set to 10bp, the body region was rescaled to 3,000bp and the flanking regions included up to 2Kb (see Supplementary Figure 2b). In addition, the “skipZeros” option was added to remove regions with no coverage. The genes were separated by quartile of expression after separating the genes with 0 TPM. The meta-profile plots were then drawn using the “plotProfile” tool of DeepTools using as input the previously created matrix. These same filters and gene grouping criteria were applied to RPKM values of Roadmap Epigenomics expression profiles (for Figure 3a).

The chromatin marks enrichment matrix associated with Roadmap Epigenomics chromHMM chromatin states (Supplementary Figure 4a) was computed as the average ChIP-seq IP over INPUT enrichment (computed with SPP package as described above) for each chromatin mark and state. Given the different range of values, the colour gradient in the heatmap has been rescaled, for each chromatin mark, to the maximum enrichment for that mark. For SAMMY-seq fractions, the

average enrichment is computed from RPKM and rescaled as a z-score to allow comparison of different SAMMY fractions with distinct enrichment ranges.

Genomic tracks visualization

The visualization of genomic tracks was performed with Gviz R library (v1.26.5) (29). The track profile was calculated using the function “DataTrack” (the input file was imported using the function “import” of the rtracklayer library) and plot using the function “plotTracks” setting the value “window = 1000”. Line plots were drawn setting the parameter type as ‘a’ and overlayed using the function “OverlayTrack”; instead, mountain plots were obtained setting the parameter type as "polygon". Extra elements of these plots, such as chromosome ideogram (on top) and genome axis, were plotted respectively using the functions “IdeogramTrack” and “GenomeAxisTrack”. The analysis was performed identically for all the datasets except for the H3K27ac tracks, where the “window” parameter was set to 10,000 to improve the signal visualization, and for Figure 4c-d, where the “window” was set to 200.

Chromatin compartments analysis

We used the Hi-C and SAMMY-seq pairwise correlation matrices of normalized contacts and enrichment signals, respectively, binned at 250 kb (Figure 2a). Bins with null contacts or signals were removed from both matrices.

Namely, for the Hi-C protocol, we first: i) balanced the raw contact matrix (cooler balance --cis-only, version 0.8.3), ii) loaded the balanced chromosome-wise contact matrix (cooler dump, version 0.8.3), and iii) normalized it by dividing each observed pairwise contact by the mean of the contacts at the same distance (*i.e.*, the expected); secondly, we computed the correlation between all pairs of bins. This step required the calculation of *n*x*n* correlations of *n*-dimensional vectors of normalized contacts, where *n* is the number of bins. For the SAMMY-seq protocols, we loaded the bin enrichment tracks (RPKM) of each fraction and computed the correlation between pairs of bins. Here, the correlation was computed between *m*-dimensional vectors of enrichment signals, where m was

the number of fractions in the specific protocol (*i.e.*, m=3 for the 3f and 10k protocols, m=4 for the 4f protocol).

For each input correlation matrix, the first eigenvector was obtained through principal component analysis decomposition in R statistical environment (prcomp, stats package, center=FALSE and scale=TRUE, rotation component of the returned object). The sign of the first eigenvector was defined using gene density: the group of bins with the highest gene density was marked as "A" compartment (positive sign), and the group with the lowest gene density was marked as "B" (negative sign). Chromosome eigenvector values were divided by the absolute maximum value for visualisation purposes. All the analyses were made using R (version 3.6.0).

Sub-compartments

We partially reimplemented CALDER algorithm (version 1.0, 2020-09-01) to accommodate the unusual format of SAMMY-seq data, maintaining the primary set of core functions (remove_blank_cols, fast_cor, generate_compartments_bed, HighResolution2Low_k_rectangle, get_PCs, bisecting_kmeans, project_to_major_axis, get_best_reorder, get_cluser_levels), their default parameters and the steps intended in the original paper (8).

As inputs, we used the balanced chromosome-wise contact matrix (cooler dump, version 0.8.3) and the pairwise distance matrix binned at 50 kb for Hi-C and SAMMY-seq protocols, respectively (Figure 5a). In particular, the pairwise distance matrix was computed using for each pair of bins the Euclidean distance (dist, stats package, method="euclidean") of the m-dimension vectors, where m was the number of fractions in the specific protocol (*i.e.*, 4 for the 4f protocol). Bins with null contacts or signals were removed from both matrices.

Briefly, for each input matrix, the algorithm: computed the pairwise correlation matrix and identified the sub-domain boundaries; computed the binary trend matrix and its decomposition using ten principal components; iteratively clustered the domains to obtain their hierarchy; sorted the domains hierarchy based on the projection of the first two components; divided them into eight groups (form the

most closed, *i.e.*, B.2.2 to the most opened, i.e., A.1.1). For more details on the original CALDER procedure and the interpretation of sub-compartments please see (8).

2.4 RESULTS

Mapping open chromatin regions by solubility with high resolution.

We previously proposed the SAMMY-seq method based on the sequential biochemical purification and sequencing of three distinct chromatin fractions, hereafter called "3f" SAMMY-seq (16). These fractions were respectively isolated after partial digestion with an endonuclease (DNase, fraction S2), extraction with high salt concentration to dissolve ionic bonds (fraction S3), and treatment with urea buffer to dissolve the remaining protein and membrane pellet (fraction S4) (Figure 1a and Supplementary Figure 1a).

Figure 1

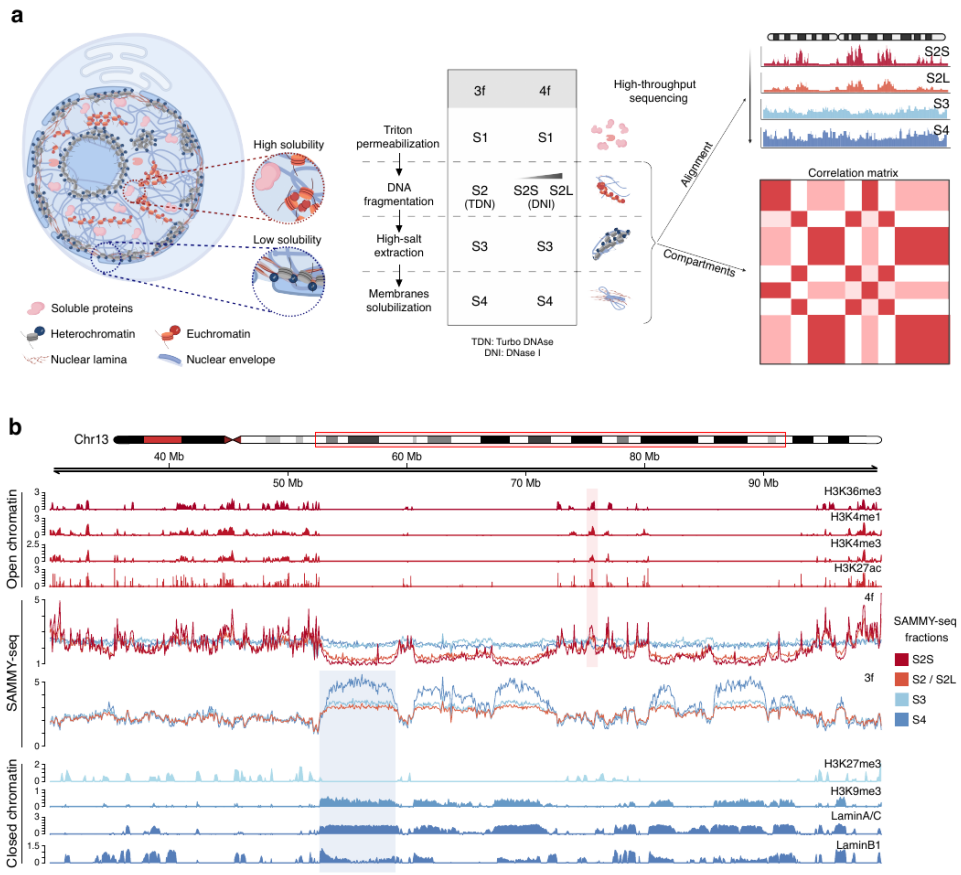
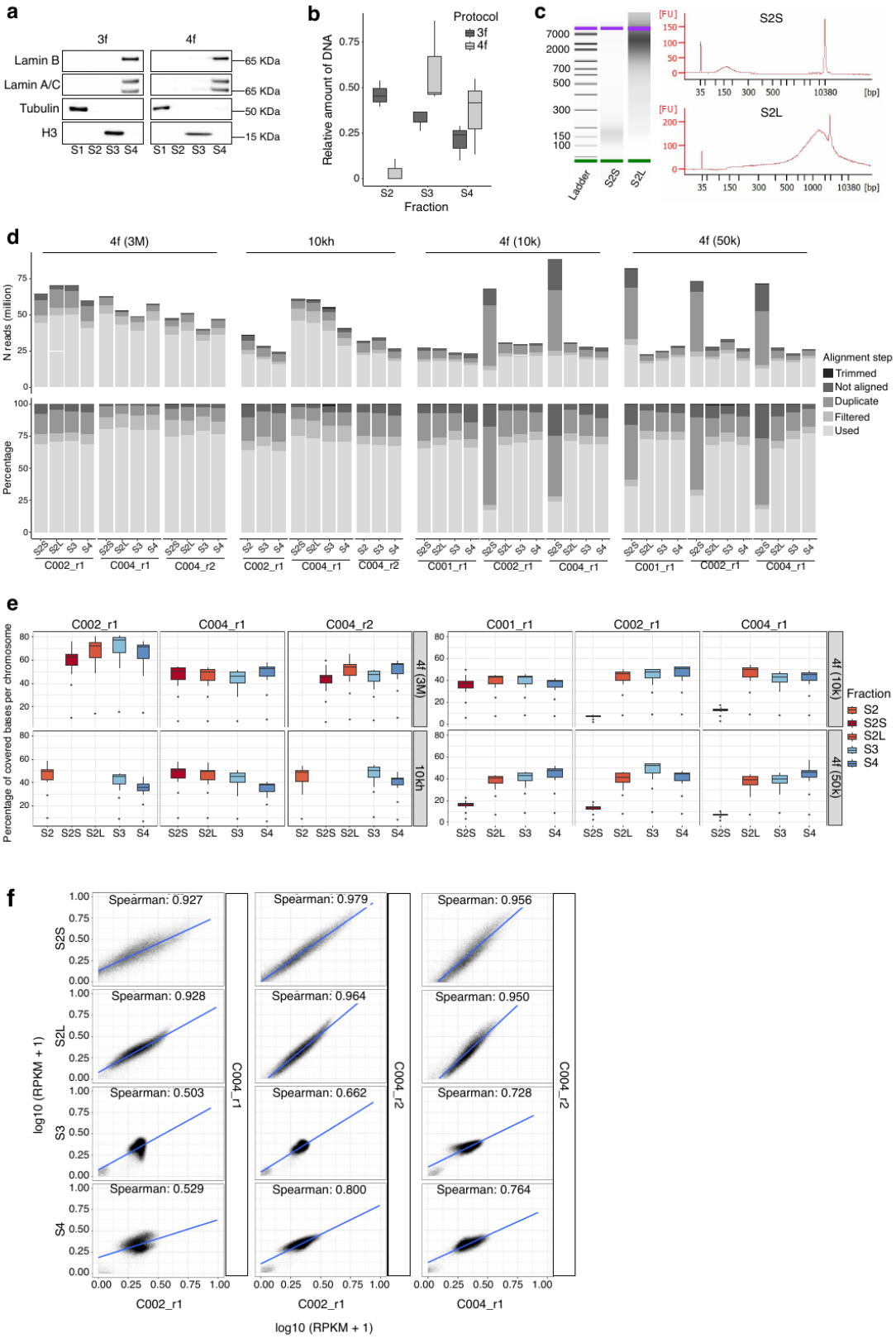


Figure 1 - 4f-SAMMY-seq maps both euchromatin and heterochromatin with high resolution. a) Schematic illustration of the SAMMY-seq protocol variants and output analysis results. From left to right, high (euchromatin) and low solubility (heterochromatin) domains correspond to genomic portions with preferential segregation in different subnuclear regions. In both "3f" and "4f" SAMMY-seq protocols, a sequential extraction of chromatin fractions (numbered

from S1 to S4) results in a separation of euchromatic and heterochromatic regions that are then mapped to their genomic coordinates by high-throughput sequencing (applied to fractions from S2 through S4). The novel 4f-SAMMY-seq has DNase I (DNI) replacing Turbo DNase (TDN) as endonuclease and the S2 fraction is size separated. Specific data analysis procedures allow reconstructing chromatin domains compartmentalization in the 3D nuclear space. **b)** Representative genomic region (chr13:30,000,000-100,000,000) showing genomic tracks for chromatin marks in human foreskin fibroblasts (E055 Roadmap Epigenomics). From top to bottom: open chromatin marks ChIP-seq enrichment profiles for H3K36me3, H3K4me1, H3K4me3, H3K27ac; reads distribution profiles for individual fractions of a representative replicate of 4f-SAMMY-seq (C004_r2) and 3f-SAMMY-seq (C004_r1); closed chromatin marks ChIP-seq enrichment profiles for H3K27me3, H3K9me3, Lamin A/C, Lamin B1. The shaded areas mark two examples of regions showing enrichment for closed (blue) or open (red) chromatin marks.

The latest fractions of 3f-SAMMY-seq (S3 and S4) are enriched for the compacted and less soluble heterochromatin domains (Figure 1b, Supplementary Figure 1a and 2a). To improve the resolution in characterizing the accessible euchromatin, usually found at actively transcribed genes and regulatory elements, we modulated the DNase digestion leading to S2 fraction isolation (Figure 1a and Supplementary Figure 1b, see Methods for details). Moreover, we introduced a size selection step on the DNA isolated from the S2 fraction, to further separate more digested small-size DNA fragments (S2-Short, S2S, <300bp) from less digested large-size DNA fragments (S2-Long, S2L) (Figure 1a and Supplementary Figure 1c). The resulting four chromatin fractions were sequenced, thus we named this the "4f" SAMMY-seq protocol. We applied 4f-SAMMY-seq on human primary fibroblasts and compared the results to 3f-SAMMY-seq data on the same cells. We confirm good and comparable quality controls on the proteins associated with each fraction (Supplementary Figure 1a) as well as on the sequencing reads, achieving good coverage and reproducibility even with moderate sequencing depth (minimum of 25 million reads per fraction) (Supplementary Figure 1d-f).

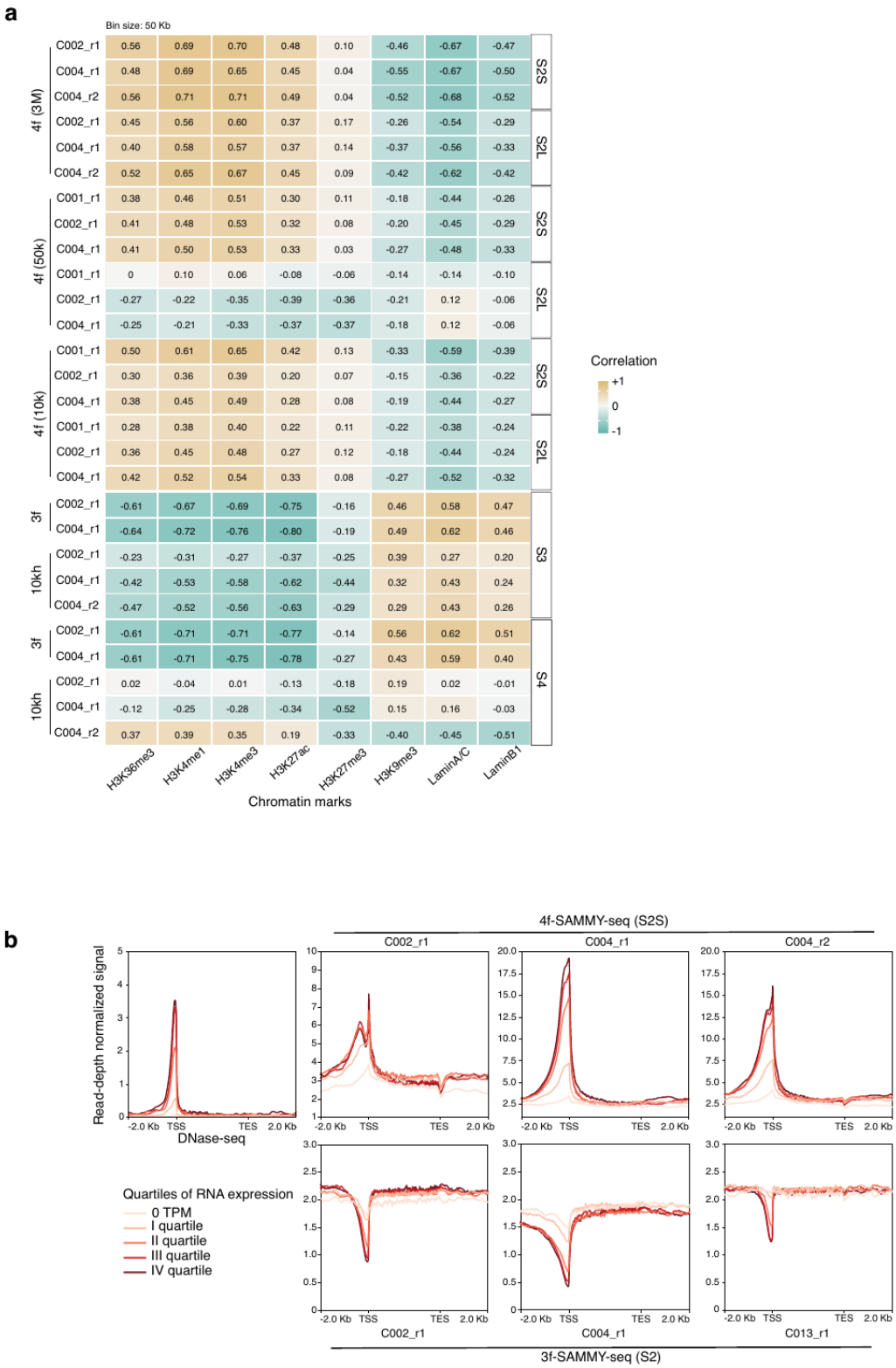
Supplementary Figure 1



Supplementary Figure 1 - Quality controls on SAMMY-seq experimental procedure. a) Representative western blots of chromatin fractions obtained with "3f" and "4f" SAMMY-seq protocols. In both cases we find soluble tubulin in S1 fraction, histone H3 in S3 fraction and lamins in S4 fraction. b) Relative abundance (y-axis) of DNA extracted from S2, S3 and S4 fractions, computed as ratio over their sum, in "3f" and "4f" SAMMY-seq protocols (dark and light grey, respectively). In the boxplots, the horizontal lines mark the median, the boxes mark the interquartile range (IQR) and whiskers extend up to 1.5 times the IQR. Relative abundance of DNA was evaluated over 3 independent biological replicates for each protocol. c) Bioanalyzer electropherograms of S2S and S2L DNA fragments showing their different size distributions. d) Stacked barplots for the number of sequencing reads obtained for each sample and chromatin fraction, divided in trimmed, not aligned, duplicates, filtered (discarded) and used reads (retained for downstream analyses). See also the associated Supplementary Table 1. e) Boxplots summarizing coverage across chromosomes for each sample and fraction. The percentage of each chromosome covered by at least one read is reported on the y-axis. In the boxplots, the horizontal lines mark the median, the boxes mark the interquartile range (IQR) and whiskers extend up to 1.5 times the IQR. f) Scatter plots of reads distribution profiles (RPKM normalized coverage over 50kb bins) for individual fractions of 4f-SAMMY-seq replicates. Fractions labels indicated on individual rows (labels on the left) and the replicates compared in each plot are indicated on bottom and right-side axes.

The S2S and S2L fractions are able to precisely recapitulate the location of histone marks associated with active chromatin (H3K36me3, H3K4me1, H3K4me3 and H3K27ac) (Figure 1b, and Supplementary Figure 2a). In particular, the most accessible S2S fraction achieves a spatial resolution comparable to DNase-seq (30) in mapping the accessibility profiles around the transcription start site (TSS) of annotated genes, also in their correlation with expression level, thus overcoming 3f-SAMMY-seq limitations to this concern (Supplementary Figure 2b).

Supplementary Figure 2



Supplementary Figure 2 - 4f-SAMMY-seq and 10kh-SAMMY-seq recapitulate open and closed chromatin domains. a) Genome-wide Spearman correlation between reads distribution

profiles for individual selected SAMMY-seq chromatin fractions and ChIP-seq enrichment profiles for histone marks and lamin proteins (x-axis labels) from human foreskin fibroblasts (E055 Roadmap Epigenomics). The "3f", "4f" or low input 4f-10k, 4f-50k and "10kh" labels on the left side indicate protocol versions for 3f-SAMMY-seq (3f), 4f-SAMMY-seq (4f), 4f-SAMMY-seq starting from 10,000 or 50,000 cells (10k and 50k, respectively), 10kh-SAMMY-seq (10kh). The labels for individual chromatin fractions (S2S, S2L, S3 and S4) are reported on the right. A row for each replicate is shown and correlation values are reported as numbers and as colour gradient. **b)** Gene centred meta-profiles for reads distribution profiles of a reference DNase-seq sample (dermal fibroblasts from ENCODE) 4f-SAMMY-seq S2S fraction (three replicates) and 3f-SAMMY-seq S2 fraction (three replicates). Genes are divided by quartiles of expression level (TPM) and a meta-profile is drawn for each quartile, as well as for genes with zero TPM (no reads), as measured by RNA-seq from (16). On the x-axis, the coordinates for the relative genomic position around the rescaled gene body are reported.

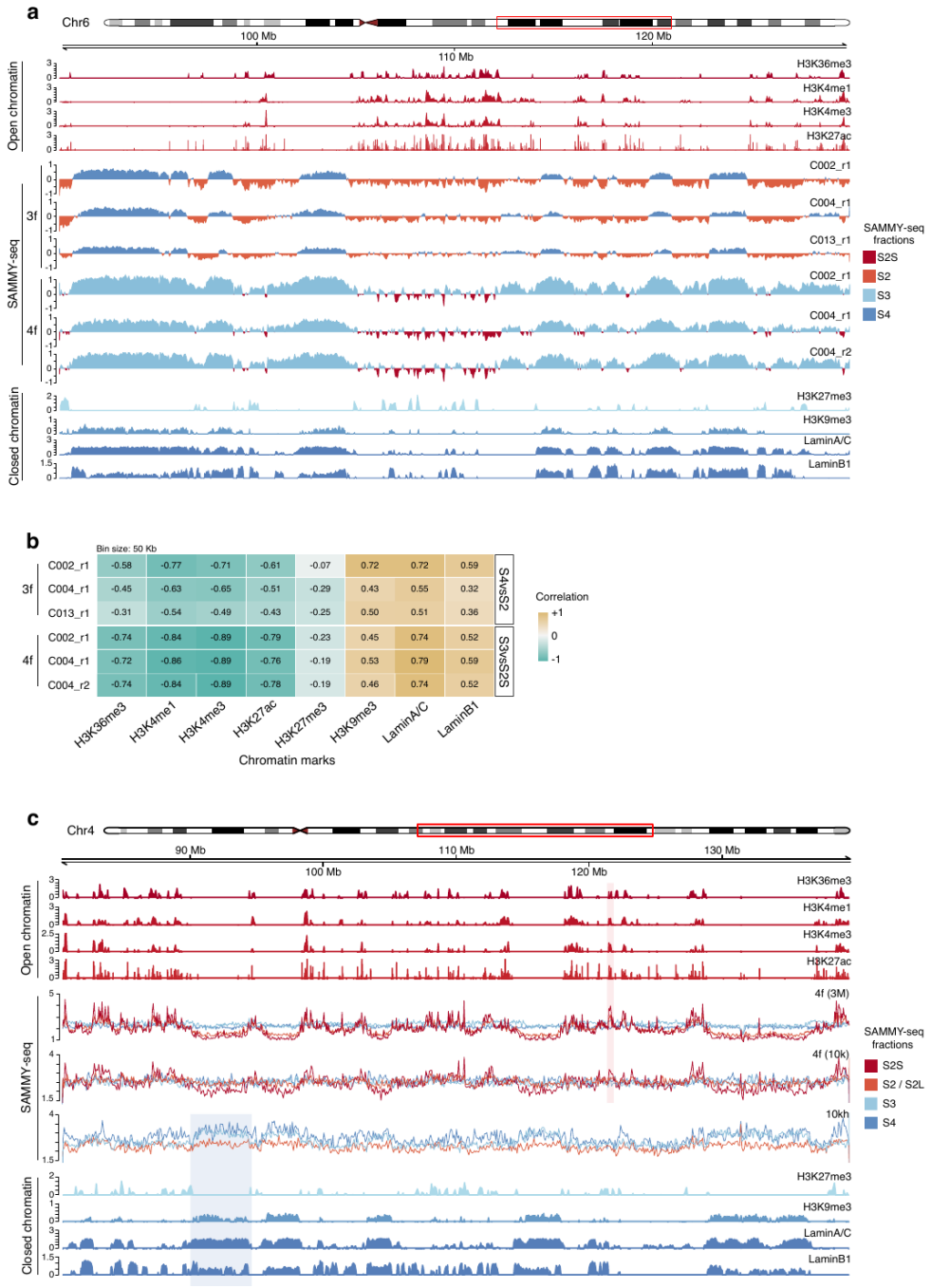
Overall, these data support our hypothesis that the combination of controlled digestion with fragments size separation allows a more precise characterization of open chromatin regions.

Active and inactive chromatin mapping on a small number of cells.

The relative sequencing reads enrichment in closed (S3) vs open (S2S) chromatin fractions in 4f-SAMMY-seq allows identifying constitutive lamina associated heterochromatin with comparable or better results with respect to 3f-SAMMY-seq (Supplementary Figure 3a,b). We noted that this result is driven by S2S reads depletion over heterochromatin regions, as S3 and S4 fractions from 4f-SAMMY-seq are not clearly enriched in closed chromatin, as they were instead in 3f-SAMMY-seq (Figure 1b). The lack of heterochromatin enrichment in insoluble S3 and S4 fractions of 4f-SAMMY-seq may be expected given the lighter endonuclease digestion, possibly leaving a mixture of open and closed chromatin in the latest fractions (Supplementary Figure 1b). To confirm this hypothesis, we performed the "4f" chromatin fractionation starting from 10,000 cells and scaling up the ratio of enzyme units per starting number of cells (see Methods section for "10kh" SAMMY-seq experiment). In this condition, we confirm that the S3 fraction is clearly enriched in heterochromatic regions (Supplementary Figures 1d-e, 2a and 3c). Of note, in the 10kh-SAMMY-seq

protocol the S4 is not as informative, possibly due to the limited amount of DNA left in the last fraction when starting from 10,000 cells (Supplementary Figure 2a).

Supplementary Figure 3



Supplementary Figure 3 - Relative comparison of chromatin fractions recapitulate heterochromatin regions. a) Representative genomic region of (chr6:62,000,000-130,000,000) showing genomic tracks for chromatin marks in human foreskin fibroblasts (E055 Roadmap Epigenomics). From top to bottom: open chromatin marks ChIP-seq enrichment profiles for H3K36me3, H3K4me1, H3K4me3, H3K27ac; relative enrichment of closed (S3 or S4) over

accessible (S2 or S2S)(see color legend) chromatin fractions from "3f" and "4f" SAMMY-seq (labels on the left) protocol variants in all replicates; closed chromatin marks ChIP-seq enrichment profiles for H3K27me3, H3K9me3, Lamin A/C, Lamin B1. **b)** Genome-wide Spearman correlation between relative enrichment profiles of closed (S3 or S4) over accessible (S2 or S2S) (as indicated in the labels on the right) chromatin fractions from "3f" and "4f" SAMMY-seq protocol variants (labels on the left) for individual replicates. **c)** Representative genomic region of (chr4:80,000,000-140,000,000) showing genomic tracks for chromatin marks in human foreskin fibroblasts (E055 Roadmap Epigenomics). From top to bottom: open chromatin marks ChIP-seq enrichment profiles for H3K36me3, H3K4me1, H3K4me3, H3K27ac; reads distribution profiles for individual fractions of a representative replicate of 4f-SAMMY-seq on 3M (C004_r2) and 10k cells (C001_r1), as well as 10kh-SAMMY-seq (C002_r1); closed chromatin marks ChIP-seq enrichment profiles for H3K27me3, H3K9me3, Lamin A/C, Lamin B1. The shaded areas mark two examples of regions showing enrichment for closed (blue) or open (red) chromatin marks, with corresponding enrichment patterns visible in ChIP-seq as well as in SAMMY-seq fractions.

We then tested the possibility of performing 4f-SAMMY-seq starting from as little as 10,000 or 50,000 cells and scaling down the DNAse enzyme units as well (see Methods section for 4f-SAMMY-seq 10k and 50k experiments, respectively). In these conditions, we recapitulate the results obtained with 3 million cells (4f-SAMMY-seq experiments) (Supplementary Figures 1d-e, 2a and 3c).

These results confirm, the applicability of 4f-SAMMY-seq on as little as 10,000 cells, and the possibility of modulating the enrichment for heterochromatin in the less soluble fractions by tuning the endonuclease partial digestion step.

SAMMY-seq based chromatin compartments consistently recapitulate chromatin states properties.

SAMMY-seq protocols provide a comprehensive panel of chromatin fractions covering a broad spectrum of biochemical properties. Chromatin compartments are commonly identified based on the long-range similarity of contact profiles measured by Hi-C for each pair of genomic loci. We hypothesized that chromatin domains spatially located in the same 3D nuclear neighbourhood are exposed to the same biochemical milieu, thus sharing similar solubility. We compared each pair of genomic regions by computing their "biochemical similarity" as the correlation in the sequencing reads coverage across SAMMY-seq fractions (see

Methods) (Figure 2a). The resulting 2D matrix is similar to the pairwise correlation matrix of normalized Hi-C data on human fibroblasts (22) (Figure 2b). The matrices eigenvector decomposition confirms that the linear segmentation in "A" and "B" compartments is mostly concordant between Hi-C and SAMMY-seq protocols, with median Jaccard Index across replicates equal to 0.73, 0.62 and 0.78 for 3f, 10kh and 4f-SAMMY-seq (3M), respectively (Figure 2c). 4f-SAMMY-seq shows the highest concordance with Hi-C, in addition to the highest reproducibility of compartment calls across replicas, also when starting from as little as 50k or 10k cells, with median Jaccard Index across replicates equal to 0.76 and 0.75, respectively (Figure 2c, d).

Figure 2

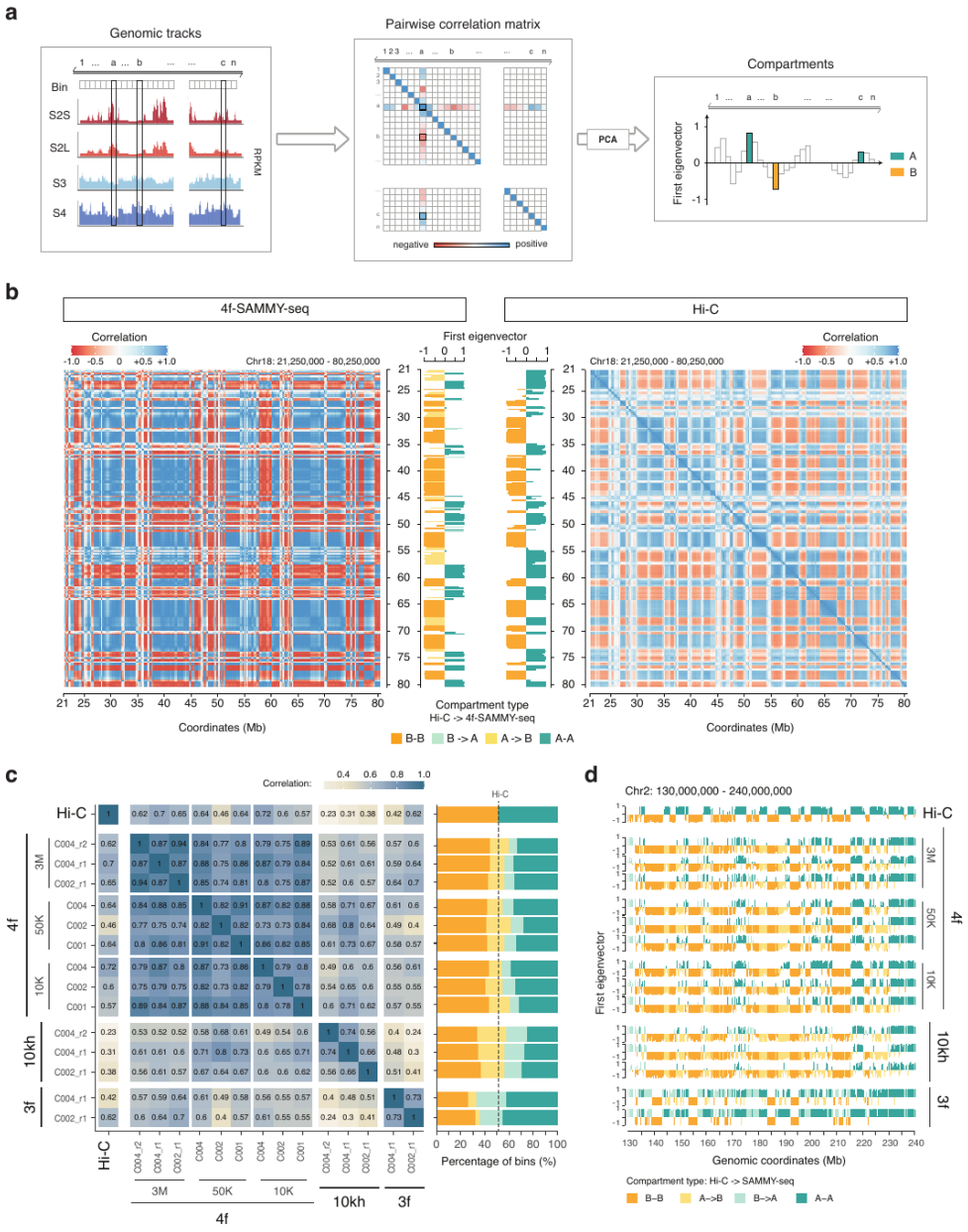


Figure 2 - 4f-SAMMY-seq detects chromatin domains segregation in compartments. a) Schematic illustration of the data analysis workflow to reconstruct chromatin compartments from SAMMY-seq data. For the "A" and "B" compartments identification, starting from reads distribution profiles for individual chromatin fractions, the Pearson correlation is computed between the vectors of reads coverage across fractions for each pair of genomic bins. The pairwise correlation matrix is computed for each chromosome separately. After performing a principal component analysis (PCA), the first component (corresponding to the first eigenvector of the matrix) is used to discriminate active "A" compartment (positive values) and inactive "B" compartment (negative values)(see also Methods). **b)** Pairwise correlation matrices of read distribution profiles for

individual 4f-SAMMY-seq (C002_r1) fractions (left) or distance normalized observed over expected Hi-C contact profiles (right) on a representative genomic region (chr18:21,250,000-80,250,000), computed at 250Kb genomic bins resolution. On the side of each matrix the respective first eigenvector is reported and coloured to mark the position of active ("A" compartment with positive eigenvalues) and inactive regions ("B" compartment with negative eigenvalues). Concordant domain classification in Hi-C and 4f-SAMMY-seq are marked in orange ("A-A") for active regions and green ("B-B") for inactive regions. In the 4f-SAMMY-seq eigenvector only, we marked differently the regions with a compartment classification discordant with Hi-C: a lighter green is used for regions classified as B in Hi-C and A in 4f-SAMMY-seq ("B->A" label), a lighter orange is used for the opposite case ("A->B" label). **c)** Genome-wide pairwise Pearson correlation of chromatin compartments eigenvectors (250kb size bins) defined by reference Hi-C and SAMMY-seq protocol variants "3f", "10kh" and "4f" starting from 3 million (3M), 50,000 (50K) or 10,000 (10K) cells with individual replicates reported. On the right side, for each sample, the stacked barplot shows the relative distribution (percentage) of genomic bins associated with concordant ("A-A" or "B-B") and discordant ("A->B" or "B->A") compartment classification with respect to Hi-C compartment calls. The chromatin compartment classification is reported using the same colouring and naming convention adopted in Figure 2b. **d)** Chromatin compartments eigenvectors for a representative genomic region (chr2:130,000,000-240,000,000). The samples order is the same as in panel b. The eigenvectors are coloured according to the same convention adopted in Figure 2b.

We then analysed in details the differences between 4fSAMMY-based and HiC-based compartmentalization. The genomic windows with a discordant compartment assignment were labelled as "A->B" if they are called as active by Hi-C and inactive by SAMMY-seq, or vice versa for those labelled "B->A" (Figure 2b-d). We examined in details the transcriptional activity in the Hi-C vs SAMMY-seq-based compartment definitions. We found that there is an evident and statistically significant pattern with SAMMY-seq correctly assigning transcribed genes to the "A" compartment (Figure 3a). In particular, we note that only SAMMY-seq consistently assigns to the "A" compartment more than 90% of the genes in both the "high" and "very high" expression groups.

We further dissected these discordant genomic regions in 4f-SAMMY-seq to verify if they are associated with chromatin states (Figure 3b) with specific epigenetic signatures (Supplementary Figure 4a). We used as reference two alternative datasets from Roadmap Epigenomics (21) and confirmed our

observations also on 50k and 10k-SAMMY-seq samples (Figure 3, Supplementary Figure 5 and 6) (see Methods).

Figure 3

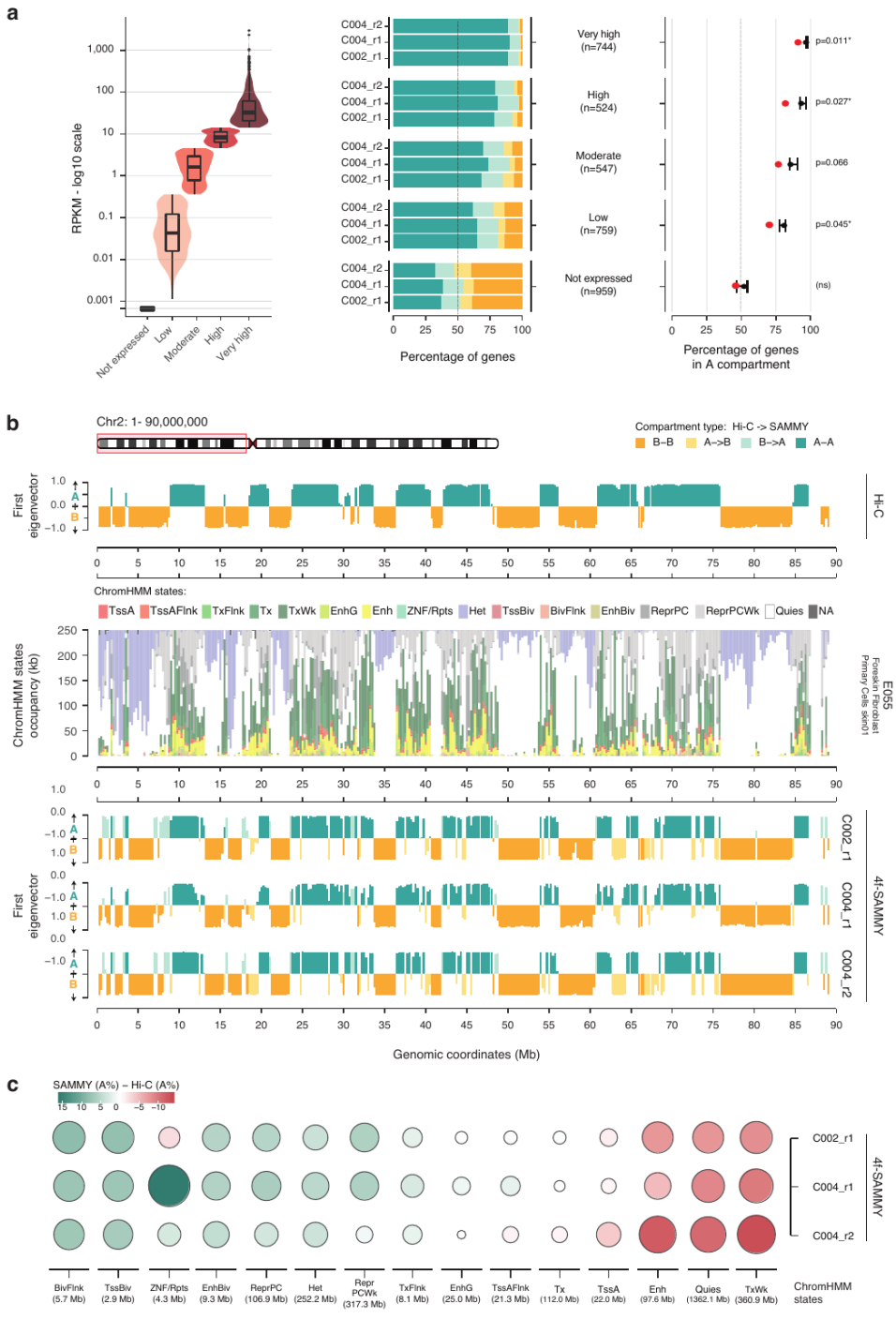


Figure 3 - 4f-SAMMY-seq based compartments provide a detailed characterization of chromatin transcription and epigenetic status. a) Compartments classification comparison with gene expression status. Genes are divided by quartiles of expression level (RPKM values

from Roadmap Epigenomic E055 sample) and labelled as "Low", "Moderate", "High" and "Very high" according to their expression level (violin and box plot on the left). Genes with zero RPKM (no reads as measured by RNA-seq) are assigned to the "Not expressed" group. The stacked barplot shows for each 4f-SAMMY-seq replicate the relative distribution of genes across concordant and discordant compartments classifications with respect to Hi-C compartments (50kb size bins - genes assigned to bins based on the TSS position). The chromatin compartment classification is reported using the same colouring and naming convention adopted in Figure 2b. In the dot and whiskers plot on right side we show the percentage of genes mapped in A compartment by Hi-C (red dot) and by each 4f-SAMMY-seq replicates (whiskers for maximum and minimum range, dot for the median replicate). The significance of differences was tested with a one-sided t-test corrected by FWER with Bonferroni procedure (corrected p-values are reported for each group on the right-hand side) **b)** Classification of "A" and "B" compartments (from Hi-C and 4f-SAMMY-seq) and chromatin states in human foreskin fibroblasts (E055 Roadmap Epigenomics) for a representative region (chr2:1-90,000,000). The eigenvector tracks (green and orange tracks) show compartments computed from Hi-C data (top row) and individual 4f-SAMMY-seq replicates (bottom three rows) coloured according to the same convention adopted in Figure 2 for concordant ("A-A" or "B-B") and discordant ("A->B" or "B->A") compartments classification (250kb size bins). The stacked barplot in the middle row summarizes the chromatin states associated with each genomic bin: 15 chromatin states model (see colour legends for states labels and Supplementary Figure 4a for associated chromatin marks signatures). **c)** Relative occupancy of 4f-SAMMY-seq vs Hi-C based compartments for each chromatin state, computed as the difference in "A" compartment percentage. For each chromatin state (the same as for panel a), positive values (green gradient) indicate a higher percentage of "A" compartment in 4f-SAMMY-seq-based classification, whereas negative values (red gradient) indicate a higher percentage of "A" compartment based on Hi-C classification (i.e. higher "B" percentage based on 4f-SAMMY-seq). The size of each dot is proportional to the absolute value in the percentage difference. Chromatin states are ordered from left (positive values) to right (negative values) based on the average percentage difference between 4f-SAMMY-seq replicates (average between the three replicates) and Hi-C.

The most prominent difference with Hi-C is on multiple chromatin states associated with Polycomb regulation, including monovalent repressive Polycomb (ReprPC, ReprPCWk) and bivalent chromatin states (BivFlnk, TssBiv, EnhBiv) (Figure 3b-c, Supplementary Figures 4b, 5a-b and 6a). It must be noted that, in line with previous literature, Hi-C generally classifies Polycomb-regulated domains as belonging to the "A" compartment (5). However, in our analyses, Hi-C classifies a few of these Polycomb-regulated regions in the inactive "B"

compartment, whereas 4f-SAMMY-seq is more consistently classifying all of these Polycomb domains in the "A" compartment (Figure 4a, Supplementary Figures 5c and 6b). We also verified that the percentage of Polycomb-related chromatin states per genomic bin is more clearly correlated to the eigenvectors of 4f-SAMMY-seq compartments rather than to the eigenvector derived from Hi-C, with Hi-C scatter plot also showing a more dispersed distribution (Figure 4b and Supplementary Figure 5d). As an example of Polycomb-regulated chromatin, we show the homeobox (HOX) gene clusters.

Figure 4

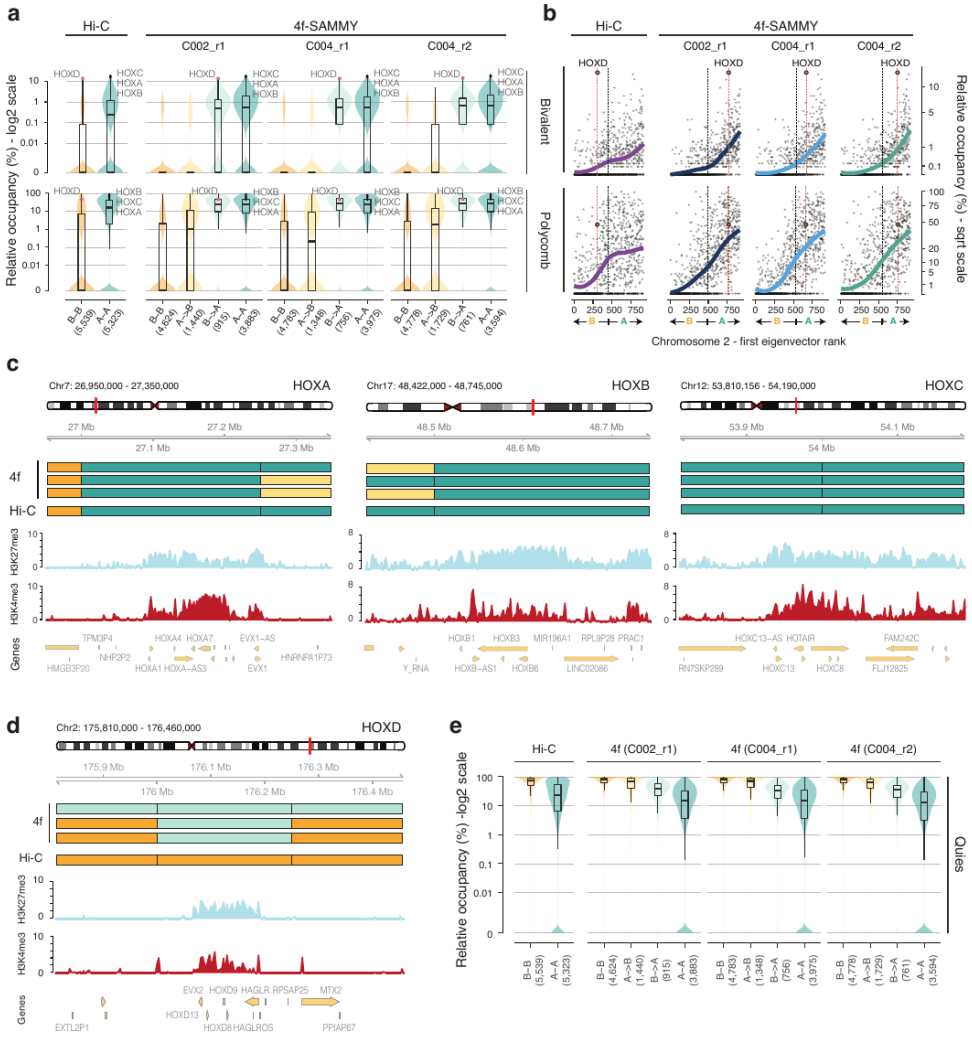


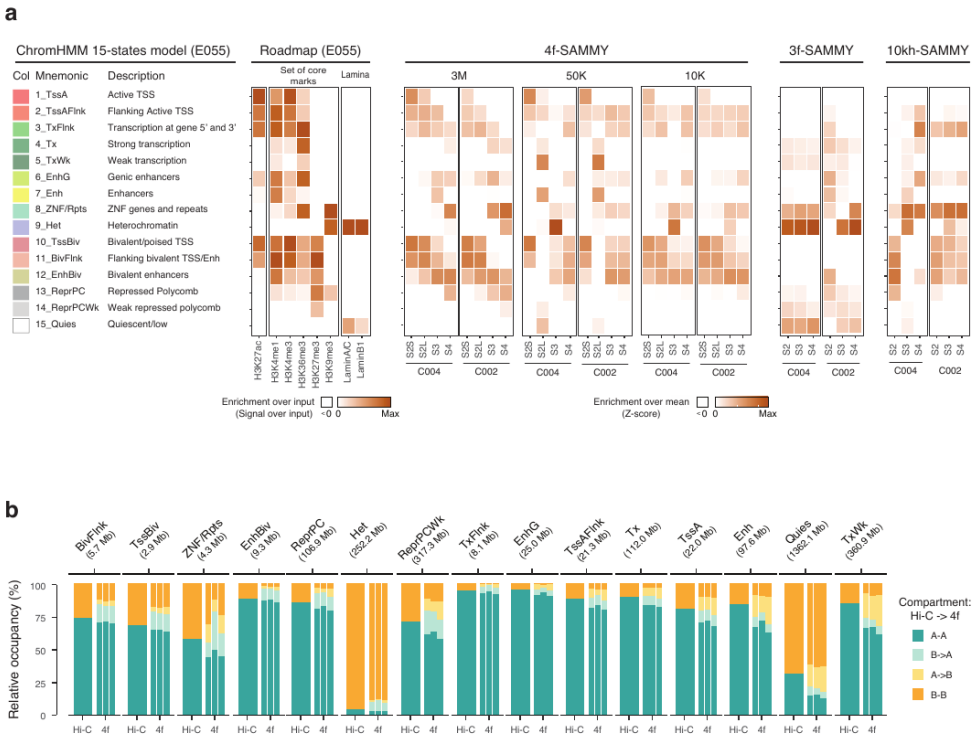
Figure 4 - 4f-SAMMY-seq based compartments consistently classify Polycomb regulated domains. a) Violin and box plots showing for the reference Hi-C dataset and individual 4f-SAMMY-seq replicates (labels on the upper margin) the distribution of Polycomb-regulated chromatin states across genomic bins of the entire genome grouped by compartments

classification ("A-A", "B-B", "A->B" or "B->A" defined as in Figure 2b). The upper violin and box plot report bivalent Polycomb states (TssBiv, EnhBiv or BivFln) the bottom ones monovalent repressive Polycomb chromatin states (ReprPC or ReprPCWk). The number of genomic bins in each group is indicated at the bottom in parentheses (x-axis). In the overlaying boxplots, the horizontal line marks the median, the box margins mark the interquartile range (IQR), and whiskers extend up to 1.5 times the IQR. Specific data points associated with HOX gene clusters are marked to show their positioning across groups and their associated high relative occupancy in Polycomb repressive and bivalent states. **b)** Scatter plots showing the relationship between the rank of the first eigenvector values used to define compartments (ranks on the x-axis) and the relative occupancy by chromatin states (percentage of each bin) in squared root (sqrt) scale (y-axis). Each data point shows a 250kb genomic bin from the representative chromosome 2. Black dashed vertical lines mark the transition point between negative and positive values of the first eigenvector (from "B" to "A" compartment). Coloured solid lines highlight the overall trend (lowess regression), for the reference Hi-C dataset (violet line) and for each 4f-SAMMY-seq replicate (blue, light blue and green; labels on upper margins). Dashed salmon circles and vertical lines highlight the HOXD gene cluster bin. Upper plots show the chromatin states occupancy for bivalent Polycomb states (TssBiv, EnhBiv or BivFlnk) and bottom plots monovalent repressive Polycomb states (ReprPC or ReprPCWk), from E055 Roadmap Epigenomics sample. **c)** Genomic regions of the HOXA, HOXB and HOXC gene clusters showing, from top to bottom: the chromatin compartments classification based on 4f-SAMMY-seq (three rows for the replicates C002_r1, C004_r1 and C004_r2, respectively); the chromatin compartments classification based on Hi-C; ChIP-seq enrichment profiles for H3K27me3 and H3K4me3; location and orientation of a subset of annotated genes (yellow arrows). The chromatin compartments are coloured using the same convention adopted in Figure 2b. **d)** Genomic region of the HOXD gene cluster showing, from top to bottom: the chromatin compartments classification based on 4f-SAMMY-seq (three rows for the replicates C002_r1, C004_r1 and C004_r2, respectively) and based on Hi-C; ChIP-seq enrichment profiles for H3K27me3 and H3K4me3; location and orientation of a subset of annotated genes (yellow arrows). The chromatin compartments are coloured using the same convention as in Figure 2b. **e)** Violin and box plots showing for the reference Hi-C dataset and for individual 4f-SAMMY-seq replicates (labels on the upper margin) the distribution of the Quiescent (Quies) chromatin states across genomic bins of the entire genome grouped by compartment classification ("A-A", "B-B", "A->B" or "B->A" defined as in Figure 2b). The number of genomic bins in each group is indicated at the bottom in parentheses (x-axis). In the overlaying boxplots, the horizontal lines mark the median, the boxes mark the interquartile range (IQR), and whiskers extend up to 1.5 times the IQR.

These are a set of developmentally regulated highly conserved gene loci expected to be all marked by Polycomb-dependent H3K27me3 histone mark in

terminally differentiated cells, or paired with H3K4me3 for the bivalent chromatin state. In humans, 39 HOX genes are organized in 4 distinct chromosomal locations: HOXA, HOXB, HOXC and HOXD gene clusters. While they all show the same chromatin marks, we notice that Hi-C classifies most of them in the "A" compartment, with the sole exception of the HOXD cluster classified in the "B" compartment. Instead, 4f-SAMMY-seq consistently classifies all of them in the "A" compartment (Figure 4c,d).

Supplementary Figure 4

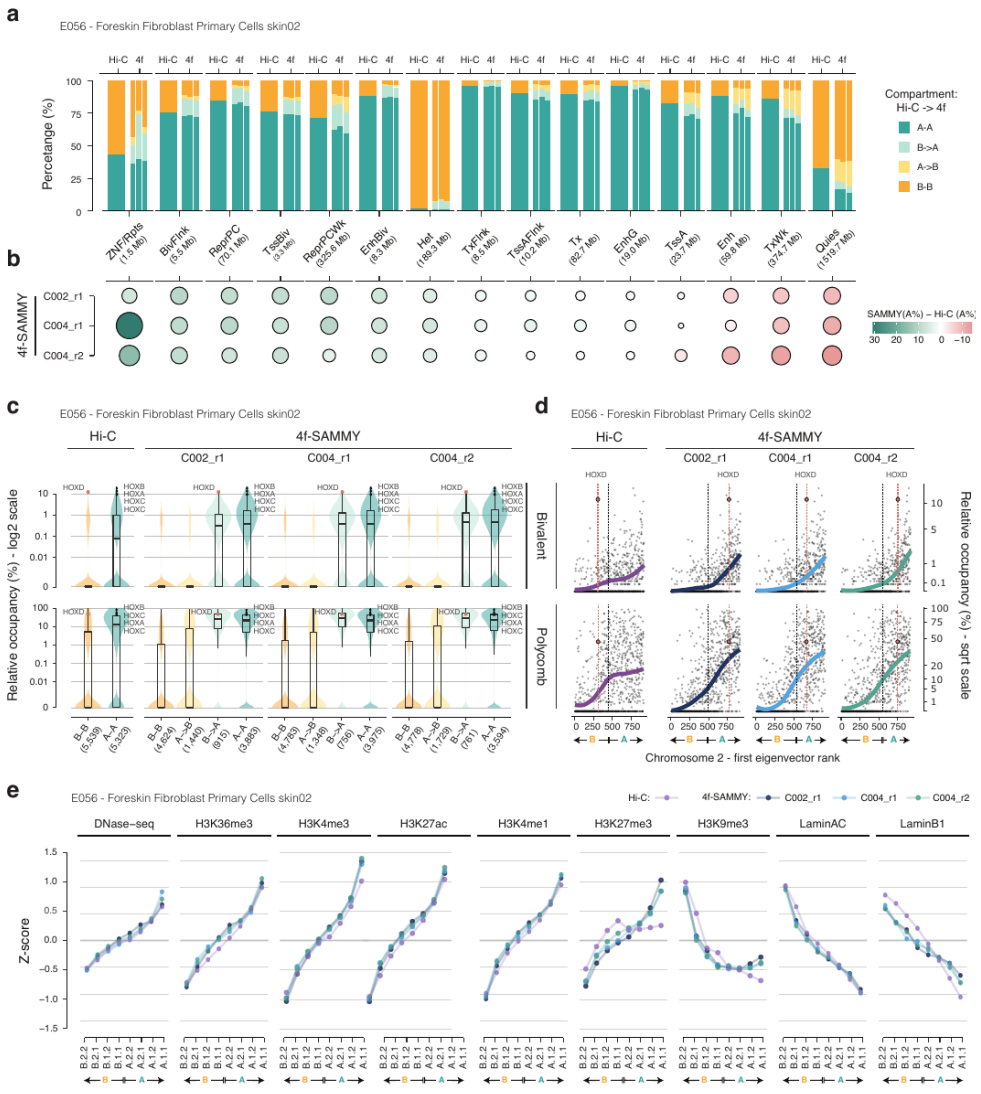


Supplementary Figure 4 - 4f-SAMMY-seq chromatin fractions and chromatin marks association to chromatin states. a) Epigenetic marks enrichment signatures associated to chromHMM chromatin states (15 states model from E055 Roadmap Epigenomics). Starting from the left: the colour code, the name and the description of the 15 chromatin states as defined by Roadmap Epigenomics, the heatmaps representing the enrichment over the input for H3K27ac, H3K4me1, H3K4me3, H3K36me3, H3K27me3 and H3K9me3, lamin A/C and B1, the heatmaps representing the enrichment over the mean (z-score) of individual SAMMY-seq fractions for protocol versions “4f” (3M, 50K and 10K), “3f” and “10kh” with two representative replicates for each of them. The histone marks originally used in the chromHMM definition of states are grouped in the "core marks" quadrant in the figure. **b)** The stacked barplot shows the percentage of regions associated with concordant ("A-A" or "B-B") and discordant ("A->B" or "B->A") compartment

classification for each chromatin state (labels on the upper margin indicating chromatin states and their total size). The chromatin compartments classification is reported using the same colouring convention adopted in Figure 2b. The barplot for 4f-SAMMY-seq compartments distribution is divided in three parts to report the specific result for the three replicates (C002_r1, C004_r1 and C004_r2, respectively). The chromatin states are ordered from left to right based on the average difference between Hi-C and SAMMY-seq based compartments, as calculated for Figure 3c.

On the opposite trend, regions with a "Quies" chromatin state, characterized by weak enrichment in lamina association (Supplementary Figure 4a), are more consistently classified in the "B" compartment based on 4f-SAMMY-seq rather than Hi-C (Figure 3c and Figure 4e).

Supplementary Figure 5

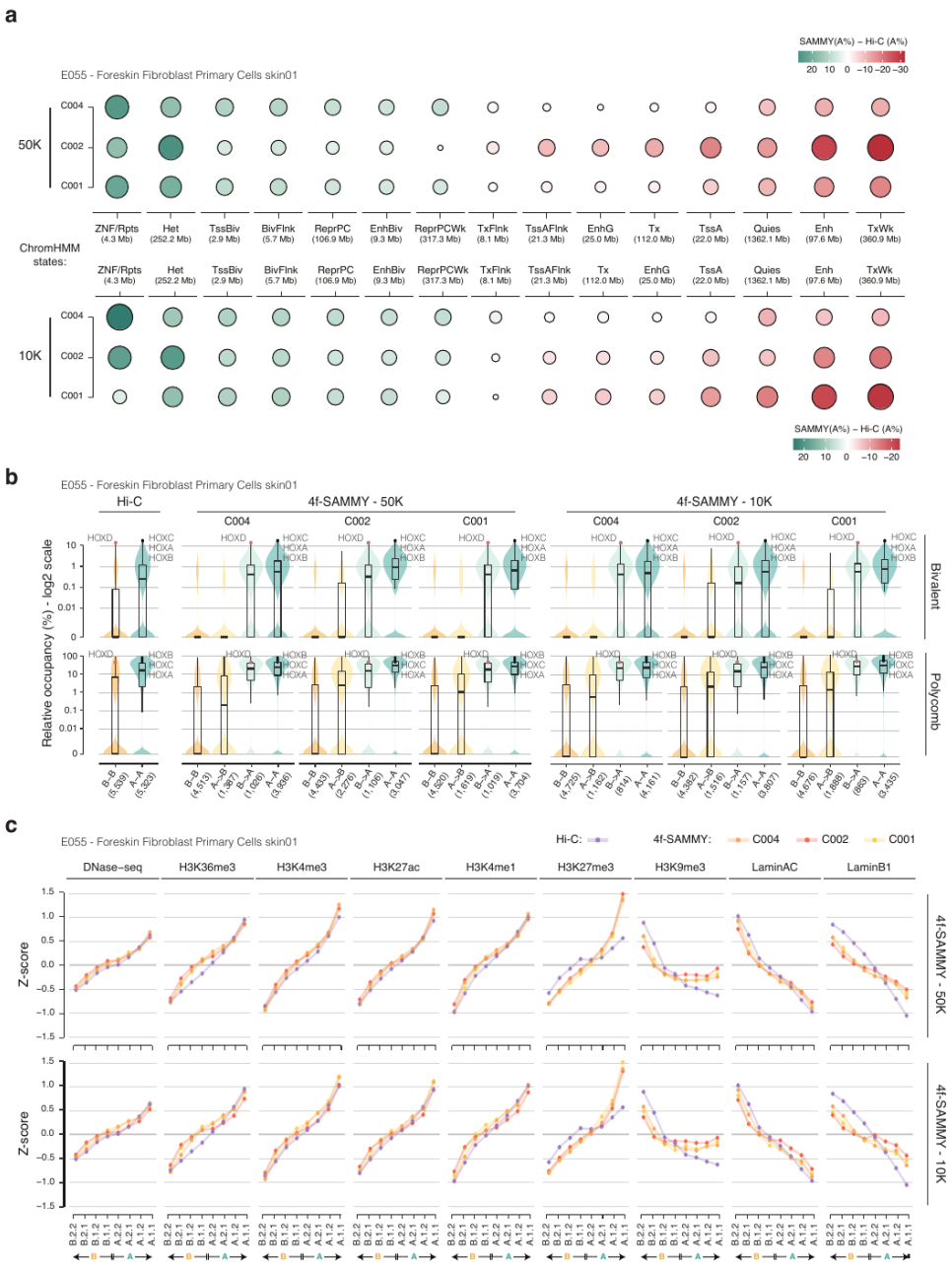


Supplementary Figure 5 - 4f-SAMMY-seq based compartments and sub-compartments on alternative epigenomic datasets. a) Distribution of "A" and "B" compartments (from Hi-C and 4f-SAMMY-seq) across multiple chromatin states from an alternative human foreskin fibroblast dataset (E056 sample data from the Roadmap Epigenomics project). The stacked barplot shows the percentage of regions associated to concordant ("A-A" or "B-B") and discordant ("A->B" or "B->A") compartments classification for each chromatin state (labels on the upper margin indicating chromatin states and their total size). The chromatin compartments classification is reported using the same colouring convention adopted in Figure 2b. The barplot for 4f-SAMMY-seq compartments distribution is divided in three parts to report the specific result for the three replicates (C002_r1, C004_r1 and C004_r2, respectively). **b)** Relative occupancy of 4f-SAMMY-seq vs Hi-C based compartments for each chromatin state (E056 Roadmap Epigenomics sample), computed as the difference in "A" compartment percentage. For each chromatin state (columns, labels and total size for each state in the central row) positive values (green gradient) indicate a higher percentage of "A" compartment in 4f-SAMMY-seq based classification, whereas negative values (red gradient) indicate a higher percentage of "A" compartment based on Hi-C classification (i.e. higher "B" percentage based on 4f-SAMMY-seq). The size of each dot is proportional to the absolute value in the percentage difference. Chromatin states are ordered from left to right based on the average percentage difference across the three 4f-SAMMY-seq replicates (on the three rows). **c)** Violin and box plots showing for the reference Hi-C dataset and for individual 4f-SAMMY-seq replicates (labels on the upper margin) the distribution of Polycomb regulated chromatin states across genomic bins of the entire genome grouped by compartments classification ("A-A", "B-B", "A->B" or "B->A" defined as above). The upper violin plot is for bivalent Polycomb states (TssBiv, EnhBiv or BivFln) the bottom one for monovalent repressive Polycomb states (ReprPC or ReprPCWk). The number of genomic bins in each group is indicated at the bottom in parentheses (x-axis). In the overlaying boxplots, the horizontal lines mark the median, the boxes margins mark the interquartile range (IQR), and whiskers extend up to 1.5 times the IQR. Specific data points associated with HOX gene clusters are marked to show their positioning across groups and their associated high relative occupancy in Polycomb repressive and bivalent states. **d)** Scatter plots showing the relationship between first eigenvector values used to assign compartments (ranks on the x-axis) and the relative occupancy by chromatin states (percentage of each bin size) in squared root (sqrt) scale (y-axis). Each data point shows a 250kb genomic bin from the representative chromosome 2. Black dashed vertical lines mark the transition point between negative and positive eigenvalues (from "B" to "A" compartments). Chromatin states occupancy for bivalent Polycomb states (TssBiv, EnhBiv or BivFlnk - in the upper plot) and monovalent repressive Polycomb states (ReprPC or ReprPCWk - in the bottom plot) (from E056 Roadmap Epigenomics sample). A scatter plot with its trend line (lowess smoothing) is reported for the reference Hi-C dataset and for each 4f-SAMMY-seq replicate (labels on the upper margin).

e) Genome-wide mean signal enrichment (centred and scaled by chromosome) in the eight sub-compartments classification defined by CALDER, using as input the Hi-C (purple points and lines) and 4f-SAMMY-seq data (blue, light blue and green points and lines for three replicates), for Roadmap DNAse-seq and ChIP-seq datasets of human foreskin fibroblasts (E056 Roadmap Epigenomics). Sub-compartments are sorted from the most closed (left, B.2.2) to the most open one (right, A.1.1).

We also found that 4f-SAMMY-seq reclassifies some regions with a weakly active state, such as enhancers (Enh) and weak transcription (TxWk), as "B" compartment (Figure 3b-c, Supplementary Figures 4b, 5a-b and 6a), while it is mostly concordant with Hi-C for other more unambiguously active (Tx, TssA, EnhG) or inactive (Het) states.

Supplementary Figure 6



Supplementary Figure 6 – Compartments and sub-compartments analysis on low input 4f-SAMMY-seq. a) Relative occupancy of 4f-SAMMY-seq (starting from either 50k or 10k cells as indicated on the left side labels) vs Hi-C based compartments computed as the difference in "A" compartment percentage. For each chromatin state (columns, labels and total size for each state in the central row) positive values (green gradient) indicate a higher percentage of "A" compartment in 4f-SAMMY-seq based classification, whereas negative values (red gradient) indicate a higher percentage of "A" compartment based on Hi-C classification (*i.e.* higher "B" percentage based on 4f-SAMMY-seq). The size of each dot is proportional to the absolute value

in the percentage difference. Chromatin states are ordered from left to right based on the average percentage difference across the three 4f-SAMMY-seq replicates (on the three rows). **b)** Violin and box plots showing for the reference Hi-C dataset and for individual 4f-SAMMY-seq replicates starting from either 50k or 10k cells (labels on the upper margin) the distribution of Polycomb regulated chromatin states across genomic bins of the entire genome grouped by compartments classification ("A-A", "B-B", "A->B" or "B->A" defined as in Figure 2b). The upper violin plot is for bivalent Polycomb states (TssBiv, EnhBiv or BivFln) the bottom one for monovalent repressive Polycomb chromatin states (ReprPC or ReprPCWk). The number of genomic bins in each group is indicated at the bottom in parentheses (x-axis). In the overlaying boxplots, the horizontal lines mark the median, the boxes margins mark the interquartile range (IQR), and whiskers extend up to 1.5 times the IQR. Specific data points associated with HOX gene clusters are marked to show their positioning across groups and their associated high relative occupancy in Polycomb repressive and bivalent states. **c)** Genome-wide mean signal enrichment (centred and scaled by chromosome) in the eight sub-compartments classification defined by CALDER, using as input the Hi-C (purple points and lines) and 4f-SAMMY-seq data obtained starting from 50K and 10K (red, orange and yellow points and lines), for Roadmap DNAse-seq and Chip-seq datasets of human foreskin fibroblasts (E055 Roadmap Epigenomics). Sub-compartments are sorted from the most closed (left, B.2.2) to the most open one (right, A.1.1).

Overall, these data suggest that 4f-SAMMY-seq derived compartments may provide a more consistent reading on the properties of chromatin in terms of their connection to gene expression and chromatin states, including the Polycomb regulated domains, which are more challenging to be captured because of their dynamic nature (31).

SAMMY-seq also achieves sub-compartments definition.

To further dissect the quantitative relation between chromatin marks and compartments, we also computed sub-compartments using an adaptation of the Calder procedure (8) in both Hi-C and 4f-SAMMY-seq (see Methods) (Figure 5a). We called eight sub-compartments spanning from the most active A.1.1 to the most inactive B.2.2 compartment.

Figure 5

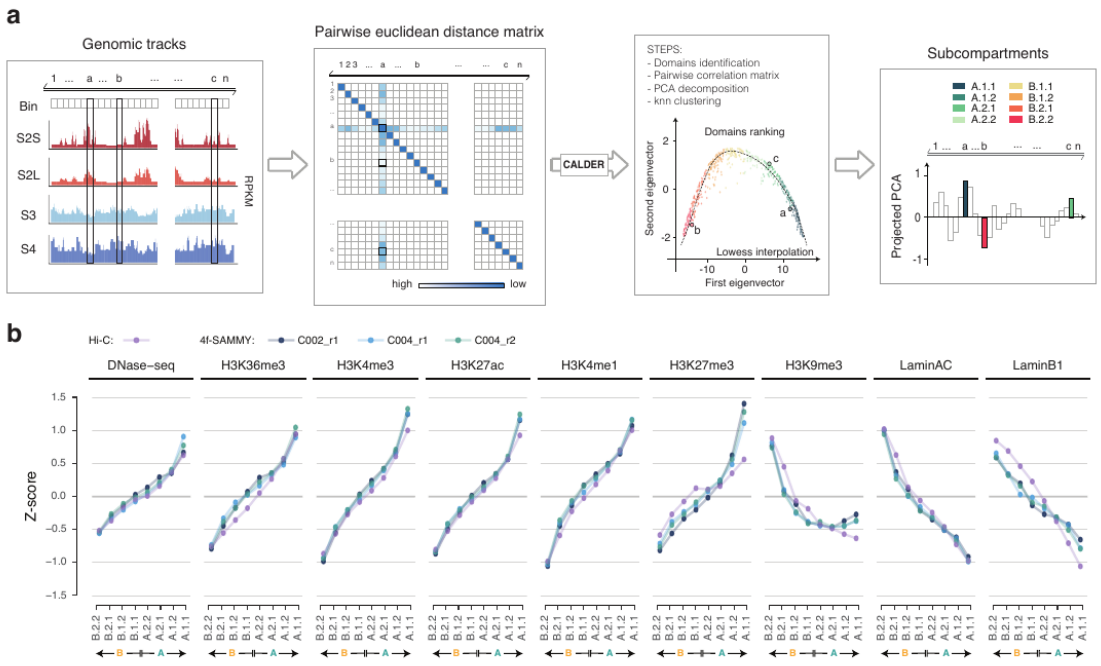


Figure 5 - 4f-SAMMY-seq allows detailed and reliable reconstruction of sub-compartments. **a)** Schematic illustration of the data analysis workflow to reconstruct chromatin sub-compartments from SAMMY-seq data. An adaptation of the CALDER procedure is applied on the pairwise Euclidean distance between genomic bins (see Methods). Starting from reads distribution profiles for individual chromatin fractions, for each pair of genomic bins the Euclidean distance is computed between the vectors of reads coverage across fractions. The method is then based on the ranking and clustering of domains (*i.e.*, group of bins) based on the lowess interpolation of the first and second eigenvector of a reduced matrix summarizing the properties of the bins. From the projected PCA values, the ranking of domains allows the discrimination of eight sub-compartments from the most compacted one (B.2.2.) to the most accessible one (A.1.1). **b)** Genome-wide mean signal enrichment (centred and scaled by chromosome) in the eight sub-compartments classification defined by CALDER, using Hi-C (purple points and lines) and 4f-SAMMY-seq data (blue, light blue and green points and lines for the three replicates), for Roadmap DNAse-seq and ChIP-seq datasets of human foreskin fibroblasts (E055 Roadmap Epigenomics). Sub-compartments are sorted from the most compacted (left, B.2.2) to the most accessible one (right, A.1.1).

Both Hi-C and 4f-SAMMY-seq sub-compartments are generally highly concordant in their association with individual chromatin marks (Figure 5b and Supplementary Figure 5e), the latter also considering 50k and 10k experiments (Supplementary Figure 6c). In fact, markers of active chromatin (DNAse-seq,

H3K36me3, H3K4me3, H3K27ac and H3K4me1) show the lowest enrichment in the most closed B2.2 sub-compartment and the highest enrichment in the most open A1.1 sub-compartment. Conversely, markers of constitutive heterochromatin (H3K9me3, Lamin AC and Lamin B) show the highest enrichment in B2.2 sub-compartment and a drop of enrichment going towards open sub-compartments. The most evident difference is again for regions marked by the Polycomb-dependent H3K27me3 histone mark. Here SAMMY-seq sub-compartments show more clear-cut H3K27me3 enrichment differences between the "A" and "B" sub-groups: this applies to both the most extreme A.1.1 vs B.2.2 and the intermediate A.2.2 vs B.1.1 compartment comparisons, as the latter shows a flattening in the H3K27me3 enrichment trend for the Hi-C-based sub-compartments definition (Figure 5b, Supplementary Figures 5e and 6c). The other main discrepancy pertains H3K9me3, for which SAMMY-seq sub-compartments are more precise in specifically confining its enrichment to the most inactive B.2.2 sub-compartment, although on the other hand the A.1.1 sub-compartment from SAMMY is not as depleted as the one from Hi-C (Figure 5b, Supplementary Figures 5e and 6c).

DISCUSSION

The genome is topologically and hierarchically organized on multiple levels in the nucleus. This 3D architecture has emerged as crucial feature for the regulation of transcription, replication, DNA repair and splicing (1). In particular, the physical separation between euchromatin, the more accessible and transcriptionally active part of the chromatin, and heterochromatin, the compacted and gene-poor part of the genome, is a hallmark of healthy cells. As such, morphological changes in genome organization are emerging as a pathological hallmark, and dysfunctional alterations of genome shape have been described in ageing, cancer and a plethora of other diseases (4,32-34). Silent chromatin can be further divided into facultative heterochromatin, regulated by the Polycomb group of repressing transcriptional factors (PcG), and constitutive heterochromatin, mainly localized at the nuclear periphery through the binding with components of the

nuclear lamina (35). Polycomb targets are heterogeneous from a functional and regulatory dynamics point of view. They include loci that are active only during development, cell cycle related genes and bivalent chromatin regions, the latter simultaneously marked by active and inactive histone modifications priming them for a rapid transcriptional activation. Moreover, different combinations of Polycomb complexes can modulate the transcriptional output as well as the context specific regulation of individual target loci (36). These features are challenging for the current methods that capture the chromatin structure.

Here we present a novel high-throughput sequencing-based protocol paired with a dedicated bioinformatic data analysis approach to single-handedly identify euchromatin and heterochromatin domains, as well as mapping their 3D compartmentalization inside the cell nucleus. The new 4f-SAMMY-seq method builds on our previously published 3f-SAMMY-seq technique, for which we have already demonstrated the ability to detect lamina-associated heterochromatin domains dynamics (16,19). Here we developed a new protocol whereby, with a combination of milder partial endonuclease digestion and size separation of the resulting DNA fragments, we can achieve higher resolution in mapping also accessible chromatin regions (Figure 1).

We then reasoned on the differences between our chromatin fractionation-based method and chromosome conformation capture techniques, such as Hi-C. Hi-C and other techniques mapping contacts between genomic loci are traditionally used to map chromatin compartmentalization. However, they are based on measuring the colocalization of genomic regions, without assessing directly their local environment in terms of associated macromolecules or subnuclear structures. On the other hand, SAMMY-seq derived methods rely on the biochemical isolation of distinct chromatin portions then assessing the differential enrichment of specific genomic domains across biochemical fractions. Based on the rationale that genomic loci close to each other will also be immersed in the same biochemical environment, we hypothesized that we could reconstruct

chromatin compartmentalization by directly measuring the similarity of biochemical properties across genomic loci with 4f-SAMMY-seq fractions.
Thus, we developed and present here dedicated bioinformatic algorithms to reconstruct the large-scale chromatin 3D organization in compartments and sub-compartments (Figure 2 and 5). The resulting compartments and sub-compartments are very similar to the ones derived from Hi-C, but with some notable differences. First of all, we show that 4f-SAMMY-seq-based compartments definitions are more consistently assigning transcribed genes to the "A" compartment (Figure 3). This is a crucial result as the first and most important discrimination between euchromatin and heterochromatin is determined by the transcriptional activity of these regions, whereas chromatin states definitions are associated afterwards to compartments, mostly correlating them to the transcription status.
In addition, our analyses indicate that 4f-SAMMY-seq is more efficient in capturing specific dynamic chromatin domains such as Polycomb regulated states (Figure 4 and 5). Namely, while Hi-C historically classifies Polycomb-regulated domains as belonging to the "A" compartment (5), we found some discrepancies (e.g. HOXD in Figure 4c,d). on the other hand, SAMMY-seq-based compartments provide a more consistent and concordant classification of Polycomb targets (Figure 4), including the highly dynamic bivalent genes (Figure 4a,b). Polycomb complexes have a central role in many physiological processes and their alterations in human diseases (37,38). In this context, the 4fSAMMY-seq is a versatile technology with unique characteristics and can be applied to several experimental models.

Moreover, the 4f-SAMMY-seq definition of chromatin compartments is achieved with several crucial practical advantages. Namely, while the standard Hi-C protocol requires at least 1 million cells, here we showed that our method works on as little as 10,000 cells, thus opening unprecedented opportunities in terms of applications to small biological samples (e.g. patients biopsies) or rare cell populations. Moreover, to achieve high resolution, Hi-C libraries are often

sequenced to more than one billion reads depth, whereas with 4f-SAMMY-seq we can use as little as 25 million reads per chromatin fraction (see Methods). Finally, the 4f-SAMMY-seq experimental protocol requires less than three hours in sample manipulation time, as opposed to at least two days for Hi-C. For all of these practical advantages, applying 4f-SAMMY-seq instead of more traditional Hi-C to map chromatin compartments would enable unparalleled scalability and applicability in the study of chromatin compartment dynamics.

DATA AVAILABILITY

The high-throughput sequencing data generated for this study are available in the NCBI GEO database with accession number “GSE226718”.

The code used to process SAMMY-seq data is available at Github/Bitbucket repository https://github.com/Clockris/SAMMY-method_4f.git. The code used to perform compartments and sub-compartments analysis on SAMMY-seq data is available at Github/Bitbucket repository https://github.com/ElisaSalviato/SAMMY-seq.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

AUTHOR CONTRIBUTIONS

FL, VR, FGo, PS, RQ and CL performed the experiments. CP, ESa and KP performed computational data analyses. EDPS, GL, IT, EP and ESe contributed to preliminary data analyses and testing of analyses procedures. LR, FGe and VV contributed to high-throughput sequencing. FL, CP, ESa, CL and FF wrote the manuscript. CL and FF jointly supervised the work and designed the study. All authors reviewed and approved the manuscript for submission.

ACKNOWLEDGEMENTS

We thank Beatrice Bodega, Gioacchino Natoli, Daniel Jost, Marina Lusic, Maria Vivo, Yinxiu Zhan, Vittore Scolari, Federica Marasca and all members of our laboratories for critical feedback on the manuscript. We acknowledge the use of dbGaP data from the study with accession number phs000791.v2.p1 (part of the Roadmap Epigenomics program). We acknowledge support from the Genomics Unit (Thelma Capra) of the European Institute of Oncology (IEO), Department of Experimental Oncology, from the Sequencing Unit (Giulia Passignani) of Fondazione IRCCS Ca’ Granda-Ospedale Maggiore Policlinico and Istituto Nazionale Genetica Molecolare (INGM) (Marco Ghilotti).

FUNDING

This work was supported by FRRB INTERSTRAT-CAD (grant #CP2_14/2018) and AIRC Start-Up (grant #16841) to FF; Cariplo Foundation (grant #2017-0649) to CL and FF; PRF (grant #2021-81) and MFAG (grant #18535) to CL; AIRC fellowships n. 26942 to FG, n. 2235 to ESa and n. 21012 to KP.

CONFLICT OF INTEREST

FL, ESe, FF and CL are co-inventor in a patent on the SAMMY-seq technique (European patent, EP3640330 A1), thus constituting a potential competing interest. The remaining authors declare no competing interests.

2.5 REFERENCES

1. Misteli, T. (2020) The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell*, **183**, 28-45.
2. Lupianez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R. *et al.* (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, **161**, 1012-1025.
3. Papathanasiou, S., Mynhier, N.A., Liu, S., Brunette, G., Stokasimov, E., Jacob, E., Li, L., Comenho, C., van Steensel, B., Buenrostro, J.D. *et al.* (2023) Heritable transcriptional defects from aberrations of nuclear architecture. *Nature*.
4. Johnstone, S.E., Reyes, A., Qi, Y., Adriaens, C., Hegazi, E., Pelka, K., Chen, J.H., Zou, L.S., Drier, Y., Hecht, V. *et al.* (2020) Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer. *Cell*, **182**, 1474-1489 e1423.
5. Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O. *et al.* (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, **326**, 289-293.

6. Pickersgill, H., Kalverda, B., de Wit, E., Talhout, W., Fornerod, M. and van Steensel, B. (2006) Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nat Genet*, **38**, 1005-1014.
7. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S. *et al.* (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665-1680.
8. Liu, Y., Nanni, L., Sungalee, S., Zufferey, M., Tavernari, D., Mina, M., Ceri, S., Oricchio, E. and Ciriello, G. (2021) Systematic inference and comparison of multi-scale chromatin sub-compartments connects spatial organization to cell phenotypes. *Nat Commun*, **12**, 2439.
9. Hsieh, T.S., Cattoglio, C., Slobodyanyuk, E., Hansen, A.S., Rando, O.J., Tjian, R. and Darzacq, X. (2020) Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol Cell*, **78**, 539-553 e538.
10. Krietenstein, N., Abraham, S., Venev, S.V., Abdennur, N., Gibcus, J., Hsieh, T.S., Parsi, K.M., Yang, L., Maehr, R., Mirny, L.A. *et al.* (2020) Ultrastructural Details of Mammalian Chromosome Architecture. *Mol Cell*, **78**, 554-565 e557.
11. Diaz, N., Kruse, K., Erdmann, T., Staiger, A.M., Ott, G., Lenz, G. and Vaquerizas, J.M. (2018) Chromatin conformation analysis of primary patient tissue using a low input Hi-C method. *Nat Commun*, **9**, 4938.
12. Falk, M., Feodorova, Y., Naumova, N., Imakaev, M., Lajoie, B.R., Leonhardt, H., Joffe, B., Dekker, J., Fudenberg, G., Solovei, I. *et al.* (2019) Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature*, **570**, 395-399.
13. Zheng, X., Hu, J., Yue, S., Kristiani, L., Kim, M., Sauria, M., Taylor, J., Kim, Y. and Zheng, Y. (2018) Lamins Organize the Global Three-Dimensional Genome from the Nuclear Periphery. *Mol Cell*, **71**, 802-815 e807.
14. Beagrie, R.A., Scialdone, A., Schueler, M., Kraemer, D.C., Chotalia, M., Xie, S.Q., Barbieri, M., de Santiago, I., Lavitas, L.M., Branco, M.R. *et al.*

(2017) Complex multi-enhancer contacts captured by genome architecture mapping. *Nature*, **543**, 519-524.

15. Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y. *et al.* (2018) Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell*, **174**, 744-757 e724.
16. Sebestyen, E., Marullo, F., Lucini, F., Petrini, C., Bianchi, A., Valsoni, S., Olivieri, I., Antonelli, L., Gregoretti, F., Oliva, G. *et al.* (2020) SAMMY-seq reveals early alteration of heterochromatin and deregulation of bivalent genes in Hutchinson-Gilford Progeria Syndrome. *Nat Commun*, **11**, 6274.
17. McCord, R.P., Nazario-Toole, A., Zhang, H., Chines, P.S., Zhan, Y., Erdos, M.R., Collins, F.S., Dekker, J. and Cao, K. (2013) Correlated alterations in genome organization, histone methylation, and DNA-lamin A/C interactions in Hutchinson-Gilford progeria syndrome. *Genome Res*, **23**, 260-269.
18. Kumar, A., Mazzanti, M., Mistrik, M., Kosar, M., Beznoussenko, G.V., Mironov, A.A., Garre, M., Parazzoli, D., Shivashankar, G.V., Scita, G. *et al.* (2014) ATR mediates a checkpoint at the nuclear envelope in response to mechanical stress. *Cell*, **158**, 633-646.
19. Frittoli, E., Palamidessi, A., Iannelli, F., Zanardi, F., Villa, S., Barzaghi, L., Abdo, H., Cancila, V., Beznoussenko, G.V., Della Chiara, G. *et al.* (2022) Tissue fluidification promotes a cGAS-STING cytosolic DNA response in invasive breast cancer. *Nat Mater*.
20. Sadaie, M., Salama, R., Carroll, T., Tomimatsu, K., Chandra, T., Young, A.R., Narita, M., Perez-Mancera, P.A., Bennett, D.C., Chong, H. *et al.* (2013) Redistribution of the Lamin B1 genomic binding profile affects rearrangement of heterochromatic domains and SAHF formation during senescence. *Genes Dev*, **27**, 1800-1808.
21. Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J. *et*

al. (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317-330.

22. Reiff, S.B., Schroeder, A.J., Kirli, K., Cosolo, A., Bakker, C., Mercado, L., Lee, S., Veit, A.D., Balashov, A.K., Vitzthum, C. *et al.* (2022) The 4D Nucleome Data Portal as a resource for searching and visualizing curated nucleomics data. *Nat Commun*, **13**, 2365.
23. Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.
24. Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754-1760.
25. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
26. Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F. and Manke, T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*, **44**, W160-165.
27. Kharchenko, P.V., Tolstorukov, M.Y. and Park, P.J. (2008) Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol*, **26**, 1351-1359.
28. Lawrence, M., Gentleman, R. and Carey, V. (2009) rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*, **25**, 1841-1842.
29. Hahne, F. and Ivanek, R. (2016) Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol*, **1418**, 335-351.
30. Boyle, A.P., Davis, S., Shulha, H.P., Meltzer, P., Margulies, E.H., Weng, Z., Furey, T.S. and Crawford, G.E. (2008) High-resolution mapping and characterization of open chromatin across the genome. *Cell*, **132**, 311-322.

31. Boettiger, A.N., Bintu, B., Moffitt, J.R., Wang, S., Beliveau, B.J., Fudenberg, G., Imakaev, M., Mirny, L.A., Wu, C.T. and Zhuang, X. (2016) Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature*, **529**, 418-422.
32. Bianchi, A., Mozzetta, C., Pegoli, G., Lucini, F., Valsoni, S., Rosti, V., Petrini, C., Cortesi, A., Gregoretti, F., Antonelli, L. *et al.* (2020) Dysfunctional polycomb transcriptional repression contributes to lamin A/C-dependent muscular dystrophy. *J Clin Invest*, **130**, 2408-2421.
33. Criscione, S.W., Teo, Y.V. and Neretti, N. (2016) The Chromatin Landscape of Cellular Senescence. *Trends Genet*, **32**, 751-761.
34. Yang, J.H., Hayano, M., Griffin, P.T., Amorim, J.A., Bonkowski, M.S., Apostolides, J.K., Salfati, E.L., Blanchette, M., Munding, E.M., Bhakta, M. *et al.* (2023) Loss of epigenetic information as a cause of mammalian aging. *Cell*, **186**, 305-326 e327.
35. van Steensel, B. and Belmont, A.S. (2017) Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell*, **169**, 780-791.
36. Kim, J.J. and Kingston, R.E. (2022) Context-specific Polycomb mechanisms in development. *Nat Rev Genet*, **23**, 680-695.
37. Liu, X. and Liu, X. (2022) PRC2, Chromatin Regulation, and Human Disease: Insights From Molecular Structure and Function. *Front Oncol*, **12**, 894585.
38. German, B. and Ellis, L. (2022) Polycomb Directed Cell Fate Decisions in Development and Cancer. *Epigenomes*, **6**.

Chapter 3 The impact of lamins on chromatin structure in mouse embryonic fibroblast

3.1 Abstract

The tridimensional chromatin structure is fundamental for the correct regulation of the genome function. Nuclear lamins are key players of the genome folding and participate in the maintenance of the chromatin structure. Lamins are divided in A-type and B-types, and have a plethora of functions in the nuclei, from mechanical properties, as resistance to forces and mechanotransduction, to the radial organization of chromatin in the nucleus. Lamin A and B interact with chromatin in specific heterochromatic regions called lamina associated domains (LADs). These domains are positioned in the periphery of the nucleus whereas the portion of chromosomes not attached to the lamina tend to go in the center where the euchromatin is preferentially found. However, the role of A and B lamins is different in the nuclei as evidenced by effects of genetic mutations. Using SAMMY-seq, we have previously characterized the effect of the mutated form of lamin A, progerin, on heterochromatin structure: in early passages fibroblasts from progeric patients we found a LAD remodeling with a detachment from the nuclear periphery. This early event leads to a deregulation of bivalent genes. Here we systematically analyze the conformation of both euchromatin and heterochromatin in the absence of lamin A or lamin B.

3.2 Introduction

The nuclear envelope (NE) is composed of a double layer membrane, and beneath is present the Nuclear Lamina (NL). The NL is a protein meshwork composed of a type V filament[1]. A and B lamins are the major constituents of the NL. A and B lamins differ for presence, nuclear positioning and functions[2,3]. Both

lamins bind heterochromatic portions of the genome called lamina associated domains (LAD)[4]. Some of the lamina-associated domains are stable in cells (defined as constitutive LADs). However, a portion of LADs change during lineage commitment and differentiation (classified as facultative LADs). The facultative LADs are cell-type specific and they determine the cell identity[4,5], by modifying the epigenetic repressive histone marks (HMs) in these regions and subsequently the transcriptional level of genes[6]. It is known that mutations of Lmna and Lmnb1/2 genes cause diseases called laminopathies[7,8]. These diseases are grouped by some defective form of lamin that accumulates and change the type of interactions with various proteins and factors[8]. However, the impact of lamins absence on the chromatin is unclear. Triple lamins KO in the mouse embryonic stem cells (mESC) model and HiC analysis suggested that lamins to not directly regulate TADs, but they are required to properly maintain inter- and intra- TAD interactions[9]. Moreover, the total absence of lamins in mESC also determines the decondensation of some facultative LADs[9]. We already reported that in some cases SAMMY-seq, based on chromatin solubility, can be more sensitive than HiC in capturing small differences in genome organization[10]. Moreover, to our knowledge, a direct comparison between Lamin A and B regarding their action the chromatin has never been addressed. We decided to analyze the role of lamins by using 4fSAMMY-seq (see Chapter 2) and immortalized mouse embryonic fibroblast (MEF) generated starting from wild type (wt), knockout for *Lmna* gene (ACKO) and double knockout for *Lmnb*1 and *Lmnb*2 (BDKO) mice.

3.3 4fSAMMY-seq reveal lamin-type specific chromatin remodeling

With this model we firstly evaluated the dynamic of the cell cycle, confirming by fluorescence-activated cell sorting (FACS) that both mutants show an alteration of the cell-cycle with a delay of the S phase (**Figure 1**).

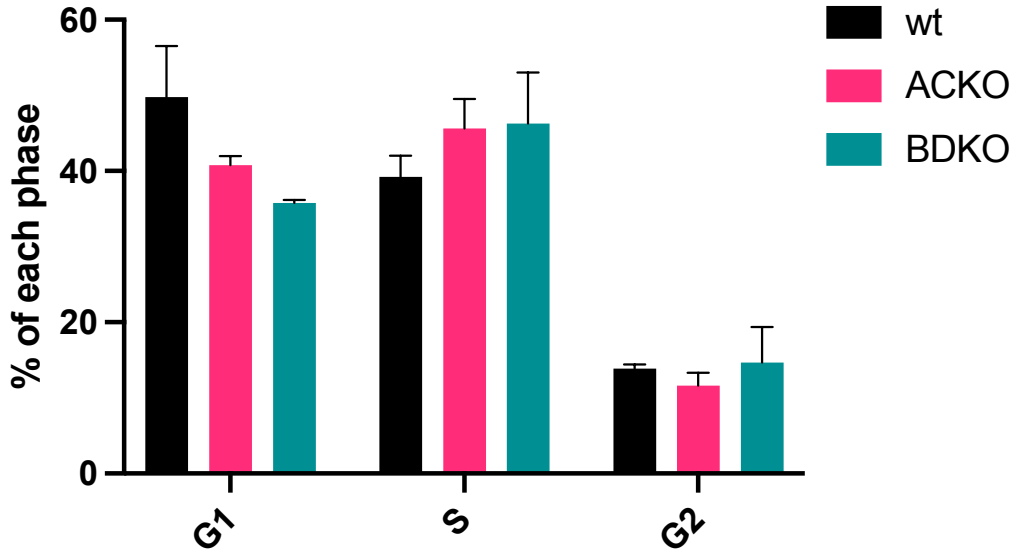


Figure 1 Cell cycle assay. Percentages of the number of cells in cell phase. The colors are black for wild type (wt), pink for ACKO and green for BDKO.

To characterize SAMMY-seq signal, we downloaded published histone marks (HM) ChIP-seq (H3K9me3, H3K27me3, H3K9ac, H3K27ac) from the short read archive (SRA) database[11] and we compared with S2SvsS3 (comparison) obtained from the differential enrichment of S2S and S3 alignments computed with SPP package[12]. To ensure maximum comparability, both histone mark IP/Input profiles and SAMMY-seq comparisons were analyzed using the same parameters. The euchromatin, identified by the H3K27ac presence is concordant with S2S enrichment of all three MEF samples (**Figure 2A,** signal above the line). On the opposite enrichment, the heterochromatic S3 regions appear to be partially covered with H3K9me3 and H3K27me3 (**Figure 2A,** signal below the line). Some of the insoluble regions captured by SAMMY-seq do not appear to have heterochromatic HM. However, it is known that stem zygotic stem cells and progenitors do not show high levels of constitutive heterochromatin as they are pluripotent cells[13–15]. They will acquire the constitutive H3K9me3 upon lineage differentiation[16,17]. Then, we calculated the spearman correlation matrix between

the comparisons profiles and histone marks (HMs) (**Figure 2B**), finding high genome wide correlations with active marks raging from 0.65 to 0,79 and lower anticorrelations with H3K9me3 enriched heterochromatin signal (between -0,25 to-.0,43). On the other hand, correlations matrix across different replicates shows high-level reproducibility raging from 0,79 to 9,92 (**Figure 2B**).

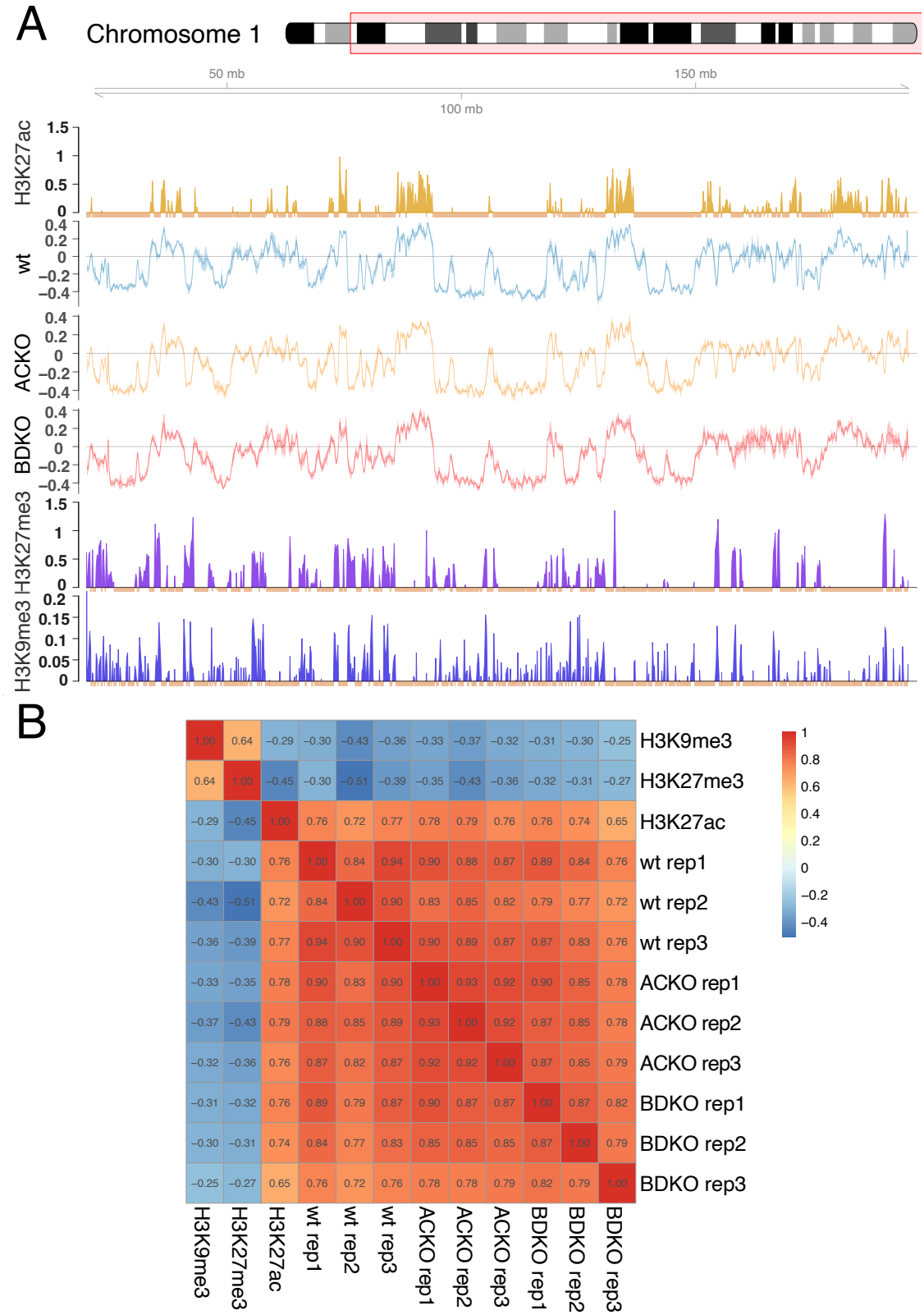


Figure 2. SAMMY-seq signal recapitulate chromatin accessibility state. Representative genomic region of chromosome 1 (chr1:20.000.000-end). **A.** The lines in light blue, orange and

red are the consensus track of 4fSAMMY-seq (S2SvsS3) for wt, *Lmna* (ACKO) and *Lmnb* (BDKO) Knock Out, respectively. Lighter color shadows for each track represent the standard deviation. Positive signal enrichments correspond to the more soluble S2S regions, whereas negative signal enrichments represent the less soluble fraction S3. The available ChIP-seq histone marks for wt MEF of euchromatin-associated H3K27ac (dark yellow), facultative heterochromatin-associated H3K27me3 (purple) and constitutive heterochromatin-associated H3K9me3 (blue) marks are shown. For ChIP-seq data, the y axis range is set to zero as minimum value. **B.** Spearman correlation for 4fSAMMY-seq (S2SvsS3) between all 4fSAMMY-seq samples and histone marks rebinned at 50Kb.

With the idea that A-type and B-type lamins have different roles in the chromatin organization, the absence of one at the time would lead to different chromatin changes. Firstly, we quantitatively explored the data by analyzing the consensus tracks of the S2SvsS3 enrichment. To remove the regions with a non-clear starting enrichment, we filtered out the regions with a mean signal less than ±0.1 value in the wild type, and therefore, the corresponding regions also in the mutants from the computation. The scatter plots show the similarity of the mutants respect to the wt. The maximum overlap expected is represented on the diagonal with an r^2 of 1. Although the entire distribution of chromatin is very similar across samples, we found an increased dispersion of the signal for the BDKO mutants with an r^2 of 0.899 for ACKO mutants and 0.848 for BDKO, suggesting that the BDKO mutant have a global stronger effect on chromatin solubility **(Figure 3 A, B)**.

Then, we decided to perform a differential enrichment analysis of the S2SvsS3 signals comparing each mutant (ACKO and BDKO) with the wt (**Figure 3 C**). We classified the differential soluble regions into 4 categories: S2S_up (red label),

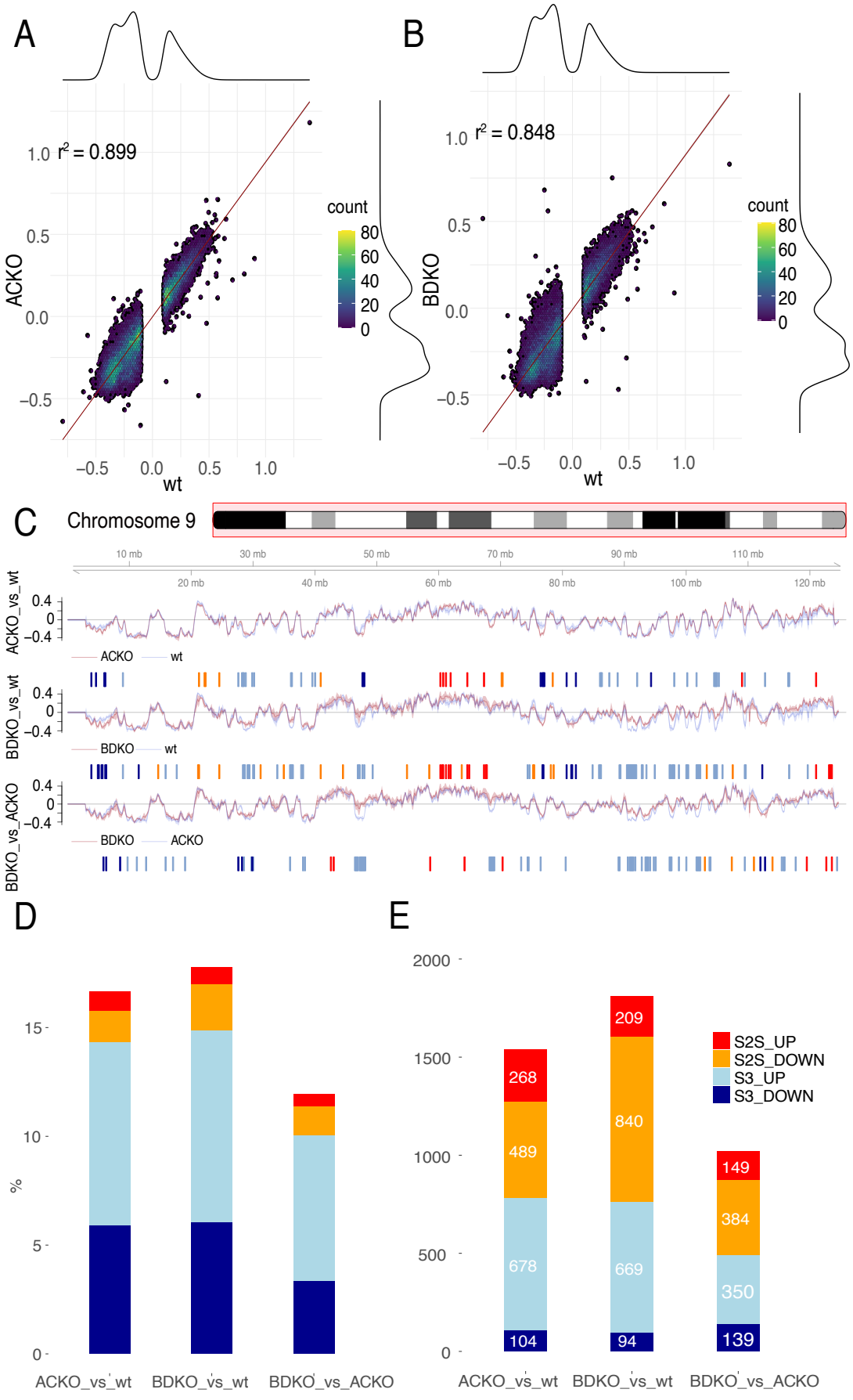
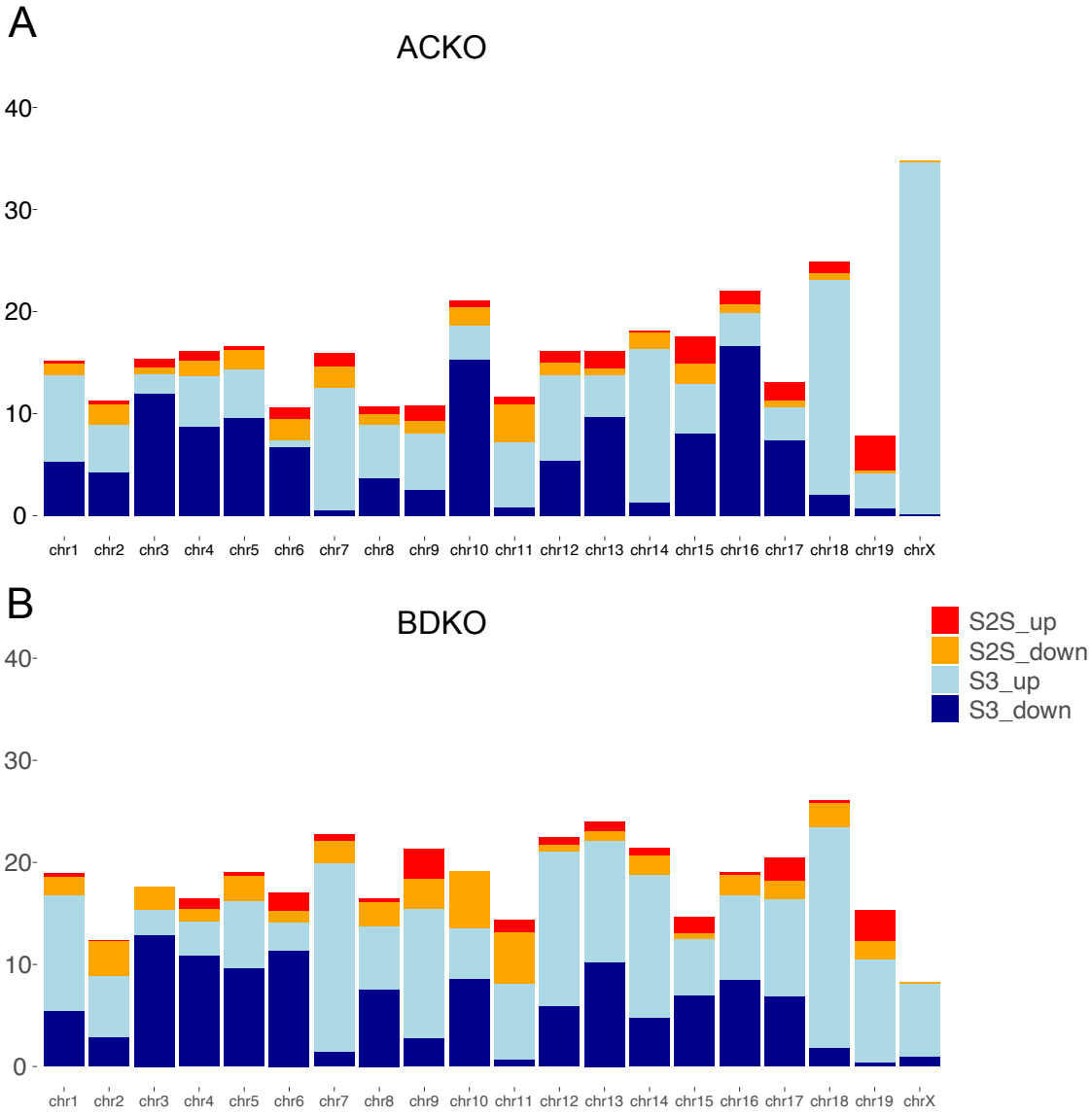

A
B
C
D
E
ACKO
BDKO
wt
count
r² = 0.899
r² = 0.848
Chromosome 9
ACKO_vs_wt
BDKO_vs_wt
BDKO_vs_ACKO
S2S_UP
S2S_DOWN
S3_UP
S3_DOWN
%
268
489
678
104
209
840
669
94
149
384
350
139

Figure 3. Chromatin remodeling is lamin-type dependent. Chromosome 9 is shown. **A and B**. Genome wide scatter plot of the rebinned consensus signal at 150Kb for wt and ACKO (**A**) and BDKO (**B**), after removal of all the bins below the signal threshold of ±0.1. The red line is the linear regression, the color code shows the density of the dots, from low amount (dark blue) to highest (yellow). The lines on top and the right represent the distribution of the signal values. **C**. Chromosome 9 is shown. The lines in blue and red are the consensus track of 4fSAMMY-seq. From the top ACKO vs wt; BDKO vs wt and BDKO vs ACKO. Lighter color shadows for each track represent the standard deviation. Below each pair tracks, the significant regions are highlighted as follow: S2S_up (red), S2S_down (orange), S3_up (light blue) and S3_down (dark blue). **D**. Stacked bar plot of the percentages of genome affected by remodeling after filtering out the low signal > ±0.1 with the differential enrichment analysis. **E**. Stacked bar plot of the number of genes that have the TSS overlapping with the regions with a range of upstream 2500bp and downstream 500 bp. Both **D** and **E** share the same color code of the **C** panel.

regions of S2S with increased accessibility in the mutant; S2S_down (orange label), regions of S2S with decreased accessibility in the mutant; S3_up (light blue label), regions of S3 with increased accessibility in the mutant and S3_down (dark blu label), regions of S3 with decreased accessibility in the mutant. Notably, with this analysis we select only genomic regions with different solubility in all analyzed mutants versus all wt samples. With the same pipeline we also compared the two ACKO and BDKO mutants. As expected by their functional role, the results show that both mutants show a strong heterochromatin remodeling (**Figure 3D**). In general, chromatin regions of the two mutants tend to lose a clear state of inaccessibility, going to a less defined state. Intriguingly, as evidenced by direct ACKO and BDKO comparisons, the affected regions are not coincident, highlighting the diverse role of lamins on genome folding.

We also checked if the remodeled regions were chromosome dependent. We found that the chromosome X in the A mutant, but not in the Lamin B mutant, displays an extensive lack of heterochromatin insolubility, showing a sharp increase of S3_up regions (**Supplementary Figure 1**). Interestingly X chromosome includes another important gene responsible for Emery Dreifuss Muscular Dystrophy: the emerin. Thus, these data suggest that the absence of Lamin A directly lead to an alteration of X chromosome compaction. The same

experiment executed in muscular cells could elucidate if this alteration can be linked to emerin dysregulation and mechanisms of dystrophy. This is important as when emerin is mutated it give the same phenotype of the emery dreifuss muscular dystrophy (EDMD)[19]. The other chromosomes whom appear to be slightly more affected than the others in loosing heterochromatin in both mutants are the chr7, chr18. loose On the other hand, the remaining chromosomes show smaller differences in quantities.



Supplementary Figure 1. Percentages of chromosome affected regions. The percentages are calculated from the total chromosome size analyzed after the removal of the genomic bins

with a signal < ±0.1. Color scheme: S2S_up (red), S2S_down (orange), S3_up (light blue), S3_down. **A.** ACKO mutant **B.** BDKO.

As expected, in S2 fraction that correspond to euchromatin, the number of gene involved in the solubility change is higher than in heterochromatin (**Figure 3D**), with BDKO affecting more genes than ACKO. In the direct comparison of B and A mutants, the differences appear slightly less pronounced, particularly in the S3 portion of the genome (**Figure 3 A B D**), accordingly to their functional common role at the NL.

Then, we calculated the degree of overlap between ACKO and BDKO altered regions by Jaccard score calculations. Globally, some accessibility alterations are retained in both mutants (20% for accessible and between 30% and 40% for inaccessible regions), thus confirming their common role on chromatin conformation (**Figure 4**). On the other hand, other regions are affected differently in both lamin types as evidenced in the comparison with wt or in the direct ACKO BDKO comparison.

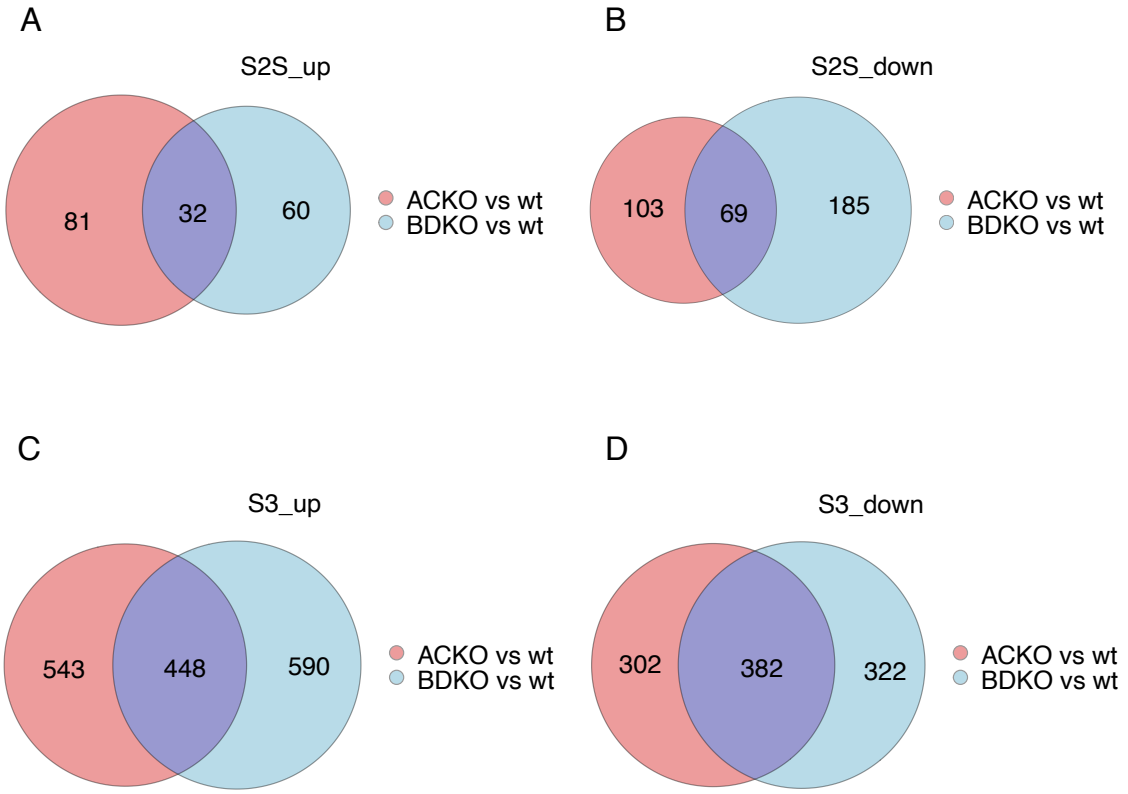


Figure 4. Venn diagrams of altered regions in each category. A. Venn diagram of S2S_up regions with a Jaccard score of 0.185 in BDKO vs wt and ACKO vs wt **B.** Venn diagram of S2S_down regions with a Jaccard score of 0.193 in BDKO vs wt and ACKO vs wt **C.** Venn diagram of S3_up regions with a Jaccard score of 0.283 BDKO vs wt and ACKO vs wt **D.** Venn diagram of S3_down regions with a Jaccard score of 0.38 BDKO vs wt and ACKO vs wt

Interestingly, the number of the common ACKO and BDKO regions differently captured by solubility respect to the wt is higher in the heterochromatin then euchromatin, accordingly with the major role of lamins in heterochromatin organization (**figure 4**). In order to address the functional role of chromatin solubility shift seen in ACKO and BDKO, we characterized the biological pathways subjected to the deregulated regions. Since available gene ontology packages are tailored on RNA-seq, to avoid inflated P-values, we set up a custom gene ontology analysis based on permutation (**see methods**).

We used this GO analysis on differentially soluble regions (S2S_up, S2S_down, S3_up, and S3_down). Despite the comparable total number of genes GO was significantly enriched only in euchromatin regions (S2S up/down). This could be due to a spreading of different genes in heterochromatin (S3 up/down), where altered regions cover larger genomic portions. In order to facilitate the view and interpretation, we did a semantic analysis with revigo[20], for each group filtering all the GOs with a false discovery rate of < 0.01 (**Figure 5**). In the analysis of euchromatin remodeling we found some pathways shared between the two mutant, as cell cycle, cell proliferation and cell death (**Figure 5 A, B**), confirmed by cell proliferation assay (**Figure 1**). The most notable clusters in the semantic analysis of S2S_up regions specific of the ACKO suggest a direct regulation on development and differentiation (**Figure 5 A**). In the other hand, BDKO GO suggests a specific positive regulation of chromatin dynamics, as the histone methylation on K4 (**Figure 5 B**). In line with previous findings[7,8,21,22], these first analyses suggest different roles in regulation of euchromatin with Lamin B involved in chromatin structure and Lamin A in cell specification. Moreover, data in BDKO suggests that lamin B KO generate specific defects in epigenetic regulation.

Analyses of S2S_down regions were more difficult, because of several GOs terms. Thus, we separated regions in 3 categories based on the Venn diagram (**Figure 4**): ACKO and BDKO specific S2S_down regions and the overlapping group (**Figure 5 C, D, E**). In ACKO, the S2S_down related terms were largely described as associated to Lamin A[23,24], including actin filament and cytoskeleton formation, protein regulation and effect on RNA splicing capacity (**Figure 5 C**). In S2S_down specific of the BDKO, we found several terms related to signaling as cellular response to cytokine. Interestingly, among signaling pathways we found the MAPK cascade containing TOR and TORC1 previously described in the lamin B dependent autophagy regulation[25]. Finally, we analyzed the shared regions, finding terms related to import/export of cells, regulation and processing of RNA and inflammation (**Figure 5 E**).

A

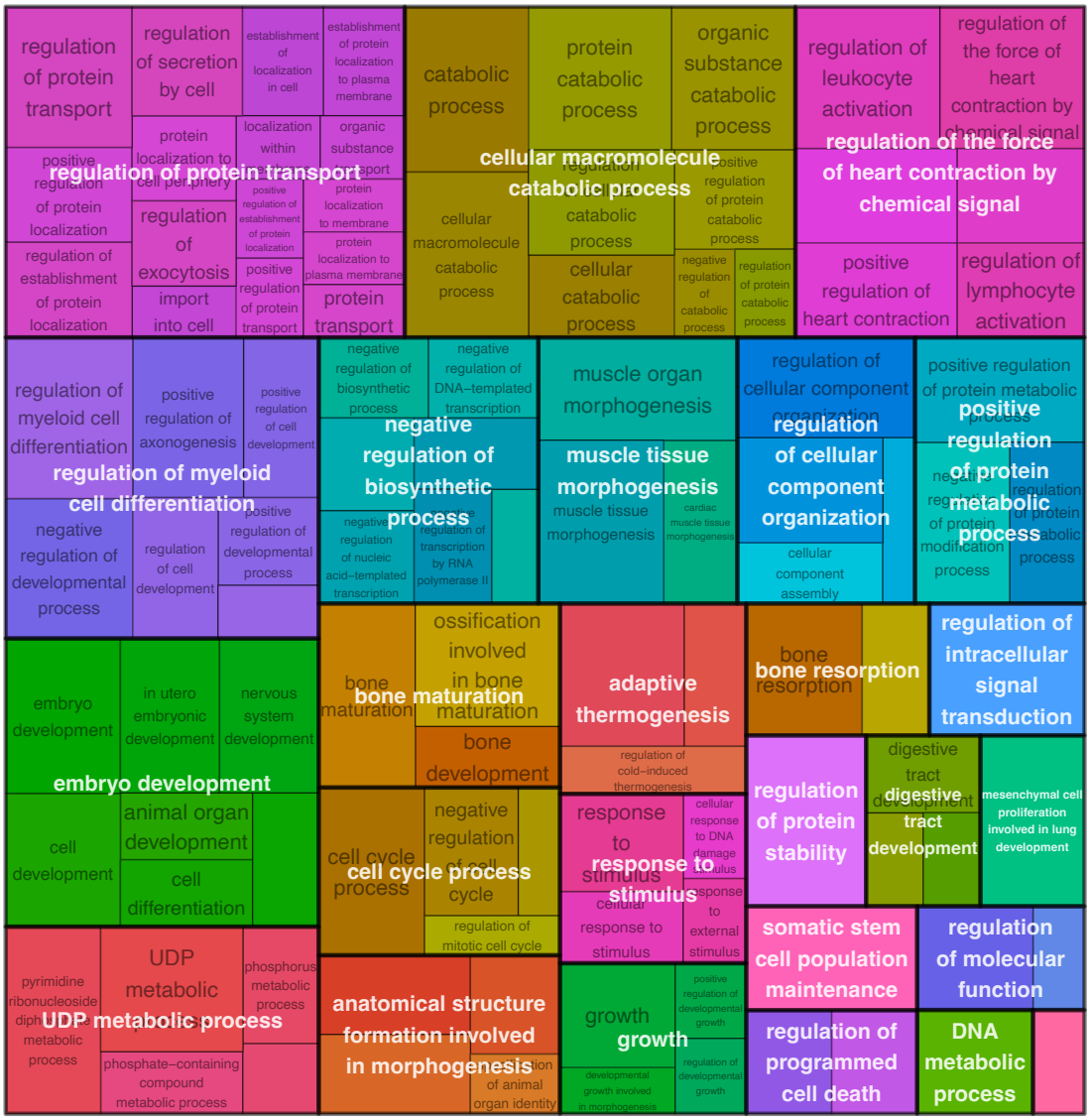


Figure 5 continues on the next page.

B

Figure 5 continues on the next page.

C

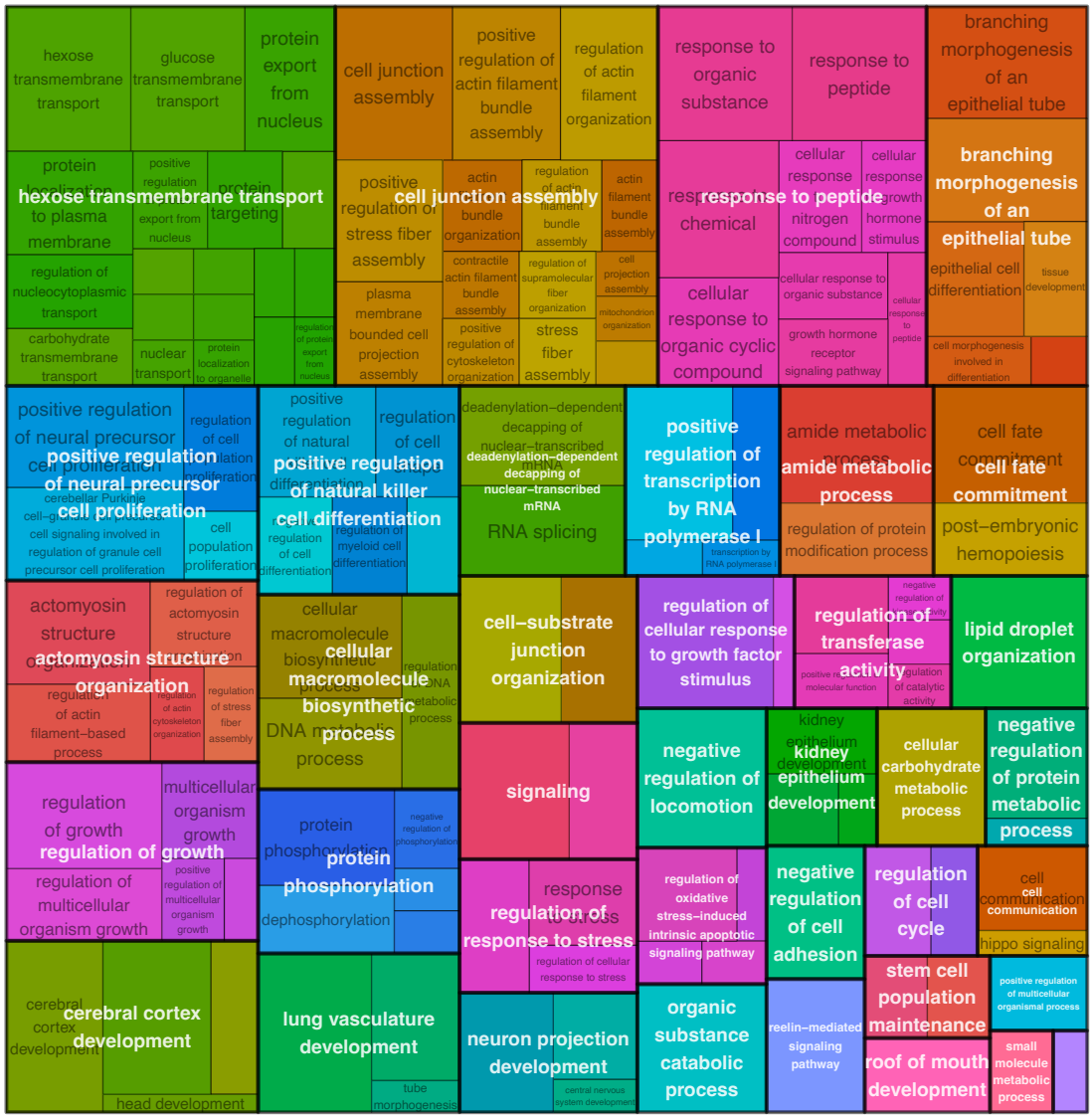


Figure 5 continues on the next page.

D

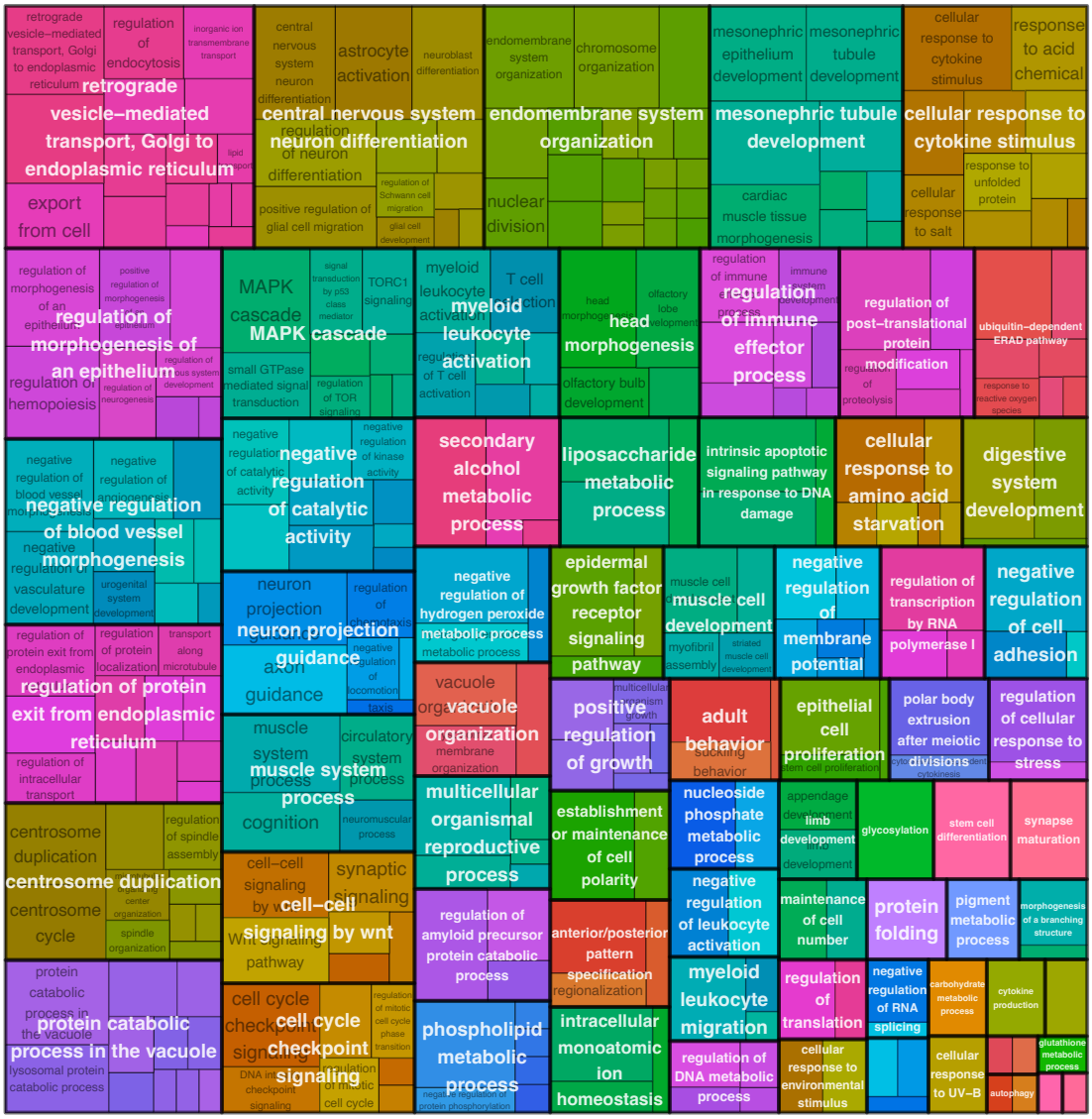


Figure 5 continues on the next page.

E

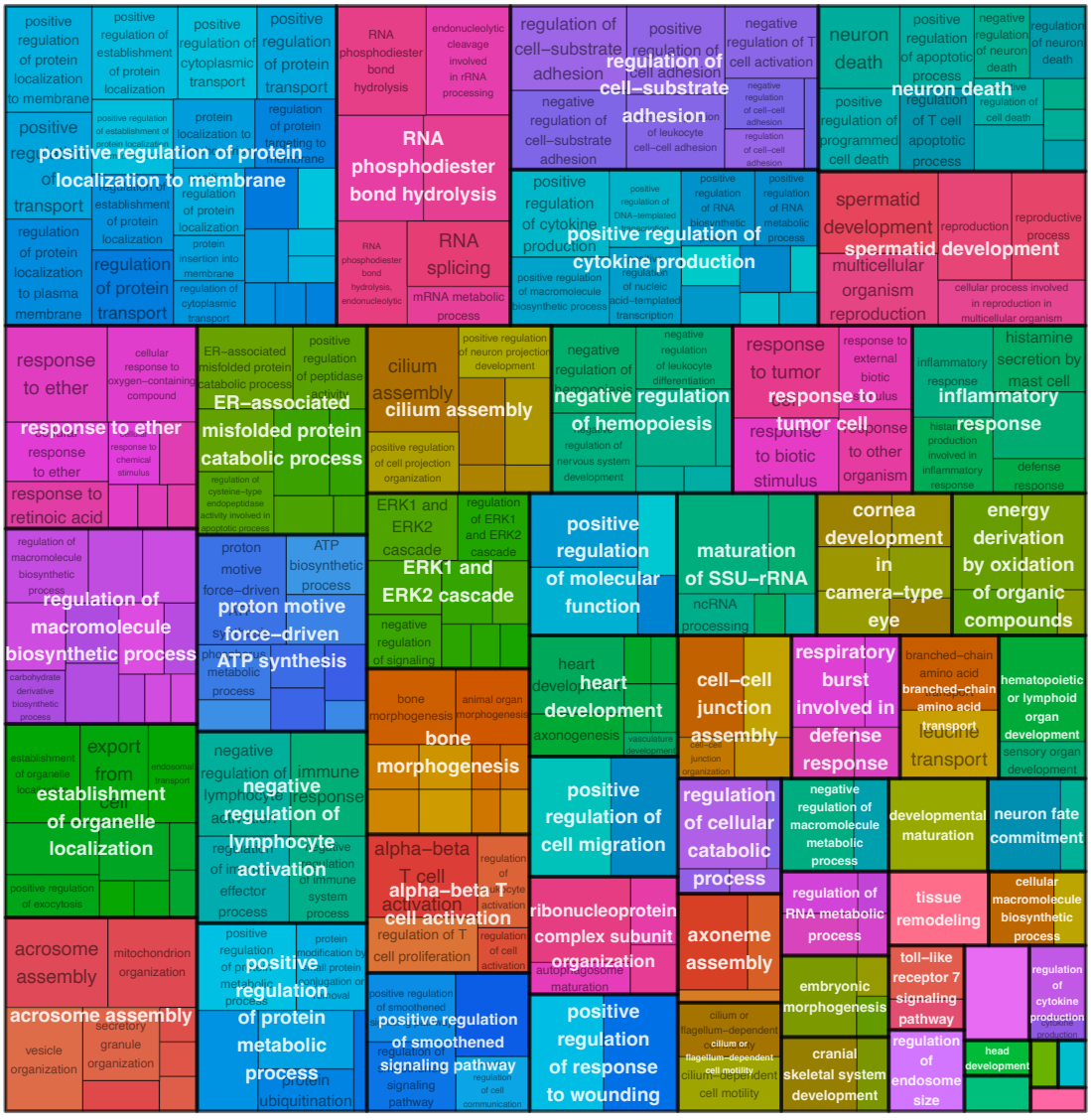


Figure 5. Semantic analysis of the Gene Ontologies derived from the genes inside remodeled genomic regions. A. S2S_up in ACKO mutant. **B.** S2S_up in BDKO mutant. **C.** S2S_down remodeled only in the ACKO mutant **D.** S2S_down remodeled only in the BDKO mutant **E**. S2S_down with same remodeling in both mutants.

To evaluate the impact of chromatin accessibility on transcription, we performed RNA-seq analysis on the same cell lines. We aligned, quantified the reads abundances and calculate the differential expressed genes with DESeq2 (**see**

methods) comparing ACKO and BDKO with the wild type. We obtained 610 Differential Expressed Genes (DEGs) in ACKO (417 overexpressed genes and 193 repressed) and 920 (602 overexpressed and 318 repressed) DEGs in BDKO, suggesting again a stronger effect of BDKO respect to ACKO. Interestingly, 30% genes upregulated and dowregulated were in common between ACKO and BDKO mutants (Jaccard score 0.35 for upregulated and 0.33 for down regulated genes). These finding suggest that a subset of genes are more sensible to the lamin dependent heterochromatin remodeling.

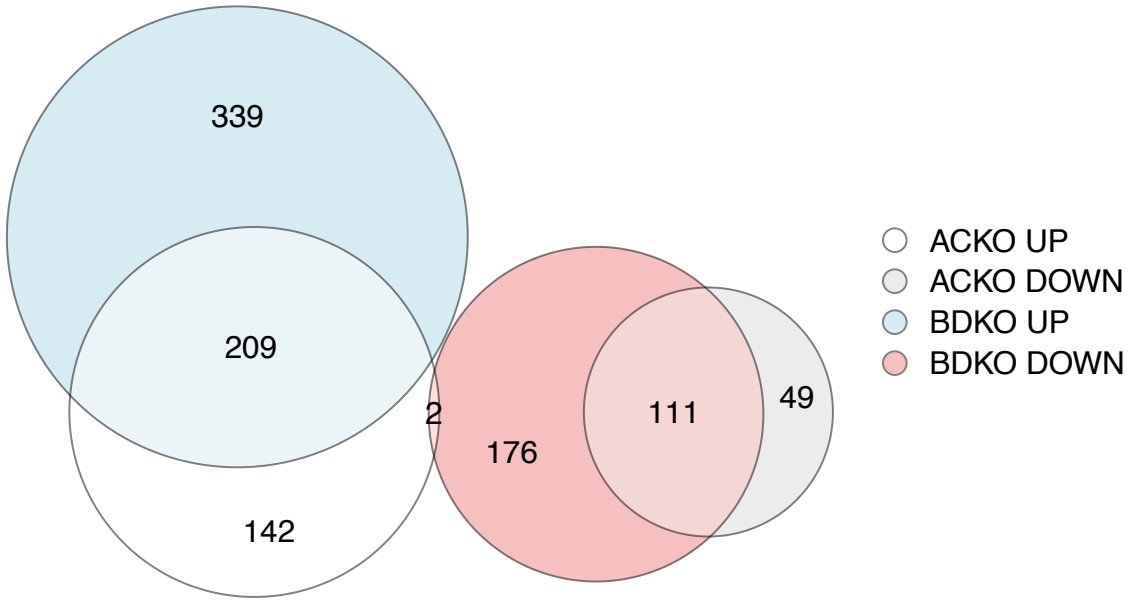


Figure 6 continues on the next page.

B

Figure 6 continues on the next page.

C

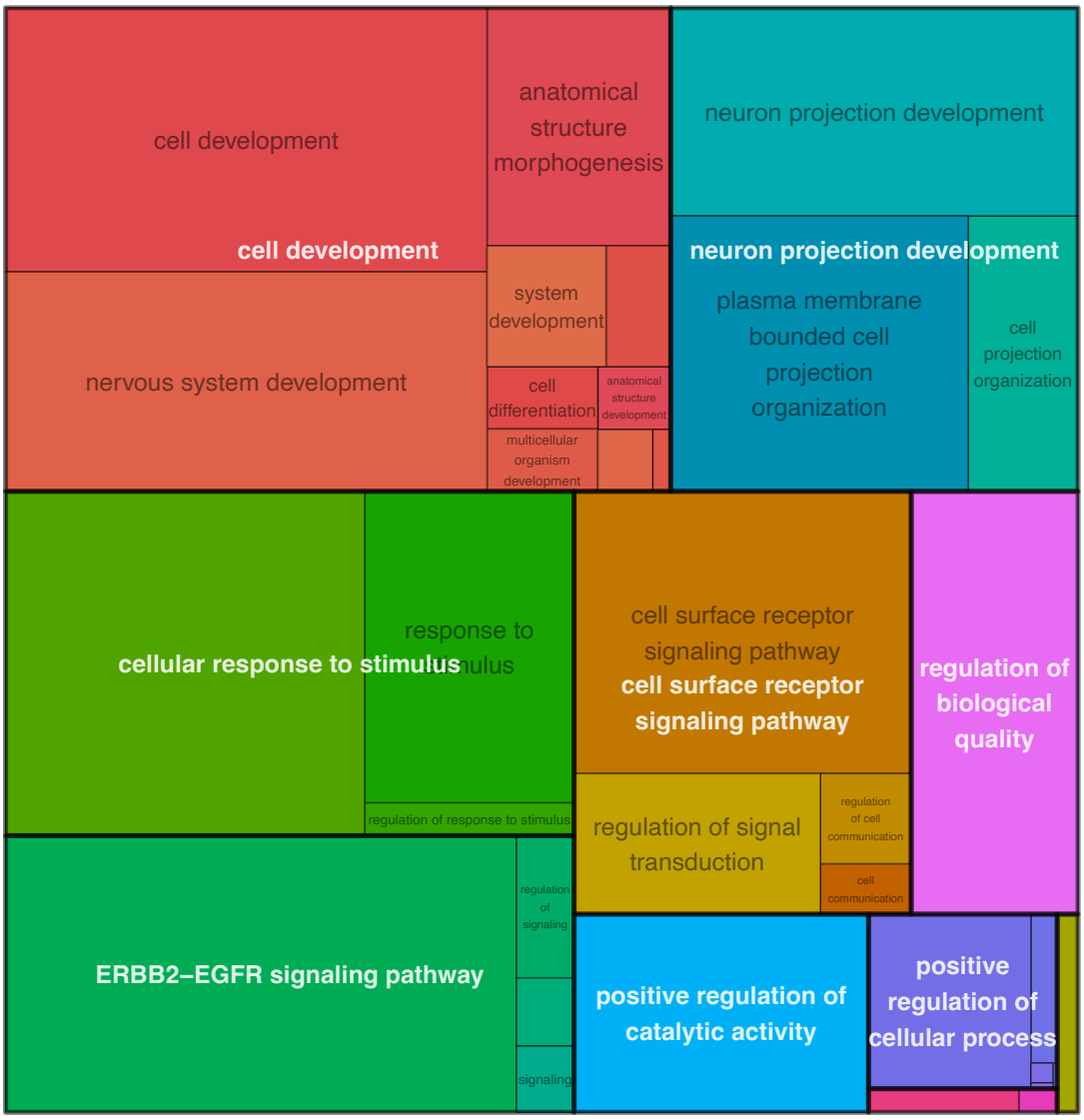


Figure 6 continues on the next page.

D

Figure 6 continues on the next page.

E

Figure 6 continues on the next page.

F

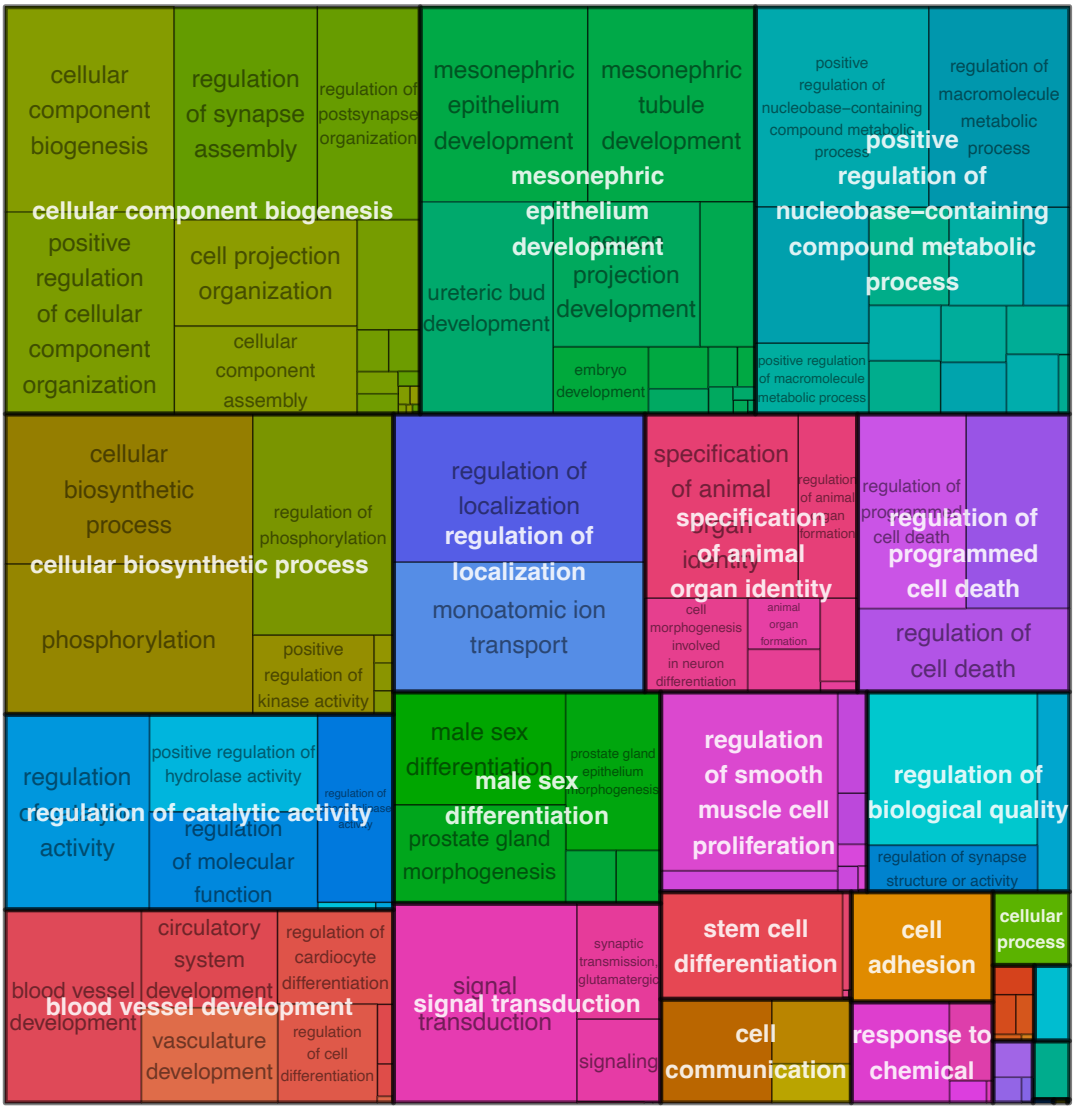


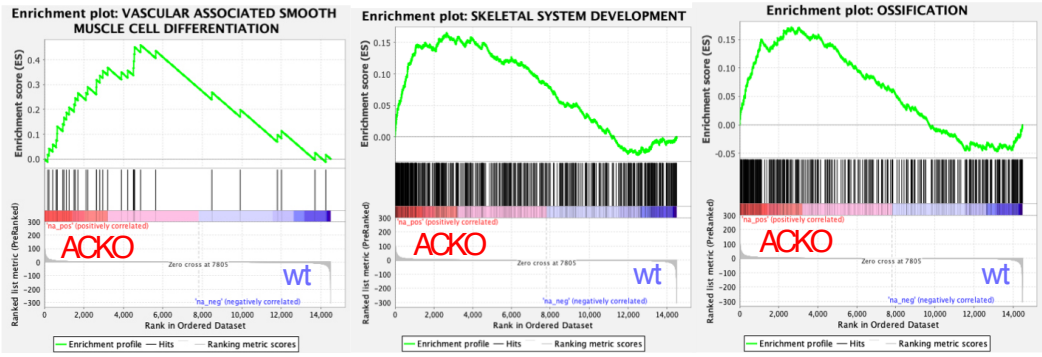
Figure 6. Overlap between DEG genes and corresponding gene ontologies A. Venn diagram of DEGs of the comparisons ACKO and BDKO vs wt **B.** Common upregulated genes in ACKO and BDKO **C.** Common downregulated genes in ACKO and BDKO **D.** Genes uniquely upregulated in ACKO **E.** Genes uniquely upregulated in BDKO **F.** Genes uniquely downregulated in BDKO. The gene ontology were calculated with g:profiler2[26], Pvalues were corrected with Benjamini Hochberg and the semantic analysis were perfomed with rrvgo[27].

The upregulated genes in common are mostly related to RNA biosynthesis and morphogenesis in various tissues (**Figure 6 B**), while the downregulated are more related with development, external stimuli and cell receptors (**Figure 6 C**).

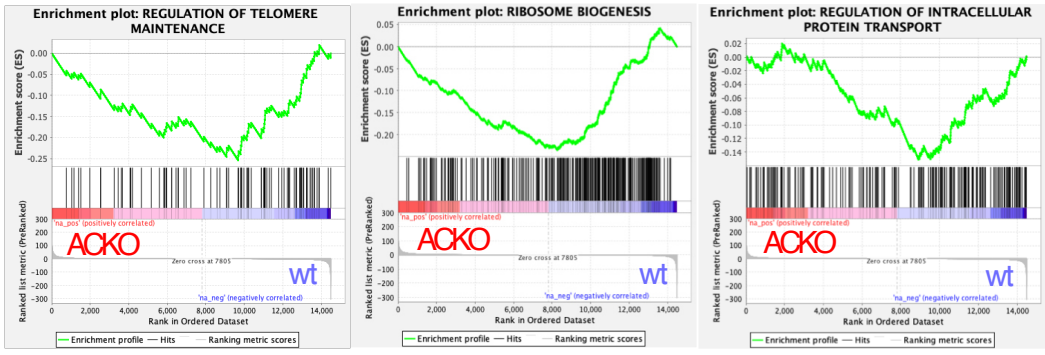
This suggest that both mutants lose the proper cell identity with a both positive and negative unbalance of differentiating pathways. This is in line with the diminished ability to correctly catch external signals to accurately differentiate [28]. Mutant specific DEGs shows that Lamin A is more involved in mechanism for cell positioning, controlling pathways of cell adhesion and motility (**Figure 6 D**) while Lamin B is a regulator of several signaling pathways and inflammation (**Figure 6 E**). Finally, BDKO specific downregulated genes are involved in cell development and differentiation (**Figure 6 F**).

To identify specific biological pathways in the two mutants ad get deeper understanding of their deregulation, we ranked the genes with the score generated with -log(Pvalue) * sign of the fold change. With this we did a preranked GSEA analysis with 1000 permutations on the biological pathways downloaded on g:profiler site. We selected 3 representative GSEA per condition negatively and positively correlated with the mutants.

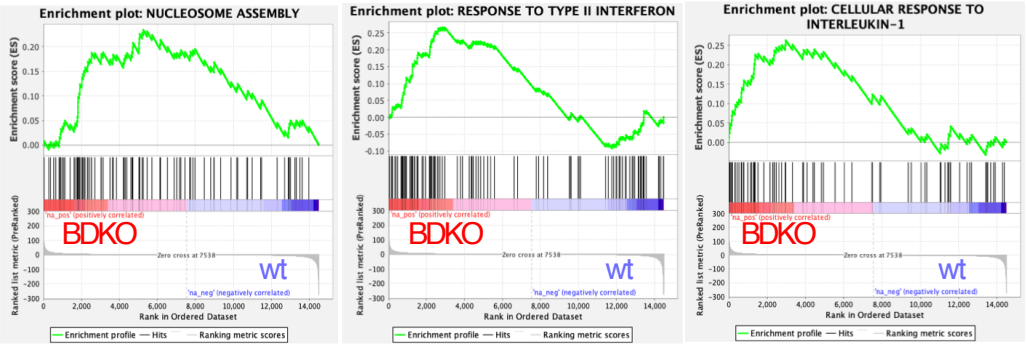
A



B



C



D

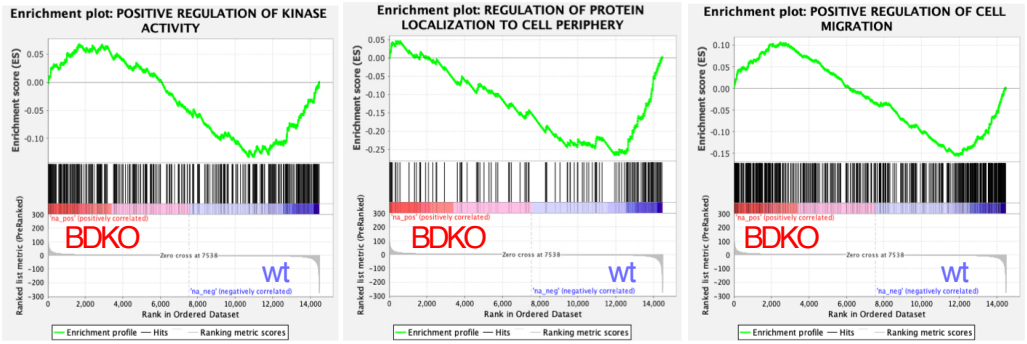


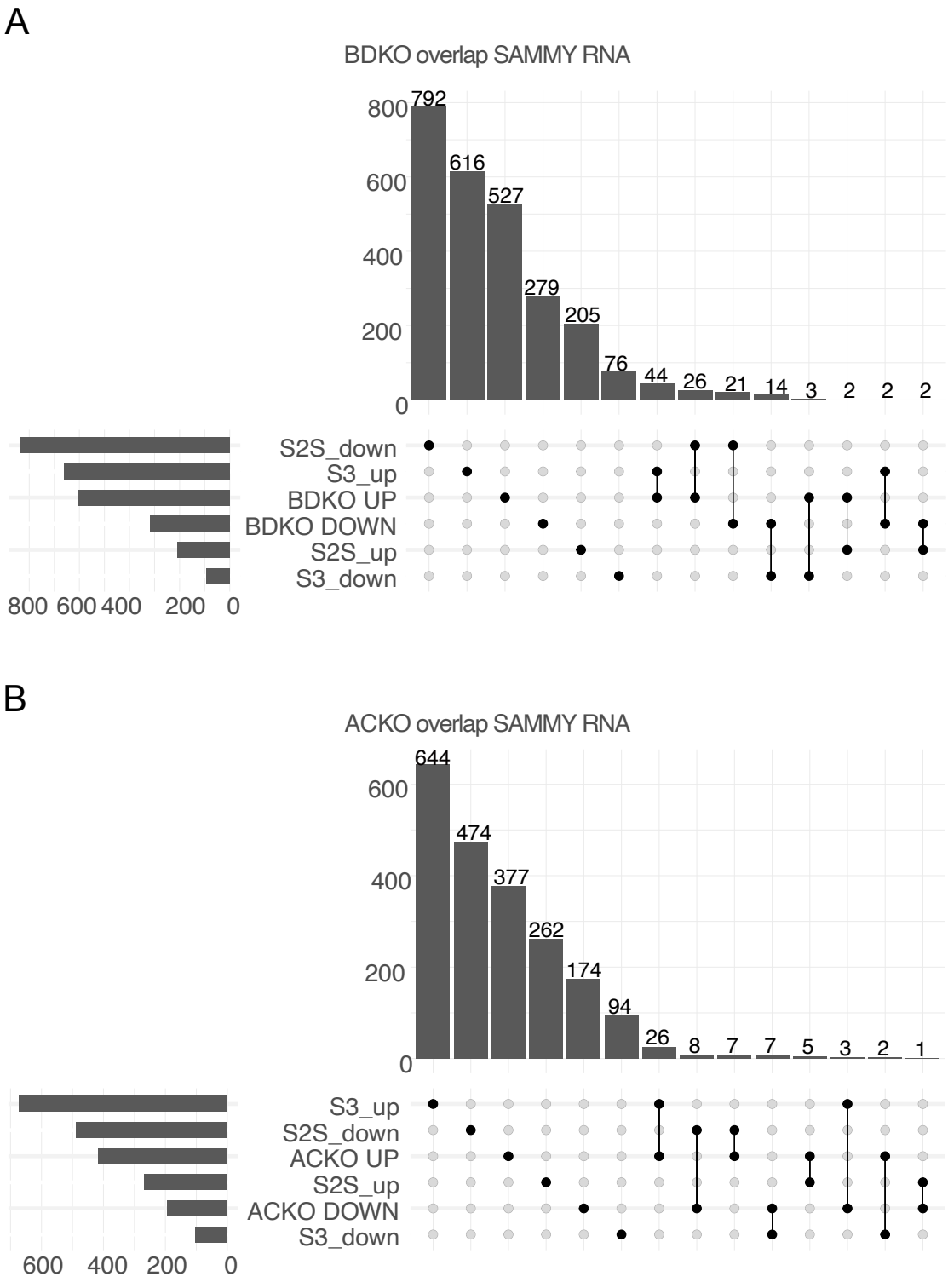
Figure 7. GSEA enrichment plots of biological pathways affected in lamin's KOs. Results from RNA-seq GSEA. **A.** Enriched and **B.** depleted pathways in ACKO **C.** Enriched and **D.** depleted pathways in BDKO.

We found, in ACKO, a pressure toward skeletal muscle differentiation and ossification as already described[2,29] (**Figure 7 A**) and a downregulation of transcripts processing and protein transportation (**Figure 7 B**). We also find a positive regulation of vascular associated smooth muscle differentiation, one of the tissues more affected in progeria[30]. Surprisingly, among the significant GSEA in ACKO downregulation we found the pathways involved in telomerase maintenance. This was not expected since used MEF are immortalized. However, given the important role of lamin A in the telomere protective cap formation and maintenance[31,32], it is possible that in the ACKO mutant there is an increased telomere attrition.

In line with gene ontology BDKO mutant shows an upregulation of inflammation (**Figure 7 C**) and a downregulation of cell signaling (**Figure 7 D**). Moreover,

Lamin B is necessary for the correct regulation of nucleosome assembly, in line with analysis performed on SAMMY-seq solubility (**Figure 3C, 5B**). These results show that most of the transcription levels are closely related to the biological pathways captured by SAMMY-seq differential solubility.

Then, we asked how many DEGs lies in the statistically significant, differential solubility regions shown in **Figure 3**. Although the two distinct analysis highlighted similar GO categories for Lamin A and B mutants, we found a minimal overlap between DEGs and SAMMY-seq changes (**Supplementary Figure 2**), suggesting that chromatin solubility alterations does not directly lead to a transcriptional deregulation.



Supplementary Figure 2. Accessibility remodeling does not directly correlate with transcriptional dysfunction. **A. B.** Upset plot of the genes between remodeled categories (S2S and S3 both up and down) and DEGs (upregulated downregulated) in ACKO (A) and BDKO (B).

On the other hand, we wondered if DEGs display a different accessibility profile. We made a metaprofile of the correspondent S2SvsS3 regions rebinned at 2Kbp (**see methods**). Then, we plotted a heatmap of the solubility signal (**Figure 8 A, C**) for each mutant.

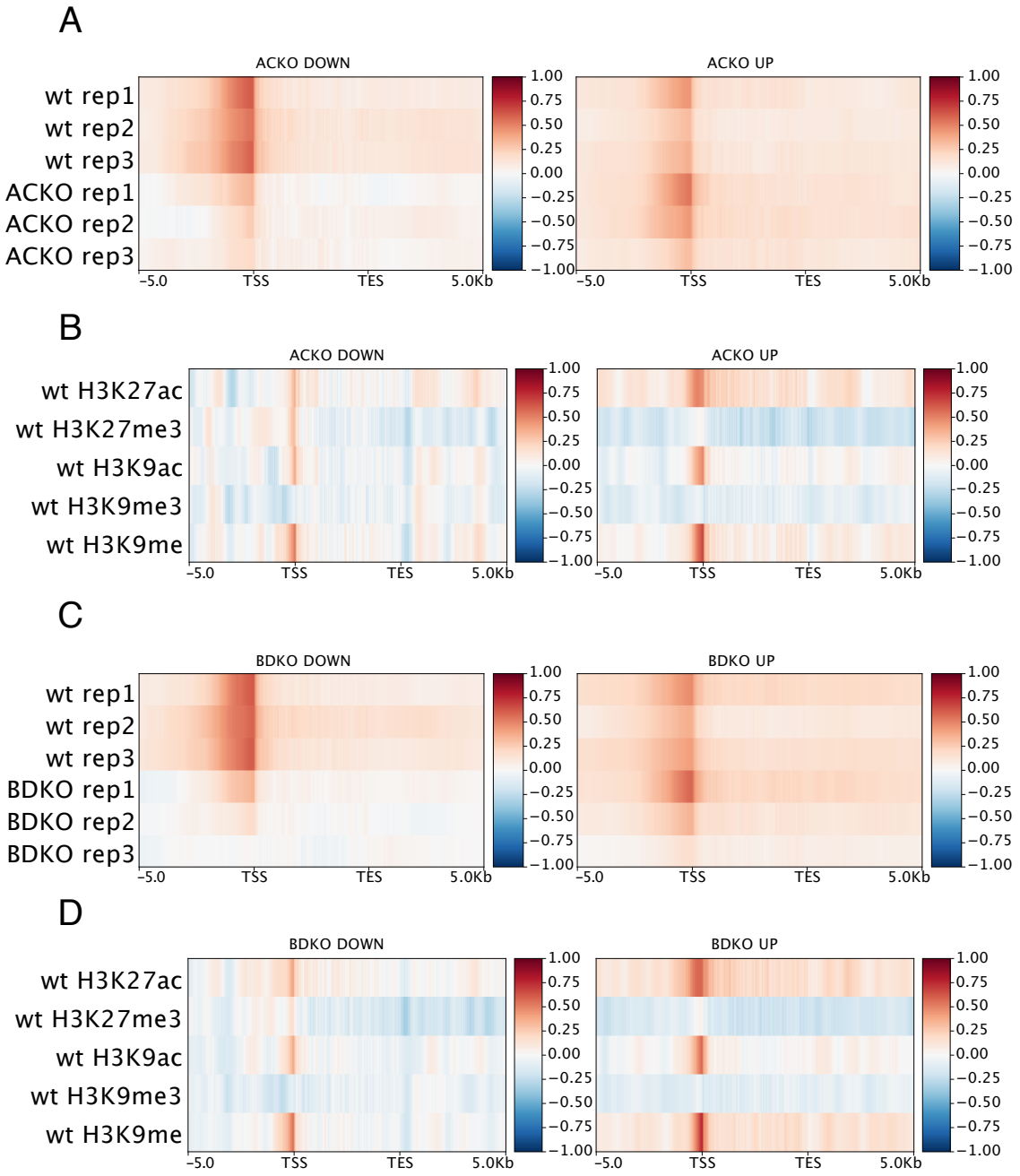
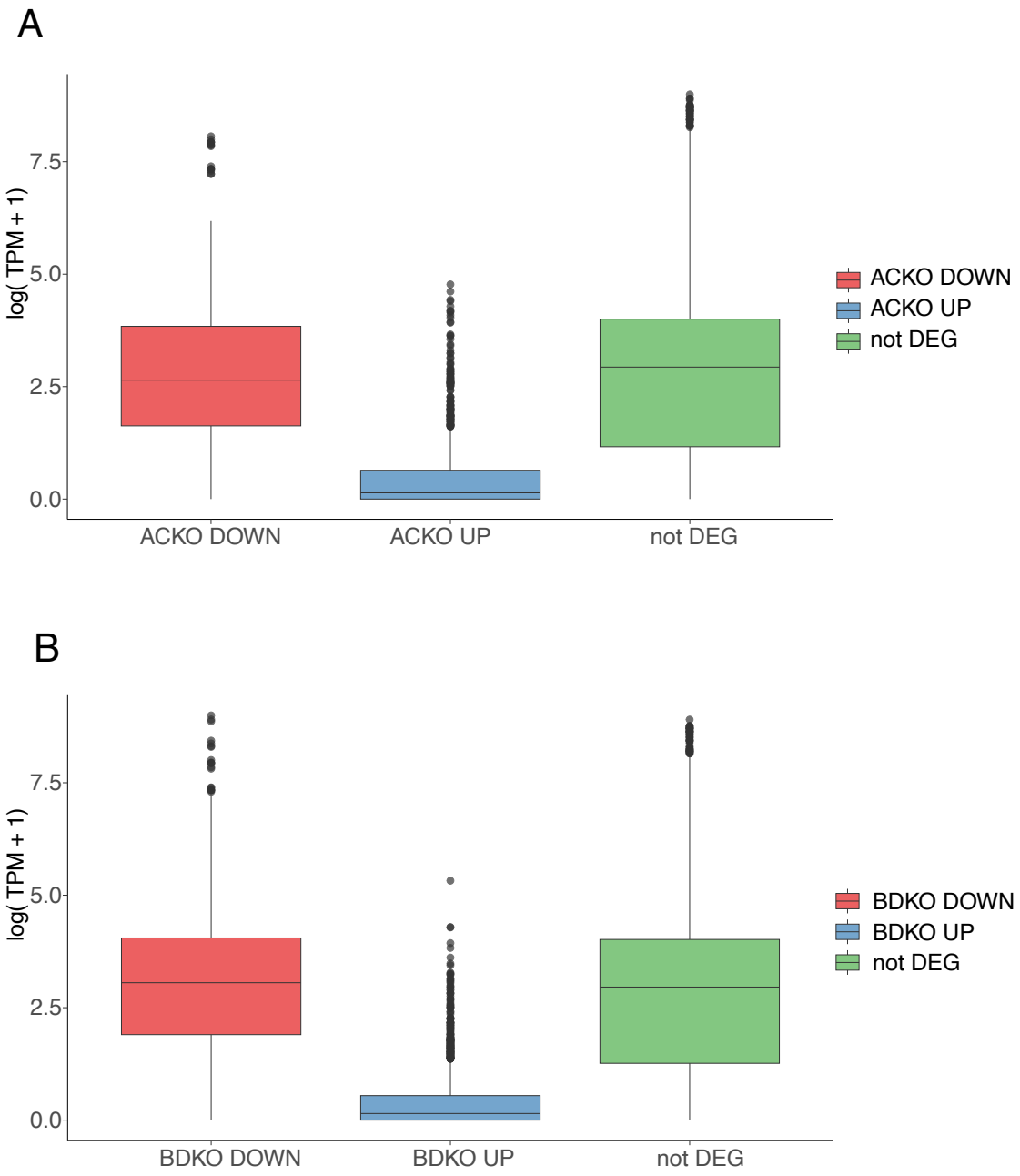


Figure 8. Metaprofiles of DEGs illustrate specific epigenetic and accessibility profile. A. Metaprofile heatmap of S2SvsS3 in wt and ACKO for up- and downregulated genes in ACKO **B.** ChIP-seq in wt of H3K27ac, H3K27me3, H3K9ac, H3K9me3, H3K9me1 for genes shown in A **C.** Metaprofile heatmap of S2SvsS3 in wt and BDKO for up- and downregulated genes in BDKO **D.** ChIP-seq in wt of H3K27ac, H3K27me3, H3K9ac, H3K9me3, H3K9me1 for genes shown in **C**.

These heatmaps show in both mutants a decrease in solubility in genes downregulated in ACKO and BDKO (**Figure 8 A, C,** left panels), suggesting a direct link between solubility and transcription (see discussion). Interestingly also upregulated genes are accompanied by a trend of increasing accessibility. To characterize the chromatin state of affected genes, we analysed the available ChIP-seq data in wt MEF and we plotted the ChIP enrichment of the ACKO and BDKO DEGs (**Figure 8 B, D**). The deregulated genes show a general enrichment in euchromatin marks. However, only downregulated genes in both mutants show an enrichment of the Polycomb dependent H3K27me3 mark at the Transcriptional Start Site (TSS). This corroborated previous data showing a crosstalk between lamins and Polycomb[2,21].



Supplementary Figure 3 Transcripts levels in wild type of deregulated genes are different. **A.** Transcripts per million (TPM) levels of the resulting DEGs in ACKO mutant are shown for the wt condition **B.** Transcripts per million (TPM) levels of the resulting DEGs in BDKO mutant are shown for the wt condition. Downregulated genes are in red, upregulated genes in blue and unaffected genes are in green.

Interestingly, as evidenced by analysis of DEGs transcription in wt condition (**Supplementary Figure 3**), upregulated genes are transcribed at the low level,

although enriched in histone marks of euchromatin (**Figure 8 B and D**). On the other hand, downregulated genes show consistent transcription in wt condition, despite the presence of the repressive H3K27me3 mark. Further analysis will clarify if these genes are bivalent.

Using the analysis described in chapter 2, we calculated the chromatin compartments starting from 4fSAMMY-seq fractions, to evaluate the knockout effects on the overall chromatin compartmentalization (**see methods**). The chromatin compartments are called with a resolution of 50Kb for all the condition replicas. The amount of chromatin compartments switches ranges from 2.5 to 4 % genome wide, with an higher impact in Lamin B KO (**Figure 9 A**). This suggests that the overall separation between euchromatin and heterochromatin is still maintained after lamins KO. This is expected since compartment switch requires a strong remodeling of chromatin.

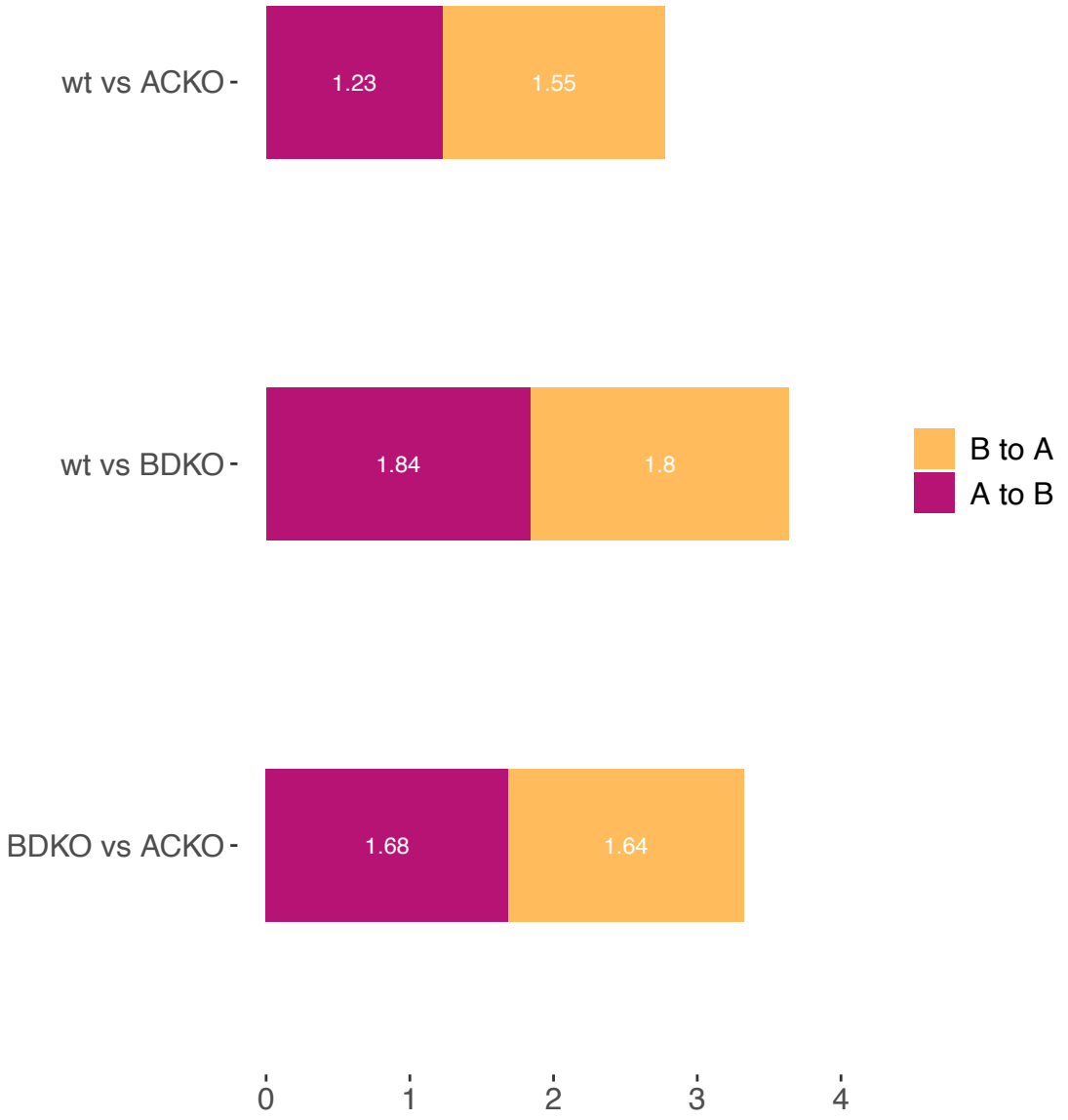


Figure 9. Compartment remodeling in lamins KO. Percentage of genome that change compartmentalization, from A to B and vice versa per contrast. The first condition (wt or BDKO) in the label name is the starting compartment.

The overlap between regions switching compartments in ACKO and BDKO mutants, calculated by Jaccard score (0.139 for A to B switch and 0.096 for B to A switch) indicates that only a small portion of the regions is commonly affected (14% of the regions that switch from A to B and 10% from B to A) (**Figure 10 A**).

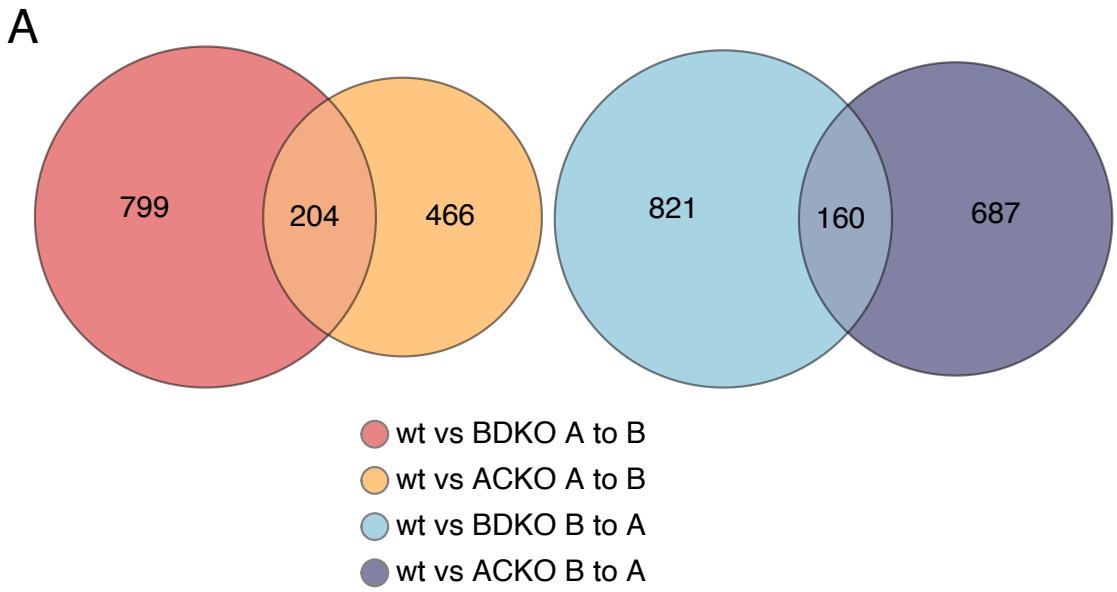


Figure 10. Lamin mutants differently remodel chromatin compartments. Venn diagram of compartment's switches in BDKO and ACKO. The A to B switches are in red and orange while B to A compartment switches are in light and dark blue.

3.4 Materials and methods

Cell cultures

Immortalized wild-type (WT), knockout for Lmna gene (LMNACKO) and double knockout for Lmnb1 and Lmnb2 (LMNBDKO) MEFs were routinely grown in DMEM (Gibco, 10566-016) supplemented with 10% (v/v) FBS (Gibco,10270106), 1% penicillin-streptomycin (Sigma-Aldrich, P0781) and 1% glutamax (Gibco,10566-016). Cell cultures have been kept under optimal growth conditions: 5% CO2 at 37°C.

RNA extraction, library preparation and sequencing

Total RNA was extracted from one million trypsinized cells for each condition in triplicate using TRI-Reagent (Sigma, T9424) following the recommended

guidelines. Total RNA was quantified by Qubit 4 fluorometer with Qubit RNA HS Assay Kit (Invitrogen, Q32852) and assessed by Agilent 2100 Bioanalyzer using Agilent RNA 6000 Nano Kit (Agilent, 5067-1511) to inspect RNA integration. For each sample, 10 ng of total RNA was used to construct strand specific RNAseq library with SMARTer Stranded Total RNA-Seq Kit - Pico Input (Takara, 634487). The quality of the libraries was assessed on Agilent 2100 Bioanalyzer using High Sensitivity DNA Kit (Agilent, 5067-4626). RNAseq libraries were sequenced on the Illumina NextSeqTM 550 system at the Ospedale Policlinico in Milan to a minimal target of 40 million for 75 bases in paired-ends mode.

Cell-cycle assay

To analyze the cell cycle profiles, 200K cells were collected, rinsed with 1× PBS and then fixed in 80% ice-cold ethanol at 4°C. The fixed cells were washed twice with ice-cold 1× PBS, followed by incubation in 1x PBS containing 25 mg/ml PI (PI, Sigma, P 4170) and 0.5 mg/ml RNase A at 4°C for 1 hour. Following staining, the cells were analyzed using flow cytometry.

4fSAMMY-seq protocol

SAMMY-seq protcol were done as explained in chapter 2

Literature data processing

The ChIP-seq data have been downloaded from the following publicly dataset available: GSE202716[33]

All the data were downloaded as raw data with fastq-dump and re-analyzed as described in followed section.

DNA sequence analysis

The results of the sequencing have been demultiplexed with bcl2fastq (https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-

conversion-software.html) (v2.20.0.422). The quality of sequencing was evaluated with fastqc[34] (version 0.11.3) and trimmed using Trimmomatic[35] (v0.39) with the following parameters: 2:30:10 for seed_mismatch, palindrome_threshold and simple_threshold, respectively; 4:15 for sliding window. The minimum threshold of 36 bp has been applied for all reads. Trimmed reads were aligned using BWA[36] (v0.7.17-r1188) setting -n 2 -k 2 and using as reference genome the mm10 downloaded from refgenie. The alignment duplicates were marked and removed with Picard (v2.23.9) (https://github.com/broadinstitute/picard) MarkDuplicates. We further filtered all the reads with mapping quality lower than 1 with Samtools[37] (v 1.11). For each alignment a coverage analysis has been performed using Deeptools[38] (version 3.5.2) bamCoverage function. The reads extended up to 250 bp and RPKM normalization method has been used. The mm10 size was considered of 2652783500 bp (value suggested in the Deeptools manual
https://deeptools.readthedocs.io/en/latest/content/feature/effectiveGenomeSize.html), from the analysis were excluded blacklisted regions obtained by ENCODE portal, (https://www.encodeproject.org/files/ENCFF547MET). The differential enrichment analysis between SAMMY fractions (comparison) was performed using spp package[12], an R (v4.3) library. The reads were additionally filtered with 'remove.local.tag.anomalies' and the differential enrichment were computed using 'get.smoothed.enrichment.mle' function setting 'tag.shift = 0' and 'background.density.scaling = TRUE'. For maximal comparability the analysis has been performed with the same parameters for all the datasets.

Track representation

The Gviz[39] R library was used for visualization purposes. Each profile was computed using the function "DataTrack"; the samples were imported using the function "import" of the rtracklayer library and plotted using the function "plotTracks" setting the value "window = 900". Visualization of the Consensus tracks were using all samples in each category setting the parameter type as 'a' and overlayed using the function "OverlayTrack". Single samples mountain plots

were computed by setting the parameter type as "polygon". Extra elements as chromosome ideogram and genome axis, were displayed using the functions “IdeogramTrack” and “GenomeAxisTrack” respectively.

SAMMY-seq comparison tracks normalization and consensus track.

The comparison signal before the analyses, is computed by rebinning at the specified windows of 150 Kb. The blacklisted regions were firstly merged together with the neighbors if the distance were less than 50Kb. After, all together the comparisons in the dataset are normalized by computing the quantile normalization with the preprocessCore[42] library by the function "normalize.quantiles". Finally, the consensus track is generated by computing the mean of the signals and the standard error as intervals.

Correlation Analysis

The smoothed differential signal enrichment is rebinned with Deeptools "multiBigwigSummary" at 50Kb. Genome-wide Spearman correlations between SAMMY-seq samples and ChIP was computed using Deeptools using the function "plotCorrelation" with the following settings: "--corMethod spearman -p heatmap --skipZeros".

Differential enrichment analysis of SAMMY-seq comparisons

To compute the differential enrichment analysis of the comparisons we firstly rebin and normalize the tracks as described above with 150 Kb rebinning size and 50Kb as merging distance of the blacklisted regions. We compute the consensus tracks for the control condition and calculate the intervals the standard deviation *2 (SD*2), obtaining consensus ± SD*2 for each genomic bin. Given the signal distribution of the consensus bins we filtered out all the bins within a range value ≤ ±0.1. After the filtering, the remaining bins are assigned to the correspondent fraction, S2S if the control condition signal is positive and S3 if is negative. To find the differential enriched bins in the test condition we sub-

selected the bins in which all samples were not in the consensus ± SD*2 intervals. The selected bins were then put in a category defined as “up” if all the test samples increased the signals, and “down” if they decreased their signals. Therefore, we obtained 4 distinct categories S2S_up, S2S_down, S3_up, S3_down.

SAMMY compartments

The computation of SAMMY-seq compartments were done using the same method explained in chapter 2**,** with 50Kb of resolution.

Gene Ontology (GO) permutation test

Available gene ontologies methods for transcriptomic analyses assume that each gene has the same probability of being represented. In genome analysis this is not the case, because we remove genes that by default that will never be encountered due to sequence characteristics (genes in blacklisted regions). Another problem to consider is: there are gene clusters with similar functions in the same genomic regions. These genes are often in same pathways. Furthermore, in SAMMY-seq, broad peaks include multiple TSS. All of these issues would systematically inflate the obtained P-values of the resulting GOs. We develop a specific analysis to address these aforementioned problems. The analysis consists in a permutation test, that relies on the assumption of exchangeability, a plausible assumption with genomic regions. The steps were: i) selection of all the genes that show an overlap of the TSS, in a range of 2500bp upstream and 500 bp downstream, with the tested genomic regions; ii) selection of all ontologies that incorporate the overlapping genes, keeping the count of the number of genes found in each ontology (X); iii) random shuffling (permutation) of all the accounted genomic regions along the genome keeping the size intact and avoiding overlapping with the blacklisted regions by masking them. Selection of genes (as explained in point i) after random repositioning and counting of number of genes associated to the GO described in point ii (x1). This procedure

was repeated 10.000 times (*B*) for each gene ontology selected at point ii (*x1*…*xB*). iv) Counting for each GOs number of times when by permutation we found the number of genes associated with a GO to be greater than the selected regions. v) Dividing by the number of permutation the P-values associated to each GOs. This can be summarized by the formula:

$$P = \frac{\sum_{j=1}^{B} I(x_j \geq X)}{B}$$

Where *I* is the function which takes the value zero when it is false, and one when the is true[43]. vi) We then correct all GOs P-value with Benjamini Hochberg method. This analysis has been performed in R (v4.3) using the GO database R library "org.Mm.eg.db" with the database updated at 2023-Mar05.

RNA-seq sequencing read analysis

The sequenced reads were analyzed with the pipeline nf-core/rnaseq version[44] (v3.8). Quality reports of raw reads and preprocessed reads are generated by FastQC[34] (v0.11.9). Quality trimming and adapter clipping are performed by cutadapt[45] (v3.4) and trimmed reads were cleaned of ribosomal RNA (rRNA) sequences with SortMeRNA[46] (v4.3.4) considering all the available databases. Reads were then mapped with STAR on mouse genome build mm10. Gene and transcript were quantified using Salmon[47] on GENCODE (M25)[48] Basic gene annotation filtered for only protein coding genes. Differential expression analyses were performed with DESeq2[49]. Due to high P-values in differential expression results we also applied the lfcShrink function with type ashr[50]

Gene Set Enrichment Analysis (GSEA)

The gene list has been ranked according to Pvalue and log2 fold change (-log10(Pvalue) * sign(log2fc)). To compute the GSEA[51,52] (v4.1.0) plot we used gmt files available at g:profiler2[26] site and computed with parameters: Number permutations: 1000; Collapse: No; seed 42; Enrichment Statistic: classic; Max size: 500; Min size:15. This has been done equally for each datataset.

RNA-seq metaprofile

To compute the accessibility metaprofile of the differential expressed genes, we rebin the comparisons and normalize them as described above (**SAMMY-seq comparison tracks normalization and consensus track** section) with a resolution of 2 Kb and 5 Kb as merging distance of the blacklisted regions. We then used Deeptools to compute the metaprofile of the enriched genes using the command "computeMatrix" from deeptools scale-regions --beforeRegionStartLength 5000 --regionBodyLength 5000 --afterRegionStartLength 5000 --skipZeros". We then used Deeptools to plot the metaprofile generated using the command "plotProfile --plotType heatmap --yMin -1 --yMax 1". To compute the ChIP-seq metaprofiles we used the tracks as they were computed by spp package and the aforementioned commands.

3.5 References

1. Turgay, Y. *et al.* The molecular architecture of lamins in somatic cells. *Nature* **543**, 261–264 (2017).
2. Cesarini, E. *et al.* Lamin A/C sustains PcG protein architecture, maintaining transcriptional repression at target genes. *Journal of Cell Biology* **211**, 533–551 (2015).
3. Shimi, T. *et al.* The A- and B-type nuclear lamin networks: microdomains involved in chromatin organization and transcription. *Genes Dev.* **22**, 3409–3421 (2008).
4. Briand, N. & Collas, P. Lamina-associated domains: peripheral matters and internal affairs. *Genome Biol* **21**, 85 (2020).
5. Meuleman, W. *et al.* Constitutive nuclear lamina–genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* **23**, 270–280 (2013).

6. van Steensel, B. & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780–791 (2017).
7. Tatli, M. & Medalia, O. Insight into the functional organization of nuclear lamins in health and disease. *Current Opinion in Cell Biology* **54**, 72–79 (2018).
8. Evangelisti, C. *et al.* The wide and growing range of lamin B-related diseases: from laminopathies to cancer. *Cell. Mol. Life Sci.* **79**, 126 (2022).
9. Zheng, X. *et al.* Lamins Organize the Global Three-Dimensional Genome from the Nuclear Periphery. *Molecular Cell* **71**, 802-815.e7 (2018).
10. Sebestyén, E. *et al.* SAMMY-seq reveals early alteration of heterochromatin and deregulation of bivalent genes in Hutchinson-Gilford Progeria Syndrome. *Nat Commun* **11**, 6274 (2020).
11. Venit, T. *et al.* Nuclear myosin 1 activates p21 gene transcription in response to DNA damage through a chromatin-based mechanism. *Commun Biol* **3**, 115 (2020).
12. Kharchenko, P. V., Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature Biotechnology* **26**, 1351–1359 (2008).
13. Probst, A. V., Santos, F., Reik, W., Almouzni, G. & Dean, W. Structural differences in centromeric heterochromatin are spatially reconciled on fertilisation in the mouse zygote. *Chromosoma* **116**, 403–415 (2007).
14. Santos, F., Peters, A. H., Otte, A. P., Reik, W. & Dean, W. Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. *Developmental Biology* **280**, 225–236 (2005).
15. Burton, A. & Torres-Padilla, M.-E. Epigenetic reprogramming and development: a unique heterochromatin organization in the preimplantation mouse embryo. *Briefings in Functional Genomics* **9**, 444–454 (2010).
16. Fadloun, A., Eid, A. & Torres-Padilla, M.-E. Mechanisms and Dynamics of Heterochromatin Formation During Mammalian Development. in *Current Topics in Developmental Biology* vol. 104 1–45 (Elsevier, 2013).

17. Burton, A. & Torres-Padilla, M.-E. Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. *Nat Rev Mol Cell Biol* **15**, 723–735 (2014).
18. Cao, K. *et al.* Rapamycin Reverses Cellular Phenotypes and Enhances Mutant Protein Clearance in Hutchinson-Gilford Progeria Syndrome Cells. *Sci. Transl. Med.* **3**, (2011).
19. Holaska, J. M. & Wilson, K. L. Multiple roles for emerin: implications for Emery-Dreifuss muscular dystrophy. *Anat Rec A Discov Mol Cell Evol Biol* **288**, 676–680 (2006).
20. Sergi Sayols. rrvgo. doi:10.18129/B9.BIOC.RRVGO.
21. Marullo, F. *et al.* Nucleoplasmic Lamin A/C and Polycomb group of proteins: An evolutionarily conserved interplay. *Nucleus* **7**, 103–111 (2016).
22. Chang, L. *et al.* Nuclear peripheral chromatin-lamin B1 interaction is required for global integrity of chromatin architecture and dynamics in human cells. *Protein Cell* **13**, 258–280 (2022).
23. Bera, M. & Sengupta, K. Nuclear filaments: role in chromosomal positioning and gene expression. *Nucleus* **11**, 99–110 (2020).
24. Khan, A. *et al.* SETD2 maintains nuclear lamina stability to safeguard the genome. *bioRxiv* 2023.09.28.560032 (2023) doi:10.1101/2023.09.28.560032.
25. Dou, Z. *et al.* Autophagy mediates degradation of nuclear lamina. *Nature* **527**, 105–109 (2015).
26. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research* **47**, W191–W198 (2019).
27. Sayols, S. rrvgo: a Bioconductor package for interpreting lists of Gene Ontology terms. *MicroPubl Biol* **2023**, (2023).
28. Bianchi, A. *et al.* Dysfunctional polycomb transcriptional repression contributes to lamin A/C–dependent muscular dystrophy. *Journal of Clinical Investigation* **130**, 2408–2421 (2020).

29. Mateos, J. *et al.* Lamin A deregulation in human mesenchymal stem cells promotes an impairment in their chondrogenic potential and imbalance in their response to oxidative stress. *Stem Cell Research* **11**, 1137–1148 (2013).
30. Del Campo, L. *et al.* Vascular smooth muscle cell-specific progerin expression in a mouse model of Hutchinson–Gilford progeria syndrome promotes arterial stiffness: Therapeutic effect of dietary nitrite. *Aging Cell* **18**, e12936 (2019).
31. Chojnowski, A. *et al.* Progerin reduces LAP2α-telomere association in Hutchinson-Gilford progeria. *eLife* **4**, e07759 (2015).
32. Lim, C. J. & Cech, T. R. Shaping human telomeres: from shelterin and CST complexes to telomeric chromatin organization. *Nat Rev Mol Cell Biol* **22**, 283–298 (2021).
33. Venit, T. *et al.* Positive regulation of oxidative phosphorylation by nuclear myosin 1 protects cells from metabolic reprogramming and tumorigenesis in mice. *Nat Commun* **14**, 6328 (2023).
34. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. (2010).
35. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
36. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at http://arxiv.org/abs/1303.3997 (2013).
37. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
38. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160–W165 (2016).
39. Hahne, F. & Ivanek, R. Visualizing Genomic Data Using Gviz and Bioconductor. in *Statistical Genomics* (eds. Mathé, E. & Davis, S.) vol. 1418 335–351 (Springer New York, 2016).
40. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).

41. Stovner, E. B. & Sætrom, P. epic2 efficiently finds diffuse domains in ChIP-seq data. *Bioinformatics* **35**, 4392–4393 (2019).

42. Ben Bolstad <Bmb@Bmbolstad. Com>. preprocessCore. (2017) doi:10.18129/B9.BIOC.PREPROCESSCORE.

43. Ruxton, G. D. & Neuhäuser, M. Improving the reporting of *P* -values generated by randomization methods. *Methods Ecol Evol* **4**, 1033–1036 (2013).

44. Harshil Patel *et al.* nf-core/rnaseq: nf-core/rnaseq v3.12.0 - Osmium Octopus. (2023) doi:10.5281/ZENODO.1400710.

45. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* **17**, 10 (2011).

46. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).

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

48. Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Research* **49**, D916–D923 (2021).

49. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

50. Stephens, M. False discovery rates: a new deal. *Biostat* kxw041 (2016) doi:10.1093/biostatistics/kxw041.

51. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267–273 (2003).

52. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).

Chapter 4 In progeria 4fSAMMY-seq reveal cell-type dependent chromatin remodeling

4.1 Abstract

Hutchinson-Gilford progeria syndrome (HGPS) is a rare, autosomic dominant human systemic disease, caused by a point mutation in Lamin A gene. Due to its key role in the control of genome architecture, it is not surprising that the structural organization and epigenetic regulation of chromatin are altered in HGPS. However, molecular studies are often focused on the terminal stage of HGPS cell life, when primary and secondary events are already overlapping. Among the epigenetic mechanisms, we are interested in the genome tridimensional structure, a key player in genome function regulation, directly regulated by Lamin A. We used the Sequential Analysis of MacroMolecules accessibilitY (SAMMY-seq), invented in our laboratory to systematically dissect chromatin conformation alterations of fibroblasts, endothelial cells and Satellite Muscular Stem cells in the HGPS mouse model at different ages. We found a cell-specific chromatin compartmentalization that is altered already in early stages of HGPS postnatal growth. Our data, showing profound changes of compartmentalization, support the hypothesis that the disease mainly affect the heterochromatin and that this alteration propagates toward euchromatin determining genome dysfunctions.

4.2 Introduction

Hutchinson-Gilford progeria syndrome (HGPS, OMIM 176670) is a rare and fatal human systemic laminopathy, estimated to be present with a frequency of 1 in 8.000.000 and caused by a point mutation in Lamin A gene[1]. This mutation produces the progerin, a truncated, splicing mutant form of the nuclear Lamin A protein. Progerin accumulation at the nucleus membrane causes genome

instability finally leading to a cellular premature senescence and a systemic aging. Currently no cures able to revert this disorder are available and treatments are mainly symptomatic and aimed in preventing secondary complications[1].

Beside its structural role, Lamin A is also an important epigenetic regulator, involved in the chromatin organization inside the nucleus[2]. Electron microscopy imaging of eukaryotic nuclei shows chromatin compartments with different levels of compaction, known as euchromatin and heterochromatin (reviewed in[3,4]). The more accessible and less condensed euchromatin is generally enriched in expressed genes. Instead, heterochromatin contains highly condensed DNA, including pericentromeric and telomeric regions[5–7], and genomic regions with unique packaging properties maintained by the Polycomb-group proteins (PcG)[8] . Lamin A preferentially interacts with the genome at specific regions called Lamina Associated Domains (LADs), enriched in H3K9me2 and H3K9me3 histone modifications that are typical of inactive heterochromatic regions[9] . LAD borders are marked by the PRC2-dependent H3K27me3 histone mark[10–12], which is characteristic of inactive PcG-regulated chromatin regions. The ensemble of lamins, chromatin marks and PcG factors around LADs creates a repressive environment[13,14], with heterochromatin and PcG target regions adjacent to each other[15,16]. In line with these observations, we previously demonstrated that Lamin A functionally interacts with PcG[17–20] , as later also reported by others[16,21,22]. In a work performed on HGPS primary fibroblasts, we detected early stage changes of heterochromatin structure and transcriptional dysregulation of PcG regulated bivalent genes, confirming that the tight crosstalk between Lamin A and PcG proteins is altered in HGPS[23]. More recently, these data have been confirmed on the mouse model and further corroborated by the analysis of LINE1, a noncoding element of the chromatin found altered in HGPS mice[24].

Here we decided to analyze the role of progerin by using 4fSAMMY-seq (see Chapter 2) and distinct cell populations coming from most affected tissues of progeric mice.

4.3 Progerin dependent chromatin remodeling captured by 4fSAMMY-seq

To analyze if progerin has a different, cell-specific impact on chromatin during the onset of symptoms, we extracted by using FACS-sorting from mice wild type (wt) and the progeric mice G609G (prg) 3 different mesenchymal derived cell types at 1 month of age (1M) and 3 months (3M) of age (**Figure 1**): endothelial CD31+ cells from aorta, fibroblasts F140+ from dorsal skin and muscle satellite alpha7+ cells from muscles. After, we applied the 4fSAMMY-seq to all the samples.

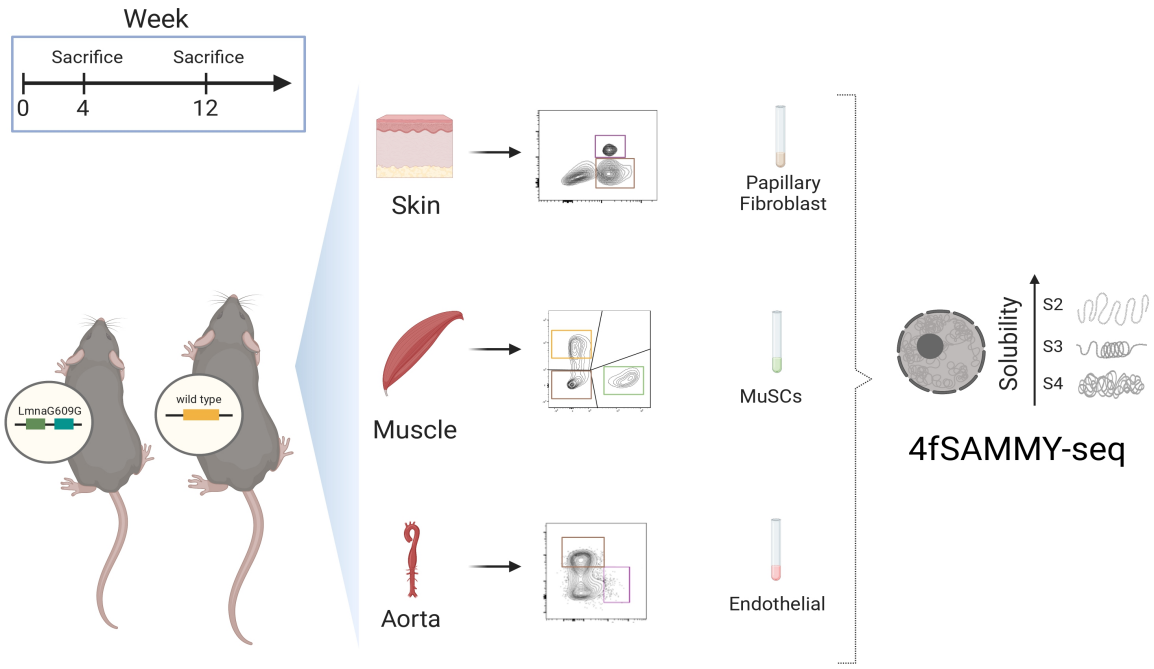


Figure 1. graphical abstract. A. S2S_up in prg1Mcd31_vs_wt1Mcd31

We firstly characterized for each time point and cell type the consensus tracks of the S2SvsS3 comparisons. We removed the genomic regions with non-clear enrichment in the wild type, by filtering out the regions with a mean enrichment less than ±0.1. The scatter plots in **Figure 2**, shows the distance of the mutants from the wt condition. The red line is the linear regression generated from the contrasts, the maximum overlap expected would be on the diagonal with the coordinates (x 0,0 and y 1,1) with an r^2 of 1.

All the cell types increase their differences from the wt condition when comparing one and three months of age, suggesting a continuous detrimental effect on chromatin solubility due to progerin accumulation in the affected cells.

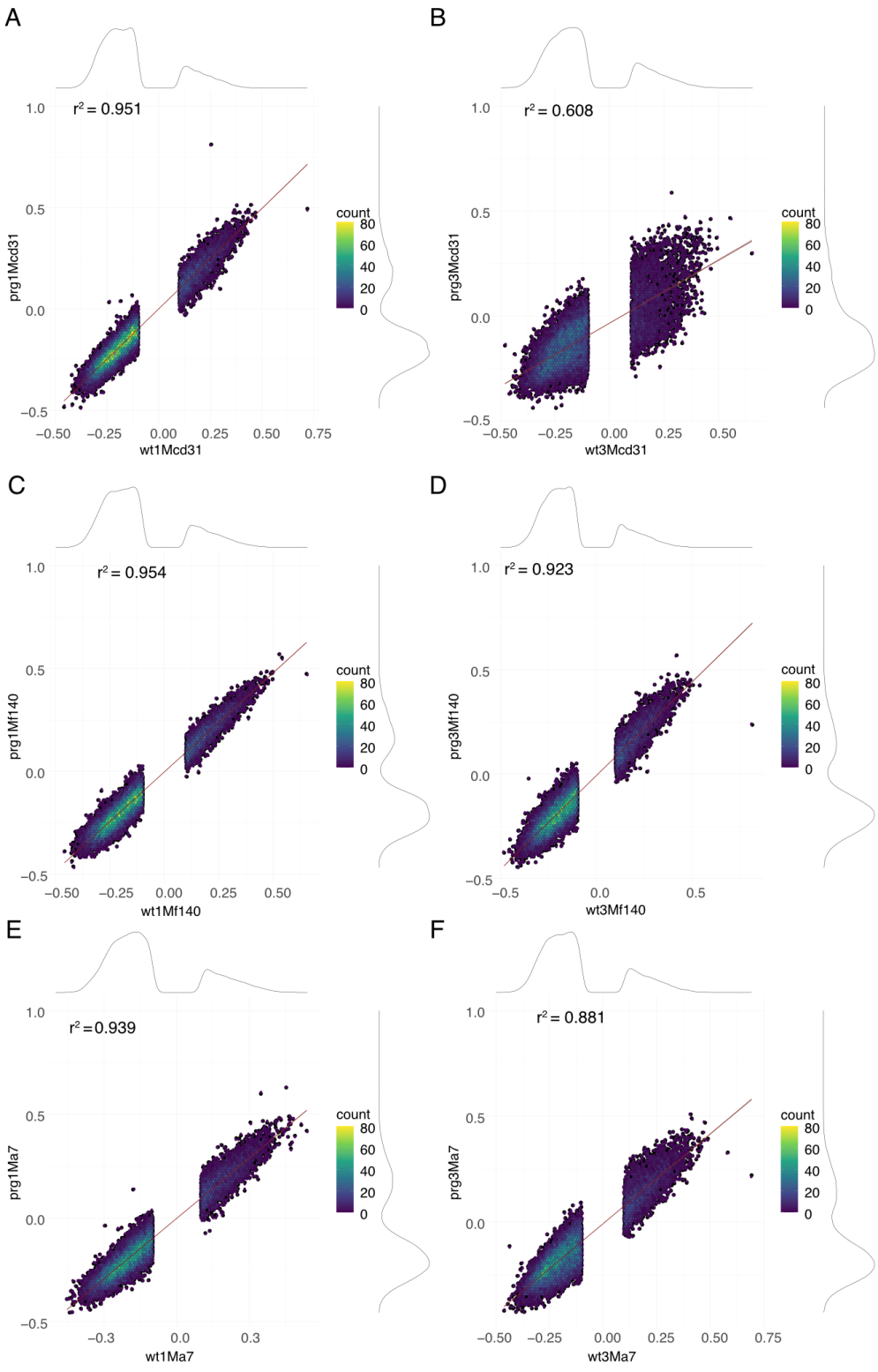


Figure 2. Scatterplot of the S2SvsS3 consensus tracks between the progeric compared to the equivalent healthy tissues. Genome wide scatter plot of the rebinned consensus signal at 150Kb, after the removal of all the bins below the signal threshold of ±0.1. The red line is the linear regression, the color code shows the density of the dots, from low amount (dark blue) to highest (yellow). The lines on top and the right represent the distribution of the signal values. **A.** prg1MCD31 vs wt1MCD31 **B.** prg3MCD31 vs wt3MCD31 **C.** prg1MF140 vs wt1MF140 **D.** prg3MF140 vs wt3MF140 **E.** prg1Ma7 vs wt1Ma7 **F.** prg3Ma7 vs wt3Ma7. Abbreviations: prg, progeric; wt, wilde type; endothelial cells, CD31; fibroblasts, F140; satellite alpha7, a7.

Interestingly, endothelial CD31+ progeric cells extracted from 3 month-old progeric aorta are the most affected (**Figure 2 A, B**), in line with the tissue involvement. On the other hand, all the analyzed cell types appear to retain most of the structure during the onset of symptoms at one month of age with a similar degree of intensity (**Figure 2 A C E**).

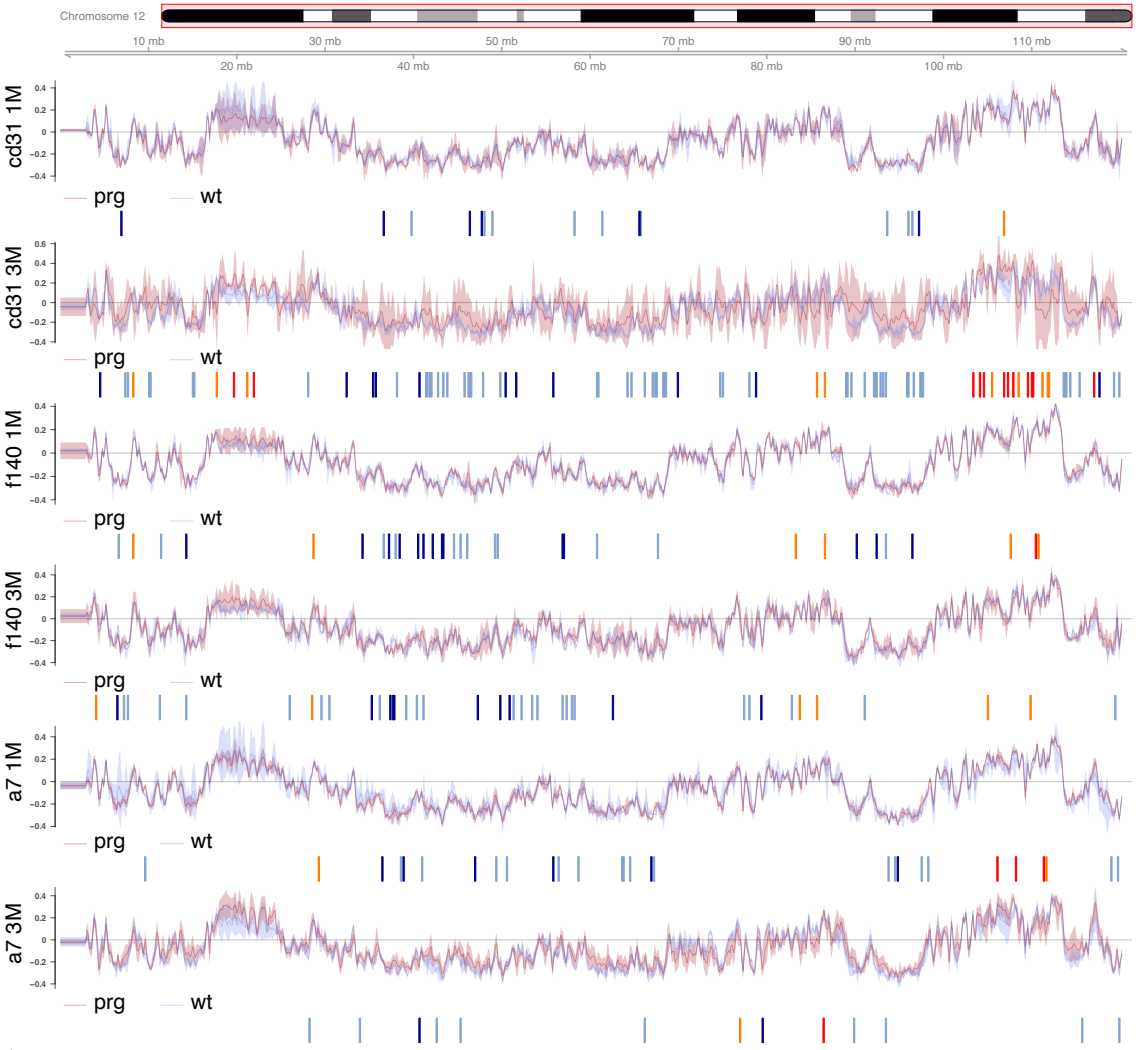


Figure 3. Solubility remodeling chromosome wide. Chromosome 9 is shown. The lines in blue and red are the consensus track of 4fSAMMY-seq between progeric samples(red) and wild type (blue). From top to bottom prg1MCD31 vs wt1MCD31, prg3MCD31 vs wt3MCD31, prg1MF140 vs wt1MF140, prg3MF140 vs wt3MF140, prg1Ma7 vs wt1Ma7, prg3Ma7 vs wt3Ma7. Lighter color shadows for each track represents the standard deviation. Below each pair tracks, the significant regions are highlighted as follow: S2S_up (red), S2S_down (orange), S3_up (light blue) and S3_down (dark blue). Abbreviations: prg, progeric; wt, wilde type; endothelial cells, CD31; fibroblasts, F140; satellite alpha7, a7.

To understand the exact degree of chromatin remodeling we performed a differential enrichment analysis for all the cell types and time points. We generated the consensus tracks for all cell types (**Figure 3**). We classified the differential soluble regions into 4 categories: S2S_up (red label), regions of S2S with increased accessibility in the progeric; S2S_down (orange label), regions of

S2S with decreased accessibility in the mutant; S3_up (light blue label), regions of S3 with increased accessibility in the mutant and S3_down (dark blu label), regions of S3 with decreased accessibility in the mutant. Notably, with this analysis we select only genomic regions with different solubility in all analysed mutants versus all wt samples.

Accordingly with the scatter plot analysis (**Figure 4**), we found an overall increase in the chromatin architecture alterations with age, confirming the dominant negative role of progerin on genome structure. As expected, we also found that heterochromatin is the most affected part pof the chromatin in all cell analysed, confirming the key role of Lamin A in heterochromatin maintenance. Surprisingly the satellite muscle stem cells show an inverse trend in solubility, with more regions affected in one month-old mice. Since these cells are the only population of quiescent cells analyzed, it is possible that the absence of the cell cycle determine an accumulation of defects at the chromatin level. As expected by low gene density in heterochromatin we found an higher number of genes in euchromatic regions (**Figure 4**).

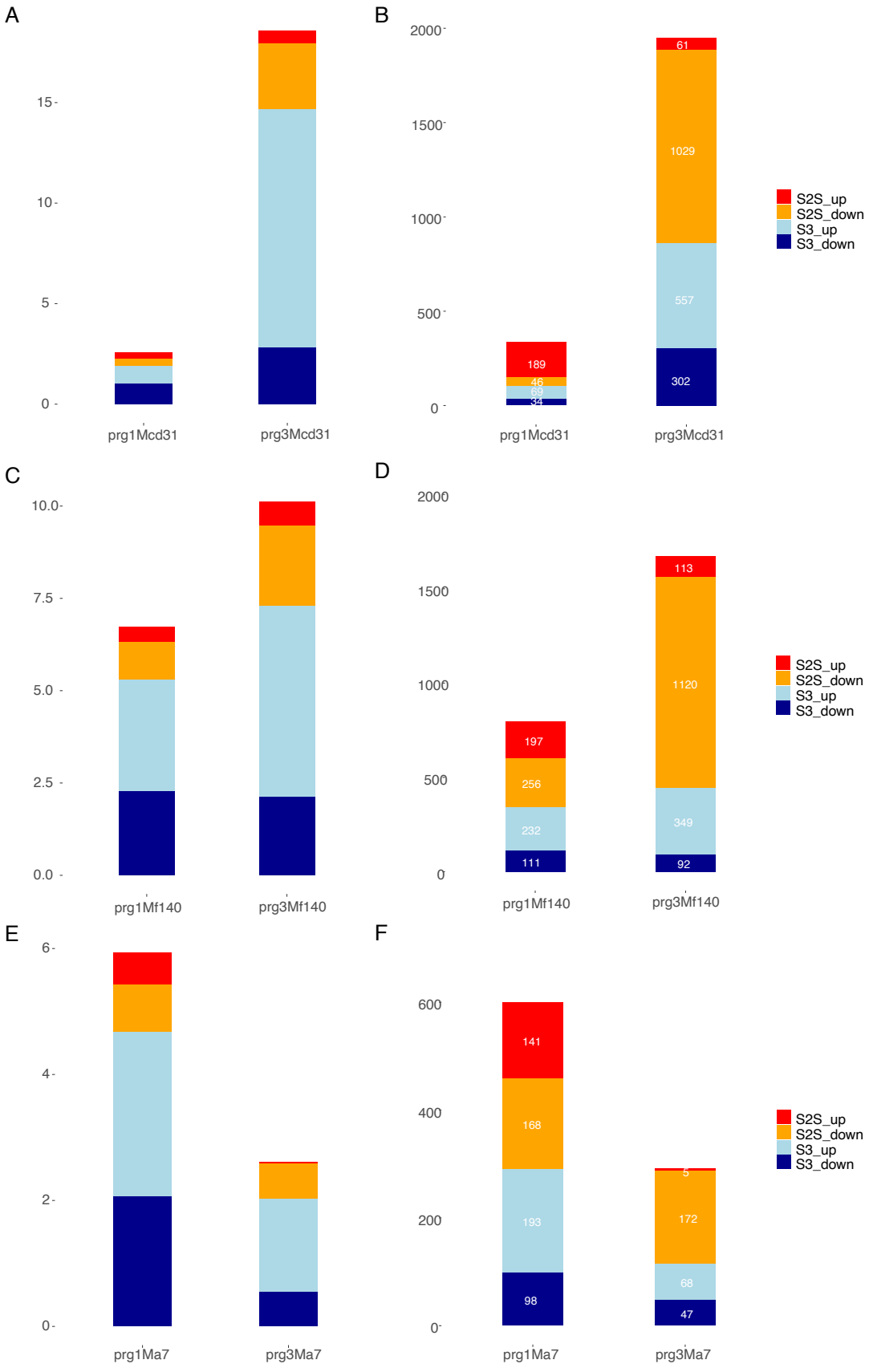
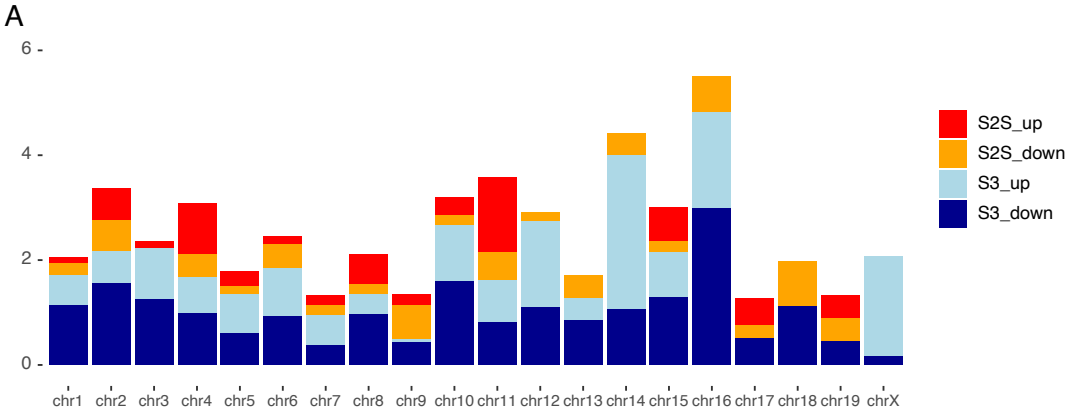
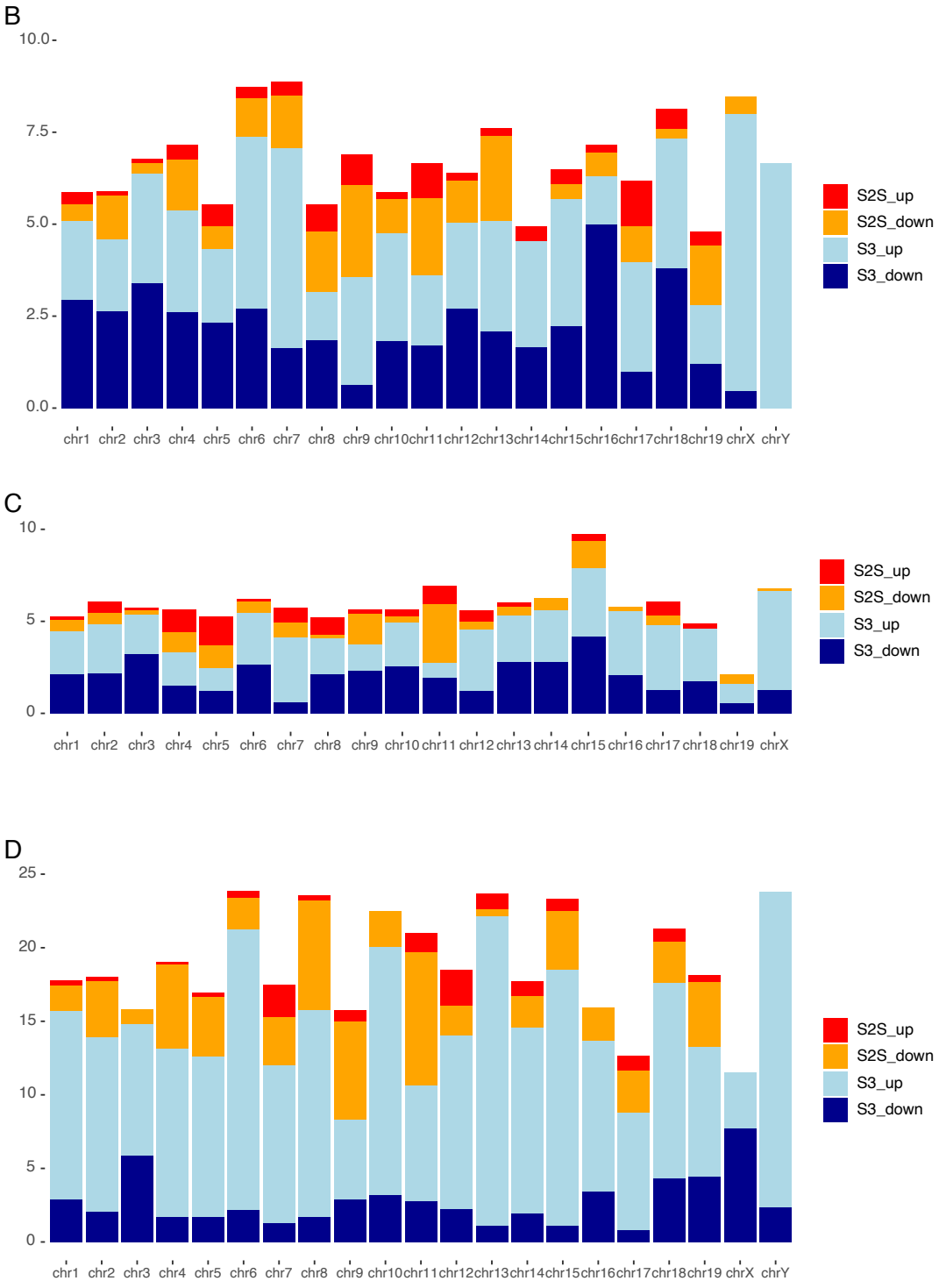

A
15
10
5
0
prg1Mcd31
prg3Mcd31
B
2000
1500
1000
500
0
189
46
61
1029
557
302
prg1Mcd31
prg3Mcd31
S2S_up
S2S_down
S3_up
S3_down
C
10.0
7.5
5.0
2.5
0.0
prg1Mf140
prg3Mf140
D
2000
1500
1000
500
0
197
256
232
111
113
1120
349
92
prg1Mf140
prg3Mf140
S2S_up
S2S_down
S3_up
S3_down
E
6
4
2
0
prg1Ma7
prg3Ma7
F
600
400
200
0
141
168
193
98
5
172
68
47
prg1Ma7
prg3Ma7
S2S_up
S2S_down
S3_up
S3_down

Figure 4. Barplot differential analysis. A. C. E. Stacked bar plot of the percentages of genome affected by remodeling after filtering out the low signal > ±0.1 with the differential enrichment analysis. **B. D. F.** Stacked bar plot of the number of genes that have the TSS overlapping with the regions with a range of upstream 2500bp and downstream 500 bp. The color code for the significant regions/genes found are highlighted as follow: S2S_up (red), S2S_down (orange), S3_up (light blue) and S3_down (dark blue). **A.** prg1MCD31 vs wt1MCD31 **B.** prg3MCD31 vs wt3MCD31 **C.** prg1MF140 vs wt1MF140 **D.** prg3MF140 vs wt3MF140 **E.** prg1Ma7 vs wt1Ma7 **F.** prg3Ma7 vs wt3Ma7. Abbreviations: prg, progeric; wt, wilde type; endothelial cells, CD31; fibroblasts, F140; satellite alpha7, a7.

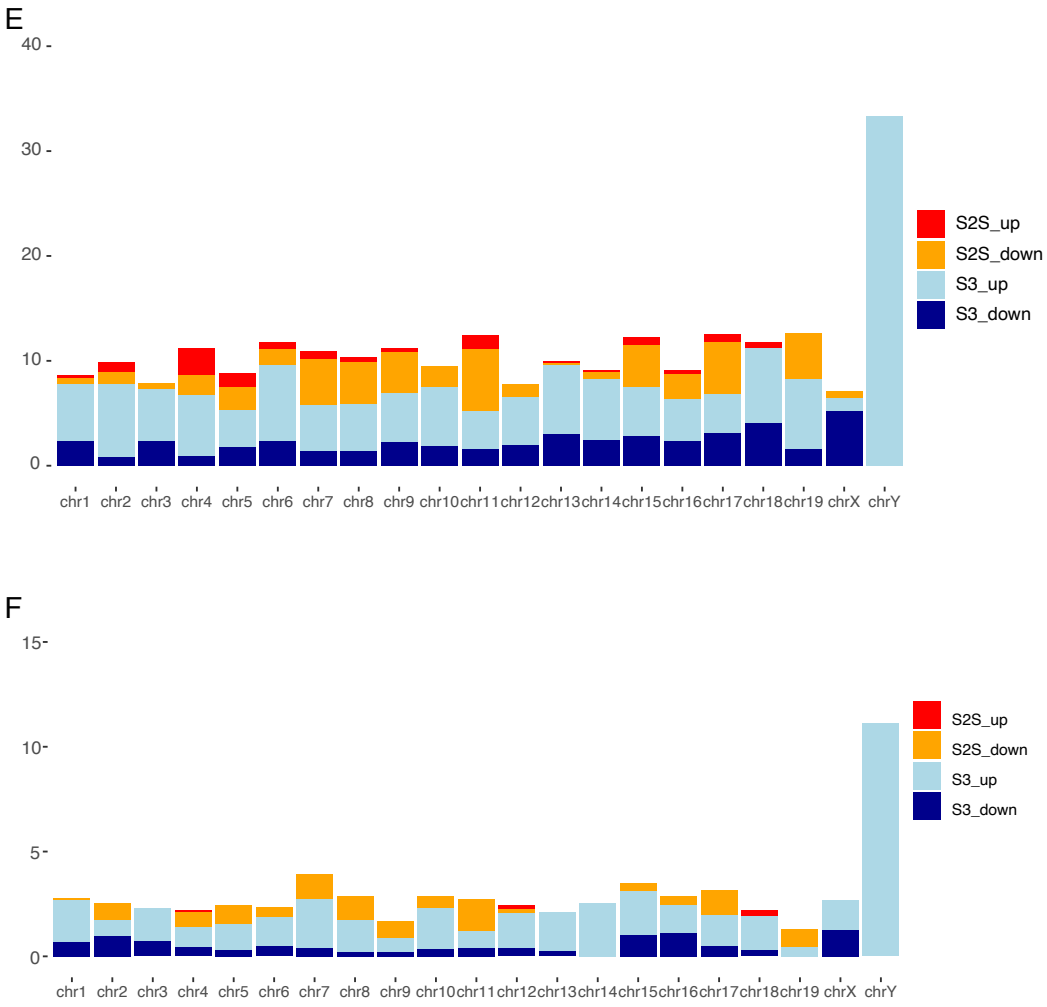
We checked if the remodeling happened with a preference for some specific chromosome. At one month no major differences across chromosomes were found (**Supplementary Figure 1 A B C**). However, at 3 months from birth all tissues displayed a major alteration of the Y chromosome, showing higher heterochromatin solubility (**Supplementary Figure 1 D E F**).



Supplementary Figure 1 continues on the next page.



Supplementary Figure 1 continues on the next page.



Supplementary Figure 1. Percentages of affected regions divided per chromosome. Each chromosome percentages is calculated from the total size analyzed after the removal of the filtered genomic bins with a signal < ±0.1. Color scheme: S2S_up (red), S2S_down (orange), S3_up (light blue), S3_down. **A.** prg1MCD31. **B.** prg1MF140 **C.** prg1Ma7 **D.** prg3MCD31 **E.** prg3MF140 **F.** prg3Ma7. Abbreviations: prg, progeric; wt, wilde type; endothelial cells, CD31; fibroblasts, F140; satellite alpha7, a7.

Then we checked if distinct tissues share some altered genomic regions by comparing the genes located in those regions (**Figure 4**). As shown in **Figure 5** we found very little overlap across distinct tissues, confirming our idea that epigenetic remodeling is tissue specific. Since available gene ontology packages are tailored on RNA-seq, to avoid inflated P-values, we set up a custom gene ontology analysis based on permutation (**see methods**).

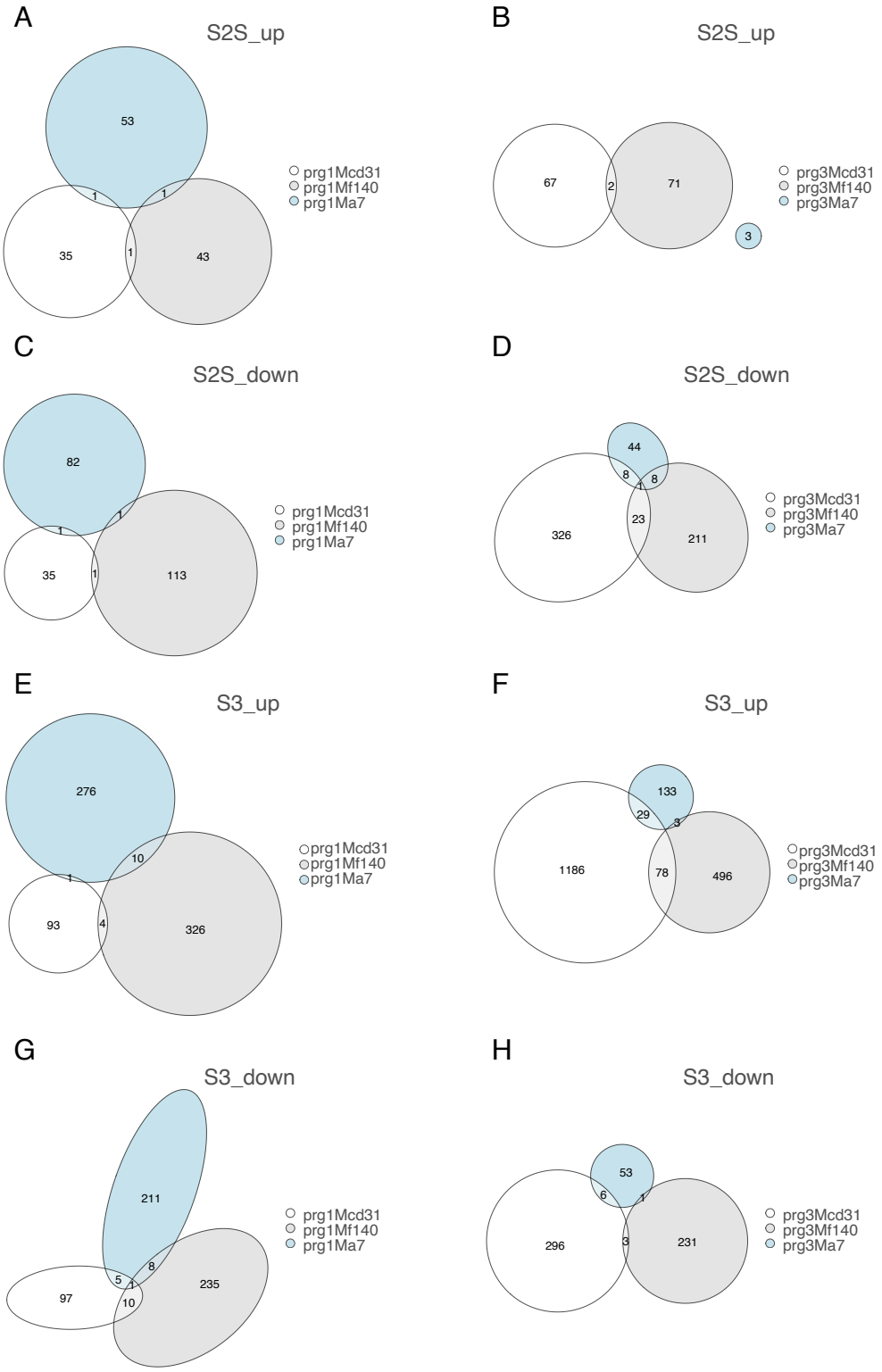


Figure 5. Venn diagrams of altered regions in each category. A. S2S_up of all cell populations at 1M. **B.** S2S_up of all cell populations at 3M **C.** S2S_down of all cell populations at 1M **D.** S2S_down of all cell populations at 3M **E**. S3_up of all cell populations at 1M. **F.** S3_up of all cell populations at 3M **G.** S3_down of all cell populations at 1M **H.** S3_down of all cell populations at

3M. Abbreviations: prg, progeric; wt, wilde type; endothelial cells, CD31; fibroblasts, F140; satellite alpha7, a7.

We used this GO analysis on differentially soluble regions (S2S_up, S2S_down, S3_up, and S3_down). Despite the comparable total number of genes GO was significantly enriched only in euchromatin regions (S2S up/down). This could be due to a spreading of different genes in heterochromatin (S3 up/down), not sufficient to enrich specific ontologies.

Endothelial cells extracted from one month old progeric mice showed an increase solubility (S2S_up) in regions involved in cell localization (cell adhesion, locomotion, cell junction, etc), filament organization (actin regulation, cytoskeleton organization) cell cycle (DNA replication, mitotic cytokinesis, etc) (**Figure 6A**). Interestingly we also found a GO specific of endothelial cells (endothelial tube morphogenesis). In the same cells regions with decrease solubility did not show any reliable GO. When analyzing the same cells extracted in 3 months old mice, we found a trend change, with significant GO only in euchromatin regions with decreased solubility (S2S_down). However, we found some categories in common with younger mice as cell adhesion and mitosis. However, we also found several new pathways involved in catabolism, membrane transportation and signaling (**Figure 6B**).

Fibroblasts of one month old progeric mice showed an increase solubility (S2S_up) in regions involved in membrane transport, inflammation (leukocytes proliferation and homeostasis, cell activation) (**Figure 6C**). Interestingly, fibroblasts, as endothelial cells, have increased solubility of euchromatin regions involved in cell localization (cell migration and adhesion) (**Figure 6C**), although they affect distinct genes (**Figure 5**). GO categories of euchromatin regions with decreased solubility mostly regards cell morphogenesis and organization of filaments (**Figure 6D**), suggesting a defect in cell shape. In 3 months old mice, as seen also in endothelial cells, significant GO of fibroblasts regards euchromatin regions with decreased solubility (S2S_down) (**Figure 6E**). The inflammation, seen altered in one month old mice in S2S_up is now in S2S_down,

but with a diversification of involved pathways (natural killer, T cell and macrophage regulation, interleukin regulation etc) (**Figure 6E**). This show how an early alteration of chromatin structure can, with age, have a cascade effect involving more and more mechanisms of the same biological process.

A

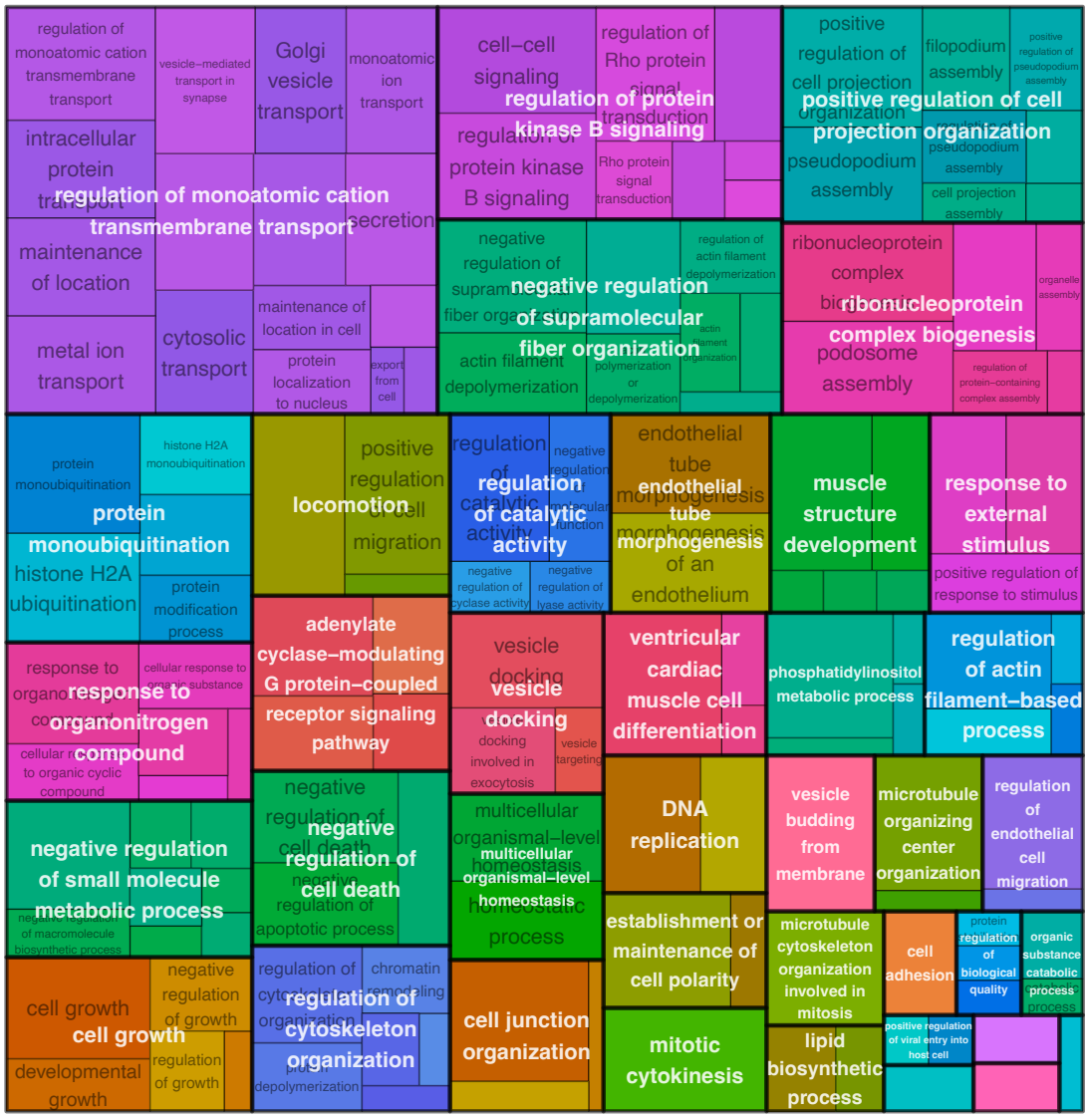


Figure 6 continues on the next page.

B

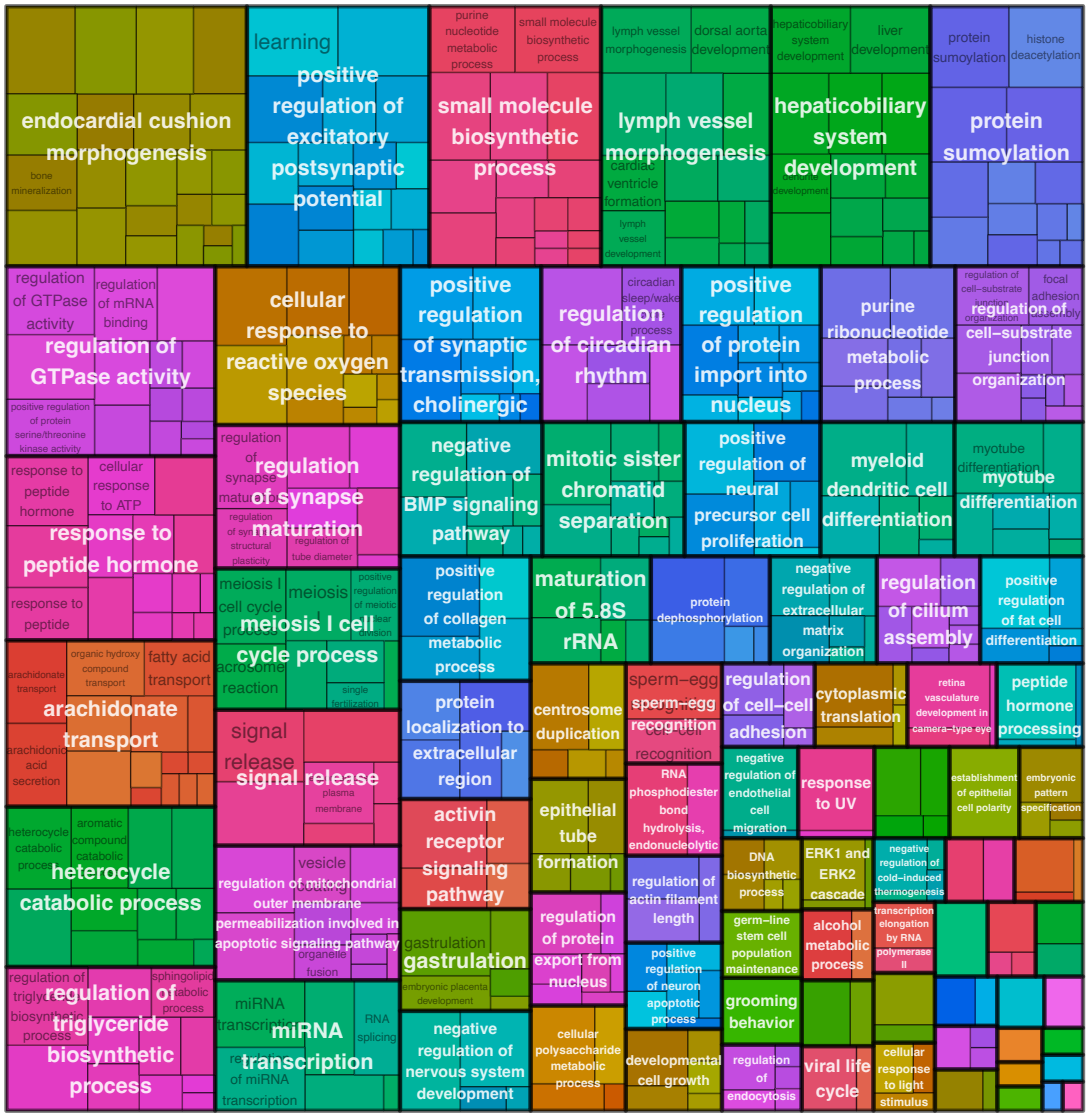


Figure 6 continues on the next page.

C

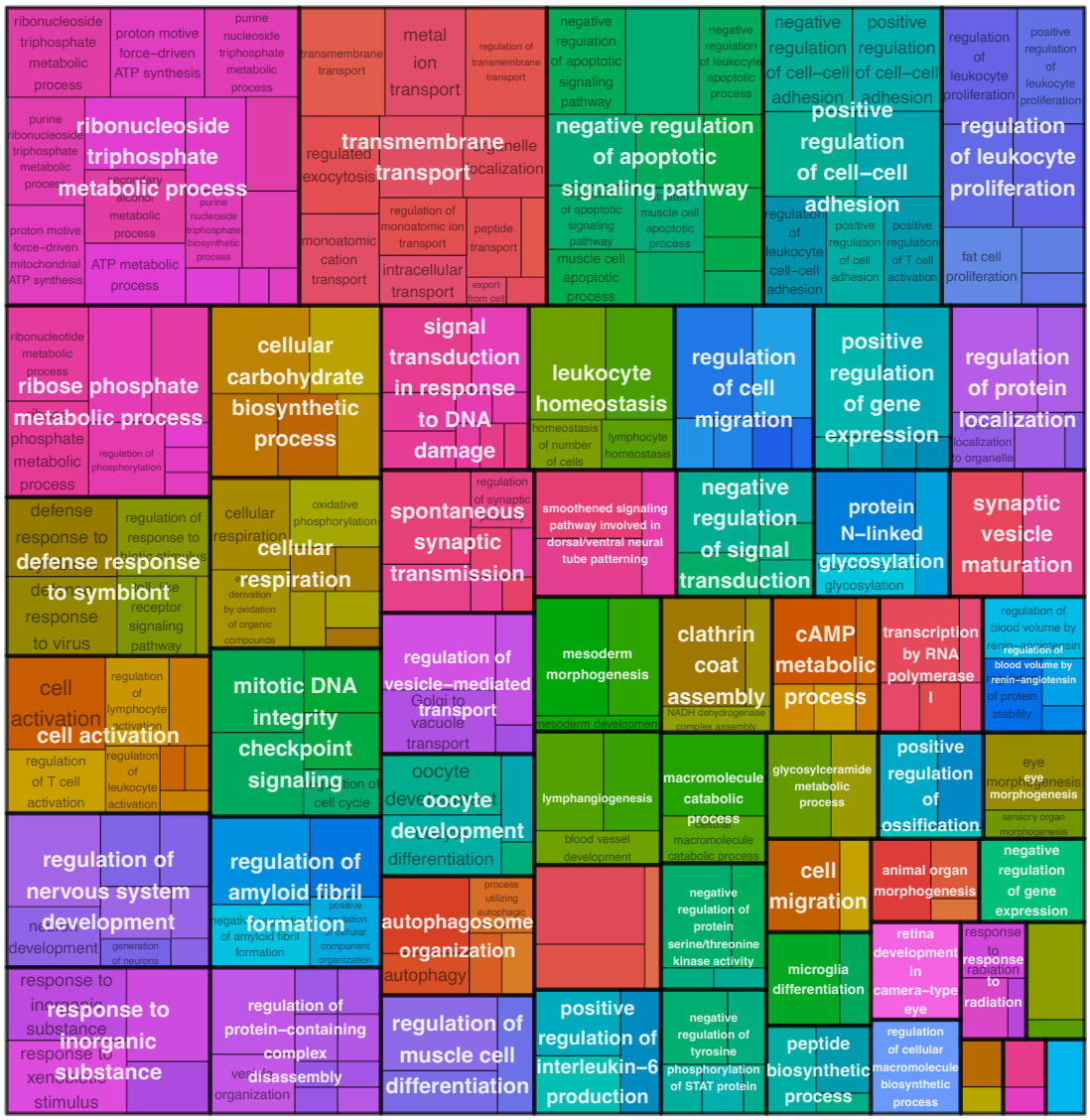


Figure 6 continues on the next page.

D

Figure 6 continues on the next page.

E

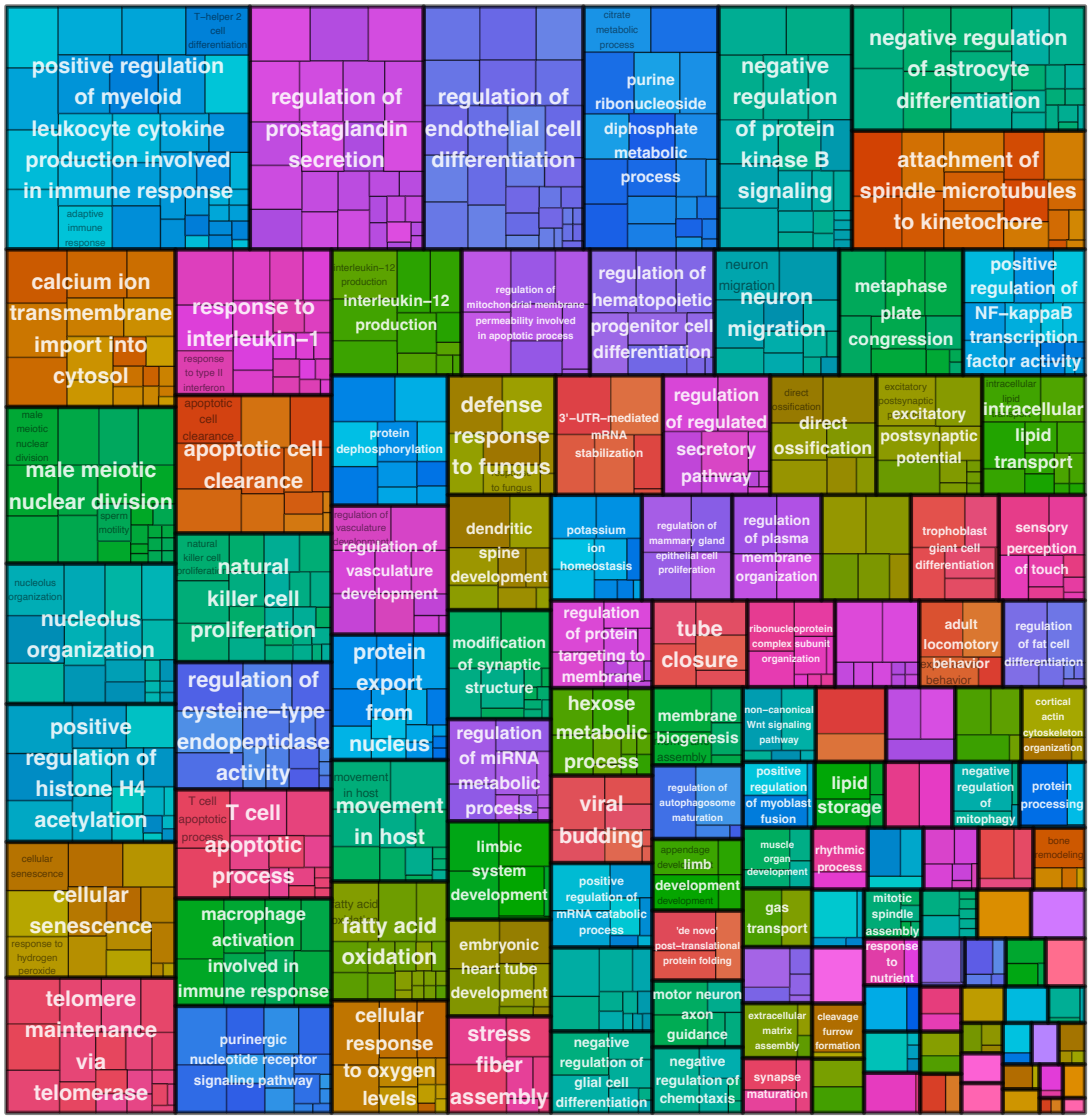


Figure 6 continues on the next page.

F

Figure 6 continues on the next page.

G

Figure 6 continues on the next page.

H

Figure 6. Semantic analysis of Gene Ontologies calculated from the genes in remodeled genomic regions. A. S2S_up in prg1MCD31. **B.** S2S_down in prg3MCD31. **C.** S2S_up in prg1MF140. **D.** S2S_down in prg1MF140 **E.** S2S_down prg3MF140 **F.** S2S_up in prg1Ma7 **G.** S2S_down prg1Ma7 **H.** S2S_down prg3Ma7. Abbreviations: prg, progeric; wt, wilde type; endothelial cells, CD31; fibroblasts, F140; satellite alpha7, a7.

Muscle stem cells (MuSC) of one month old progeric mice showed an increase solubility (S2S_UP) in regions involved in cell differentiation and muscle regeneration (**Figure 6F**), confirming their strong commitment in muscle

formation and suggesting a premature activation in progeric mice. On the other hand, these cells showed decreased solubility in regions involved in cell localization (**Figure 6G**). As other cell types described above, MuSC extracted from three months old progeric mice showed GO significance only in regions with decreased solubility (S2S_down). Also in these cells we observed an increase of the pathways involved with age. As seen in other cell type, we found cell adhesion and localization, cell cycle (**Figure 6H**), suggesting that this common pathways id directly dependent on progerin. However, some specific pathways important for MuSC activation and homeostasis were found (regulation of MAPK cascade, kinase activity, fat cell differentiation). Inflammation, that in muscle is a key mechanisms for satellite activation and muscle regeneration, is also highlighted (cytokine production, immune response, regulation of lymphocyte proliferation and migration).

Using the analysis described in chapter 2, we calculated the chromatin compartments starting from 4fSAMMY-seq fractions, to evaluate the progerin effects on the overall chromatin compartmentalization (**see methods**). The chromatin compartments are called with a resolution of 250Kb for all the replicas. We also analyzed chromatin compartments and relative switches (A-to-B and B-to-A) (**Figure 7**). In this analysis, different from the previous one, we used all the 4 fractions together to construct the solubility maps and to call compartments. In line with other analysis, the amount of chromatin compartments switches change with age, increasing over time, also in muscle stem cells. Changes are very little in one month-old mice. This is expected since compartment switch requires a strong remodeling of chromatin. Further analysis are ongoing to integrate solubility data (**Figure 2**) and compartmentalization (**Figure 7**). RNA-seq of the same pool of cells are ongoing and tailored bioinformatic analyses will provide a complete picture of the genome structure and function in healthy and changes imposed by pathological conditions.

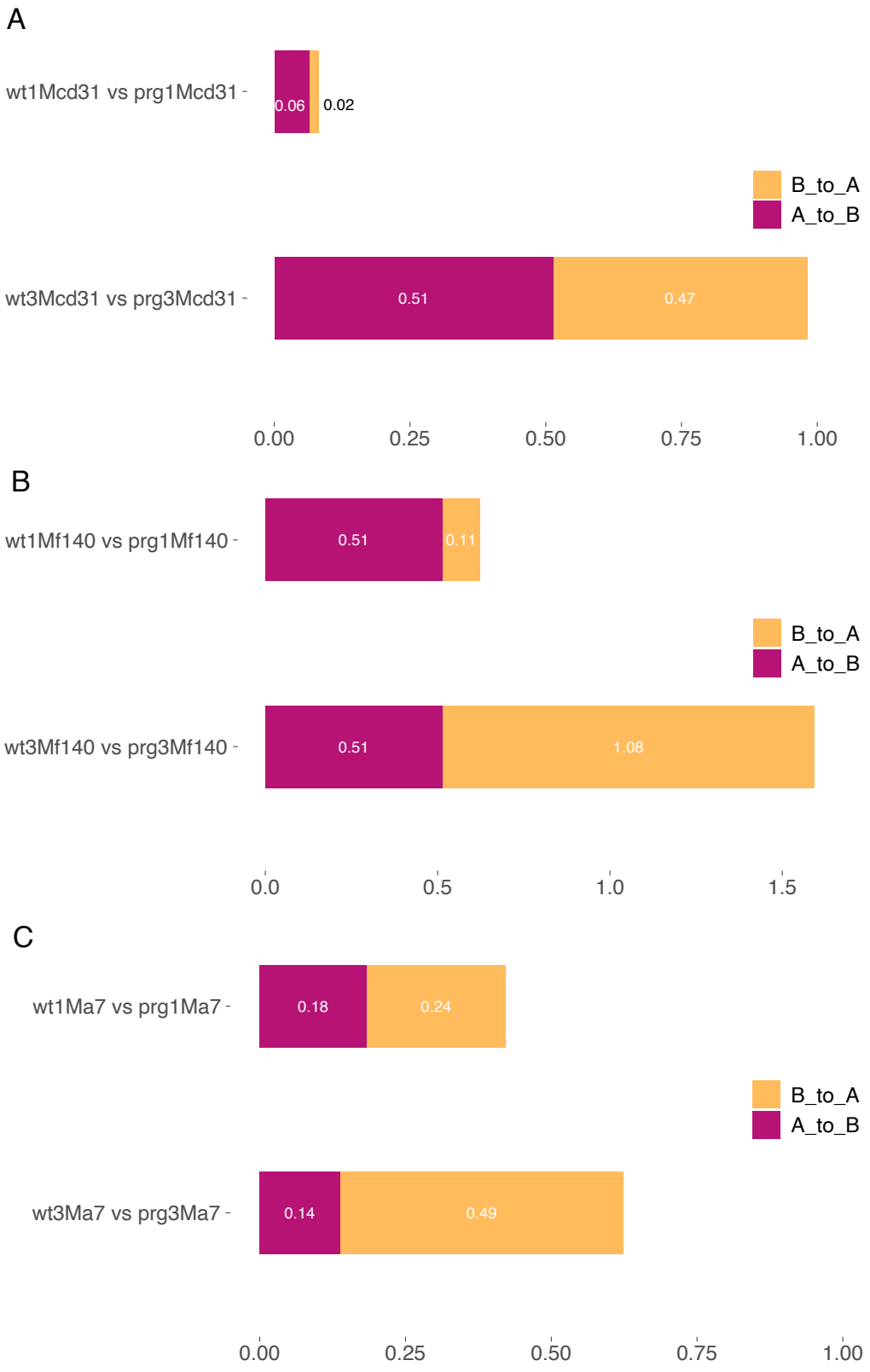

A
wt1Mcd31 vs prg1Mcd31
0.06
0.02
B_to_A
A_to_B
wt3Mcd31 vs prg3Mcd31
0.51
0.47
0.00
0.25
0.50
0.75
1.00
B
wt1Mf140 vs prg1Mf140
0.51
0.11
B_to_A
A_to_B
wt3Mf140 vs prg3Mf140
0.51
1.08
0.0
0.5
1.0
1.5
C
wt1Ma7 vs prg1Ma7
0.18
0.24
B_to_A
A_to_B
wt3Ma7 vs prg3Ma7
0.14
0.49
0.00
0.25
0.50
0.75
1.00

Figure 6. Compartment remodeling in different tissues. Percentage of the genome subjected to recompartmentalization, from A to B and vice versa per contrast. The first condition in the label name is the wild type starting compartment. **A.** CD31cells. **B.** F140 cells **C.** a7 cells

4.4 Materials and methods

Extraction and FACS sorting

Each used tissue were isolated from sacrificed mice and digested 60/75 minutes in 2,4 U/ml of Dispase II (Roche, 04942078001), 2 mg/ml of Collagenase A (Roche, 1013586001), 0,4 mM CaCl2 (Sigma, C5670), 5 mM MgCl2 (Sigma, M8266), 0,1 mg/mL DNase I (Roche, 1014159001) in PBS 1X at 37°C in a water bath. The samples were resuspended in HBSS (Gibco, 14025-050) and implemented with 0,2% BSA (Sigma, A7030) to stop the enzymatic digestion. The cell suspension was serially filtered with 70 µm and 40 µm cells strainers and resuspended in HBBS +++ (0,2% BSA, 1% DNase I, 1% PenStrep (Euroclone, ECB3001) for the night. The day after, cell suspensions were stained 30 minutes at 4°C with the following antibodies: muscle digestion- PB- CD45 1:50 (eBioscience 48-0451), PB-CD31 1:50 (eBioscience 48-0311), PB-Ter119 1:50 (eBioscience 48-5921), FITC-Sca1 1:50 (eBioscience 11-5981), APC 7integrin 1:100 (AbLab, 67- 001-05) and sorted with BD FACS ARIA SORP for: PB-CD45-/ PB-CD31-/ PB-Ter119-/ FITC-Sca1-/APC-7integrin+ (Muscle Stem Cell); aorta- PB-CD31 1:50 (eBioscience 48-0311), viability dye making TO-PRO™-3 and sorted with BD FACS ARIA SORP for PB-CD31+; skin- PB- CD45 1:100 (eBioscience 48-0451), PB-CD31 1:100 (eBioscience 48-0311), PB-Ter119 1:100 (eBioscience 48-5921), PB-CD326 1:100, PE-Vio® 770-CD140b (Miltenyi Biotec, 130-105118) 1:25, FITC- ER-TR7 1:100, viability dye making TO-PRO™-3 and sorted with BD FACS ARIA SORP for PE-Vio® 770-CD140b+.

Chromatin fractionation

Sorted cell types was put in Trypsin-EDTA solution at 37°C, 5% CO2. After two washes in cold PBS, the cells were resuspended in 600 µL of CSK-Triton buffer (10 mM PIPES pH 6.8, 100 mM NaCl, 1 mM EGTA, 300 mM Sucrose, 3 mM MgCl2, 1mM PMSF, 1 mM DTT, 0.5% Triton X-100, with protease inhibitors). After 10 min incubation on a wheel at 4°C, soluble proteins and the cytoskeletal structure were separated from the nuclei by centrifugation at 900g for 3 min at 4°C; the supernatant was labelled as S1 fraction. The pellet was then washed with an additional volume of CSK-Triton buffer, resuspended in 100 µL of CSK buffer (10 mM PIPES pH 6.8, 100 mM NaCl, 1 mM EGTA, 300 mM Sucrose, 3 mM MgCl2, 1mM PMSF, with protease inhibitors) and incubated for 60 min at 37°C with 25 U of RNase–free DNase I (Invitrogen, AM2222). To stop DNA digestion, ammonium sulphate was added in the CSK buffer to a final concentration of 250 mM. After 5 min incubation on ice, the sample was pelleted at 900g for 3 min at 4°C; the supernatant, containing digested chromatin fragments, was labelled as S2 fraction. Afterwards, the pellet was washed with 200 µL of CSK buffer and pelleted at 3000g for 3 min at 4°C, then resuspended in 100 µL of CSK-NaCl buffer (CSK buffer with NaCl final concentration increased to 2 M) and incubated 10 min on a wheel at 4°C. At the end of the incubation, the sample was centrifuged at 2300g for 3 min at 4°C and the supernatant was labelled as S3 fraction. Finally, after two washes in 200 µL of CSK-NaCl buffer followed by centrifugation at 3000g for 3 min at 4°C, the pellet was solubilized in 100 µL of 8 M urea; the final suspension was labelled as S4 fraction.

DNA sonication and sequencing

Fractions S2, S3 and S4 were diluted in TE (10 mM TrisHCl pH 7.5, 1mM EDTA pH 8.0) to a final volume of 200 *µ*l and then incubated 90 min at 37°C with 6 µL of RNase cocktail (Ambion, AM2286), followed by 150 minutes at 55°C with Proteinase K (Invitrogen, AM2548) to a final concentration of 0.2 *µ*g/*µ*L. Next, DNA was isolated through phenol:chloroform:isoamyl alcohol (Sigma, 77617) extraction, precipitated in 70% ethanol, 0.3M sodium acetate and 20*µ*g glycogen overnight at -20°C or 1 hour in dry ice and resuspended in nuclease-free water.

S2 from 4f-SAMMY-seq was additionally purified using PCR DNA Purification Kit (Qiagen, 28106) and DNA fragments in this fraction were separated using AMPure XP paramagnetic beads (Beckman Coulter, A63880) to obtain S2S (< 300 bp) and S2L (> 300 bp) fractions. Beads were added to the S2 fraction in a 0.95x (v/v) ratio to bind fragments larger than 300bp. Magnetic separation of beads from supernatant allowed the physical separation of larger fragments (on the beads) from shorter ones (in the supernatant). Larger fragments bound on beads were then washed in 85% ethanol, resuspended in water and magnetically separated from the beads (S2L fraction). Shorter fragments in the supernatant of the first step were bound to beads by adding a further 0.85x (v/v) beads ratio to the suspension; after washing in 85% ethanol and resuspension in water, they were also detached from beads (S2S fraction). Separation of S2S and S2L from S2 fraction of 10kh-SAMMY-seq S2 was also tested (C004_r1); since the enrichment profile of S2S and S2L was identical, this passage was later avoided (C002_r1, C004_r2). After DNA isolation, S2 (from 10kh-SAMMY-seq), S2L (from 4f-SAMMY-seq), S3 and S4 (from both 10kh- and 4f-SAMMY-seq) fractions were transferred to screw cap microTUBEs (Covaris, 004078) and sonicated in a Covaris M220 focused-ultrasonicator to obtain a smear of DNA fragments peaking at 200 bp (settings: water bath 20°C, peak power 30.0, duty factor 20.0, cycles/burst 50; duration: 125 sec for S2 and S2L, 175 sec for S3 and S4). DNA in the fractions was then quantified using Qubit dsDNA HS Assay Kit (Invitrogen, Q32854) and a Qubit 4.0 fluorometer; quality control was performed by run on an Agilent 2100 Bioanalyzer System using the High Sensitivity DNA Kit (Agilent, 5067-4626). Libraries were then created from each fraction using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645L) and the Unique Dual Index NEBNext Multiplex Oligos for Illumina (NEB, E6440S); final qualitative and quantitative controls were performed through an Agilent 2100 Bioanalyzer System and a Qubit 4.0 fluorometer. Libraries with distinct adapter indexes were multiplexed and, after cluster generation on FlowCell, sequenced for 100 bases in single-end mode on an Illumina NextSeq 2000 instrument at Ospedale

Policlinico in Milan. A sequencing depth of at least 24.9 million raw sequencing reads was obtained for each sample.

described in following section .

DNA sequence analysis

The results of the sequencing have been demultiplexed with bcl2fastq (https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html)(v2.20.0.422). The quality of sequencing was evaluated with fastqc[36] (version 0.11.3) and trimmed using Trimmomatic[37] (v0.39) with the following parameters: 2:30:10 for seed_mismatch, palindrome_threshold and simple_threshold, respectively; 4:15 for sliding window. The minimum threshold of 36 bp has been applied for all reads. Trimmed reads were aligned using BWA[38] (v0.7.17-r1188) setting -n 2 -k 2 and using as reference genome the mm10 downloaded from refgenie. The alignment duplicates were marked and removed with Picard (v2.23.9) (https://github.com/broadinstitute/picard) MarkDuplicates. We further filtered all the reads with mapping quality lower than 1 with Samtools[39] (v 1.11). For each alignment a coverage analysis has been performed using Deeptools[40] (version 3.5.2) bamCoverage function. The reads extended up to 250 bp and RPKM normalization method has been used. The mm10 size was considered of 2652783500 bp (value suggested in the Deeptools manual
https://deeptools.readthedocs.io/en/latest/content/feature/effectiveGenomeSize.html), from the analysis were excluded blacklisted regions obtained by ENCODE portal, (https://www.encodeproject.org/files/ENCFF547MET). The differential enrichment analysis between SAMMY fractions (comparison) was performed using spp package[41], an R (v4.3) library. The reads were additionally filtered with 'remove.local.tag.anomalies' and the differential enrichment were computed using 'get.smoothed.enrichment.mle' function setting 'tag.shift = 0' and 'background.density.scaling = TRUE'. For maximal comparability the analysis has been performed with the same parameters for all the datasets.

Track representation

The Gviz[42] R library was used for visualization purposes. Each profile was computed using the function “DataTrack”; the samples were imported using the function “import” of the rtracklayer library and plotted using the function “plotTracks” setting the value “window = 900”. Visualization of the Consensus tracks were using all samples in each category setting the parameter type as ‘a’ and overlayed using the function “OverlayTrack”. Single samples mountain plots were computed by setting the parameter type as "polygon". Extra elements as chromosome ideogram and genome axis, were displayed using the functions “IdeogramTrack” and “GenomeAxisTrack” respectively.

SAMMY-seq normalization and consensus track generation

The comparison signal before the analyses, is computed by rebinning at the specified windows of 150 Kb. The blacklisted regions were firstly merged together with the neighbors if the distance were less than 50Kb. After, all together the comparisons in the dataset are normalized by computing the quantile normalization with the preprocessCore[46] library by the function "normalize.quantiles". Finally, the consensus track is generated by computing the mean of the signals and the standard error as intervals.

Rebinning and Normalization of the comparison tracks

The comparison signal of all samples in a dataset is rebinned at the specified window. The blacklisted regions were merged together with the neighbor’s regions if the distance were less than a specified distance. After, to avoid biased results due only to fractionation efficiency, all comparisons in each dataset are normalized together by computing the quantile normalization with the preprocessCore library by the function "normalize.quantiles".

Differential enrichment analysis of SAMMY-seq comparisons

To compute the differential enrichment analysis of the comparisons we firstly rebin and normalize the tracks as described above with 150 Kb rebinning size and 50Kb as merging distance of the blacklisted regions. We compute the consensus tracks for the control condition and calculate the intervals the standard deviation *2 (SD*2), obtaining consensus ± SD*2 for each genomic bin. Given the signal distribution of the consensus bins we filtered out all the bins within a range value ≤ ±0.1. After the filtering, the remaining bins are assigned to the correspondent fraction, S2S if the control condition signal is positive and S3 if is negative. To find the differential enriched bins in the test condition we sub-selected the bins in which all samples were not in the consensus ± SD*2 intervals. The selected bins were then put in a category defined as “up” if all the test samples increased the signals, and “down” if they decreased their signals. Therefore, we obtained 4 distinct categories S2S_up, S2S_down, S3_up, S3_down.

SAMMY compartments

For the SAMMY-seq protocols, we loaded the bin enrichment tracks (RPKM) of each fraction and computed the correlation between pairs of bins. Here, the correlation was computed between *m*-dimensional vectors of enrichment signals, where m was the number of fractions in the specific protocol (*i.e.*, m=3).
For each input correlation matrix, the first eigenvector was obtained through principal component analysis decomposition in R statistical environment (prcomp, stats package, center=FALSE and scale=TRUE, rotation component of the returned object). The sign of the first eigenvector was defined using gene density: the group of bins with the highest gene density was marked as "A" compartment (positive sign), and the group with the lowest gene density was marked as "B" (negative sign). Chromosome eigenvector values were divided by the absolute maximum value for visualization purposes. All the analyses were made using R. We partially reimplemented CALDER algorithm (version 1.0, 2020-09-01) to accommodate the unusual format of SAMMY-seq data, maintaining the primary set of core functions (remove_blank_cols, fast_cor,

generate_compartments_bed, HighResolution2Low_k_rectangle, get_PCs, bisecting_kmeans, project_to_major_axis, get_best_reorder, get_cluser_levels), their default parameters and the steps intended in the original paper[47] (8).
The pairwise distance matrix binned at 250 kb. In particular, the pairwise distance matrix was computed using for each pair of bins the Euclidean distance (dist, stats package, method="euclidean") of the m-dimension vectors, where m was the number of fractions in the specific protocol (*i.e.*, 3). Bins with null contacts or signals were removed from both matrices.
Briefly, for each input matrix, the algorithm: computed the pairwise correlation matrix and identified the sub-domain boundaries; computed the binary trend matrix and its decomposition using ten principal components; iteratively clustered the domains to obtain their hierarchy; sorted the domains hierarchy based on the projection of the first two components; divided them into 2 groups (form the most closed, *i.e.*, B to the most opened, i.e., A). For more details on the original CALDER procedure and the interpretation of sub-compartments please see[47].

Gene Ontology (GO) permutation test

Available gene ontologies methods for transcriptomic analyses assume that each gene has the same probability of being represented. In genome analysis this is not the case, because we remove genes that by default that will never be encountered due to sequence characteristics (genes in blacklisted regions). Another problem to consider is: there are gene clusters with similar functions in the same genomic regions. These genes are often in same pathways. Furthermore, in SAMMY-seq, broad peaks include multiple TSS. All of these issues would systematically inflate the obtained P-values of the resulting GOs. We develop a specific analysis to address these aforementioned problems. The analysis consists in a permutation test, that relies on the assumption of exchangeability, a plausible assumption with genomic regions. The steps were: i) selection of all the genes that show an overlap of the TSS, in a range of 2500bp upstream and 500 bp downstream, with the tested genomic regions; ii) selection of all ontologies that incorporate the overlapping genes, keeping the count of the

number of genes found in each ontology (*X*); iii) random shuffling (permutation) of all the accounted genomic regions along the genome keeping the size intact and avoiding overlapping with the blacklisted regions by masking them. Selection of genes (as explained in point i) after random repositioning and counting of number of genes associated to the GO described in point ii (x1). This procedure was repeated 10.000 times (*B*) for each gene ontology selected at point ii (*x1*…*xB*). iv) Counting for each GOs number of times when by permutation we found the number of genes associated with a GO to be greater than the selected regions. v) Dividing by the number of permutation the P-values associated to each GOs. This can be summarized by the formula:

$$P = \frac{\sum_{j=1}^{B} I(x_j \geq X)}{B}$$

Where *I* is the function which takes the value zero when it is false, and one when the is true[48]. vi) We then correct all GOs P-value with Benjamini Hochberg method. This analysis has been performed in R (v4.3) using the GO database R library "org.Mm.eg.db" with the database updated at 2023-Mar05.

4.5 References

1. Lai, W. & Wong, W. Progress and trends in the development of therapies for Hutchinson–Gilford progeria syndrome. *Aging Cell* **19**, e13175 (2020).
2. Misteli, T. The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* **183**, 28–45 (2020).
3. Gilbert, N. *et al.* Chromatin Architecture of the Human Genome. *Cell* **118**, 555–566 (2004).
4. Huisinga, K. L., Brower-Toland, B. & Elgin, S. C. R. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma* **115**, 110–122 (2006).

5. Blasco, M. A. The epigenetic regulation of mammalian telomeres. *Nat Rev Genet* **8**, 299–309 (2007).
6. Gilbert, N. & Allan, J. Distinctive higher-order chromatin structure at mammalian centromeres. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11949–11954 (2001).
7. Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**, 272–285 (2007).
8. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).
9. van Steensel, B. & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780–791 (2017).
10. Kind, J. *et al.* Single-Cell Dynamics of Genome-Nuclear Lamina Interactions. *Cell* **153**, 178–192 (2013).
11. Kind, J. & Van Steensel, B. Genome–nuclear lamina interactions and gene regulation. *Current Opinion in Cell Biology* **22**, 320–325 (2010).
12. Lund, E. *et al.* Lamin A/C-promoter interactions specify chromatin state–dependent transcription outcomes. *Genome Res.* **23**, 1580–1589 (2013).
13. Collas, P., Lund, E. G. & Oldenburg, A. R. Closing the (nuclear) envelope on the genome: How nuclear lamins interact with promoters and modulate gene expression: Prospects & Overviews. *BioEssays* **36**, 75–83 (2014).
14. Meuleman, W. *et al.* Constitutive nuclear lamina–genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* **23**, 270–280 (2013).
15. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
16. Zheng, X. *et al.* Lamins Organize the Global Three-Dimensional Genome from the Nuclear Periphery. *Molecular Cell* **71**, 802-815.e7 (2018).
17. Bianchi, A. *et al.* Dysfunctional polycomb transcriptional repression contributes to lamin A/C–dependent muscular dystrophy. *Journal of Clinical Investigation* **130**, 2408–2421 (2020).

18. Pegoli, G., Lucini, F., Mozzetta, C. & Lanzuolo, C. Single Myofiber Isolation and Culture from a Murine Model of Emery-Dreifuss Muscular Dystrophy in Early Post-Natal Development. *JoVE* 61516 (2020) doi:10.3791/61516.

19. Cesarini, E. *et al.* Lamin A/C sustains PcG protein architecture, maintaining transcriptional repression at target genes. *Journal of Cell Biology* **211**, 533–551 (2015).

20. Kolb, T., Maaß, K., Hergt, M., Aebi, U. & Herrmann, H. Lamin A and lamin C form homodimers and coexist in higher complex forms both in the nucleoplasmic fraction and in the lamina of cultured human cells. *Nucleus* **2**, 425–433 (2011).

21. Briand, N. & Collas, P. Laminopathy-causing lamin A mutations reconfigure lamina-associated domains and local spatial chromatin conformation. *Nucleus* **9**, 216–226 (2018).

22. Oldenburg, A. *et al.* A lipodystrophy-causing lamin A mutant alters conformation and epigenetic regulation of the anti-adipogenic *MIR335* locus. *Journal of Cell Biology* **216**, 2731–2743 (2017).

23. Sebestyén, E. *et al.* SAMMY-seq reveals early alteration of heterochromatin and deregulation of bivalent genes in Hutchinson-Gilford Progeria Syndrome. *Nat Commun* **11**, 6274 (2020).

24. Della Valle, F. *et al. LINE-1* RNA causes heterochromatin erosion and is a target for amelioration of senescent phenotypes in progeroid syndromes. *Sci. Transl. Med.* **14**, eabl6057 (2022).

Chapter 5 Chromatin regulation in muscle satellite cells in aging.

5.1 Abstract

It is known that the aging process hinder the homeostatic gene expression which is modified by multiple factors. One of the most crucial is the three-dimensional conformation of the genome which at homeostatic condition models the spatiotemporal gene expression. One of the most dangerous effects during aging is sarcopenia, characterized by muscle fat infiltration and reduced skeletal muscle performance. Sarcopenia is normally caused by a concomitant set of intrinsic and extrinsic factors, among them also the reduction and impairment function of muscle stem cell pool. To deeply characterize chromatin changes along the entire lifespan, we used a cohort of mice at representative stages of the aging process, starting from postnatal, to geriatric stage. Our findings highlight how MuSCs over life exhibit a progressive modification of chromatin solubility that may interfere with proper pool maintenance.

5.2 Introduction

Aging, is defined as a gradual deterioration of functionality that affects practically every organism[1]. In particular, one affected tissue which become pathological over time is the skeletal muscle. This progressive deterioration in muscle strength and functionality is called sarcopenia[2]. It is characterized by fat infiltration in muscles (marbling), a low-level of inflammation, increased level of ROS, weakened stress response capacity, and muscle stem cell exhaustion[3]. All these factors are considered hallmarks of aging and are tightly intertwined together[1]. Muscle stem cells or muscle satellite cells (MuSCs) are a set of almost unipotent stem cells that lie beneath the basal lamina of the muscle[4,5]. At homeostatic conditions these cells remain quiescent; upon damage MuSCs activate and by symmetric and asymmetric division they both ensure pool replenishment and muscle restoration[6–8]. The cellular and molecular events which contribute to aging process and concluding in senescence are plenty. In fact, during aging the microenvironment in the muscle changes keeping the MuSCs in a continued state of pre-activation, which, if not reverted, ultimately lead to senescence[9]. Molecularly, aged MuSCs are subjected to the disfunction of the Notch-p53 signaling pathway which leads to defects in cell cycle. When the signaling pathway is activated it start a mitotic catastrophe, which impairs the regenerative

and self-renewal potential of the cells[10]. In addition, the p38/MAPK with the Janus kinases (JAKs) and signal transducer and activator of transcription proteins (STATs) signaling (JAK/STAT) hinders the asymmetric division, impairing regeneration and leading to pool exhaustion[11,12]. Another fundamental process that have a great influence on the homeostasis of the cells during the aging process is the chromatin remodeling as its 3D structure govern cell identity and spatiotemporal gene expression[13,14]. It is already reported that during aging, there is a widespread loss of compaction in heterochromatin. The cells face HM loss such as H3K9me3 and H3K27me3 due to a progressive depletion of lamin B1 that cause LAD detachment and gradually prompt cellular senescence[1,15,16]. Various studies described a general increase in chromatin accessibility[17–20] in old MuSCs versus young MuSCs, accompanied by a 4% change in the general chromatin compartmentalization[20].

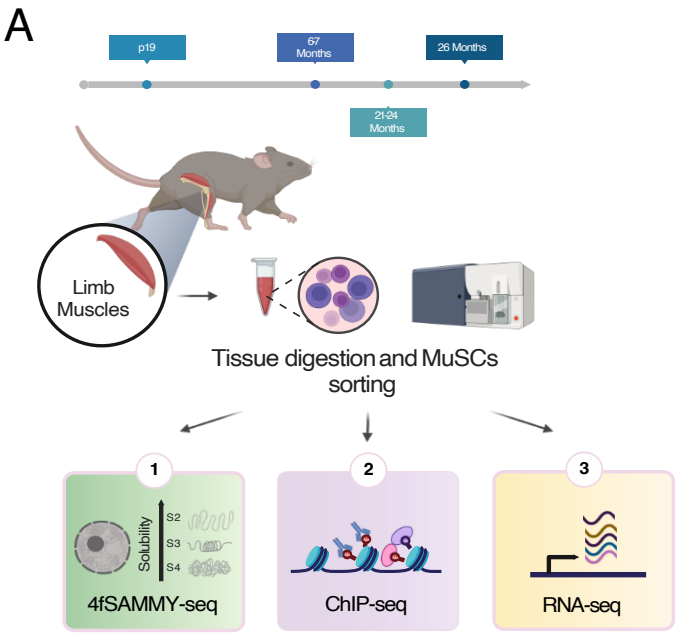
Other works, using transcriptomics and proteomics methods applied to young and old MuSCs, have shown a dysregulation of histone mark profile during aging [17,21]. Mass-spectrometry-base analysis in old and young MuSCs highlighted increased global levels of H3K9me2 and H3K27me3 in old cells, confirming Liu's noteworthy paper in muscle aging[22]. This data, apparently in contrast with the heterochromatin lost, is accompanied by an increase in open chromatin marks during activation, such as H3 and H4 acetylation[21], suggesting a remodelling in both euchromatin and heterochromatin compartments. However, also LaminA/C play a role in interaction and compartmentalization of chromatin as it is involved in the PcG proteins foci establishment. This interaction leads to correct position of PcG and ensures a correct muscle differentiation[23]. Genetic mutations of lamin A/C cause diseases called laminopathies; among them there are a class that affect the skeletal muscle with a dystrophy, as in the case of Emery Dreifuss Muscular Dystrophy (EDMD). Studies on the mouse model of EDMD showed that, in MuSCs, a redistribution of H3K27me3 on off-target as the cell cycle inhibitors ($p16^{ink4a}$) drives the cell to a loss of cell identity and premature senescence[24]. Here, we will use our innovative technique SAMMY-seq on limited amount of MuSCs to analyze single mice as samples. We set up a systematic

analysis of the most representative stages of the mouse lifespan: post-natal stage (19-21 days after birth), adult (7 months), old (21-24 months) and geriatric (28 months). In addition, we will integrate RNA-seq and ChIP-seq datasets to elucidate the epigenome remodeling in muscular aging.

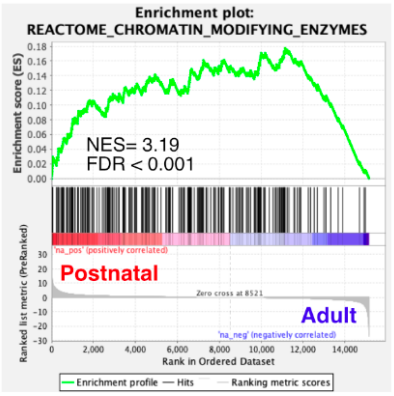
5.3 Epigenetic remodeling activity in aging.

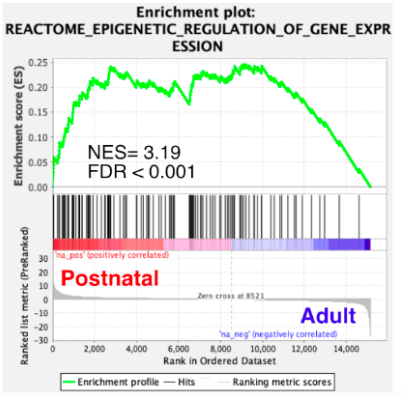
We investigated the chromatin remodeling on muscle stem cells (MuSCs) at four different life stages: postnatal (19-21 days after birth), adult (7 months), old (21-24 months), and geriatric (28 months). We collected at least 3 samples from each stage (biological replicates) with the exception of the old for which we have only 2. For each sample, we collected muscle from the hind-limb muscle, enzymatically digested and fluorescently stained the cells. We then used the fluorescence-activated cell sorting (FACS) to separate the muscle stem cells from the other cell types. After sorting, we performed genome-wide analyses, such as RNA-seq and SAMMY-seq, on all samples(**Figure 1A**). First of all, we did a differential expression analysis of the RNA-seq. In the passage from postnatal to adult the GSEA shown a change in epigenetic modifiers, suggesting a genome remodeling (**Figure 1B**). This could be dependent by the high number of activated MuSCs in postnatal age[25]. To investigate the epigenome regulation along the ageing MuSCs we analyzed some chromatin remodelers gene activity involved in the deposition or removal of HMs (**Figure 1C**). The heatmap display a generalized increased level of chromatin remodelers' transcripts in the postnatal stage which confirm a strong remodeling in differentiation. Notably, not only lamin A/C but both lamin types shows an increased RNA transcription in the postnatal stage which suggest their implication in the MuSCs differentiation[23]. Then, chromatin remodelers are transcriptionally downregulated until the geriatric stage, where we found a trend of upregulation. To better characterize the epigenetic differences in adult and postnatal MuSCs, we selected already available ChIP-seq from already published scientific papers and produced the missing HM which were not already done de novo for H3K36me3, H3K4me3 (as

euchromatin marks) or H3K9me3, H3K27me3 (as constitutive and facultative heterochromatin marks, respectively).



B





C

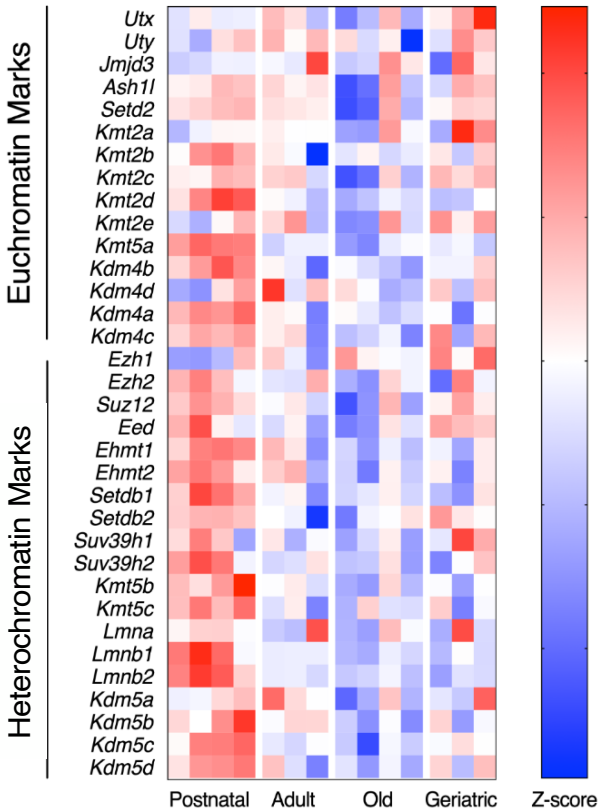


Figure 1. RNA expression in the passage from postnatal to adult show an upregulation of chromatin remodeling pathways. A. Graphical abstract **B.** GSEA enriched patwhays from Reactome database[26], downloaded from Gprofiler site[27]. **C.** Heatmap of chromatin remodelers genes for eu and hetero chromatin expressed in Z-score calculated from TPM.

The already available ChIP-seq were re-analyzed from the raw data with the same tools as the newly produce ones to ensure maximal comparability. Then, we called the peaks for each HM and associate the genes affected by each histone mark. The association between genes were performed differently for each HM based on their distribution along the genome (**see methods**). Then, to evaluate the degree of remodeling we overlapped the affected genes by each HM between conditions (**Figure 2**). We found a great overlapping of euchromatin HMs in the two conditions (**Figure 2A, B**), with an overall increase of H3K36me3 in the adult stage.

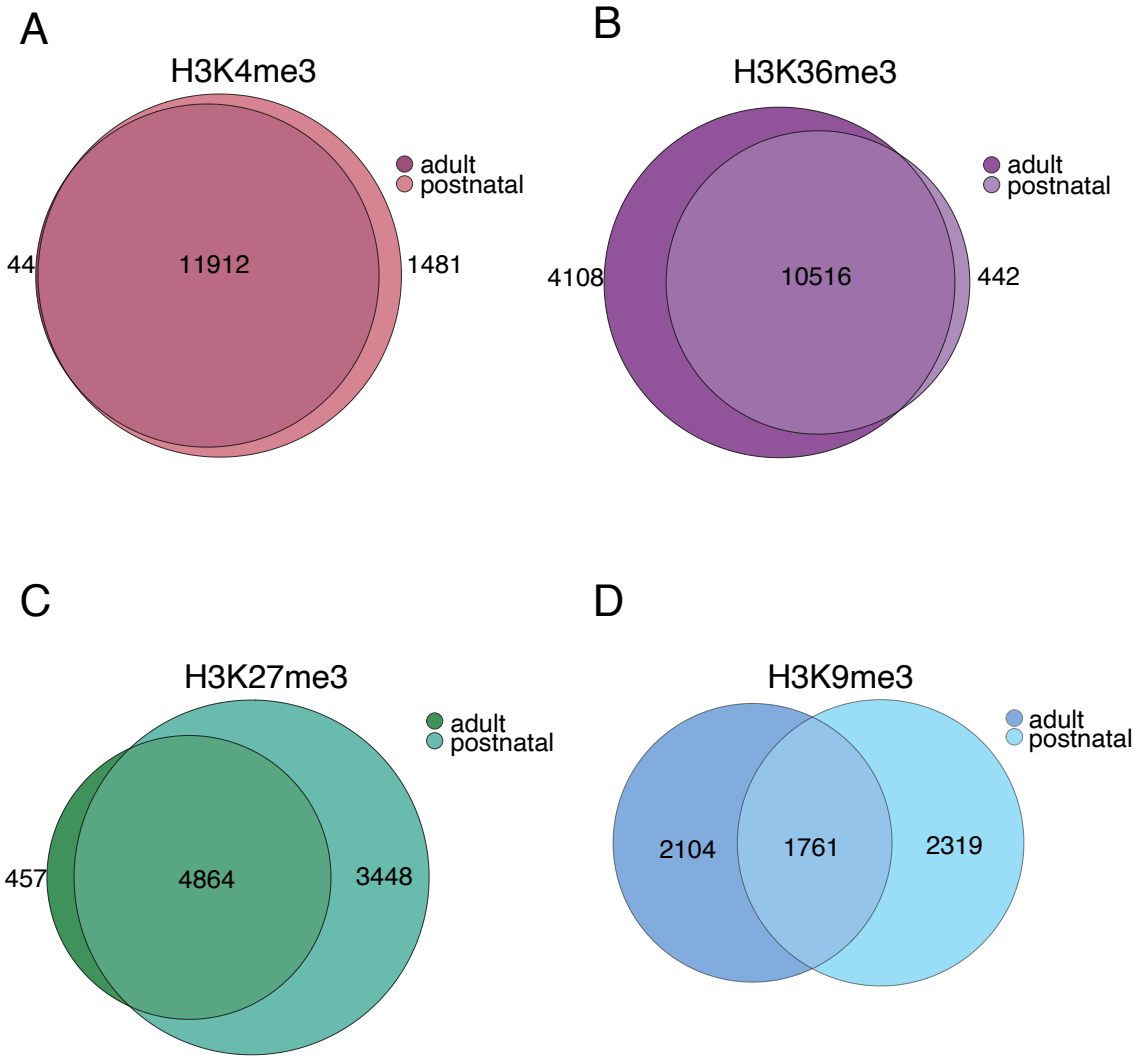


Figure 2. Venn diagrams of associated genes to each HM between adult and postnatal. A. H3K4me3 Jaccard Score 0.89 **B.** H3K36me3 Jaccard Score 0.7 **C.** H3K27me3 Jaccard Score 0.55 **D.** H3K9me3 Jaccard Score 0.28.

Heterochromatin marks behave differently: H3K27me3 peaks in postnatal are more than those in adult stage. This is in line with the increase of heterochromatin due to cell differentiation, necessary to repress other cell lineage genes. H3K9me3, mark of constitutive heterochromatin is the more variable, covering different genomic regions in postnatal and adult. However, the genome coverage is quite different with a decrease in the adult stage (38.83% in postnatal and 32.50% in adult). To confirm this trend we performed some immunofluorescence on MuSCs extracted from postnatal and adult mice (**Figure 3A**). We confirmed a

slight decrease in H3K9me3 signal intensity. Overall, these analysis suggest that the heterochromatin undergoes to an epigenetic remodeling in the passage from postnatal to adult.

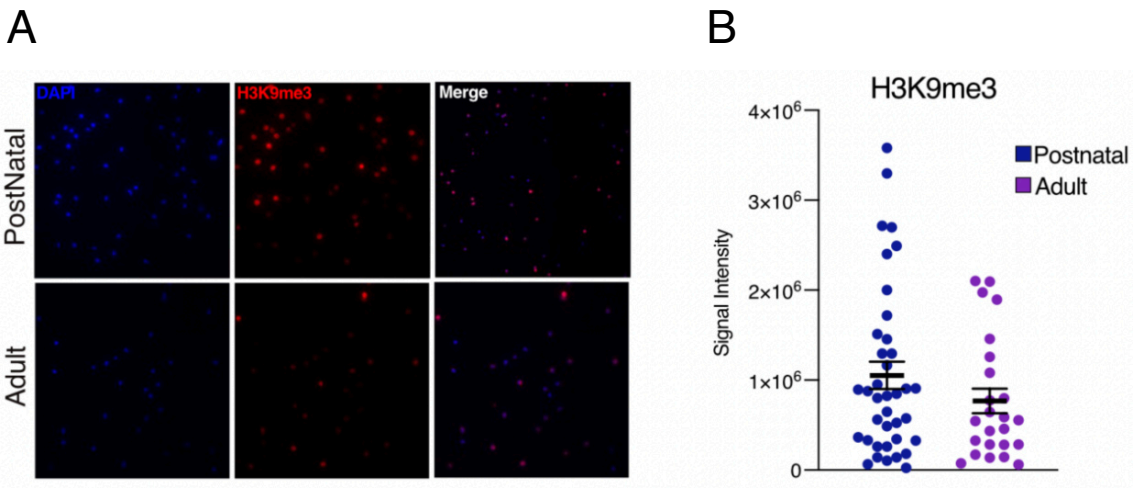


Figure 3. Immunofluorescence assay of the H3K9me3 in postnatal and adult satellite cells. **A.** H3K9me3 IF in postnatal and adult MuSCs, **B.** Quantification of the IF signal intensity. Data are shown as mean ± SEM (Standard Error of the Mean).

5.4 SAMMY-seq profile during aging

Considering that our technology is able to discriminate euchromatin and heterochromatin within a single sample and with low number of cells (see chapter 2) we decided to analyze the chromatin accessibility in all the aging steps taking advantage also from the epigenetic profiles that we generated in adult and postnatal MuSCs. Firstly, we evaluate the quality of SAMMY-seq postnatal

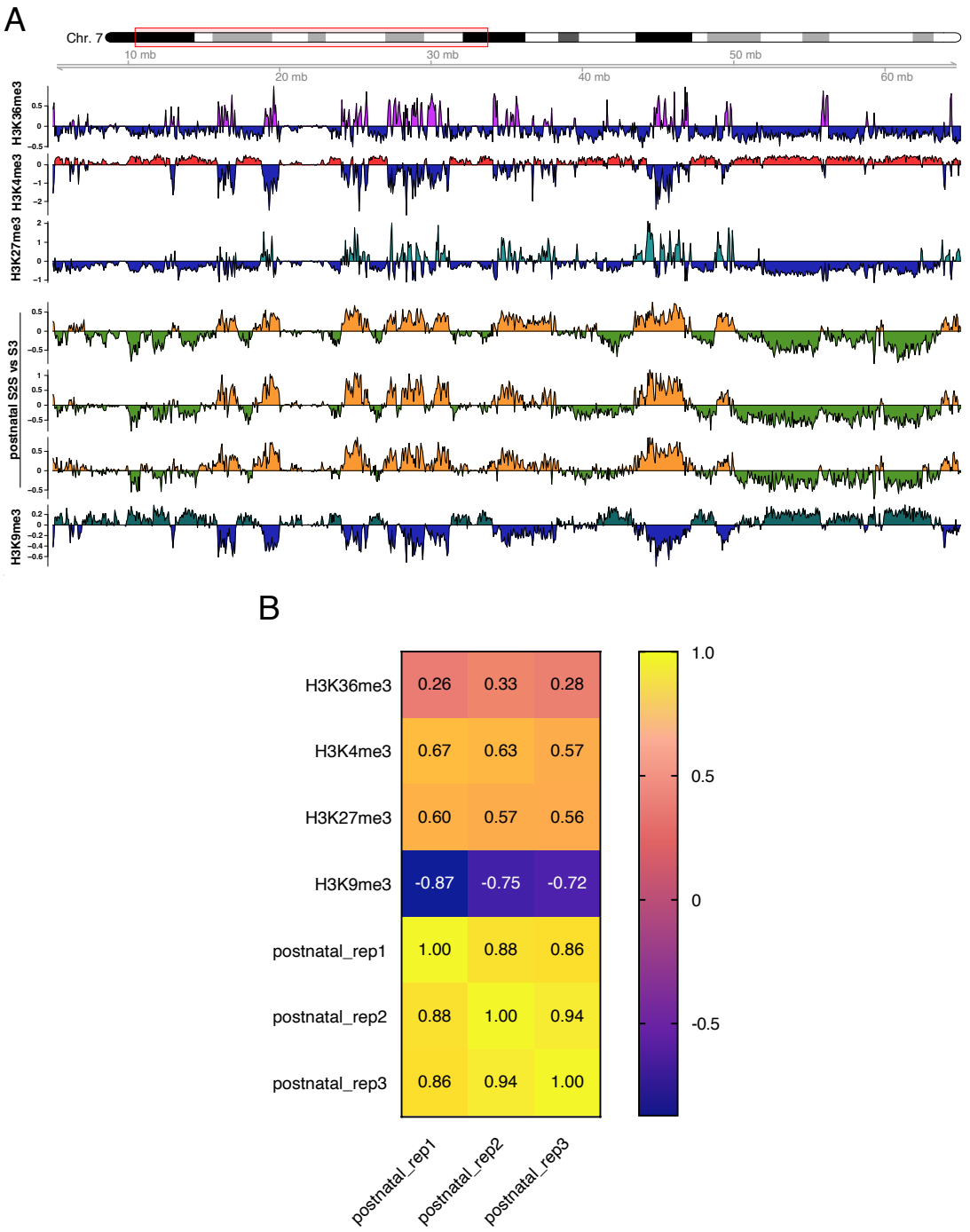


Figure 4. 4fSAMMY-seq comparison signal recapitulate postnatal chromatin accessibility state. Representative genomic region of chromosome 7 (chr7:4,070,191-68,514,866). **A.** Smoothed differential signal for adult ChIP-seq and SAMMY-seq. From top to bottom: H3K4me3 (red), H3K36me3 (purple), H3K27me3 (aqua green), 4fSAMMY-seq (S2SvsS3) comparisons of mouse replicates (orange and light green), H3K9me3 (dark green). **B.** Genome-wide Spearman correlation heatmap of the H3K4me3, H3K36me3, H3K27me3, H3K9me3 and postnatal SAMMY-seq samples.

Comparisons S2SvsS3 with respective ChIP-seq (**Figure 4A**). To evaluate the reproducibility and the fractionation quality, we did a genome wide correlation analysis (**Figure 4B**). The correlation matrix showed a high level of reproducibility between samples. In addition, SAMMY-seq postnatal samples showed a strong anticorrelation with H3K9me3 (from -0.72 to -0.87) which confirmed the good performance of the technique with heterochromatin. On the other hand also the S2S enriched regions displayed a good correlation with H3K4me3 mark, from 0,56 to 0,67 of range and weaker correlation with H3K36me3 mark (range from 0.26 to 0.33, **Figure 4B**). Interestingly, facultative heterochromatin, marked by H3K27me3, correlated with open and soluble genome portion from 0,56 to 0,60. This suggest that H3K27me3 enriched genomic regions adopt a specific state of solubility. This hypothesis is in line with other studies that describe H3K27me3 heterochromatin highly dynamic[28] . Then, a similar analysis was done in the adult stage (**Figure 5**). Again we found an high level of SAMMY-seq reproducibility among samples, with a range from 0.83 to 0.94 (**Figure 5B**). However, we found different trends of correlations between SAMMY-seq and the HMs (**Figure 5B**). Correlation with H3K36me3 increased to a range of 0.78-0.9 while the H3K4me3 decreased to a range of 0.40-0.44. In parallel, anticorrelations of H3K9me3 decreased in a range from -0.39 to -0.49 (**Figure 3B**). H3K27me3 display no correlation with SAMMY-seq having a range between -0.10 to -0.14. The high correlations of adult and postnatal SAMMY-seq samples (**Figure 5B**) suggest that in healthy condition whereas the degree of accessibility is not greatly affected, the epigenetic remodeling is subjected to various smaller changes.

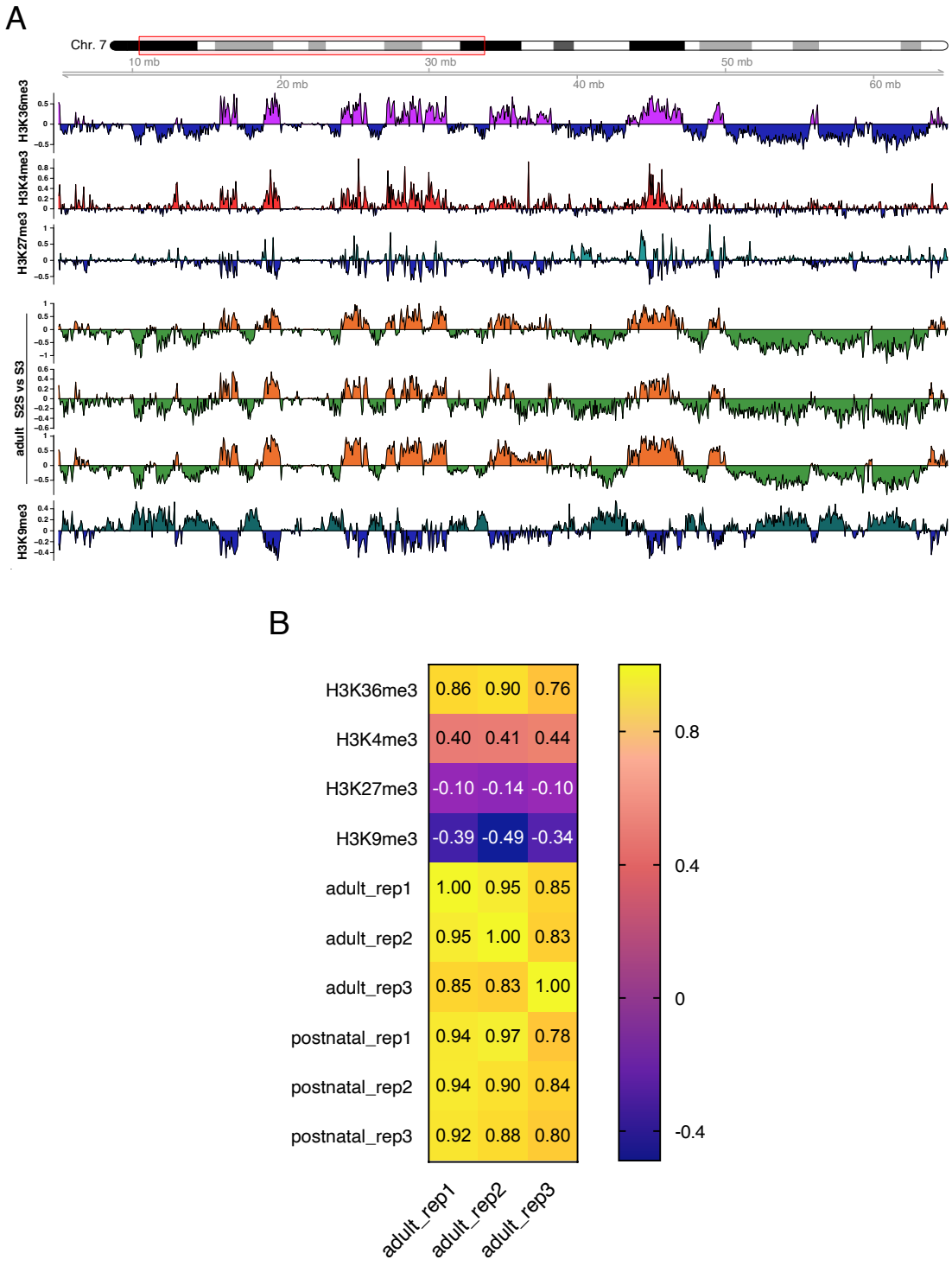


Figure 5. SAMMY-seq signal also recapitulate adult chromatin accessibility state. Representative genomic region of chromosome 7 (chr7:4,070,191-68,514,866). **A.** Smoothed differential signal for adult ChIP-seq and SAMMY-seq. From top to bottom: H3K4me3 (red), H3K36me3 (purple), H3K27me3 (aqua green), 4fSAMMY-seq (S2SvsS3) comparisons of mouse

replicates (orange and light green), H3K9me3 (dark green). **B.** Genome-wide Spearman correlation heatmap of the H3K4me3, H3K36me3, H3K27me3, H3K9me3, postnatal and adult SAMMY-seq samples.

To analyze these changes, we deeply analyzed the SAMMY-seq of the postnatal, adult, old and geriatric time points by using the consensus tracks. We decided to compare adult vs postnatal, adult vs old and geriatric vs old. We removed the regions with a non-clear enrichment, filtering the regions with a mean signal less than ±0.1 value in the wild type, and therefore, the corresponding regions in the other time point. As shown in (**Figure 6**), the scatter plot of the S2SvsS3 consensus tracks display different levels of solubility dispersion. The maximum overlap expected if identical would be represented on the diagonal with an r^2 of 1.

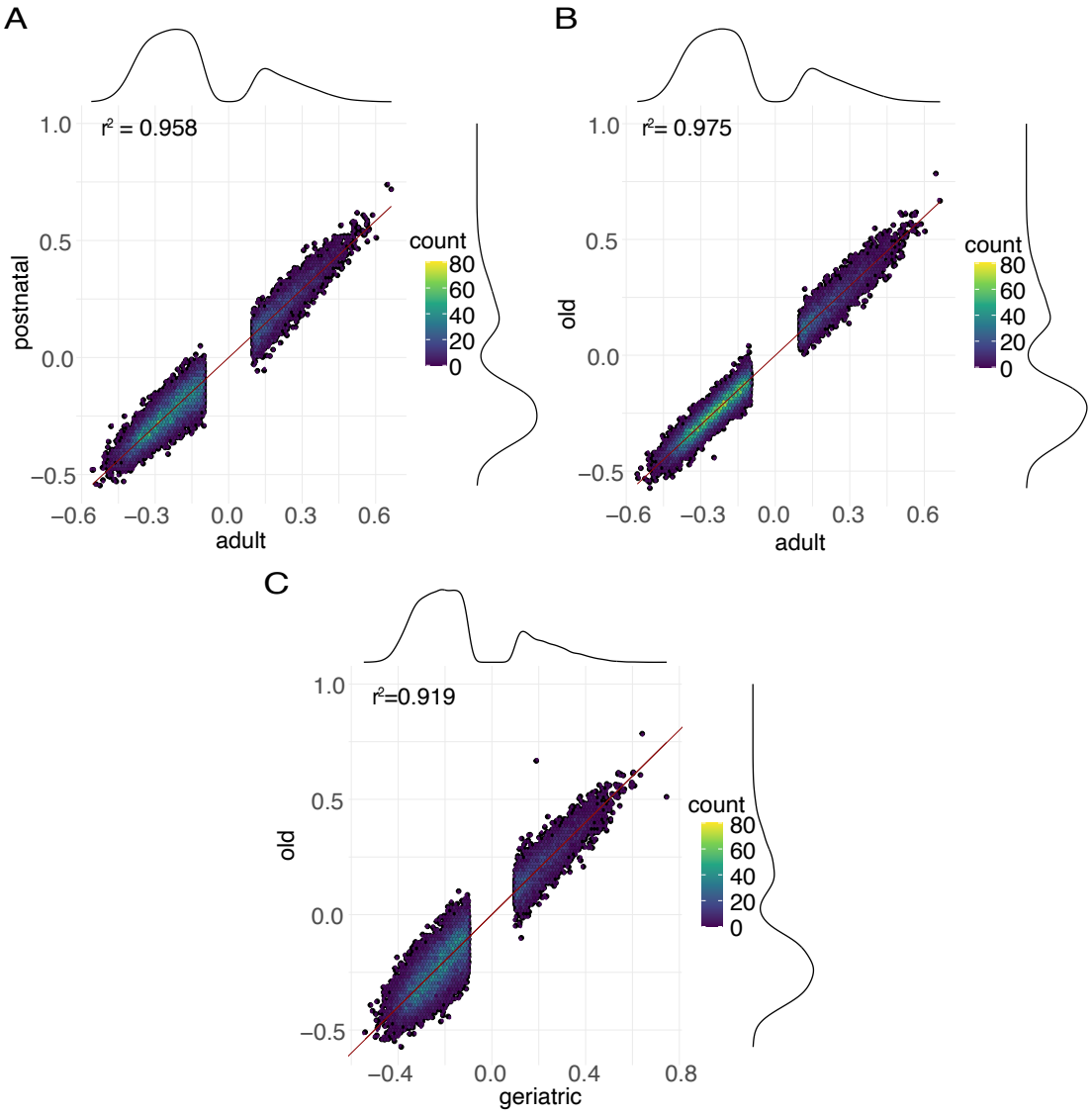


Figure 6. Scatter plot of the consensus signal between contrast do not show an increased dispersion with age. The consensus signal of the 4FSAMMY-seq comparison at genome wide level, rebinned at 150Kb for each time point, then compared for each contrast. After discarding all data points below a threshold of 0.1, the red line shows the linear regression model. The color of each dot represents its density, from low (dark blue) to high (yellow). The lines at the top and right of the plot show the distribution of each consensus signal values. **A.** adult vs postnatal **B.** old vs adult **C.** geriatric vs old.

Whereas the contrast between adult vs postnatal show a r^2 of 0.958 (**Figure 6A**), the lowest dispersion signal is the adult vs old with a r^2 of 0.975 (**Figure 6B**). This suggests that SAMMY-seq can capture a difference due to the activated state of

a subset of postnatal MuSCs. The most dramatic remodeling was seen in the comparison between the old vs geriatric which shows the highest level of dispersion, suggesting a pathological change in chromatin solubility.

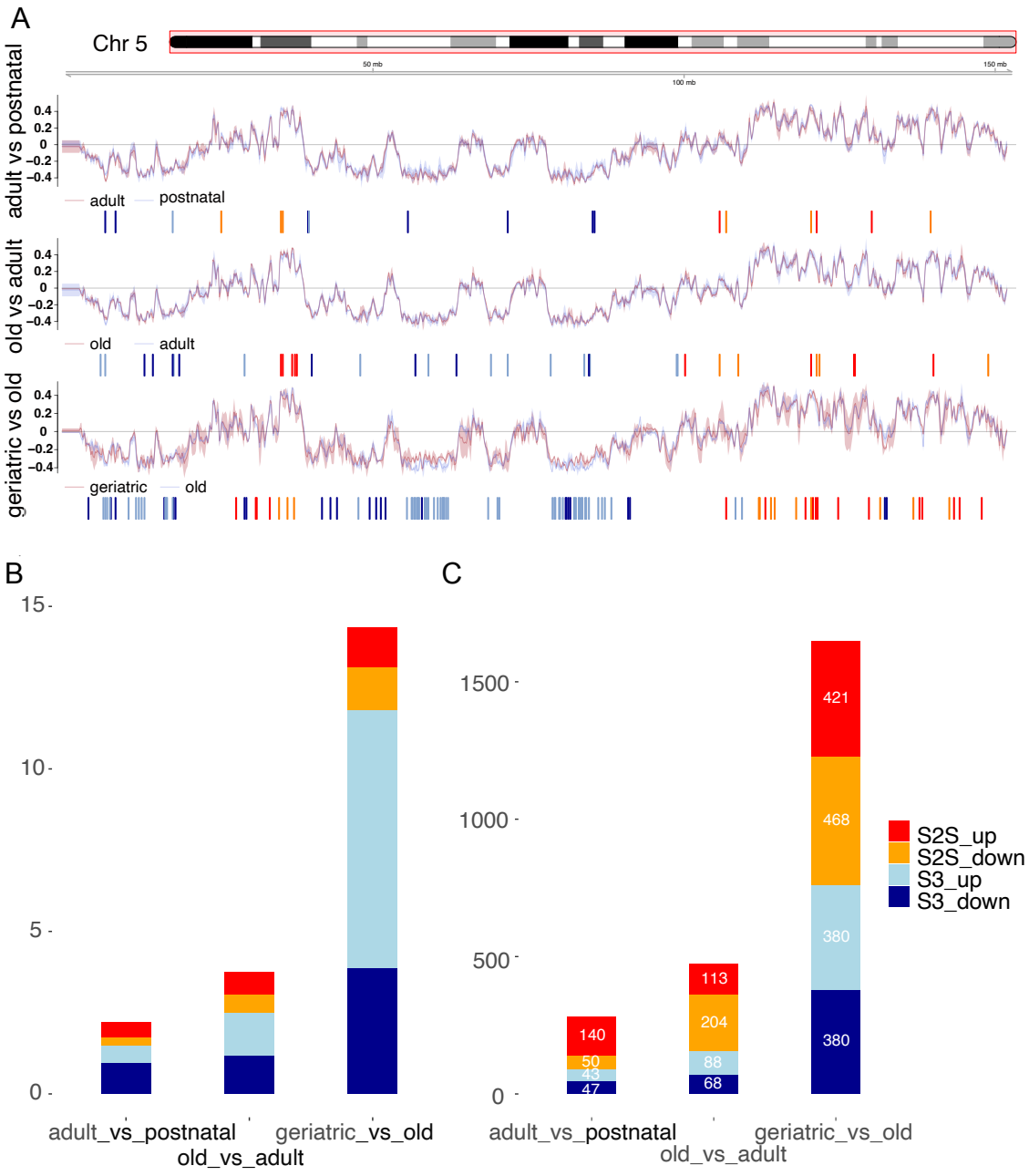
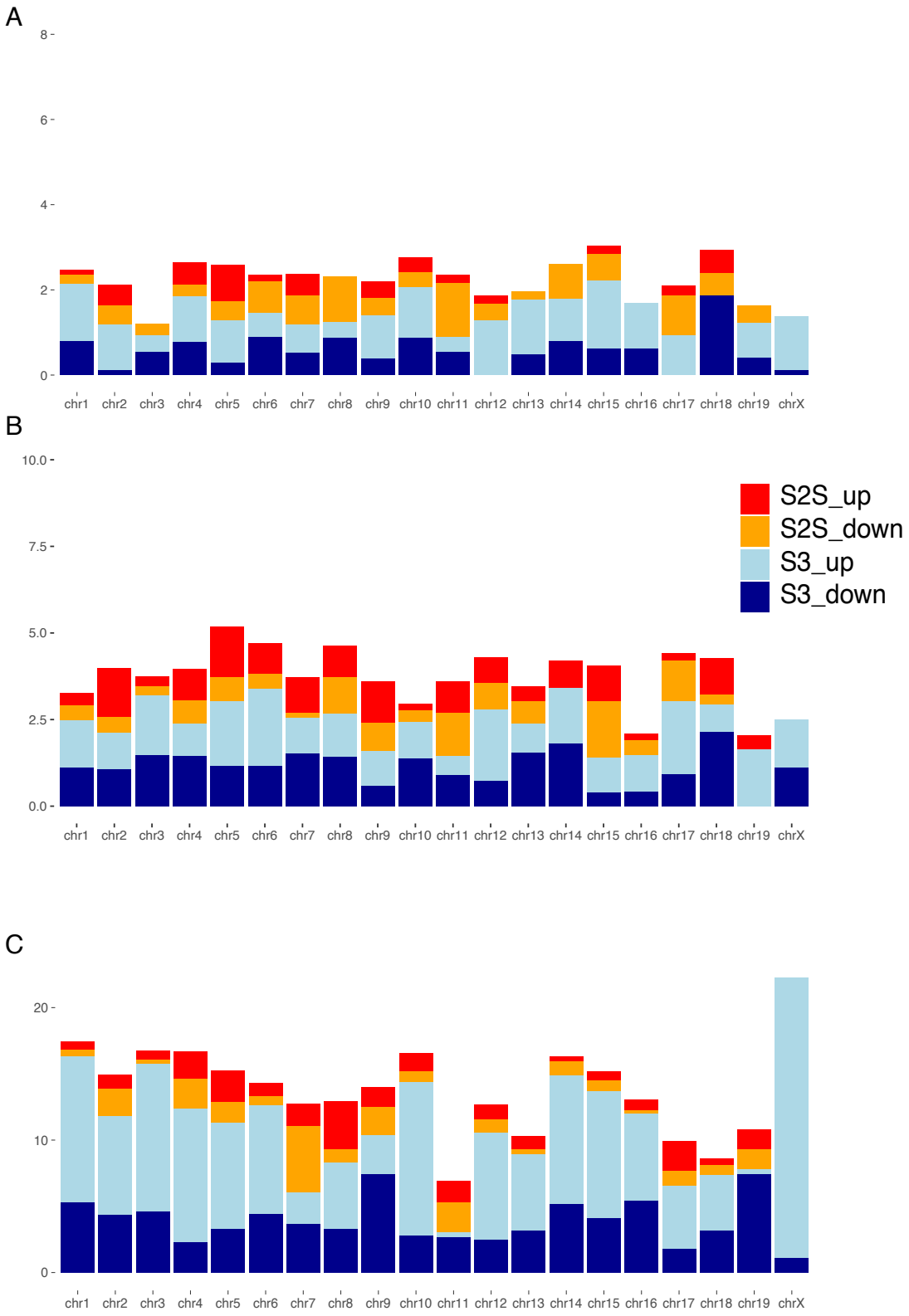


Figure 7. Chromatin remodeling increase with ageing. A. Chromosome 5 is shown as representative chromosome. The lines in blue and red are the consensus track of 4fSAMMY-seq. From top to bottom adult vs postnatal; old vs adult and geriatric vs old. The lighter color shadows

for each track represent the standard deviation of the consensus track. Below each track pairs, the differential enriched regions are highlighted as follow: S2S_up (red), S2S_down (orange), S3_up (light blue) and S3_down (dark blue). **B**. Stacked bar plot of the of genome percentage affected by the remodeling type. Color scheme: S2S_up (red), S2S_down (orange), S3_up (light blue) and S3_down (dark blue). **C**. Stacked bar plot of the number of genes with the TSS overlapping with the regions with a range of upstream 2500bp and downstream 500 bp. Color scheme: S2S_up (red), S2S_down (orange), S3_up (light blue) and S3_down (dark blue).

We then proceeded to quantify the differential enrichment for each contrast, with a representative chromosome (**Figure 7 A**). We classified the differential soluble regions into 4 categories: S2S_up (red), S2S increased accessibility; S2S_down (orange), S2S with decreased accessibility; S3_up (light blue), S3 with increased accessibility and S3_down (dark blue), S3 with decreased accessibility. Notably, with this analysis we select only genomic regions with different solubility in all analysed condition versus all reference samples. The postnatal vs adult contrast shows less genomic portion interested by a chromatin accessibility remodeling compared to the adult vs old comparison (**Figure 7B**). We found a clear trend of solubility alterations going from postnatal toward old and geriatric mouse. Considering the dispersion shown in **Figure 6A**, this result could appear in contrast. However, the two analysis are different from a technical point of view and local differences can be underestimated in the genome wide analysis. Geriatric MuSCs showed the most extensive remodeling of chromatin solubility affecting ~15% of the genome (**Figure 7B**). Interestingly, the heterochromatin, whose changes were shown with light and dark blue staining, was the most affected by the remodeling becoming generally more accessible. Then, we checked the distribution of chromatin solubility changes on chromosomes, finding no differences between adult and postnatal or old and adult (**Supplementary Figure 1 A B**).



Supplementary Figure 1. Percentages affected regions per chromosome. The percentages are calculated from the total chromosome sizes after the filtering of the signals < ±0.1. Color

scheme: S2S_up (red), S2S_down (orange), S3_up (light blue), S3_down, S2S (light grey) and S3 (dark grey). **A.** adult vs postnatal **B.** old vs adult **C.** geriatric vs adult

On the other hand, we found in geriatric MuSCs a preferential increase of heterochromatin solubility of the chromosome X (**Supplementary Figure 1**). Considering our results on Lamin A knock out described in the chapter 3 (**chapter 3 Supplementary Figure 1**), where we see a similar trend, this result suggests an involvement of Lamin A in the geriatric MuSC chromatin remodeling.

We then used the GO analysis with permutation on the differential soluble regions (S2S_up, S2S_down, S3_up, and S3_down). Since available gene ontology packages are tailored on RNA-seq, to avoid inflated P-values, we set up a custom gene ontology analysis based on permutation (**see methods**). However, only few sets of genomic regions lead to significantly enriched GOs (**Figure 8**). This could be due to the lower density of genes in heterochromatin (S3 up/down), where altered regions cover larger genomic portions. In addition, the lower quantity of genes inside the postnatal genomic regions would not end up to significant GOs. Nonetheless, were the GOs were too many, we performed the semantic analysis with revigo[29] and for each group we filtered out the GOs with a FDR < 0.01 (**Figure 8**). The **Figure 8A** show the biological pathways in the old vs adult contrast in the S2S_down gene set. The old MuSCs showed a globally lowered solubility in genes involved in a correct transcription and biosynthesis of nucleic acids which are processes fundamental for correct transcriptional activities and regulation[1]. It also showed an impairment in solubility in the cell surface receptors which are fundamental for a correct sensitivity for cell activation and differentiation processes[30]. Finally, it showed impairment in entering the cell cycle with a focus in chromosome segregation which is closely related to lamin B correct function[31]. Similar pathways were, especially those focused on chromosome integrity and segregation were found in semantic analysis of the S2S_down in geriatric (**Figure 8B**). In addition, in geriatric condition we found several terms that regards cell specification, including muscle related pathways. This suggest a loss of cell identity which resemble the premature ageing

observed in the EDMD[24]. In **Figure 8C** the semantic analysis of S2S_up in geriatric vs old contrast show an activation of the satellite cells yet with increased stresses and apoptotic pathways.

A

Figure 8 continues on the next page.

B

Figure 8 continues on the next page.

C

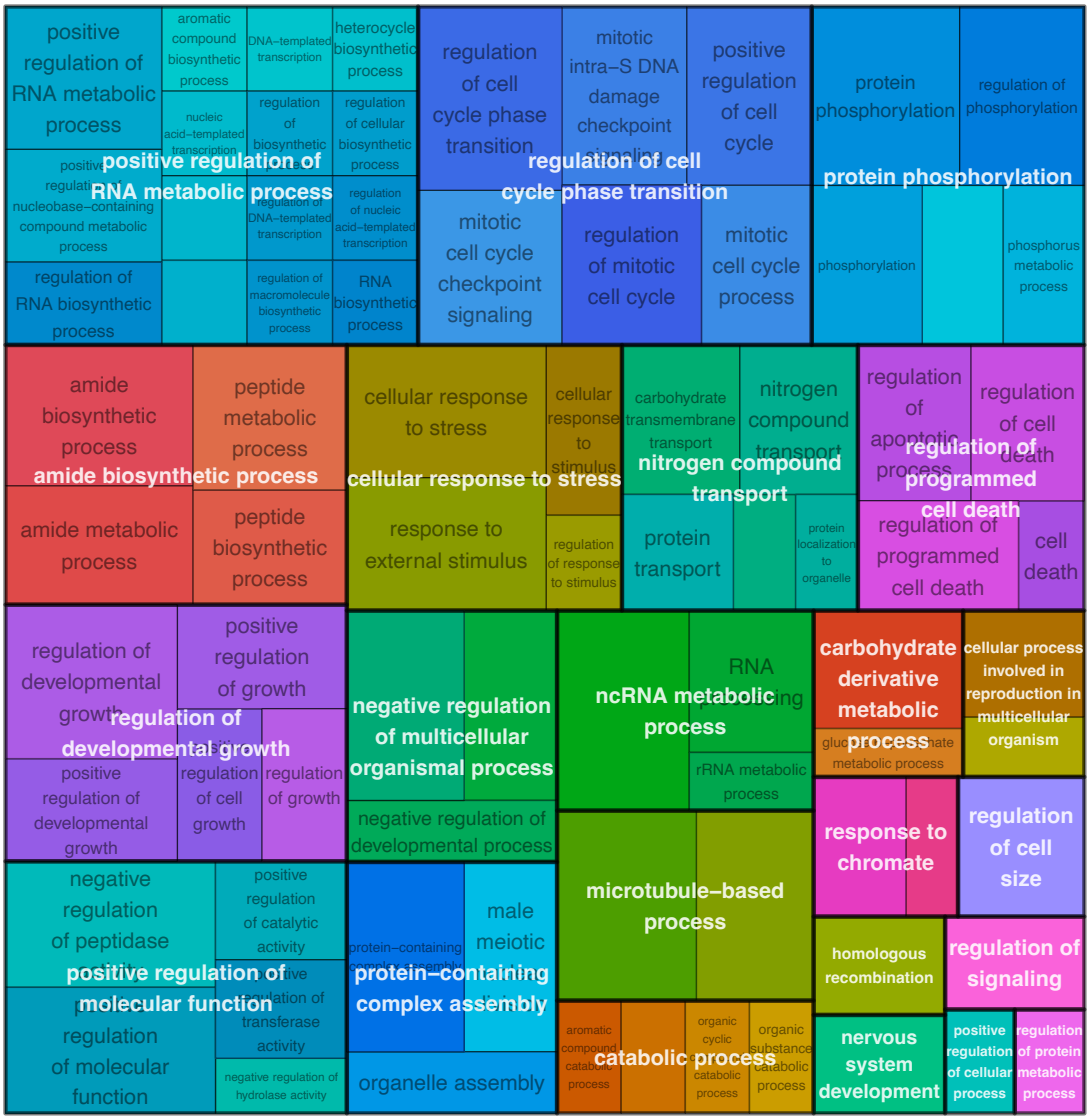


Figure 8 Semantic analysis of the Gene Ontologies computed from the genes overlapping the remodeled genomic regions. A. S2S_down in old vs adult. **B.** S2S_down in geriatric vs old **C.** S2S_up in geriatric vs old.

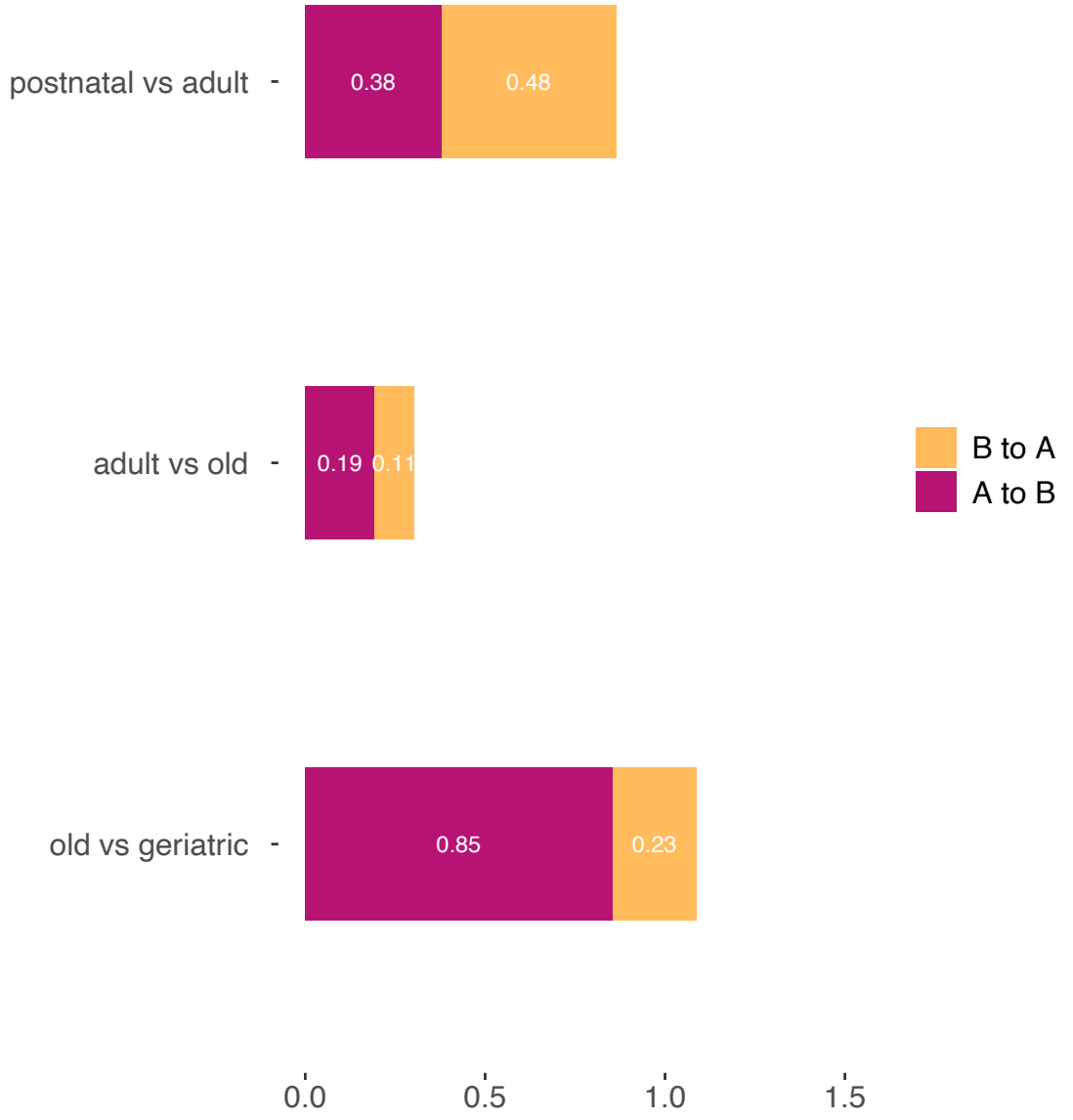


Figure 9 Compartmentalization remodeling during ageing. Percentage of genome that change compartmentalization, from A to B and vice versa per contrast. The first condition (postnatal, adult, old) in the label name is the starting compartment.

Using the analysis described in chapter 2, we also investigated the chromatin compartments in all ageing steps datasets calculating the chromatin compartments starting from 4fSAMMY-seq fractions (**see chapter 2 methods**).

Due to the lack in some cases of the fraction S4 we called compartments with a resolution of 250Kb considering only 3 fractions, S2S, S2L and S3. The genomic occupancy of the chromatin compartments that switch from A to B or vice-versa span from 0.11 to 0.85 % (**Figure 9A**). As expected from previous results (**Figure 7B**), the highest difference was found in the old vs geriatric comparison. Integration of RNA-seq and SAMMY-seq are ongoing to elucidate the functional impact of the observed chromatin remodeling. All together, these preliminary results give us the picture of chromatin remodeling during the most relevant stages in the lifespan.

5.5 Materials and methods

Mice

Wild-type C57BL/6 mice from "The Jackson Laboratories" were used. Mice were bred and maintained according to the standard facility procedure (San Raffaele Hospital, DIBIT 1, Milan), and all the experimental protocols were approved accordingly by the Italian Ministry of Health (IACUC 952). Mice were sacrificed at different times (P19-21, 3-7 months, 21-24 months, 28 months) by cervical dislocation.

Immunofluorescence on Muscle Stem Cells

To preserve the integrity of chromatin architecture, Skeletal muscle cell suspension was fixed in 1:10 formaldehyde solution (50mM Hepes-KOH pH7.5, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 11% formaldehyde - in H2O) for 9 minutes at room temperature in mild agitation and quenched with 125 mM Glycine before FACS sorting. Sorted muscle stem cells (MuSCs) were adhered on poly-L-lysine (Sigma, P8920) pre-coated coverslips for 30 minutes at RT. Then, they were fixed with 4% paraformaldehyde (PFA) (Sigma 30525-89) dissolved in PBS, for 7 minutes at RT. After 3 washes in PBS of 5 minutes each, MuSCs were permeabilized in 0.5% Triton X-100/PBS for 10 minutes at RT in mild agitation.

Blocking was performed with 5% BSA/PBS for 1 hour at RT. Samples were incubated with a rabbit primary antibody anti-H3K9me3 (Abcam, ab8898) diluted 1:1000 in 50% of blocking solution at 4 °C overnight. After 6 washes in PBS of 3 minutes each, samples were incubated with an anti- rabbit secondary antibody conjugated with Alexa Fluor 568 (Invitrogen, A11036) diluted 1:1000 in Blocking solution, for 2 hours at RT in the dark. After 6 washes in PBS of 3 minutes each, DNA was stained with DAPI (1:1000) in PBS for 10 minutes at RT in the dark. After 6 washes with PBS, coverslips were mounted on a slide with Prolong diamond antifade mountant (Thermo Fisher Scientific, P36961). Images were acquired using a Nikon Crest microscope using 40x objectives and analyzed using Fiji[32]

Muscle Stem Cell extraction and FACS sorting

Hind and Fore-limb muscles were isolated from sacrificed mice and digested 60/75 minutes in 2,4 U/ml of Dispase II (Roche, 04942078001), 2 mg/ml of Collagenase A (Roche, 1013586001), 0,4 mM CaCl2 (Sigma, C5670), 5 mM MgCl2 (Sigma, M8266), 0,1 mg/mL DNase I (Roche, 1014159001) in PBS 1X at 37°C in a water bath. The samples were resuspended in HBSS (Gibco, 14025-050) and implemented with 0,2% BSA (Sigma, A7030) to stop the enzymatic digestion. The cell suspension was serially filtered with 70 µm and 40 µm cells strainers and resuspended in HBBS +++ (0,2% BSA, 1% DNase I, 1% PenStrep (Euroclone, ECB3001) for the night. The day after, cell suspension was stained 30 minutes at 4°C with the following antibodies: PB- CD45 1:50 (eBioscience 48-0451), PB-CD31 1:50 (eBioscience 48-0311), PB-Ter119 1:50 (eBioscience 48-5921), FITC-Sca1 1:50 (eBioscience 11-5981), APC 7integrin 1:100 (AbLab, 67-001-05) and sorted with BD FACS ARIA SORP for: PB-CD45-/ PB-CD31-/ PB-Ter119-/ FITC-Sca1-/APC-7integrin+ (Muscle Stem Cell).

Sammy-seq protocol

Chromatin fractionation on sorted Muscle Stem Cells (20k-50k) was performed with minor adaptations to the protocol described in Sebestien et al.[33]. Cells were counted, washed in cold PBS and resuspended in cold cytoskeleton buffer CSK: 10 mM PIPES pH 6,8; 100 mM NaCl; 1 mM EGTA; 300 mM Sucrose; 3 mM MgCl2; 1X Protease Inhibitor Cocktail (Roche, 04693116001); 1 mM PMSF (Sigma-Aldrich, 93482) supplemented with 1 mM DTT and 0,5% Triton X-100. After 10 minutes on a wheel at 4°C, samples were centrifugated for 3 minutes at 900g at 4°C and cytoplasmic and nucleoplasmic components were collected as S1 fraction. Pellets were washed for 10 minutes on the wheel at 4°C with an additional volume of the same CSK buffer (supplemented with 1 mM DTT and 0,5% Triton X-100). Chromatin was then digested by using 25 U DNase I (Invitrogen, AM2222) in CSK buffer for 60 minutes at 37°C. To stop digestion, ammonium sulfate was added to samples to a final concentration of 250 mM and, after 5 minutes on ice, samples were pelleted at 900g for 3 minutes at 4°C and the supernatant was collected as S2 fraction. After a wash in CSK buffer, the pellet was further extracted with 2M NaCl in CSK buffer for 10 minutes at 4°C, centrifuged at 2300 g for 3 minutes at 4°C and the supernatant was conserved as S3 fraction. Pellets were washed twice for 10 minutes on the wheel at 4°C with a double volume of 2M NaCl CSK buffer. Finally, after 3 minutes of 3000g centrifugation at 4°C, pellets were solubilized in 8M urea for 10 minutes at RT to denature any remaining protein and dissolve membranes and labeled as S4. Fractions were stored at – 80°C until DNA extraction.

DNA extraction, library preparation and sequencing

Fractions were diluted 1:2 in TE buffer (10mM Tris-HCl pH 8.0, 1 mM EDTA) and incubated with 61,5 U of RNAse cocktail (Ambion, AM2286) at 37° for 90 minutes, followed by 40µg of Proteinase K (Invitrogen, AM2548), at 55° for 150 minutes. Genomic DNA was then isolated using phenol/chloroform/isoamyl (Sigma-Aldrich, 77617) extraction followed by back extraction of phenol/chloroform/isoamyl with an additional volume of TE buffer. DNA was precipitated by adding 20ug glycogen (Ambion AM9510) in 0.3M sodium acetate

with 3 volumes of cold ethanol. Precipitating DNA were then incubated for 1 hour in dry ice or overnight at -20° and centrifuged 30 minutes at 23000g. After a wash in 80% ethanol, dry pellets were resuspended in 50 µl (S2) or 15 ul (S3 and S4) of nuclease-free water and incubated at 4°C overnight. On the next day, S2 was further purified using PCR DNA Purification Kit (Qiagen, 28106) and separated using AMPure XP paramagnetic beads (Beckman Coulter, A63880) with the ratio of 0,90/0,95 to obtain smaller fragments conserved as S2S (< 300 bp) and larger fragments labelled as S2L (> 300bp) fractions. Both were suspended in 15 ul of nuclease-free water. S2L, S3 and S4 fractions were sonicated in a Covaris M220 focused-ultrasonicator using screw cap microTUBEs (Covaris, 004078) to obtain a smear of DNA fragments peaking at 150-350 bp (water bath 20°C, peak power 30.0, duty factor 20.0, cycles/burst 50). 150 seconds for S2L and 175 seconds for S3 and S4, 150 seconds for S4. Fractions were quantified using a Qubit 4 fluorometer with Qubit dsDNA HS Assay Kit (Invitrogen, Q32854) and run on an Agilent 2100 Bioanalyzer using High Sensitivity DNA Kit (Agilent, 5067-4626). Libraries were created from each sample using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645L) and Unique Dual Index NEBNextMultiplex Oligos for Illumina (NEB, E6440S); libraries were then qualitatively and quantitatively checked on Bioanalyzer 2100 using High Sensitivity DNA Kit (Agilent, 5067- 4626). Libraries with distinct adapter indexes were then multiplexed and, after cluster generation on FlowCell, sequenced for 50 bases in paired-ends mode on an IlluminaNovaSeq 6000 instrument at the IEO Genomic Unit in Milan or in one-end mode at the Division of Pathology of Fondazione IRCCS Ca' Granda- Ospedale Maggiore, Policlinico of Milan. A sequencing depth of at least 25-35 million reads was obtained for each sample.

RNA extraction, library preparation and sequencing

10-20K of Muscle Stem cell from the FACS-sorting were stabilized in 200µl of 1-Thioglycerol/Homogenization Solution of the Maxwell® RSC miRNA Tissue Kit (Promega, AS1460) and stored frozen at - 80°C for later total RNA automated purification using Maxwell® RSC 48 Instrument (Promega, AS8500) according to

manufacturer's instructions. Total RNA was quantified by Qubit 4 fluorometer with Qubit RNA HS Assay Kit (Invitrogen, Q32852) and assessed by Agilent 2100 Bioanalyzer using Agilent RNA 6000 Pico Kit (Agilent, 5067-1513) to inspect RNA integration. For each sample, 1 ng of total RNA was used to construct strand-specific RNAseq library with SMARTer Stranded Total RNA-Seq Kit - Pico Input (Takara, 634487). The yield and quality of the libraries were evaluated on Agilent 2100 Bioanalyzer using High Sensitivity DNA Kit (Agilent, 5067-4626). RNAseq libraries were sequenced on the Illumina NextSeqTM 550 system at the sequencing facilities of Humanitas or Division of Pathology of Fondazione IRCCS Ca' Granda-Ospedale Maggiore, Policlinico of Milan to a minimum target of 40 million for 75-100 bases in paired-ends mode.

Chromatin Immunoprecipitation, library preparation and sequencing

To preserve the integrity of chromatin architecture, Skeletal muscle cell suspension was fixed as described in the above paragraph ("Immunofluorescence on Muscle Stem Cells") before FACS sorting. Isolated Muscle Stem cells derived from different mice of the same age were pulled together and stored at -80°C. For ChIP analysis, 1.5 million fixed cells (derived from 15-20 mice) were thawed on ice and resuspended in fresh SDS buffer (10 mM Tris-HCl pH 8.0, 2 mM EDTA, 0.25% SDS, 1X PMSF, 1X protease inhibitors) and sonication was performed using a Covaris M220 focused-ultrasonic with following settings: water bath 7 °C, peak power 75.0, duty factor 10.0, cycles/burst 250, duration: 720 seconds. A small aliquot of chromatin (5ul) was purified and used for a chromatin shearing check. For IP, equilibration buffer (10mM Tris pH 8.0, 233mM NaCl, 1.66% Triton X-100, 0.166% DOC, 1mM EDTA, 1X PMSF, 1X protease inhibitors) was added to the shared chromatin, followed by the incubation with 2 µg of the antibody of interest on a rotating wheel at 4°C overnight. A minimum of 1% of the chromatin was stored as an input sample. Primary antibodies: H3K9me3 (Abcam, ab8898), H3K36me3 (Abcam, ab9050)). The next day, protein G beads (Life Technology,1004D) were added, and the samples were incubated for an additional 2 hours on the rotating wheel

at 4°C. Samples containing the beads-antibody-protein/DNA complexes were washed twice with IP buffer, high-salt IP buffer (500 mM NaCl in IP buffer), RIPA-LiCl buffer (10 mM TrisHCl pH 8.0, 1mM EDTA, 250 mM LiCl, 0.5% DOC, 0.5% NP-40, 1x PMSF, 1x Protease inhibitors), and TrisHCl pH 8,0. Between each wash, the samples were incubated for 5 minutes at 4 °C on the wheel. Finally, protein-DNA complexes were eluted with the Elution buffer (10 mM TrisHCl pH 8.0, 5 mM EDTA, 300 mM NaCl, 0.4% SDS). The cross-links were reverted by incubating the samples (included the input previously stored) overnight at 65°C. DNA was isolated through standard phenol/chloroform extraction, followed by precipitation and resuspension in 10 mM Tris-HCl pH 8.0. Fractions were quantified using Qubit 4 fluorometer with Qubit dsDNA HS Assay Kit (Invitrogen, Q32854) and run on an Agilent 2100 Bioanalyzer using High Sensitivity DNA Kit (Agilent, 5067-4626). Libraries were created from each sample using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645L) and Unique Dual Index NEBNextMultiplex Oligos for Illumina (NEB, E6440S); libraries were then qualitatively and quantitatively checked on Bioanalyzer 2100. Libraries with distinct adapter indexes were then multiplexed and, after cluster generation on FlowCell, sequenced for 50 bases in paired-ends mode on an IlluminaNovaSeq 6000 instrument at the IEO Genomic Unit in Milan. A sequencing depth of at least 30 million reads was obtained for each sample.

Literature data processing

The ChIP-seq data have been downloaded from the following publicly datasets available: GSE123725 (H3K4me3, H3K27me3)[24], GSE103163 (H3K27me3, H3K4me3)[34], GSE47362 (H3K27me3, H3K36me3, H3K4me3)[35].

All the data were downloaded as raw data with fastq-dump and re-analyzed as described in following section .

DNA sequence analysis

The results of the sequencing have been demultiplexed with bcl2fastq (https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html)(v2.20.0.422). The quality of sequencing was evaluated with fastqc[36] (version 0.11.3) and trimmed using Trimmomatic[37] (v0.39) with the following parameters: 2:30:10 for seed_mismatch, palindrome_threshold and simple_threshold, respectively; 4:15 for sliding window. The minimum threshold of 36 bp has been applied for all reads. Trimmed reads were aligned using BWA[38] (v0.7.17-r1188) setting -n 2 -k 2 and using as reference genome the mm10 downloaded from refgenie. The alignment duplicates were marked and removed with Picard (v2.23.9) (https://github.com/broadinstitute/picard) MarkDuplicates. We further filtered all the reads with mapping quality lower than 1 with Samtools[39] (v 1.11). For each alignment a coverage analysis has been performed using Deeptools[40] (version 3.5.2) bamCoverage function. The reads extended up to 250 bp and RPKM normalization method has been used. The mm10 size was considered of 2652783500 bp (value suggested in the Deeptools manual https://deeptools.readthedocs.io/en/latest/content/feature/effectiveGenomeSize.html), from the analysis were excluded blacklisted regions obtained by ENCODE portal, (https://www.encodeproject.org/files/ENCFF547MET). The differential enrichment analysis between SAMMY fractions (comparison) was performed using spp package[41], an R (v4.3) library. The reads were additionally filtered with 'remove.local.tag.anomalies' and the differential enrichment were computed using 'get.smoothed.enrichment.mle' function setting 'tag.shift = 0' and 'background.density.scaling = TRUE'. For maximal comparability the analysis has been performed with the same parameters for all the datasets.

Track representation

The Gviz[42] R library was used for visualization purposes. Each profile was computed using the function "DataTrack"; the samples were imported using the function "import" of the rtracklayer library and plotted using the function "plotTracks" setting the value "window = 900". Visualization of the Consensus

tracks were using all samples in each category setting the parameter type as ‘a’ and overlayed using the function “OverlayTrack”. Single samples mountain plots were computed by setting the parameter type as "polygon". Extra elements as chromosome ideogram and genome axis, were displayed using the functions “IdeogramTrack” and “GenomeAxisTrack” respectively.

ChiP-seq Peak calling and annotation

Before using the ChIP-seq each sample, downloaded and created was previously analyzed for quality evaluation with ChIC tool[43]. Only samples which passed the quality test were used for peak calling and analysis. Peak calling for H3K4me3 datasets was performed using macs2[44] (v2.2.7.1) with options -q 0.05 to filter peaks with q-values higher than 0.05. For broad ChIP-seq data (H3K9me3, H3K36me3, H3K27me3), peak calling was performed using epic2[45] (0.0.52), setting the FDR ≤ 0.05. Unwanted chromosomes were removed with the option -d '(GLlJHlM)'. The following settings were used for each histone mark analysis: H3K36me3 peaks --bin-size and --gaps-allowed were left to auto-mode; H3K27me3, the --bin-size 200 and --gaps-allowed 3; H3K9me3 --bin-size 900 --gaps-allowed 30. ChIP-seq peaks are annotated on the GENCODE vM25 basic gene annotation. For H3K4me3 and H3K27me3 histone marks, peaks were associated with the genes on ±1Kb around TSS. For H3K36me3, we selected genes based on the presence of peaks inside gene bodies. Finally, for H3K9me3, genes are selected if the TSS are inside a peak. Peak biological replicas, where present, were merged to ensure reliability.

SAMMY-seq normalization and consensus track generation

The comparison signal before the analyses, is computed by rebinning at the specified windows of 150 Kb. The blacklisted regions were firstly merged together with the neighbors if the distance were less than 50Kb. After, all together the comparisons in the dataset are normalized by computing the quantile normalization with the preprocessCore[46] library by the function

"normalize.quantiles". Finally, the consensus track is generated by computing the mean of the signals and the standard error as intervals.

Rebinning and Normalization of the comparison tracks

The comparison signal of all samples in a dataset is rebinned at the specified window. The blacklisted regions were merged together with the neighbor's regions if the distance were less than a specified distance. After, to avoid biased results due only to fractionation efficiency, all comparisons in each dataset are normalized together by computing the quantile normalization with the preprocessCore library by the function "normalize.quantiles".

Correlation Analysis

The smoothed differential signal enrichment is rebinned with Deeptools "multiBigwigSummary" at 50Kb. Genome-wide Spearman correlations between SAMMY-seq samples and ChIP was computed using Deeptools using the function "plotCorrelation" with the following settings: "--corMethod spearman -p heatmap --skipZeros".

Differential enrichment analysis of SAMMY-seq comparisons

To compute the differential enrichment analysis of the comparisons we firstly rebin and normalize the tracks as described above with 150 Kb rebinning size and 50Kb as merging distance of the blacklisted regions. We compute the consensus tracks for the control condition and calculate the intervals the standard deviation *2 (SD*2), obtaining consensus ± SD*2 for each genomic bin. Given the signal distribution of the consensus bins we filtered out all the bins within a range value ≤ ±0.1. After the filtering, the remaining bins are assigned to the correspondent fraction, S2S if the control condition signal is positive and S3 if is negative. To find the differential enriched bins in the test condition we sub-selected the bins in which all samples were not in the consensus ± SD*2 intervals. The selected bins were then put in a category defined as "up" if all the test

samples increased the signals, and “down” if they decreased their signals. Therefore, we obtained 4 distinct categories S2S_up, S2S_down, S3_up, S3_down.

SAMMY compartments

For the SAMMY-seq protocols, we loaded the bin enrichment tracks (RPKM) of each fraction and computed the correlation between pairs of bins. Here, the correlation was computed between *m*-dimensional vectors of enrichment signals, where m was the number of fractions in the specific protocol (*i.e.*, m=3).

For each input correlation matrix, the first eigenvector was obtained through principal component analysis decomposition in R statistical environment (prcomp, stats package, center=FALSE and scale=TRUE, rotation component of the returned object). The sign of the first eigenvector was defined using gene density: the group of bins with the highest gene density was marked as "A" compartment (positive sign), and the group with the lowest gene density was marked as "B" (negative sign). Chromosome eigenvector values were divided by the absolute maximum value for visualization purposes. All the analyses were made using R. We partially reimplemented CALDER algorithm (version 1.0, 2020-09-01) to accommodate the unusual format of SAMMY-seq data, maintaining the primary set of core functions (remove_blank_cols, fast_cor, generate_compartments_bed, HighResolution2Low_k_rectangle, get_PCs, bisecting_kmeans, project_to_major_axis, get_best_reorder, get_cluser_levels), their default parameters and the steps intended in the original paper[47] (8).

The pairwise distance matrix binned at 250 kb. In particular, the pairwise distance matrix was computed using for each pair of bins the Euclidean distance (dist, stats package, method="euclidean") of the m-dimension vectors, where m was the number of fractions in the specific protocol (*i.e.*, 3). Bins with null contacts or signals were removed from both matrices.

Briefly, for each input matrix, the algorithm: computed the pairwise correlation matrix and identified the sub-domain boundaries; computed the binary trend matrix and its decomposition using ten principal components; iteratively clustered

the domains to obtain their hierarchy; sorted the domains hierarchy based on the projection of the first two components; divided them into 2 groups (form the most closed, *i.e.*, B to the most opened, i.e., A). For more details on the original CALDER procedure and the interpretation of sub-compartments please see[47].

Gene Ontology (GO) permutation test

Available gene ontologies methods for transcriptomic analyses assume that each gene has the same probability of being represented. In genome analysis this is not the case, because we remove genes that by default that will never be encountered due to sequence characteristics (genes in blacklisted regions). Another problem to consider is: there are gene clusters with similar functions in the same genomic regions. These genes are often in same pathways. Furthermore, in SAMMY-seq, broad peaks include multiple TSS. All of these issues would systematically inflate the obtained P-values of the resulting GOs. We develop a specific analysis to address these aforementioned problems. The analysis consists in a permutation test, that relies on the assumption of exchangeability, a plausible assumption with genomic regions. The steps were: i) selection of all the genes that show an overlap of the TSS, in a range of 2500bp upstream and 500 bp downstream, with the tested genomic regions; ii) selection of all ontologies that incorporate the overlapping genes, keeping the count of the number of genes found in each ontology (X); iii) random shuffling (permutation) of all the accounted genomic regions along the genome keeping the size intact and avoiding overlapping with the blacklisted regions by masking them. Selection of genes (as explained in point i) after random repositioning and counting of number of genes associated to the GO described in point ii (x1). This procedure was repeated 10.000 times (B) for each gene ontology selected at point ii ($x1 \ldots xB$). iv) Counting for each GOs number of times when by permutation we found the number of genes associated with a GO to be greater than the selected regions. v) Dividing by the number of permutation the P-values associated to each GOs. This can be summarized by the formula:

$$P = \frac{\sum_{j=1}^{B} I(x_j \geq X)}{B}$$

Where *I* is the function which takes the value zero when it is false, and one when the is true[48]. vi) We then correct all GOs P-value with Benjamini Hochberg method. This analysis has been performed in R (v4.3) using the GO database R library "org.Mm.eg.db" with the database updated at 2023-Mar05.

RNA-seq sequencing read analysis

The sequenced reads were analyzed with the pipeline nf-core/rnaseq version[49] (v3.8). Quality reports of raw reads and preprocessed reads are generated by FastQC[36] (v0.11.9). Quality trimming and adapter clipping are performed by cutadapt[50] (v3.4) and trimmed reads were cleaned of ribosomal RNA (rRNA) sequences with SortMeRNA[51] (v4.3.4) considering all the available databases. Reads were then mapped with STAR on mouse genome build mm10. Gene and transcript were quantified using Salmon[52] on GENCODE (M25)[53] Basic gene annotation filtered for only protein coding genes. Differential expression analyses were performed with DESeq2[54].

Gene Set Enrichment Analysis (GSEA)

The gene list has been ranked according to Pvalue and log2 fold change (-log10(Pvalue) * sign(log2fc)). To compute the GSEA[55,56] (v4.1.0) plot we used gmt files available at g:profiler2[27] site and computed with parameters: Number permutations: 1000; Collapse: No; seed 42; Enrichment Statistic: classic; Max size: 500; Min size:15. This has been done equally for each datataset.

5.6 References

1. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. Hallmarks of aging: An expanding universe. *Cell* **186**, 243–278 (2023).
2. Cesari, M., Landi, F., Vellas, B., Bernabei, R. & Marzetti, E. Sarcopenia and Physical Frailty: Two Sides of the Same Coin. *Front. Aging Neurosci.* **6**, (2014).
3. Woo, J. Sarcopenia. *Clinics in Geriatric Medicine* **33**, 305–314 (2017).
4. Fuchs, E. & Blau, H. M. Tissue Stem Cells: Architects of Their Niches. *Cell Stem Cell* **27**, 532–556 (2020).
5. Alessandri, G. *et al.* Isolation and culture of human muscle-derived stem cells able to differentiate into myogenic and neurogenic cell lineages. *The Lancet* **364**, 1872–1883 (2004).
6. Evano, B., Khalilian, S., Le Carrou, G., Almouzni, G. & Tajbakhsh, S. Dynamics of Asymmetric and Symmetric Divisions of Muscle Stem Cells In Vivo and on Artificial Niches. *Cell Reports* **30**, 3195-3206.e7 (2020).
7. Robinson, D. C. L. *et al.* Negative elongation factor regulates muscle progenitor expansion for efficient myofiber repair and stem cell pool repopulation. *Developmental Cell* **56**, 1014-1029.e7 (2021).
8. Scaramozza, A. *et al.* Lineage Tracing Reveals a Subset of Reserve Muscle Stem Cells Capable of Clonal Expansion under Stress. *Cell Stem Cell* **24**, 944-957.e5 (2019).
9. Sousa-Victor, P., Perdiguero, E. & Muñoz-Cánoves, P. Geroconversion of aged muscle stem cells under regenerative pressure. *Cell Cycle* **13**, 3183–3190 (2014).
10. Liu, L. *et al.* Impaired Notch Signaling Leads to a Decrease in p53 Activity and Mitotic Catastrophe in Aged Muscle Stem Cells. *Cell Stem Cell* **23**, 544-556.e4 (2018).
11. Bernet, J. D. *et al.* p38 MAPK signaling underlies a cell-autonomous loss of stem cell self-renewal in skeletal muscle of aged mice. *Nat Med* **20**, 265–271 (2014).
12. Price, F. D. *et al.* Inhibition of JAK-STAT signaling stimulates adult satellite cell function. *Nat Med* **20**, 1174–1181 (2014).

13. Girelli, G. *et al.* GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nat Biotechnol* **38**, 1184–1193 (2020).
14. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).
15. Zhu, X. *et al.* Inflammation, epigenetics, and metabolism converge to cell senescence and ageing: the regulation and intervention. *Sig Transduct Target Ther* **6**, 245 (2021).
16. Lee, J.-H., Kim, E. W., Croteau, D. L. & Bohr, V. A. Heterochromatin: an epigenetic point of view in aging. *Exp Mol Med* **52**, 1466–1474 (2020).
17. Shcherbina, A. *et al.* Dissecting Murine Muscle Stem Cell Aging through Regeneration Using Integrative Genomic Analysis. *Cell reports* **32**, 107964 (2020).
18. Larouche, J. A. *et al.* Murine muscle stem cell response to perturbations of the neuromuscular junction are attenuated with aging. *eLife* **10**, (2021).
19. Dong, A. *et al.* Global chromatin accessibility profiling analysis reveals a chronic activation state in aged muscle stem cells. *iScience* **25**, 104954 (2022).
20. Yang, B. A. *et al.* Three-dimensional chromatin re-organization during muscle stem cell aging. *Aging cell* **22**, e13789 (2023).
21. Schworer, S. *et al.* Epigenetic stress responses induce muscle stem-cell ageing by Hoxa9 developmental signals. *Nature* **540**, 428–432 (2016).
22. Liu, L. *et al.* Chromatin modifications as determinants of muscle stem cell quiescence and chronological aging. *Cell Rep* **4**, 189–204 (2013).
23. Cesarini, E. *et al.* Lamin A/C sustains PcG protein architecture, maintaining transcriptional repression at target genes. *Journal of Cell Biology* **211**, 533–551 (2015).
24. Bianchi, A. *et al.* Dysfunctional polycomb transcriptional repression contributes to lamin A/C–dependent muscular dystrophy. *Journal of Clinical Investigation* **130**, 2408–2421 (2020).

25. Gattazzo, F., Laurent, B., Relaix, F., Rouard, H. & Didier, N. Distinct Phases of Postnatal Skeletal Muscle Growth Govern the Progressive Establishment of Muscle Stem Cell Quiescence. *Stem Cell Reports* **15**, 597–611 (2020).

26. Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Research* **50**, D687–D692 (2022).

27. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research* **47**, W191–W198 (2019).

28. Boettiger, A. N. *et al.* Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* **529**, 418–422 (2016).

29. Sergi Sayols. rrvgo. doi:10.18129/B9.BIOC.RRVGO.

30. Sousa-Victor, P., García-Prat, L. & Muñoz-Cánoves, P. Control of satellite cell function in muscle regeneration and its disruption in ageing. *Nat Rev Mol Cell Biol* **23**, 204–226 (2022).

31. Kuga, T. *et al.* Lamin B2 prevents chromosome instability by ensuring proper mitotic chromosome segregation. *Oncogenesis* **3**, e94–e94 (2014).

32. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682 (2012).

33. Sebestyén, E. *et al.* SAMMY-seq reveals early alteration of heterochromatin and deregulation of bivalent genes in Hutchinson-Gilford Progeria Syndrome. *Nat Commun* **11**, 6274 (2020).

34. Machado, L. *et al.* In Situ Fixation Redefines Quiescence and Early Activation of Skeletal Muscle Stem Cells. *Cell Rep* **21**, 1982–1993 (2017).

35. Liu, L. *et al.* Chromatin modifications as determinants of muscle stem cell quiescence and chronological aging. *Cell Rep* **4**, 189–204 (2013).

36. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. (2010).

37. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

38. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at http://arxiv.org/abs/1303.3997 (2013).

39. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).

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

41. Kharchenko, P. V., Tolstorukov, M. Y. & Park, P. J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature Biotechnology* **26**, 1351–1359 (2008).

42. Hahne, F. & Ivanek, R. Visualizing Genomic Data Using Gviz and Bioconductor. in *Statistical Genomics* (eds. Mathé, E. & Davis, S.) vol. 1418 335–351 (Springer New York, 2016).

43. Livi, C. M. *et al. A ChIC solution for ChIP-seq quality assessment*. http://biorxiv.org/lookup/doi/10.1101/2020.05.19.103887 (2020) doi:10.1101/2020.05.19.103887.

44. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).

45. Stovner, E. B. & Sætrom, P. epic2 efficiently finds diffuse domains in ChIP-seq data. *Bioinformatics* **35**, 4392–4393 (2019).

46. Ben Bolstad <Bmb@Bmbolstad. Com>. preprocessCore. (2017) doi:10.18129/B9.BIOC.PREPROCESSCORE.

47. Liu, Y. *et al.* Systematic inference and comparison of multi-scale chromatin sub-compartments connects spatial organization to cell phenotypes. *Nat Commun* **12**, 2439 (2021).

48. Ruxton, G. D. & Neuhäuser, M. Improving the reporting of *P* -values generated by randomization methods. *Methods Ecol Evol* **4**, 1033–1036 (2013).

49. Harshil Patel *et al.* nf-core/rnaseq: nf-core/rnaseq v3.12.0 - Osmium Octopus. (2023) doi:10.5281/ZENODO.1400710.

50. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* **17**, 10 (2011).

51. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).

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

53. Frankish, A. *et al.* GENCODE 2021. *Nucleic Acids Research* **49**, D916–D923 (2021).

54. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

55. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267–273 (2003).

56. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).

Chapter 6 Conclusions

Chromatin organization, which involves multiple levels of folding, is fundamental for the homeostatic regulation of the cells[1]. Perturbations in chromatin folding influence the transcriptional regulation of the genome. Variations of chromatin structure can be physiological, for example, those accompanying lineage specification or stress response. On the other hand, diseases are often associated to a pathological remodeling of chromatin that can be tissue specific or systemic. One of the most studied phenomenon of pathological chromatin remodeling is the cellular senescence[2]. The induction of cellular senescence can be triggered by many factors, such as DNA damage accumulation, critically short telomeres, mitochondrial defects, ROS and nutrient imbalance[3,4].

The induction of cellular senescence lead to chromatin remodeling, which consequently lead to a metabolic cellular reprogramming[5]. However, other factors could play a role triggering cellular senescence and/or chromatin remodeling. The lamin A for example, a fundamental nuclear protein which lies beneath the nuclear membrane, have a role in DNA damage response, telomeres correct capping, and correct distribution of the H3K27me3 heterochromatin mark on target genes, all genomic features of cellular senescence[3,4]. With our 3fSAMMY-seq technology we already demonstrated that we can observe early heterochromatin solubility changes in HGPS cells[6]. As the solubility is different from the classical notion of chromatin interaction introduced by other techniques, we can now use it as a new parameter to better understand chromatin function[6]. However, 3fSAMMY-seq has the limitation in capturing only heterochromatin. In recent years we further improve the method, using a milder DNAse digestion and fragment separation by size, and we successfully isolate also euchromatin fragments (**Figure 1**, Chapter 2). With the new improved method, the 4fSAMMY-seq, we can describe both euchromatin and heterochromatin solubility starting from 10K cells. Moreover, with a dedicated bioinformatic algorithm we can successfully reconstruct chromatin 3D organization in compartments and sub-compartments (**Figure 2**, Chapter 2). These advancements in the technique were

crucial during my PhD thesis to study chromatin dynamics during early senescence in various cellular models. While the vast majority of existing quantitative bioinformatics tools were developed for RNA-seq analyses, few tools are available for DNA analysis, and none of them specifically suited for SAMMY-seq which capture ~80% of the genome in large domains. Thus, during my thesis I modified and/or developed various tools and pipelines to study SAMMY-seq data. Working on multiple biological samples we created a consensus track to simultaneously evaluate at the genome-wide level the mean solubility and the variability due to the biological samples. This facilitated the identification of genomic regions mostly affected by chromatin remodeling. Then, to systematically compare different conditions, we set up a method to point out the genomic regions that change their solubility. To analyze the role of lamins A and B and their detrimental effect on chromatin, we used the immortalized mouse embryonic fibroblast (MEF) generated starting from mice knockout for *Lmna* gene and double knockout for *Lmnb1* and Lmnb2. We found that both lamin, A and B, are required for the heterochromatin maintenance (**Figure 3**, Chapter 3). Moreover, our data showed that lack of lamin A or lamin B has different effects on the genome solubility, with Lamin B affecting larger genomic regions. To investigate the regions differently enriched in mutants we used a modified version of the Gene Ontology (GO) analysis. Existing GO methods for transcriptome analysis assume that all genes are equally likely to be represented in the dataset. This is not true for the entire genome, for multiple reasons: genes could be included in a blacklisted region, excluded from the analysis; each broad SAMMY peak can include several genes that could confound the analysis; gene clusters with similar functions normally share the genomic localization, increasing the GO p-value. Hence, we set up a permutation analysis on the SAMMY-seq genomic regions selected to find the significant enriched GOs. From the GO analysis of affected chromatin regions, we found different roles in regulation of euchromatin with Lamin B involved in chromatin structure and Lamin A in cell specification (**Figure 5**, Chapter 3). It is already reported that the relative amounts of lamin A and B influences the differentiation in a specific lineage differentiation[7]. Mouse

embryonic fibroblasts (MEF), used in this project are still plastic cells, able to differentiate in different lineages[8]. Our data confirmed the idea that different laminA/laminB ratios can trigger distinct lineage specification. We extensively studied the link between chromatin solubility and transcription. Although GO analysis of differential solubility regions and of differential expressed genes (DEGs) overlap, showing several common pathways, we found that only few DEGs are localized in regions with different solubility (**Supplementary Figure 2**, Chapter 3). However, when we specifically analyzed the metaprofile of the mutant DEGs, we found a clear solubility change in the downregulated genes, congruent with the transcriptional switch (**Figure 8**, Chapter 3). From the technical point of view, this incongruity could be explained by differences in genomic resolution of the two experiments (**see methods**, Chapter 3). On the other hand, these data suggest that while the solubility create large windows of opportunity in several genomic regions, only in some cases this will determine the transcriptional switch. Using the same bioinformatic tools that I developed in the project described in Chapter 3, we also analyze different tissues of progeric mice at distinct stage of disease: early (1 month-old mice) and advanced (3 months-old mice). We selected the three mainly affected tissues in progeria: aorta, skin, and striated muscles. From these tissues, we extracted the cells derived from mesenchymal lineage, normally more affected by progerin: endothelial cells (CD31+) from aorta, dorsal skin fibroblasts (F140+), and muscle satellite cells (a7+)[9]. In progeric cells we found a progressive increase of chromatin remodeling during age in endothelial CD31+ cells and in skin fibroblasts (**Figure 3**, Chapter 4). This trend was not observed in muscle stem cells (MuSC) which show an opposite trend. However, other analysis based on chromatin compartmentalization or by the consensus dispersion showed also in MuSC the same trend (**Figure 2 and 6**, Chapter 4). This discrepancy can be justified in the nature of MuSC, quiescent, not cycling cells, that are activated upon muscle damage[10]. In one month-old mice the MuSC enter the quiescent forming the stem cell niche, while in adult 3 months-old mice MuSC are stably quiescent. The effect that we see only in the solubility profile can be due to a higher sensibility of chromatin to progerin in that

stage. Further analysis on transcriptome derived from the same cells will elucidate if those solubility changes are accompanied by transcriptional deregulation. Analysis of GO terms associated to solubility changes suggest that remodeled regions are cell type specific (**Figure 5**, Chapter 4). However, we also found some biological pathways altered in all cellular types such as cell adhesion and localization and cell cycle, suggesting that these pathways could be directly dependent on progerin.

In my last project, we analyzed the chromatin evolution during muscle physiological ageing to understand the development of sarcopenia. Sarcopenia is characterized by muscle fat infiltration and reduced skeletal muscle performance[11]. Among the reasons that cause sarcopenia, muscle stem cell exhaustion is one of the most important ones[12]. To achieve the aim of the project, we used muscle satellite cells with four most representative life stages: postnatal, adult, old and geriatric. The close relationship between aging and epigenetic mechanisms has been highlighted in adult stem cells[13,14]. In muscle, aging-associated alterations of stem cell niche affect extrinsic signals that sustain MuSC activation and differentiation[15]. In parallel, MuSC-intrinsic molecular mechanisms might provoke normal muscle stem cell function failure[16]. Extrinsic and intrinsic causes determine a change in the MuSC niche during aging, with a decline of genuine, stemness-like quiescent satellite cells[17]. Molecularly, it has been reported a general increase in chromatin accessibility in geriatric MuSC[18–21]. We first compared postnatal stage, characterized by an intense MuSC activity required for muscle growth and adult in which mostly MuSC are quiescent. We found few solubility changes, compatible with differentiating stage of these cells (**Figure 7**, Chapter 5). On the other hand, analysis of adult, old and geriatric stage highlighted a progressive change in the chromatin solubility, especially affecting the heterochromatin (**Figure 7**, Chapter 5).

The GO terms analysis associated to solubility changes shown that aged MuSCs a decreased solubility in the cell receptors which can obstruct the activation and differentiation processes[22] (**Figure 8**, Chapter 5). In geriatric stage show a global picture which resemble the loss of cell identity already described in our previous

work on MuSC premature ageing caused by lack of lamin A [23] (**Figure 8**, Chapter 5).

The transcriptional results show us an increased level of chromatin remodelers' transcripts in both postnatal and geriatric. In the first stage, this confirm a chromatin remodeling in differentiation. In the latter, we found a trend of upregulation which suggest a dysregulation of the homeostatic quiescence processes (**Figure 8**, Chapter 5). Overall, our results propose chromatin solubility as a new epigenetic parameter that can describe chromatin function. Moreover, here we presented several insights that chromatin remodeling may facilitate the triggering of senescence in a cell specific manner.

Chapter 6 references

1. Misteli, T. The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* **183**, 28–45 (2020).
2. Di Micco, R., Krizhanovsky, V., Baker, D. & d'Adda di Fagagna, F. Cellular senescence in ageing: from mechanisms to therapeutic opportunities. *Nat Rev Mol Cell Biol* **22**, 75–95 (2021).
3. Gorgoulis, V. *et al.* Cellular Senescence: Defining a Path Forward. *Cell* **179**, 813–827 (2019).
4. Zhu, X. *et al.* Inflammation, epigenetics, and metabolism converge to cell senescence and ageing: the regulation and intervention. *Sig Transduct Target Ther* **6**, 245 (2021).
5. Huang, W., Hickson, L. J., Eirin, A., Kirkland, J. L. & Lerman, L. O. Cellular senescence: the good, the bad and the unknown. *Nat Rev Nephrol* **18**, 611–627 (2022).
6. Sebestyén, E. *et al.* SAMMY-seq reveals early alteration of heterochromatin and deregulation of bivalent genes in Hutchinson-Gilford Progeria Syndrome. *Nat Commun* **11**, 6274 (2020).
7. Swift, J. *et al.* Nuclear Lamin-A Scales with Tissue Stiffness and Enhances Matrix-Directed Differentiation. *Science* **341**, 1240104 (2013).

8. Dastagir, K. *et al.* Murine embryonic fibroblast cell lines differentiate into three mesenchymal lineages to different extents: new models to investigate differentiation processes. *Cell Reprogram* **16**, 241–252 (2014).
9. Zhang, J. *et al.* A Human iPSC Model of Hutchinson Gilford Progeria Reveals Vascular Smooth Muscle and Mesenchymal Stem Cell Defects. *Cell Stem Cell* **8**, 31–45 (2011).
10. Schmidt, M., Schüler, S. C., Hüttner, S. S., von Eyss, B. & von Maltzahn, J. Adult stem cells at work: regenerating skeletal muscle. *Cell Mol Life Sci* **76**, 2559–2570 (2019).
11. Cesari, M., Landi, F., Vellas, B., Bernabei, R. & Marzetti, E. Sarcopenia and Physical Frailty: Two Sides of the Same Coin. *Front. Aging Neurosci.* **6**, (2014).
12. Woo, J. Sarcopenia. *Clinics in Geriatric Medicine* **33**, 305–314 (2017).
13. Shaban, H. A. & Gasser, S. M. Dynamic 3D genome reorganization during senescence: defining cell states through chromatin. *Cell death and differentiation* (2023) doi:10.1038/s41418-023-01197-y.
14. Liu, Z., Belmonte, J. C. I., Zhang, W., Qu, J. & Liu, G.-H. Deciphering aging at three-dimensional genomic resolution. *Cell insight* **1**, 100034 (2022).
15. Sousa-Victor, P., García-Prat, L. & Muñoz-Cánoves, P. Control of satellite cell function in muscle regeneration and its disruption in ageing. *Nature reviews. Molecular cell biology* **23**, 204–226 (2022).
16. Sousa-Victor, P., Garcia-Prat, L., Serrano, A. L., Perdiguero, E. & Munoz-Canoves, P. Muscle stem cell aging: regulation and rejuvenation. *Trends Endocrinol Metab* **26**, 287–296 (2015).
17. García-Prat, L. *et al.* FoxO maintains a genuine muscle stem-cell quiescent state until geriatric age. *Nature cell biology* **22**, 1307–1318 (2020).
18. Shcherbina, A. *et al.* Dissecting Murine Muscle Stem Cell Aging through Regeneration Using Integrative Genomic Analysis. *Cell reports* **32**, 107964 (2020).
19. Larouche, J. A. *et al.* Murine muscle stem cell response to perturbations of the neuromuscular junction are attenuated with aging. *eLife* **10**, (2021).

20. Dong, A. *et al.* Global chromatin accessibility profiling analysis reveals a chronic activation state in aged muscle stem cells. *iScience* **25**, 104954 (2022).
21. Yang, B. A. *et al.* Three-dimensional chromatin re-organization during muscle stem cell aging. *Aging cell* **22**, e13789 (2023).
22. Sousa-Victor, P., García-Prat, L. & Muñoz-Cánoves, P. Control of satellite cell function in muscle regeneration and its disruption in ageing. *Nat Rev Mol Cell Biol* **23**, 204–226 (2022).
23. Bianchi, A. *et al.* Dysfunctional polycomb transcriptional repression contributes to lamin A/C–dependent muscular dystrophy. *Journal of Clinical Investigation* **130**, 2408–2421 (2020).