SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

School of Medicine and Surgery

PhD program in Translational and Molecular Medicine (DIMET)
XXXII Cycle

Unconventional nuclear architecture in CD4$^+$ T lymphocytes uncouples chromatin solubility from function

Lucini Federica

Mat. 823290

Tutor: Dr. Chiara Lanzuolo

Coordinator: Prof. Andrea Biondi

ACADEMIC YEAR 2018/2019

CHAPTER 1.
GENERAL INTRODUCTION

The tens of trillions of cells that make up our body have in their nuclei the same genome. Each of them, however, can conduct specific functions by using only a small part of the entire genetic information[1,2]. The selection of the sequences that need to be transcribed and translated is regulated through the ensemble of mechanisms composing the epigenome[3].

Epigenetic mechanisms modulate the organized compaction of the linear genomic sequence into the small nuclear volume and guarantee the correct transcriptional state to every genomic region[4]. They prevent DNA damage[5], reduce errors during replication and cell division[6] and safeguard the transcriptional state throughout the cell cycle.

1.1 The epigenome

As a first level of genome compaction, 145-147 base pairs (bp) of DNA wrap around an octamer of four core histone proteins (H2A, H2B, H3 and H4, each in a double copy), generating the nucleosome. This nucleoprotein complex, occurring essentially every 200 bp in all eukaryotic genomes[7], is then bound by linker histones. They stabilize its structure and mediate its interaction with neighbouring nucleosomes, allowing the formation of chromatosomes and condensed chromatin fibres[8]. The extreme degree of compaction (50-fold and higher) and the constant dynamics of the fibres have hampered so far the resolution of their precise internal organization[9]. The interaction of the fibres between them and with several nuclear RNAs and proteins results in the formation of loops and other higher-order chromatin structures key for the modulation of transcription (as reviewed in [10–13]). The heterogeneous chromatin distribution inside the interphase nucleus is clearly revealed by electron microscopy (EM) imaging, which highlights areas of higher and lower

density, called heterochromatin and euchromatin, respectively[14] (Fig. 1). At each level of genome compaction, a variety of regulating factors contribute to the final transcriptional output of a single genomic region, making it tightly controlled.

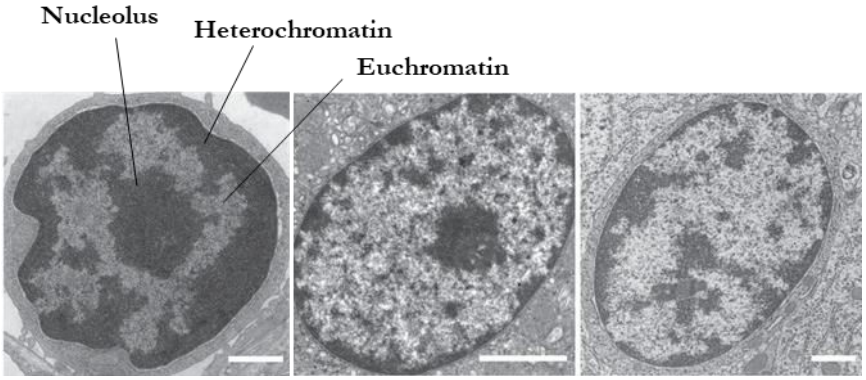


Figure 1 – Electron microscopy visualization of interphase nuclei. Electron micrographs of mammalian nuclei: lymphocyte (left), epithelial cell (centre), glial cell (right). Heterochromatin and nucleolus are darkly stained, while euchromatin is lightly stained. Scale bars: 1 μm. Adapted from[15]

Covalent modifications and histone code

At the primary level of DNA sequence, the only covalent modification with a known epigenetic effect in mammals is the methylation of C5 position of cytosine in CpG sites. Operated by DNA methyl transferases (DNMTs), this modification is associated with long-term repression of transcription, being for example implicated in X-chromosome inactivation, genomic imprinting and silencing of repetitive DNA elements[16,17]. On the other hand, countless described mechanisms control chromatin compaction and transcription through the regulation of histone covalent modification, nucleosome composition and its positioning. The positioning of nucleosomes, primary determinant of DNA accessibility,

results from a combination of DNA sequence preference, thermal motion, competitive protein binding and ATP-dependent remodelling[18]. Nucleosome stability is further influenced by histone octamer composition and post-translational modifications (PTMs) of the basic residues (mainly lysine and arginine) of the intrinsically disordered N-terminal protruding histone tails[18] (Fig. 2).

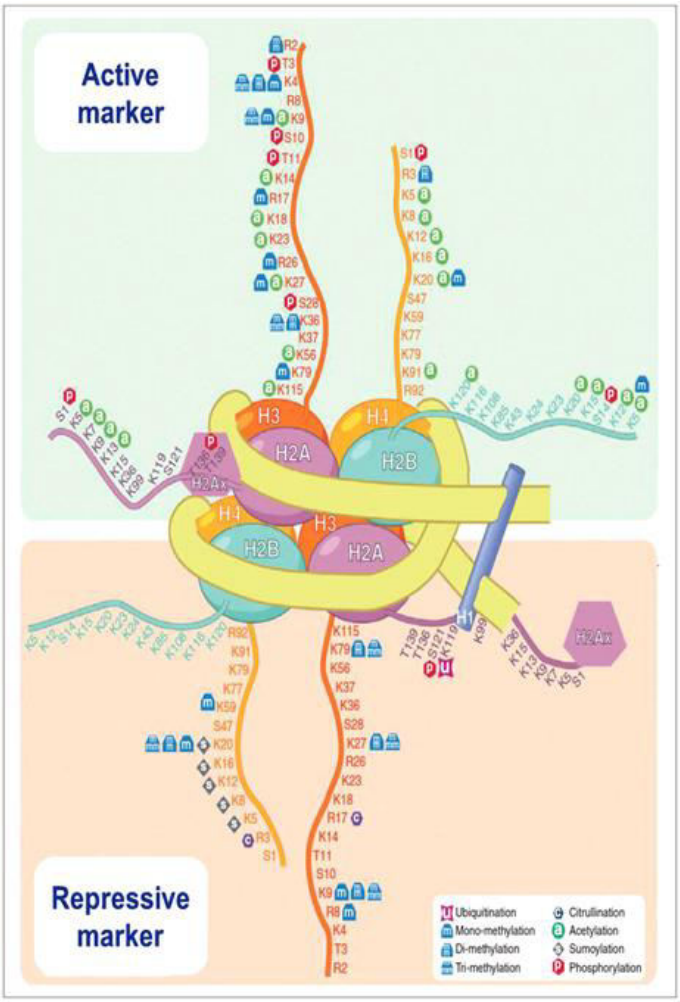


Figure 2 – Representation of nucleosome and histone code. The octamer of core histone proteins (H2A, H2B, H3 and H4) is wrapped by 145bp of DNA (yellow) stabilized by the linker histone H1 (blue). On the

protruding tail of each histone protein are indicated the basic residues - lysine (K), arginine (R), serine (S), and threonine (T) - together with their described post-translational modifications (PTMs). Ubiquitination (pink), mono-, di- and trimethylation (blue), citrullination (purple), acetylation (green), sumoylation (gray), phosphorylation (red). Activating PTMs are shown in the upper part of the picture, while repressive PTMs are shown in the lower part. Adapted from[19]

The human genome has more than 20 genes encoding for histone variants[20,21]. The replacement of canonical octamer histones with variant forms can produce subtle differences in the structure and interaction surfaces of the nucleosome, endowing specific genomic regions with peculiar functions. For instance, H2A.Z - variant of H2A - is predominantly localized around the transcription start site (TSS) of actively transcribed genes[22], while CENP-A - variant of H3 - is incorporated into the centromeric chromatin and seem to play a pivotal role in centromere propagation and in the formation of the assembly scaffold for centromere/kinetochore proteins[21]. Another H3 variant, H3.3, integrated at specific genomic loci in a replication-independent manner, assists the rapid activation of genes that were silenced by H3 modifications[23]. Acetylation, methylation, phosphorylation, ubiquitination, sumoylation and ADP-ribosylation (reviewed in [24]) are the most common among PTMs. They are generated by histone modifying enzymes and can produce structural alterations inside the nucleosome, affecting inter-nucleosomal contacts and finally resulting in the modification of chromatin fibre structure. Furthermore, they serve as docking sites for various chromatin modifiers and multiple PTMs coexisting on the same histone tail are able to dictate specific biological readouts[25].

Overall, the influence of nucleosomes on genome conformation and gene expression is widely recognized[26,27]. Studies in yeast, later complemented

in mammals, evidenced that active genes have reduced nucleosome occupancy[28,29] and that their functional regions are characterized by specific histone modifications: promoters are enriched in histone acetylation[30–33], TSSs in H3K4 methylation, while the 3' end in H3K36me3[34]. On the other hand, elevated levels of H3K27 and H3K9 methylation are correlated with gene repression[34,35]. At present, the connection between specific histone modifications and genomic function is so established that PTMs are often used to identify genomic elements[36]: H3 acetylation and H3K4me1, for example, are indicators of enhancers[37,38]. Moreover, changes in the histone modification pattern of genomic regions during biological processes are used to locate alterations in genome function. In fact, the combinatorial genome-wide analysis of few histone marks has been demonstrated to allow a good prediction of the transcriptional activity of genomic regions[39–41]. In embryonic stem cells, for example, specific sites featuring both the activating H3K4me3 and the repressive H3K27me3 were found in correspondence with developmental and differentiation genes poised for activation[42] and they were named "bivalent domains".

RNA interactions

The large repertoire of RNAs produced by each cell also holds responsibility on chromatin organization and gene expression[43]. Noncoding RNAs (ncRNAs), for instance, are involved in the establishment of X-chromosome inactivation in female mammals, guaranteeing gene dosage compensation between sexes[44]. The ncRNA Xist, retained in the nucleus, initiates a chain of chromatin-remodelling events on X-chromosome, encompassing hypoacetylation of H3 and H4,

methylation of H3K9, H3K27 and H4K20, incorporation of the H2A variant macroH2A. Other ncRNAs, on the contrary, contribute to gene transcription by targeting activating proteins to a specific locus. This is the case of HOTTIP, long noncoding RNA (lncRNA) which targets WDR5/MLL complexes on HOXA locus, driving trimethylation of H3K4 and gene transcription[45]. Thanks to the technological innovations allowing their precise detection, a lot more functions in the regulation of genome architecture and transcription are currently being ascribed to chromatin-associated RNAs and their binding proteins (RBPs)[11].

Three-dimensional chromatin organization

The interaction and folding of nucleosomes into complex higher-order structures allows long-distance contacts between genomic sequences, enabling the segregation of specific nuclear functions into different nuclear regions. A first macro-discrimination between repressed heterochromatin regions and active chromatin hubs (ACHs) is already evident by electron microscopy observation of interphase nuclei[14]. Furthermore, the nucleus possesses a plethora of smaller foci where specific RNAs and proteins associate to the genome to implement dedicated activities[46–48]. Heterochromatin, for example, is present in large domains positioned along the nuclear lamina (LADs)[49] or around the nucleoli (NADs)[50], but also in small and more dynamic foci in the nucleoplasm, associated with Polycomb repressive complexes (PRCs)[51]. Euchromatin encompasses even more subtypes of condensates deputed to specific functions. Among them, nucleoli, histone locus bodies (HLBs) and transcription factories. Other nuclear structures, such as the nuclear pore complexes (NPCs), characterized by the interaction with both

euchromatin and heterochromatin, contribute to a highly regulated transcription through euchromatin looping[52]. In all of the mentioned nuclear regions, multiple genes gather together through looped structures and interact with proteins responsible for a specific genomic function[12,53,54]. These numerous and highly dynamic structures are another fundamental level of control on genome function[48,55] (Fig. 3).

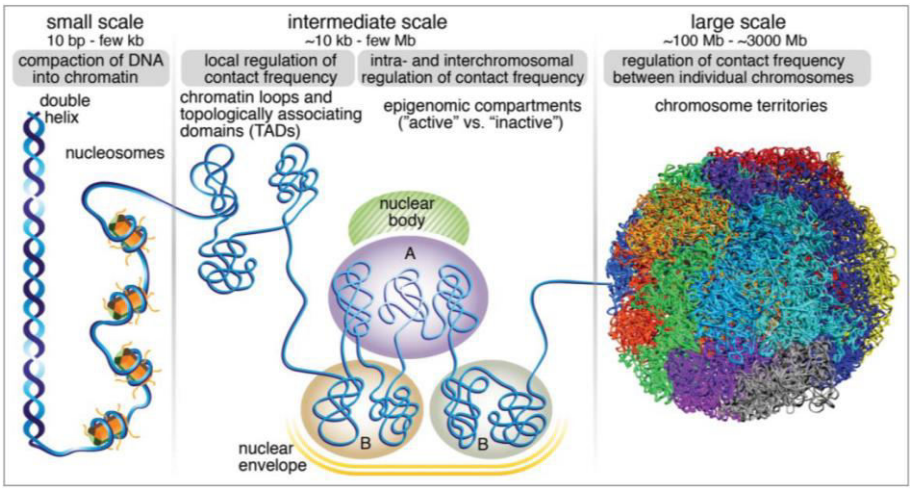


Figure 3 – Three-dimensional chromatin organization. After the first DNA-proteins association in the nucleosome (left), further medium- and long-range interactions (centre-left) guide the formation of loops and Topologically Associating Domains (TADs), critically controlling enhancer-promoter interactions. Interacting with nuclear structures, such as nuclear bodies and nuclear envelope, TADs organize into transcriptionally active (A) and inactive (B) compartments (centre-right). At the largest scale (right), chromosomes tend to associate between them and with the nuclear structure in a conserved way, allowing the recognition of stereotyped chromosome territories inside the cell nucleus. Adapted from[56]

Nuclear architecture

The conserved geographic positioning of chromosomes in the nucleus has a further important role in the transcriptional control of gene

expression[57,58] (Fig. 4).

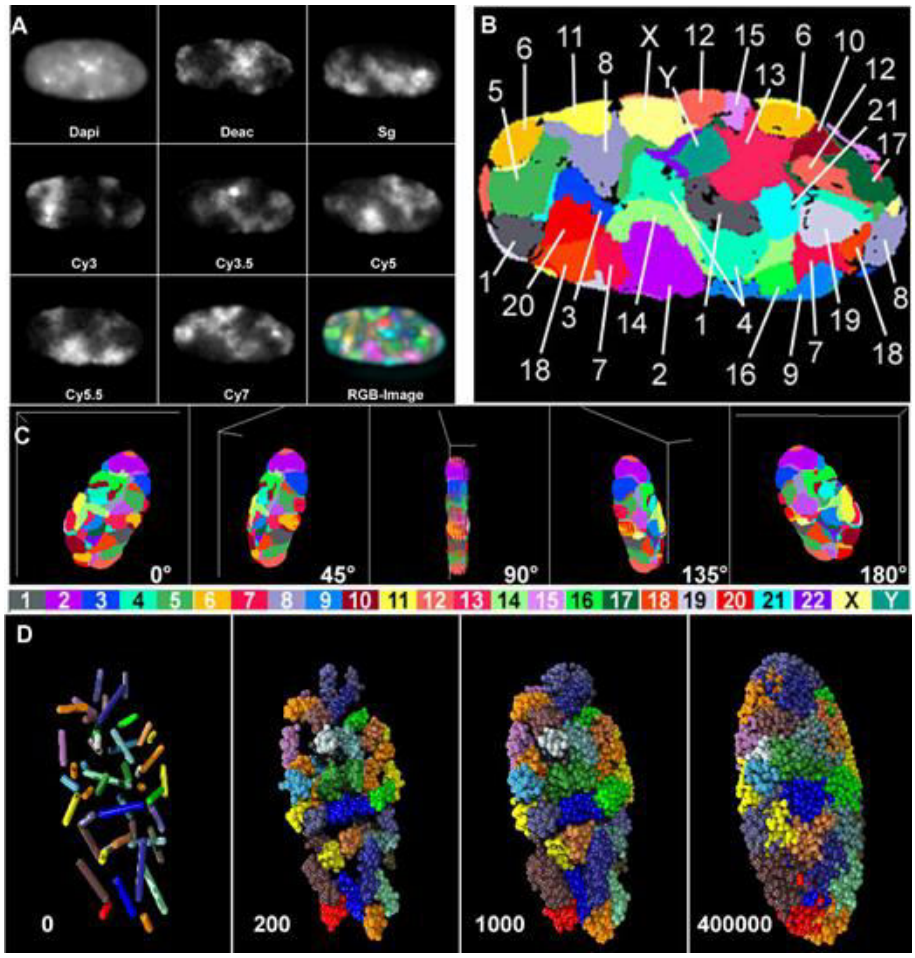


Figure 4 – Chromosome territories in a human fibroblast nucleus in interphase. 24-Color 3D FISH approach for the identification of chromosome territories. **A,** Deconvoluted mid-plane nuclear section images obtained in eight channels: one for DAPI (DNA counterstain) and seven for fluorochromes painting a chromosome territory subset. RGB image of the 24 differently labelled chromosome types (bottom-right) was produced by superposition of the seven channels. **B,** False colour representation of all CTs visible in the mid-section in (A). **C,** 3D reconstruction of the complete CT arrangement in the nucleus viewed from different angles. **D,** Simulation of a human fibroblast model nucleus. In the first image 46 statistically placed rods represent the 46 human chromatids (left). The next images simulate the decondensation process

and show the resulting chromosome territory arrangement. Adapted from[59]

Indeed, it has been observed that some chromosomes recurrently localize toward the nuclear periphery, while others preferentially locate near to the nuclear centroid. The interaction with the nuclear proteins present in these different microenvironments ensures a final level of transcriptional regulation[60]. The Nuclear Lamina (NL), dense fibrillary network localized under the inner nuclear membrane, for example, interacting with nearly 40% of the genome[61] governs the principal transcriptionally inactive nuclear environment. On the contrary, in the internal nucleoplasm various proteins engage with the genome favouring transcription. NPCs, as mentioned, at nuclear periphery interact with both repressed and actively transcribed regions[62,63] contributing to both chromatin organization[64] and gene expression regulation[65].

1.2 Epigenome and cell identity

The ensemble of all the epigenetic mechanisms described here is fundamental for the orchestration of DNA-based biological processes and for the maintenance of cell identity. Evidences indicating a connection between mutations in histone modifying enzymes and cancer development are in constant increase, to the point that drugs targeting these enzymes are already in clinical practice[66–68]. Alteration in genes coding for nuclear structural proteins are also well recognized to originate pathologies, with one clear example being laminopathies[69,70]. However, in both cases, the molecular mechanisms that lead from the mutation to the uncontrolled proliferation or loss of cell identity are still poorly understood. A deregulated gene expression, caused by alterations in

nucleosome structure and by displacement of genes from their nuclear environment, has been proposed as the main driving force for pathology development. Recent studies, however, evidenced that changes in nuclear localization are not always connected to gene expression alteration[71–73].

1.3 Epigenome dynamics

Given the fundamental importance of the epigenome for the maintenance of cell identity, it could be easy to think that it should be refractory to any modification. On the contrary, the epigenetic machinery needs to sense and to be flexible to external changes, to constantly reach the best fitting for the survival of the organism[74].

In physiological conditions, the more extensive changes in the epigenome are induced upon cell cycle and cell differentiation.

During cell cycle, genome replication and its segregation into two daughter cells imposes a complete reorganization of the epigenome. In these critical steps, a lot of mechanisms activate to preserve and transmit the correct epigenetic signature[75].

A great number of studies on the epigenome dynamics have been performed on differentiating cells, especially myocytes and neurons. In this process cells not only divide but also change their epigenome, silencing "stem-genes" and timely activating new genes, useful for the final cell phenotype[76]. For example, evidence reports that muscular stem cells (MuSCs) undertaking differentiation repress the stem transcription factor Pax7 and activate the early muscular-commitment factor MyoD, initiating a cascade of activation of muscle-specific genes[77].

1.4 Mechanotransduction

Several studies demonstrate that the connection between cyto- and nucleoskeleton is fundamental for the conveyance of external mechanical stimuli to the nucleus in the form of biochemical signals (reviewed in [78,79]). The dynamics of epigenome are therefore strictly related to the components of the mechanotransduction machinery. A-type Lamins in particular, being exactly at the interface between the nuclear membrane and chromatin, occupy a primary position[70]. They are in fact physical transducers of mechanical signals[80], conveyed on chromatin through distortion of higher-order structure and changes in transcription factor accessibility[81].

1.5 The immune system

In the everyday life, our bodies come into contact with multiple threats to our health. We touch, inhale and ingest potential external pathogens at any time[82]. Furthermore, we create our own threats: the accumulation of inherited and acquired genomic mutations stochastically originates transformed cells, with growth or survival advantage, whose uncontrolled proliferation leads to cancer. The immune system owns the delicate balance between the elimination of dangerous agents and the tolerance of self or innocuous non-self[83,84]. An alteration in this balance results either in immunodeficiency or in hypersensitivity reactions and autoimmune diseases. In a proper functioning immune system, a lot of different components collaborate to maintain the homeostasis of organism[85] (Fig.5). First of all, the entrance of the pathogens is obstructed by physical and chemical barriers, namely skin, microbial peptides, cilia and gastric

acid. For the pathogens that manage to overcome these barriers, there are two further lines of defence waiting inside: the innate and the adaptive immune system.

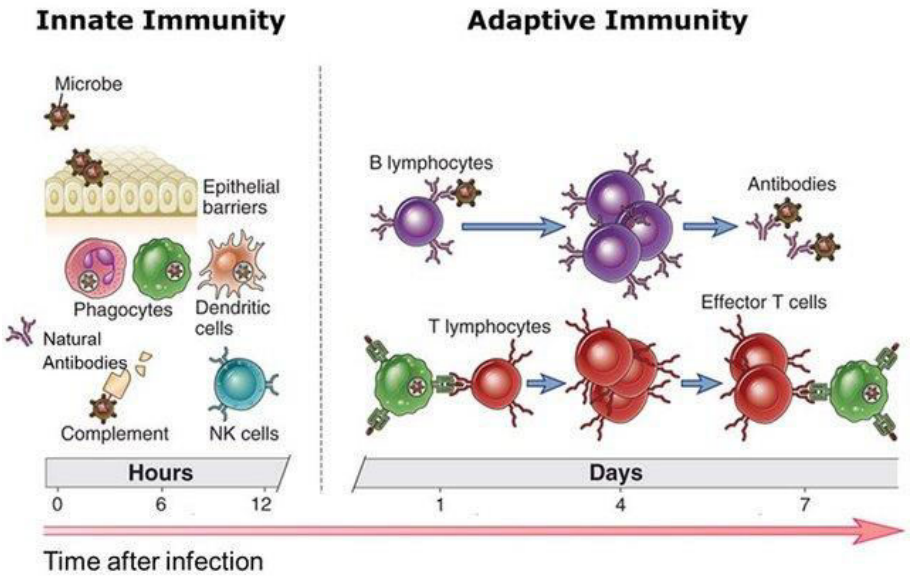


Figure 5 - The immune system. Representation of the arms of the immune system. On the left: innate immunity, comprising physical/chemical barriers, molecules and cells. On the right: adaptive immunity, comprising T and B lymphocytes. Under the boxes, approximate kinetics of their response. Adapted from[86]

The innate immune system

The innate immune system comprises a molecular and a cellular division. Molecules include pentraxins, naturally occurring antibodies and the complement system[87], while cells encompass granulocytes, monocytes, macrophages, mast cells, dendritic cells (DCs) and natural killer (NK) cells. All of these cells possess pattern recognition receptors (PRRs) able to recognize common microbial molecules, such as viral nucleic acids or bacterial and fungal cell-wall components – collectively called pathogen-

associated molecular patterns (PAMPs) – as well as molecules deriving from damaged cells – damage-associated molecular patterns (DAMPs)[88].

The adaptive immune system

On the other hand, the adaptive immune system is composed of B and T lymphocytes, cells that are unique in the organism. Their surface antigen receptors (BCR and TCR) are in fact the product of a cell-specific DNA recombination occurring at definite genomic loci during lineage maturation[89]. Recombination engenders a huge repertoire of antigen specificities, endowing the organism with the ability to potentially recognize every exogenous pathogen. Furthermore, adaptive immunity is characterized by the ability of lymphocytes to clonally expand upon encounter with their cognate antigen and to differentiate into the specialized subsets that best counteract the occurring infection[90]. Finally, a key advantage of adaptive over innate immunity resides in the ability of lymphocytes to retain memory of the encountered pathogens, improving efficacy and speed of the response to secondary exposures.

Crosstalk between innate and adaptive immune system

Upon engagement of the PRRs, cells of the innate immunity rapidly activate an inflammatory response to contain the infection[88]. They produce cytokines to recruit circulating leukocytes while they directly kill the pathogens through phagocytosis. Furthermore, some of them act as professional antigen presenting cells (APCs), activating and directing adaptive immune response[91]. Innate immunity, in fact, collects information about the microbial presence – starting with its location and continuing with its viability, replication and virulence[83] – that is

fundamental for the choice of the appropriate T lymphocyte effector response.

For the clearance of viral infection, for example, it is necessary to activate $CD8^+$ cytotoxic T lymphocytes[92], whereas the defence against intracellular bacteria and protozoa is mediated by $CD4^+$ T helper 1 (Th1) cells together with $CD8^+$ cytotoxic T cells[93]. Moreover, Th1 and Th17 are deputed to fungal pathogens elimination[94] and Th17 subset is also specialized in eliminating extracellular bacteria[95].

So, based on the characteristics of the APCs and of the environment in which they were activated, naïve T lymphocytes differentiate into the phenotype more appropriate for pathogen elimination.

1.6 T lymphocytes

Naïve T cells develop from the thymus, where progenitor cells undergo VDJ recombination to achieve the expression of a productive TCR. Mechanisms of positive and negative selection acting at this level ensure that mature T lymphocytes will be able to recognize peptides loaded into self-MHC molecules but that they will not activate in response to self-peptides[96]. As a last step of thymic maturation, TCR-rearranged $CD4^+CD8^+$ double-positive T lymphocytes repress one of the two co-receptors to became single positive $CD4^+$ or $CD8^+$ Naïve T cells. These mature cell subsets are then released in blood and lymph, where they can meet their cognate antigens loaded into either an MHC class I ($CD8^+$ T cells) or an MHC class II molecule ($CD4^+$ T cells). After activation, Naïve $CD8^+$ T cells develop a cytotoxic function, aimed at the direct killing of affected cells. On the other hand, Naïve $CD4^+$ T cells differentiate into one of multiple possible helper or regulatory phenotypes and orchestrate

the mounting immune response, by inducing corresponding specific phenotypes in the other innate and adaptive immune cells. $CD4^+$ T helper cells, for example, help B cells in the production of effective antibodies[97], enhance and maintain responses of $CD8^+$ T cells[98], regulate macrophage function[99] and suppress immune activation to control the magnitude and persistence of responses[100].

The fundamental role played by $CD4^+$ T lymphocytes in the immune system is stressed by the evidence that persistent $CD4^+$ T cell lymphopenia is associated with two serious malignant conditions: HIV-dependent immunodeficiency[101] and the very similar idiopathic $CD4^+$ lymphocytopenia[102].

1.7 Naive $CD4^+$ T lymphocyte activation

For productive activation, Naïve $CD4^+$ T lymphocytes need to receive from APCs at least two signals (Fig. 6): an antigen-specific signal through the interaction of the TCR with a peptide-bound MHC class II (pMHCII) and an APC-deriving signal via the binding of CD28 to CD80 or CD86[103,104].

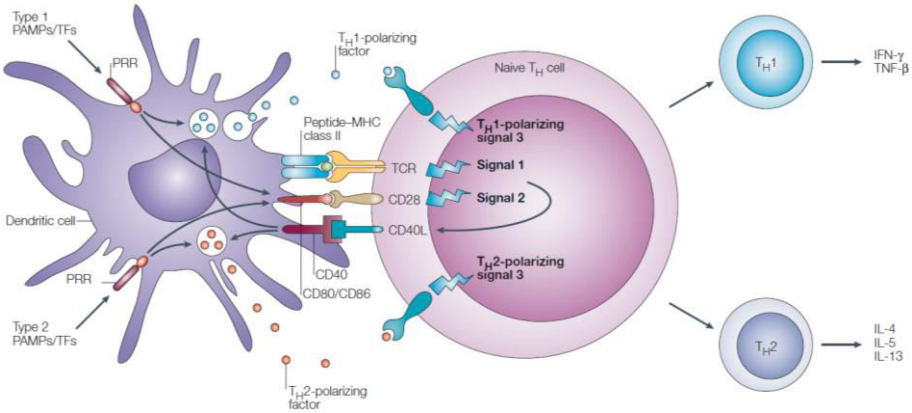


Figure 6 - The immunological synapse. APCs provide signals for optimal Naïve $CD4^+$ T lymphocytes activation: TCR engagement via

peptide-MHC II complex, co-stimulation through CD28 binding and secretion of polarizing factors. Adapted from[105]

The absence of CD28 co-stimulation, usually associated with MHC carrying self-peptides, prevent productive activation, resulting in the generation of unresponsive anergic T cells and tolerogenic Tregulatory (Treg) cells[106,107].

On the contrary, the delivery a complete stimulation rapidly initiates a cascade of signalling pathways that leads to efficient T cell clonal expansion and differentiation[108,109]. TCR, CD3 and costimulatory receptor molecules concentrate at the central supramolecular activation cluster (cSMAC), which is surrounded by a ring of actin and integrins named peripheral SMAC (pSMAC)[110]. In this situation, A-type lamins expression, absent in T Naïve quiescent cells, is transiently increased. Probably it serves, together with nesprins and SUN proteins, as connector between process at the plasma membrane and the nucleus[111–113]. F-actin, accumulating at the immunological synapse, has been held responsible for providing a scaffold for the assembly of signalling complexes[114], while the microtubule-organizing center (MTOC), directed to the centre of the immunological synapse, polarizes the Golgi apparatus towards the APC[115,116]. Overall, this immunological synapse structure is fundamental to allow stable cell-cell contacts, essential for efficient activation, and the organization of cell polarity, useful for the directional release of cytokines[117].

On the inner side of the plasma membrane, cytoplasmic tail of coreceptor molecules attract kinases and adaptor proteins initiating the signalling cascade that finally leads to a substantial increase in the intracellular Ca^{2+} levels[118]. Elevated levels of cytoplasmic Ca^{2+} activate calcium-dependent enzymes, such as calmodulin, PKC, IKK and CaMKII/CaMKIV, which

drive nuclear translocation of several transcription factor fundamental for T cell clonal expansion and differentiation, including CREB, AP-1, c-Myc, NFAT, NF-kB and DEC1[119–123]. Furthermore, high intracellular Ca^{2+} concentration have been shown to reduce mobility and induce rounding of otherwise polymorphic T cells, allowing the formation of a stable and effective immunological synapse[124]

Once effective activation has been achieved, basing on the third signal delivered by the APCs - the cytokine milieu – Naïve $CD4^+$ T lymphocyte begin the differentiation towards one their possible effector fates[125–129].

Th1 phenotype, induced by IL-12 is characterized by the production of the signature cytokine interferon-g (IFN-g)[127,130–133]; Th2, induced by IL-4, is characterized by GATA-binding protein 3 (Gata3) and IL-13[134–137], whereas Th17, induced by a combination of transforming growth factor-beta 1 (TGFb-1) and the cytokines IL-6 and IL-1 is characterized by the production of IL-17[128,138] and IL-21. Finally TGFb-1, IL-2 and retinoic acid can induce the transcription factor forkhead box P3 (FOXP3), generating regulatory T (Treg) cells that contribute to the maintenance of peripheral tolerance[139].

After the elimination of the pathogen, most of the clonally expanded cells die apoptotically. However, a reduced subset acquires a resting state and persists, preserving memory of the activation and providing faster and more effective response to a repeated challenge[140].

1.8 The epigenome of $CD4^+$ T lymphocytes

Emerging works provided association between the epigenome of immune cells and their functions. For example, the evolutionary repressor factor Ezh2 and its mark H3K27me3 have been demonstrated to be critical for

both T and B proper development[141,142]. Multiple epigenetic mechanisms at all levels of regulation, from the histone code up to chromatin organization in the nucleus, are also involved in T cell activation and differentiation. As an example, the correct development of Treg cells requires MLL4-mediated long-range chromatin interactions[143].

EM analysis allowed to observe that resting $CD4^+$ T cells have small nuclei with a high fraction of heterochromatin, which is lost upon activation[10], concomitantly with a modification of nucleosome positioning[144]. In line with this observation, lymphocytes in the resting state have lower levels of transcription compared to the active state[145]. Interestingly, recent evidence indicates that some RNAs necessary for T cell activation are already present in resting lymphocytes and they only require a translational switch upon TCR engagement[146].

Other works show that resting $CD4^+$ T cells contain Pol II islands at around 37% of their silent promoters, almost all of which are also marked by H3K4me3[34]. This suggests a poised state, possibly deriving from memory of past transcription or aiming at a rapid new activation. Furthermore, genes controlled by bivalent promoters, among which the master transcription factors Tbx21 and Gata3, have low transcriptional levels and do not present a uniform histone modification[147], suggesting they could serve as flexible switches, allowing plasticity of supposedly terminally differentiated $CD4^+$ T cells[148].

All this evidence, together with the absent or low levels of A-type lamins[149,150], suggests that T lymphocytes have a peculiar structure of chromatin.

1.9 Scope of the thesis

In our work we study chromatin organization in three different models and describe the interplay between nuclear architecture, positioning of Polycomb epigenetic repressors and cell identity.

In the first project (chapter 2), focusing on the murine model of Emery Dreifuss Muscular Dystrophy, we show that the lack of the structural protein Lamin A/C in muscle stem cells leads to an aberrant differentiation programme. In particular, the loss of chromatin higher-order structures causes the spreading of Polycomb group (PcG) of proteins repressors from their target genes over the flanking regions. The release of epigenetic repression results in premature exhaustion of quiescent stem cells and accumulation of fat in the muscles, determining muscular dystrophy progression.

In our second project (chapter 3) we use SAMMY-seq, a new high-throughput sequencing-based technique, to compare the chromatin structure of healthy and early-stages Hutchinson-Gilford Progeria Syndrome (HGPS) primary fibroblasts. We show that low levels of Progerin, a truncated form of Lamin A/C, interfere with the peripheral heterochromatin structure of Lamina Associated Domains (LADs), generating site-specific variations in the PcG-dependent H3K27me3 mark and transcriptional dysregulation of PcG target genes.

In the third project (chapter 4) the advances in SAMMY-seq technology allow us to describe an unconventional genomic organization in human primary $CD4^+$ T lymphocytes. The susceptibility of heterochromatin to DNAse digestion and the resistance of euchromatin to serial processes of extraction clearly distinguish lymphocytes from any other cell population analysed, opening unprecedented perspectives on the relation between chromatin organization and function.

References to chapter 1

1. Calero-Nieto, F. J. *et al.* Key regulators control distinct transcriptional programmes in blood progenitor and mast cells. *EMBO J.* **33**, 1212–1226 (2014).

2. Schoenfelder, S. *et al.* Divergent wiring of repressive and active chromatin interactions between mouse embryonic and trophoblast lineages. *Nat. Commun.* **9**, 4189 (2018).

3. D'Urso, A. & Brickner, J. H. Mechanisms of epigenetic memory. *Trends Genet.* **30**, 230–236 (2014).

4. Gorkin, D. U., Leung, D. & Ren, B. The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* **14**, 762–775 (2014).

5. Takata, H. *et al.* Chromatin Compaction Protects Genomic DNA from Radiation Damage. *PLoS One* **8**, e75622 (2013).

6. Piskadlo, E., Tavares, A. & Oliveira, R. A. Metaphase chromosome structure is dynamically maintained by condensin I-directed DNA (de)catenation. *Elife* **6**, 1–22 (2017).

7. Richmond, T. J., Luger, K., Mäder, A. W., Richmond, R. K. & Sargent, D. F. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).

8. Fyodorov, D. V., Zhou, B. R., Skoultchi, A. I. & Bai, Y. Emerging roles of linker histones in regulating chromatin structure and function. *Nat. Rev. Mol. Cell Biol.* **19**, 192–206 (2018).

9. Tremethick, D. J. Higher-Order Structures of Chromatin: The Elusive 30 nm Fiber. *Cell* **128**, 651–654 (2007).

10. Pombo, A. & Dillon, N. Three-dimensional genome architecture: Players and mechanisms. *Nat. Rev. Mol. Cell Biol.* **16**, 245–257 (2015).

11. Li, X. & Fu, X. D. Chromatin-associated RNAs as facilitators of functional genomic interactions. *Nat. Rev. Genet.* **20**, (2019).

12. Bantignies, F. & Cavalli, G. Polycomb group proteins: Repression in 3D. *Trends Genet.* **27**, 454–464 (2011).

13. Lanzuolo, C. & Orlando, V. Memories from the Polycomb Group Proteins. *Annu. Rev. Genet.* **46**, 561–589 (2012).

14. Ou, H. D. *et al.* ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science (80-.).* **357**, (2017).

15. Pueschel, R., Coraggio, F. & Meister, P. From single genes to entire genomes: The search for a function of nuclear organization. *Dev.* **143**, 910–923 (2016).

16. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).

17. Klose, R. J. & Bird, A. P. Genomic DNA methylation: The mark and its mediators. *Trends Biochem. Sci.* **31**, 89–97 (2006).

18. Bell, O., Tiwari, V. K., Thomä, N. H. & Schübeler, D. Determinants and dynamics of genome accessibility. *Nat. Rev. Genet.* **12**, 554–564 (2011).

19. Kim, Y. Z. Altered Histone Modifications in Gliomas. *Brain Tumor Res. Treat.* **2**, 7 (2014).

20. Talbert, P. B. *et al.* A unified phylogeny-based nomenclature for histone variants. *Epigenetics Chromatin* **5**, 7 (2012).

21. Koyama, M. & Kurumizaka, H. Structural diversity of the nucleosome. *J. Biochem.* **163**, 85–95 (2018).

22. Millar, C. B. Organizing the genome with H2A histone variants. *Biochem. J.* **449**, 567–579 (2013).

23. Ahmad, K. & Henikoff, S. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* **9**, 1191–1200 (2002).

24. Kouzarides, T. Chromatin Modifications and Their Function. *Cell* **128**, 693–705 (2007).

25. Rothbart, S. B. & Strahl, B. D. Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1839**, 627–643 (2014).

26. Prakash, K. & Fournier, D. Evidence for the implication of the histone code in building the genome structure. *BioSystems* **164**, 49–59 (2018).

27. Fischle, W., Wang, Y. & Allis, C. D. Histone and chromatin cross-talk. *Curr. Opin. Cell Biol.* **15**, 172–183 (2003).

28. Sekinger, E. A., Moqtaderi, Z. & Struhl, K. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell* **18**, 735–748 (2005).

29. Lai, B. *et al.* Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nature* **562**, 281–285 (2018).

30. Guo, X., Tatsuoka, K. & Liu, R. Histone acetylation and transcriptional regulation in the genome of Saccharomyces cerevisiae. *Bioinformatics* **22**, 392–399 (2006).

31. Roh, T.-Y. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* **19**, 542–552 (2005).

32. Pokholok, D. K. *et al.* Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122**, 517–527 (2005).

33. Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876–880 (2005).

34. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**, 823–837 (2007).

35. Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).

36. Won, K. J., Chepelev, I., Ren, B. & Wang, W. Prediction of regulatory elements in mammalian genomes using chromatin signatures. *BMC Bioinformatics* **9**, 1–18 (2008).

37. Roh, T. -y., Wei, G., Farrell, C. M. & Zhao, K. Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. *Genome Res.* **17**, 74–81 (2006).

38. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).

39. Kharchenko, P. V. *et al.* Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**, 480–486 (2011).

40. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* **28**, 817–825 (2010).

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

42. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125**, 315–326 (2006).

43. Bernstein, E. & Allis, C. D. RNA meets chromatin. *Genes Dev.* **19**,

1635–1655 (2005).

44. Loda, A. & Heard, E. Xist RNA in action: Past, present, and future. *PLOS Genet.* **15**, e1008333 (2019).

45. Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).

46. Mao, Y. S., Zhang, B. & Spector, D. L. Biogenesis and function of nuclear bodies. *Trends Genet.* **27**, 295–306 (2011).

47. Staněk, D. & Fox, A. H. Nuclear bodies: news insights into structure and function. *Curr. Opin. Cell Biol.* **46**, 94–101 (2017).

48. Zhu, L. & Brangwynne, C. P. Nuclear bodies: The emerging biophysics of nucleoplasmic phases. *Curr. Opin. Cell Biol.* **34**, 23–30 (2015).

49. van Steensel, B. & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780–791 (2017).

50. van Koningsbruggen, S. *et al.* High-Resolution Whole-Genome Sequencing Reveals That Specific Chromatin Domains from Most Human Chromosomes Associate with Nucleoli. *Mol. Biol. Cell* **21**, 3735–3748 (2010).

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

52. Sun, J., Shi, Y. & Yildirim, E. The Nuclear Pore Complex in Cell Type-Specific Chromatin Structure and Gene Regulation. *Trends Genet.* **35**, 579–588 (2019).

53. Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F. & Orlando, V.

Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat. Cell Biol.* **9**, 1167–1174 (2007).

54. Rieder, D., Trajanoski, Z. & McNally, J. G. Transcription factories. *Front. Genet.* **3**, 469–489 (2012).

55. Dundr, M. Nuclear bodies: Multifunctional companions of the genome. *Curr. Opin. Cell Biol.* **24**, 415–422 (2012).

56. Hansen, A. S., Cattoglio, C., Darzacq, X. & Tjian, R. Recent evidence that TADs and chromatin loops are dynamic structures. *Nucleus* **9**, 20–32 (2018).

57. Cremer, T. & Cremer, C. Chromosome Territories , Nuclear Architecture and Gene Regulation in Mammalian Cells. *Genetics* **2**, 292–301 (2001).

58. Sexton, T. & Cavalli, G. The role of chromosome domains in shaping the functional genome. *Cell* **160**, 1049–1059 (2015).

59. Bolzer, A. *et al.* Three-Dimensional Maps of All Chromosomes in Human Male Fibroblast Nuclei and Prometaphase Rosettes. *PLoS Biol.* **3**, e157 (2005).

60. Shachar, S. & Misteli, T. Causes and consequences of nuclear gene positioning. *J. Cell Sci.* **130**, 1501–1508 (2017).

61. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).

62. Fišerová, J. *et al.* Chromatin organization at the nuclear periphery as revealed by image analysis of structured illumination microscopy data. *J. Cell Sci.* **130**, 2066–2077 (2017).

63. Gilbert, N. *et al.* Chromatin Architecture of the Human Genome. *Cell* **118**, 555–566 (2004).

64. Tan-Wong, S. M., Wijayatilake, H. D. & Proudfoot, N. J. Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev.* **23**, 2610–2624 (2009).

65. Raices, M. & D'Angelo, M. A. Nuclear pore complexes and regulation of gene expression. *Curr. Opin. Cell Biol.* **46**, 26–32 (2017).

66. Souroullas, G. P. *et al.* An oncogenic Ezh2 mutation induces tumors through global redistribution of histone 3 lysine 27 trimethylation. *Nat. Med.* **22**, 632–640 (2016).

67. Huang, X. *et al.* Targeting Epigenetic Crosstalk as a Therapeutic Strategy for EZH2-Aberrant Solid Tumors. *Cell* **175**, 186-199.e19 (2018).

68. Kelly, A. D. & Issa, J. P. J. The promise of epigenetic therapy: reprogramming the cancer epigenome. *Curr. Opin. Genet. Dev.* **42**, 68–77 (2017).

69. Ho, R. & Hegele, R. A. Complex effects of laminopathy mutations on nuclear structure and function. *Clin. Genet.* **95**, 199–209 (2019).

70. Bianchi, A., Manti, P. G., Lucini, F. & Lanzuolo, C. Mechanotransduction, nuclear architecture and epigenetics in emery dreifuss muscular dystrophy: Tous pour un, un pour tous. *Nucleus* **9**, 321–335 (2018).

71. Wijchers, P. J. *et al.* Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments. *Mol. Cell* **61**, 461–473 (2016).

72. Ghavi-Helm, Y. *et al.* Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **51**, 1272–1282 (2019).

73. Deniaud, E. & Bickmore, W. A. Transcription and the nuclear periphery: edge of darkness? *Curr. Opin. Genet. Dev.* **19**, 187–191 (2009).

74. Harvey, A., Caretti, G., Moresi, V., Renzini, A. & Adamo, S. Interplay between Metabolites and the Epigenome in Regulating Embryonic and Adult Stem Cell Potency and Maintenance. *Stem Cell Reports* **13**, 573–589 (2019).

75. Lanzuolo, C., Lo Sardo, F. & Orlando, V. Concerted epigenetic signatures inheritance at PcG targets through replication. *Cell Cycle* **11**, 1296–1300 (2012).

76. Habibi, E. & Stunnenberg, H. G. Transcriptional and epigenetic control in mouse pluripotency: lessons from in vivo and in vitro studies. *Curr. Opin. Genet. Dev.* **46**, 114–122 (2017).

77. Sartorelli, V. & Juan, A. H. Sculpting chromatin beyond the double helix: Epigenetic control of skeletal myogenesis. *Curr. Top. Dev. Biol.* **96**, 57–83 (2011).

78. Osmanagic-Myers, S., Dechat, T. & Foisner, R. Lamins at the crossroads of mechanosignaling. *Genes Dev.* **29**, 225–237 (2015).

79. Isermann, P. & Lammerding, J. Nuclear Mechanics and Mechanotransduction in Health and Disease. *Curr. Biol.* **23**, R1113–R1121 (2013).

80. Haque, F. *et al.* SUN1 Interacts with Nuclear Lamin A and Cytoplasmic Nesprins To Provide a Physical Connection between the Nuclear Lamina and the Cytoskeleton. *Mol. Cell. Biol.* **26**, 3738–3751 (2006).

81. Gruenbaum, Y. & Foisner, R. Lamins: Nuclear Intermediate Filament Proteins with Fundamental Functions in Nuclear Mechanics and Genome Regulation. *Annu. Rev. Biochem.* **84**, 131–164 (2015).

82. Parkin, J. & Cohen, B. An overview of the immune system. **357**, 1777–1789 (2001).

83. Iwasaki, A. & Medzhitov, R. Control of adaptive immunity by the innate immune system. *Nat. Immunol.* **16**, 343–353 (2015).

84. Schwartz, R. H. Historical overview of immunological tolerance. *Cold Spring Harb. Perspect. Biol.* **4**, (2012).

85. Marques, R. E., Marques, P. E., Guabiraba, R. & Teixeira, M. M. Exploring the homeostatic and sensory roles of the immune system. *Front. Immunol.* **7**, 1–7 (2016).

86. Abbas, A., Litchman, A. & Pillai, S. *Cellular and molecular immunology*. (2012).

87. Shishido, S. N., Varahan, S., Yuan, K., Li, X. & Fleming, S. D. Humoral innate immune response and disease. *Clin. Immunol.* **144**, 142–158 (2012).

88. Kimbrell, D. A. & Beutler, B. The evolution and genetics of innate immunity. *Nat. Rev. Genet.* **2**, 256–267 (2001).

89. Roth, D. B. V(D)J Recombination: Mechanism, Errors, and Fidelity. in *Mobile DNA III* 313–324 (American Society of Microbiology, 2014). doi:10.1128/microbiolspec.MDNA3-0041-2014

90. Boehm, T. & Swann, J. B. Origin and Evolution of Adaptive Immunity. *Annu. Rev. Anim. Biosci.* **2**, 259–283 (2014).

91. Banchereau, J. & Steinman, R. M. Dendritic cells and the control of immunity. [Review] [103 refs]. *Nature* **392**, 245–252 (1998).

92. Wherry, E. J. *et al.* Molecular Signature of CD8+ T Cell Exhaustion during Chronic Viral Infection. *Immunity* **27**, 670–684 (2007).

93. Mashayekhi, M. *et al.* CD8α + Dendritic Cells Are the Critical Source

of Interleukin-12 that Controls Acute Infection by Toxoplasma gondii Tachyzoites. *Immunity* **35**, 249–259 (2011).

94. Romani, L. Cell mediated immunity to fungi: A reassessment. *Med. Mycol.* **46**, 515–529 (2008).

95. Hernández-Santos, N. & Gaffen, S. L. Th17 cells in immunity to Candida albicans. *Cell Host Microbe* **11**, 425–435 (2012).

96. Zúñiga-Pflücker, J. C. T-cell development made simple. *Nat. Rev. Immunol.* **4**, 67–72 (2004).

97. Qin, L. *et al.* Insights into the molecular mechanisms of T follicular helper-mediated immunity and pathology. *Front. Immunol.* **9**, 1–21 (2018).

98. Laidlaw, B. J., Craft, J. E. & Kaech, S. M. The multifaceted role of CD4+ T cells in CD8+ T cell memory. *Nat. Rev. Immunol.* **16**, 102–111 (2016).

99. Mark Doherty, T. T-cell regulation of macrophage function. *Curr. Opin. Immunol.* **7**, 400–404 (1995).

100. Dario A. A. Vignali, L. W. C. and C. J. W. How regulatory T cells work. *Nat rev immunol* **8**, 523–532 (2008).

101. Okoye, A. A. & Picker, L. J. CD4 + T cell depletion in HIV infection: mechanisms of immunological failure. *Immunol. Rev.* **254**, 54–64 (2013).

102. Yarmohammadi, H. & Cunningham-Rundles, C. Idiopathic CD4 lymphocytopenia. *Ann. Allergy, Asthma Immunol.* **119**, 374–378 (2017).

103. Lenschow, D. J., Walunas, T. L. & Bluestone, J. A. Cd28/B7 System of T Cell Costimulation. *Annu. Rev. Immunol.* **14**, 233–258 (1996).

104. Bour-Jordan, H. *et al.* Intrinsic and extrinsic control of peripheral T-

cell tolerance by costimulatory molecules of the CD28/B7 family. *Immunol. Rev.* **241**, 180–205 (2011).

105. Kapsenberg, M. L. Dendritic-cell control of pathogen-driven T-cell polarization. *Nat. Rev. Immunol.* **3**, 984–993 (2003).

106. Sakaguchi, S., Yamaguchi, T., Nomura, T. & Ono, M. Regulatory T Cells and Immune ToleranceSakaguchi, S., Yamaguchi, T., Nomura, T., and Ono, M., 2008. Regulatory T Cells and Immune Tolerance. Cell, 133 (5), 775–787. *Cell* **133**, 775–787 (2008).

107. Knoechel, B. *et al.* Functional and Molecular Comparison of Anergic and Regulatory T Lymphocytes. *J. Immunol.* **176**, 6473–6483 (2006).

108. Diehn, M. *et al.* Genomic expression programs and the integration of the CD28 costimulatory signal in T cell activation. *Proc. Natl. Acad. Sci.* **99**, 11796–11801 (2002).

109. Riley, J. L. *et al.* Modulation of TCR-induced transcriptional profiles by ligation of CD28, ICOS, and CTLA-4 receptors. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11790–11795 (2002).

110. Fooksman, D. R. *et al.* Functional Anatomy of T Cell Activation and Synapse Formation. *Annu. Rev. Immunol.* **28**, 79–105 (2010).

111. Starr, D. A. & Fridolfsson, H. N. Interactions Between Nuclei and the Cytoskeleton Are Mediated by SUN-KASH Nuclear-Envelope Bridges. *Annu. Rev. Cell Dev. Biol.* **26**, 421–444 (2010).

112. Burke, B. & Stewart, C. L. The nuclear lamins: Flexibility in function. *Nature Reviews Molecular Cell Biology* **14**, 13–24 (2013).

113. Robson, M. I. *et al.* Constrained release of lamina-associated enhancers and genes from the nuclear envelope during T-cell activation facilitates their association in chromosome compartments. *Genome Res.* **27**, 1126–1138 (2017).

114. González-Granado, J. M. *et al.* Nuclear envelope lamin-A couples actin dynamics with immunological synapse architecture and T cell activation. *Sci. Signal.* **7**, (2014).

115. Billadeau, D. D., Nolz, J. C. & Gomez, T. S. Regulation of T-cell activation by the cytoskeleton. *Nat. Rev. Immunol.* **7**, 131–143 (2007).

116. Vicente-Manzanares, M. & Sánchez-Madrid, F. Role of the cytoskeleton during leukocyte responses. *Nat. Rev. Immunol.* **4**, 110–122 (2004).

117. Wulfing, C. & Davis, M. M. A Receptor/Cytoskeletal Movement Triggered by Costimulation During T Cell Activation. *Science (80-.).* **282**, 2266–2269 (1998).

118. Cantrell, D. A. T-cell antigen receptor signal transduction. *Immunology* **105**, 369–374 (2002).

119. Jain, J. *et al.* The T-cell transcription factor NFATp is a substrate for calcineurin and interacts with Fos and Jun. *Nature* **365**, 352–355 (1993).

120. Kempiak, S. J., Hiura, T. S. & Nel, A. E. The Jun kinase cascade is responsible for activating the CD28 response element of the IL-2 promoter: Proof of cross-talk with the IκB kinase cascade. *J. Immunol.* **162**, 3176–3187 (1999).

121. Rao, S., Gerondakis, S., Woltring, D. & Shannon, M. F. c-Rel Is Required for Chromatin Remodeling Across the IL-2 Gene Promoter. *J. Immunol.* **170**, 3724–3731 (2003).

122. Marinari, B., Costanzo, A., Marzano, V., Piccolella, E. & Tuosto, L. CD28 delivers a unique signal leading to the selective recruitment of RelA and p52 NF- B subunits on IL-8 and Bcl-xL gene promoters. *Proc. Natl. Acad. Sci.* **101**, 6098–6103 (2004).

123. Sánchez-Valdepeñas, C., Martín, A. G., Ramakrishnan, P., Wallach, D. & Fresno, M. NF-κB-Inducing Kinase Is Involved in the Activation of the CD28 Responsive Element through Phosphorylation of c-Rel and Regulation of Its Transactivating Activity. *J. Immunol.* **176**, 4666–4674 (2006).

124. Negulescu, P. A., Krasieva, T. B., Khan, A., Kerschbaum, H. H. & Cahalan, M. D. Polarity of T cell shape, motility, and sensitivity to antigen. *Immunity* **4**, 421–430 (1996).

125. Ansel, K. M., Djuretic, I., Tanasa, B. & Rao, A. Regulation of Th2 differentiation and Il4 locus accessibility. *Annu. Rev. Immunol.* **24**, 607–656 (2006).

126. Murphy, K. M. & Reiner, S. L. The lineage decisions of helper T cells. *Nat. Rev. Immunol.* **2**, 933–944 (2002).

127. Szabo, S. J., Sullivan, B. M., Peng, S. L. & Glimcher, L. H. Molecular mechanism regulating Th1 immune responses. *Annu. Rev. Immunol.* **21**, 713–758 (2003).

128. Weaver, C. T., Hatton, R. D., Mangan, P. R. & Harrington, L. E. IL-17 Family Cytokines and the Expanding Diversity of Effector T Cell Lineages. *Annu. Rev. Immunol.* **25**, 821–852 (2007).

129. Zhu, J. & Paul, W. E. CD4 T cells : fates , functions , and faults ASH 50th anniversary review CD4 T cells : fates , functions , and faults. *Immunobiology* **112**, 1557–1569 (2009).

130. Afkarian, M. *et al.* T-bet is a STAT1-induced regulator for IL-12R expression in naïve CD4 + T cells. *Nat. Immunol.* **3**, 549–557 (2002).

131. Müller, L. S. M. *et al.* Genome organization and DNA accessibility control antigenic variation in trypanosomes. *Nature* **563**, 121–125 (2018).

132. Szabo, S. J. *et al.* A novel transcription factor, T-bet, directs Th1 lineage commitment. *Cell* **100**, 655–669 (2000).

133. Zhang, L. *et al.* Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**, 268–272 (2018).

134. Kaplan, M. H., Schindler, U., Smiley, S. T. & Grusby, M. J. Stat6 is required for mediating responses to IL-4 and for the development of Th2 cells. *Immunity* **4**, 313–319 (1996).

135. Fukada, S. *et al.* Molecular Signature of Quiescent Satellite Cells in Adult Skeletal Muscle. *Stem Cells* **25**, 2448–2459 (2007).

136. Takeda, K. *et al.* Essential role of Stat6 in IL-4 signalling. *Nature* **380**, 627–630 (1996).

137. Zheng, W. & Flavell, R. A. The Transcription Factor GATA-3 Is Necessary and Sufficient for Th2 Cytokine Gene Expression in CD4 T Cells. *Cell* **89**, 587–596 (1997).

138. Harrington, L. E., Mangan, P. R. & Weaver, C. T. Expanding the effector CD4 T-cell repertoire: the Th17 lineage. *Curr. Opin. Immunol.* **18**, 349–356 (2006).

139. Zheng, Y. & Rudensky, A. Y. Foxp3 in control of the regulatory T cell lineage. *Nat. Immunol.* **8**, 457–462 (2007).

140. Garrod, K. R. *et al.* Dissecting T Cell Contraction In Vivo Using a Genetically Encoded Reporter of Apoptosis. *Cell Rep.* **2**, 1438–1447 (2012).

141. Mandal, M. *et al.* Epigenetic repression of the Igk locus by STAT5-mediated recruitment of the histone methyltransferase Ezh2. *Nat. Immunol.* **12**, 1212–1220 (2011).

142. Su, I. H. *et al.* Ezh2 controls B cell development through histone H3

methylation and Igh rearrangement. *Nat. Immunol.* **4**, 124–131 (2003).

143. Placek, K. *et al.* MLL4 prepares the enhancer landscape for Foxp3 induction via chromatin looping. **18**, (2017).

144. Schones, D. E. *et al.* Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell* **132**, 887–898 (2008).

145. Frangini, A. *et al.* The Aurora B Kinase and the Polycomb Protein Ring1B Combine to Regulate Active Promoters in Quiescent Lymphocytes. *Mol. Cell* **51**, 647–661 (2013).

146. Ricciardi, S. *et al.* The Translational Machinery of Human CD4+ T Cells Is Poised for Activation and Controls the Switch from Quiescence to Metabolic Remodeling. *Cell Metab.* **28**, 961 (2018).

147. Wei, G. *et al.* Global Mapping of H3K4me3 and H3K27me3 Reveals Specificity and Plasticity in Lineage Fate Determination of Differentiating CD4+ T Cells. *Immunity* **30**, 155–167 (2009).

148. Roh, T. Y., Cuddapah, S., Cui, K. & Zhao, K. The genomic landscape of histone modifications in human T cells. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15782–15787 (2006).

149. Röber, R. A., Sauter, H., Weber, K. & Osborn, M. Cells of the cellular immune and hemopoietic system of the mouse lack lamins A/C: distinction versus other somatic cells. *J. Cell Sci.* **95 (Pt 4)**, 587–98 (1990).

150. Guilly, M.-N., Kolb, J.-P., Gosti, F., Godeau, F. & Courvalin, J.-C. Lamins A and C are not expressed at early stages of human lymphocyte differentiation. *Exp. Cell Res.* **189**, 145–147 (1990).

CHAPTER 2.
POLYCOMB DYSFUNCTIONAL TRANSCRIPTIONAL REPRESSION CONTRIBUTES TO LAMIN A/C DEPENDENT MUSCULAR DYSTROPHY

Bianchi Andrea[§], Mozzetta Chiara[§], Lucini Federica, Valsoni Sara, Pegoli Gloria, Rosti Valentina, Petrini Cristiano, Cortesi Alice, Antonelli Laura, Gregoretti Francesco, Oliva Gennaro, De Bardi Marco, Rizzi Roberto, Bodega Beatrice, Pasini Diego, Ferrari Francesco, Bearzi Claudia[*] and Lanzuolo Chiara[*]

submitted

Polycomb dysfunctional transcriptional repression contributes to Lamin A/C dependent muscular dystrophy

Bianchi Andrea[1, 2*], Mozzetta Chiara[1*], Lucini Federica[2], Valsoni Sara[1,3], Pegoli Gloria[3], Rosti Valentina[4], Petrini Cristiano[5], Cortesi Alice[2], Antonelli Laura[6], Gregoretti Francesco[6], Oliva Gennaro[6], De Bardi Marco[3], Rizzi Roberto[1], Bodega Beatrice[2], Pasini Diego[7], Ferrari Francesco[5, 8], Bearzi Claudia[1, 9] and Lanzuolo Chiara[1, 2, 3]

1. CNR Institute of Cell Biology and Neurobiology, Rome, Italy
2. Istituto Nazionale di Genetica Molecolare "Romeo ed Enrica Invernizzi", Milan, Italy
3. IRCCS Santa Lucia Foundation, Rome, Italy
4. CNR Institute of Biomedical Technologies, Milan, Italy
5. IFOM, the FIRC Institute of Molecular Oncology, Milan, Italy
6. CNR Institute for High Performance Computing and Networking, Naples, Italy
7. European Institute of Oncology, Milan, Italy
8. Institute of Molecular Genetics, National Research Council, Pavia, Italy
9. MultiMedica Scientific and Technology Pole, Milan, Italy

* equal contribution

Conflict of interest statement

The authors have declared that no conflict of interest exists.

Keywords

Muscle Stem Cells (MuSCs), Lamin A, Polycomb, cell fate, differentiation, premature senescence, muscular dystrophy.

Running head

PcG dysfunctions in lamin dystrophy

Abstract

Lamin A is a component of the inner nuclear membrane that, together with epigenetic factors, organizes the genome in higher order structures required for transcriptional control. Mutations in the Lamin A/C gene cause several diseases, belonging to the class of laminopathies, including muscular dystrophies. Nevertheless, molecular mechanisms involved in the pathogenesis of Lamin A-dependent dystrophies are still largely unknown. Polycomb group of proteins (PcG) are epigenetic repressors and Lamin A interactors, primarily involved in the maintenance of cell identity. Using different murine models of Emery-Dreifuss Muscular Dystrophy (EDMD), we show here that Lamin A loss deregulates PcG positioning in muscle satellite stem cells leading to de-repression of non-muscle specific genes and $p16^{INK4a}$, a senescence driver encoded in the Cdkn2a locus. This aberrant transcriptional programme causes impairment in self-renewal, loss of cell identity and premature exhaustion of quiescent satellite cell pool. Genetic ablation of Cdkn2a locus restores muscle repair and muscle stem cell properties in Lamin A/C null dystrophic mice. Our findings establish a direct link between Lamin A and PcG epigenetic silencing and indicate that Lamin A-dependent muscular dystrophy can be ascribed to intrinsic epigenetic dysfunctions of muscle stem cells.

Introduction

The nuclear lamina (NL) is located in the inner part of the nuclear membrane and is made up of a complex network of type V filament proteins, the lamins (1, 2). In vertebrates lamin proteins are divided into A and B types, based on sequence homologies. A growing body of evidence suggests that lamins are directly involved in the functional control of the genome, by organizing its three dimensional positioning in the nuclear space, through the association with transcriptionally repressed large genomic regions, called Lamina-associated domains (LADs) (3). The crucial function of lamins is attested by an entire class of genetic diseases, called *laminopathies*, where specific components of the NL are altered (4). In particular, the study of Lamin A/C is gaining an increasing interest for three reasons: i) Lamin A/C plays an undisputed role in several cellular processes from mechanotransduction to cell differentiation; ii) Lamin A/C has a peculiar intranuclear distribution being present in the nucleoplasm as well as in the nuclear periphery (5); iii) Lamin A/C interacts with several epigenetic factors, exerting a functional control on transcriptional regulation (3, 6). One of the most studied Lamin A/C dependent cellular process is myogenesis because mutations in the *LMNA* gene lead to muscular dystrophies, as in the case of Emery Dreifuss Muscular Dystrophy (EDMD) (7). However, epigenetic mechanisms involved in lamin-dependent dystrophy are still largely unknown. PcG proteins are epigenetic repressors originally discovered for their central roles in development and cell differentiation (8) and recently described as functional partners of Lamin A/C (9-13). In the last years several evidence demonstrated that PcG proteins are involved in the regulation of adult stem cells (14-17), safeguarding cell identity and preventing cell fate transition (18). In multipotent stem cells, PcG proteins ensure the correct balance between self-renewal and lineage-specific differentiation, promptly responding to the environmental changes. At the molecular level this is achieved through PcG binding at bivalent domains, genomic regions containing active and repressive epigenetic signatures and a poised RNA polymerase II (19). This

epigenetic condition allows a rapid transition from one transcriptional state to another, ensuring the correct expression of unique and specific cell lineage genes. Defects in these fine-tuned mechanisms lead to lack of cell identity (20) or pathological reprogramming (21).

Given their key role in regulating stem cells fate decisions and tissue homeostasis, it is conceivable that PcG dysfunctions contribute to lamin-dependent, tissue-specific human diseases. Here, we examined how the absence of Lamin A/C impacts muscle stem (satellite) cells (MuSCs) *in vivo*, and the role of PcG proteins in lamin muscular dystrophy. We found that MuSCs lacking LaminA/C redistribute PcG-dependent histone marks leading to transcriptional upregulation of crucial PcG-target genes, such as non-muscle related genes. This leads to lack of MuSC cell identity and cellular senescence, determining a premature exhaustion of the muscular stem cell niche. Genetic ablation of the PcG-regulated Cdkn2a locus in lamin dystrophic mice restores MuSCs regenerative defects.

Results

Lamin A is required to preserve quiescent muscle stem cells (MuSC) pool.

We used two dystrophic mouse models for EDMD: the knock-in *Lmna* (H222P/H222P) mice that present a mild skeletal muscle alterations (22) and the *Lmna* Δ8-11 mice showing an aggressive, extensive and early onset muscular dystrophy, that leads to premature death at 1-2 months after birth (23). Quantification of the relative amounts of quiescent (QSCs; PAX7+/MYOD-) and activated (ASCs; PAX7+/MYOD+) MuSCs in three months-old H222P/H222P mice did not reveal any defect in the total amount of MuSCs (Figure S1A and S1B, Control), in the cross-sectional area of the muscle fibers (Figure S1A and S1C, Control) or in the balance of QSCs vs ASCs (Figure S1A and S1D, Control) compared to wild type mice. Also, upon damage caused by repeated injuries with cardiotoxin (CTX) on tibialis anterior (TA) muscles (Figure S1E) no differences in the regeneration efficiency of wt and

H222P/H222P littermates could be found (Figure S1A, S1B, S1C and S1D; Injured) thus suggesting that no defects in MuSCs activation in the muscle stem cell niche occur in this genetic context. We then analysed the stem cell niche composition during dystrophy progression in the severe dystrophic *Lmna* Δ8-11 -/- mice (homozygous, hom), together with their unaffected littermates, wild type (wt, *Lmna* Δ8-11 +/+) or heterozygous (het, *Lmna* Δ8-11 +/-), during dystrophy progression at 10, 14, 16 and 19 days after birth (labelled as P10, P14, P16 and P19, respectively). In early stages of post-natal growth (P10 and P14) no differences were found in the relative amounts of QSCs and ASCs (Figure 1A, 1B and S2A) among distinct genotypes. Conversely, starting from P16, an unbalance of MuSCs becomes evident in *Lmna* Δ8-11 -/- muscles, with a decreased proportion of QSCs compared to ASCs (Figure 1A, 1B and S2A), mirroring a decline in myofibers cross sectional area (CSA) (Figure 1C). Of note, the overall amount of Pax7+ MuSCs was not significantly altered across the different genotypes (Figure S2B) and Ki67 staining at P19 confirmed that in *Lmna* Δ8-11 -/- muscles a lower amount of QSCs (Pax7+/Ki67-) is present (Figure S2C). These findings suggested that *Lmna* Δ8-11 -/- MuSCs may be deficient in self-renewal capacity. To test this hypothesis, we isolated single myofibers at P19 and cultured them for 96h, monitoring their ability to give rise to QSCs (PAX7+/MYOD-), ASCs (PAX7+/MYOD+) and differentiating (PAX7-/MYOD+) cells (Figure 1D and 1E). In fibers isolated from *Lmna* Δ8-11 -/- muscles, we observed a decrease in the number of QSCs compared to wt, paralleled by a diminished number of differentiating cells (MYOD+/PAX7-) and an increased number of ASCs. This suggests a defect in self-renewal and muscle differentiation (24), resulting in a lower number of MuSCs within different clusters accompanied by proliferation defects *in vitro* (Figure S2D). Interestingly, the healthy heterozygous *Lmna* Δ8-11 +/- mice, although not developing muscular dystrophy (23), presents a lower amount of QSCs compared to wt (Figure 1D and 1E), suggesting that proper Lamin A levels are important for MuSCs homeostasis to preserve their self-renewal

capacity. To further address this, we performed repeated muscle injuries (as in Figure S1E) on adult heterozygous *Lmna* Δ8-11 +/- mice, showing less Lamin A at mRNA and protein level (Figure S2E and S2F). Analysis of MuSC populations revealed a lower amount of QSCs in *Lmna* Δ8-11 +/- muscles upon repeated injuries (Figure 1F and 1G; Injured) and a decline in Pax7+ cells (Figure 1H; Injured), suggesting that Lamin A affects MuSCs self-renewal in a dose dependent manner and that normal levels of Lamin A are required to replenish the QSCs pool during regeneration and to preserve MuSCs self-renewal.

***Lmna* Δ8-11 -/- dystrophic MuSCs displays PcG binding alterations**

Our recent results showed a LaminA/C-PcG crosstalk along *in vitro* myogenesis (10). We thus wondered if the altered MuSCs balance observed in *Lmna* Δ8-11 -/- muscles might be ascribed to aberrant PcG functions. We first performed immunostaining of Ezh2, the catalytic subunit of Polycomb Repressive Complex 2 (PRC2) (Figure S3A) in P19 MuSCs. We fixed MuSCs before FACS-Sorting to preserve the nuclear architecture of Lamin A-deficient cells (see methods). We found a general intranuclear diffusion of Ezh2 in *Lmna* Δ8-11 -/- MuSCs, ascertained by measuring PcG bodies parameters (25) (Figure S3A, S3B and S3C) further corroborating previous findings (10, 11). Then, to gain further insights into possible PcG-dependent transcriptional defects, we performed RNA-sequencing on freshly isolated MuSCs at P19, finding 1424 upregulated genes and 1842 downregulated genes in the *Lmna* Δ8-11 -/- MuSCs with respect to the wt (Figure S3D). Interestingly, performing a Gene Set Enrichment Analysis (GSEA) based on differential expression generated after conditional ablation of Ezh2 in MuSCs (26) and *Lmna* Δ8-11 -/- up-regulated genes, we found a significant association between the two datasets suggesting that Lamin A absence impairs Ezh2 function (Figure S3E). We also followed the deposition of the Ezh2 dependent H3K27me3 histone mark in *Lmna* Δ8-11 mice by quantitative spike-in ChIP-seq (27) (Figure S3F and

S3G). Integrative analysis of RNA-seq and ChIP-seq revealed that up-regulated genes in Lmna Δ8-11 -/- condition are significantly enriched for H3K27me3 targets (identified in wt condition) (Figure 2A, Fisher's Exact Test p-value = 2.38e-05). Indeed, analysis of H3K27me3 distribution around the Transcription Start Sites (TSS) and along the body of genes indicated a global decrease of this repressive mark in *Lmna* Δ8-11 -/- MuSCs with respect to wt (Figure 2B and 2C), accompanied by a mild decrease in H3K27me3 global levels detected by western blot (Figure S3H and S3I). In contrast, a deep analysis of H3K27me3 ChIP-seq reads coverage in the intervening genomic regions between the known H3K27me3 enrichment peaks interestingly showed a higher average coverage in the *Lmna* Δ8-11 -/- MuSCs with respect to wt counterparts (Figure 2D). These results are compatible with a diffusion of PcG proteins along the chromatin fibers rather than a complete PcG displacement. To identify the PcG targets mostly affected by Lamin A deficiency, genes were grouped according to their transcription level in wt MuSCs. We thus defined one class of genes with no coverage by RNA-seq reads plus 4 equally sized groups of genes based on expression level quartiles (Figure 2E). For each expression category, we reanalysed the H3K27me3 distribution along the body of genes and the TSS and the percentage of upregulated genes in the *Lmna* Δ8-11 -/- MuSCs (Figure 2E, 2F and S3J). In quartiles 0 and I we found only a small percentage (respectively 0.02 and 0.65%) of upregulated genes in *Lmna* Δ8-11 -/-, suggesting that the H3K27me3 decrease/redistribution is not sufficient to activate transcription in highly repressed genes (Figure 2F, S3J and S4A). In contrast, quartiles containing transcribed genes (II, III and IV) are more affected by the diminished H3K27me3 levels in *Lmna* Δ8-11 -/- (Figure 2F, S3J and S4B), showing a percentage of upregulated genes between 5 and 9%. Specifically, we noticed that in wt MuSCs, H3K27me3 ChIP-seq signal enrichment around the TSS and within the body of genes is progressively lower in quartiles of higher expression, as expected. However,

for *Lmna* Δ8-11 -/- mice the decrease of H3K27me3 signal inside the gene body is relatively less marked than in wt mice (Figure 2F, III and IV quartiles and S3J). We quantified and confirmed this observation by considering the ratio of H3K27me3 ChIP-seq enrichment signal at the TSS and 2.5Kb downstream of the TSS, for each gene, in wt and *Lmna* Δ8-11 -/- mice (Figure S4C), showing that this ratio is significantly different (lower in *Lmna* Δ8-11 -/- than wt MuSCs) for higher expression quartiles (Figure S4D).

LaminA-dependent PcG redistribution determines de-repression of non-muscle related bivalent genes

The altered PcG binding observed in *Lmna* Δ8-11 -/- MuSCs, prompted us to examine more in details the bivalent genes, a subgroup of PcG targets whose expression is more susceptible to variations of PcG occupancy (28). Bivalent genes are characterized by the concurrent presence of both H3K27me3 and H3K4me3 marks around TSS and have an intermediate gene expression state (29, 30). We first performed H3K27me3 and H3K4me3 ChIP-seq in wt MuSCs (Figure S5A and S5B) and we defined bivalent genes using the parameters described in (19) for the H3K4me3 window at TSS (Figure S5C). Then, we tested the association between up-regulated and bivalent genes in the Lmna Δ8-11 -/- by means of the Fisher exact test. We observed a significant (Fisher Exact Test p-value = 4.57e-07) over-representation of bivalent genes among up-regulated ones in the Lmna Δ8-11 -/- MuSCs (Figure 3A). To gain more insights into the biological relevance of deregulated genes in the mutant mice, we performed semantic similarity analysis of all GO terms associated with upregulated genes (Figure 3B) together with GSEA (Figure S5D and S5E). These analyses showed a negative correlation with muscle specification (Figure S5D) together with an acquisition of markers related to lipid metabolic processes (Figure 3B and S5E). Notably, Fisher exact test analysis

highlighted a significant overlap between genes with bivalent promoter and the Lmna Δ8-11 -/- MuSCs up-regulated genes involved in adipogenesis (Figure 3C, Fisher Exact Test p-value = 6.73e-06), suggesting that Lamin A is involved in preserving MuSCs identity by ensuring the correct PcG-mediated transcriptional repression.

PPARγ is aberrantly expressed in *Lmna* Δ8-11 -/- dystrophic MuSCs

Given this strong association between bivalent reactivation and adipogenesis markers (Figure 3C), we analysed the expression of PPARγ (Peroxisome proliferator-activated receptor gamma), the master transcription factor for adipose cell differentiation (31, 32), which in our genome-wide analyses has the bivalent signature (Figure 2E and S6A). We stained muscles with PAX7 and PPARγ to directly test if MuSCs displayed aberrant expression of PPARγ in the absence of Lamin A. Strikingly, we found about 10% of *Lmna* Δ8-11 -/- MuSCs that simultaneously express both muscular and adipogenesis markers being PAX7+/ PPARγ+ (Figure 4A and 4B). Accordingly, the genomic region of the PPARG gene showed a decrease of H3K27me3 enrichment around the TSS in the *Lmna* Δ8-11 -/- MuSCs, accompanied by a transcriptional up-regulation (Figure 4C). Considering the key role of PcG proteins in mediating the formation of chromatin loop structures (33, 34), we reasoned that the loss of H3K27me3 and transcriptional up-regulation of PPARG locus could be related to the alteration of chromatin 3D structure. The genome 3D architecture is organized in structurally separated Topologically Associated Domains (TADs), chromosomal structures that favour intra-domain looping interactions (35). TADs can be identified by genome-wide chromosome conformation capture (Hi-C) and are largely conserved across different cell types (36). We verified that the PPARG locus is included in a TAD encompassing a region extending also upstream of the PPARG locus itself, using high-resolution Hi-C data on mouse embryonic stem cells (37) and the 3D Genome Browser (38) (Figure S6B).

Then we performed 3D multicolor DNA FISH analysis on pre-fixed MuSCs using one BAC probe overlapping the PPARG promoter and a second probe annealing at the TAD border. We observed an overlap of the signals from the two regions in the wt *Lmna* Δ8-11 +/+ MuSCs that indicates the presence of a DNA looping *in cis* (Figure 4D and 4E). By contrast, in *Lmna* Δ8-11 -/- MuSCs the distance between the signals was higher, definitely suggesting the lack of DNA/DNA interaction. Indeed, from the analysis of H3K27me3 ChIPseq tracks we noticed in the *Lmna* Δ8-11 -/- MuSCs a reduction of H3K27me3 peaks upstream the PPARG locus (Figure S6B). FISH analysis also highlighted that in wt the entire genomic region is close to the nuclear periphery (Figure 4D and 4F) while in *Lmna* Δ8-11 -/- is re-located in the nuclear interior, suggesting that Lamin A absence interferes with chromatin anchoring to the nuclear lamina and PcG dependent DNA remodelling.

To evaluate the relevance of our findings in human EDMD, we generated iPS cells starting from healthy donors and EDMD patient's myoblasts, carrying R453W mutation in LMNA gene (Figure S7A, S7B and S7C). Cells were then differentiated *in vitro* toward myotubes (Figure S7D) and several markers were measured at different time points. In parallel with a regular downregulation of the OCT4 stemness marker, we found, in the EDMD clones an upregulation of Pax7 and muscle markers (Figure S7E and S7F), resembling what we previously found in the murine model *in vitro* upon Lamin A depletion (10). Furthermore, in the EDMD clones, PPARγ was aberrantly upregulated during EDMD muscle differentiation (Figure S7G), thus recapitulating the lack of cell identity observed in murine *Lmna* Δ8-11 -/- MuSCs.

***Lmna* Δ8-11 -/- MuSCs undergo premature senescence**

Taken together, these results clearly point towards a role of LaminA in mediating PcG-transcriptional repression in MuSCs to safeguard their identity and regenerative capacity. This lack of cell identity and the impairment of self-renewal displayed by *Lmna* Δ8-11 -/- MuSCs are all features reminiscent of the phenotype described for

Ezh2-null MuSCs (26). Moreover, the impairment in self-renewal and the progressive decline of MuSCs pool are also typical traits of aged MuSCs (39, 40) in which both Lamin A/C and PcG proteins play a key role (41, 42). A key cellular mechanism that ensures self-renewal and hence the maintenance of the MuSC pool is the asymmetric division (43). At the molecular level, in aged mice, the accumulation of activated form of P38 (phospho-p38, Ph-P38) and its symmetric distribution in MuSC doublets heavily compromise the self-renewal capacity leading to MuSCs functional decline (44-47). To test whether premature exhaustion of quiescent *Lmna* Δ8-11 -/- MuSCs cells could be ascribed to a defective asymmetric division, we stained myofibers-associated MuSCs for Ph-p38 after 48h of culture, a time in which the activated MuSC divides and forms, in most cases, a cell doublet (48, 49) (Figure 5A). In contrast to wt, *Lmna* Δ8-11 -/- MuSC doublets showed a preferential symmetric distribution of ph-p38 (Figure 5A and 5B) mostly accompanied by a planar orientation with respect to myofibers (Figure 5C). This highlights problems in asymmetric division, which should be instead characterized by apico-basal orientation (50). In line with this result, in *Lmna* Δ8-11 -/- muscle sections we found higher amount of Ph-P38+ (Figure S8A and S8B) MuSCs and sign of genomic instability, as measured by increased γH2AX DNA repair signal foci (Figure S8C and S8D). To test if defective asymmetric division is associated to premature senescence we then analysed RNAseq to determine if *Lmna* Δ8-11 -/- MuSCs share the same transcriptional signature of MuSCs isolated from aging mice. We performed a GSEA using two different RNA datasets from MuSCs of 24 months old mice (51, 52) and *Lmna* Δ8-11 -/- up-regulated genes. In line with our hypothesis, we found that 19 days-old *Lmna* Δ8-11 -/- MuSCs presents an upregulated transcriptome similar to 20-24-months-old MuSCs (Figure 5D and S8E), but different from geriatric 28-32-months-old MuSCs (Figure S8F). At the molecular level, the senescence program is supported by upregulation of some PcG-regulated Cyclin-dependent Kinase Inhibitors (CDKIs) (53) as p21, involved in cellular senescence and in cell cycle

arrest. p21 maintains the viability of DNA damage-induced senescent cells (54) and aberrant expression of p21 has been observed in EDMD derived human myoblasts (55). ChIP-seq and RNA-seq analysis of Cdkn1a/p21 locus showed a displacement of Ezh2 from the promoter accompanied by an upregulation of p21 transcript in *Lmna* Δ8-11 -/- MuSCs (Figure 5E). These findings suggest a role for LaminA-PcG-mediated repression in preserving MuSCs proliferative and regenerative capacities, thus preventing premature senescence.

Genetic ablation of Cdkn2a locus partially rescues self-renewal defects in *Lmna* Δ8-11 -/- dystrophic mice.

To further corroborate our findings, we also analysed the Cdkn2a locus, a PcG target primarily involved in muscular senescence (52) (Figure 6A). Two transcripts, $p16^{INK4a}$ and $p19^{ARF}$ originate from Cdkn2a locus (56). Interestingly, it has been recently reported that $p16^{INK4a}$ expression is a second event, subsequent to p21 upregulation, in the cellular senescence progression (57). In line with these observations, $p16^{INK4a}$ expression is specifically induced in geriatric 28-32-months-old MuSCs (not in 24-months-old MuSCs) (52). Moreover, depletion of $p16^{INK4a}$ is sufficient to reduce senescence-associated gene expression in geriatric MuSCs. RNA-seq and qRT-PCR on 19-days old *Lmna* Δ8-11 -/- MuSCs did not reveal any transcription of $p16^{INK4a}$ in wt or *Lmna* Δ8-11 -/- (Figure 6B). However qRT-PCR analysis performed in older mice (26 days from birth), revealed higher levels of $p16^{INK4a}$ transcript in *Lmna* Δ8-11 -/- MuSCs compared to the wt counterpart (Figure 6B), suggesting a transition during dystrophy progression toward geriatric condition. We thus decided to test whether the genetic ablation of Cdkn2a locus could reverse *Lmna* Δ8-11 -/- MuSCs premature aging, by crossing *Lmna* Δ8-11 +/- with *Cdkn2a* -/- mice (58). Analysis performed in *Lmna* Δ8-11 +/+ background showed no differences in the percentage of QSCs and ASCs, nor on CSA, suggesting that Cdkn2a is dispensable for post-natal muscle development (Figure 6C, 6D and 6E; *LMNA+/+*). On the other hand,

Cdkn2a Lmna Δ8-11 -/- mice partially rescued the quiescent MuSC pool and CSA defects and CSA defects observed in the absence of *LMNA* (Figure 6C, 6D and 6E; *LMNA-/-;* S9A and S9B) emphasizing that Lmna-dependent muscular dystrophy might be due to progressive MuSCs functional decline caused by acquisition of premature aging features.

Discussion

LaminA-dependent muscular dystrophy pathogenesis has been classically ascribed to nuclear fragility that renders myonuclei more prone to mechanical stress and damage imposed by myofibre contraction. However, the evidence that Lamin A/C is expressed also by MuSCs has led to suggest that satellite cell dysfunction might contribute to EDMD progression (59), yet experimental evidence of this hypothesis was still lacking.

Cell fate choice during muscle differentiation is governed by epigenetic factors controlling the sequential restriction of transcriptional programs (60). Any dysfunction in this finely tuned epigenetic regulation could lead to impaired or aberrant cell fate determination (61). Here, we show that Lamin A/C is indeed crucial to preserve MuSCs identity and regenerative capacity. We demonstrate that cell-autonomous Lamin A-dependent Polycomb dysfunction leads to MuSCs functional decline, which culminate with impaired regenerative capacity and dystrophic phenotype (Figure 1). Traditionally, the role of Lamin A/C in muscle differentiation has been considered to cause defects in muscle differentiation (62, 63). However, in other conditions, MuSCs from Lamin A/C null mice showed a normal ability to differentiate and to form myotubes (24, 64). By moving the viewpoint from differentiation to cell identity we now propose that in the absence of Lamin A/C, a portion of MuSCs derails from their fate affecting the quiescent MuSCs pool. At the molecular level, we described a mechanism of Lamin A dependent deregulation of PcG targets showing the spreading of repressive marks along the chromatin fiber (Figure 2), with lack of the

expected higher order structures and consequent de-repression of bivalent genes (Figure 3, 4 and 7). Recently it has been proposed that PcG domains can have distinct size and boundaries characteristics (65): upon differentiation, loci directly involved in fate specification lose PcG mediated looping interactions, allowing new active promoter/enhancer interactions. In parallel, other PcG domains, such as the Hox clusters, do not change their 3D architecture. Our findings further corroborate this hypothesis showing that the stability of PcG interacting domains correlates with PcG occupancy and depends on Lamin A (Figure 2). PcG dysfunctions drive *Lmna* Δ8-11 -/- MuSCs toward two fates not mutually exclusive (Figure 7): lack of cell identity, highlighted by the presence of MuSCs co-expressing muscle and adipogenic markers (Figure 4), and premature senescence, as shown by defects in asymmetric division and accumulation of Ph-P38 and γH2AX (Figure 5 and S8). These epigenetic alterations determine a progressive decline in MuSCs self-renewal that accompanies the muscular dystrophy progression (Figure 1), ultimately leading to a geriatric condition characterized by the expression of $p16^{INK4a}$ from Cdkn2a locus (40) (Figure 6B). Genetic ablation of Cdkn2a locus can recover some muscular dystrophy defects of the *Lmna* Δ8-11 -/- mouse (Fig. 6) thus supporting the hypothesis that dystrophic and aging muscles share a dysfunction of epigenetic mechanisms controlling cell cycle and fate decision of MuSCs. Our findings corroborate recent evidence on PcG dysfunction in human disease (21), showing that PcG alterations contribute to pathology progression and severity in EDMD. This will further stimulate future studies on the role of PcG proteins in the dynamics of stem cell niche, when embedded in a pathological environment.

Methods

Study approval

Mice were bred and maintained according with the standard facility procedure (San Raffaele hospital, DIBIT 1, Milan) and Santa Lucia Foundation (Rome) and all the experimental protocol were approved accordingly with Italian Ministry of Health (protocol n. 617/2015-PR). Heterozygous B6.129S1(Cg)-*Lmnatm1Stw*/BkknJ mice (*Lmna* Δ8-11 +/-)(23), Lmna H222P/H222P knock-in mice(22) and Cdkn2a +/- mice(58) were used.

Immunofluorescence

On muscle sections: *Tibialis Anterior* (TA) muscles were embedded in Killik (Bio-Optica, 05-9801), immediately frozen in pre-cooled isopentane (Sigma, 277258) and criosectioned at 8 μm thin. Sections were fixed 20 minutes in paraformaldehyde (PFA) 4% (Sigma, P6148) and washed 3x5 minutes in PBS1X (Euroclone, ECB4004L). To permeabilize tissues, pre-cooled methanol (Sigma, 322415) at -20°C was added for 6 minutes. Antigen retrieval was performed 2x5 minutes in hot citric acid 80°C (Sigma, C0759) pH6.0 and washed 2x5 minutes in PBS1X. Sections were blocked 1 h in BSA 4% (Sigma, A7030) followed by incubation 45 minutes with FAM mouse fragment 1:100 (Jackson Immuno Research, 115-007-003)/ PBS1X. Primary antibodies were diluted 1:100 in blocking solution, except for pparγ diluted 1:75, and incubated O/N at 4°C. The day after, sections were washed 3x10 minutes in PBS1X/0,1% BSA and incubated with secondary antibodies in blocking solution 1:200 1h at RT in the dark. Then, sections were washed 3x10 minutes in 0,1% BSA/PBS1X and incubated 2h at RT with Laminin and Pax7 (1:20). After washing 3x10 minutes in PBS1X/0,1% BSA sections were incubated 45 minutes with Cy5 (1:300) and Biotin (1:500) for Pax7 signal amplification. After washing 3x10 minutes in PBS1X, sections were incubated 45 minutes with secondary antibodies Cy3-streptavidin (1:1250). The sections were finally washed 3x10 minutes in PBS1X, stained 5 minutes with dapi (Sigma, D9542), briefly washed twice in PBS1X and mounted on slide with a drop of Prolong Antifade (Life P36930).

For multiple injury experiment, 20 μl of Cardiotoxin (CTX) 10 μM (Latoxan, L8102) were injected in TA muscle each week for 3 weeks. TA muscle was harvested one week after the last CTX injection.

On single myofibers: *Tibialis Anterior* (TA) and *Exstensor Digitorium Longus* (EDL) were isolated from mice and digested 75 minutes in 0,35% collagenase type I (Sigma, C0130)/DMEM (Gibco, 10569010) at 37°C. 3 rounds of myofibers washes

were performed in pre-coated dishes with 20% FBS (Gibco, 10500064)/DMEM. Finally the myofibers were let grown in DMEM supplemented with 20% FBS, 1% Chicken Embryo extract (Seralab, CE650-DL) and 1% Penicillin/Streptomycin (Euroclone, ECB 3001) for 48h or 96h changing the medium each 48h. Myofibers were collected in 2ml tubes pre-coated with 10% FBS/PBS1X and fixed 15 minutes with 4%PFA followed by 3 washes in PBS1X. Permeabilization was performed 8 minutes with 0,5% Triton X-100 (Sigma, 93443)/PBS1X followed by 2 washes in PBS1X. Myofibers were incubated 1h in blocking solution (10% of FBS/PBS1X). Primary antibodies was incubated in blocking solution ON at 4°C. The day after myofibers were washed in 0,1% FBS/PBS1X twice and incubated 60 minutes with secondary antibodies in blocking solution. Fibers were washed in PBS1X, incubate 5 min with dapi, briefly washed twice in PBS1X and mounted on slide with a drop of Prolong Antifade.

On satellite cells: In order to preserve the integrity of chromatin architecture, Muscular stem (Satellite) Cells (MuSCs) suspension was fixed in 1% PFA for 9 minutes and quenched with 125 mM Glycine (Sigma, 8898) before FACS sorting. SC cells were placed on pre- Poly-L-lisined coverslips (Sigma, P8920) at density of 100.000/mL for 30 minutes at RT. Coverslips were fixed with PFA at 4%/PBS for 10 minute at RT. Then, cells were washed in 2X5 minutes in 0,05% Triton X-100/PBS 1X, permeabilized with 0,5% Triton X-100/PBS1X for 10 minutes and rinsed in PBS 1X. The slides were let in 20% Glycerol (Sigma, G5516)/PBS 1X at least 60 minutes followed by 4 round of freeze and thaw: freezed on dry ice (30 seconds) and thawed in 20% Glycerol/PBS 1X at RT. Slides were washed 2X5 minutes in 0,05% Triton X-100 in PBS 1X, 1x5 minutes 0,5% Triton X-100/PBS1X, incubated in HCl 0,1M (Sigma, H1758) for 15 minutes and rinsed in PBS 1X. Aspecific signals were blocked with 1% BSA/PBS1X for 30 minutes at room temperature. Reaction with primary antibodies Ezh2 diluted 1:100 in blocking solution was performed 12–16 h at 4°C; Lamin A/C diluted 1:200 in blocking solution was performed the day after at room temperature for 2 hour. Secondary antibodies were diluted 1:200 in blocking solution for 1 h at room temperature. Washes were done in PBS1X. DNA was counterstained with dapi, and glasses were mounted in Prolong Antifade.

On iPS:

Coverslips with cells were fixed with PFA at 4%/PBS for 10 minute at RT. Then, cells were washed in 2X5 minutes in 0,05% Triton X-100/PBS 1X, permeabilized with 0,5% Triton X-100/PBS1X for 10 minutes and rinsed in PBS 1X. Reaction with primary antibodies diluted 1:10 in blocking solution was performed 12–16 h at 4°C. Secondary antibodies were diluted 1:100 in blocking solution for 1 h at room

temperature. Washes were done in PBS1X. DNA was counterstained with dapi, and glasses were mounted in Prolong Antifade.

Primary antibodies: MYOD, Santacruz, SC-760; phospho-p38, Cell Signaling, 4511; γ-H2aX, Cell Signalling, 2577; ki67, Abcam, ab 11580; PPARγ, Abcam, ab118521; LAMININ, Sigma, L9393; PAX7, Developmental Studies Hybridoma Bank; BIOTIN, Jackson Immuno Research, 115-065-205; EZH2, Cell Signaling Technology, 5246; LAMIN A/C, Santacruz, sc6215; OCT4, SOX2, TRA-1-60, SSEA4 (Thermofisher, A24881), MF20 (Developmental Studies Hybridoma Bank, MF 20-c).

Secondary antibodies: Alexa 488, Jackson Immuno Research, 711-545-152; Cy5 1:300, Jackson Immuno Research, 111-115-144; Cy3-streptavidin, Jackson Immuno Research, 016-160-084; Alexa 594, Jackson Immuno Research, 711-545-150.

Satellite cells extraction and FACS-sorting

Hind-limb muscle were isolated from sacrificed mice and digested 120 min in 2,4 U/ml of Dispase II (Roche, 04942078001), 2 ug/ml of Collagenase A (Roche, 1013586001), 0,2 mM CaCl2 (Sigma, C5670), 4 mM MgCl2 (Sigma, M8266), 10 ng/ml DNase I (Roche, 1014159001) in PBS1X (Euroclone, ECB4004L) at 37°C in a water bath. The sample were resuspended in HBSS (Gibco, 14025-050) supplemented with 0,1% BSA (Sigma, A7030). Cell suspension was serially filtered with 70 (Falcon, 352350) and 40 μm (Falcon, 352340), stained with 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-α7integrin 1:200 (AbLab, 67-001-05) and sorted with BD FACS ARIA III for: PB-CD45$^-$/ PB-CD31$^-$/ PB-Ter119$^-$/ FITC-Sca1$^-$/ APC-α7integrin$^+$.

Three-dimensional multicolor DNA FISH

To produce probes for 3D multicolor DNA FISH we used the following BAC DNA clones (BACPAC Resources Program, CHORI): CH29-101F16 (for Pparγ) and CH29-555O5 (for the upstream region of Pparγ). 1-3 μg of BAC DNA were labelled with dig-dUTP (Roche, 11093088910) (for the upstream region of Pparγ) or cy3-dUTP (Thermo Fisher Scientific, C11401) (for Pparγ) through nick translation in 50 μl of Labelling mix buffer (0.02 mM C-G-A dNTPs: Euroclone, EMR273025, EMR274025, EMR272025 respectively, 0.01 mM dTTP Euroclone EMR275025, 0.01 mM labelled dUTP, 50 mM Tris-HCl pH 7.8, 5 mM MgCl2, 10 mM b-mercaptoethanol, 10 ng/μl Bovine serum albumin (BSA), 0.05-0.1 U/μl DNA Polymerase I (Thermo Fisher Scientific, 18010-017), 0.004-0.001 U/μl Amplification

Grade DNase I (Sigma, D5307) for 30 min-2 h at 16 °C, to obtain an average probe size of 50 bp. Probes were collected by ethanol precipitation, resuspended in 10 mM Tris-HCl pH 7.5 and then quantified using a Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific). For a single experiment 100-300 ng of each probe was precipitated with 3.5 µg of Mouse Cot-1 DNA (Thermo Fisher Scientific, 18440-016) and 20 µg of Deoxyribonucleic acid, single stranded from salmon testes (Sigma, D7656), and then resuspended in 6 µl of Hybridization solution (50% formamide pH 7.0 (FA)/2X SSC/10% Dextran sulfate). Pre-fixed Satellite cells Cells were plated directly on poly-L-Lisyned coverslips and fixed with 4% Paraformaldehyde (PFA) in 1X PBS and TWEEN 20 0.1% (PBS-T) for 10 min at room temperature. During the last minute few drops of 0.5% Triton X-100 in 1X PBS (PBS) were added and then cells were washed with 0.01% Triton X-100 in PBS three times for 3 min at room temperature. Cells were first permeabilized with 0.5% Triton X-100 in PBS for 10 min at room temperature. In order to remove RNA, samples were treated with RNase Cocktail Enzyme Mix (Thermo Fisher Scientific, AM2288) for 1 h at 37 °C. Cells were subjected to other steps of permeabilization with 20% Glycerol in PBS overnight at room temperature, followed by four cycles of freeze and thaw interleaved by soak with 20% Glycerol in PBS. Permeabilized cells were washed with PBS three times for 10 min at room temperature. Cells were then incubated in 0.1 M HCl for 5 min at room temperature, followed by a rinse with 2X SSC and then incubated in 50% FA in 2X SSC for at least 30 min at room temperature. Slides were equilibrated in 2X SSC for 2 min, washed in PBS for 3 min and then treated with 0.0025-0.0075% pepsin in 0.01-0.03 N HCl for 2-4 min at room temperature to eliminate cytoskeleton. Pepsin was inactivated with 50 mM MgCl2 in PBS twice for 5 min. Nuclei were post-fixed with 1% PFA in PBS for 1 min, washed with PBS for 5 min and with 2X SSC twice, and then back to 50% FA in 2X SSC for at least 30 min at room temperature. Hybridization solution was loaded on a clean microscopic slide, coverslip with nuclei was turned upside down on the drop of hybridization mixture and sealed with rubber cement. Samples were denatured for 4 min at 75 °C and leaved to hybridize in a metallic box floating in a 37 °C water bath overnight. Samples were washed with 2X SSC three times for 5 min at 37 °C and with 0.1X SSC three times for 5 min at 60 °C, followed a rinse with 0.2% TWEEN 20 in 4X SSC. Aspecific binding sites were blocked with Blocking solution (4% BSA in 4X SSC, 0.2% TWEEN 20) for 20 min at 37 °C. Samples were then incubated in the appropriate concentration of Streptavidin, DyLight 488 Labeled Anti-Digoxigenin/Digoxin (Vector Laboratories, DI-7488) (1:100) diluted in Blocking solution for 35 min in a dark and wet chamber at 37 °C. Samples were washed with 0.2% TWEEN 20 in 4X SSC three times for 3 min at 37 °C,

equilibrated in PBS and post-fixed with 2% Formaldehyde in PBS for 10 min at room temperature. Finally, the 3D-fixed nuclei were washed with PBS three times for 5 min at room temperature, counterstained with 1 ng/µl DAPI in PBS for 10 min at room temperature and washed with PBS two times for 5 min at room temperature. Coverslips were mounted. An Eclipse Ti-E (Nikon Instruments) microscope was used to scan the nuclei, with an axial distance of between 0.2-0.25 micron consecutive sections.

Image analysis

Image analysis derived from(25) with some adaptation. The algorithm performs the 2D segmentation of cell nuclei and the detection of PcG bodies for each slice of the stack, followed by the 3D reconstruction and identification of nuclei and PcG bodies. It then measures the volume of nuclei and the number and volume of the PcG bodies. The algorithm has been implemented in MATLAB following this scheme:

```
[nuclei_n,avg_Ezh2,n] = nuclei_seg(I_dapi,lamin,n,I_Ezh2,n) %Performs 2D
nuclei segmentation of the middle section n of the stack
I_avg,n = imfilter(I_Ezh2,n,fspecial('average', [3,3]))
I_filt,n = I_Ezh2,n - I_avg,n
t_n = isodata_thresh(I_filt,n,avg_Ezh2,n) %Applies ISODATA method to
separate PcG bodies from nuclei regions in the middle section
n of the stack
for each slice n of the stack
      nuclei_n = nuclei_seg(I_dapi,lamin,n) %Performs 2D nuclei
segmentation
      I_avg,n = imfilter(I_Ezh2,n,fspecial('average', [3,3]))
      I_filt,n = I_Ezh2,n - I_avg,n
      pcg_n = I_filt,n > t_n %Performs thresholding operation for 2D
      PcG bodies detection
      pcg_vol(:,:,n) = pcg_n(:,:)
      nuclei_vol(:,:,n) = nuclei_n(:,:)
endfor
nuclei_CC = bwconncomp(nuclei_vol)
nuclei_L = labelmatrix(nuclei_CC)
compute volume for each nucleus object in nuclei_CC
exclude nuclei whose volume is less than 10% of mean volumes
{NCL}_M <- identified 3D nuclei
for each nucleus m in {NCL}_M
      NCL_m.PcG = pcg_vol .* {NCL}_M %3D positions of detected PcG
      bodies within the nucleus NCL_m
      NCL_m.PcG = bwareaopen(NCL_m.PcG,17,6);
      PcG_CC = bwconncomp(NCL_m.PcG)
      compute number of PcG bodies in PcG_CC
      compute volume for each PcG body in PcG_CC
endfor
```

$I_{dapi,lamin}$ is the image obtained by the sum of both images showing the fluorescence of nucleus and lamin.

The function *nuclei_seg* performs a partition of cell image $I_{dapi,lamin}$ in nuclei regions and background implementing a region based segmentation algorithm[1]. avg_{Ezh2} is the mean intensity value of the nuclei regions in the image I_{Ezh2} that shows the fluorescence of PcG bodies.

In order to better enhance PcG areas we subtract from the original I_{Ezh2} image its smoothed version obtained by applying an averaging filter of size 3, obtaining the image I_{filt}. The function *isodata_thresh* implements the ISODATA classification algorithm[2] and uses relevant values computed by *nuclei_seg* function in order to extract PcG bodies from the nuclei regions. It sets the initial threshold value of ISODATA method as avg_{Ezh2}. For each slice of the stack, the algorithm separates PcG bodies from nuclei regions by means of a thresholding operation using the threshold value estimated by the function *isodata_thresh* applied to the middle section. *pcg_vol* and *nuclei_vol* are 3D arrays that contain the positions of the detected PcG bodies and nuclei from all slices. 3D reconstructions of nuclei are obtained through the connected components algorithm (*bwconncomp* MATLAB function, using a connectivity of 6). 3D nuclei are then labeled by applying the *labelmatrix* MATLAB function so they can easily separated each from the others.
The algorithm computes the volume of each 3D reconstruction, discarding objects whose volume is less than 10% of mean volumes which are just noise. The algorithm uses the *bwareopen* function In order to discard too small detected PcG objects which are probably just noise. 3D reconstructions of PcG bodies are obtained through the connected components algorithm (*bwconncomp* MATLAB function, using a connectivity of 6). The algorithm computes the number of the PcG bodies and the volume of any PcG body.

Histone extraction and Western blot

Total proteins were prepared starting from quadriceps muscle. Muscles were homogenized on ice with tissue ruptor (Qiagen, 902756) in 1ml of extraction buffer (50 mM Tris HCl, pH 7.5, 150mM NaCl, 1 mM EDTA pH 8.0, 0,1% SDS, 1% Np40 (Sigma, 74385), 0,5% Sodium deoxycholate (Sigma, D6750) 1× protease Inhibitors (Roche, 04693132001) 2 mM PMSF (Sigma, 93482), 1 mM NaF (Sigma, s7920); 1 mM Na3VO4). The homogenized extract was let on ice for 30 minutes. Three pulses of 10 seconds sonication at 30% amplitude were performed with Branson digital sonifier 250 to allow dissociation of protein from chromatin and solubilization. The extract was centrifuged at 12000 rcf 15 min at 4°C and the supernatant was

collected. Extracts were analyzed by SDS-PAGE using a 8–10% gel (29:1 acrylamide solution Sigma, 01708). For histone extractions approximately 2.000.000 of freshly isolated MuSCs was collected and resuspended in 15uL of 0.2 N HCl with 1× protease Inhibitors (Roche, 04693132001) 2 mM PMSF (Sigma, 93482), 1 mM NaF (Sigma, s7920); 1 mM Na3VO4) and incubated overnight at 4°C with constant rocking. The supernatant was run on 4-12% bis-tris Acrylamide gel (Thermo Fisher Scientific, NW04125).

RNA-seq

Total RNA from freshly isolated satellite cells from 3 mice for each condition (2 mice for the het condition) was extracted in TRI-Reagent (Sigma, T9424) following the relative guidelines. Libraries were prepared using the Illumina TruSeq Stranded Total RNA with Ribo-Zero GOLD kit and were quality controlled with an Agilent Bioanalyzer at the Sequencing Facility of the Institute of Applied Genomics (IGA, Udine, Italy). 125 bp reads were produced using an Illumina HiSeq2500 machine in paired-end mode to reach a sequencing depth of about 25 million reads for each sample. 5 million reads were added to have a total of 30 million of sequenced reads. RTPCR were performed on total RNA extracted from single mouse. Primers sequences: gapdh: gtatgtcgtggagtctactgg, tcgtggttcacacccatcac; p16: gtgtgcatgacgtgcggg, cagttcgaatctgcaccgtag; p19: gctctggctttcgtgaacatg, tcgaatctgcaccgtagttgag; p21: acggaggcagaccagcct, acacagagtgagggctaagg.

ChIP-seq

In order to preserve the integrity of chromatin architecture, muscular cell suspension in DMEM high glucose (Euroclone, ECB7501L), 10% FBS (Thermo Fisher Scientific, 10270106) was fixed in 1% formaldehyde solution (5mM HepesKOH pH7.5, 10mM NaCl, 0.1mM EDTA, 0.05mM EGTA, 1.1% formaldehyde) for 9 minutes, then quenched with 125 mM Glycine for 5 minutes before fluorescence activated cell sorting (FACS). Isolated satellite cells were resuspended in SDS buffer (100 mM NaCl, 50mM TrisHCl pH8.1, 5 mM EDTA, 0.2% NaN_3, 0.5% SDS) and stored at -80°C (100 µL/10^6 cells). For ChIP analysis, satellite cells from pools of 6-9 mice were used. 2-2.5*10^6 of fixed cells/experiment were thawed on ice, resuspended in fresh SDS buffer and incubated at 4°C in mild agitation for 3 hours, passing them through a 0.50x16mm syringe needle every hour. Solution was then adjusted to IP Buffer composition (100 mM Tris pH 8.6, 0.3% SDS, 1.7% Triton X-100, 5 mM EDTA) and cells were sonicated with Branson Digital Sonifier to shear chromatin to 200 bp fragments. For spike-in experiments, a 5% of

sheared drosophila chromatin was added to each sample(27). The 2% of the total volume from each sample was taken as input chromatin. The remaining fragmented chromatin was incubated with 1 mM PMSF (Sigma, 93482) and 4 µg of the antibody of interest on a rotating wheel at 4°C overnight. Primary antibodies used were: H3K27me3 (Millipore 07-449) and H3K4me3 (Millipore 07-473). The next day, protein G beads (Life Technology, 1004D) were added (80 µL) and samples were incubated for additional 2 hours on the rotating wheel at 4°C. The beads were washed with Low Salt solution (150 mM NaCl, 20 mM TrisHCl pH 8.0, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), High Salt solution (500 mM NaCl, 20 mM TrisHCl pH 8.0, 2 mM EDTA, 0.1% SDS, 1% Triton X-100), Low Salt solution and then TE NaCl (50 mM NaCl, 10 mM TrisHCl pH8.0, 1 mM EDTA). Crosslinking was reversed by incubating the beads at 65°C overnight in Elution buffer (50 mM TrisHCl pH 8.0, 10 mM EDTA, 1% SDS). Input chromatin was also decrosslinked in Elution buffer overnight at 65°C. The next day, all samples were diluted with one volume of TE 10:1, treated with 0.2 ug/mL RNase A (Sigma R6513) for 2 hours at 37°C and then with 0.2ug/mL Proteinase K (Sigma P2308) for 2 hours at 55°C. DNA was isolated through standard phenol/chloroform extraction, followed by precipitation and resuspension in 31 µL of 10 mM Tris-HCl pH 8.0. Before library construction, ChIP were validated using the primers: hoxd9 ggataatcgcctaggtgtgactt, catctcttcttgcctctctggg and pax7prom gcgacccccctgaggaaaa, cgaaaagaagtctccaacgagtatt. Libraries for IP and reference input DNA were created for each sample using the automation instrument Biomek FX (Beckman Coulter), then qualitatively and quantitatively checked using Agilent High Sensitivity DNA Kit (Agilent Technologies, 5067-4627) on a Bioanalyzer 2100 (Agilent Technologies). Libraries with distinct adapter indexes were multiplexed and, after cluster generation on FlowCell, were sequenced for 50 bases in the single read mode on a HiSeq 2000 sequencer at the IEO Genomic Unit in Milan.

TAD information

The HiC matrix, showing the TAD information in the region chr6:114,760,000-116,080,000 (UCSC coordinates style) has been downloaded form the “3D Genome Browser website” (http://promoter.bx.psu.edu/hi-c/). The query has been inserted in the “HiC” section with the following parameters: “Hi-C” as “Dataset Type” (step 1), “mouse” as Species and “mm10” as “assembly” (step 2), the checkbox “Browse Available Hi-C Data” has been flagged and “mESC” as “Tissue” (step 3), “Bonev_2017-raw” as “Type” and “10kb” as “Resolution” have been chosen, finally

“chr6” as “Chromosome”, “114760000” as “Start” and “116080000” as “End” have been chosen in the box “Option 2: Search by Location” (step 4).

Bioinformatics analysis

RNA-seq: the overall quality of the sequenced reads was assessed using FastQC tool (version 0.11.2; http:/www.bioinformatics.babraham.ac.uk/projects/fastqc), then reads were trimmed and the adapters removed using Trimmomatic software (version 0.33;(66)). The remaining reads were aligned to mouse genome (version GRCm38; Ensembl release 87) providing the murine transcriptome (Ensembl release 87) with Spliced Transcripts Alignment to a Reference (STAR; version 2.5.2b;(67)). STAR alignments were used as input for HTSeq (version 0.6.1;(68)) to quantify the amount of reads per gene in the annotation with -s reverse and -t exon parameters. The differentially expressed genes (DEGs) between Lmna homozigous and wild type mice were identified using edgeR (version 3.16.0;(69) in R environment version 3.3.2); the significance threshold was set to False Discovery Rate (FDR) ≤ 0.05. Raw expression values normalization was computed calculating $\log_2$ count per million (cpm) with the same tool. Z-scores of hom vs. wt DEGs were calculated and visualized using *gplots* R package (version 3.0.1 in R environment version 3.3.2). A functional enrichment on DEGs between Lmna homozygous and wild type mice was conducted using the Gene Set Enrichment Analysis (GSEA) software(70) and the mouse Gene Ontology (GO) terms of the *“Biological Processes”* section retrieved from the R package *“gskb”* (version 1.6.1 in R environment 3.3.2; (71)). GSEA tool was used in pre-ranked mode as suggested for RNA-seq experiments using as input the log fold change (log FC) for all the genes in the annotation and the *“classic”* metric for gene ranking. The significance threshold was set to FDR ≤ 0.05. To compare our results with public data a GSEA analysis was performed on Liu et al., 2013 dataset(51), Sousa-Victor et al., 2014 dataset(52) and Juan et al., 2011 dataset(26). The first was retrieved from Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/; GSE47177), the second one from supplementary data, whereas the latter was obtained from the authors in form of raw data. On Liu et al., 2013: gene expression level of quiescent satellite cells from old and young mice was calculated starting from raw data using affy and annotate R packages (gene-level summarization using Brainarray Entrez gene custom cdf for mouse gene 1.0 st microarrays;
http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/CDF_download.asp) and was used as expression dataset, up-regulated genes in homozygous mice filtered for log FC ≥ 1 were added to Biocarta mouse pathways retrieved from

gskb R package and used as gene sets database; moreover the old vs young class comparison was used as phenotype labels and *Signal2Noise* as metric for gene ranking as defined in(70). On Sousa-Victor et al., 2014: differential expressed genes from old vs. young mice and geriatric vs. young mice were filtered for q-value ≤ 0.05 and added to Biocarta mouse pathways retrieved from gskb R package and used as gene sets database and hom vs. wt log FC of all genes for a pre-ranked analysis with classic as metric for gene ranking. For Juan et al., 2011: raw data were analyzed using affy and annotate R packages (gene-level summarization using Brainarray Entrez gene custom cdf for mouse 430.2 microarrays; http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/CDF_download.asp); log FC were calculated for all genes to perform a pre-ranked analysis with classic as metric for gene ranking; up-regulated genes in homozygous mice filtered for log FC ≥ 1 were added to Biocarta mouse pathways retrieved from gskb R package and used as gene sets database. To correlate H3K27me3 changes with gene expression, genes were divided into 5 groups based on the expression level in wt condition. In this case we quantified gene expression using kallisto (version 0.44.0; (72)). Genes with 0 coverage, i.e. 0 transcript per million (tpm) in all of the 3 replicates were classified as *"zero-expressed"* (i.e. *"0"* category). This group includes genes not expressed and not detected for technical reasons, i.e. either low abundance resulting in technical dropout, or low coverage due to low mappability of the sequence. Remaining genes were divided into 4 equally sized groups (i.e. *"I"*, *"II"*, *"III"* and *"IV"* categories) based on quartiles of the mean expression values in the 3 replicates. RNA-seq genome coverage was determined using *bamCoverage* function of *deepTools* suite (version 2.5.1;(73)) with 1 base bin size and effective genome size normalization (excluding X and Y chromosomes) and visualized using *pyGenomeTracks* (version 2.0;(73)). To refine the functional analysis on DEGs, an additional enrichment of GO terms was performed using the *"FGNet"* R package (version 3.8;(74)) followed by a semantic similarity analysis on significantly enriched terms given and FDR threshold of 0.05 taking advantage of the *"GOSemSim"* R package (version 2.0.4;(75)). Then the macro-categories grouping GO terms taking part of the same biological process were identified using Revigo web tool(76). Similarity scores heatmap was visualized using *gplots* R package (version 3.0.1 in R environment version 3.3.2). To test the enrichment of up-regulated genes among those with a bivalent promoter and among H3K27me3 target genes a Fisher exact test was performed using *fisher.test* function of *stats* R package (R environment version 3.3.2; the p-value for significant threshold was set to 0.05) considering only mappable genes as reference background (i.e. genes with at least one reported

alignment in one of the RNA-seq/ChIP-seq samples). Moreover, the same procedure was used to test the enrichment of Pparγ-related up-regulated genes among those with a bivalent promoter selecting up-regulated genes in significantly enriched adipogenesis-related GO terms. The expected number of genes was determined by means of the product of marginal values method and the $\log_2$ ratio of observed over expected number of genes was computed and visualized using *gplots* R package (version 3.0.1 in R environment version 3.3.2).

ChIP-seq: FastQC (version 0.11.3; http:/www.bioinformatics.babraham.ac.uk/projects/fastqc) and Trimmomatic (version 0.32;(66)) were respectively used for the quality control of the sequenced reads and trimming/adapter removal steps. Reads were aligned to mouse genome (GRCm38; Ensembl release 87) using Bowtie (version 1.0.1;(77)) and retaining only those reads that map to unique loci. PCR duplicates were removed using samtools (version 0.1.19;(78)). Peak calling step was performed with Model-based Analysis of ChIP-Seq 2 (MACS2; version 2.1.1;(79)) using input samples as control and *--broad* parameter for H3K27me3 ChIP-seq samples. Signal tracks were constructed calculating the linear fold enrichment (IP/input) with the same tool. Annotation of called peaks was performed with *ChIPpeakAnno* R package (version 3.8.9 in R environment version 3.3.2;(80)) in *both* mode (nearest start plus overlapping annotation) and calculating the distance between the middle point of peaks and the transcription start site (TSS) of genes. Genes with bivalent promoter were identified as those genes having an H3K27me3 peak that overlap for at least the 5% an H3K4me3 peak in a window between -0.5kb and +2.0kb around the TSS (similarly to what originally proposed by (81)) in wt condition (*bedtools* version 2.24.0;(82)). For ChIP-seq with spike-in approach reads were aligned also to Drosophila melanogaster genome (BDGP6; Ensembl release 87) and filtered as those mapping to mouse genome; then ambiguous reads aligning to both genomes were removed using *BBMap* tool (BBMap - Bushnell B. - sourceforge.net/projects/bbmap/). Reads mapping to Dm6 genome were used to calculate spike-in normalization factors for IP samples as in(27) which were employed as scaling factors to normalize signal tracks (MACS2 version 2.1.1). H3K27me3 target genes were identified as those genes with a H3K27me3 peak in a window of +/- 5 kb around the TSS in the wt condition. We defined intergenic and interpeaks regions as those outside H3K27me3 peaks and outside annotated genes (*bedtools* version 2.24.0).

Average IP/input ChIP-seq profiles on the indicated lists of genes were calculated and visualized using *computeMatrix* and *plotProfile* functions of *deepTools* suite (version 2.5.1; (73)). Heatmap of average IP/input ChIP-seq signal on indicated lists

of genes were calculated and visualized using *computeMatrix* and *plotHeatmap* functions of *deepTools* suite (version 2.5.1; (73)). Both average profiles and average heatmaps were performed after excluding genes with length less than 2kb and genes overlapping each other without respect of the strand in the window of visualization. Signal tracks on specific loci were visualized using *pyGenomeTracks* tool (version 2.0;(73)).

iPSCs generation, characterization and differentiation

To induce reprogramming, human adult muscular myoblasts from patient affected by EDMD and from healthy donor were exposed to SENDAI virus (Thermofisher, A16518) carrying the embryonic genes Oct4, Sox2, Klf4, cMyc under defined-media conditions completely free of non-human, animal materials (Thermofisher, A1517001) and expanded on Geltrex matrix (Thermofisher, A1569601). From each reprogrammed sample, 3 iPSC lines were expanded and characterized by immunostaining for the expression of OCT4, SOX2, TRA-1-60, SSEA4 and quantitative RT-PCR for *OCT4, SOX2, NANOG, LIN28.* Primers sequences: hgadph: agaggcagggatgatgttct, tctgctgatgcccccatgtt; hoct4: cctgaagcagaagaggatcacc, aaagcggcagatggtcgtttgg; hsox2: agctacagcatgatgcagga, ggtcatggagttgtactgca; hnanog: cccaaaggcaaacaacccacttct, agctgggtggaagagaacacagtt; hlin28: agtaagctgcacatggaagg, attgtggctcaattctgtgc; hpax7: aagaaggccaaacacagcatcgac, aggtcaggttccgactccacat; hmhc: ttcattggggtcttggacat, aacgtccactcaatgccttc; hmyog: aatgcagctctcacagcgcctc, tcagccgtgagcagatgatcc; hn/cam: atggaaactctattaaagtgaacctg, tagacctcatactcagcattccagt; hpparg2; gctgttatgggtgaaactctg,ataaggtggagatgcaggctc; hc/ebp: gcaagagccgcgacaag, ggctcgggcagctgctt. iPSC clones were sequenced for the EDMD mutation p.r453w (c.1357c>t), caused by the substitution of a cytosine with a thymine in position 1357 of the coding sequence. Primers sequences: lamina: ttctcacagcacgcacgca, tgccatccttctgtcttgcc.

iPSC Myogenic differentiation: to differentiate iPSC into the myotube-like cells (MTs), we employed a combination of GSK3 inhibitor (CHIR99021) (Sigma-Aldrich, SML 1046) and freshly isolated EVs derived from differentiated myotube cultures (MT-derived EV). iPSC were induced toward mesodermal commitment in Essential 6 medium (Thermofisher, A15165-01) and ITS (Thermofisher, 41400-045) supplemented with 10uM CHIR. After 2 days, medium was replaced with Essential 6 medium enriched with ITS, bFGF (Thermofisher, 13256-029), IGF (ThermoFisher Scientific, PHG0078) and 50ug/mL MT-derived EV. Between day 8 and 10, cells were expanded on Collagen Type I matrix. In the final step of differentiation all the

growth factors were removed. To verify the completed differentiation process immunostaining for MF20 was performed.

Data availability

The datasets generated during the current study are available in the GEO repository with accession number GSE123725.

Statistics

All the data are represented as mean with S.E.M. using Graph Pad prism 6. The sample size (n) is described for each experiment in the relative figure legend. Multiple comparison between three o more groups were made using one way Anova or two way Anova with significance as P value <0,05 (*), P value < 0,01 (**), P value <0,001 (***).

Author Contributions

C.L. conceived the project and designed experiments. A.B. and C.M. performed experiments and analysed data. F. L. and D.P. performed and interpreted ChIP-seq experiments. S.V. performed bioinformatic analysis. G.P. performed experiments in mice. A.C. and B.B. performed and interpreted FISH experiments. L.A., F.G. and G.O. quantified immunofluorescence and FISH images. M.D.B. performed FACS. R.R. and C.B. generated and analysed iPS. F.F. and C.P. analysed ChIPseq data. C.L. supervised the project and analysed data. C.L. and A.B. wrote the manuscript. All authors edited and approved the manuscript.

Acknowledgements

We thank Maria Vivo, Gisèle Bonne, Federica Marasca, Pierluigi Manti, Davide Gabellini, Daniela Palacios, Giovanna Lattanzi, the Italian network of Laminopathies and members of the laboratory for stimulating discussions and constructive criticisms. We gratefully acknowledge Gisèle Bonne and Nicolas Vignier for providing the mouse H222P. We thank Dr Peter Meinke for providing primary human EDMD myoblasts. We are grateful to Chiara Cordiglieri and INGM

Imaging Facility for assistance during images acquisition and to Mariacristina Crosti, Monica Moro and INGM FACS Facility for assistance in cell sorting. This work was supported by grants from the Italian Minister of Health n. GR-2013-02355413 to C.L., My First AIRC Grant (MFAG) n. 18535 to C.L., AFM-Telethon n. 21030 to C.L. and F.F. and Cariplo 2017-0649 to C.L. and F.F.

References

1. Shimi T, Kittisopikul M, Tran J, Goldman AE, Adam SA, Zheng Y, Jaqaman K, and Goldman RD. Structural organization of nuclear lamins A, C, B1, and B2 revealed by superresolution microscopy. *Mol Biol Cell.* 2015;26(22):4075-86.
2. Turgay Y, Eibauer M, Goldman AE, Shimi T, Khayat M, Ben-Harush K, Dubrovsky-Gaupp A, Sapra KT, Goldman RD, and Medalia O. The molecular architecture of lamins in somatic cells. *Nature.* 2017;543(7644):261-4.
3. van Steensel B, and Belmont AS. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell.* 2017;169(5):780-91.
4. Zaremba-Czogalla M, Dubinska-Magiera M, and Rzepecki R. Laminopathies: the molecular background of the disease and the prospects for its treatment. *Cell Mol Biol Lett.* 2011;16(1):114-48.
5. Kolb T, Maass K, Hergt M, Aebi U, and 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.* 2011;2(5):425-33.
6. Serebryannyy L, and Misteli T. Protein sequestration at the nuclear periphery as a potential regulatory mechanism in premature aging. *J Cell Biol.* 2018;217(1):21-37.
7. Dubinska-Magiera M, Zaremba-Czogalla M, and Rzepecki R. Muscle development, regeneration and laminopathies: how lamins or lamina-associated proteins can contribute to muscle development, regeneration and disease. *Cell Mol Life Sci.* 2013;70(15):2713-41.
8. Schuettengruber B, Bourbon HM, Di Croce L, and Cavalli G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell.* 2017;171(1):34-57.
9. Briand N, and Collas P. Laminopathy-causing lamin A mutations reconfigure lamina-associated domains and local spatial chromatin conformation. *Nucleus.* 2018;9(1):216-26.
10. Cesarini E, Mozzetta C, Marullo F, Gregoretti F, Gargiulo A, Columbaro M, Cortesi A, Antonelli L, Di Pelino S, Squarzoni S, et al. Lamin A/C sustains PcG protein architecture, maintaining transcriptional repression at target genes. *J Cell Biol.* 2015;211(3):533-51.
11. Marullo F, Cesarini E, Antonelli L, Gregoretti F, Oliva G, and Lanzuolo C. Nucleoplasmic Lamin A/C and Polycomb group of proteins: an evolutionarily conserved interplay. *Nucleus.* 2016:0.
12. Oldenburg A, Briand N, Sorensen AL, Cahyani I, Shah A, Moskaug JO, and Collas P. A lipodystrophy-causing lamin A mutant alters conformation and epigenetic regulation of the anti-adipogenic MIR335 locus. *J Cell Biol.* 2017;216(9):2731-43.
13. Zheng X, Hu J, Yue S, Kristiani L, Kim M, Sauria M, Taylor J, Kim Y, and Zheng Y. Lamins Organize the Global Three-Dimensional Genome from the Nuclear Periphery. *Mol Cell.* 2018;71(5):802-15 e7.
14. Chiacchiera F, Rossi A, Jammula S, Piunti A, Scelfo A, Ordonez-Moran P, Huelsken J, Koseki H, and Pasini D. Polycomb Complex PRC1 Preserves Intestinal Stem Cell Identity by Sustaining Wnt/beta-Catenin Transcriptional Activity. *Cell Stem Cell.* 2016;18(1):91-103.
15. Lapthanasupkul P, Feng J, Mantesso A, Takada-Horisawa Y, Vidal M, Koseki H, Wang L, An Z, Miletich I, and Sharpe PT. Ring1a/b polycomb proteins regulate the mesenchymal stem cell niche in continuously growing incisors. *Dev Biol.* 2012;367(2):140-53.
16. Lopez-Arribillaga E, Rodilla V, Pellegrinet L, Guiu J, Iglesias M, Roman AC, Gutarra S, Gonzalez S, Munoz-Canoves P, Fernandez-Salguero P, et al. Bmi1 regulates murine intestinal stem cell proliferation and self-renewal downstream of Notch. *Development.* 2015;142(1):41-50.
17. Singarapu N, Ma K, Reeh KAG, Shen J, Lancaster JN, Yi S, Xie H, Orkin SH, Manley NR, Ehrlich LIR, et al. Polycomb Repressive Complex 2 is essential for development and maintenance of a functional TEC compartment. *Sci Rep.* 2018;8(1):14335.
18. Aloia L, Di Stefano B, and Di Croce L. Polycomb complexes in stem cells and embryonic development. *Development.* 2013;140(12):2525-34.
19. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell.* 2006;125(2):315-26.

20. Morata G, and Herrera SC. Cell reprogramming during regeneration in Drosophila: transgression of compartment boundaries. *Curr Opin Genet Dev.* 2016;40(11-6.
21. Lu TT, Heyne S, Dror E, Casas E, Leonhardt L, Boenke T, Yang CH, Sagar, Arrigoni L, Dalgaard K, et al. The Polycomb-Dependent Epigenome Controls beta Cell Dysfunction, Dedifferentiation, and Diabetes. *Cell Metab.* 2018;27(6):1294-308 e7.
22. Arimura T, Helbling-Leclerc A, Massart C, Varnous S, Niel F, Lacene E, Fromes Y, Toussaint M, Mura AM, Keller DI, et al. Mouse model carrying H222P-Lmna mutation develops muscular dystrophy and dilated cardiomyopathy similar to human striated muscle laminopathies. *Hum Mol Genet.* 2005;14(1):155-69.
23. Sullivan T, Escalante-Alcalde D, Bhatt H, Anver M, Bhat N, Nagashima K, Stewart CL, and Burke B. Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy. *J Cell Biol.* 1999;147(5):913-20.
24. Cohen TV, Gnocchi VF, Cohen JE, Phadke A, Liu H, Ellis JA, Foisner R, Stewart CL, Zammit PS, and Partridge TA. Defective skeletal muscle growth in lamin A/C-deficient mice is rescued by loss of Lap2alpha. *Hum Mol Genet.* 2013;22(14):2852-69.
25. Gregoretti F, Cesarini E, Lanzuolo C, Oliva G, and Antonelli L. An Automatic Segmentation Method Combining an Active Contour Model and a Classification Technique for Detecting Polycomb-group Proteinsin High-Throughput Microscopy Images. *Methods Mol Biol.* 2016;1480(181-97.
26. Juan AH, Derfoul A, Feng X, Ryall JG, Dell'Orso S, Pasut A, Zare H, Simone JM, Rudnicki MA, and Sartorelli V. Polycomb EZH2 controls self-renewal and safeguards the transcriptional identity of skeletal muscle stem cells. *Genes Dev.* 2011;25(8):789-94.
27. Orlando DA, Chen MW, Brown VE, Solanki S, Choi YJ, Olson ER, Fritz CC, Bradner JE, and Guenther MG. Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep.* 2014;9(3):1163-70.
28. Minoux M, Holwerda S, Vitobello A, Kitazawa T, Kohler H, Stadler MB, and Rijli FM. Gene bivalency at Polycomb domains regulates cranial neural crest positional identity. *Science.* 2017;355(6332).
29. Bernstein E, Duncan EM, Masui O, Gil J, Heard E, and Allis CD. Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol.* 2006;26(7):2560-9.
30. Thalheim T, Herberg M, Loeffler M, and Galle J. The Regulatory Capacity of Bivalent Genes-A Theoretical Approach. *Int J Mol Sci.* 2017;18(5).
31. Barak Y, Nelson MC, Ong ES, Jones YZ, Ruiz-Lozano P, Chien KR, Koder A, and Evans RM. PPAR gamma is required for placental, cardiac, and adipose tissue development. *Mol Cell.* 1999;4(4):585-95.
32. Rosen ED, Sarraf P, Troy AE, Bradwin G, Moore K, Milstone DS, Spiegelman BM, and Mortensen RM. PPAR gamma is required for the differentiation of adipose tissue in vivo and in vitro. *Mol Cell.* 1999;4(4):611-7.
33. Bantignies F, Roure V, Comet I, Leblanc B, Schuettengruber B, Bonnet J, Tixier V, Mas A, and Cavalli G. Polycomb-Dependent Regulatory Contacts between Distant Hox Loci in Drosophila. *Cell.* 2011;144(2):214-26.
34. Lanzuolo C, Roure V, Dekker J, Bantignies F, and Orlando V. Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat Cell Biol.* 2007;9(10):1167-74.
35. Eagen KP. Principles of Chromosome Architecture Revealed by Hi-C. *Trends Biochem Sci.* 2018;43(6):469-78.
36. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, and Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature.* 2012;485(7398):376-80.
37. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot JP, Tanay A, et al. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell.* 2017;171(3):557-72 e24.
38. Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, Li D, Choudhary MNK, Li Y, Hu M, et al. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.* 2018;19(1):151.
39. Chakkalakal JV, Jones KM, Basson MA, and Brack AS. The aged niche disrupts muscle stem cell quiescence. *Nature.* 2012;490(7420):355-60.

40. Sousa-Victor P, Perdiguero E, and Munoz-Canoves P. Geroconversion of aged muscle stem cells under regenerative pressure. *Cell Cycle.* 2014;13(20):3183-90.
41. Gonzalo S, Kreienkamp R, and Askjaer P. Hutchinson-Gilford Progeria Syndrome: A premature aging disease caused by LMNA gene mutations. *Ageing Res Rev.* 2017;33(18-29.
42. Jacobs JJ, Kieboom K, Marino S, DePinho RA, and van Lohuizen M. The oncogene and Polycomb-group gene bmi-1 regulates cell proliferation and senescence through the ink4a locus. *Nature.* 1999;397(6715):164-8.
43. Chang NC, Chevalier FP, and Rudnicki MA. Satellite Cells in Muscular Dystrophy - Lost in Polarity. *Trends Mol Med.* 2016;22(6):479-96.
44. Bernet JD, Doles JD, Hall JK, Kelly Tanaka K, Carter TA, and Olwin BB. p38 MAPK signaling underlies a cell-autonomous loss of stem cell self-renewal in skeletal muscle of aged mice. *Nat Med.* 2014;20(3):265-71.
45. Cosgrove BD, Gilbert PM, Porpiglia E, Mourkioti F, Lee SP, Corbel SY, Llewellyn ME, Delp SL, and Blau HM. Rejuvenation of the muscle stem cell population restores strength to injured aged muscles. *Nat Med.* 2014;20(3):255-64.
46. Hausburg MA, Doles JD, Clement SL, Cadwallader AB, Hall MN, Blackshear PJ, Lykke-Andersen J, and Olwin BB. Post-transcriptional regulation of satellite cell quiescence by TTP-mediated mRNA decay. *eLife.* 2015;4(e03390.
47. Troy A, Cadwallader AB, Fedorov Y, Tyner K, Tanaka KK, and Olwin BB. Coordination of satellite cell activation and self-renewal by Par-complex-dependent asymmetric activation of p38alpha/beta MAPK. *Cell Stem Cell.* 2012;11(4):541-53.
48. Ono Y, Urata Y, Goto S, Nakagawa S, Humbert PO, Li TS, and Zammit PS. Muscle stem cell fate is controlled by the cell-polarity protein Scrib. *Cell Rep.* 2015;10(7):1135-48.
49. Zammit PS, Golding JP, Nagata Y, Hudon V, Partridge TA, and Beauchamp JR. Muscle satellite cells adopt divergent fates: a mechanism for self-renewal? *J Cell Biol.* 2004;166(3):347-57.
50. Kuang S, Kuroda K, Le Grand F, and Rudnicki MA. Asymmetric self-renewal and commitment of satellite stem cells in muscle. *Cell.* 2007;129(5):999-1010.
51. Liu L, Cheung TH, Charville GW, Hurgo BM, Leavitt T, Shih J, Brunet A, and Rando TA. Chromatin modifications as determinants of muscle stem cell quiescence and chronological aging. *Cell Rep.* 2013;4(1):189-204.
52. Sousa-Victor P, Gutarra S, Garcia-Prat L, Rodriguez-Ubreva J, Ortet L, Ruiz-Bonilla V, Jardi M, Ballestar E, Gonzalez S, Serrano AL, et al. Geriatric muscle stem cells switch reversible quiescence into senescence. *Nature.* 2014;506(7488):316-21.
53. Beguelin W, Rivas MA, Calvo Fernandez MT, Teater M, Purwada A, Redmond D, Shen H, Challman MF, Elemento O, Singh A, et al. EZH2 enables germinal centre formation through epigenetic silencing of CDKN1A and an Rb-E2F1 feedback loop. *Nat Commun.* 2017;8(1):877.
54. Yosef R, Pilpel N, Papismadov N, Gal H, Ovadya Y, Vadai E, Miller S, Porat Z, Ben-Dor S, and Krizhanovsky V. p21 maintains senescent cell viability under persistent DNA damage response by restraining JNK and caspase signaling. *EMBO J.* 2017;36(15):2280-95.
55. Kandert S, Wehnert M, Muller CR, Buendia B, and Dabauvalle MC. Impaired nuclear functions lead to increased senescence and inefficient differentiation in human myoblasts with a dominant p.R545C mutation in the LMNA gene. *Eur J Cell Biol.* 2009;88(10):593-608.
56. Quelle DE, Ashmun RA, Hannon GJ, Rehberger PA, Trono D, Richter KH, Walker C, Beach D, Sherr CJ, and Serrano M. Cloning and characterization of murine p16INK4a and p15INK4b genes. *Oncogene.* 1995;11(4):635-45.
57. Ito T, Teo YV, Evans SA, Neretti N, and Sedivy JM. Regulation of Cellular Senescence by Polycomb Chromatin Modifiers through Distinct DNA Damage- and Histone Methylation-Dependent Pathways. *Cell Rep.* 2018;22(13):3480-92.
58. Serrano M, Lee H, Chin L, Cordon-Cardo C, Beach D, and DePinho RA. Role of the INK4a locus in tumor suppression and cell mortality. *Cell.* 1996;85(1):27-37.
59. Gnocchi VF, Ellis JA, and Zammit PS. Does satellite cell dysfunction contribute to disease progression in Emery-Dreifuss muscular dystrophy? *Biochem Soc Trans.* 2008;36(Pt 6):1344-9.

60. Robinson DCL, and Dilworth FJ. Epigenetic Regulation of Adult Myogenesis. *Curr Top Dev Biol.* 2018;126(235-84.
61. Tosic M, Allen A, Willmann D, Lepper C, Kim J, Duteil D, and Schule R. Lsd1 regulates skeletal muscle regeneration and directs the fate of satellite cells. *Nat Commun.* 2018;9(1):366.
62. Frock RL, Kudlow BA, Evans AM, Jameson SA, Hauschka SD, and Kennedy BK. Lamin A/C and emerin are critical for skeletal muscle satellite cell differentiation. *Genes Dev.* 2006;20(4):486-500.
63. Solovei I, Wang AS, Thanisch K, Schmidt CS, Krebs S, Zwerger M, Cohen TV, Devys D, Foisner R, Peichl L, et al. LBR and lamin A/C sequentially tether peripheral heterochromatin and inversely regulate differentiation. *Cell.* 2013;152(3):584-98.
64. Melcon G, Kozlov S, Cutler DA, Sullivan T, Hernandez L, Zhao P, Mitchell S, Nader G, Bakay M, Rottman JN, et al. Loss of emerin at the nuclear envelope disrupts the Rb1/E2F and MyoD pathways during muscle regeneration. *Hum Mol Genet.* 2006;15(4):637-51.
65. Kundu S, Ji F, Sunwoo H, Jain G, Lee JT, Sadreyev RI, Dekker J, and Kingston RE. Polycomb Repressive Complex 1 Generates Discrete Compacted Domains that Change during Differentiation. *Mol Cell.* 2017;65(3):432-46 e5.
66. Bolger AM, Lohse M, and Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-20.
67. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29(1):15-21.
68. Anders S, Pyl PT, and Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31(2):166-9.
69. Robinson MD, McCarthy DJ, and Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139-40.
70. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545-50.
71. Lai Liming HJ, Bares Valerie, Son Eun Woo, Ban Yuguang, Wang Wei, Qi Jianli, Jiang Gaixin, Liberzon Arthur, Ge Steven. GSKB: A gene set database for pathway analysis in mouse *bioRxiv*. 2016.
72. Bray NL, Pimentel H, Melsted P, and Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 2016;34(5):525-7.
73. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, and Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 2016;44(W1):W160-5.
74. Aibar S, Fontanillo C, Droste C, and De Las Rivas J. Functional Gene Networks: R/Bioc package to generate and analyse gene networks derived from functional enrichment and clustering. *Bioinformatics.* 2015;31(10):1686-8.
75. Yu G, Li F, Qin Y, Bo X, Wu Y, and Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics.* 2010;26(7):976-8.
76. Supek F, Bosnjak M, Skunca N, and Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One.* 2011;6(7):e21800.
77. Langmead B, Trapnell C, Pop M, and Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10(3):R25.
78. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078-9.
79. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):R137.
80. Zhu LJ, Gazin C, Lawson ND, Pages H, Lin SM, Lapointe DS, and Green MR. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics.* 2010;11(237.

81. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, et al. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* 2008;4(10):e1000242.
82. Quinlan AR, and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841-2.
83. Jahn D, Schramm S, Schnolzer M, Heilmann CJ, de Koster CG, Schutz W, Benavente R, and Alsheimer M. A truncated lamin A in the Lmna -/- mouse line: implications for the understanding of laminopathies. *Nucleus.* 2012;3(5):463-74.

Figure Legends

Figure 1. Lamin A regulates MuSC self-renewal.

(**A**) Immunohistochemical staining in *Lmna* Δ8-11 mice of PAX7 (red) and MYOD (green) markers at the indicated days of post-natal growth (P10-P19). Basement membrane of muscle fibers was stained with Laminin and nuclei with dapi. Activated, ASCs (PAX7+/MYOD+) and self-renewing, QSCs (PAX7+/MYOD-) MuSCs are shown. Scale bars, 50 μm. (**B**) Quantification of MuSCs pool composition in (A); n =3-6 animals per genotype (**C**) Quantification of myofibre size during post-natal growth, evaluated by the Cross Sectional area (CSA). n> 350 muscle fibers, n =3-4 animals per group. (**D**) Immunohistochemical staining of single myofibers extracted from *Lmna* Δ8-11 mice at P19 and cultured 96h. Activated, ASCs (PAX7+/MYOD+), self-renewing, QSCs (PAX7+/MYOD-) and differentiating cells (PAX7-/MYOD+) are shown. Nuclei were stained with dapi. Scale bars, 20 μm. (**E**) Quantification of MuSCs pool composition in (D); n= 4-5 animals per group. (**F**) Immunohistochemical staining in adult, injured *Lmna* Δ8-11 mice of PAX7 (red) and MYOD (green) markers at the indicated days of post-natal growth. Basement membrane of muscle fibers was stained with Laminin and nuclei with dapi. Activated, ASCs (PAX7+/MYOD+) and self-renewing, QSCs (PAX7+/MYOD-) MuSCs are shown. Scale bars, 100 μm. **G**, Quantification of MuSCs pool composition in (F); n= 3-6 animals per group. **H**, Percentage of Pax7+ MuSCs on fibers for the experiment shown in (F). **B**, **C**, **E**, **G, H**, Data are mean ± s.e.m. Statistics by analysis of variance (ANOVA) with multiple comparisons. Statistical comparisons between controls and injured (G, H) were not shown. * P < 0.05, * * P < 0.01, * * * P < 0.001; wt= *Lmna* Δ8-11 +/+; het= *Lmna* Δ8-11 +/-; hom= *Lmna* Δ8-11 -/-.

Figure 2. *Lmna* Δ8-11 -/- dystrophic MuSCs display PcG displacement.

(**A**) Heatmap reporting log_2 ratios of observed over expected (colorbar) number of genes in the intersections between H3K27me3 targets identified in *Lmna* Δ8-11 +/+ mice and the up-regulated genes in *Lmna* Δ8-11 -/- mice. The expected counts in the joint probability are estimated by the products of marginal distributions. Fisher Exact Test p-value = 2.38e-05. (**B, C, D**) Average profile of H3K27me3 ChIP-seq signal calculated as the IP/input ratio over annotated mouse genes (**B**) Average profile of H3K27me3 signal around the TSS. (**C**) Average profile of H3K27me3 signal along the gene body. TES= annotated transcript end. (**D**) Average profile of H3K27me3 signal in regions outside H3K27me3 peaks and outside annotated genes. (**E**) Normalized expression distribution of genes stratified using wt expression level in the 3 biological replicates (see methods). Distribution of average log_2 transcripts per million (TPM + 0.1) values in the 3 biological replicates is plotted for wt and hom. (**F**) Average profile of H3K27me3 signal (IP/input) along the gene body using gene categories as in panel (E). wt= *Lmna* Δ8-11 +/+; hom= *Lmna* Δ8-11 -/-.

Figure 3. Lamin A-PcG-mediated transcriptional repression preserves MuSCs identity.

(**A**) Heatmap reporting log_2 ratios of observed over expected number of genes (colorbar) in the intersections between bivalent promoters identified in wt satellite cells (see methods) and the up-regulated genes in *Lmna* Δ8-11 -/- mice. Fisher's Exact Test p-value = 4.57e-07. (**B**) Semantic similarity analysis on Gene Ontology terms enriched in up-regulated genes in hom vs wt comparison (FDR < 0.05) with macro-categories identified using Revigo webtool (see methods). (**C**) Heatmap reporting log_2 ratios of observed over expected number of genes (colorbar) in the intersections between up-regulated genes in *Lmna* Δ8-11 -/- mice in Pparγ-related GO terms (see methods) and the bivalent genes identified as above. Fisher's Exact Test p-value = 6.73e-06. wt= *Lmna* Δ8-11 +/+; hom= *Lmna* Δ8-11 -/-.

Figure 4. Deregulation of Pparγ locus in *Lmna* Δ8-11 -/- MuSCs.

(**A**) Immunohistochemical staining in *Lmna* Δ8-11 mice of PAX7 (red) and PPARγ (green) markers at 19 days of post-natal growth. Nuclei of muscle fibers were stained with dapi. Scale bars, 50 μm. (**B**) Quantification of PPARγ+ MuSCs in (A); n= 4-5 animals per genotype. (**C**) ChIP-seq of H3K27me3 mark as the spike-in normalized linear ratio IP/input and RNA-seq signal tracks as the effective genome size normalized coverage on *PPARG* locus. PPARG promoter is highlighted by a light blue rectangle. (**D**) Representative image of FISH analysis of fixed and sorted MuSCs with probes indicated in Supplementary Figure 6b. Scale bars, 2 μm. (**E**) Quantification of FISH probes (represented in Figure S6B) distances (x-axis) by cumulative frequency distributions (y-axis). Only probes with distance ≤ $0.5\mu m$ are reported. (**F**) Quantification of FISH probes positioning with respect to the nuclear envelope. (**B**) Data are mean ± s.e.m. Statistics by analysis of variance (ANOVA) with multiple comparisons. (**E, F**) Comparisons by Kolmogorov-Smirnov test. * P < 0.05, * * * P < 0.001; wt= *Lmna* Δ8-11 +/+; het= *Lmna* Δ8-11 +/-; hom= *Lmna* Δ8-11 -/-.

Figure 5. *Lmna* Δ8-11 -/- MuSCs acquire senescence transcriptional traits.

(**A**) Representative image of asymmetric stem cell division in myofibers from *Lmna* Δ8-11 mice immunostained for Ph-P38 (green) and PAX7 (red) after 48 h of culture. Nuclei were stained with dapi. Scale bars, 20 μm. An apicobasal asymmetric stem cell division was observed in wt; a planar symmetric division was observed in hom. (**B**) Quantification of asymmetric divisions by Ph-P38 distribution. (**C**) Quantification of asymmetric apico-basal division vs symmetric planar divisions. n=46± 6 doublets of MuSCs per group, n= 5-6 mice per group. (**D**) GSEA on expression data from old and young mice quiescent satellite cells(51). Up-regulated (log Fold Change – logFC > 1) genes in hom vs wt comparison added to Biocarta mouse pathways from gskb R

package were used as gene sets (NES = 4.70, FDR < 10e-4). (**E**) ChIP-seq of H3K27me3 mark and RNA-seq signal tracks on Cdkn1a/p21 locus. Promoter regions are highlighted by light blue rectangles. (**B, C**) Data are mean ± s.e.m. Statistics by analysis of variance (ANOVA) with multiple comparisons. * P < 0.05; wt= *Lmna* Δ8-11 +/+; het= *Lmna* Δ8-11 +/-; hom= *Lmna* Δ8-11 -/-.

Figure 6. Cdkn2a genetic ablation restores regenerative capacity of *Lmna* Δ8-11 -/- dystrophic mice.

(**A**) ChIP-seq signal tracks of H3K27me3 mark as the spike-in normalized linear ratio IP/input on Cdkn2a locus. Promoters are highlighted by a light blue rectangle. (**B**) Transcriptional analysis of $p16^{INK4}$ and $p19^{ARF}$ at Cdkn2a locus in *Lmna* Δ8-11 mice at 19 and 26 days of postnatal growth. The p21 CDKI is our positive control of up-regulation in hom. Nd, not detected. (**C**) Immunohistochemical staining in *Cdkn2a*/*Lmna* Δ8-11 mice of PAX7 (red) and MYOD (green) markers at 19 days of post-natal growth. Basement membrane of muscle fibers was stained with Laminin and nuclei with dapi. Activated, ASCs (PAX7+/MYOD+) and self-renewing, QSCs (PAX7+/MYOD-) MuSCs are shown. Scale bars, 50 μm. (**D**) Quantification of MuSCs pool composition in (C). (**E**) Quantification of myofibre size, evaluated by the Cross Sectional area (CSA). n= 600 muscle fibers. n=4-5 animals per genotype. (**B**) Data are mean ± s.e.m. Comparisons by unpaired *t-test*. (**D, E**) Data are mean ± s.e.m. Statistics by analysis of variance (ANOVA) with multiple comparisons. * P < 0.05, * * P < 0.01, * * * P < 0.001; wt= *Lmna* Δ8-11 +/+; hom= *Lmna* Δ8-11 -/-.

Figure 7. Lamin A/C-Polycomb crosstalk in Lamin dependent muscular dystrophy.

Absence of Lamin A/C determines PcG displacement and relaxation of PcG-mediated higher-order chromatin structure. In *Lmna* Δ8-11 -/- MuSCs PcG

dysfunction causes a lack of cellular identity and premature senescence finally leading to an exhaustion of the stem cell niche.

Supplementary Figure 1. H222P mice do not show altered balance in the muscle MuSC niche.

(**A**) Immunohistochemical staining in adult, injured *H222P* mice of PAX7 (red) and MYOD (green) markers. Activated, ASCs (PAX7+/MYOD+) and self-renewing, QSCs (PAX7+/MYOD-) MuSCs are shown. Basement membrane of muscle fibers was stained with Laminin and nuclei with dapi. Scale bars, 100 μm. (**B**) Quantification of PAX7+ MuSCs per fiber in (A). (**C**) Quantification of myofibre size in (A), evaluated by the Cross Sectional area (CSA). n= 480 muscle fibers. (**D**) Quantification of MuSCs pool composition in (A); n= 4-5 animals per group. (**E**) Scheme for the persistent muscular injury. (**B, C, D**) Data are mean ± s.e.m. Statistics by analysis of variance (ANOVA) with multiple comparisons. Statistical comparisons between controls and injured were not shown.

Supplementary Figure 2. Dystrophic mice show unbalanced stem cell niche.

(**A**) Immunohistochemical staining in *Lmna* Δ8-11 mice of PAX7 (red) and MYOD (green) markers at the indicated days of post-natal growth. Basement membrane of muscle fibers was stained with Laminin and nuclei with dapi. Scale bars, 50 μm. (**B**) Quantification of PAX7+ MuSCs per fiber at the indicated days of post-natal growth in (A). n =3-6 animals per genotype. (**C**) Immunohistochemical staining in *Lmna* Δ8-11 mice of Activated, ASCs (PAX7+/KI67+) and self-renewing, QSCs (PAX7+/Ki67-) of MuSCs at 19 days of post-natal growth and relative quantification (on the right). Nuclei were stained with dapi. n= 4-6 animals per group. Scale bars, 50 μm. (**D**) Quantification of the number of cells per cluster in single myofibers extracted from P19 *Lmna* Δ8-11 mice and cultured 96h. n= 4-5 animals per group. (**E**) RNA-seq

signal tracks as the effective genome size normalized coverage of each biological replicate of *Lmna* Δ8-11 mice on Lmna locus. Neomycine cassette is indicated as a dark blue rectangle. (**F**) Western blot of total protein extracted from the whole *Lmna* Δ8-11 muscles hybridized with indicated antibodies. α-tubulin was used as loading control. (**B, C, D**) Data are mean ± s.e.m. Statistics by analysis of variance (ANOVA) with multiple comparisons. * * P < 0.01, * * * P < 0.001. wt= *Lmna* Δ8-11 +/+; het= *Lmna* Δ8-11 +/-; hom= *Lmna* Δ8-11 -/-.

Supplementary Figure 3. *Lmna* Δ8-11 dystrophic mice show alteration of PcG nuclear architecture and transcriptional regulation.

(**A**) Representative images of Ezh2/Lamin A immunofluorescence analysis on fixed and sorted *Lmna* Δ8-11 MuSCs. On the right, 3D reconstruction of PcG bodies (see methods). Of note, in some *Lmna* Δ8-11 -/- nuclei we could see low levels of the truncated form of Lamin A/C(83) recognizable by the Lamin A/C antibody. (**B**) Quantification of number of PcG bodies/nucleus (x-axis) by cumulative frequency distributions (y-axis). (**C**) Quantification of number of PcG body volumes (x-axis) by cumulative frequency distributions (y-axis). (**D**) Heatmap of differentially expressed genes Z-scores in *Lmna* Δ8-11 -/- vs *Lmna* Δ8-11 +/+ comparison (FDR < 0.05). (**E**) pre-ranked GSEA on log Fold Change from Ezh2 KO vs WT satellite cells(26). Up-regulated (logFC > 1) in hom vs wt comparison added to Biocarta mouse pathways from *gskb* R package were used as gene sets (NES = 2.31, FDR = 0.020). (**F, G**) heatmaps showing H3K27me3 spike-in normalized linear ratio IP/input signal along the gene body (F) or around the TSS (G) (coordinates along x-axis) of individual annotated mouse genes (y-axis). Genes are sorted by computing the average signal for each gene and then sorting them in descending order based on the mean value. TES= annotated transcript end. (**H**) Western blot of total histones extracted from *Lmna* Δ8-11 MuSCs hybridized with indicated antibodies. The * indicates the truncated form of Lamin A detected in mutant mice(83). (**I**) Band density

quantification of H3K27me3 normalized on H3 levels. n=2. Ns: not statistical. (**J**) Average profile of H3K27me3 signal as the spike-in normalized linear ratio IP/input at the TSS using genes grouped as in panel (Figure 2E). (**J**) Data are mean ± s.e.m. Comparisons by unpaired *t-test*. (**B, C**) Comparisons by Kolmogorov-Smirnov test. * * * P < 0.001. wt= *Lmna* Δ8-11 +/+; hom= *Lmna* Δ8-11 -/-.

Supplementary Figure 4. Aberrant H3K27me3 deposition in *Lmna* Δ8-11 -/- MuSC cells

(**A**) ChIP-seq of H3K27me3 mark as the spike-in normalized linear ratio IP/input and RNA-seq signal tracks as the effective genome size normalized coverage on Hoxb genes, belonging to the 0/I/II quartiles of expression. (**B**) ChIP-seq of H3K27me3 mark as the spike-in normalized linear ratio IP/input and RNA-seq signal tracks as the effective genome size normalized coverage on Eno2 locus, belonging to the III quartile of expression. Eno2 promoter is highlighted by a light blue rectangle. (**C**) Cartoon to illustrate the approach used to measure relative difference in H3K27me3 ChIP over input enrichment at TSS (A point) vs 2.5Kb downstream (B point). For each sample (wt or hom), for each individual gene, the ratio o H3K27me3 ChIP over input enrichment at point A over point B is computed. (**D**) The boxplots report the distribution of A/B ChIP-seq enrichment ratios (defined as in panel C) in each sample, for each gene, grouped as defined in panel (Figure 2E). The p-values reported are based on non-parametric Wilcoxon test.

Supplementary Figure 5. Bivalent genes deregulation in *Lmna* Δ8-11 -/- MuSC cells

(**A, B**) Average profile and heatmap of H3K27me3 (A) and H3K4me3 (B) ChIP-seq signal along the gene body calculated as the ratio IP/input over annotated mouse genes. TES= annotated transcript end. (**C**) ChIP-seq signal tracks of H3K27me3 and H3K4me3 marks as the library size normalized linear ratio IP/input on Lonrf2 locus,

showing a bivalent promoter. Lonrf2 promoter is highlighted by a light blue rectangle. (**D, E**) Result of pre-ranked GSEA using log fold change from RNA-seq differential analysis in hom vs wt comparison. Representative muscle development-related NES = -3.11, FDR = 1.86e-5 (**D**) and lipid metabolism-related NES = 2.91, FDR = 1.54e-4 (**E**) gene sets among Gene Ontology terms from gskb R package are shown.

Supplementary Figure 6. PPARG is a bivalent gene.

(**A**) ChIP-seq signal tracks of H3K27me3 and H3K4me3 marks as the library size normalize linear ratio ChIP/input on PPARG locus, showing the bivalent promoter. PPARG promoter is highlighted by a light blue rectangle. (**B**) At the top, the Hi-C contact map of mouse embryonic stem cells ((37), resolution 10kb) showing the TADs compartmentalization (horizontal light yellow and blue bars) in the region chr6:114,760,000-116,080,000. At the bottom, H3K27me3 ChIP-seq spike-in normalized linear IP/input ratio of the same genomic region. In the middle of two dashed lines are highlighted as dark blue rectangles the mapping positions of BAC clones used for FISH experiments described in Figure 4D. H3K27me3 peak lost in the *Lmna* Δ8-11 -/- MuSCs is highlighted by light red rectangle.

Supplementary Figure 7. EDMD derived iPS recapitulate transcriptional defects of lamin deficient MuSCs.

(**A**) Representative images of immunochemical analysis performed with indicated pluripotency markers on iPS colonies derived from wt and EDMD myoblasts. Myoblasts phase contrast images are placed above. (**B**) Transcriptional analysis with pluripotency markers of distinct iPS clones derived from wt and EDMD myoblasts. (**C**) Scheme of the human Lamin A/C gene on chromosome 1 and sequences of control and EDMD derived clones. Asterisks indicated the mutated nucleotide in EDMD. (**D**) Representative images of immunochemical analysis performed with indicated muscular markers on myotubes differentiated from wt and EDMD iPS. (**E,**

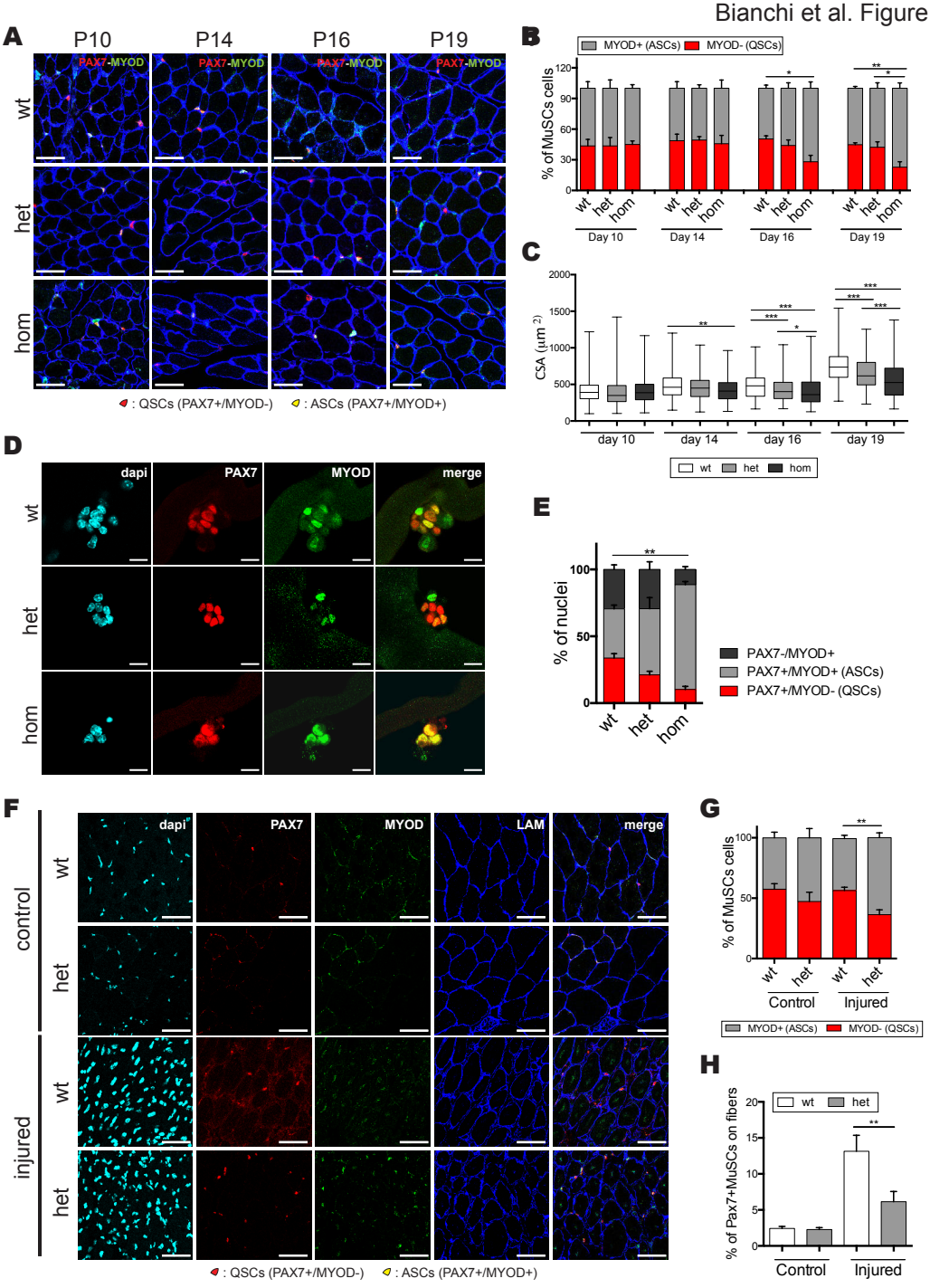
F, G) Transcriptional analysis with markers of pluripotency (E), muscle differentiation (F) and adipogenesis (G) of two independent iPS clones derived from wt and EDMD myoblasts. 0 represents the iPS state and E represents the differentiation endpoint.

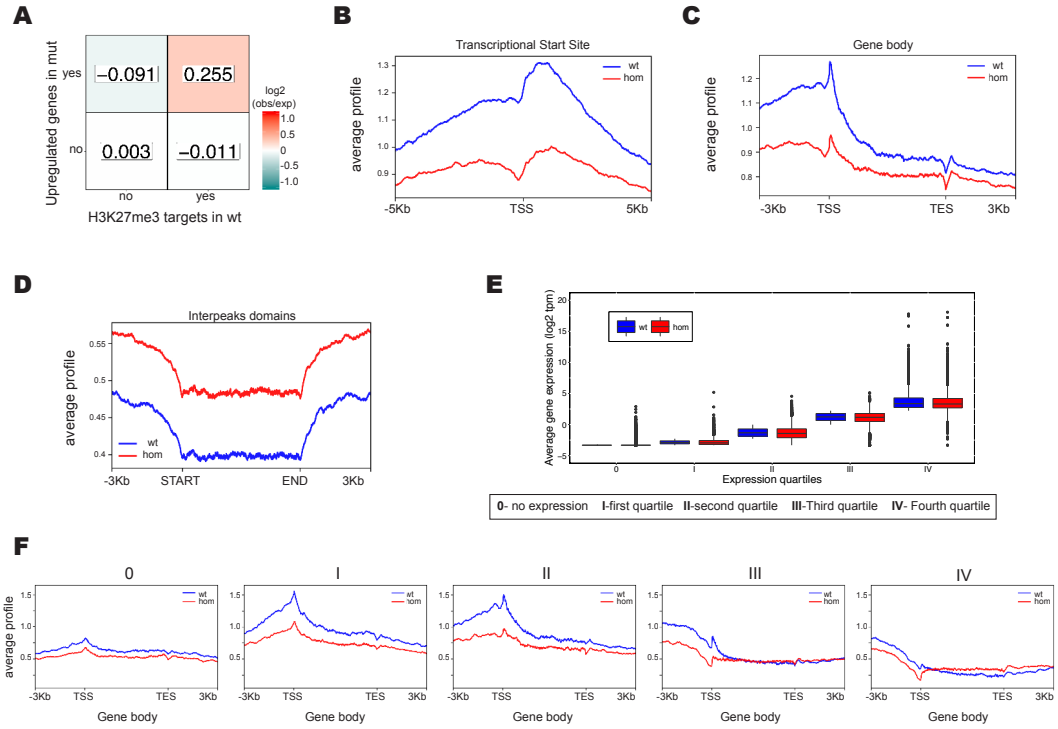
Supplementary Figure 8. *Lmna* Δ8-11 -/- SC cells undergo premature senescence.

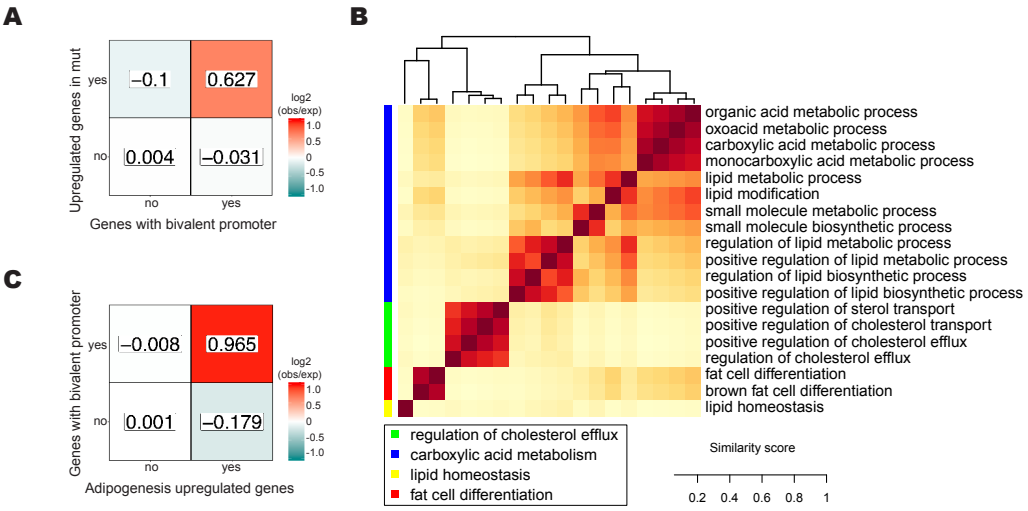
(**A**) Immunohistochemical staining in *Lmna* Δ8-11 mice of PAX7 (red) and Ph-P38 (green) markers at 19 days of post-natal growth. Basement membrane of muscle fibers was stained with Laminin and nuclei with dapi. Scale bars, 50 μm. (**B**) Quantification of MuSCs pool in (A); n= 5 animals per genotype. (**C**) Immunohistochemical staining in *Lmna* Δ8-11 mice of PAX7 (red) and γH2AX (green) markers at 19 days of post-natal growth. Basement membrane of muscle fibers was stained with Laminin and nuclei with dapi. Scale bars, 50 μm. (**D**) Quantification of MuSCs pool in (C); n= 3 animals per genotype. (**E, F**) pre-ranked GSEA on log Fold Change from RNA-seq differential analysis in hom vs wt comparison. Up-regulated genes (q-value < 0.05) in old vs young (E) and geriatric vs young (F) mice comparisons from (52) were added to Biocarta mouse pathways from *gskb* R package and used as gene sets. (**E**) old vs. young mice up-regulated genes are enriched in hom vs wt up-regulated ones (NES = 2.86, FDR = 3.81e-04). (**F**) geriatric vs young mice up-regulated genes are enriched in hom vs wt down-regulated ones (NES = -5,09, FDR < 10e-4). (**B, D**) Data are mean ± s.e.m. Statistics by analysis of variance (ANOVA) with multiple comparisons. * * P < 0.01. wt= *Lmna* Δ8-11 +/+; het= *Lmna* Δ8-11 +/-; hom= *Lmna* Δ8-11 -/-.

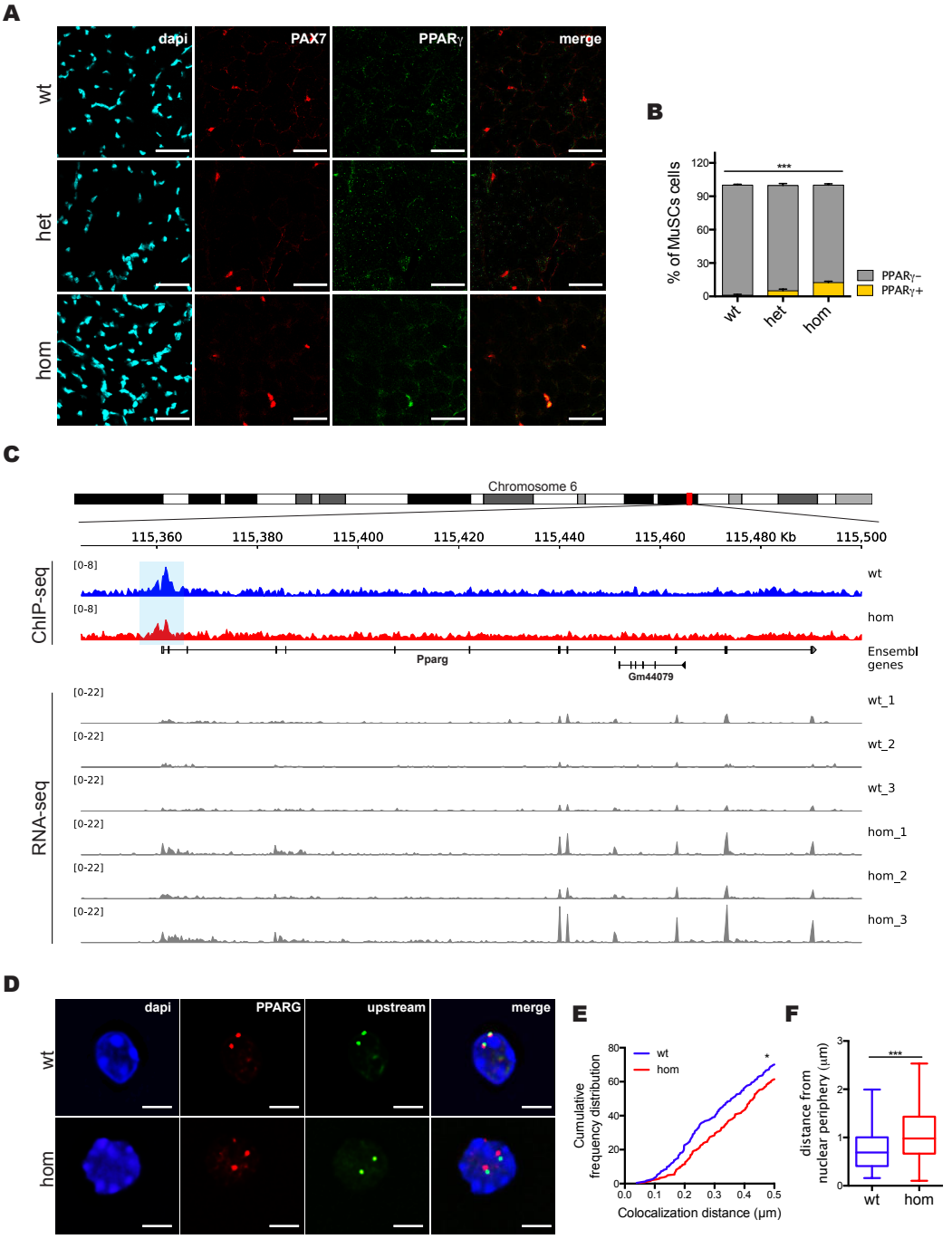
Supplementary Figure 9. Cdkn2a absence improves quiescence maintenance in *Lmna* Δ8-11 -/- mice.

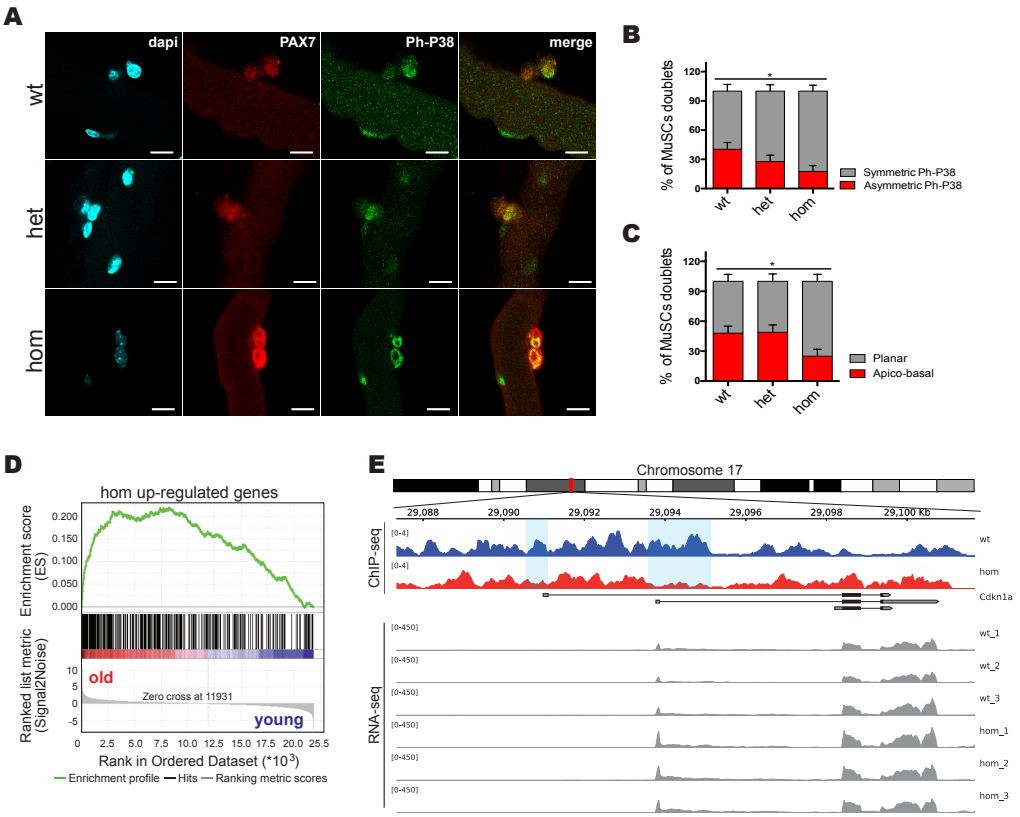
(**A**) Immunohistochemical staining in *Cdkn2a*/*Lmna* Δ8-11 mice of Activated, ASCs (PAX7+/KI67+) and self-renewing, QSCs (PAX7+/Ki67-) of MuSCs at 19 days of post-natal growth. Nuclei were stained with dapi. Scale bars, 50 μm. (**B**) Quantification of MuSCs pool in (A). n=4-5 animals per genotype. Data are mean ± s.e.m. Statistics by analysis of variance (ANOVA) with multiple comparisons. * $P < 0.05$, ** $P < 0.01$.

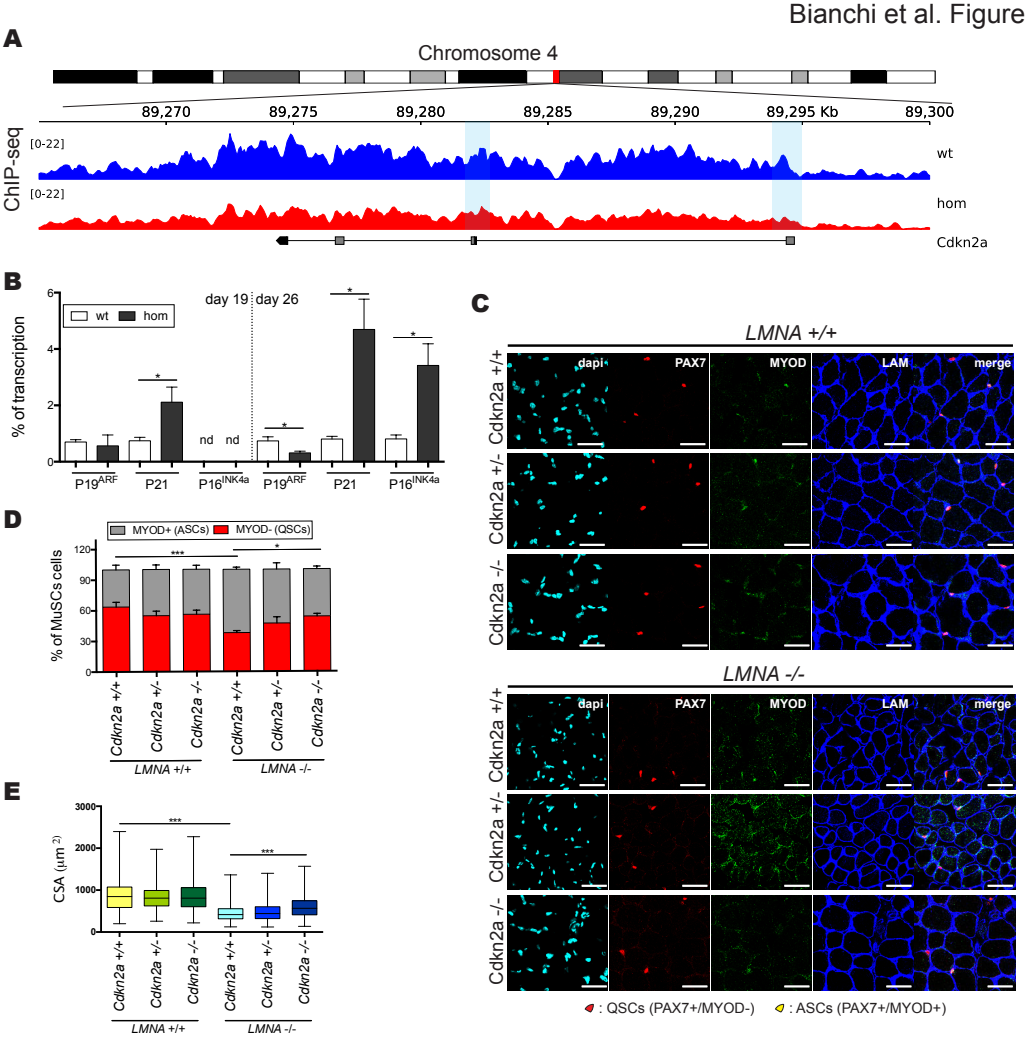

A
P10
P14
P16
P19
PAX7-MYOD
wt
het
hom
: QSCs (PAX7+/MYOD-)
: ASCs (PAX7+/MYOD+)
B
MYOD+ (ASCs)
MYOD- (QSCs)
% of MuSCs cells
wt
het
hom
Day 10
Day 14
Day 16
Day 19
C
CSA (µm 2)
day 10
day 14
day 16
day 19
wt
het
hom
D
dapi
PAX7
MYOD
merge
E
% of nuclei
PAX7-/MYOD+
PAX7+/MYOD+ (ASCs)
PAX7+/MYOD- (QSCs)
F
control
injured
LAM
G
% of MuSCs cells
Control
Injured
H
% of Pax7+MuSCs on fibers

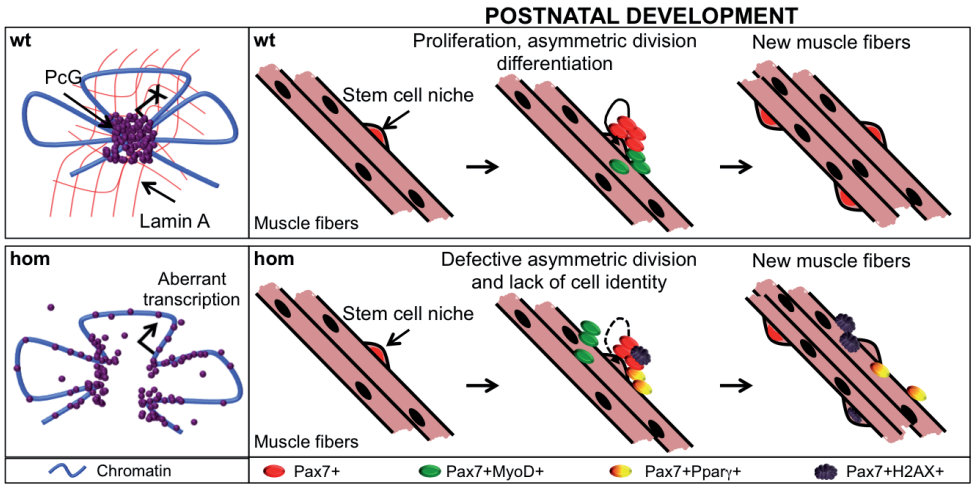

A
Upregulated genes in mut
yes
no
−0.091
0.255
0.003
−0.011
no
yes
H3K27me3 targets in wt
log2 (obs/exp)
1.0
0.5
0
-0.5
-1.0
B
Transcriptional Start Site
average profile
wt
hom
-5Kb
TSS
5Kb
C
Gene body
average profile
wt
hom
-3Kb
TSS
TES
3Kb
D
Interpeaks domains
average profile
wt
hom
-3Kb
START
END
3Kb
E
Average gene expression (log2 tpm)
wt
hom
0
I
II
III
IV
Expression quartiles
0- no expression
I-first quartile
II-second quartile
III-Third quartile
IV- Fourth quartile
F
0
I
II
III
IV
average profile
wt
hom
-3Kb
TSS
TES
3Kb
Gene body

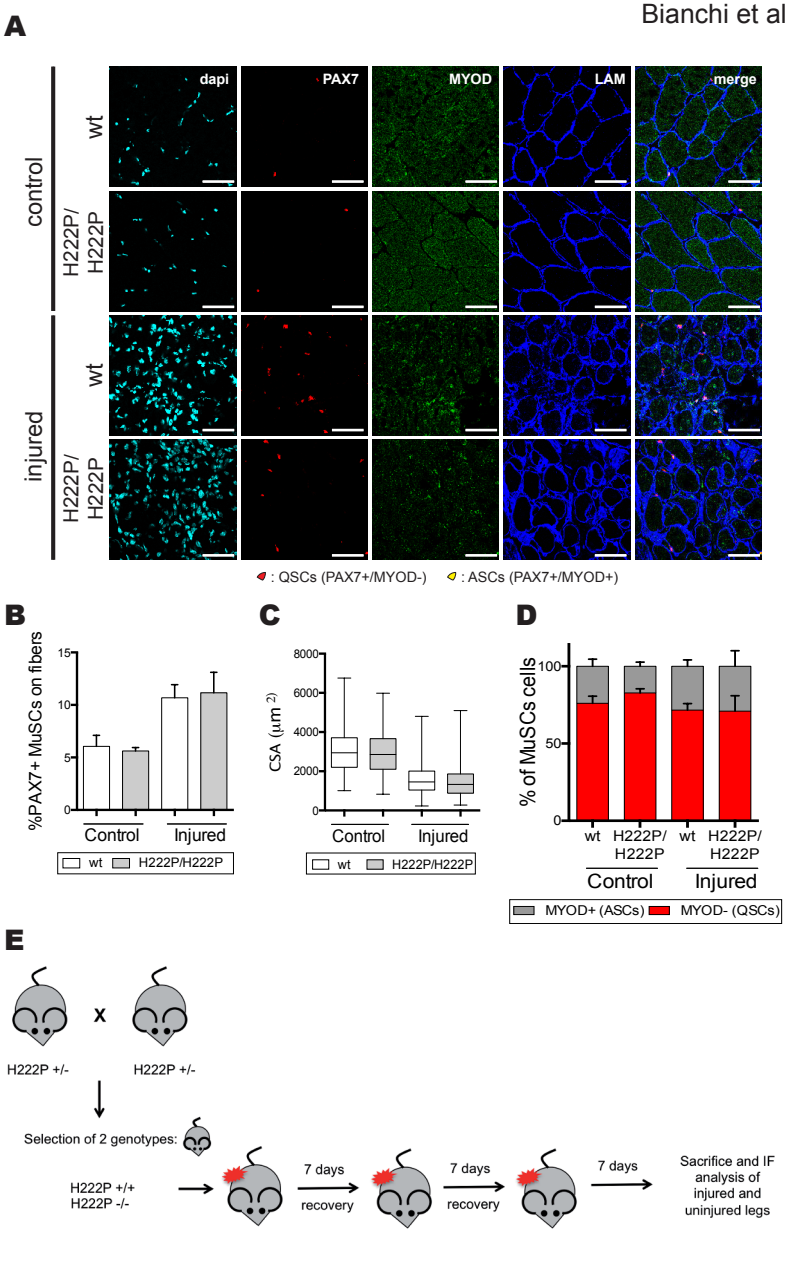

A
Upregulated genes in mut
yes
no
−0.1
0.627
0.004
−0.031
no
yes
Genes with bivalent promoter
log2 (obs/exp)
1.0
0.5
0
-0.5
-1.0
C
Genes with bivalent promoter
yes
no
−0.008
0.965
0.001
−0.179
no
yes
Adipogenesis upregulated genes
log2 (obs/exp)
1.0
0.5
0
-0.5
-1.0
B
organic acid metabolic process
oxoacid metabolic process
carboxylic acid metabolic process
monocarboxylic acid metabolic process
lipid metabolic process
lipid modification
small molecule metabolic process
small molecule biosynthetic process
regulation of lipid metabolic process
positive regulation of lipid metabolic process
regulation of lipid biosynthetic process
positive regulation of lipid biosynthetic process
positive regulation of sterol transport
positive regulation of cholesterol transport
positive regulation of cholesterol efflux
regulation of cholesterol efflux
fat cell differentiation
brown fat cell differentiation
lipid homeostasis
regulation of cholesterol efflux
carboxylic acid metabolism
lipid homeostasis
fat cell differentiation
Similarity score
0.2
0.4
0.6
0.8
1

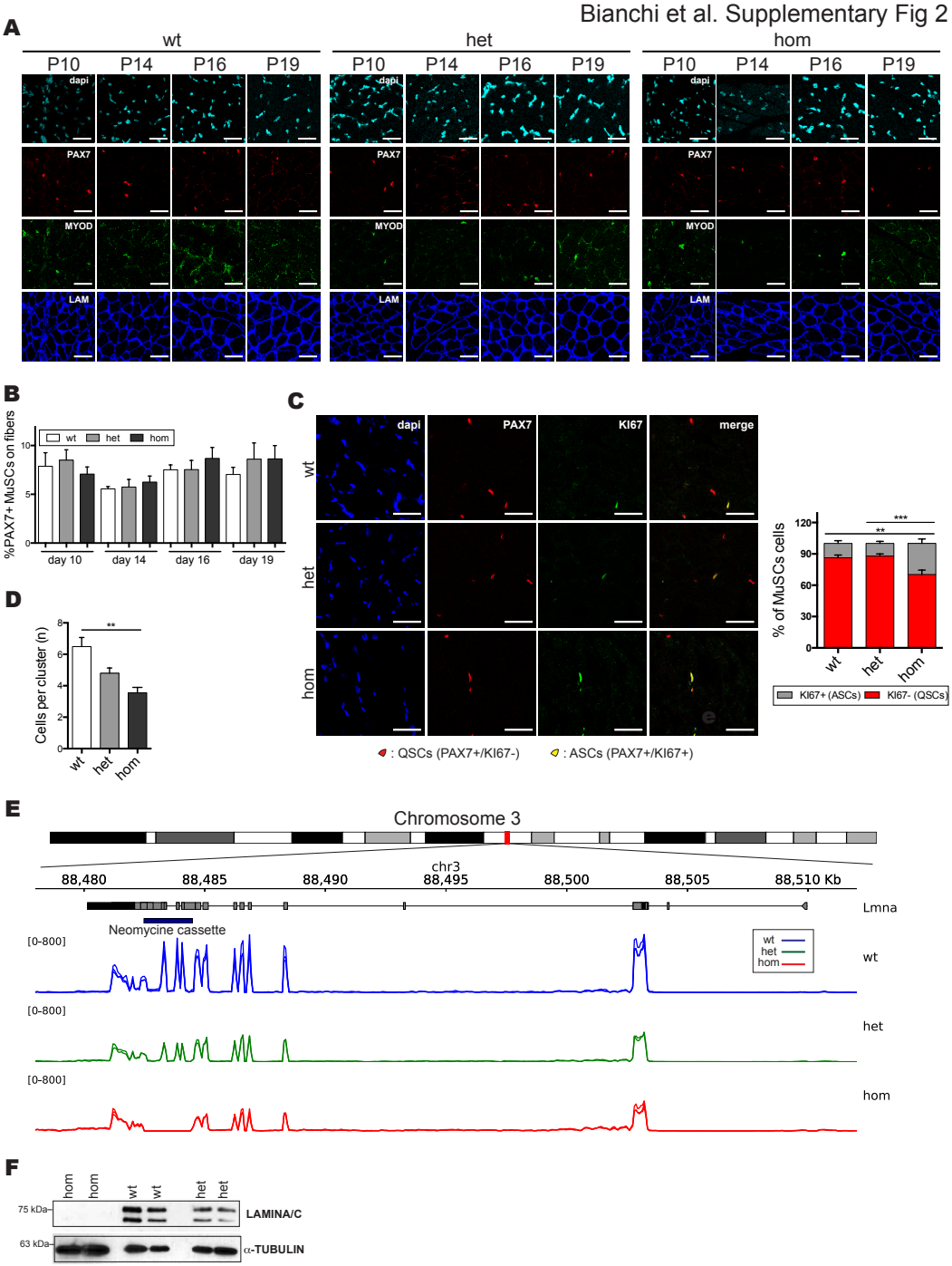

A
dapi
PAX7
PPARγ
merge
wt
het
hom
B
% of MuSCs cells
120
90
60
30
0
***
wt
het
hom
PPARγ−
PPARγ+
C
Chromosome 6
115,360
115,380
115,400
115,420
115,440
115,460
115,480 Kb
115,500
ChIP-seq
[0-8]
wt
[0-8]
hom
Pparg
Gm44079
Ensembl genes
RNA-seq
[0-22]
wt_1
[0-22]
wt_2
[0-22]
wt_3
[0-22]
hom_1
[0-22]
hom_2
[0-22]
hom_3
D
dapi
PPARG
upstream
merge
wt
hom
E
Cumulative frequency distribution
wt
hom
*
0.0
0.1
0.2
0.3
0.4
0.5
Colocalization distance (µm)
F
distance from nuclear periphery (µm)
***
wt
hom

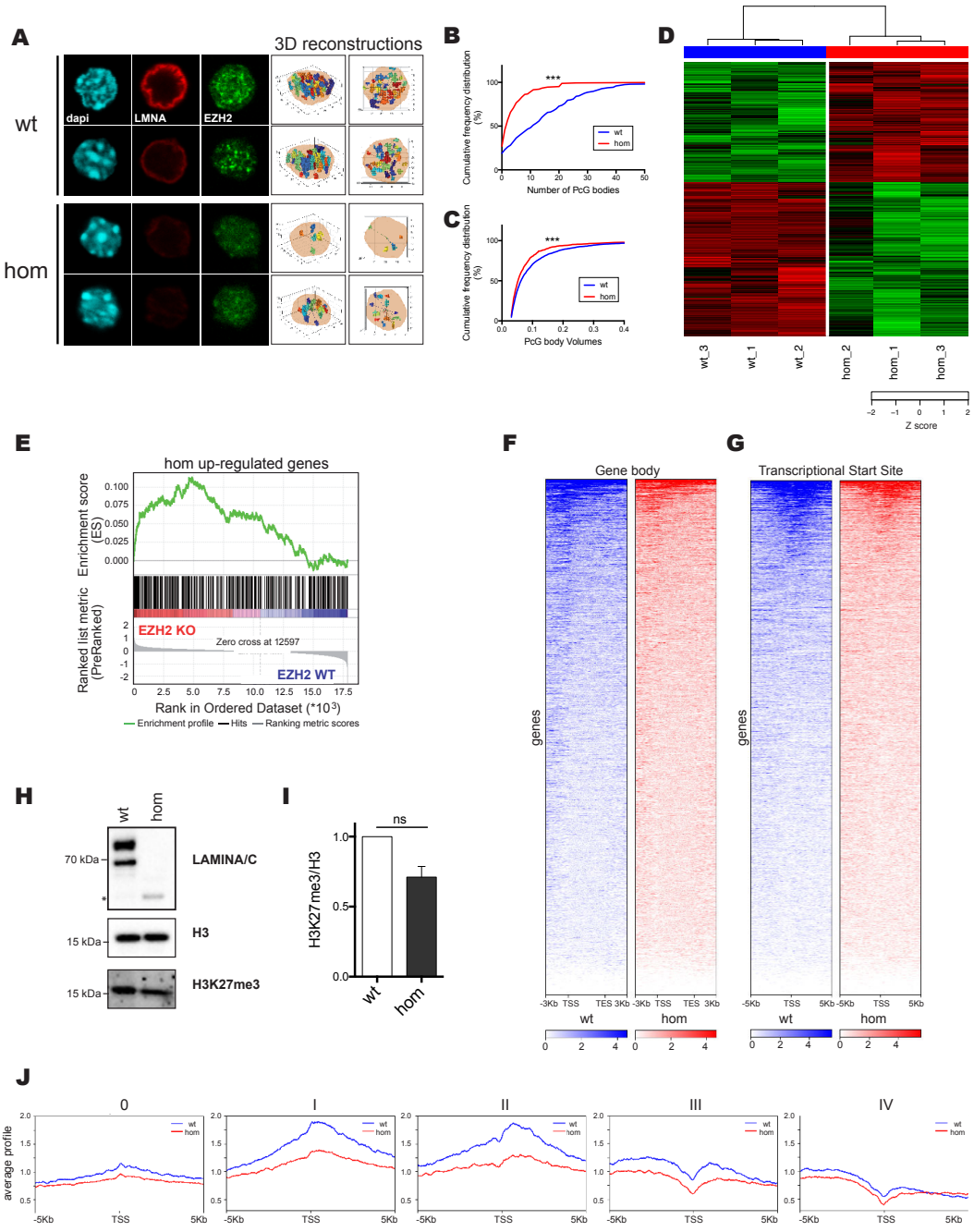

A
dapi
PAX7
Ph-P38
merge
wt
het
hom
B
% of MuSCs doublets
Symmetric Ph-P38
Asymmetric Ph-P38
C
% of MuSCs doublets
Planar
Apico-basal
D
hom up-regulated genes
Enrichment score (ES)
Ranked list metric (Signal2Noise)
old
young
Zero cross at 11931
Rank in Ordered Dataset (*10^3)
Enrichment profile
Hits
Ranking metric scores
E
Chromosome 17
ChIP-seq
RNA-seq
Cdkn1a
wt
hom
wt_1
wt_2
wt_3
hom_1
hom_2
hom_3

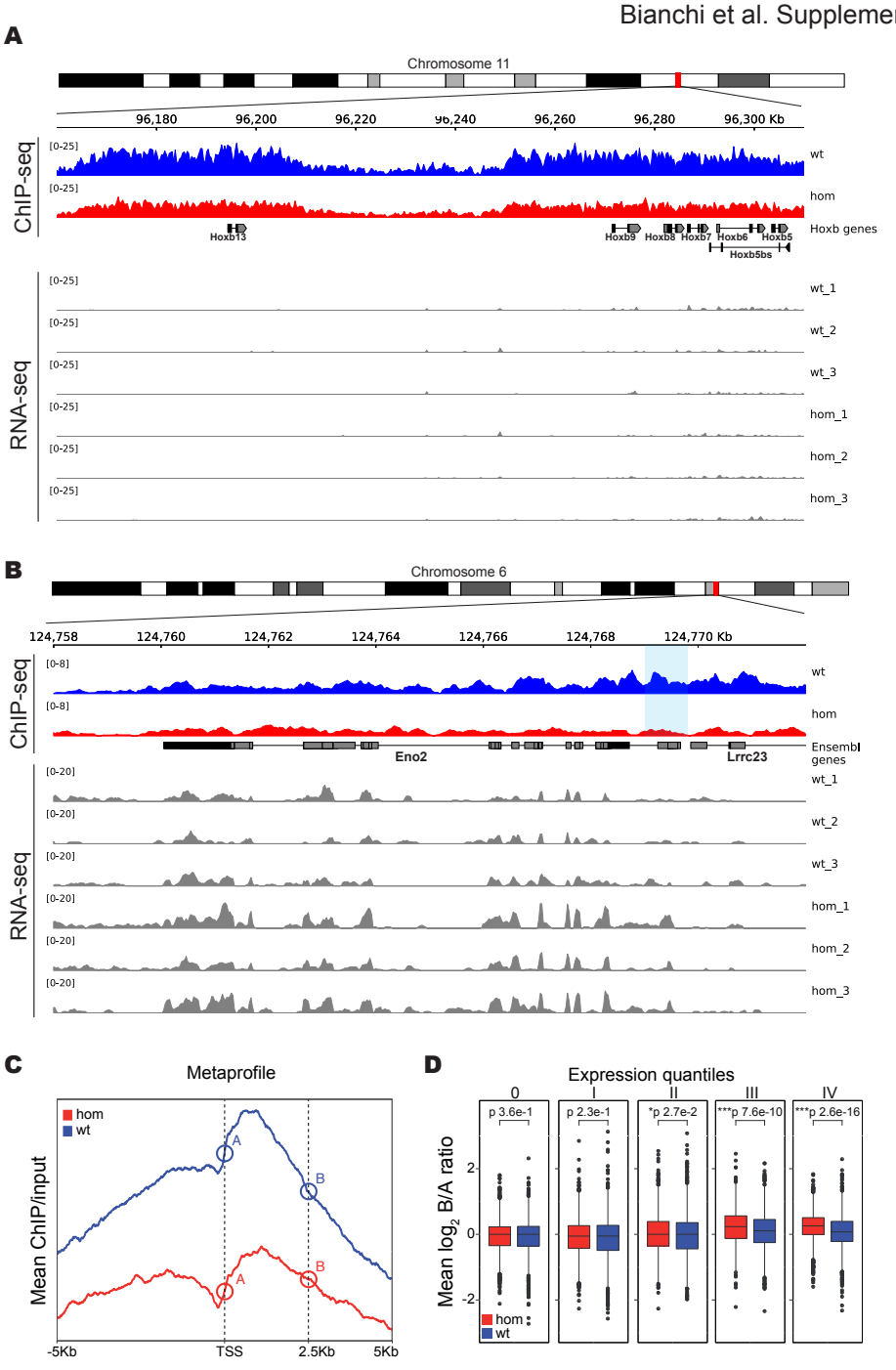

A
Chromosome 4
89,270
89,275
89,280
89,285
89,290
89,295 Kb
89,300
ChIP-seq
[0-22]
wt
[0-22]
hom
Cdkn2a
B
% of transcription
wt
hom
day 19
day 26
nd
nd
P19ARF
P21
P16INK4a
P19ARF
P21
P16INK4a
C
LMNA +/+
LMNA -/-
Cdkn2a +/+
Cdkn2a +/-
Cdkn2a -/-
dapi
PAX7
MYOD
LAM
merge
: QSCs (PAX7+/MYOD-)
: ASCs (PAX7+/MYOD+)
D
MYOD+ (ASCs)
MYOD- (QSCs)
% of MuSCs cells
Cdkn2a +/+
Cdkn2a +/-
Cdkn2a -/-
LMNA +/+
LMNA -/-
E
CSA (µm 2)
Cdkn2a +/+
Cdkn2a +/-
Cdkn2a -/-
LMNA +/+
LMNA -/-


POSTNATAL DEVELOPMENT
wt
PcG
Lamin A
wt
Proliferation, asymmetric division
differentiation
New muscle fibers
Stem cell niche
Muscle fibers
hom
Aberrant
transcription
hom
Defective asymmetric division
and lack of cell identity
New muscle fibers
Stem cell niche
Muscle fibers
Chromatin
Pax7+
Pax7+MyoD+
Pax7+Pparγ+
Pax7+H2AX+

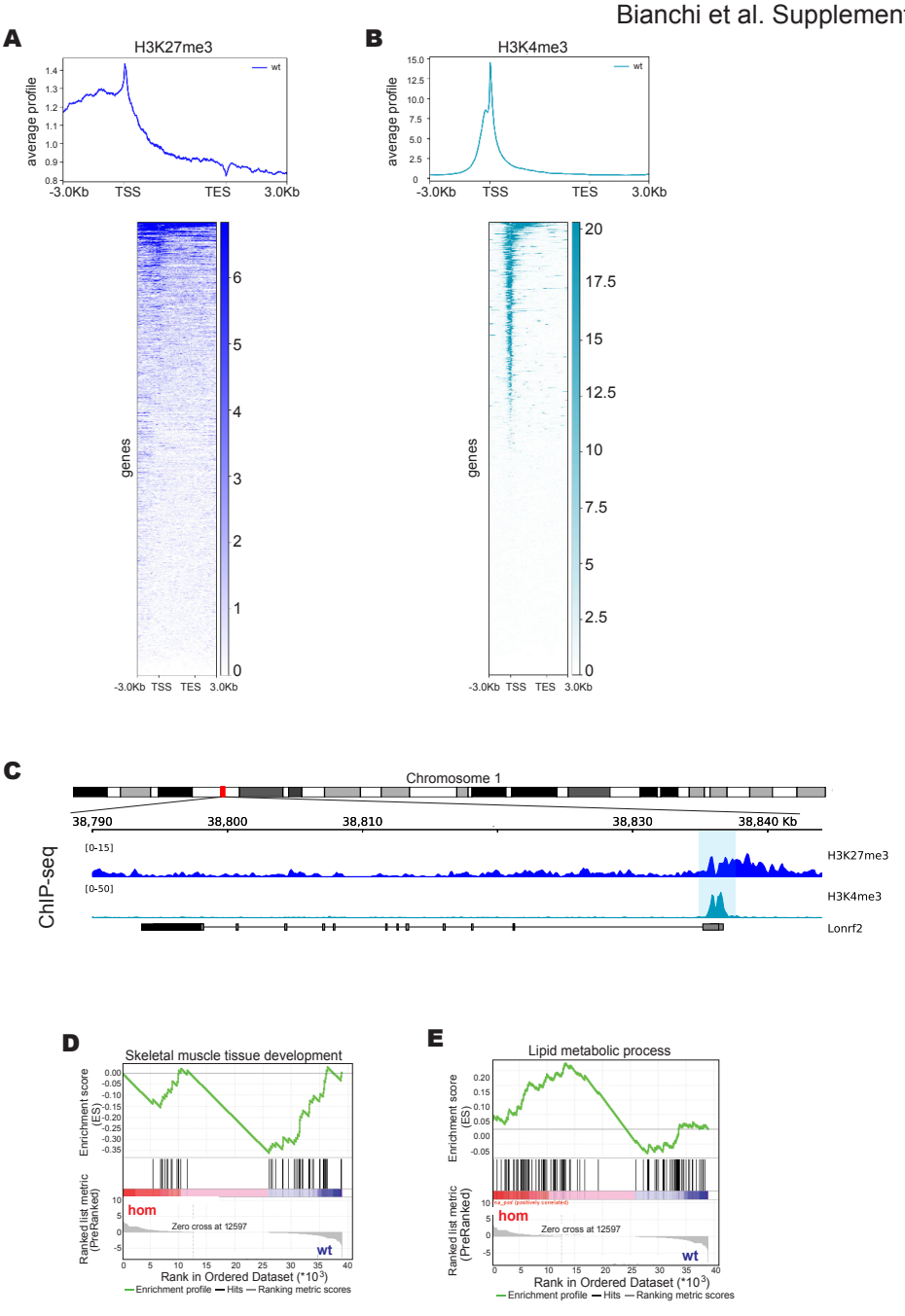

A
dapi
PAX7
MYOD
LAM
merge
control
injured
wt
H222P/ H222P
: QSCs (PAX7+/MYOD-)
: ASCs (PAX7+/MYOD+)
B
%PAX7+ MuSCs on fibers
Control
Injured
wt
H222P/H222P
C
CSA (µm-2)
D
% of MuSCs cells
MYOD+ (ASCs)
MYOD- (QSCs)
E
H222P +/-
H222P +/-
Selection of 2 genotypes:
H222P +/+
H222P -/-
7 days
recovery
Sacrifice and IF analysis of injured and uninjured legs

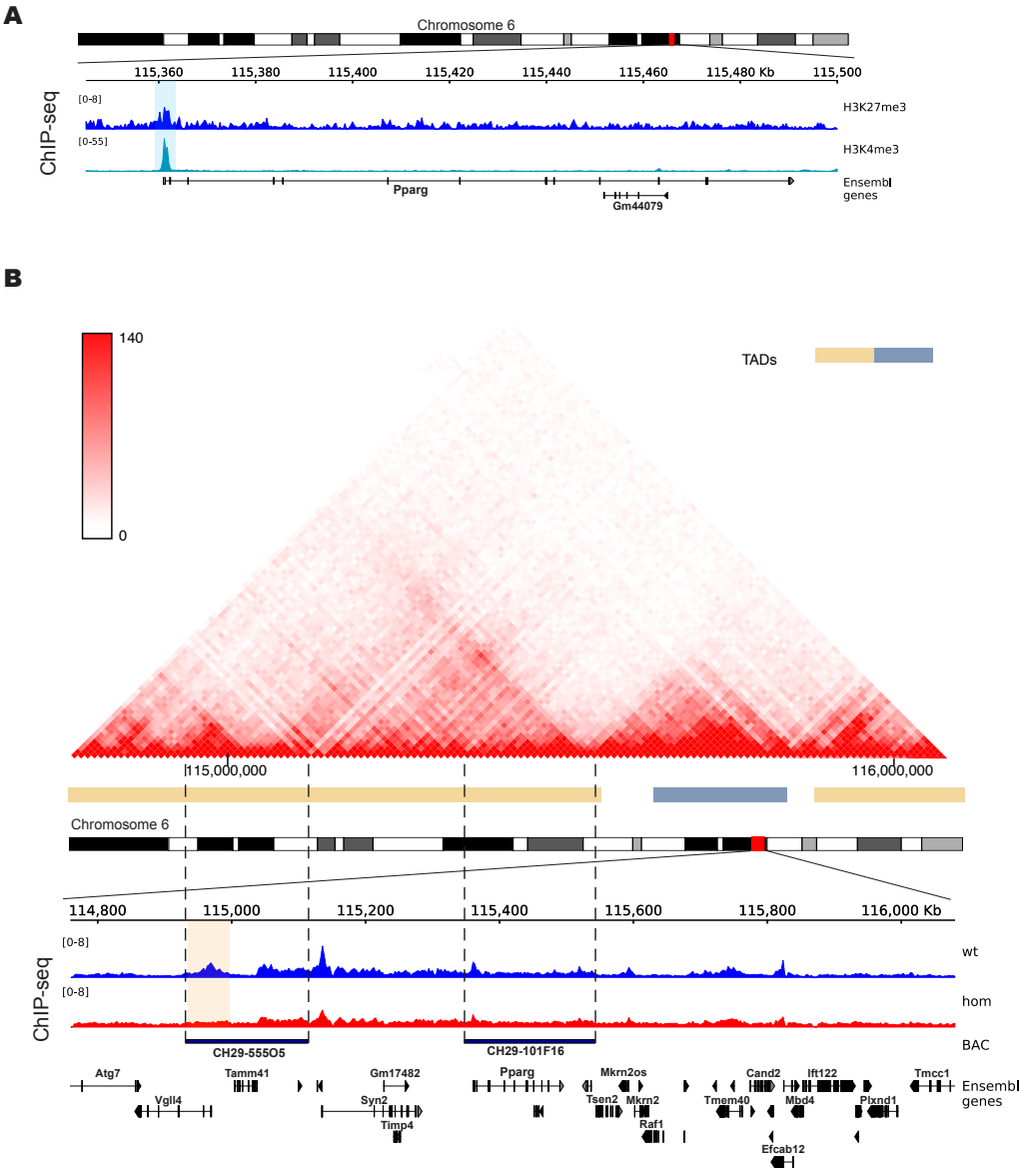

A
wt
het
hom
P10
P14
P16
P19
dapi
PAX7
MYOD
LAM
B
%PAX7+ MuSCs on fibers
wt
het
hom
day 10
day 14
day 16
day 19
C
dapi
PAX7
KI67
merge
wt
het
hom
: QSCs (PAX7+/KI67-)
: ASCs (PAX7+/KI67+)
% of MuSCs cells
KI67+ (ASCs)
KI67- (QSCs)
D
Cells per cluster (n)
E
Chromosome 3
chr3
88,480
88,485
88,490
88,495
88,500
88,505
88,510 Kb
Lmna
Neomycine cassette
[0-800]
F
75 kDa
63 kDa
LAMINA/C
α-TUBULIN

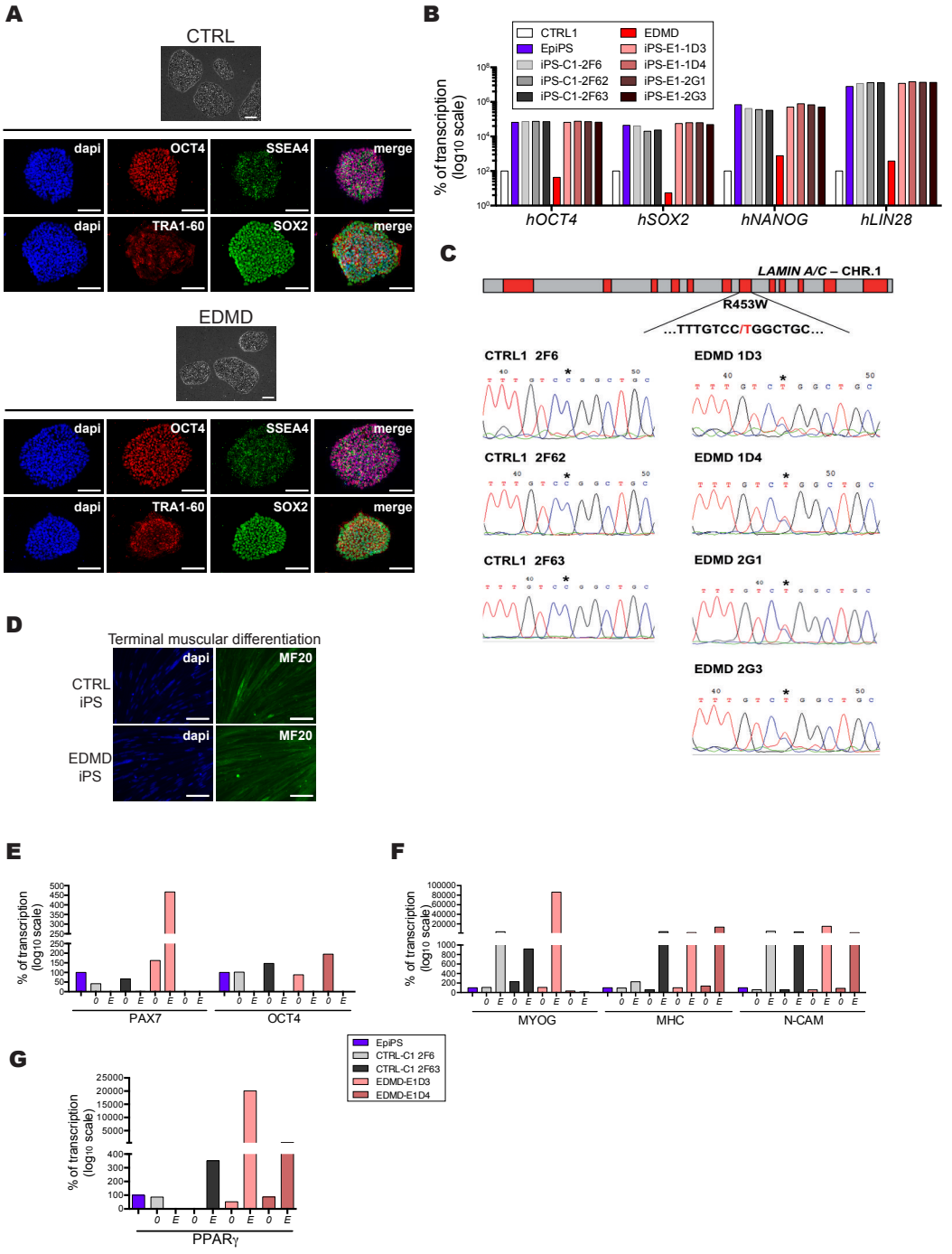

A
3D reconstructions
wt
hom
dapi
LMNA
EZH2
B
Cumulative frequency distribution (%)
Number of PcG bodies
***
wt
hom
C
Cumulative frequency distribution (%)
PcG body Volumes
***
wt
hom
D
wt_3
wt_1
wt_2
hom_2
hom_1
hom_3
Z score
E
hom up-regulated genes
Enrichment score (ES)
Ranked list metric (PreRanked)
EZH2 KO
Zero cross at 12597
EZH2 WT
Rank in Ordered Dataset (*10^3)
Enrichment profile
Hits
Ranking metric scores
F
Gene body
genes
wt
hom
G
Transcriptional Start Site
genes
wt
hom
H
wt
hom
70 kDa
LAMINA/C
15 kDa
H3
15 kDa
H3K27me3
I
ns
H3K27me3/H3
wt
hom
J
0
I
II
III
IV
average profile
wt
hom
-5Kb
TSS
5Kb

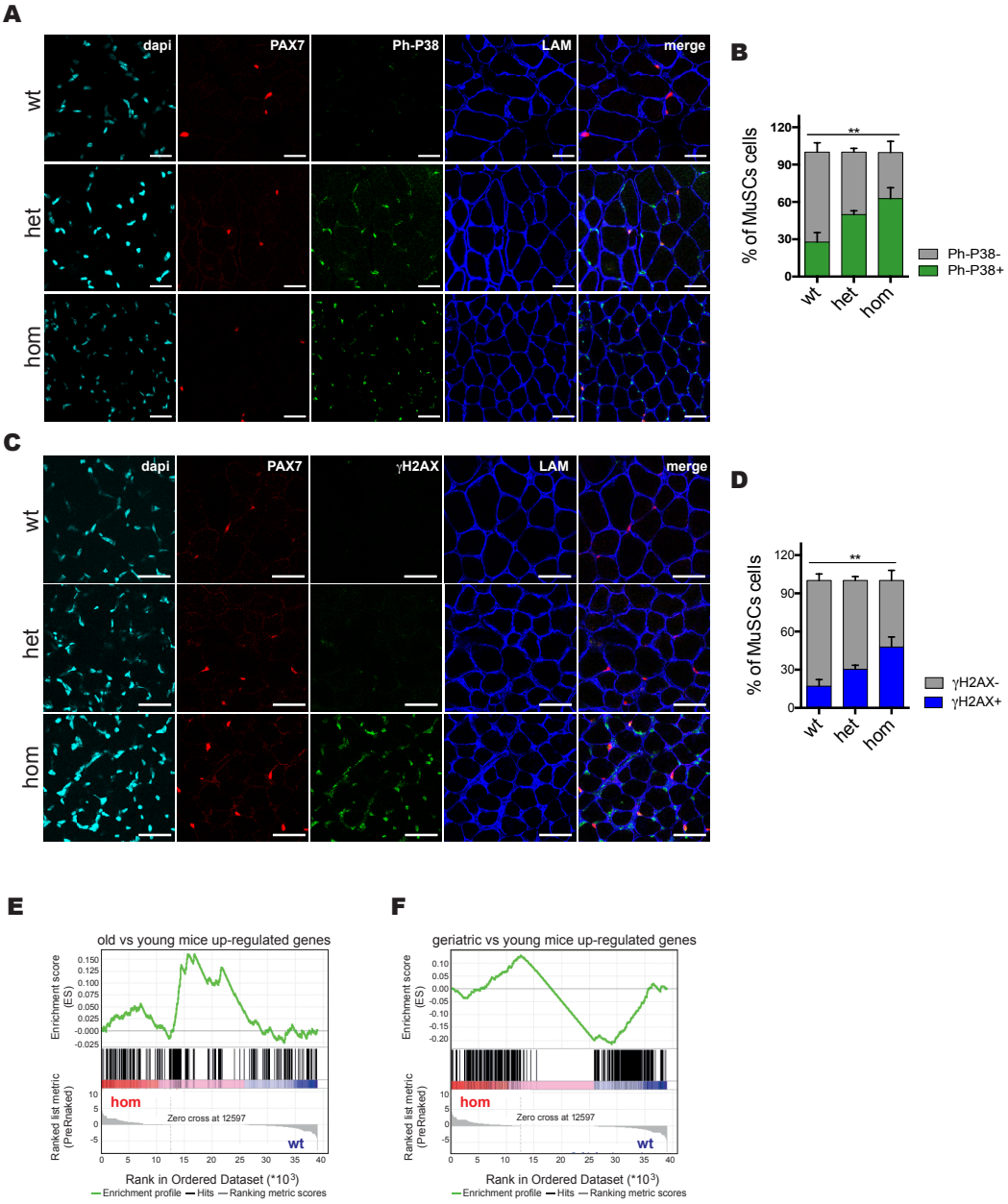

A
Chromosome 11
96,180
96,200
96,220
96,240
96,260
96,280
96,300 Kb
ChIP-seq
[0-25]
wt
hom
Hoxb genes
Hoxb13
Hoxb9
Hoxb8
Hoxb7
Hoxb6
Hoxb5
Hoxb5bs
RNA-seq
wt_1
wt_2
wt_3
hom_1
hom_2
hom_3
B
Chromosome 6
124,758
124,760
124,762
124,764
124,766
124,768
124,770 Kb
[0-8]
[0-20]
Eno2
Lrrc23
Ensembl genes
C
Metaprofile
hom
wt
Mean ChIP/input
A
B
-5Kb
TSS
2.5Kb
5Kb
D
Expression quantiles
0
I
II
III
IV
p 3.6e-1
p 2.3e-1
*p 2.7e-2
***p 7.6e-10
***p 2.6e-16
Mean log2 B/A ratio
2
0
-2
hom
wt

Bianchi et al. Supplementary Fig 5

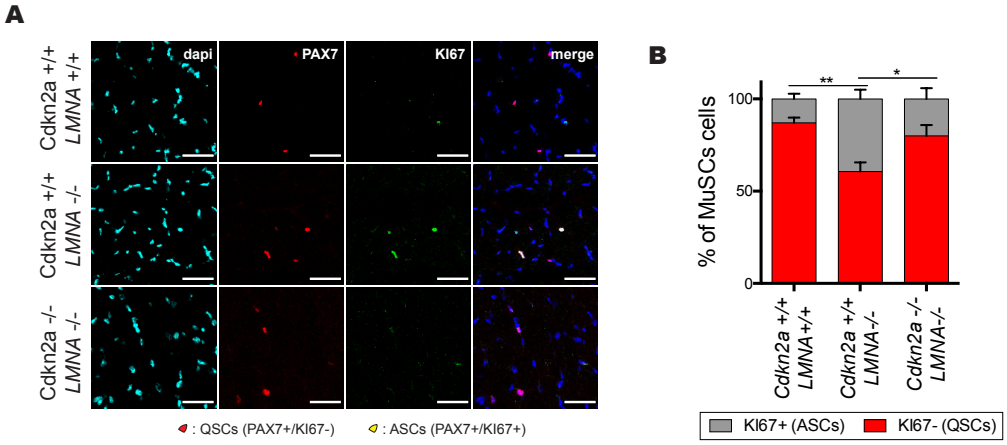



A
Chromosome 6
115,360
115,380
115,400
115,420
115,440
115,460
115,480 Kb
115,500
ChIP-seq
[0-8]
[0-55]
H3K27me3
H3K4me3
Pparg
Gm44079
Ensembl genes
B
140
0
TADs
115,000,000
116,000,000
Chromosome 6
114,800
115,000
115,200
115,400
115,600
115,800
116,000 Kb
ChIP-seq
[0-8]
[0-8]
wt
hom
BAC
CH29-555O5
CH29-101F16
Atg7
Tamm41
Gm17482
Pparg
Mkrn2os
Cand2
Ift122
Tmcc1
Vgll4
Syn2
Timp4
Tsen2
Mkrn2
Raf1
Tmem40
Mbd4
Plxnd1
Efcab12
Ensembl genes


A
CTRL
dapi
OCT4
SSEA4
merge
dapi
TRA1-60
SOX2
merge
EDMD
dapi
OCT4
SSEA4
merge
dapi
TRA1-60
SOX2
merge
B
CTRL1
EpiPS
iPS-C1-2F6
iPS-C1-2F62
iPS-C1-2F63
EDMD
iPS-E1-1D3
iPS-E1-1D4
iPS-E1-2G1
iPS-E1-2G3
% of transcription (log10 scale)
hOCT4
hSOX2
hNANOG
hLIN28
C
LAMIN A/C – CHR.1
R453W
...TTTGTCC/TGGCTGC...
CTRL1 2F6
EDMD 1D3
CTRL1 2F62
EDMD 1D4
CTRL1 2F63
EDMD 2G1
EDMD 2G3
D
Terminal muscular differentiation
CTRL iPS
EDMD iPS
dapi
MF20
dapi
MF20
E
% of transcription (log10 scale)
PAX7
OCT4
F
% of transcription (log10 scale)
MYOG
MHC
N-CAM
G
EpiPS
CTRL-C1 2F6
CTRL-C1 2F63
EDMD-E1D3
EDMD-E1D4
% of transcription (log10 scale)
PPARγ


A
dapi
PAX7
Ph-P38
LAM
merge
wt
het
hom
B
% of MuSCs cells
**
Ph-P38-
Ph-P38+
wt
het
hom
C
dapi
PAX7
γH2AX
LAM
merge
wt
het
hom
D
% of MuSCs cells
**
γH2AX-
γH2AX+
wt
het
hom
E
old vs young mice up-regulated genes
Enrichment score (ES)
Ranked list metric (PreRnaked)
hom
Zero cross at 12597
wt
Rank in Ordered Dataset (*10^3)
Enrichment profile
Hits
Ranking metric scores
F
geriatric vs young mice up-regulated genes
Enrichment score (ES)
Ranked list metric (PreRnaked)
hom
Zero cross at 12597
wt
Rank in Ordered Dataset (*10^3)
Enrichment profile
Hits
Ranking metric scores

Bianchi et al. Supplementary Fig 9



CHAPTER 3.
EARLY POLYCOMB-TARGET DEREGULATIONS IN HUTCHINSON-GILFORD PROGERIA SYNDROME REVEALED BY HETEROCHROMATIN ANALYSIS

Sebesyén Endre[§], Marullo Fabrizia[§], Lucini Federica, Bianchi Andrea, Petrini Cristiano, Valsoni Sara, Olivieri Ilaria, Antonelli Laura, Gregoretti Francesco, Oliva Gennaro, Ferrari Francesco* and Lanzuolo Chiara*

BioRxiv 799668 [Preprint]. October 10, 2019. doi: 10.1101/799668

Submitted

Early Polycomb-target deregulations in Hutchinson-Gilford Progeria Syndrome revealed by heterochromatin analysis

Endre Sebestyén[1§#], Fabrizia Marullo[2§], Federica Lucini[3], Andrea Bianchi[2,4], Cristiano Petrini[1], Sara Valsoni[4,5], Ilaria Olivieri[2], Laura Antonelli[5], Francesco Gregoretti[5], Gennaro Oliva[5], Francesco Ferrari[1,6*] and Chiara Lanzuolo[3,7*]

[1] IFOM, the FIRC Institute of Molecular Oncology, Milan, Italy

[2] Institute of Cell Biology and Neurobiology, National Research Council, Rome, Italy

[3] Istituto Nazionale Genetica Molecolare "Romeo ed Enrica Invernizzi", Milan, Italy

[4] IRCCS Santa Lucia Foundation, Rome, Italy

[5] Institute for High Performance Computing and Networking, Naples, Italy

[6] Institute of Molecular Genetics, National Research Council, Pavia, Italy

[7] Institute of Biomedical Technologies, National Research Council, Milan, Italy

[#] Current address: 1st Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary

§ Equal contribution

* Equal contribution and correspondence: chiara.lanzuolo@cnr.it; francesco.ferrari@ifom.eu

Lead Contact: Chiara Lanzuolo

Running head: Chromatin alterations in HGPS

Keywords: Heterochromatin, Genome structure, Lamin A/C, Progeria, Lamin Associated Domains, Transcription regulation.

Abstract

Hutchinson-Gilford progeria syndrome (HGPS) is characterized by the progressive accumulation of progerin, an aberrant form of Lamin A. This leads to chromatin structure disruption, in particular by interfering with Lamina Associated Domains. Although several cellular and molecular alterations have been characterized, it is still unclear how chromatin structural changes translate into premature senescence in HGPS. Moreover, early events in chromatin remodeling have not been detected so far. We developed a new high-throughput sequencing-based method, named SAMMY-seq, for genome-wide characterization of heterochromatin accessibility changes. Using SAMMY-seq, we detected early stage alterations of chromatin structure in HGPS primary fibroblasts. Of note, these structural changes do not disrupt the distribution of H3K9me3 but are associated with site-specific H3K27me3 variations and transcriptional dysregulation of Polycomb target genes. Our results show that SAMMY-seq represents a novel and sensitive tool to characterize heterochromatin alterations. Moreover, we found that the assembly of lamin associated domains is strictly connected to the correct Polycomb repression, rapidly lost in early HGPS pathogenesis.

In eukaryotic nuclei, DNA shows different levels of compaction recognizable by electron microscopy, known as euchromatin and heterochromatin (reviewed in[1,2]). The more accessible and less condensed euchromatin is generally enriched in expressed genes. Whereas heterochromatin contains highly condensed DNA, including pericentromeric and telomeric regions[3-5], and genomic regions with unique packaging properties targeted by the Polycomb-group proteins (PcG)[6]. These proteins are developmentally regulated transcriptional factors acting as parts of multimeric complexes such as Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2)[7].

Another key player of chromatin organization is the nuclear lamina (NL). NL is mainly composed of A- and B-type lamins, intermediate filaments of 3.5 nm thickness which form a dense network at the nuclear periphery[8,9]. NL preferentially interacts with the genome at specific regions called Lamina Associated Domains (LADs), with sizes ranging from 100Kb to 10Mb[10]. LADs are enriched in H3K9me2 and H3K9me3 histone modifications that are typical of inactive heterochromatic regions. LAD borders are marked by the PRC2-dependent H3K27me3 histone mark[11-13], which is characteristic of inactive PcG-regulated chromatin regions. The ensemble of lamins, chromatin marks and PcG factors around LADs creates a repressive environment[14,15], with heterochromatin and PcG target regions adjacent to each other[16,17]. In line with these observations, we previously demonstrated that Lamin A functionally interacts with PcG[18,19], as later also reported by others[17,20-22].

In physiological conditions, the association or detachment of LADs from the NL are related to the spatiotemporal regulation of gene expression in cell differentiation processes[13,23,24]. The crucial function of these dynamics is attested by an entire class of genetic diseases, called laminopathies, where specific components of the NL are altered[13,21,25]. Among them, Hutchinson-Gilford Progeria Syndrome (HGPS, OMIM 176670) is a rare and fatal disease caused by a *de novo* point mutation in the LMNA gene[26], resulting in a truncated splicing mutant form of Lamin A protein, known as progerin. The disease phenotype is characterized by early

onset of several symptoms resulting in a severe premature aging in pediatric patients. At the cellular level, the accumulation of progerin induces progressive alterations of nuclear shape and cellular senescence[27]. Notably, inhibition of PcG functions can trigger “nuclear blebbing”, an alteration of the nuclear shape typically observed in Lamin A deficient cells[28], suggesting that chromatin architecture and nuclear shape strictly depend on each other. Consistently, LADs organization and PcG-dependent H3K27me3 mark are altered in HGPS[29,30]. However, while extensive disruption of chromatin structure has been reported by Hi-C in late-passage HGPS fibroblasts, this could not be detected in early passage cells, commonly considered a model to identify early molecular alterations in the disease[29].

In recent years, the widespread adoption of experimental techniques based on high-throughput sequencing (NGS) has been instrumental in advancing the knowledge of chromatin structure and function[31 32]. Notably, several of these techniques were originally designed to map active chromatin regions, including ChIP-seq[33], ATAC-seq[34], DNase-seq[35] and MNase-seq[36]. On the other hand, few options are available for genome-wide mapping of heterochromatin and LADs. These include ChIP-seq or DamID-seq[37] targeting NL components or ChIP-seq for the heterochromatin associated histone marks (H3K9 methylation). Such techniques suffer major limitations as ChIP-seq relies on crosslinking, causing technical biases, and antibodies, with several potential issues of sensitivity or cross-reactivity. Furthermore, DamID-seq relies on the exogenous expression of a transgene and it can’t be applied on primary cells or tissues.

Here, we present a novel high-throughput sequencing based method to map lamina associated heterochromatin regions. This technique relies on the sequential isolation and sequencing of multiple chromatin fractions, enriched for differences in accessibility. Our method is robust, fast, easy-to-perform, it does not rely on crosslinking or antibodies, and it can be applied on primary cells. In normal human fibroblasts, we can reliably identify genomic regions with heterochromatin marks (H3K9me3) and LADs.

In early-passage HGPS fibroblasts we found patient-specific changes of chromatin accessibility. Although HGPS chromatin remodeling is not accompanied by H3K9me3 changes, differentially compacted domains affect H3K27me3 distribution and the transcriptional state of PcG regulated bivalent genes. These findings suggest that chromatin structural changes are an early event in HGPS nuclear remodeling and interfere with proper PcG control. We named our technology SAMMY-seq (Sequential Analysis of MacroMolecules accessibilitY), as a tribute to Sammy Basso, the founder of the Italian Progeria Association, a passionate and inspiring advocate of research on laminopathies.

Results

SAMMY-seq allows mapping of lamina associated heterochromatic regions

We developed the SAMMY-seq method for genome-wide mapping of chromatin fractions separated by accessibility. The method is based on the sequential extraction of distinct nuclear fractions[18,38,39] containing: soluble proteins (S1 fraction), DNase-sensitive chromatin (S2 fraction), DNase-resistant chromatin (S3 fraction) and the most condensed and insoluble portion of chromatin (S4 fraction) (Fig. 1a and Supplementary Fig. 1a). We further adapted the method to leverage high throughput DNA sequencing for genome-wide mapping of the distinct chromatin fractions (see Methods).

We first applied SAMMY-seq to 3 independent normal skin primary fibroblast cell lines, originating from 3 different individuals (control samples). The average number of reads in the different fractions was 72 million (S2), 62 million (S3) and 65 (S4) million (Fig. 1b, Supplementary Fig. 1b, Supplementary Table 1), and on average 58% (S2), 54% (S3) and 52% (S4) of the genome was covered by at least one read (Supplementary Fig. 1c). The genome-wide correlation of read coverage confirms a high reproducibility of signal for the S3 and S4 fractions, enriched for more compacted chromatin, whereas the S2 fraction is more variable across biological replicates (Supplementary Fig. 1d). We noticed that correlation computed over genomic bins of variable sizes, ranging from 10Kb to 5Mb, becomes more stable with larger bin sizes (Supplementary Fig. 1e). At 1Mb resolution, the mean Spearman correlation is 0.27 for the S2 fraction, 0.92 for S3, and 0.79 for S4. Visual inspection of read coverage profiles shows megabase-scale "bumps" in S3 and, more prominently, in S4 fractions (Fig. 1b). To detect relative enrichment of specific genomic regions between chromatin fractions, we applied the Enriched Domain Detector (EDD)[40] algorithm to our data. EDD was specifically designed to identify broad regions of enrichment in high-throughput sequencing results. We looked for differentially enriched genomic regions using pairwise comparisons of less accessible vs more accessible fractions used as baseline: S3 vs S2, S4 vs S3 and S4 vs S2 (Fig. 1c). We found a well-defined set of regions (SAMMY-seq domains) consistently enriched in the less accessible

fractions. These domains are conserved across samples, with the S4 vs S2 comparison being the most consistent, and showing an average of 70.18% conservation (Supplementary Fig. 1f). To confirm the robustness of our results we repeated the differential enrichment analysis after down-sampling to 50% of sequencing reads. We obtained similar SAMMY-seq domains, with an average overlap of 80.64 % for S3 vs S2, 82.22 % for S4 vs S3 and 83.42 % for S4 vs S2 across samples (Supplementary Fig 1g). Further down-sampling to 25% of reads still led to an average of 61.92 % overlapping domains across samples and comparisons (Supplementary Fig 1g). SAMMY-seq domains have an average size of 2,26 Mb and cover on average 527 Mb (Supplementary Fig. 1h, i, Supplementary Table 2). Of note, other chromatin fractionation-based NGS methods have been previously described[41,42], however they are not directly comparable to our technology as they adopt different conditions, enzymes and cellular models.

To functionally characterize the SAMMY-seq enriched regions, we compared the genome-wide read coverage with reference chromatin mark profiles for the same cell type. Using data from the Roadmap Epigenomics consortium[43] and other publications[29,44,45] we noticed that SAMMY-seq signal (S4 vs S2 enrichment) is highly consistent across biological replicates (CTRL002, CTRL004, CTRL013), it is inversely correlated with open chromatin marks (ATAC-seq and DNase-seq) and is positively correlated with Lamin A/C and B ChIP-seq signal (Fig. 2a). We confirmed this is a consistent genome-wide pattern using StereoGene kernel correlation[46] for the unbiased comparison of different types of chromatin marks (Fig. 2b). Similarly, when considering histone modifications, we observed a positive association with heterochromatin mark (H3K9me3) and inverse correlation with active chromatin (H3K4me1) as well as Polycomb-regulated chromatin (H3K27me3) (Fig. 2a-b). The latter observation is compatible with previous reports that H3K27me3 is located at the border of heterochromatic LAD regions[11-13,17]. The genome-wide correlation patterns were also consistent when considering individual SAMMY-seq fractions (Supplementary Fig. 2a), confirming that we were progressively enriching for more compact chromatin in the sequential extraction steps.

To quantify the overlap between SAMMY-seq enrichment regions and LADs, we used Lamin ChIP-seq datasets (Lamin A/C and Lamin B)[29,44] and compared them with SAMMY-seq domains (S4 vs S2, S4 vs S3, S3 vs S2). On average 79% of SAMMY-seq domains across different comparisons and samples overlap with Lamin A/C LADs (Supplementary Table 3). All of the comparisons showed a significant overlap (average Jaccard Index 0.56 for Lamin A/C in S4 vs S2), as assessed by randomizing either LADs or SAMMY-seq domains (Fig. 2c and Supplementary Fig. 2b). Interestingly, the overlap with Lamin A/C associated domains was 21% higher than with Lamin B associated domains, suggesting that SAMMY-seq preferentially enriches for regions interacting with Lamin A/C. This is also consistent with the observed higher genome-wide correlation with Lamin A/C compared to Lamin B (Fig. 2b and Supplementary Fig. 2a). As a further confirmation that SAMMY-seq specifically detects the most compact heterochromatin, such as the lamina associated heterochromatic regions, we observed that SAMMY-seq domains have an even higher overlap with regions defined as both LADs and H3K9me3 enriched (77.84% average overlap), rather than considering each of the two markers independently (64.9% and 42.69% average overlap, respectively) (Supplementary Fig. 2c).

SAMMY-seq detects chromatin changes in early passage progeria fibroblasts.

We applied SAMMY-seq on early passage HGPS fibroblasts, a cellular model to identify early molecular alterations in Progeria, as progerin accumulates during passaging (Supplementary Fig. 3a). According to recent literature, primary HGPS fibroblasts entering senescence show alterations in DNA interactions with Lamin A/C[29]. However, at early passages Hi-C experiments for genome-wide mapping of 3D chromatin architecture failed to detect alterations in the pairwise contact frequency between genomic loci. These were evident only at later-passage cells, representing more advanced stages of senescence, when cells exhibit severe nuclear blebs and inhibited proliferation[29].

We confirmed that early-passage HGPS fibroblasts do not show senescence-associated features such as beta-galactosidase positivity (Supplementary Fig. 3b), reduction in the

proliferation rate (Supplementary Fig. 3c), changes in nuclear area and morphology (Supplementary Fig. 3d, e, f). We used SAMMY-seq to investigate chromatin changes in early-passage (passage number from 10 to 12) skin fibroblasts from 3 independent HGPS patients (Supplementary Fig. 1a, b). Compared with analysis showing a consistent overlap (S4 vs S2) across control replicates, in HGPS patients the SAMMY-seq domains appeared scattered and with generally lower signal-to-noise ratio (Fig. 3a). This would be compatible with a general loss of chromatin organization, resulting in more random distribution of genomic regions among the different fractions. Pairwise comparisons confirmed that SAMMY-seq domains are largely overlapping among control samples (average Jaccard Index 0.67) whereas this overlap is lost in progeria samples, most notably in the S4 vs S2 comparison (average Jaccard Index 0.11) (Fig. 3b and Supplementary Table 4). Moreover, SAMMY-seq enrichment in control samples is matching lamins and H3K9me3 ChIP-seq enrichment, whereas this pattern is lost in HGPS SAMMY-seq samples (Fig. 3a). We confirmed this is a genome-wide trend by analyzing chromatin marks at SAMMY-seq domain borders (Fig. 3c). SAMMY-seq domains from control samples (S4 vs S2, S4 vs S3, S3 vs S2) correspond to regions normally enriched in Lamin A/C, Lamin B and H3K9me3 ChIP-seq reads, as opposed to flanking regions outside of domain borders. Progeria SAMMY-seq domains (S3 vs S2) are similar to controls. However, when considering more compacted S4 fraction, the association to reference LADs and heterochromatin marks is lost in 2 patients for the S4 vs S2 and 3 patients for the S4 vs S3 comparisons, respectively (Fig. 3c). Here, both the lamins and the H3K9me3 signals are flattened out suggesting no specific pattern.

Early changes of chromatin structure in HGPS fibroblasts is not accompanied by alterations of H3K9me3 patterns

To further investigate how changes in chromatin accessibility in early-passage HGPS fibroblasts affect the epigenetic regulation, we examined H3K9me3 distribution by ChIP-seq on the same set of control and HGPS fibroblasts analyzed by SAMMY-seq (Fig. 4a). Visual exploration of

sequencing reads distribution revealed that, despite chromatin remodeling was detected by SAMMY-seq, H3K9me3 profiles were unaffected in HGPS. H3K9me3 enriched regions are largely conserved between controls and HGPS cells as shown by the genome-wide overlap (between 69% and 80%) (Fig. 4b). Analysis of H3K9me3 patterns at SAMMY-seq domain borders (Supplementary Fig. 4a) and Jaccard index score (Fig. 4c and Supplementary Fig. 4b) confirmed that only controls exhibit a correlation between SAMMY-seq and H3K9me3.

In line with genomic data, at cellular level early-passage HGPS cells and controls showed comparable H3K9me3 global levels measured by western blot (Supplementary Fig. 4c, d). Immunofluorescence analysis with H3K9me3 antibody revealed that heterochromatin foci were detected in both controls and early-passage HGPS fibroblasts with high variability across samples (Fig. 4d) but without significant differences in their number per cell (Supplementary Fig. 4e). Instead, their intranuclear localization was different, with a shift towards the nuclear center in HGPS (Fig. 4d, e), supporting the hypothesis of heterochromatin detachment from the nuclear periphery (Fig. 3). Notably, as previously shown[30], later passages (p20) HGPS fibroblasts showed an aberrant pattern of H3K9me3 foci (Fig. 4d), further indicating that chromatin remodeling in HGPS precedes H3K9me3 loss.

Large-scale chromatin structural changes do not affect expression of LAD-associated genes

Using RNA-seq on the same set of control and HGPS fibroblasts we assessed the expression at the gene or transcript level detecting respectively 260 differentially expressed genes and 256 transcripts, suggesting that early stages of senescence in HGPS cells do not cause large scale changes in the transcriptome (Supplementary Fig. 5a). To clarify whether chromatin structural changes detected with SAMMY-seq have an effect on transcription, we examined the expression of protein coding genes located in SAMMY-seq domains. As expected[16], genes within SAMMY-seq domains (S4 vs S2) have lower expression levels than those outside in all control samples (Supplementary Fig. 5b). More specifically, in the genomic regions flanking

SAMMY-seq domain borders we noticed a clear pattern with a transition from lower expression within the domain to higher expression outside in all of the control samples (Supplementary Fig. 5c). A similar pattern was observed in HGPS samples when considering the same genomic regions (Supplementary Fig. 5b, c). This suggests that heterochromatin detachment from NL seen in HGPS (Fig 3 and 4e) was not sufficient to trigger gene expression deregulation inside SAMMY-seq domains, where H3K9me3 is also largely unchanged as shown above (Fig. 4a, b). Further confirming this hypothesis, the transcriptional activity within H3K9me3 domains of each sample is repressed in HGPS as expected (Supplementary Fig. 5d).

HGPS chromatin remodeling affects expression of PcG regulated bivalent genes

As we noticed that each HGPS sample has unique SAMMY-seq domains (Fig.3), we compared individual HGPS expression profiles against the group of three controls to account for sample-specific patterns (see Online methods). To verify which biological functions are affected we performed a pathway enrichment analysis using the MSigDB database[47] and selected the pathways commonly deregulated in all HGPS samples (Fig. 5a). Many of them are related to stem cells and cancer, in line with the hypothesis that accelerated aging affect stem cell niches[48]. Interestingly, we also found several pathways related to PcG regulation (H3K27me3 and PRC2) (Fig. 5a). To further investigate PcG role in early passage HGPS fibroblasts, we examined the total amounts of PRC1-subunit Bmi1 and PRC2-subunit Ezh2 (Supplementary Fig. 6a, b), as well as H3K27me3 levels (Supplementary Fig. 6c, d), which showed no significant differences, despite some level of inter-individual variability. We also checked intranuclear architecture of Ezh2 (Supplementary Fig. 6e, f) and H3K27me3 foci (Supplementary Fig. 6g, h) and we found that their number and size do not change at early-passage HGPS cells.

Finally, we performed ChIP-seq to dissect H3K27me3 distribution at the genome-wide level. By visual inspection we noticed that while the main peaks were present in both controls and HGPS samples, H3K27me3 signal was spread over flanking regions in HGPS (Fig. 5b). The progeria-associated differences in H3K27me3 profile were also confirmed by multi-dimensional scaling of ChIP-seq data (Fig. 5c), showing that control samples are close to each other and distant from the progeria samples, the latter also showing larger between sample distances. Analysis of TSS distribution of H3K27me3 mark revealed a slight decrease of signal in HGPS samples (Supplementary Fig. 6i). This difference was even more evident when considering the subset of upregulated genes in HGPS, where we observed a drop in H3K27me3 signal (Fig. 6a).

To further dissect the role of chromatin regulators in the disrupted expression patterns, we examined the overlap of SAMMY-seq domains with the chromatin states of normal skin fibroblasts[43]. SAMMY-seq domains of control samples are mostly overlapping "Quiescent" or "Heterochromatin" states (Fig. 6b), whereas SAMMY-seq domains of HGPS cells comprise regions transcribed or regulated by PcG proteins in normal skin fibroblasts. To analyze the correlation between gene expression, chromatin states and altered genome structure, we grouped SAMMY-seq domain genes on the basis of their chromatin state in normal skin fibroblasts and examined their expression difference between control and HGPS cells (Fig. 6c and Supplementary Fig. 6k). This analysis revealed that PcG regulated and bivalent genes are especially deregulated at the transcriptional level in progeria with respect to controls. These findings overall suggest that early chromatin remodelling has an impact on a subset of PcG targets, the bivalent genes, more susceptible to variations of PcG occupancy[49] (Fig. 6d).

To verify the hypothesis of a PcG-progerin crosstalk we used the proximity ligation assay (PLA), which detects interactions between proteins in close proximity (< 30nm)[50]. We found that Lamin A interacts with PcG proteins in control or HGPS fibroblasts (Fig. 6 e, f), and that progerin, expressed in late-passage HGPS, interacts with both Ezh2 and Bmi1 (Fig. 6 e, f). To investigate the aberrant PcG-progerin interaction in HGPS progression we tested PcG protein

compartmentalization, in early and late-passage HGPS (Fig. 6g, h). Comparing controls and early-passage HGPS fibroblasts we did not observe major differences in PcG nuclear distribution across chromatin fractions. On the other hand, PcG proteins were predominantly localized in the S4 fraction at late passage HGPS, supporting the previous observation that PcG regulated regions shift towards the insoluble S4 fraction in HGPS (Fig. 6b). Altogether, our results indicate that increased amount of progerin may alter PcG proteins nuclear compartmentalization and function.

Discussion

Within the eukaryotic cell nucleus, chromatin acquires a specific structure which is fundamental for the correct spatiotemporal regulation of gene expression[51]. In particular, heterochromatin is characterized by highly compact structure and specific nuclear compartmentalization. Important heterochromatin organizing regions include LADs, along with pericentromeric and telomeric regions, which are crucial for preserving chromosome structure[3,4,10]. Alterations in heterochromatin are associated with developmental defects and cancer, while its proper conformation is a hallmark of healthy cells[52,53]. As such, reliable methods to characterize heterochromatin and its alterations are a crucial need for the biomedical scientific community.

Here we present a new technology called SAMMY-seq, for mapping lamina associated heterochromatic regions (Fig. 1). The protocol is based on the sequential extraction of multiple chromatin fractions, corresponding to increasingly compacted and less accessible chromatin regions, which are identified by the relative comparison of high-throughput sequencing reads from different fractions (Fig. 2). We show that SAMMY-seq can reproducibly and reliably identify lamina associated heterochromatic regions as the less accessible portions of chromatin in control wild-type fibroblasts. SAMMY-seq represents a significant improvement in the field of chromatin characterization as it provides novel information complementary and not overlapping with other high-throughput sequencing based methods commonly used to study chromatin structure and function. Of note, our protocol is overcoming several major limitations of other methods for mapping lamina associated heterochromatic regions. First of all, the procedure can be applied on primary cells, as it doesn't require exogenous gene expression as in DamID-seq. Then, SAMMY-seq does not involve chemical modifications of chromatin, which might cause artifacts and high background after sequencing[54,55]. Moreover, it does not rely on antibodies for enriching specific chromatin fractions, thus avoiding issues related to antibody specificity, production, lot-variability and cross-reactivity. This is particularly important when studying chromatin changes in cells where protein levels of chromatin associated factors could be

altered, thus allowing more flexibility in terms of experimental design compared to antibody-based techniques. Currently, we are successfully applying SAMMY-seq to many other experimental systems such as human epithelial and myogenic cells, human lymphocyte subpopulations or murine muscle satellite stem cells, in some cases scaling down to 10.000 cells (data not shown).

Emerging studies revealed that chromatin tethering at nuclear lamina and histone marks deposition are independent processes[56,57] and that nuclear relocation is not sufficient to trigger a transcriptional switch[57-59]. Overall, our findings support these assumptions in HGPS model, further demonstrating that the pathological chromatin remodeling trigger a transcriptional dysfunction at PcG bivalent target, more sensitive to the epigenetic environment. In particular, by applying our SAMMY-seq technology on early passage HGPS fibroblasts we were able to detect chromatin structure alterations, in terms of accessibility changes (Fig. 3), despite the recent failure to do so by Hi-C[29]. Indeed Hi-C is designed to measure pairwise interactions between genomic loci, but they are not able to distinguish the intra-nuclear compartments where such interactions occur[60]. Notably, changes in chromatin accessibility seen in early-passage HGPS are not accompanied by H3K9me3 alterations or transcriptional deregulation in the same genomic regions (Fig. 4, Supplementary Fig. 5b, c), further indicating that chromatin remodeling detected by SAMMY-seq precedes H3K9me3 decrease observed at later passages (Fig. 4d). On the other hand, we found that chromatin accessibility changes in HGPS functionally affect PcG-regulated genes, by interfering with H3K27me3 genomic distribution (Fig. 5). In particular, the bivalent genes, involved in stemness maintenance and cell identity specification, are mostly affected in early-passage HGPS fibroblasts (Fig. 6), providing a mechanism for the previously postulated idea that the reduced potential for tissue homeostasis, commonly associated with progeria, is due to a misregulation of stem cell compartments[48].

Methods

Cell cultures

Primary fibroblast cell lines were cultured in DMEM High glucose with glutamax supplemented with 15% FBS and 1% Pen/Strep. HGADFN164 (HGPS164), HGADFN167 (HGPS167), HGADFN169 (HGPS169), HGADFN188 (HGPS188), HGADFN271 (HGPS271) human dermal fibroblasts derived from HGPS patients were provided by the Progeria Research Foundation (PRF). AG08498 (CTRL001) and AG07095 (CTRL002) human dermal fibroblasts were obtained from the Coriell Institute. Preputial fibroblast strain #2294 (CTRL004) was a generous gift from the Laboratory of Molecular and Cell Biology, Istituto Dermopatico dell'Immacolata (IDI)-IRCCS, Rome, Italy", while control dermal fibroblast CTRL013 was kindly provided by the Italian Laminopathies Network.

Histochemistry, immunofluorescence assay and PLA analysis

Beta-galactosidase assay: cells were fixed at room temperature for 10min in paraformaldehyde at 4%, washed twice with 1×PBS, and incubated with fresh staining solution at 37 °C for 16 hrs in dark room. Then cells were washed twice with 1×PBS, overlaid with 70% glycerol/PBS, and photographed.

BrdU (5-Bromo-2¢-deoxy-uridine, Sigma B9285) labeling: cells were grown for 8 hrs in the presence of 50microM of BrdU, and then fixed in 4% PFA. After Triton X-100 treatment, cells were incubated for 2 min at RT in 0.07N NaOH, briefly rinsed twice in PBS and blocked in PBS/1% BSA. Reaction with BrdU antibody (1:10, Becton Dickinson 347580) was performed at room temperature for 1 hour in a PBS solution containing BSA 1%.

Immunofluorescence assay: coverslips were fixed with paraformaldehyde at 4% in PBS for 10 minutes. Then, cells were permeabilized with 0,5% triton X-100 in PBS and non-specific signals were blocked with 1% BSA in PBS for 30 minutes at room temperature. The following antibodies were used: Bmi1 (Millipore 05-637, mouse) diluted 1:100; Lamin A/C (Santa Cruz sc-6215,

goat) diluted 1:200; Ezh2 (Cell signaling AC22 3147S, mouse) diluted 1:100; H3K9me3 (Abcam ab8898, rabbit) diluted 1:500; H3K27me3 (Millipore 07-449, rabbit) diluted 1:100. Incubation was performed at 12-16 hrs at 4°C or at room temperature for 2 hours for Lamin A/C. Primary antibodies were diluted in a PBS solution containing BSA 1%. Cells were stained with appropriate secondary antibodies, diluted 1:200 for 1 h at room temperature. Washes were done in PBT. As secondary antibodies we used Alexa Fluor 488 Donkey anti-mouse IgG (715-545-150), Alexa Fluor 594 Donkey anti-goat IgG (705-585-003) from Jackson ImmunoResearch Laboratories and Alexa Fluor 647 Chicken anti-goat IgG (Invitrogen, A21469). Finally, DNA was counterstained with DAPI, and glasses were mounted in Vectashield Antifade (Vector Laboratories) or ProLong Gold Antifade Reagent (Invitrogen). For PLA experiments, coverslips were fixed first with paraformaldehyde at 4% in PBS for 10 minutes. Then, cells were permeabilized with 0,5% triton X-100 in PBS and blocked with 1% BSA in PBS for 1 hour at room temperature. Incubation with progerin (Alexis human mAb, 13A4, ALX-804-662-R200) diluted 1:20 or Lamin A/C (Santa Cruz sc-6215, goat) diluted 1:200 was performed for 12-16 hrs at 4°C. After PBT washes, cells were incubated with Ezh2 (Cell signaling 4905S, rabbit 1:100) or Bmi1 (Abcam ab85688 rabbit 1:100) for 12-16 hrs at 4°C. Primary antibodies were diluted in a PBS solution containing BSA 1%. Finally, detection of protein interactions was performed using the Duolink system (Sigma) following manufacturer's instructions.

Image processing and analysis

Fluorescent images were taken with a Nikon ECLIPSE 90i microscope (1006 objective), equipped with a digital camera (Nikon Coolpix 990) and NIS Element software or with confocal Leica SP5 supported by LAS-AF software (version 2.6.0). Confocal image size was 1024 · 1024. PLA blobs quantification was performed using Cell Profiler 2.0 as previously described[18]. In order to automatically detect and quantify PcG bodies in fluorescence cell images we developed an algorithm that implements a method derived from a previous work [65,] with variants and adaptations. The algorithm performs the segmentation of cell nuclei and the detection of PcG

bodies of each image. It then measures the area and eccentricity of nuclei as well as the number and the area of the PcG bodies. It also provides for each PcG body a measure of its closeness to nuclear periphery (proximity). The algorithm has been implemented in MATLAB following this pseudocode:

```
[nuclei,avg_Bmi1] = nuclei_seg(I_dapi,lamin, I_Bmi1) %Performs nuclei segmentation
I_avg = imfilter(I_Bmi1,fspecial('average', [7,7]))
I_filt = I_Bmi1 - I_avg
[pcg,t] = isodata_thresh(I_filt,avg_Bmi1) %Applies ISODATA method to
separate PcG bodies from nuclei regions
nuclei_CC = bwconncomp(nuclei)
nuclei_L = labelmatrix(nuclei_CC)
compute area for each nucleus object in nuclei_CC
exclude nuclei whose area is less than 10% of mean areas
{NCL}_M <- identified nuclei
for each nucleus m in {NCL}_M
     NCL_m.metric = 4*pi*NCL_m.area/NCL_m.perimeter^2 %compute eccentricity
     of nucleus NCL_m to provide an estimate of its 'roundness'
     NCL_m.PcG = pcg .* {NCL}_M %positions of detected PcG bodies within
     the nucleus NCL_m
     NCL_m.PcG = bwareaopen(NCL_m.PcG,3,8)
     PcG_CC = bwconncomp(NCL_m.PcG)
     compute number N of PcG bodies in PcG_CC
     for each PcG body n in NCL_m.{PcG_CC}_N
          NCL_m.PcG_n.area = area of PcG body
          NCL_m.PcG_n.mindist = minimum distance of PcG body from the
          nuclear periphery
          NCL_m.PcG_n.closestP = the point on nuclear periphery closest
          to PcG body
          NCL_m.PcG_n.distC2P = distance of nuclear centroid from
          closestP
          NCL_m.PcG_n.proximity = NCL_m.PcG_n.mindist / NCL_m.PcG_n.distC2P
     endfor
    endfor
```

$I_{dapi,lamin}$ is the image obtained by the sum of both images showing the fluorescence of nucleus and lamin.

The function *nuclei_seg* performs a partition of cell image $I_{dapi,lamin}$ in nuclei regions and background implementing a region based segmentation algorithm[66]. `nuclei` is the binary image defining nuclei regions while avg_{Bmi1} is the mean intensity value of the nuclei regions in the image I_{Bmi1} that shows the fluorescence of PcG bodies.

In order to better enhance PcG areas we subtract from the original image I_{Bmi1} its smoothed version obtained by applying an averaging filter of size 7, producing the image I_{filt}.

The function *isodata_thresh* implements the ISODATA classification algorithm[67] and uses relevant values computed by *nuclei_seg* function in order to extract PcG bodies from the nuclei regions. It sets the initial threshold value of ISODATA method as avg_{Bmi1}. `pcg` is the binary image defining PcG bodies.

Reconstructions of nuclei are obtained through the connected components algorithm (*bwconncomp* MATLAB function, using a connectivity of 6). Nuclei are then labeled by applying the *labelmatrix* MATLAB function so they can be easily separated from each other.

The algorithm computes the area of each nucleus, discarding objects whose area is less than 10% of mean areas which are just noise.

The algorithm uses the *bwareaopen* function in order to discard too small detected PcG objects which are probably just noise.

Reconstructions of PcG bodies are obtained through the connected components algorithm (*bwconncomp* MATLAB function, using a connectivity of 6).

The algorithm computes the number of PcG bodies, the area of any PcG body, and the minimum euclidean distance of any PcG body from the nuclear periphery.

Protein extraction and Western Blot analyses

Total proteins were prepared by resuspending $1x10^6$ cells in extraction buffer (50 mM TrisHCl pH 7.6; 0.15 M NaCl; 5 mM EDTA; 16 Protease Inhibitors; 1% Triton X-100). One 40sec pulse of sonication (UP100H manual sonicator, Hielscher) at 40% amplitude was performed to allow dissociation of protein from chromatin and solubilization. Extracts were analyzed by SDS-PAGE using an 8% gel (37.5:1 Acryl/Bis Acrylamide). The following primary antibodies were used: Beta-Actin (Santa-Cruz sc1616, rabbit 1:4000), H3 total (Abcam ab1791, rabbit 1:6000), Lamin A/C (Santa Cruz sc-6215, goat 1:4000), Lamin B (Santa Cruz sc6216, goat 1:2000), progerin (13A4 mouse, Abcam 66587, mouse 1:1000), Ezh2 (AC22 Cell Signaling 3147S, mouse 1:1000), Bmi1 (D42B3 Cell signaling, rabbit 1:1000), H3K9me3 (Abcam ab8898, rabbit 1:1000), H3K27me3 (Millipore 07-449 rabbit 1:1000).HRP-conjugated secondary antibodies were revealed with the ECL chemiluminescence kit (Thermo Fisher Scientific).

Chromatin fractionation

Chromatin fractionation was carried out as described in[38] with minor adaptions. Briefly, 4 million fibroblasts were washed in PBS 1X, and extracted in cytoskeleton buffer (CSK: 10 mM PIPES pH 6,8; 100 mM NaCl; 1 mM EGTA; 300 mM Sucrose; 3 mM $MgCl_2$; 1X protease Inhibitors by Roche Diagnostics; 1 mM PMSF) supplemented with 1 mM DTT and 0,5% Triton X-100. After 5 min at 4°C the cytoskeletal structure was separated from soluble proteins by centrifugation at 3000 rpm for 3 min, and the supernatant was labeled as S1 fraction. The pellets were washed with an additional volume of cytoskeleton buffer. Chromatin was solubilized by DNA digestion with 25U of RNase–free DNase (Turbo DNAse; Invitrogen AM2238) in CSK buffer for 60 min at 37°C. To stop digestion, ammonium sulphate was added in CSK buffer to a final concentration of 250 mM and, after 5 min at 4°C samples were pelleted at 5000 rpm for 3 min at 4°C and the supernatant was labeled as S2 fraction. After a wash in CSK buffer, the pellet was further extracted with 2M NaCl in CSK buffer for 5 min at 4°C, centrifuged at 5000 rpm 3 min at 4°C and the supernatant was labeled as S3 fraction. This treatment removed the majority of histones from chromatin. After 2 washing in NaCl 2M CSK, the pellets were solubilized in 8M urea buffer to remove any remaining protein component by applying highly denaturing conditions. This remaining fraction was labeled as S4. Supernatants from each extraction step were quantified and analyzed by SDS-PAGE and immunoblotting. Anti-tubulin alpha (Sigma T5168, mouse 1:10000), H3 (Abcam ab1791, rabbit 1:4000), Beta-Actin (Santa-Cruz sc1616, rabbit 1:4000), Lamin A/C (Santa Cruz sc-6215, goat 1:4000), Lamin B (Santa Cruz sc-6216, goat 1:2000), progerin (13A4 mouse, Abcam 66587, mouse 1:1000), Ezh2 (AC22 Cell Signaling 3147S, mouse 1:1000), Bmi1 (D42B3 Cell signaling, rabbit 1:1000) were used as primary antibodies. HRP-conjugated secondary antibodies were revealed with the ECL chemiluminescence kit (Thermo Fisher Scientific).

DNA sonication and sequencing for chromatin fractionation

For DNA fractionation, we took an aliquot corresponding to 30% of volume of S2, S3, S4 fractions and added TE buffer to reach a final volume of 300µL. After incubation with RNAse A (Roche) (90 minutes at 37°) and Proteinase K (Sigma) (150 minutes at 55°), DNA was extracted from beads by standard phenol/chloroform extraction, precipitated and resuspended in 25µl milliQ H_2O. After Nanodrop (260/280 = 1,7-1,9; 260/230 ≥ 2) and Qubit HS DNA quantification, we added H_2O to a final volume of 105µL. Then samples were transferred to 96 well plates and sonicated 4 times with Bioruptor sonicator (10 minutes 30 seconds ON - 30 seconds OFF, High Power). The DNA profiles were finally checked by capillary electrophoresis (Agilent 2100 Bioanalyzer). Finally, DNA libraries were prepared by using NuGEN Ovation Ultralow Library Prep System kit and then sequenced using an Illumina HiSeq 2500 instrument according to manufacturer's instructions (Illumina).

Chromatin immunoprecipitation and sequencing

Cells were cross-linked with 1% HCHO for 12 minutes at room temperature, lysed and chromatin sheared. IP were performed overnight on a wheel at 4° with 2.4µg of H3K9me3 antibody (ab8898, Abcam) or 2µg of H3K27me3 (07-449, Millipore). The following day, antibody-chromatin immunocomplexes were loaded onto Dynabeads Protein G (Invitrogen 10004D). The bound complexes were washed once in Low Salt Solution (0,1% SDS, 2mM EDTA, 1% Triton X-100, 20mM Tris pH 8, 150 mM NaCl), once in High Salt Solution (0,1% SDS, 2mM EDTA, 1% Triton X-100, 20mM Tris pH 8, 500 mM NaCl), once again in Low Salt Solution and once in Tris/EDTA 50 mM NaCl. Crosslinking was reversed at 65°C overnight in Elution Buffer (50mM Tris pH 8, 20mM EDTA, 1%SDS), DNA was purified by standard phenol/chloroform extraction, precipitated and resuspended in 30µl of 10mM Tris pH 8. ChIP efficiency was tested by qPCR reactions, performed in triplicate using QuantiTect SYBR Green master mix (Qiagen) on a StepOnePlus™ Real-Time PCR System (Applied Biosystems). Relative enrichment was calculated as IP / Input ratio. Libraries for sequencing were created using the automation instrument Biomek FX (Beckman Coulter), then qualitatively and

quantitatively checked using Agilent High Sensitivity DNA Kit (Agilent Technologies, 5067-4627) on a Bioanalyzer 2100 (Agilent Technologies). Libraries with distinct adapter indexes were multiplexed and, after cluster generation on FlowCell, were sequenced for 50bp in the single read mode on a HiSeq 2000 sequencer at the IEO Genomic Unit in Milan.

RNA sequencing

For high-throughput sequencing, cDNA libraries were prepared from total RNA, extracted with Trizol, by using Illumina TruSeq Stranded Total RNA Kit with Ribo-Zero GOLD. cDNA fragments of ~300 bp were purified from each library and were sequenced for 125bp, using an Illumina HiSeq 2500 instrument according to manufacturer's instructions (Illumina).

SAMMY-seq sequencing read analysis

Sequencing reads were aligned to the hg38-noalt reference human genome available in the bcbio-nextgen pipeline, using bwa aln[61] (version 0.7.12) with options -n 2 -k 2 and saved the results in sam format with bwa samse. The sam files were converted to bam and name sorted with samtools[62] (version 1.3.1). We marked PCR duplicates using the biobambam2 toolset[63] (version 2.0.54). We discarded reads mapping to non-autosomal chromosomes, PCR duplicates, qcfail, multimapping and unmapped reads with samtools. We converted the bam files to bedgraph using bedtools[64] (version 2.25.0) and bedgraph to bigWig using UCSC's bedgraphToBigWig tool (version 4) for reads distribution visualization. Read coverage was normalized by the total sequencing library size, before converting bedgraph files to bigWig. Sample quality was assessed using fastqc (version 0.11.5) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). For the down sampling analysis, we sampled the raw fastq files to 50% and 25% using the seqtk sample command (https://github.com/lh3/seqtk) and ran the same pipeline, as above.

ChIP-seq sequencing read analysis

Sequencing reads were aligned, filtered, converted and quality checked using the same tools as for the SAMMY-seq reads. Before sequence alignment we did an additional trimming step, using Trimmomatic[65] (version 0.32) in single end mode for the H3K9me3 data. We used the Enriched Domain Detector (EDD)[40] tool to call H3K9me3 peaks, with the following options: --fdr 0.1 –gap-penalty 10 –bin-size 100 –write-log-ratios –write-bin-scores and also excluding blacklisted genomic regions containing telomeric, centromeric, and certain heterochromatic regions[66]. We also changed the required_fraction_of_informative_bins parameter to 0.98. We calculated genome wide ChIP-seq signal for H3K9me3 and H3K27me3 data, using the SPP package (version 1.15.4)[67]. We imported bam files into the R (version 3.3.1) statistical environment, and selected informative reads with the get.binding.characteristics and select.iformative.tags functions, removed anomalous positions with extremely high number of reads using the remove.tag.anomalies function, and calculated the differential signal, smoothed by a Gaussian kernel, using the get.smoothed.tag.density function with the default bandwidth and tag.shift parameters. In the case of H3K27me3 data, we also set the scale.by.dataset.size = TRUE parameter for the get.smoothed.tag.density function.

MDS analysis was done using the plotMDS function of edgeR (version 3.24.3)[68] after importing the 10k genomic region read counts for all samples into the R (version 3.5.1) statistical environment. We dropped all genomic regions where the log count-per-million (cpm) value did not reach at least 1 in at least 2 samples. We calculated the normalization factor of samples using the calcNormFactors and estimated dispersion using the estimateDisp function of edgeR. We used the top = 1000 and gene.selection = “common” parameters for the plotMDS function. The gene.selection parameter here does not refer to genes, but the imported genomic regions.

We used deepTools (version 3.2.1) [69] to visualize the ChIP-seq signal around TSS regions, with the following parameters: --referencePoint TSS -b 5000 -a 5000 –missingDataAsZero –averageTypeBins median –binSize 100.

Literature data processing

We collected publicly available datasets from the following sources: ATAC-seq from[45], Lamin A/C ChIP-seq from[29], Lamin B ChIP-seq from[44]. Sequencing reads were aligned, filtered and converted using the same tools as for the SAMMY-seq reads. We calculated the genome wide ChIP-seq signal for Lamin A/C and Lamin B data, using the SPP package (version 1.15.4)[67]. We imported bam files into the R (version 3.3.1) statistical environment, and selected informative reads with the get.binding.characteristics and select.informative.tags functions, removed anomalous positions with extremely high number of reads using the remove.local.tag.anomalies function, and calculated the differential signal, smoothed by a Gaussian kernel, using the get.smoothed.tag.density function with the default bandwidth and tag.shift parameter. We used the filtered read count as the genome-wide signal for the ATAC-seq sample.

Additionally, we downloaded from Roadmap Epigenomics[43] the genome-wide signal coverage tracks for the DNAse-seq, as well as histone marks ChIP-seq datasets (H3K9me3, H3K27me3, H3K4me1, H3K36me3, H3K27ac, H3K4me3) for the foreskin fibroblast sample E055. We converted bigwig files to bedgraph with the bigWigToBedGraph tool, lifted over genomic coordinates from hg19 to hg38 with the liftOver tool and converted back bedgraph to bigwig with the bedGraphToBigWig tool using the UCSC toolkit[70] (version 4).

We calculated genome-wide correlations between SAMMY-seq samples and all of the public datasets listed and processed above using StereoGene[46] (version 1.73). StereoGene uses kernel correlation to calculate genome-wide correlation for spatially related but not completely overlapping features irrespective of their discrete or continuous nature.

We called lamina associated domains with the EDD tool. For the Lamin A/C dataset we used the following options: –gap-penalty 25 –bin-size 200 –write-log-ratios –write-bin-scores, and for the Lamin B dataset we used: –gap-penalty 5 –bin-size 100 –write-log-ratios –write-bin-scores. We also changed the required_fraction_of_informative_bins parameter to 0.98.

SAMMY-seq domain calling, signal calculation, overlap analysis and visualization

We performed relative comparisons of SAMMY-seq fractions within each sample using EDD, optimized to call very broad enrichment domains. EDD was originally designed for lamin ChIP-seq data, comparing IP and input samples. As SAMMY-seq data also shows broad enrichment regions, we used EDD to select significantly enriched SAMMY-seq domains by comparing less accessible fractions to more accessible ones (S3 vs S2, S4 vs S3 and S4 vs S2 comparisons) in each sample, with the following options: --gap-penalty 25 --bin-size 50 --write-log-ratios --write-bin-scores and also excluding blacklisted genomic regions containing telomeric, centromeric, and certain heterochromatic regions[66]. We also changed the required_fraction_of_informative_bins parameter to 0.98. We used the same set of parameters for the down sampled data, except changing –bin-size to 100 for 50% down sampling or 200 for 25% down sampling.

Additionally, we calculated the genome wide differential signal for all comparisons, using the SPP package (version 1.15.4)[67]. We imported bam files into the R (version 3.3.1) statistical environment, and selected informative reads with the get.binding.characteristics and select.informative.tags functions, removed anomalous positions with extremely high number of reads using the remove.local.tag.anomalies function, and calculated the differential signal, smoothed by a Gaussian kernel, using the get.smoothed.tag.density function with the default bandwidth parameter and tag.shift = 0.

We used bedtools[64] (version 2.25.0) and the bedtools jaccard or bedtools fisher command to calculate Jaccard Index or Fisher-test p-values for the various overlap analyses. We randomized SAMMY-seq domains and LADs for the SAMMY-seq – LAD overlap analysis using the bedtools shuffle with -noOverlapping option and also excluding blacklisted genomic regions containing telomeric, centromeric, and certain heterochromatic regions[66] with the -exclude option.

We used the gViz[71] and karyoploter[72] Bioconductor packages to visualize SAMMY-seq read coverages, differential signals and domains. We used ggplot2[73] for additional plotting, and the GAM method[74] from the mgcv package (version 1.8-12) for smoothing data while plotting

SAMMY-seq border region signals.

RNA-seq sequencing read analysis

Transcript and gene level quantification was done with Kallisto[75] (version 0.43.0) to estimate transcript level read counts and TPM (Transcripts Per Million) values. The GENCODE v27 annotation was used to build the transcriptome index. Kallisto was run with the --bias option to perform sequence based bias correction on each sample. We used sleuth[76] (version 0.29.0) to calculate gene or transcript level differential expression comparing all three controls against all three HGPS samples, where the linear model used in sleuth included the sample type (control or HGPS), sex and library id. Additionally, we analyzed transcript level differential expression, comparing the three controls with a single progeria sample, where the linear model used in sleuth included the sample type (control or HGPS) and sex. For both cases, we used sleuth's Wald test to calculate p and q-values. After differential expression, we grouped transcripts according to MSigDB pathways and aggregated transcript p-values using the Lancaster aggregation method from the aggregation R package (version 1.0.1) motivated by a recently described analysis pipeline in[77]. Lancaster p-values were corrected for multiple testing using the Benjamini-Hochberg method. Alternatively, we aggregated transcript p-values based on their Roadmap chromatin states and differential expression direction. We used two different criteria: for the Tx, TxWk, ReprPC, ReprPCWk, Het, and ZNF/Rpts chromatin states, we filtered for transcripts overlapping at least 50% with the state. For the EnhBiv, TssBiv and BivFlnk regions we required that the 200nt region around the TSS region overlaps at least 50% with the state. We aggregated p-values of up-regulated (sleuth b-value > 0) or down-regulated (sleuth b-value < 0) transcripts separately, according to their chromatin state overlap and additionally filtering for being in a specific (S4 vs S3 or S4 vs S2) enrichment region.

Data availability

The datasets generated during the current study are available in the GEO repository with accession number PRJNA483177.

.

Acknowledgements

We thank Giovanna Lattanzi, Sammy Basso, the Italian network of Laminopathies and members of the laboratory for stimulating discussions and constructive criticisms. We thank Beatrice Bodega, Marina Lusic, Maria Vivo, Daniela Palacios, Chiara Mozzetta, Judith Hariprakash, Koustav Pal, Paolo Maiuri, Mattia Forcato and Fabrizio d'Adda di Fagagna for critical feedback on the manuscript. We are grateful to Mariangela Panetta and Valentina Saccone for providing support with cell cultures. We thank Elisa Salviato for advice on statistical analyses. We gratefully acknowledge the Progeria Research Foundation for providing primary human fibroblasts of HGPS patients. This work was supported by grants from the flagship CNR projects, (Epigen and Interomics) to C.L. and G.O, the Italian Ministry of Health n. GR-2013-02355413 to C.L., My First AIRC Grant (MFAG) n. 18535 to C.L., AFM-Telethon n. 21030 to C.L. and F.F.; by AIRC Start-up grant 2015 n.16841 to F.F.; and Cariplo 2017-0649 to C.L. and F.F. E.S. was supported by the Structured International Post-doc Program of SEMM (SIPOD) and the AFM-TELETHON fellowship n. 21835.

Author Contributions

C. L. designed and F. M., F. L., A. B. and I.O. performed the experiments. E. S., C. P. and F. F. analyzed all sequencing data. S. V. contributed to analyze ChIP-seq data. L.A, F.G and G.O performed image processing and analysis. E. S., F. M., C. L. and F. F. interpreted the results and wrote the manuscript. All authors reviewed and approved the manuscript for submission.

Competing Interests statement

The authors declare SAMMY-seq technique is currently under patenting, thus constituting a potential competing interest.

References

1. Gilbert, N. et al. Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* **118**, 555-66 (2004).
2. Huisinga, K.L., Brower-Toland, B. & Elgin, S.C. The contradictory definitions of heterochromatin: transcription and silencing. *Chromosoma* **115**, 110-22 (2006).
3. Blasco, M.A. The epigenetic regulation of mammalian telomeres. *Nat Rev Genet* **8**, 299-309 (2007).
4. Gilbert, N. & Allan, J. Distinctive higher-order chromatin structure at mammalian centromeres. *Proc Natl Acad Sci U S A* **98**, 11949-54 (2001).
5. Slotkin, R.K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**, 272-85 (2007).
6. Boettiger, A.N. et al. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* (2016).
7. Schuettengruber, B., Bourbon, H.M., Di Croce, L. & Cavalli, G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **171**, 34-57 (2017).
8. Turgay, Y. et al. The molecular architecture of lamins in somatic cells. *Nature* **543**, 261-264 (2017).
9. Shimi, T. et al. Structural organization of nuclear lamins A, C, B1, and B2 revealed by superresolution microscopy. *Mol Biol Cell* **26**, 4075-86 (2015).
10. van Steensel, B. & Belmont, A.S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780-791 (2017).
11. Kind, J. et al. Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153**, 178-92 (2013).
12. Kind, J. & van Steensel, B. Genome-nuclear lamina interactions and gene regulation. *Curr Opin Cell Biol* **22**, 320-5 (2010).
13. Lund, E. et al. Lamin A/C-promoter interactions specify chromatin state-dependent transcription outcomes. *Genome Res* (2013).
14. 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. *Bioessays* **36**, 75-83 (2014).
15. Meuleman, W. et al. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res* **23**, 270-80 (2013).
16. Guelen, L. et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948-51 (2008).
17. Zheng, X. et al. Lamins Organize the Global Three-Dimensional Genome from the Nuclear Periphery. *Mol Cell* **71**, 802-815 e7 (2018).
18. Cesarini, E. et al. Lamin A/C sustains PcG protein architecture, maintaining transcriptional repression at target genes. *J Cell Biol* **211**, 533-51 (2015).
19. Marullo, F. et al. Nucleoplasmic Lamin A/C and Polycomb group of proteins: an evolutionarily conserved interplay. *Nucleus*, 0 (2016).
20. Briand, N. & Collas, P. Laminopathy-causing lamin A mutations reconfigure lamina-associated domains and local spatial chromatin conformation. *Nucleus* **9**, 216-226 (2018).
21. Oldenburg, A. et al. A lipodystrophy-causing lamin A mutant alters conformation and epigenetic regulation of the anti-adipogenic MIR335 locus. *J Cell Biol* **216**, 2731-2743 (2017).
22. 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).
23. Mattout, A. et al. An EDMD mutation in C. elegans lamin blocks muscle-specific gene relocation and compromises muscle integrity. *Curr Biol* **21**, 1603-14 (2011).
24. Peric-Hupkes, D. et al. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* **38**, 603-13 (2010).
25. Perovanovic, J. et al. Laminopathies disrupt epigenomic developmental programs and cell fate. *Sci Transl Med* **8**, 335ra58 (2016).
26. Ullrich, N.J. & Gordon, L.B. Hutchinson-Gilford progeria syndrome. *Handb Clin Neurol* **132**, 249-64 (2015).

27. Vidak, S. & Foisner, R. Molecular insights into the premature aging disease progeria. *Histochem Cell Biol* **145**, 401-17 (2016).
28. Stephens, A.D. et al. Chromatin histone modifications and rigidity affect nuclear morphology independent of lamins. *Mol Biol Cell* **29**, 220-233 (2018).
29. 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**, 260-9 (2013).
30. 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**, 8703-8 (2006).
31. Spielmann, M., Lupianez, D.G. & Mundlos, S. Structural variation in the 3D genome. *Nat Rev Genet* **19**, 453-467 (2018).
32. Dirks, R.A., Stunnenberg, H.G. & Marks, H. Genome-wide epigenomic profiling for biomarker discovery. *Clin Epigenetics* **8**, 122 (2016).
33. Johnson, D.S., Mortazavi, A., Myers, R.M. & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497-502 (2007).
34. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-8 (2013).
35. Boyle, A.P. et al. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311-22 (2008).
36. Schones, D.E. et al. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887-98 (2008).
37. van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol* **18**, 424-8 (2000).
38. He, D.C., Nickerson, J.A. & Penman, S. Core filaments of the nuclear matrix. *J Cell Biol* **110**, 569-80 (1990).
39. Marasca, F., Marullo, F. & Lanzuolo, C. Determination of Polycomb Group of Protein Compartmentalization Through Chromatin Fractionation Procedure. *Methods Mol Biol* **1480**, 167-80 (2016).
40. Lund, E., Oldenburg, A.R. & Collas, P. Enriched domain detector: a program for detection of wide genomic enrichment domains robust against local variations. *Nucleic Acids Res* **42**, e92 (2014).
41. Baudement, M.O. et al. High-salt-recovered sequences are associated with the active chromosomal compartment and with large ribonucleoprotein complexes including nuclear bodies. *Genome Res* **28**, 1733-1746 (2018).
42. Henikoff, S., Henikoff, J.G., Sakai, A., Loeb, G.B. & Ahmad, K. Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res* **19**, 460-9 (2009).
43. Roadmap Epigenomics, C. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
44. Sadaie, M. et al. Redistribution of the Lamin B1 genomic binding profile affects rearrangement of heterochromatic domains and SAHF formation during senescence. *Genes Dev* **27**, 1800-8 (2013).
45. Smith, D.K., Yang, J., Liu, M.L. & Zhang, C.L. Small Molecules Modulate Chromatin Accessibility to Promote NEUROG2-Mediated Fibroblast-to-Neuron Reprogramming. *Stem Cell Reports* **7**, 955-969 (2016).
46. Stavrovskaya, E.D. et al. StereoGene: rapid estimation of genome-wide correlation of continuous or interval feature data. *Bioinformatics* **33**, 3158-3165 (2017).
47. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).
48. Scaffidi, P. & Misteli, T. Lamin A-dependent misregulation of adult stem cells associated with accelerated ageing. *Nat Cell Biol* **10**, 452-9 (2008).
49. Minoux, M. et al. Gene bivalency at Polycomb domains regulates cranial neural crest positional identity. *Science* **355**(2017).
50. Fredriksson, S. et al. Protein detection using proximity-dependent DNA ligation assays. *Nat Biotechnol* **20**, 473-7 (2002).

51. Sewitz, S.A., Fahmi, Z. & Lipkow, K. Higher order assembly: folding the chromosome. *Curr Opin Struct Biol* **42**, 162-168 (2017).
52. Criscione, S.W., Teo, Y.V. & Neretti, N. The Chromatin Landscape of Cellular Senescence. *Trends Genet* **32**, 751-761 (2016).
53. Madakashira, B.P. & Sadler, K.C. DNA Methylation, Nuclear Organization, and Cancer. *Front Genet* **8**, 76 (2017).
54. Teytelman, L., Thurtle, D.M., Rine, J. & van Oudenaarden, A. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proc Natl Acad Sci U S A* **110**, 18602-7 (2013).
55. Waldminghaus, T. & Skarstad, K. ChIP on Chip: surprising results are often artifacts. *BMC Genomics* **11**, 414 (2010).
56. Towbin, B.D. et al. Step-wise methylation of histone H3K9 positions heterochromatin at the nuclear periphery. *Cell* **150**, 934-47 (2012).
57. Wijchers, P.J. et al. Cause and Consequence of Tethering a SubTAD to Different Nuclear Compartments. *Mol Cell* **61**, 461-473 (2016).
58. Ghavi-Helm, Y. et al. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat Genet* **51**, 1272-1282 (2019).
59. Hug, C.B., Grimaldi, A.G., Kruse, K. & Vaquerizas, J.M. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell* **169**, 216-228 e19 (2017).
60. Falk, M. et al. Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395-399 (2019).
61. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60 (2009).
62. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
63. Tischler, G. & Leonard, S. biobambam: tools for read pair collation based algorithms on BAM files. *Source Code for Biology and Medicine* **9**, 13 (2014).
64. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).
65. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20 (2014).
66. Li, H. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics* **30**, 2843-51 (2014).
67. Kharchenko, P.V., Tolstorukov, M.Y. & Park, P.J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* **26**, 1351-9 (2008).
68. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40 (2010).
69. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-5 (2016).
70. Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S. & Karolchik, D. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* **26**, 2204-7 (2010).
71. Hahne, F. & Ivanek, R. Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol* **1418**, 335-51 (2016).
72. Gel, B. & Serra, E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**, 3088-3090 (2017).
73. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis.*, (2016).
74. Nakamura, A., Osonoi, T. & Terauchi, Y. Relationship between urinary sodium excretion and pioglitazone-induced edema. *J Diabetes Investig* **1**, 208-11 (2010).
75. Bray, N.L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-7 (2016).
76. Pimentel, H., Bray, N.L., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat Methods* **14**, 687-690 (2017).
77. Yi, L., Pimentel, H., Bray, N.L. & Pachter, L. Gene-level differential analysis at transcript-level resolution. *Genome Biol* **19**, 53 (2018).

Figure Legends

Figure 1 - **SAMMY-seq isolates specific DNA regions in fibroblast samples. a**, Schematic representation of SAMMY-seq. Chromatin fractions are sequentially isolated after DNase treatment, high salt and urea buffers extractions. The associated genomic DNA is extracted, sonicated and processed for high-throughput sequencing. **b**, Distribution of SAMMY-seq reads along a representative region (85Mb region in chr4:55480346-141005766). Library size normalized read counts over 10Kb genomic bins are shown for each sequenced fraction of a control fibroblast sample (CTRL002). **c**, Differential reads distribution across pairwise comparisons of SAMMY-seq fractions in a control fibroblast sample (CTRL002) along the genomic region in (a). Less accessible fractions are compared to more accessible ones used as reference (S3 vs S2; S4 vs S3; S4 vs S2). Regions of signal enrichment or depletion over the reference samples are marked in yellow or purple, respectively. The smoothed differential signal is calculated with SPP, and significantly enriched regions (SAMMY-seq domains) are called with EDD (see Online methods for details) and reported as grey boxes under the enrichment signal track.

Figure 2 - SAMMY-seq preferentially isolates heterochromatic domains. a, Visualization of SAMMY-seq enrichment signal along with multiple chromatin marks on a representative region (8Mb region in chr5:158000000-170000000). From top to bottom: tracks for open chromatin (ATAC-seq and DNase-seq – dark green); association to lamina (Lamin A/C and Lamin B ChIP-seq – dark blue); ChIP-seq for histone marks associated to PcG regulation (H3K27me3 – light green), active chromatin (H3K4me1 – light green) or heterochromatin (H3K9me3 – light blue); SAMMY-seq enrichment signal (S4 vs S2) in 3 control samples (yellow or purple colored track for enrichment or depletion, respectively, over the S2 reference baseline). Grey boxes under each SAMMY-seq track show the SAMMY-seq domains. See Online methods for details of each type of signal processing. **b**, Genome-wide kernel correlation (StereoGene) for SAMMY-seq of

all control samples (S4 vs S2 enrichment) compared to ATAC-seq, DNase-seq, and ChIP-seq (Lamin A/C, Lamin B, H3K27me3, H3K4me1 and H3K9me3). **c**, Overlap (Jaccard Index - JI) of control samples SAMMY-seq domains (S4 vs S2) with LADs (Lamin A/C or Lamin B ChIP-seq enrichment domains). The observed JI (blue bar) is compared to mean JI across 10,000 randomizations of SAMMY-seq domains (pink bar) or LADs (yellow bar) positions along the genome. All Jaccard Indexes are significantly higher than the randomized values (all empirical p-values < 0.0001). The black rectangle over the randomized values shows the +/- 2 standard error range.

Figure 3 - HGPS fibroblasts show early changes in chromatin accessibility. a, Genomic tracks for SAMMY-seq in 3 control and 3 HGPS samples along with chromatin marks in a representative region (48 Mb region in chr10:46000000-94000000). From top to bottom: tracks for association to lamina (Lamin A/C and Lamin B ChIP-seq – dark blue); heterochromatin (H3K9me3 ChIP-seq – light blue); SAMMY-seq enrichment signal (S4 vs S2) in 3 control and 3 HGPS samples (yellow or purple colored track for enrichment or depletion, respectively, over the S2 reference baseline). Grey boxes under each SAMMY-seq track show the SAMMY-seq domains. **b**, Pairwise overlap of SAMMY-seq domains (S3 vs S2, S4 vs S3, S4 vs S2) between control (black dots) or HGPS (red dots) sample pairs (JI on y-axis, number of overlapping regions on x-axis). Upper-tail Fisher-test p-value < 0.01 for all overlaps, except for HGPS S4 vs S2 pairwise overlaps. **c**, Smoothed average ChIP-seq enrichment signal for chromatin marks around SAMMY-seq domain borders. Average ChIP-seq signal for Lamin A/C (upper row), Lamin B (middle row) and H3K9me3 (bottom row) is reported around the SAMMY-seq domain borders start (left side plots) or end (right side plots) for domains detected in each control (black lines) and progeria (red lines) sample. Results for each set of SAMMY-seq enrichment domains are reported (S3 vs S2 on the left, S4 vs S3 center, S4 vs S2 on the right) using a +/-50 bins window (10Kb bin size) centered on the start or end domain border positions (vertical dashed grey line).

Figure 4 – Early chromatin structure disruption in progeria is not accompanied by alterations in H3K9me3. a, Genomic tracks for paired H3K9me3 ChIP-seq and SAMMY-seq in 3 control and 3 HGPS samples for a representative region (40Mb region in chr4:109000000-149000000). Enrichment signals (computed using SPP) for each sample are shown (H3K9me3 ChIP-seq – light blue; yellow or purple colored track for SAMMY-seq S4 vs S2 enrichment or depletion, respectively). Grey boxes under each SAMMY-seq track show the SAMMY-seq domains. **b**, Percentage of H3K9me3 domains conserved across all 3 control or all 3 progeria samples (computed as percent over per sample total H3K9me3 domains size). **c**, Overlap of H3K9me3 enriched domains and SAMMY-seq domains (S4 vs S3, S4 vs S2) for control (black dots) or HGPS (red dots) samples (JI on y-axis, number of overlapping regions on x-axis). All overlaps are significant in controls (for both S4 vs S3 and S4 vs S2 SAMMY-seq domains) (upper-tail Fisher-test p-value < 0.01 for all overlaps) and in HGPS169 (only for S4 vs S2 SAMMY-seq domains). **d**, Representative images of H3K9me3/Lamin A immunofluorescence analysis on control or HGPS fibroblasts. **e**, Percentage of H3K9me3 foci localized within 1 um from nuclear periphery in distinct cell populations. Comparisons were done using a two-tailed t-test. Statistically significant differences are marked * $p < 0.05$

Figure 5 – HGPS fibroblasts show an alteration of PRC2 distribution. a, MSigDB pathways consistently deregulated in all three progeria samples (FDR corrected aggregated p-value < 0.05), based on comparison of individual HGPS samples to the group of control samples (transcript level analysis - see Online methods). Pathways are ranked based on the sum of their $-\log_{10}$ transformed p-values across all 3 progeria samples. Dot colors highlight different categories of pathways as per graphical legend: related to stem cells or cancer, H3K27me3/PRC2 regulation, extracellular matrix, aging or other categories. Dot size is proportional to the $-\log_{10}$ transformed p-value. **b**, Genomic track for H3K27me3 ChIP-seq signal

in 2 control and 3 HGPS samples for a representative region (2.5Mb region in chr7:88000000-90500000). Enrichment signals (computed using SPP) for each sample are shown. **c**, Multi-dimensional scaling of H3K27me3 ChIP-seq IP experiments using read counts in 10k genomic regions across all autosomes. Black dots show control, while red dots show progeria samples. The 2D distance of dots is representative of the pairwise distance of samples using the 10k genomic region read counts.

Figure 6 - Chromatin structural changes in HGPS affect PRC2 regulated bivalent gens.

a, H3K27me3 signal distribution around the TSS region of genes upregulated in HGPS samples compared to controls calculated by deepTools using the genome-wide signal from SPP (see Online methods for details). The x-axis represents relative genomic position around the TSS (+/-5Kb), and the y-axis represents average signal intensity. **b**, Overlap of SAMMY-seq domains from each sample to Roadmap Epigenomics chromatin states for normal fibroblasts (E055). The overlaps with each chromatin state are reported as percentage over the total of SAMMY-seq domains (S4 vs S3 or S4 vs S2) for each sample. **c**, Transcripts differential expression (up-regulation) was assessed by comparing individual HGPS samples against the group of controls, then p-values were aggregated based on their chromatin state, considering only genes in SAMMY-seq domains (S4 vs S2) (see Online methods for details). The plot reports the number of transcripts (x-axis) and significance (Lancaster method aggregated p-values – y-axis) for each chromatin state (color legend) as defined by[43]. **d**, Representative track of H3K27me3 ChIP-seq signal in 2 control and 3 HGPS samples around the bivalent EN1 gene (chr2: 118810000-118880000), that was upregulated in HGPS samples. **e**, **f**, Representative images of PLA experiments in CTRL004 and HGPS167 at late passages (p18). Each fluorescent dot represents the co-localization of Lamin A/C or Progerin and Ezh2 (panel e) or Bmi1 (panel f). Nuclei were stained with dapi. All data were generated from an average of three independent experiments, whiskers represent SEM. **g**, Western blot on chromatin fractionation experiments of CTRL004 and HGPS167 at early (left) and late (right) passages. Equal amounts of each

fraction were hybridized with indicated antibodies. Alpha-tubulin, histone H3 and Lamin A/C were used as loading controls respectively for S1, S2 and S3, S4. **h**, The graph shows quantifications of Ezh2 and Bmi1 in S4 fraction normalized on S2. Data points were generated from an average of at least two biological replicates, whiskers represent SEM.

Comparisons were done using a two-tailed t-test in (e, f, h). Statistically significant differences are marked * $p < 0.05$; * * $p < 0.01$.

Figure 1

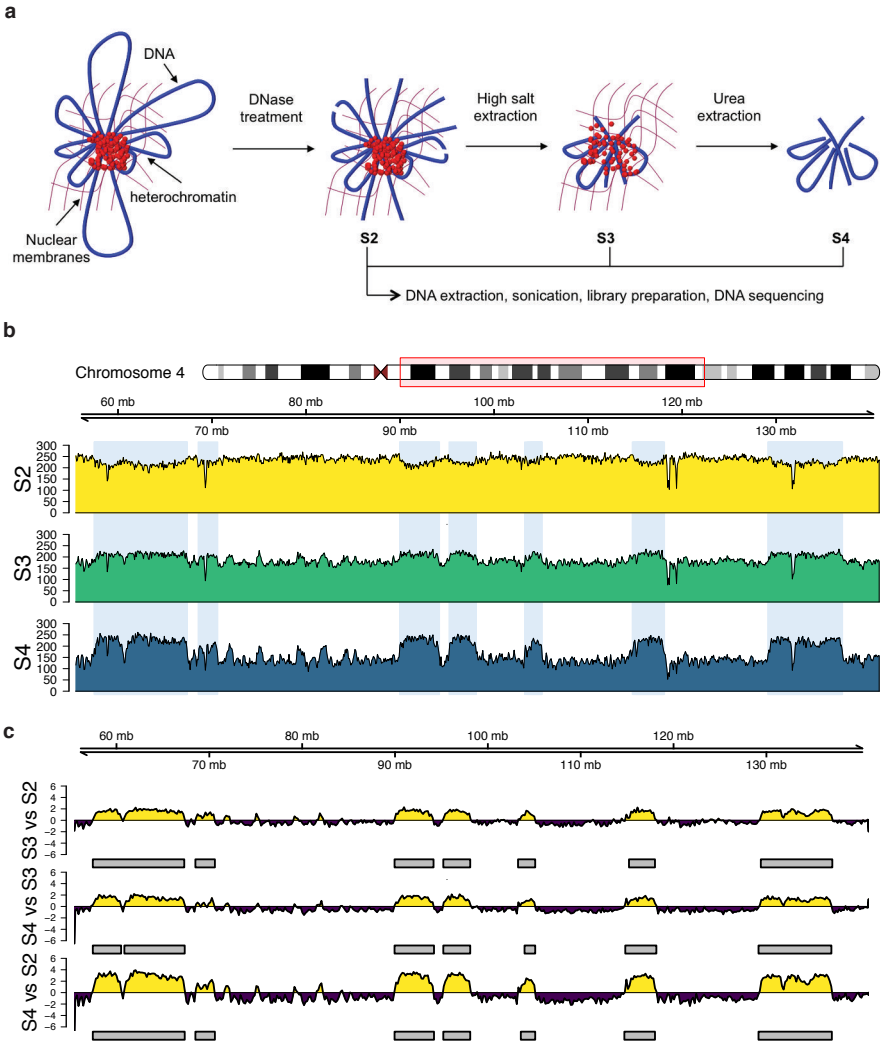


Figure 2

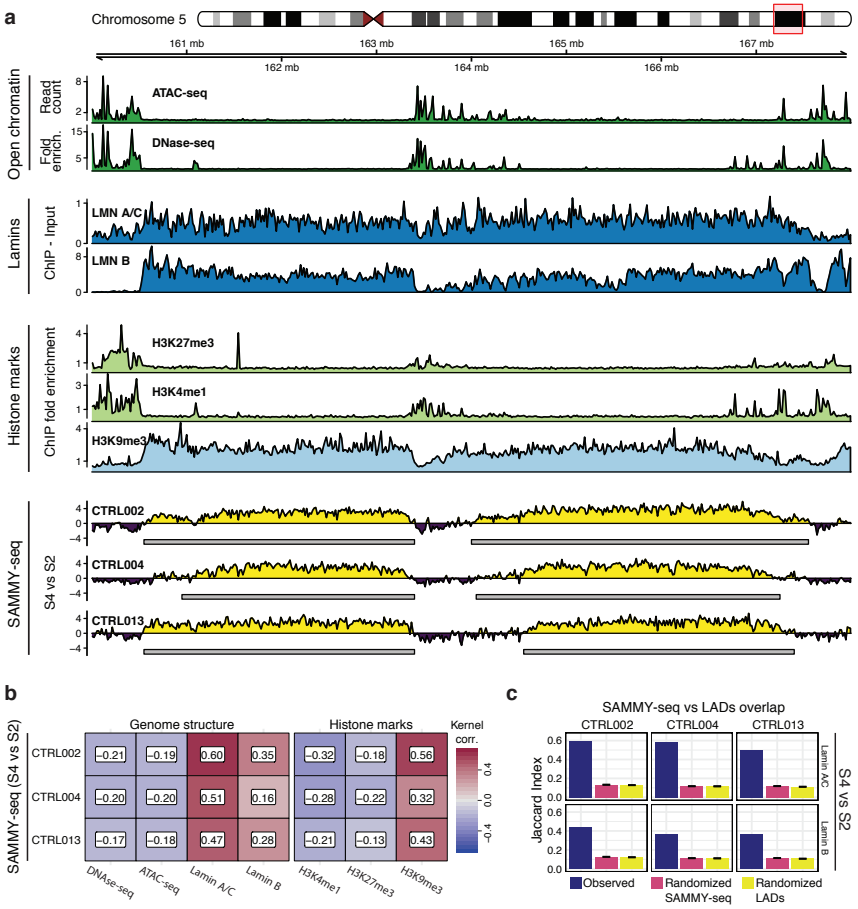


Figure 3

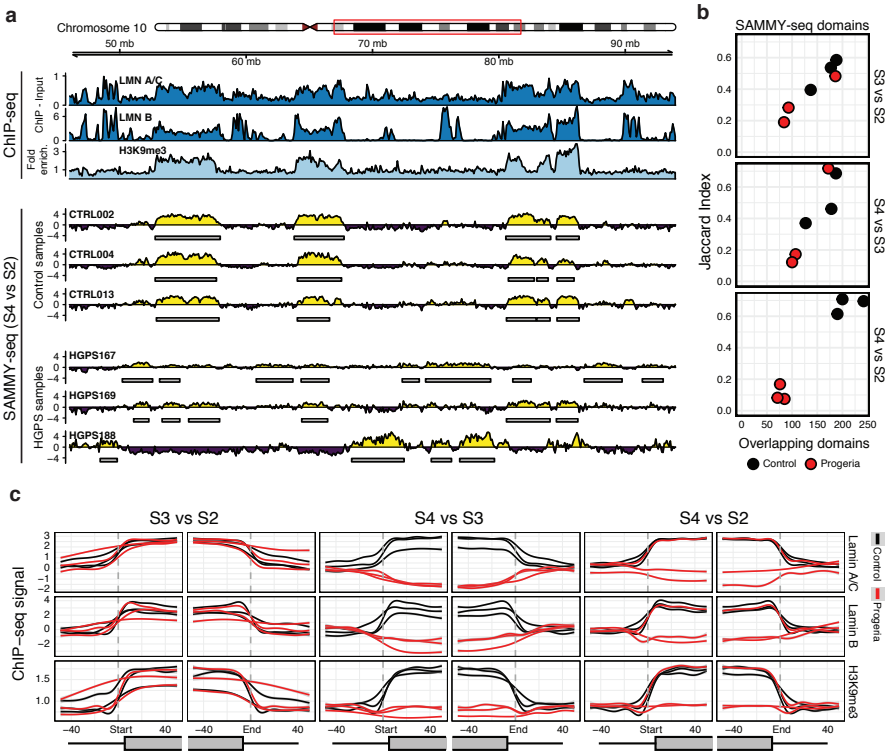


Figure 4

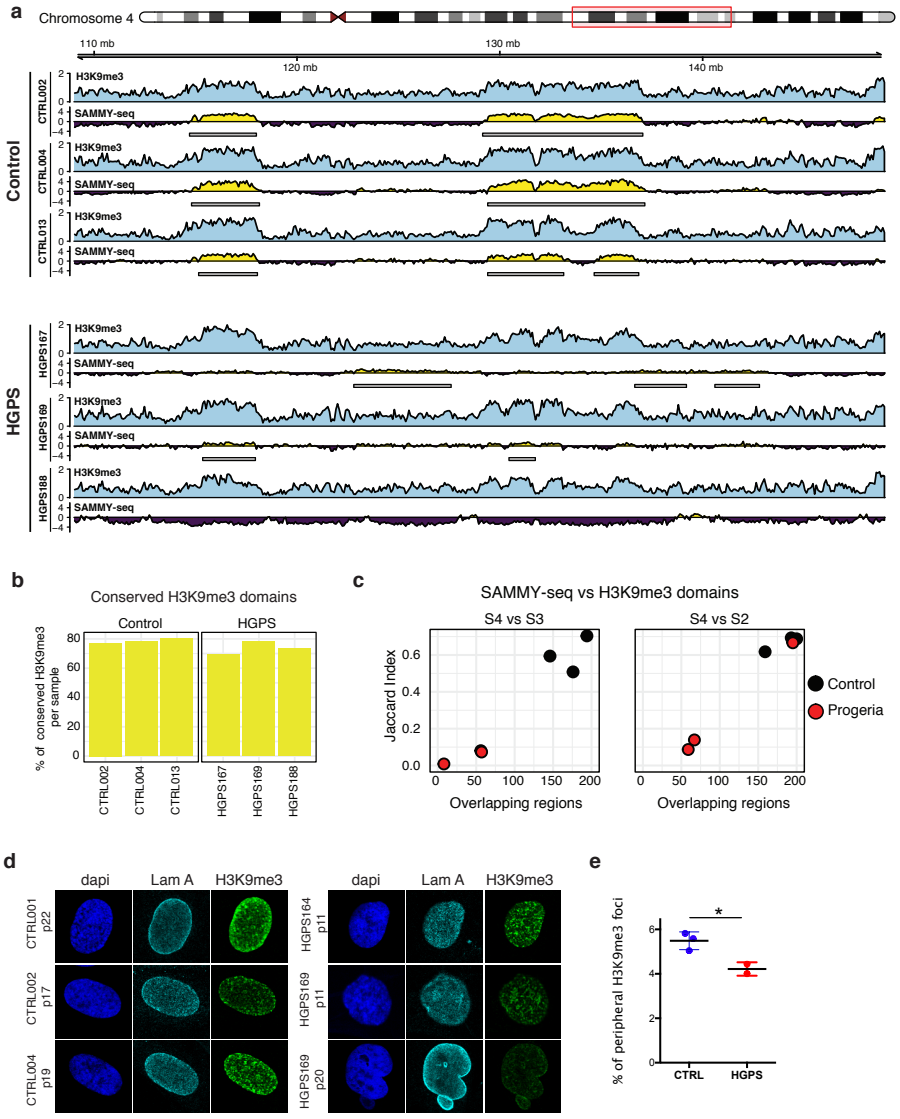


Figure 5

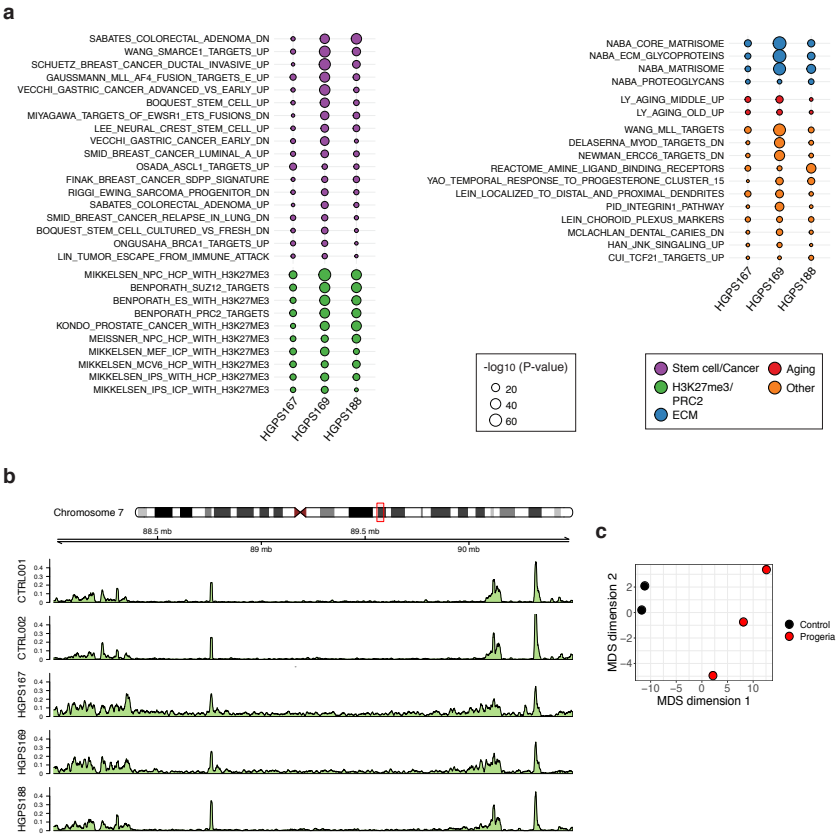
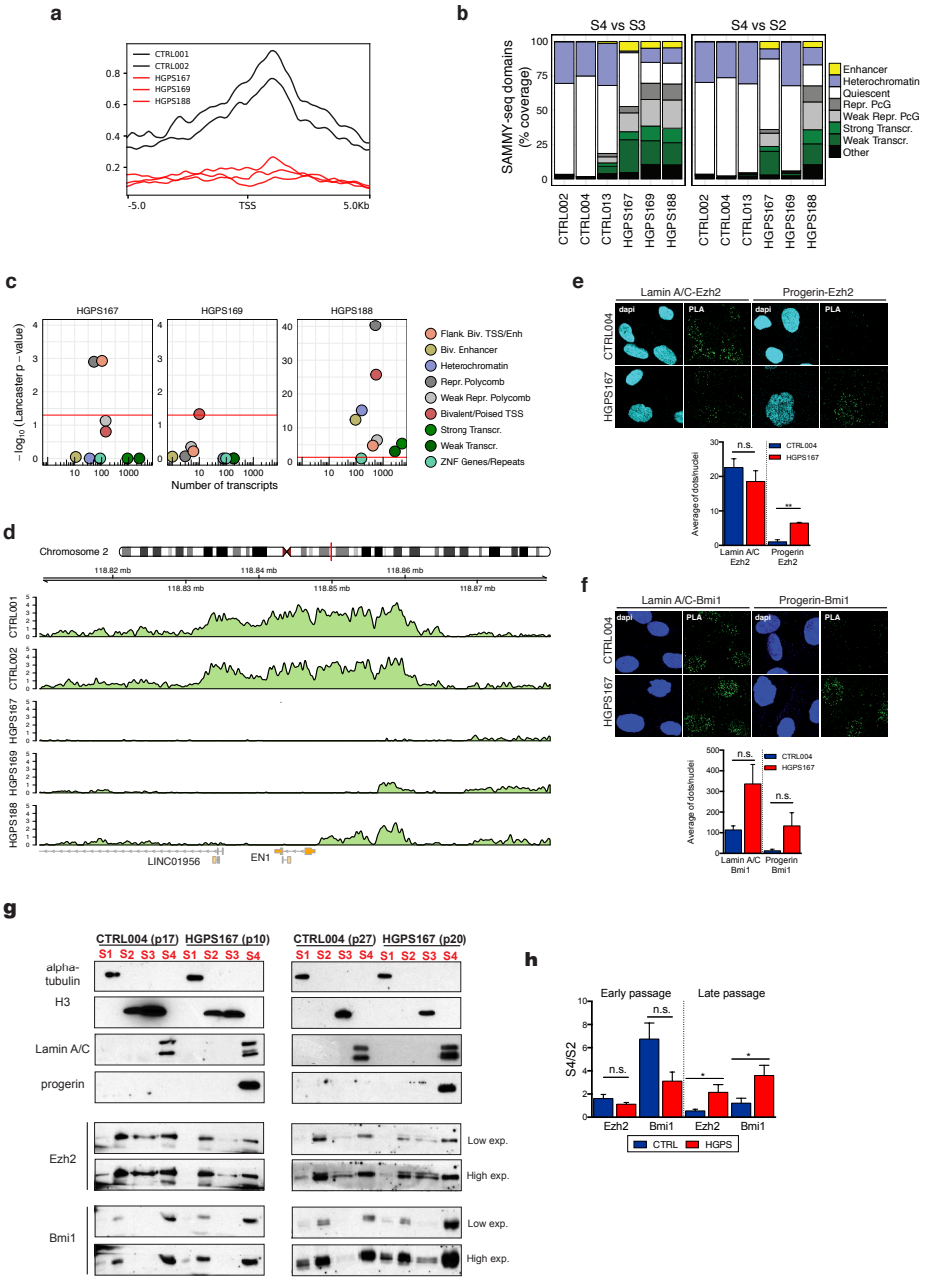


Figure 6



Supplementary material for:

Early Polycomb-target deregulations in Hutchinson-Gilford Progeria Syndrome revealed by heterochromatin analysis

Endre Sebestyén[1§#], Fabrizia Marullo[2§], Federica Lucini[3], Andrea Bianchi[2,4], Cristiano Petrini[1], Sara Valsoni[4,5], Ilaria Olivieri[2], Laura Antonelli[5], Francesco Gregoretti[5], Gennaro Oliva[5], Francesco Ferrari[1,6*] and Chiara Lanzuolo[3,7*]

[1] IFOM, the FIRC Institute of Molecular Oncology, Milan, Italy
[2] Institute of Cell Biology and Neurobiology, National Research Council, Rome, Italy
[3] Istituto Nazionale Genetica Molecolare "Romeo ed Enrica Invernizzi", Milan, Italy
[4] IRCCS Santa Lucia Foundation, Rome, Italy
[5] Institute for High Performance Computing and Networking, Naples, Italy
[6] Institute of Molecular Genetics, National Research Council, Pavia, Italy
[7] Institute of Biomedical Technologies, National Research Council, Milan, Italy
[#] Current address: 1st Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary

§ Equal contribution
* Equal contribution and correspondence: chiara.lanzuolo@cnr.it; francesco.ferrari@ifom.eu

Table of Contents

Sample ID	Fraction	Read number
CTRL002	S2	78102212
CTRL002	S3	61341611
CTRL002	S4	55731571
CTRL004	S2	78683296
CTRL004	S3	60438514
CTRL004	S4	54864540
CTRL013	S2	59635985
CTRL013	S3	65360683
CTRL013	S4	85335120
HGPS167	S2	58983114
HGPS167	S3	73501327
HGPS167	S4	61985833
HGPS169	S2	79209359
HGPS169	S3	77731616
HGPS169	S4	92435639
HGPS188	S2	68067662
HGPS188	S3	104711260
HGPS188	S4	54037646

Supplementary Table 1. Total number of sequencing reads in all samples and fractions for the SAMMY-seq sequencing runs.

Sample ID	Comparison	Total length	Domain count	Domain average length	Autosome %
CTRL002	S3 vs S2	605,736895	273	2,218816465	21,06909824
CTRL002	S4 vs S2	616,186907	285	2,162059323	21,43257672
CTRL002	S4 vs S3	582,086898	276	2,1090105	20,24649008
CTRL004	S3 vs S2	374,950204	204	1,837991196	13,04173932
CTRL004	S4 vs S2	499,250205	205	2,435366854	17,36521533
CTRL004	S4 vs S3	477,000188	188	2,537235043	16,5913021
CTRL013	S3 vs S2	644,500193	193	3,339379238	22,41738615
CTRL013	S4 vs S2	510,136889	267	1,910625052	17,7438824
CTRL013	S4 vs S3	435,23686	238	1,828726303	15,13866538
HGPS167	S3 vs S2	178,500104	104	1,716347154	6,208695983
HGPS167	S4 vs S2	549,050231	231	2,376840827	19,09738923
HGPS167	S4 vs S3	487,250195	195	2,498718949	16,9478239
HGPS169	S3 vs S2	782,150186	186	4,205108527	27,20520946
HGPS169	S4 vs S2	465,712191	260	1,791200735	16,19867633
HGPS169	S4 vs S3	629,494655	176	3,576674176	21,89545467
HGPS188	S3 vs S2	540,336884	262	2,062354519	18,79431645
HGPS188	S4 vs S2	650,582792	197	3,302450721	22,6289547
HGPS188	S4 vs S3	693,286821	199	3,483853372	24,11431144

Supplementary Table 2. Characteristics of SAMMY-seq domains, detected by EDD. The table contains the following columns. *Sample ID*: patient ID, *Comparison*: fractions used in the EDD run as the "IP" and "input" sample, *Total length*: total length of SAMMY-seq domains in megabase, *Domain count*: number of SAMMY-seq domains, *Domain average length*: average length of SAMMY-seq domains in megabase, *Autosome %*: % of autosomes covered by SAMMY-seq domains.

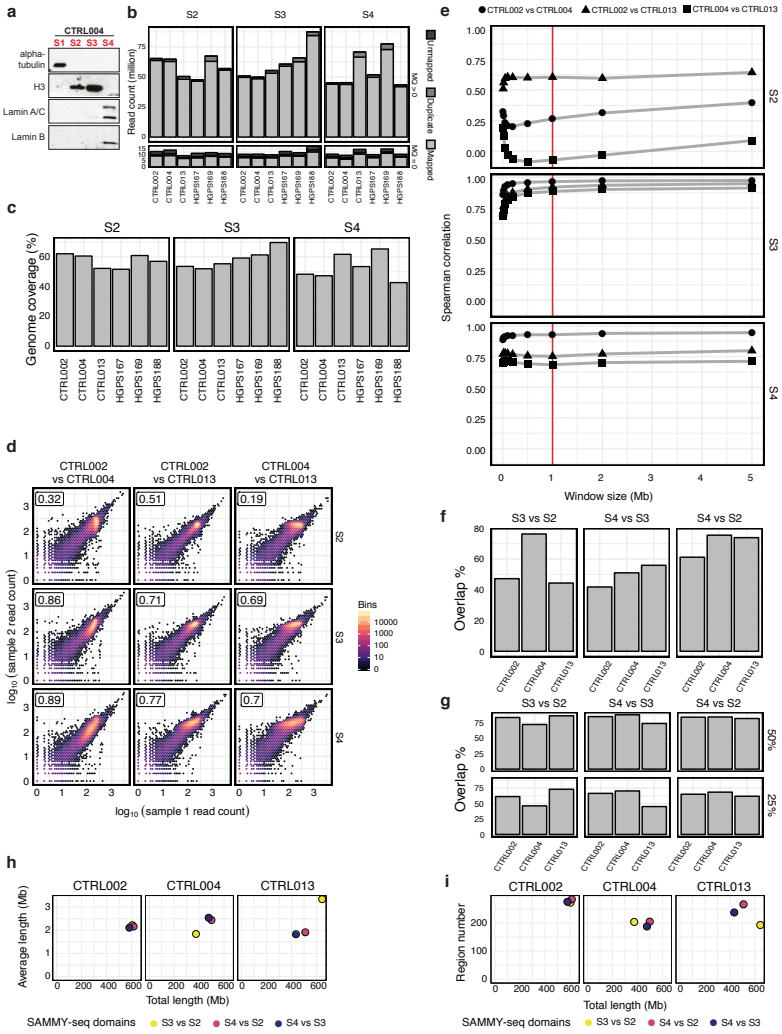
Sample ID	Comparison	Lamin A overlap JI	Lamin A overlap %	Lamin B overlap JI	Lamin B overlap %
CTRL002	S3 vs S2	0,609681	79,95224036	0,437656	60,95883699
CTRL002	S4 vs S2	0,598586	78,33665691	0,439887	60,6553253
CTRL002	S4 vs S3	0,584696	79,55001584	0,43488	61,92377326
CTRL004	S3 vs S2	0,406604	80,78410455	0,292558	59,28790505
CTRL004	S4 vs S2	0,58281	86,44967451	0,363631	59,09864797
CTRL004	S4 vs S3	0,603878	90,76519916	0,356267	59,70649895
CTRL013	S3 vs S2	0,630849	79,06904577	0,38012	53,49107836
CTRL013	S4 vs S2	0,502858	77,59685993	0,367625	58,87638469
CTRL013	S4 vs S3	0,31422	60,87493253	0,277669	52,05214556

Supplementary Table 3. Statistics on on control sample SAMMY-seq domains overlapping Lamina Associated Domains (LADs) based on literature data. The table contains the following columns. *Sample ID*: patient ID, *Comparison*: fractions used in the EDD run as the “IP” and “input” sample, *Lamin A overlap JI*: Jaccard index of the SAMMY-seq domain – Lamin A/C LAD comparison as calculated by bedtools, *Lamin A overlap %*: % of SAMMY-seq domains overlapping with Lamin A/C LADs, *Lamin B overlap JI*: Jaccard index of the SAMMY-seq domain – Lamin B LAD comparison as calculated by bedtools, *Lamin B overlap %*: % of SAMMY-seq domains overlapping with Lamin B LADs.

Sample ID 1	Sample ID 2	Comparison	Overlap JI	Overlap number
CTRL002	CTRL004	S3 vs S2	0,536316	177
CTRL002	CTRL013	S3 vs S2	0,585113	188
CTRL004	CTRL013	S3 vs S2	0,39536	137
HGPS167	HGPS169	S3 vs S2	0,190102	84
HGPS167	HGPS188	S3 vs S2	0,282751	93
HGPS169	HGPS188	S3 vs S2	0,482547	186
CTRL002	CTRL004	S4 vs S3	0,686345	187
CTRL002	CTRL013	S4 vs S3	0,46086	178
CTRL004	CTRL013	S4 vs S3	0,370164	127
HGPS167	HGPS169	S4 vs S3	0,172627	107
HGPS167	HGPS188	S4 vs S3	0,121931	100
HGPS169	HGPS188	S4 vs S3	0,717687	172
CTRL002	CTRL004	S4 vs S2	0,707946	200
CTRL002	CTRL013	S4 vs S2	0,695666	242
CTRL004	CTRL013	S4 vs S2	0,614278	190
HGPS167	HGPS169	S4 vs S2	0,168257	76
HGPS167	HGPS188	S4 vs S2	0,0748601	85
HGPS169	HGPS188	S4 vs S2	0,0823542	71

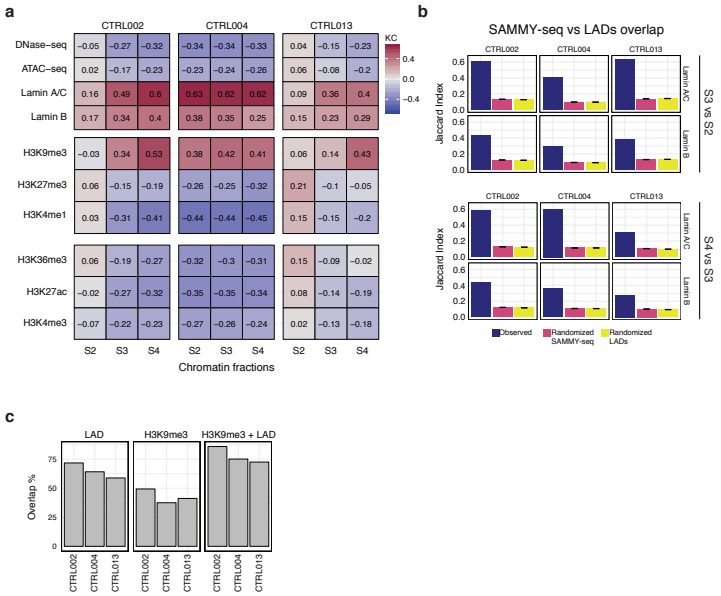
Supplementary Table 4. Pairwise comparison of SAMMY-seq domains across control or progeria samples. The table contains the following columns. *Sample ID 1*: patient ID for comparison, *Sample ID 2*: patient ID for comparison, *Comparison*: fractions used in the EDD run as the “IP” and “input” sample, *Overlap JI*: Jaccard index of the SAMMY-seq domain overlaps as calculated by bedtools, *Overlap number*: number of overlapping SAMMY-seq domains.

Supplementary Figure 1



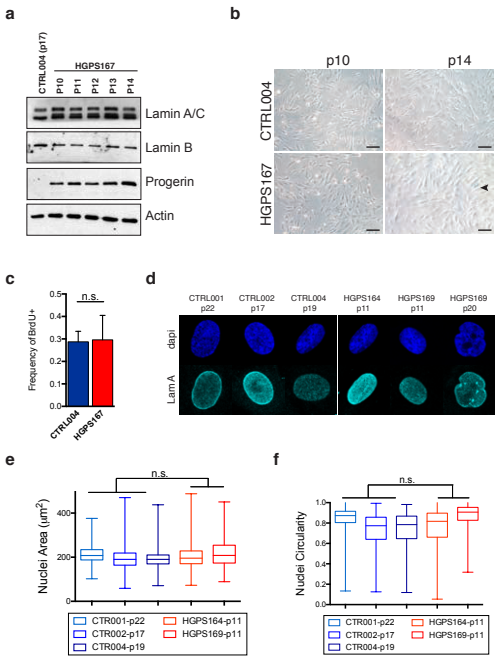
Supplementary Figure 1 - Characteristics and reliability of SAMMY-seq libraries. a, Western blots for chromatin fractionation experiments of CTRL004. Sequential extractions were performed to isolate soluble proteins (S1-fraction), DNase-sensitive chromatin (S2-fraction), DNase-resistant chromatin (S3-fraction), and the most compact and inaccessible chromatin (S4-fraction). Equal amounts of each fraction were hybridized with indicated antibodies. **b**, Total read counts for each SAMMY-seq sample. The stacked bar plot shows mapped, unmapped, and PCR duplicate reads with bwa mapping quality equal to zero (MQ = 0, non-unique mapping) or larger than zero (MQ > 0, uniquely mapped reads). The average number of sequencing reads in the different fractions were 72 million (S2), 62 million (S3) and 65 (S4) million. **c**, Percent of the human genome (hg38-noalt build) covered by at least one read in each SAMMY-seq sample. On average 58% (S2), 54% (S3) and 52% (S4) of the genome was covered by at least one read. **d**, Scatter plots comparing read counts over 10Kb bins in the S2, S3 and S4 fractions (rows) across three pairwise comparisons of the control samples (columns). The color gradient represents the number of genomic bins with the same (x,y) values. Spearman correlation values are reported in the top left corner of each subpanel. **e**, Spearman correlation (y-axis) of read counts between control samples pairs (shape of dots) over different bin sizes (x-axis) in the S2, S3 and S4 fractions. The red line shows the 1Mb bin size. **f**, Percentage of SAMMY-seq domains (S3 vs S2, S4 vs S3 and S4 vs S2) conserved across all 3 control samples (computed as percent over per sample total SAMMY-seq domains size). **g**, Percentage of SAMMY-seq domains (S3 vs S2, S4 vs S3 and S4 vs S2) detected after 50% or 25% down sampling of sequencing reads (computed as percent over per sample total SAMMY-seq domains size detected with 100% of sequencing reads). **h**, Average (y-axis) and total (x-axis) size of SAMMY-seq domains (S3 vs S2, S4 vs S3 and S4 vs S2) for each control sample. Average size across samples is 2.47Mb for S3 vs S2, 2.16Mb for S4 vs S3 and 2.17Mb for S4 vs S2. Average genome coverage across samples is 18.84% S3 vs S2, 17.33% for S4 vs S3 and 18.85% for S4 vs S2. **i**, Number (y-axis) and total size (x-axis) of SAMMY-seq domains (S3 vs S2, S4 vs S3 and S4 vs S2) for each control sample.

Supplementary Figure 2



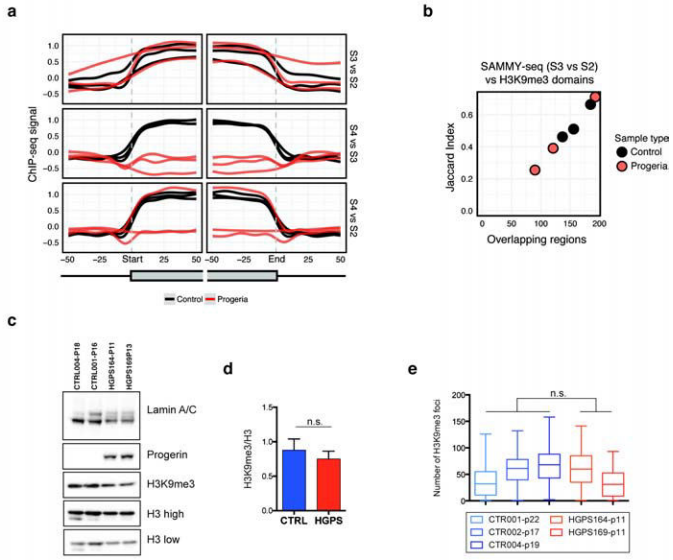
Supplementary Figure 2 - Detailed characteristics and genome-wide association of SAMMY-seq fractions. **a**, Genome-wide kernel correlation calculated by StereoGene, between read coverage in individual control samples chromatin fractions against ChIP-seq and other enrichment signals in ATAC-seq, DNase-seq, Lamin A/C and Lamin B; H3K27me3, H3K4me1 and H3K9me3; H3K36me3, H3K27ac and H3K4me3. In most samples the correlation is progressively increasing for closed chromatin marks or decreasing for open chromatin marks when considering fractions from S2 to S4. **b**, Overlap (Jaccard Index - JI) of control samples SAMMY-seq domains (S3 vs S2 or S4 vs S3) with LADs (Lamin A/C or Lamin B ChIP-seq enrichment domains). The observed JI (blue bar) is compared to median JI across 10,000 randomizations of SAMMY-seq domains (pink bar) or LADs (yellow bar) positions along the genome. All Jaccard Indexes are significantly higher than the randomized values (all empirical p-values < 0.0001). The black rectangle over the randomized values shows the +/- 2 standard error range. **c**, Percentage of LAD, H3K9me3 or LAD+H3K9me3 regions detected by the S4 vs S2 SAMMY-seq domains. LADs were defined based on[27], while H3K9me3 regions were originating from the Roadmap Epigenomics E055 sample[32].

Supplementary Figure 3



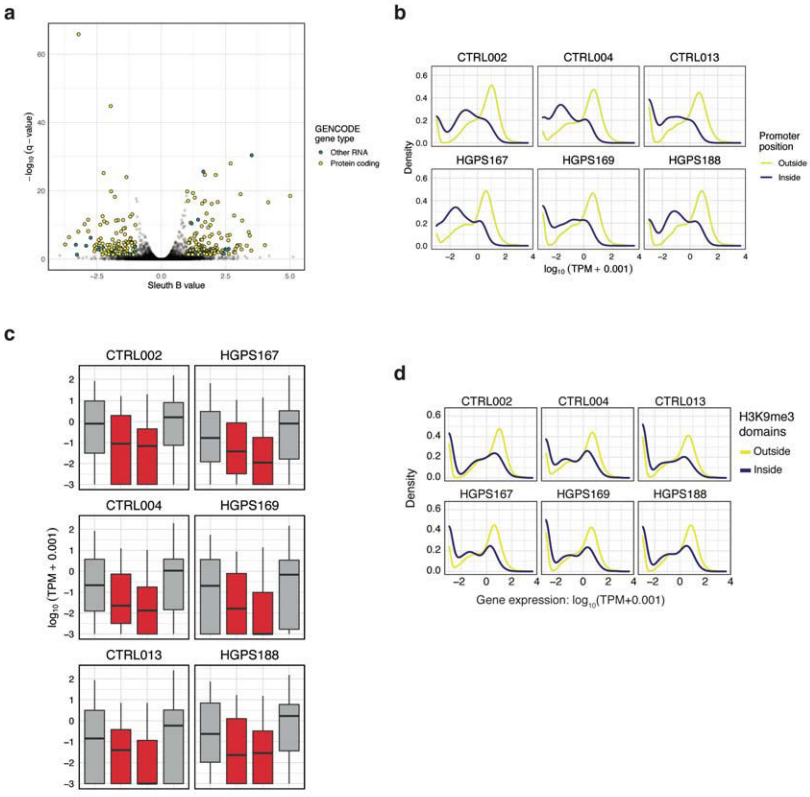
Supplementary Figure 3 - Characteristics of control and progeria fibroblast cells. a, Western blots for total extract from CTRL and HGPS fibroblasts at different cell culture passages (P10-14) hybridized with indicated antibodies. **b**, Representative fields of SA-β-gal activity in CTRL and HGPS fibroblasts at p10 and p14. Scale bar, 50 μm. **c**, Bar plot reporting the proliferation rate in control vs HGPS fibroblasts estimated as percentage of BrdU positive cells with respect to the total number of nuclei (n > 550). The bars report averages of two independent experiments, whiskers represent standard deviation. **d**, Representative images of Lamin A/DAPI immunofluorescence analysis on control or HGPS fibroblasts at various passages (indicated with p). **e**, **f**, Quantification of nuclei area (panel e) and circularity (panel f) of CTRL and HGPS fibroblasts. Number of analyzed nuclei: CTRL001: 660; CTRL002: 1253; CTRL004: 885; HGPS164: 990; HGPS169-p11: 312; HGPS169-p20: 150.
Comparisons were done using a two-tailed t-test in (c) and nested anova in (e, f). n.s.: not significant.

Supplementary Figure 4



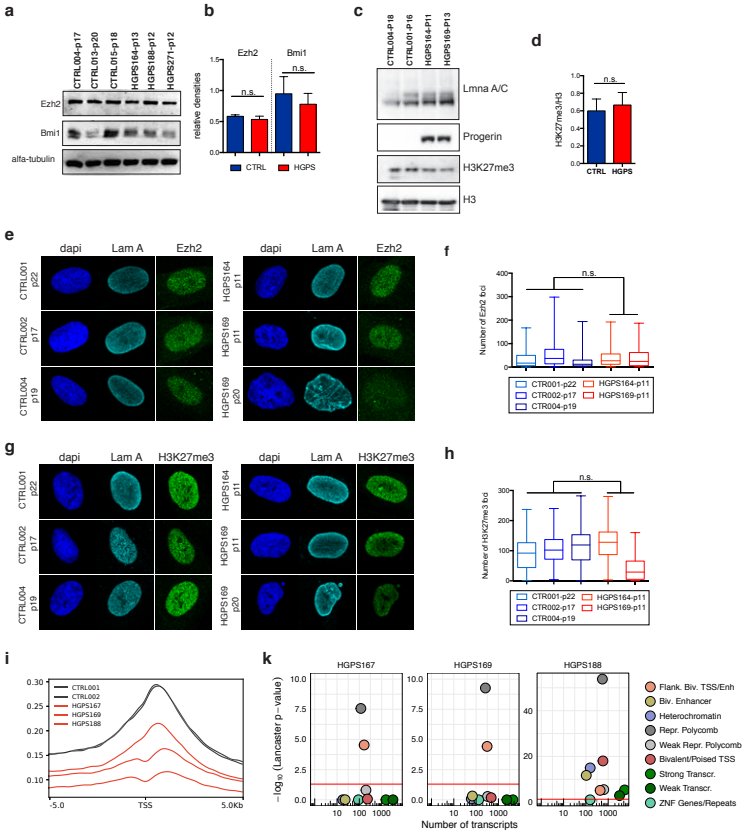
Supplementary Figure 4 - Additional analysis on H3K9me3 patterns. a, Average ChIP-seq enrichment signal for H3K9me3 is reported around the SAMMY-seq domain borders start (left side plots) or end (right side plots) for domains detected in each control (black lines) and progeria (red lines) sample. H3K9me3 ChIP-seq was obtained for each sample individually. Results for each set of SAMMY-seq enrichment domains are reported (S3 vs S2 top, S4 vs S3 middle, S4 vs S2 bottom) using a +/-50 bins window (10Kb bin size) centered on the start or end domain border positions (vertical dashed grey line). **b**, Overlap of H3K9me3 enriched domains and SAMMY-seq domains (S3 vs S2) for control (black dots) or HGPS (red dots) samples (JI on y-axis, number of overlapping regions on x-axis). **c**, Western blots for total extract from control or HGPS fibroblasts at early passages hybridized with indicated antibodies. Histone H3 was used as loading controls for H3K9me3 quantification. **d**, The graph shows quantifications of H3K9me3 levels normalized on H3. Data points were generated from an average of at least 10 biological replicates. **e**, Quantification of number per nucleus of H3K9me3 foci of CTRL and HGPS fibroblasts. Number of analyzed nuclei: CTRL001: 242; CTRL002: 649; CTRL004: 518; HGPS164: 550; HGPS169: 65.
Comparisons were done using a two-tailed t-test in (d) and nested anova in (e). n.s.: not significant.

Supplementary Figure 5



Supplementary Figure 5 – Expression analysis in control and progeria samples. a, Volcano plot of gene level differential expression. The x axis shows the sleuth b value, while the y axis shows $-\log_{10}$(q-value) of genes. Genes highlighted in green or yellow show significant changes ($|b| > 1$ and q-value < 0.05). The yellow color refers to protein coding genes based on GENCODE v27 annotation, green refers to all other types of non protein coding genes. **b**, Normalized gene expression distribution for protein coding genes separated based on their promoter being inside or outside of the consensus SAMMY-seq domains (S4 vs S2) of all controls. Genes with promoters outside the domains have higher expression (Wilcoxon rank sum test p-values < 0.01). **c**, Boxplot of normalized gene expression distribution for protein coding genes flanking consensus SAMMY-seq domain borders (S4 vs S2) of all controls. The gray boxplots show genes located in the outer 500Kb flanking regions (upstream of start or downstream of end), the red boxplots show genes in the inner 500Kb flanking regions (all Wilcoxon rank sum test p-values < 0.01). **d**, Gene expression distribution for protein coding genes separated based on their position being inside or outside of the H3K9me3 peaks in each sample. All samples show a significant difference in expression (Wilcoxon rank sum test p-values < 0.01).

Supplementary Figure 6



Supplementary Figure 6 - Additional analysis on H3K27me3 patterns. a, Western blots for total extract from control or HGPS fibroblasts at early passages hybridized with indicated antibodies. Alpha-tubulin was used as loading control. **b**, The graph shows the average of Ezh2 and Bmi1 levels in (a) normalized on Alpha-tubulin. **c**, Western blots for total extract from control or HGPS fibroblasts at early passages hybridized with indicated antibodies. Histone H3 was used as loading controls for H3K27me3 quantification. **d**, The graph shows quantifications of H3K27me3 levels normalized on H3. Data points were generated from an average of at least 4 biological replicates. **e**, Representative images of Ezh2/Lamin A immunofluorescence analysis on control or HGPS fibroblasts. **f** Quantification of number per nucleus of Ezh2 foci of CTRL and HGPS fibroblasts. Number of analyzed nuclei: CTRL001: 225; CTRL002: 265; CTRL004: 202; HGPS164: 188; HGPS169: 118. **g**, Representative images of H3K27me3/Lamin A immunofluorescence analysis on control or HGPS fibroblasts. **h**, Quantification of number per nucleus of H3K27me3 foci of CTRL and HGPS fibroblasts. Number of analyzed nuclei: CTRL001: 221; CTRL002: 339; CTRL004: 165; HGPS164: 251; HGPS169: 129. **i**, H3K27me3 signal distribution around the TSS region of all protein coding genes, based on GENCODE v27 annotation, calculated by deepTools using the genome-wide signal from SPP (see Online methods for details). The x-axis represents relative genomic position around the TSS (+/- 5Kb), and the y-axis represents average signal intensity. **k**, Transcripts differential expression (up-regulation) was assessed by comparing individual HGPS samples against the group of controls, then p-values were aggregated based on their chromatin state, considering only genes in SAMMY-seq domains (S4 vs S3) (see Online methods for details). The plot reports the number of transcripts (x-axis) and significance (Lancaster method aggregated p-values – y-axis) for each chromatin state (color legend) as defined by Roadmap Epigenomics[32].
Comparisons were done using a two-tailed t-test in (b, d) and nested anova in (f, h). n.s.: not significant.

CHAPTER 4.
UNCONVENTIONAL NUCLEAR ARCHITECTURE IN $CD4^+$ T LYMPHOCYTES UNCOUPLES CHROMATIN SOLUBILITY FROM FUNCTION

4.1 SAMMY-seq isolates heterochromatic domains in fibroblast samples

We recently developed SAMMY-seq (under patenting, application n. 18200482.0–1118), a technology for the genome-wide mapping of chromatin fractions with different solubility. The method is based on the sequential extraction of distinct nuclear fractions containing: soluble proteins (S1 fraction), DNase-sensitive chromatin (S2 fraction), DNase-resistant chromatin (S3 fraction) and the most condensed and insoluble portion of chromatin (S4 fraction) (Fig. 1a and Suppl. Fig. 1a). Adapting the original technology, designed for the analysis of proteins distribution[1–3], we obtained genome-wide mapping of the DNA content of each fraction (Fig. 1b). SAMMY-seq performed on 3 independent healthy skin primary fibroblast cell lines[4] shows a high reproducibility of signal for the less soluble S3 and S4 fractions (mean Spearman correlation coefficient 0.92 and 0.79, respectively) (Suppl. Fig. 1b), whereas the S2 fraction results more variable across biological replicates (mean Spearman correlation 0.27). Visual inspection of read coverage profiles reveals megabase-scale "bumps" in S3 and, more prominently, in S4 fraction (Fig. 1b). The Enriched Domain Detector (EDD) algorithm[5], specifically designed to identify broad regions of enrichment in high-throughput sequencing results, was used on our data to evaluate the relative enrichment of specific genomic regions across fractions. Pairwise comparisons between less accessible and more accessible fractions (S3vsS2, S4vsS3 and S4vsS2) highlight a well-defined set of regions (SAMMY-domains) consistently enriched in the less accessible fractions (Fig. 1c). These domains are conserved across samples, with the S4vsS2 comparison being the most consistent (Suppl. Fig. 1c). To functionally characterize the insoluble

regions, we compared the genome-wide read coverage with reference chromatin mark profiles for the same cell type. Using data from the Roadmap Epigenomics consortium[6] and other publications[7–9] we noticed that SAMMY-domains are inversely correlated with open chromatin marks (ATAC-seq, DNase-seq, H3K4me1) and polycomb-regulated H3K27me3 mark, while they positively correlate with H3K9me3, Lamin A/C and Lamin B heterochromatin-associated ChIP-seq signals (Fig. 2).

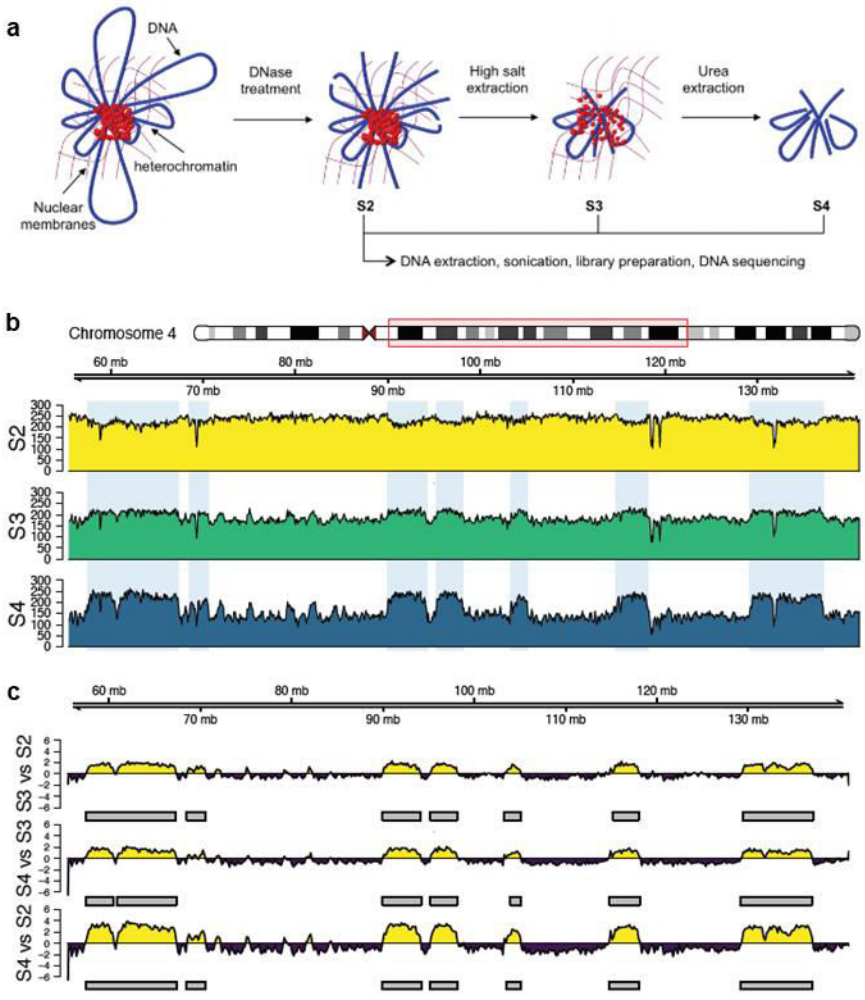
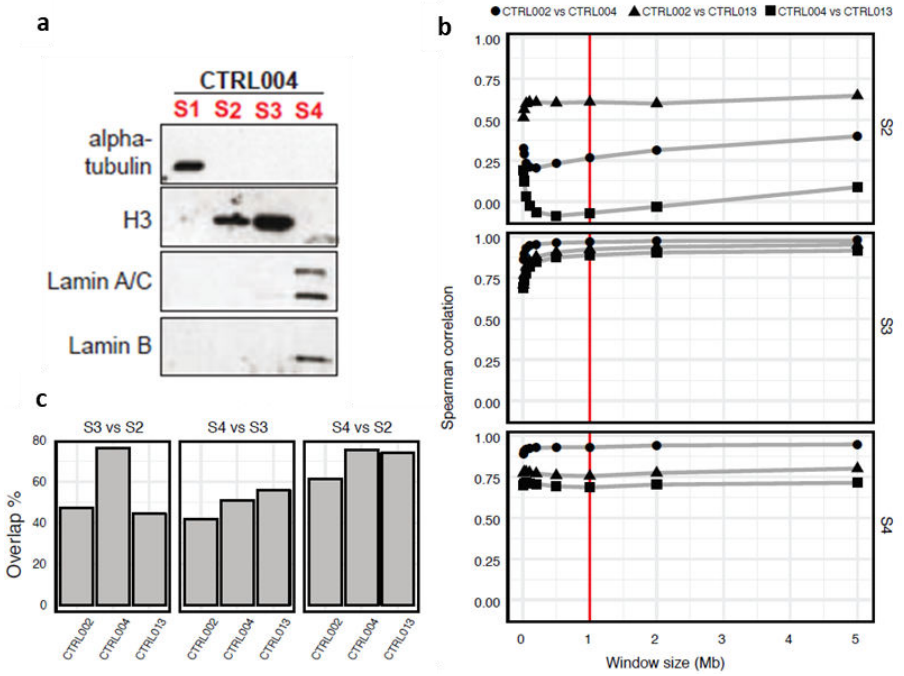


Figure 1 - SAMMY-seq isolates insoluble domains in fibroblast samples. a, Schematic representation of SAMMY-seq. Chromatin

fractions are sequentially isolated after DNase treatment (S2) and after high-salt (S3) and urea (S4) extractions. The associated genomic DNA is purified, sonicated and processed for high-throughput sequencing. **b**, Distribution of SAMMY-seq reads in each sequenced fraction of an healthy skin primary fibroblast sample (CTRL002) along a representative region (85Mb on chr4:55,480,346-141,005,766). Library size normalized read counts over 10Kb genomic bins are shown. **c**, Differential reads distribution across pairwise comparisons of SAMMY-seq fractions in CTRL002 along the genomic region in (b). Less accessible fractions are compared to more accessible ones used as reference (S3vsS2; S4vsS3; S4vsS2). Regions of signal enrichment or depletion over the reference are marked in yellow or purple, respectively. The smoothed differential signal is calculated with SPP. Significantly enriched regions (SAMMY-domains) are called with EDD and reported as grey boxes under the enrichment signal tracks.



Supplementary Figure 1 – Characterization of fibroblast chromatin fractions and SAMMY-domains. a, Western blot of chromatin fractionation experiment on a healthy skin primary fibroblast sample (CTRL004). Sequential extractions isolate soluble proteins (S1 fraction), DNase-sensitive chromatin (S2 fraction), DNase-resistant chromatin (S3 fraction), and the most compact and inaccessible chromatin (S4 fraction).

Equal amounts of each fraction were hybridized with indicated antibodies. **b**, Spearman correlation (y-axis) of read counts between healthy skin primary fibroblast sample pairs (shape of dots) over different bin sizes (x-axis) in the S2, S3 and S4 fractions. The red line shows the 1Mb bin size. **c**, Percentage of SAMMY-domains (S3vsS2, S4vsS3 and S4vsS2) conserved across all 3 healthy samples (computed as percent over per sample total SAMMY-domains size).

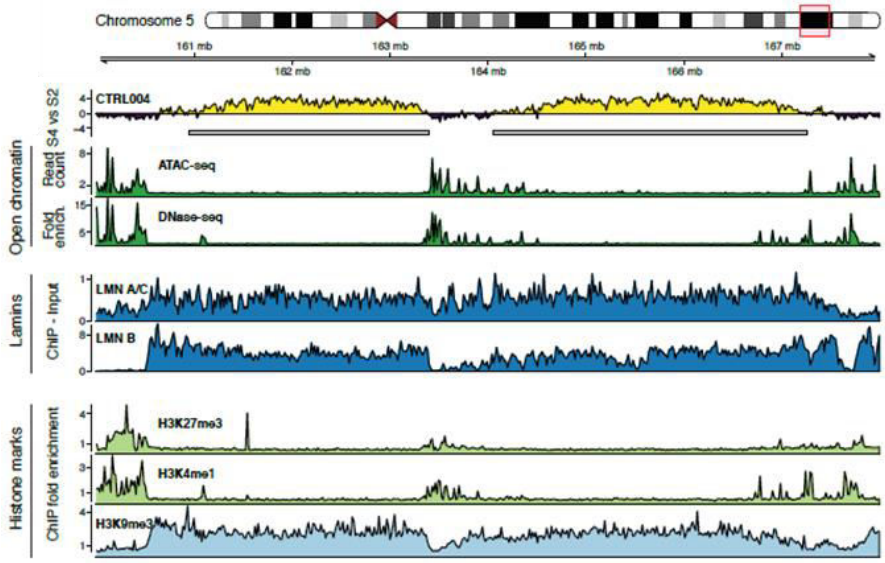
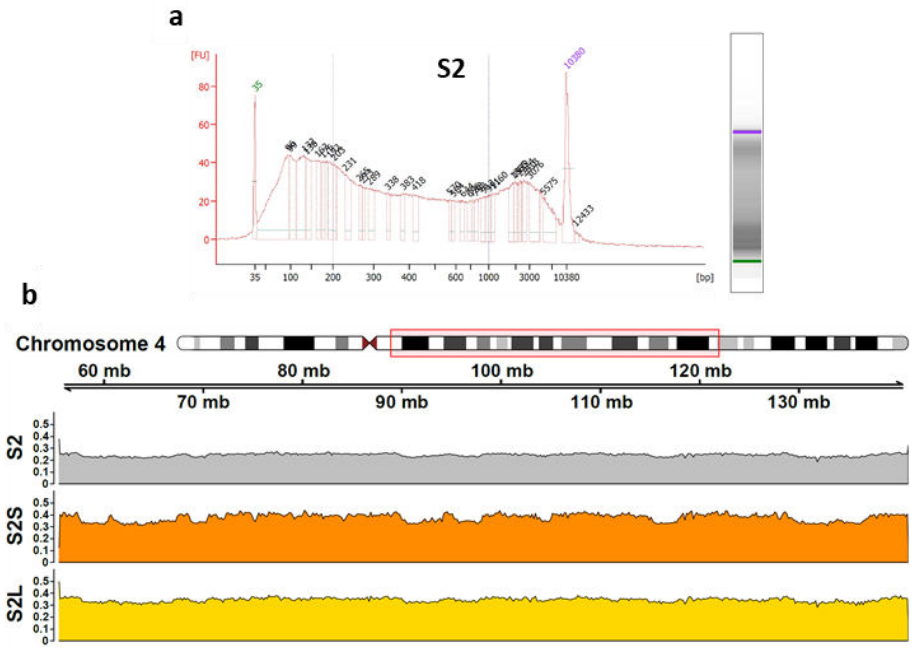


Figure 2 – Insoluble SAMMY-domains in fibroblast samples correlate with heterochromatin. Visualization of SAMMY-seq enrichment signal along with multiple chromatin marks on a representative region (8Mb on chr5:160,000,000-168,000,000). From top to bottom: SAMMY-seq enrichment signal (S4vsS2) in an healthy skin primary fibroblast sample (CTRL004) with grey boxes under the track showing the SAMMY-domains; tracks for open chromatin (ATAC-seq and DNase-seq – dark green); ChIP-seq tracks for lamins (Lamin A/C and Lamin B – dark blue); ChIP-seq tracks for histone marks (HMs) associated to PcG regulation (H3K27me3 – light green), active chromatin (H3K4me1 – light green) or heterochromatin (H3K9me3 – light blue).

4.2 A modified SAMMY-seq protocol isolates both euchromatic and heterochromatic domains in fibroblast and myoblast samples

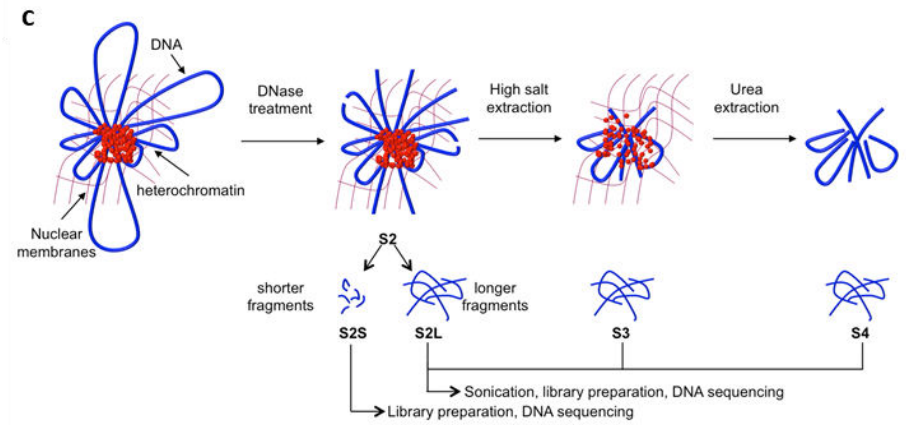
To further improve SAMMY-seq technology, we modified the DNAse step and we split the S2 fraction into two subfractions: S2 small (S2S) containing DNA fragments ranging from 50 to 300 bps and S2 large (S2L) gathering longer fragments (Suppl. Fig. 2a).

The new version of the SAMMY-seq technology (Suppl. Fig. 2c), applied on healthy skin primary fibroblasts, produced S2 subfractions (S2S and S2L) with a read coverage less uniformly distributed with respect to the previous S2 (Suppl. Fig. 2b), suggesting an improvement in the enrichment specificity.



Supplementary Figure 2 continued on the next page.

Supplementary Figure 2 continued.



Supplementary Figure 2 – Separation of S2S and S2L subfractions. **a**, Representative bioanalyzer run of S2 purified DNA before size-selection. **b**, Distribution of SAMMY-seq reads in total S2, S2S and S2L of a healthy skin primary fibroblast sample (CTRL004) along a representative region (85Mb on chr4:55,500,000-141,000,000). Library size normalized read counts are shown. **c,** Schematic representation of the new SAMMY-seq protocol.

Moreover, the "bumps" previously observed in S3 and S4 disappear, substituted by "valleys" in S2L and, more prominently, in S2S (Fig. 3a). Comparison of less accessible versus more accessible fractions still highlights a set of regions consistently enriched in the less accessible fractions and positively correlated with H3K9me3 and Lamin ChIP-seq signals (Fig.3b). In addition, with the opposite comparison of more accessible versus less accessible fractions, we isolate regions correlating with DNase-seq peaks, H3K36me3 and H3K4me3 ChIP-seq signals (Fig. 3c).

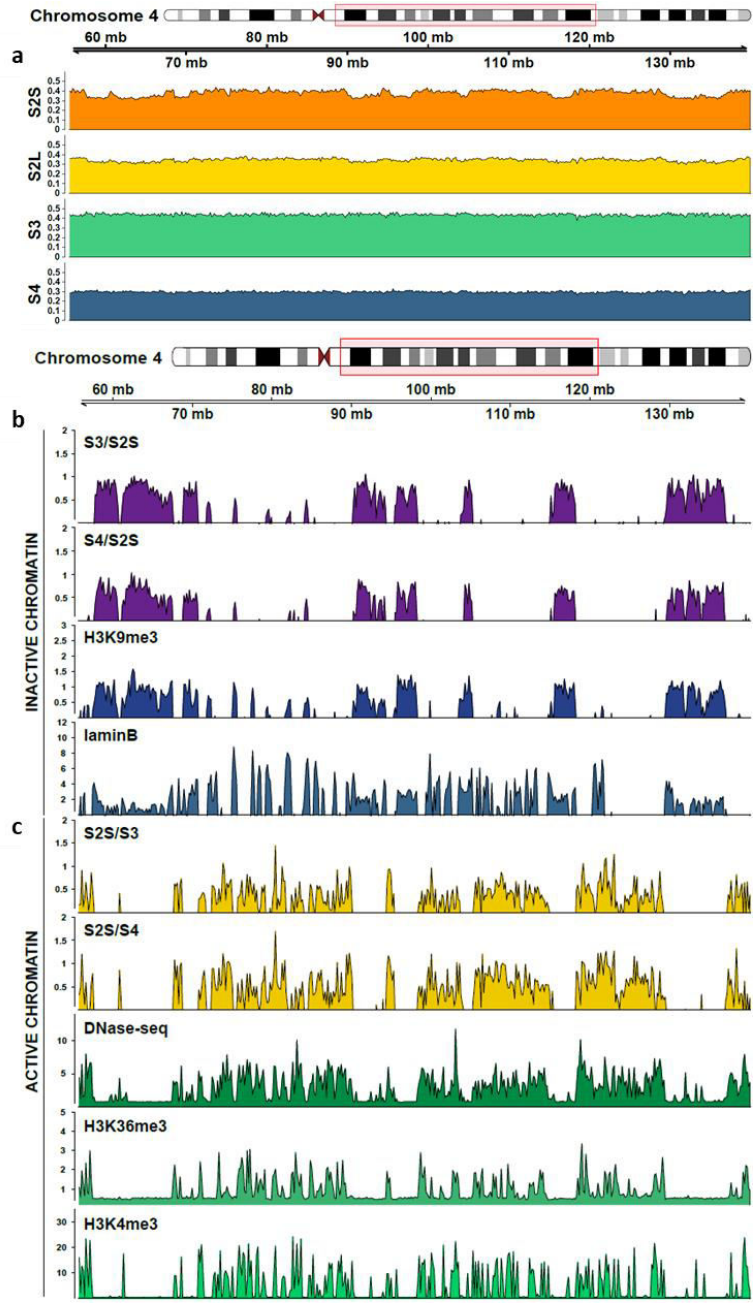


Figure 3 – Modified SAMMY-seq isolates heterochromatic and euchromatic domains in a fibroblast sample. a, Distribution of SAMMY-seq reads in each sequenced fraction of an healthy skin primary fibroblast sample (CTRL004) along a representative region (85Mb on chr4:55,500,000-141,000,000). Library size normalized read counts are

shown. **b**, Visualization of SAMMY-seq enrichment signal in insoluble compared to soluble fractions (S3vsS2S and S4vsS2S) of CTRL004 along the genomic region in (a). The smoothed differential signal is calculated with SPP and regions of signal enrichment over the reference are marked in purple. ChIP-seq tracks of heterochromatin mark H3K9me3 and Lamin B in the same region are also shown (blue). **c**, Visualization of SAMMY-seq enrichment signal in soluble compared to insoluble fractions (S2SvsS3 and S2SvsS4) of CTRL004 along the genomic region in (a). The smoothed differential signal is calculated with SPP and regions of signal enrichment over the reference are marked in yellow. DNase-seq tracks and ChIP-seq tracks of euchromatin marks H3K36me3 and H3K4me3 in the same region are also shown (green).

This finding suggested that the new version of SAMMY-seq protocol identifies, in a single assay on living cells, both euchromatic and heterochromatic domains, reconstructing the genome functional organization of the cells.

The application of the same protocol to C2C12 murine myoblast cells, a second cell population of great interest for the lab[3], confirmed the advantages of SAMMY-seq. As for human fibroblasts, in fact, regions enriched in insoluble domains (S4vsS2S) correlate with heterochromatin marks (H3K9me3 and Lamin B) while regions enriched in soluble domains (S2SvsS3) with euchromatin marks (H3K36me3 and H3K4me3) (Fig.4). Clustering of SAMMY-seq ratios and chromatin marks using the pairwise genome-wide Spearman correlation confirmed this observation, indicating the S2SvsS3 ratio as active and the S4vsS2s as inactive C2C12 compartments (Suppl. Fig. 3)

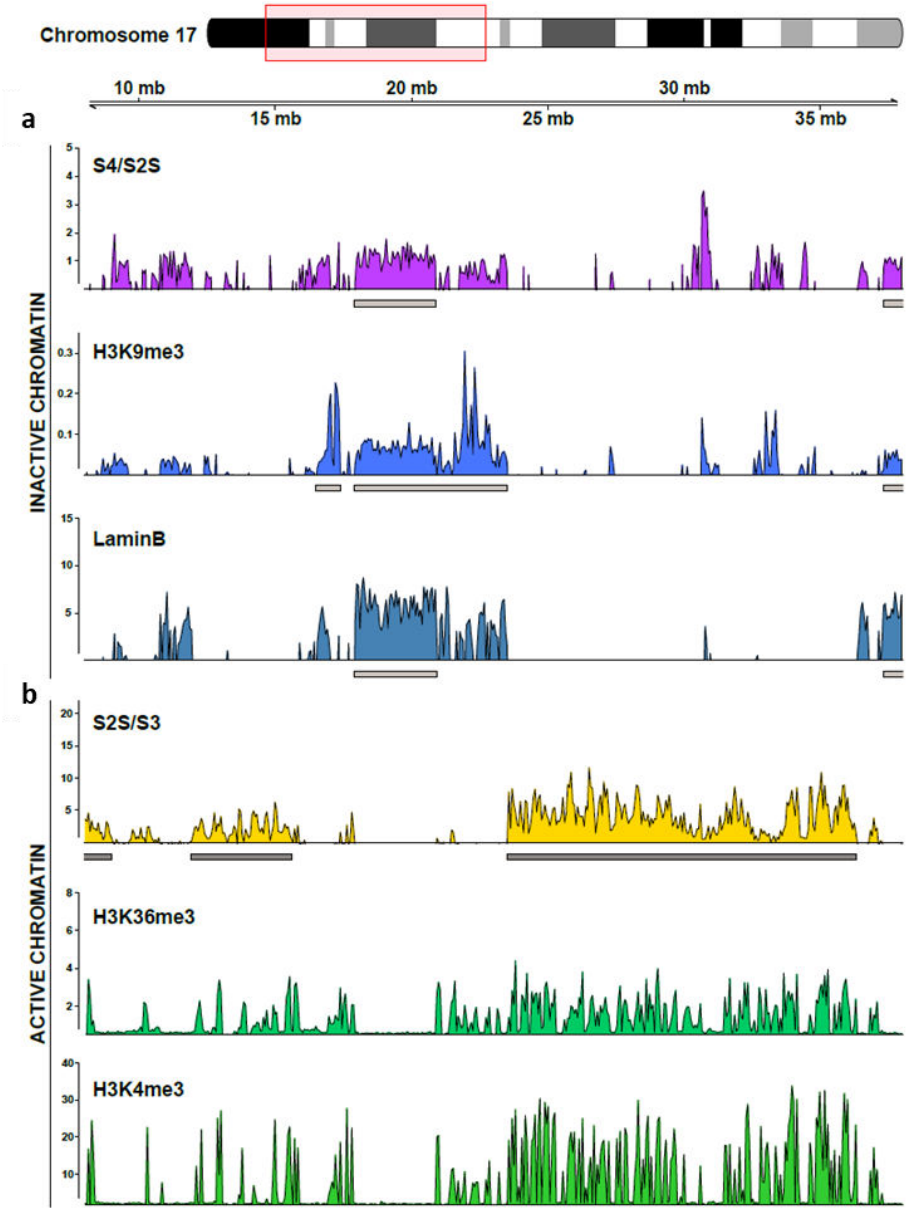
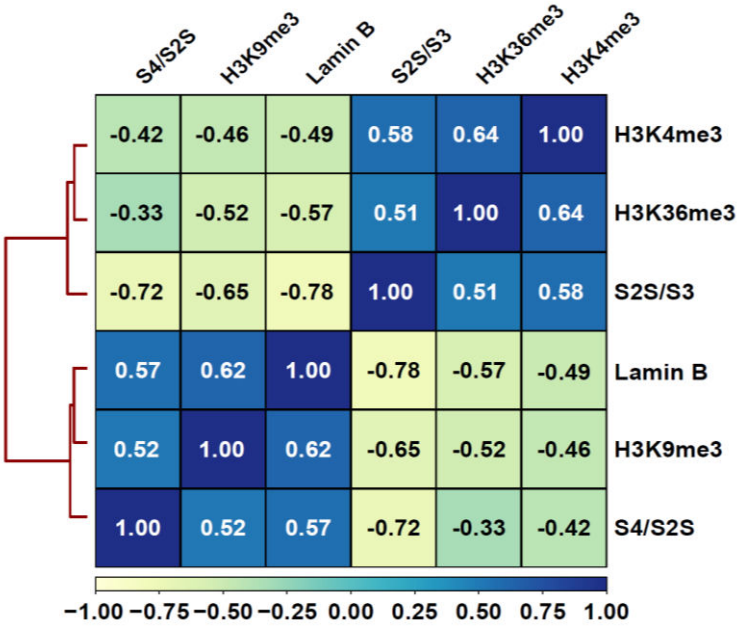


Figure 4 - SAMMY-seq isolates heterochromatic and euchromatic domains in a murine myoblast sample. a, Differential reads distribution across pairwise comparisons of SAMMY-seq fractions in a C2C12 sample along a representative region (30Mb on chr17:8,000,000-38,000,000). The less accessible fraction is compared to the more accessible one used as reference (S4vsS2S). Signal enrichment over the reference sample is marked in purple. The smoothed differential signal is

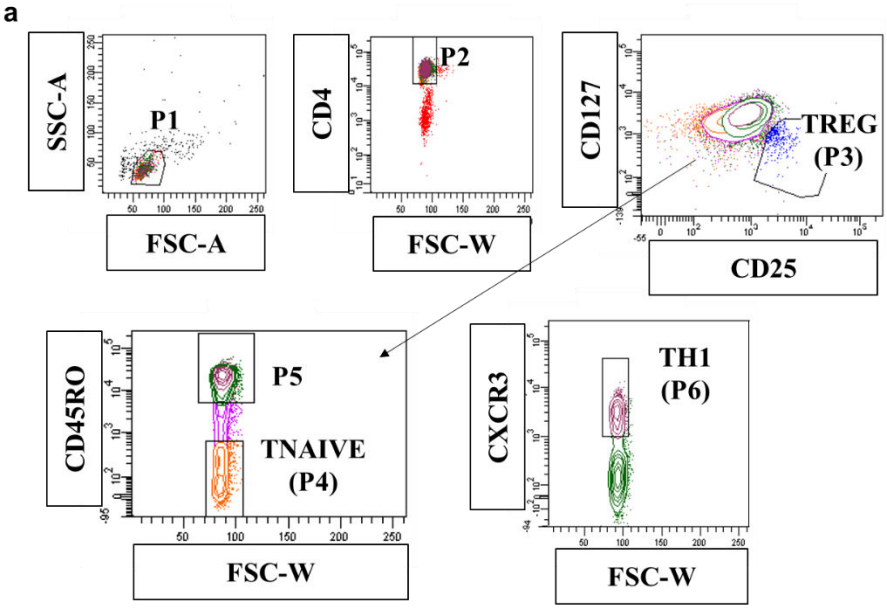
calculated with SPP, and significantly enriched regions (SAMMY-domains) are called with EDD and reported as grey boxes under the enrichment signal track. ChIP-seq signal tracks for heterochromatin-associated H3K9me3 (blue) and Lamin B (light blue) are also shown, together with their EDD domains. **b**, Differential reads distribution between the more and the less accessible fraction (S2SvsS3) in the same C2C12 sample along the region in (a). Signal enrichment over the reference sample is marked in yellow. The smoothed differential signal is calculated with SPP, and significantly enriched regions (SAMMY-domains) are called with EDD and reported as grey boxes under the enrichment signal track. ChIP-seq signal tracks for euchromatin-associated H3K36me3 (light green) and H3K4me3 (green) are also shown.



Supplementary Figure 3 – Clustering of SAMMY-seq ratios and chromatin marks discriminate active from inactive myoblast compartments. Clustering of soluble (S2SvsS3) and insoluble (S4vsS2S) SAMMY-seq comparisons with HMs and Lamin B ChIP-seq enrichments using the pairwise genome-wide Spearman correlation.

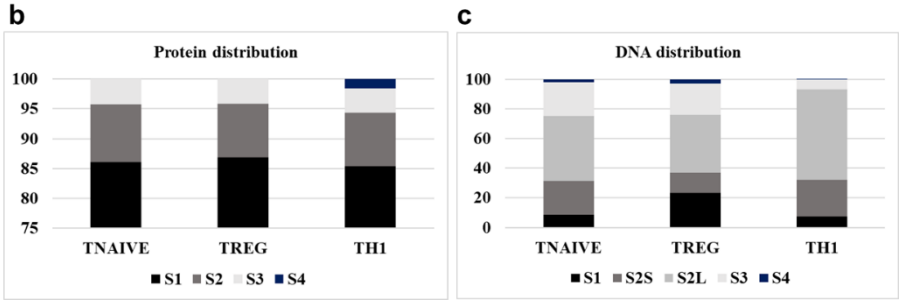
4.3 Chromatin fractionation reveals peculiar histone H3 distribution in CD4⁺ T lymphocytes

Central to all the clinical applications of immunotherapy[10,11], the immune system comprises extremely plastic phenotypes, each one resulting from the integration of a multitude of environmental cues. Although several cellular and molecular studies provided evidence for the multiple functions of immune subsets[12–16], their epigenetic mechanisms driving response to different stimuli are still far from being completely understood. To investigate the correspondence between chromatin conformation and gene expression programmes in T lymphocytes, we decided to take advantage of SAMMY-seq technology. We selected from the peripheral blood of healthy donors the three populations of Naïve (Tnaïve), Helper 1 (Th1) and Regulatory (Treg) cells (Suppl. Fig. 4a) and we fractionated their chromatin.



Supplementary Figure 4 continued on the next page.

Supplementary Figure 4 continued.



Supplementary Figure 4 – Human primary lymphocytes sorting panel and macromolecules distribution across chromatin fractions. **a,** Lymphocytes were gated for their physical parameters and for the presence of CD4 surface antigen (top left). Then, $CD25^{+}$ $CD127^{dim}$ Treg cells were firstly isolated (top right). From the remaining population $CD45RO^{-}$ Tnaive (bottom left) and $CD45RO^{+}$ $CXCR3^{+}$ Th1 cells (bottom right) were extracted. **b**, Relative abundance of proteins in each chromatin fraction of lymphocyte samples (average of at least two independent assays) **c**, Relative abundance of DNA in each SAMMY-seq fraction of lymphocyte samples (average of at least three independent assays).

Both protein and DNA analysis revealed a clear shortage of material in the S4 fraction, representing 1% or less of the total protein and DNA extracted (Suppl. Fig. 4b and c). However, western blot analysis of proteins distribution confirmed the proper fractionation: alpha tubulin is retrieved in the cyto/nucleosolic extract (S1), histone H3 in DNase (S2) and high-salt (S3) extracted fractions, and finally Lamin B in the highly denaturating urea solution (S4) (Fig. 5a). However, differently from what observed in fibroblasts (Fig. 5b and c, Suppl. Fig. 1a), the highest abundance of H3 is found in S2 fraction (Fig. 5a and c). Going into the detail of H3 covalently modified forms, we noticed that heterochromatin-associated H3K9me3 resides prevalently in the more soluble chromatin fraction (Fig. 5a and d). On an opposite trend, a substantial portion of the transcription-associated H3K36me3 and of the polycomb-mediated

H3K27me3 are retained through DNase digestion (Fig. 5a and d). As H3K27me3, also Bmi1, member of Polycomb Repressive Complex 1 (PRC1), was found prevalently in S2 and S3 (Fig. 5a), while fibroblasts retain it prevalently in S4 (Fig. 5b). Finally, the nuclear structural protein Lamin A/C, completely absent from Tnaive cells as described in literature[17], shows in Th1 and Treg cells an expected presence in S4, but also a weak signal throughout S2 and S3 fractions (Fig.5a).

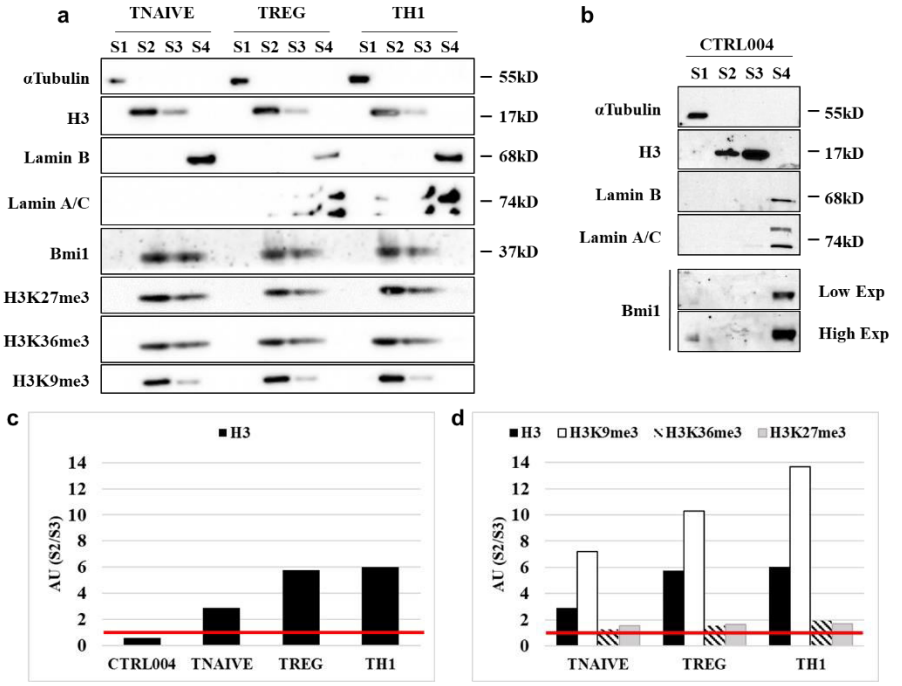


Figure 5 – Protein distribution in lymphocytes and fibroblasts chromatin fractions. a, Western blot on chromatin fractions of Tnaive, Treg and Th1 samples. Equal amounts of each fraction (except for S4, that was only quantifiable in Th1) were hybridized with indicated antibodies. Alpha-tubulin, histone H3 and Lamin B were used as loading controls respectively for S1, S2 and S3, S4. **b**, Western blot on chromatin fractions of a healthy skin primary fibroblast sample (CTRL004). Equal amounts of each fraction were hybridized with indicated antibodies. **c, d**, Graphs showing quantification of H3 in S2 fraction of CTRL004, Tnaive, Treg and Th1 cells (c) and quantification of H3, H3K9me3, H3K36me3 and H3K27me3 in S2 fraction of Tnaive Treg and Th1 cells (d). Values

normalized on S3. The red line indicates the value 1, obtained in case of an equal quantification of S2 and S3.

4.4 SAMMY-seq isolates reproducible insoluble domains in CD4$^+$ T lymphocytes

Visual inspection of SAMMY-seq linear tracks evidenced, in all lymphocytes populations analysed, a scattered S4 profile. Bumps and valleys that in fibroblasts are typical of S2 (Fig. 3a and Suppl. Fig. 2c) seem here to be shifted in the S4 fraction (Fig. 6a, Suppl. Fig. 5a and b). Furthermore, some regions of S4 are filled with sharp and high spikes (Fig. 6a). Pairwise comparison of less and more soluble fractions revealed that lymphocytes, as well as fibroblasts and myoblasts, retain some specific genomic regions until S4 extraction (Fig. 6b and Suppl. Fig. 5a and b). We used EDD algorithm to detect these domains and we calculated their mean overlap between different donors using Jaccard index (Fig. 6c). The results, ranging from 0.5 to 0.69, evidence a high conservation of insoluble SAMMY-domains in all the lymphocytes subsets analysed.

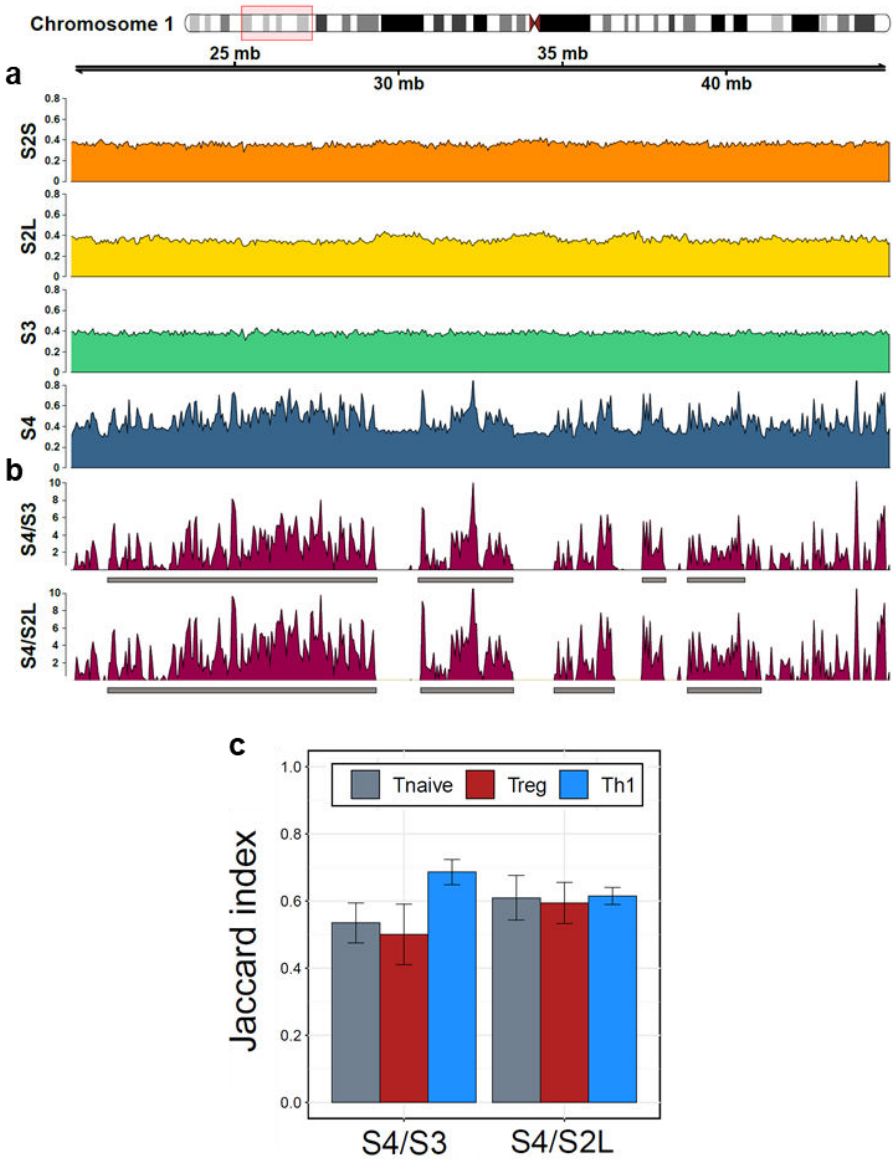
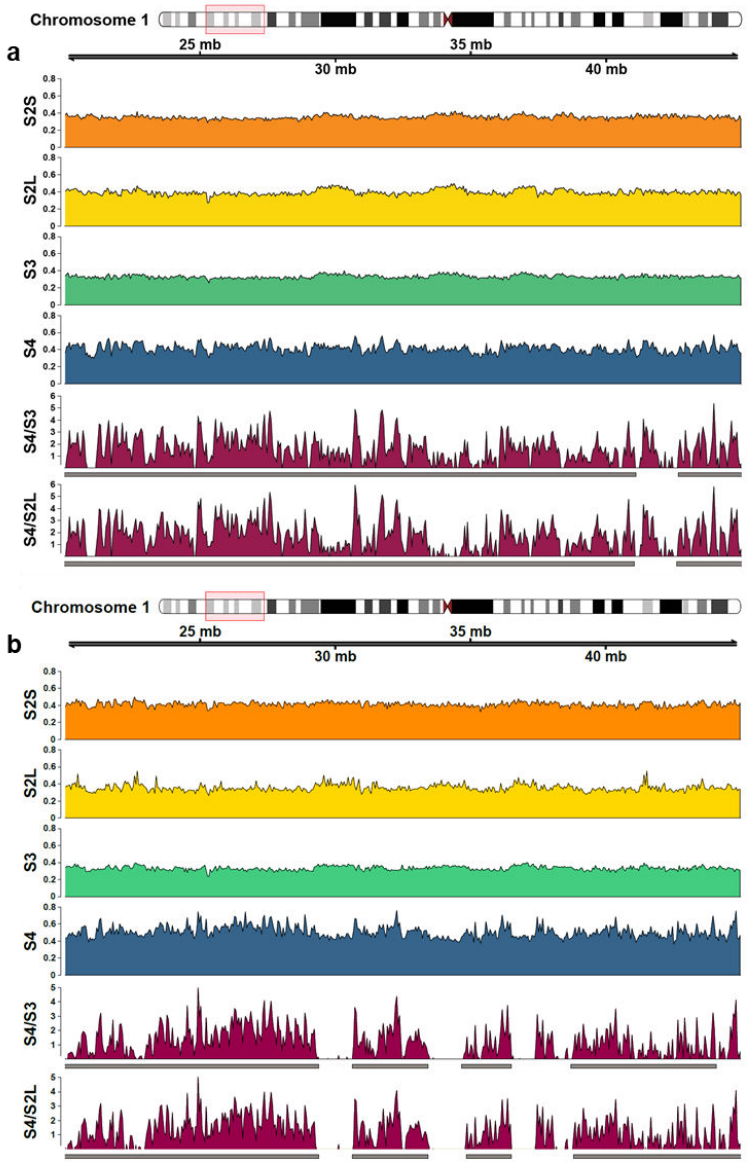


Figure 6 – Lymphocytes SAMMY-seq fractions and the conservation of insoluble domains in lymphocytes samples from different donors. a, Distribution of SAMMY-seq reads in each sequenced fraction of a Th1 sample (donor 3) along a representative region (25Mb on chr1:20,000,000-45,000,000). Library size normalized read counts are shown. **b**, Visualization of SAMMY-seq enrichment signal in insoluble compared to soluble fractions (S4vsS2L, S4vsS3) of the sample in (a) along the same region. The smoothed differential signal is calculated with SPP and regions of signal enrichment over the reference are marked in purple.

Significantly enriched regions (SAMMY-domains) are called with EDD and reported as grey boxes under the enrichment signal tracks. **c**, Overlap (Jaccard Index - JI) between SAMMY-domains S4vsS3 and S4vsS2L of differ donors. Values derive from the analysis of at least three independent donors for each population.

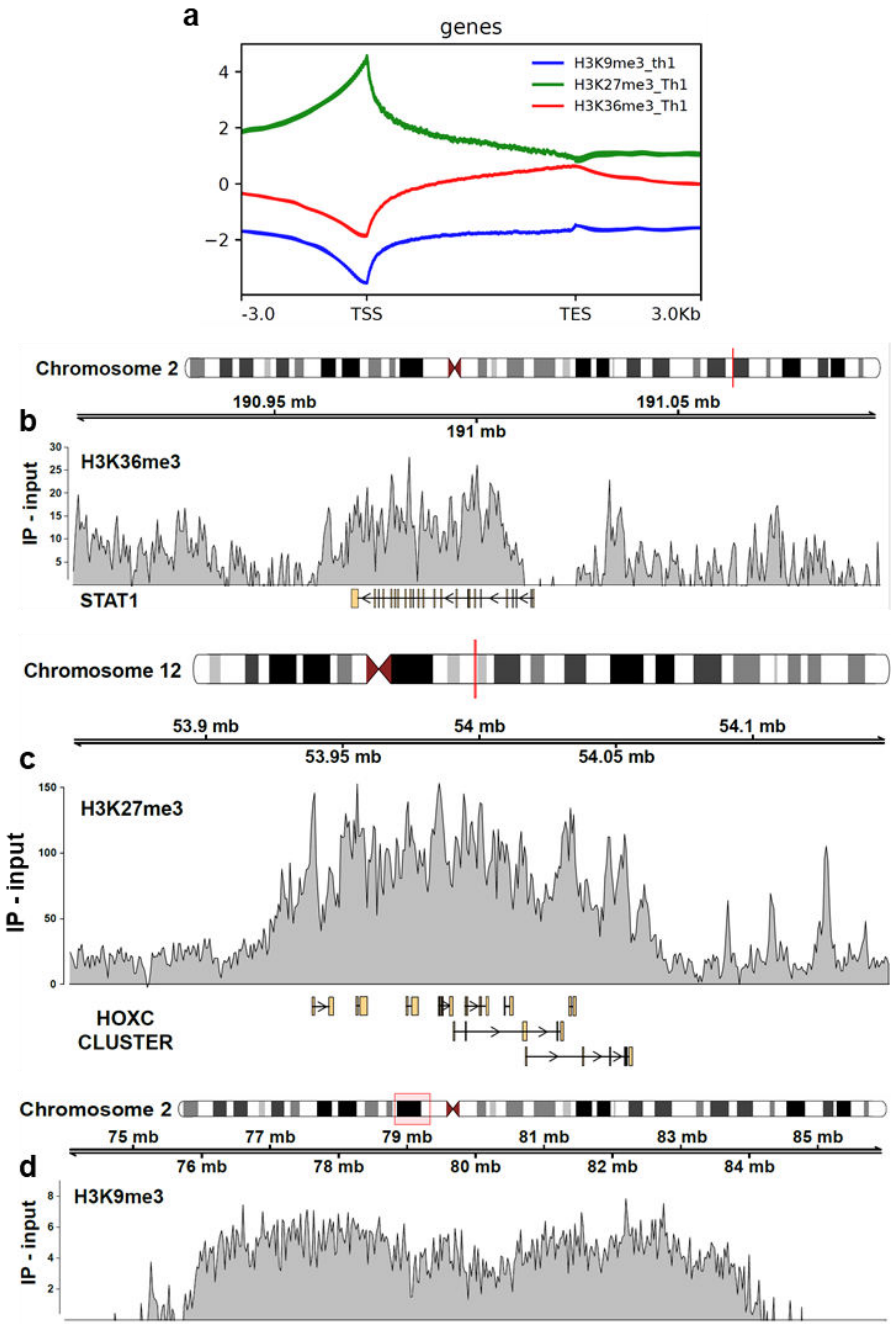


Supplementary Figure 5 – SAMMY-seq fractions and insoluble domains in Tnaive and Treg lymphocytes samples. a, **b** Distribution of SAMMY-seq reads in each sequenced fraction of a Tnaive sample

(donor 3) (a) and a Treg sample (donor 2) (b) along a representative region (25Mb on chr1:20,000,000-45,000,000). Library size normalized read counts are shown. SAMMY-seq enrichment signal in insoluble compared to soluble fractions (S4vsS2L, S4vsS3) along the same region is also shown. The smoothed differential signal is calculated with SPP and regions of signal enrichment over the reference are marked in purple. Significantly enriched regions (SAMMY-domains) are called with EDD and reported as grey boxes under the enrichment signal tracks.

4.5 Lymphocytes insoluble chromatin matches euchromatin marks

To analyse the function of insoluble domains, we compared SAMMY-seq ratio tracks with ChIP-seq tracks from the same cell populations. We used publicly available data on Tnaive and Treg[6], while we produced our own datasets on Th1 cells. We performed H3K27me3, H3K9me3 and H3K36me3 ChIP-seq on Th1 and checked their signal profile (IP/input) both at genome-wide level and on specific genomic regions (Suppl. Fig. 6). As expected, we found depletion of H3K9me3 from coding regions, enrichment of H3K27me3 on the TSS and enrichment of H3K36me3 along the gene body (Suppl. Fig. 6a). To further check ChIP-seq quality, we analysed enrichment profiles on target regions, confirming H3K36me3 enrichment on the body of the expressed STAT1 gene (Suppl. Fig. 6b), H3K27me3 enrichment on the HOXC cluster (Suppl. Fig. 6c) and a large H3K9me3 domain in a gene-desert genomic region (Suppl. Fig. 6d).



Supplementary Figure 6 – Th1 ChIP-seq validation. a, ChIP-seq metaprofile enrichment of H3K9me3 (blue), H3K36me3 (red) and H3K27me3 (green) marks in a Th1 sample. The enrichment signal of the IP over the input is calculated with SPP, averaged along the entire hg38

gene annotation and visualized using deepTools suite. **b**, Representative H3K36me3 track (IP/input) on a highly expressed gene (STAT1) (chr2:190,900,000-191,100,000). **c**, Representative H3K27me3 track (IP/input) on the PcG target region of the HOXC cluster (chr12:53,800,000-54,100,000). **d**, Representative H3K9me3 track (IP/input) on a representative gene desert region (12Mb on chr2:74,000,000-86,000,000).

The comparison of SAMMY-seq enrichments with ChIP-seq tracks highlighted that in lymphocytes, differently from myoblasts and fibroblasts (Fig. 7a and b), insoluble regions do not correlate with H3K9me3 mark (Fig. 7c). Instead, regions enriched in S4 fraction look to be mutually exclusive with H3K9me3 domains and to correlate with active chromatin marks such as H3K36me3 (Fig. 8a, Suppl. Fig. 7a and b).

Good correlation with heterochromatin mark H3K9me3 was found in lymphocytes soluble regions, extracted in DNase and high-salt buffers (S2 and S3) (Fig. 8b and Suppl. Fig. 7a and b). Soluble SAMMY-domains are highly conserved between donors (Fig. 9a), with Jaccard similarity indexes between 0.5 and 0.76 in all lymphocytes subsets (Fig. 9b).

The analysis of PcG-mediated H3K27me3 mark revealed apparently no correlation with either accessible or inaccessible SAMMY-domains (Fig. 8b and Suppl. Fig. 7a and b).

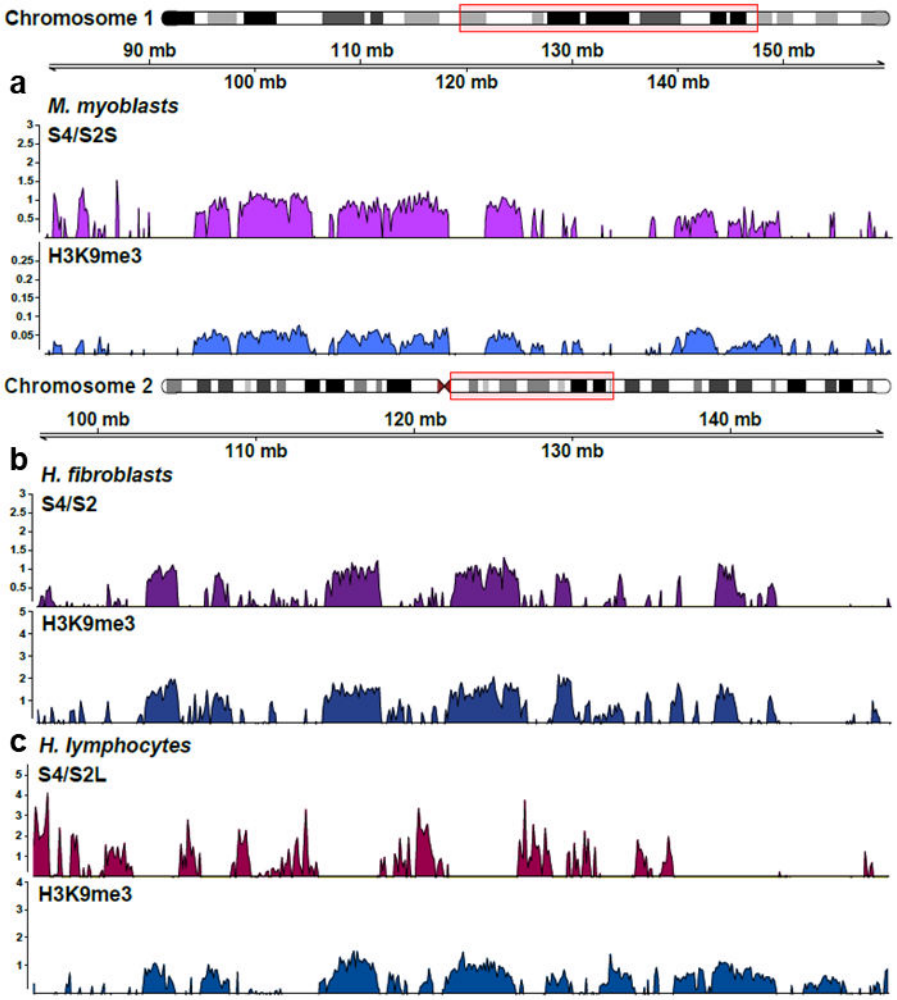


Figure 7 – Lymphocytes insoluble regions anti-correlate with constitutive heterochromatin. a, Visualization of SAMMY-seq enrichment signal in insoluble compared to soluble fraction (S4vsS2S) of a C2C12 sample along a representative region (80Mb on chr1:80,000,000-160,000,000). The smoothed differential signal is calculated with SPP and regions of signal enrichment over the reference are marked in purple. ChIP-seq track for H3K9me3 in the same region is also shown (blue). **b**, **c** Visualization of SAMMY-seq enrichment signal in insoluble compared to soluble fraction (S4vsS2) of a fibroblast sample (CTRL004) (b) and a Tnaive sample (donor 3) (c) along a representative region (54Mb on chr2:96,000,000-150,000,000). The smoothed differential signal is calculated with SPP and regions of signal enrichment over the reference are marked in purple. ChIP-seq tracks for H3K9me3 in the same region are also shown (blue).

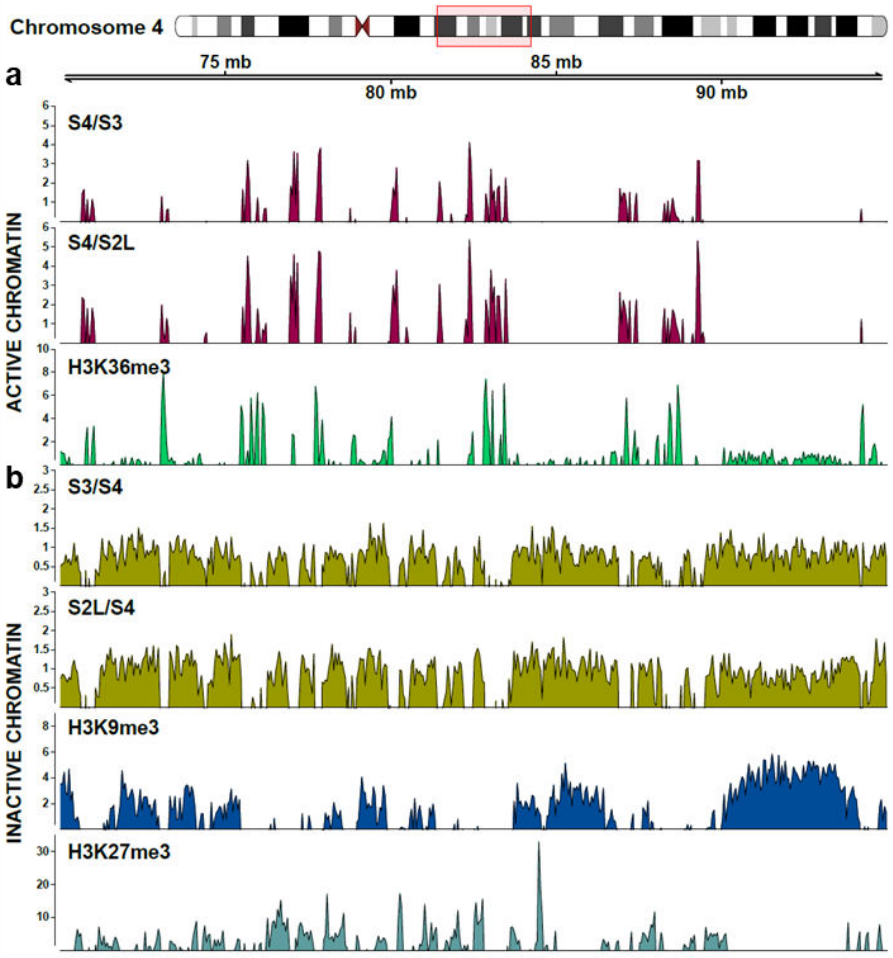
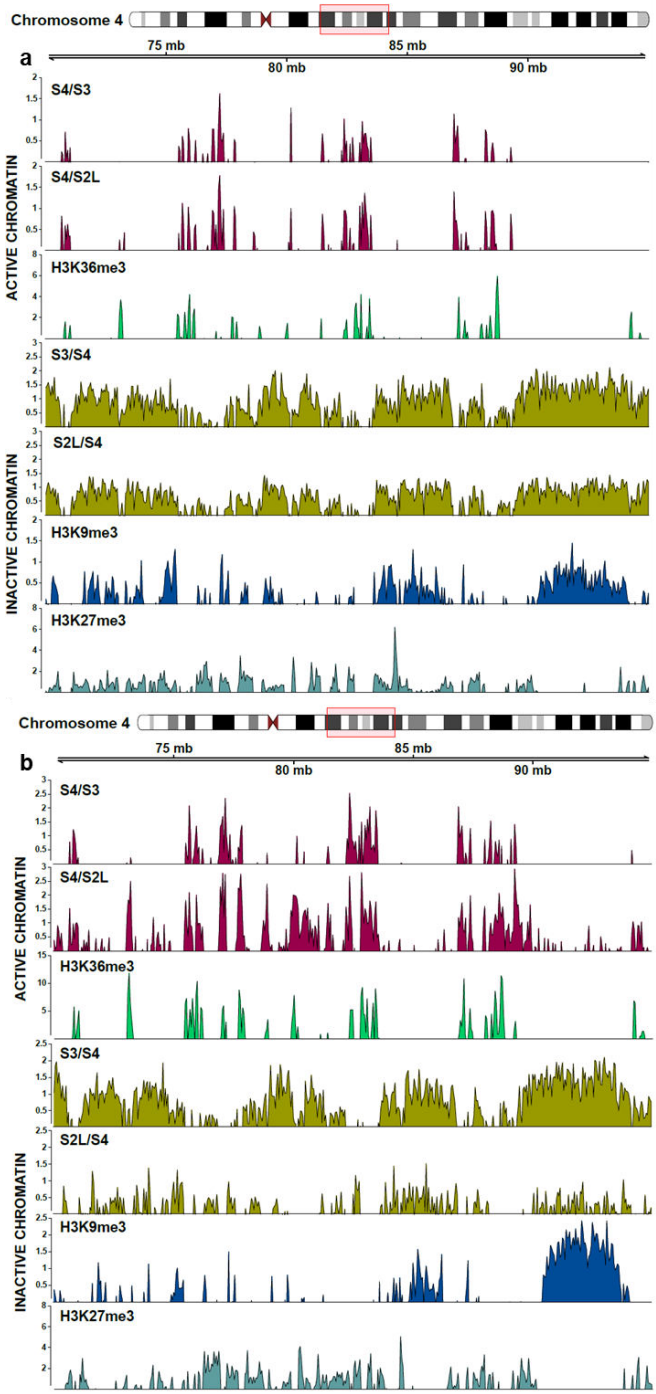


Figure 8 – Th1 insoluble regions correlate with euchromatin marks, while heterochromatin marks correlate with soluble regions. a, b Differential reads distribution across pairwise comparisons of SAMMY-seq fractions in a Th1 sample (donor 3) along a representative genomic region (45Mb on chr4:70,000,000-95,000,000). Enrichment in less accessible fractions (S4vsS3; S4vsS2L) is marked in purple (a), while enrichment in more accessible fractions (S2LvsS4, S3vsS4) is marked in yellow-green (b). The smoothed differential signal is calculated with SPP. ChIP-seq signal tracks for transcription-associated H3K36me3 (green), heterochromatin-associated H3K9me3 (blue) and polycomb-associated H3K27me3 (cyan) are also shown.



Supplementary Figure 7 – Tnaive and Treg cells also show association of soluble regions with heterochromatin and insoluble regions with euchromatin. a, b Differential reads distribution across

pairwise comparisons of SAMMY-seq fractions in a Tnaive sample (donor 3) (a) and a Treg sample (donor 4) (b) along a representative genomic region (45Mb on chr4:70,000,000-95,000,000). Enrichment in less accessible fractions (S4vsS3; S4vsS2L) is marked in purple, while enrichment in more accessible fractions (S2LvsS4, S3vsS4) is marked in yellow-green. The smoothed differential signal is calculated with SPP. ChIP-seq signal tracks for transcription-associated H3K36me3 (green), heterochromatin-associated H3K9me3 (blue) and polycomb-associated H3K27me3 (cyan) are also shown.

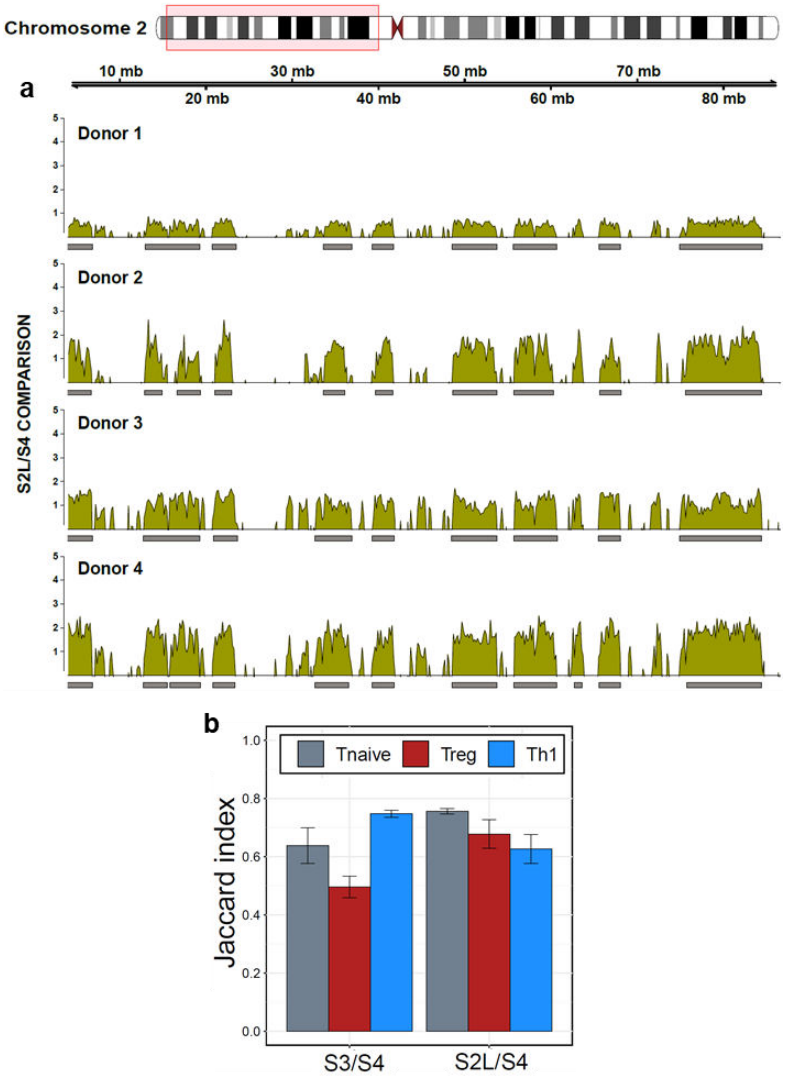


Figure 9 – Soluble SAMMY-domains are highly conserved among donors. a, Differential reads distribution across S2LvsS4 ratio in Th1

samples from four different donors along a representative genomic region (83Mb on chr2:4,000,000-86,600,000). The smoothed differential signal is calculated with SPP and the signal enrichment over the reference sample is marked in yellow-green. Significantly enriched regions (SAMMY-domains) are called with EDD and reported as grey boxes under the enrichment signal tracks. **b,** Overlap (Jaccard Index - JI) between SAMMY-domains (S2LvsS4 and S3vsS4) identified in different donors. Values derive from the analysis of at least three independent donors for each population.

Genome-wide Spearman correlation between SAMMY-seq fractions read coverage and histone marks ChIP-seq enrichments, evaluating on a scale 0 to 1 the pairwise similarity, supported the observations derived from visual inspection: in Tnaive and Treg cells the first three fractions clearly correlate with heterochromatic H3K9me3, while only the last one correlates with active H3K36me3 and H3K4me3 (Fig. 10). Conversely, in Th1 cells both euchromatin and heterochromatin marks strongly correlate with S2S fraction. In all lymphocyte subsets H3K27me3 mildly correlates with all the fractions, in particular S2L and S3.

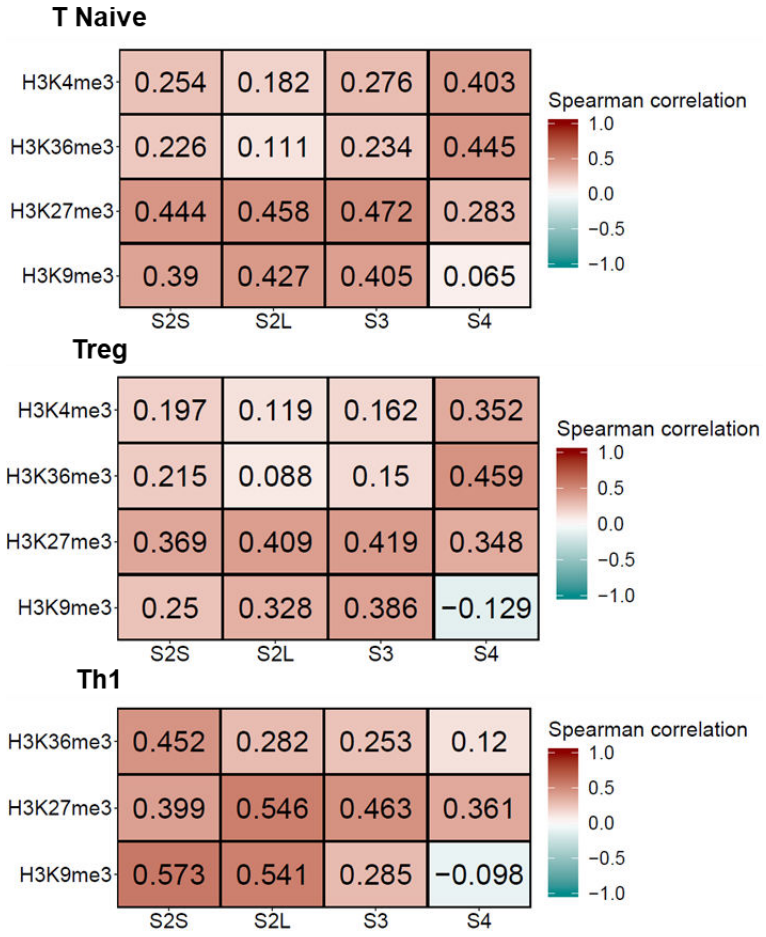
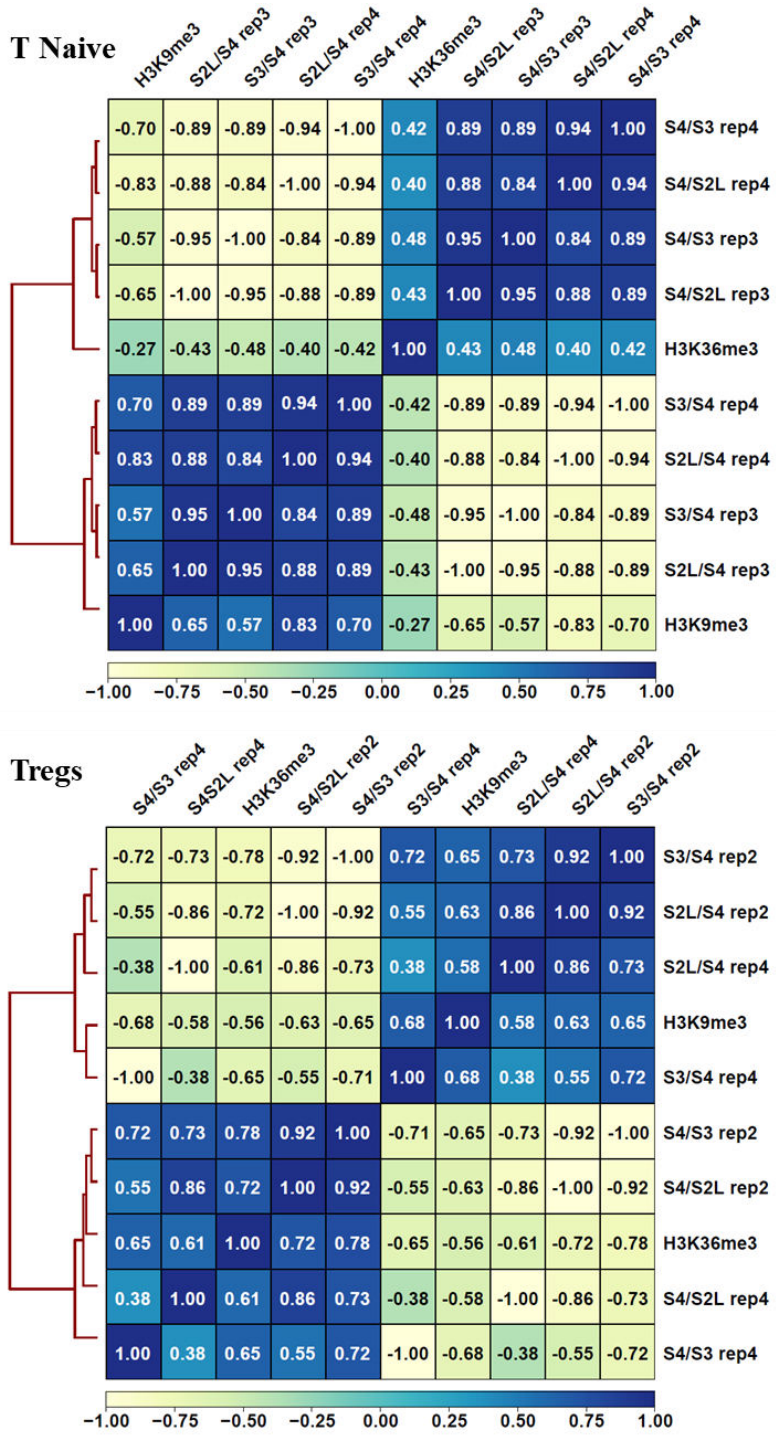


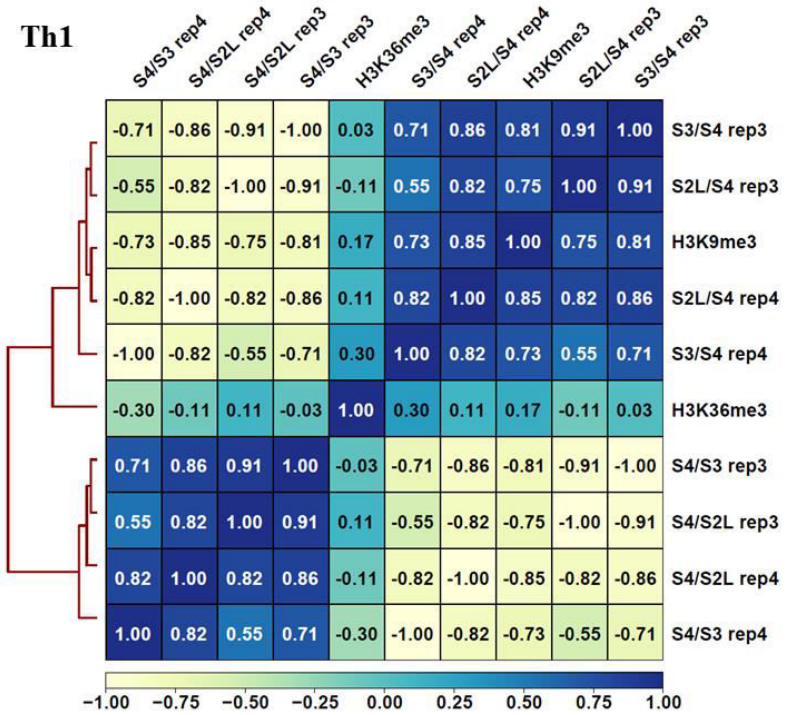
Figure 10 – SAMMY-seq fractions represent different chromatin states. a, Genome-wide Spearman correlation between read coverage in individual SAMMY-seq fractions and HMs ChIP-seq enrichments (IP/input) in Tnaive, Treg and Th1 lymphocytes.

Moving to the correlation of HMs with the ratio between SAMMY-fractions, we confirmed a neat clustering of soluble regions (S2LvsS4 and S3vsS4) with H3K9me3 ChIP-seq tracks (Suppl. Fig. 8). On the other hand, H3K36me3 ChIP-seq tracks cluster with insoluble chromatin (S4vsS3 and S4vsS2L) in Treg and, less evidently, in Tnaive. In Th1, H3K36me3 substantially divides the soluble from the insoluble cluster, with no preference for either of them.



Supplementary Figure 8 continued on the next page.

Supplementary Figure 8 continued.



Supplementary Figure 8 – Clustering of SAMMY-seq ratios and chromatin marks discriminate functional chromatin compartments in CD4$^+$ T lymphocyte samples. Clustering of soluble (S2LvsS4 and S3vsS4) and insoluble (S4vsS2L and S4vsS3) SAMMY-seq comparisons and HMs ChIP-seq in Tnaive, Treg and Th1 lymphocytes. Pairwise genome-wide Spearman correlation. For each population, comparison from two independent donors are shown.

Principal component analysis (PCA) on SAMMY-domains further proved the diversity between soluble (S2LvsS4 and S3vsS4) and insoluble (S4vsS3 and S4vsS2L) domains, which clearly separated on the principal component 1 (PC1), accounting for 88% (Fig. 11a) or even 95% (Fig. 11b) of the diversity. Of note, both the pair of soluble (S2LvsS4 and S3vsS4) and the pair of insoluble domains (S4vsS3 and S4vsS2L) demonstrated instead similarity among them (data not shown).

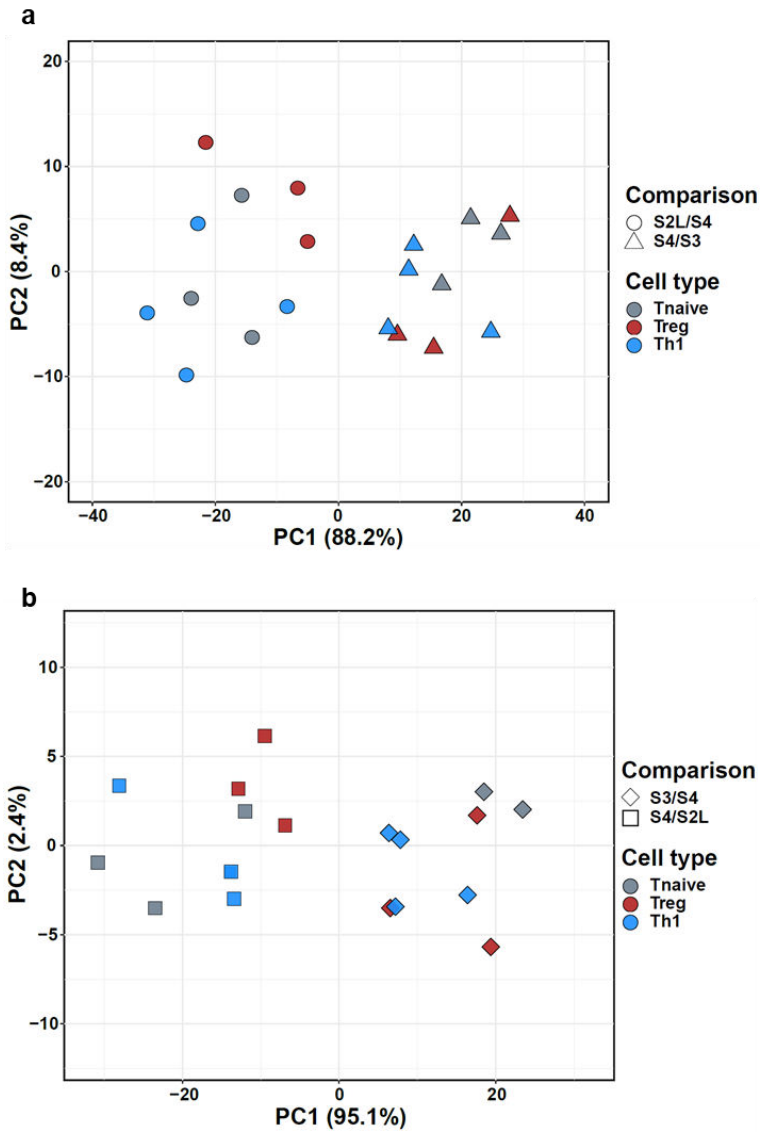
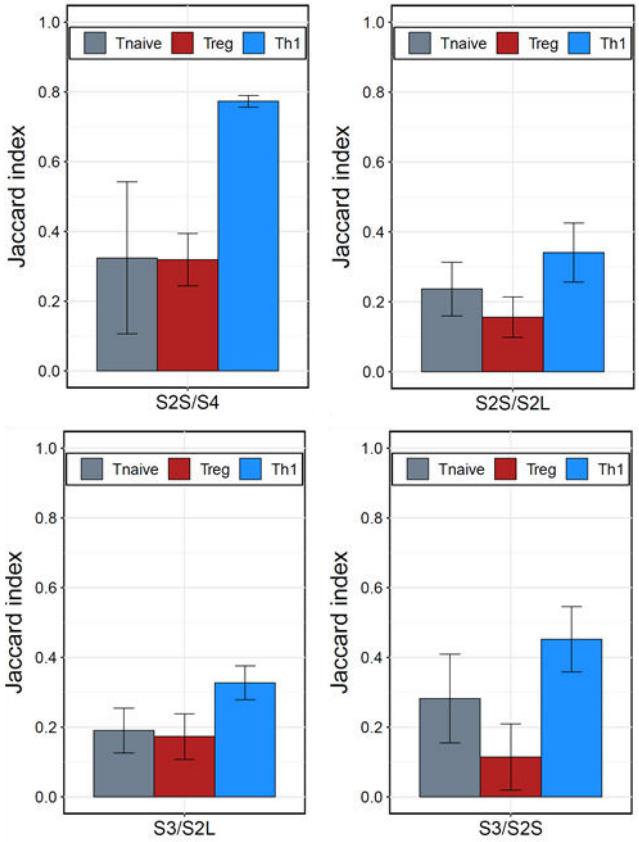


Figure 11 – Soluble and insoluble SAMMY-domains evidence different chromatin states. **a, b** Principal Component Analysis (PCA) on two pairs of soluble and insoluble domains: S2LvsS4-S4vsS3 (a) and S3vsS4-S4vsS2L (b). The analysis is performed examining the differential signal calculated by SSP on the intersection between the enriched regions called by EDD in at least 3 donors for each cell type and merged along the 4 comparisons investigated.

We then examined SAMMY-domains extracted by EDD from other possible ratios between chromatin fractions. Jaccard index, measure of domains consistency between different donors, always scored higher in Th1 cells (Suppl. Fig. 9). In particular, Th1-domains enriched in the most soluble over the most insoluble fraction (S2SvsS4) showed an almost complete conservation between different donors.

As a further characterization of the more consistent SAMMY-domains individuated by Jaccard index, we evaluated the genome coverage of the intersection of domains detected in all donors (Table 1). We noticed that insoluble domains are lower in number and cover a minor portion of the genome (8-13%) with respect to soluble domains (12-28%).

Given the similar coverage of the new class of SAMMY-domains identified in Th1 (S2SvsS4) with other soluble domains, we considered a possible redundancy. Indeed, we found that S2SvsS4 domains share 22% of genome coverage with S2LvsS4 (data not shown). All together, these data indicate both a high stability of soluble fractions in Th1 cells from different donors and a similarity between the two S2 subfractions in this cell type.



Supplementary Figure 9 – Consistency of different SAMMY-domains between donors. Overlap (Jaccard Index - JI) between donors SAMMY-domains S2SvsS4 and S2SvsS2L, S3vsS2L and S3vsS2S. A minimum of three independent donors for each population were analysed.

a

S4vsS2L	# of domains	coverage (% of genome)	# genes
Tnaive	138	11%	12817
Treg	94	8%	9590
Th1	100	10%	12273

S4vsS3	# of domains	coverage (% of genome)	# genes
Tnaive	147	13%	13645
Treg	120	10%	12317
Th1	83	8%	10228

b

S2LvsS4	# of domains	coverage (% of genome)	# genes
Tnaive	199	21%	3425
Treg	155	12%	3042
Th1	174	26%	4200

S3vsS4	# of domains	coverage (% of genome)	# genes
Tnaive	225	28%	4760
Treg	188	23%	3951
Th1	96	20%	3348

S2SvsS4	# of domains	coverage (% of genome)	# genes
Tnaive	36	3%	565
Treg	43	5%	687
Th1	144	29%	5456

Table 1 – Genome coverage of SAMMY-domains inversely correlate with gene density. a, b Number, percentage of genome covered and number of contained genes for soluble (a) and insoluble (b) SAMMY-domains detected in Tnaive, Treg and Th1 cells. Only domains detected in all donors are considered.

For a more precise investigation of the relationship between histone marks and SAMMY-domains, we analysed their genome-wide pattern along domains and at their borders (Fig. 11). H3K9me3 showed a trend of correlation with soluble domains and anti-correlation with insoluble domains, which becomes especially significant in Th1 cells (Fig. 11a). The transcription-associated H3K36me3 is not correlated to any SAMMY-domain in Th1 cells, while it follows a trend opposite to H3K9me3 in Tnaive and especially in Treg cells (Fig. 11b). The polycomb-associated H3K27me3 displays no correlation to any domain and cell type, except for an anti-correlation with Th1 soluble domains (Fig. 11c). Rather, signs of correlation can be detected at the border of domains, especially in Th1

cells: small and sharp peaks rise close to the border of soluble fractions, while wider and higher areas appear in the flanking regions of insoluble domains.

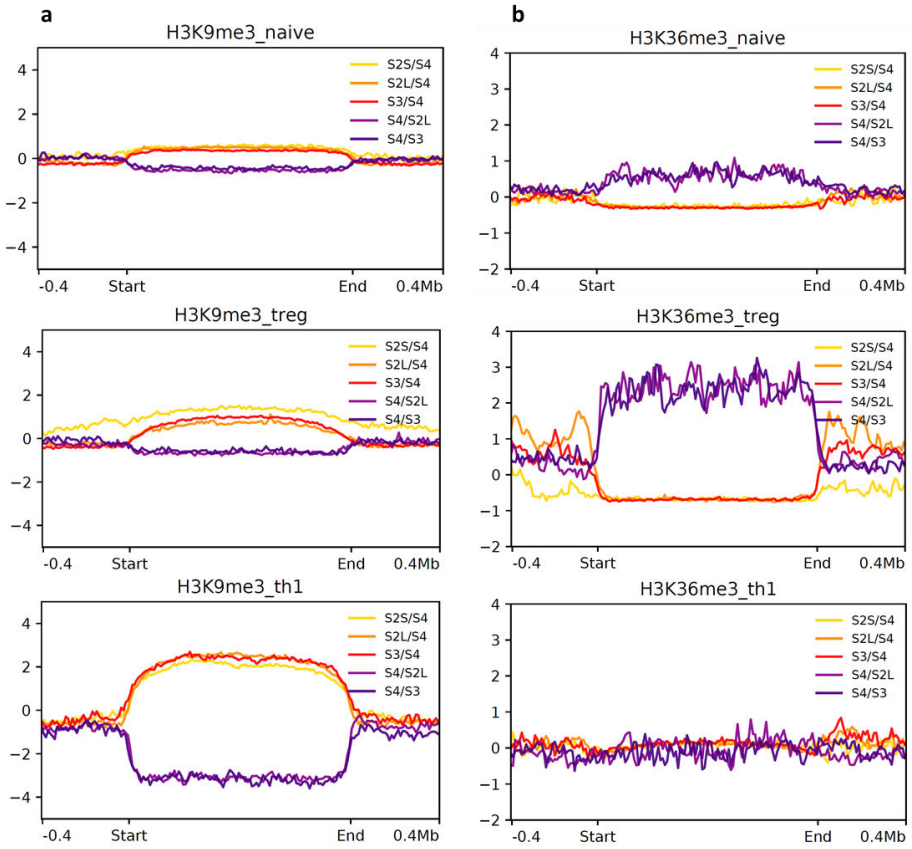


Figure 12 continued on the next page.

Figure 12 continued.

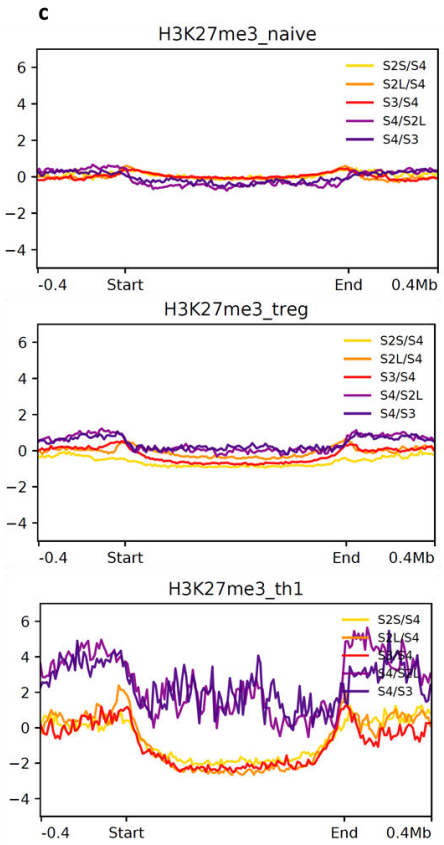
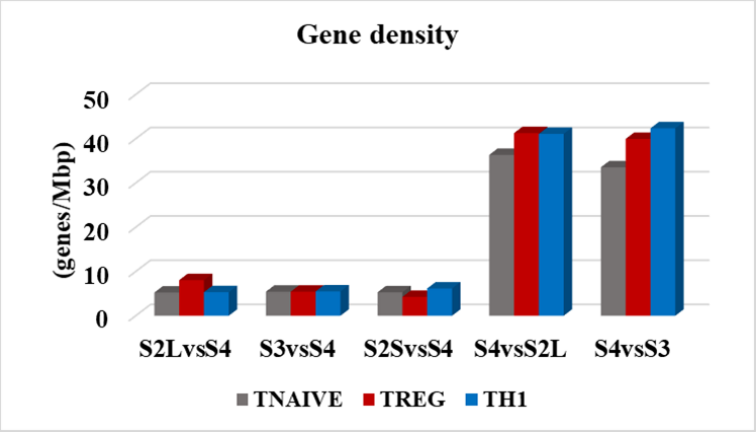


Figure 12 – HMs discriminate soluble and insoluble SAMMY-domains. a, b, c ChIP-seq metaprofile enrichment of H3K9me3 (a), H3K36me3 (b) and H3K27me3 (c) marks in Tnaive (top), Treg (middle) and Th1 (bottom) on soluble (S2SvsS4, S2LvsS4 and S3vsS4) and insoluble (S4vsS2L and S4vsS3) SAMMY-domains. The enrichment signal of the IP over the input is calculated with SPP, averaged along the intersection of domains of at least 3 donors and visualized using deepTools suite.

4.6 Lymphocytes insoluble domains are gene-rich and highly expressed

The highly specific association with function-related histone marks, prompted us to investigate the transcriptional activity of the regions

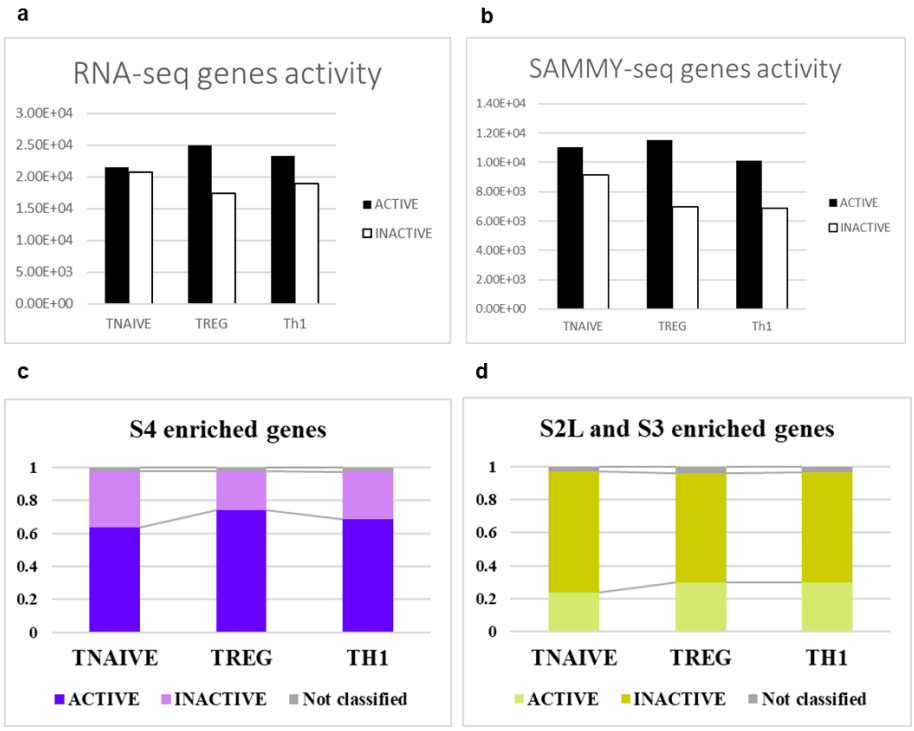
identified in the distinct chromatin compartments. Comparing gene content with genome coverage (Table 1), we noticed that gene density in insoluble domains is 7 to 9 times higher than in soluble domains (Fig.10), suggesting that we are indeed recovering euchromatin and heterochromatin.



Supplementary Figure 10 – Gene density of soluble and insoluble SAMMY-domains suggest genomic function. Average number of genes for each megabase (Mbp) of genome in soluble (S2LvsS4, S3vsS4, S2SvsS4) and insoluble (S4vsS2L, S4vsS3) SAMMY-domains in Tnaive, Treg and Th1 cells.

To directly evaluate the activity of the regions highlighted by SAMMY-domains, we used RNA-seq information from five biological replicates for each population[16] and we divided genes into two broad group of repressed and transcribed following the method described in[18]. On total transcriptome, Th1 and Treg cells revealed to have a slight prevalence of expressed genes (55% and 59%, respectively) while Tnaive cells showed an equal representation of active and inactive genes (51% and 49%) (Suppl. Fig. 11a). Within soluble and insoluble SAMMY-domains we retrieve, for each population, around 50% of the total annotated genes. Their distribution between transcribed and repressed resembles what

observed on total RNA-seq data, with only a slight preference for the recovery of active genes (Suppl. Fig. 11b).



Supplementary Figure 11 – Activity of genes in SAMMY-domains support correlation between HMs and transcriptional activity. a, Number of active and inactive genes in Tnaive, Treg and Th1 cells (data derived from[16] and classified in the two groups following the method described in[18]). **b**, Number of active and inactive genes recovered in SAMMY-domains in Tnaive, Treg and Th1 cells. **c, d** Classification of insoluble (c) and soluble (d) genes into active and inactive in Tnaive, Treg and Th1 cells. A few genes lack classification because they are not annotated in kallisto.

In insoluble domains, we observed a prevalence of active genes (Fig. 11c), while soluble domains contain prevalently silent genes (Fig.11d), supporting the correlation between transcriptional activity and HMs, independently from their solubility.

4.7 Fraction enrichment does not directly reflect nuclear localization of lymphocytes chromatin

The unusual correspondence between solubility and activity of genomic regions initially induced us to think to an inversion of classical euchromatin and heterochromatin intranuclear positioning, a condition already described for rod photoreceptors of nocturnal mammals[19]. To test this hypothesis, we studied two well-characterized chromosomes: the 18, gene-poor and with an evolutionarily conserved peripheral nuclear localization, and the 19, gene-rich and localized toward the nuclear centre[20]. As expected from genome-wide correlation analysis, chromosome 18, characterized by H3K9me3 mark, is prevalently enriched in the soluble fractions (Fig. 13a, Suppl. Fig. 12a and b). Chromosome painting confirmed its positioning in nuclear periphery, suggesting that chromatin solubility does not correlate with intranuclear localization (Fig. 13b). On the other hand, chromosome 19, enriched in S4 and exhibiting a prevalence of H3K36me3 mark (Fig. 13c and Suppl. Fig. 12a and b), showed a mixed pattern, being peripheral in some nuclei and central in others (Fig.13d). The measure of the distance between chromosome centroid and nuclear centroid further stressed a peripheric confinement of chromosome 18 in all the subsets (Fig. 13e). The position of chromosome 19, highly variable, was statistically different from chromosome 18 in Tnaive and Th1, but not in Treg cells, suggesting a less compartmentalized chromatin.

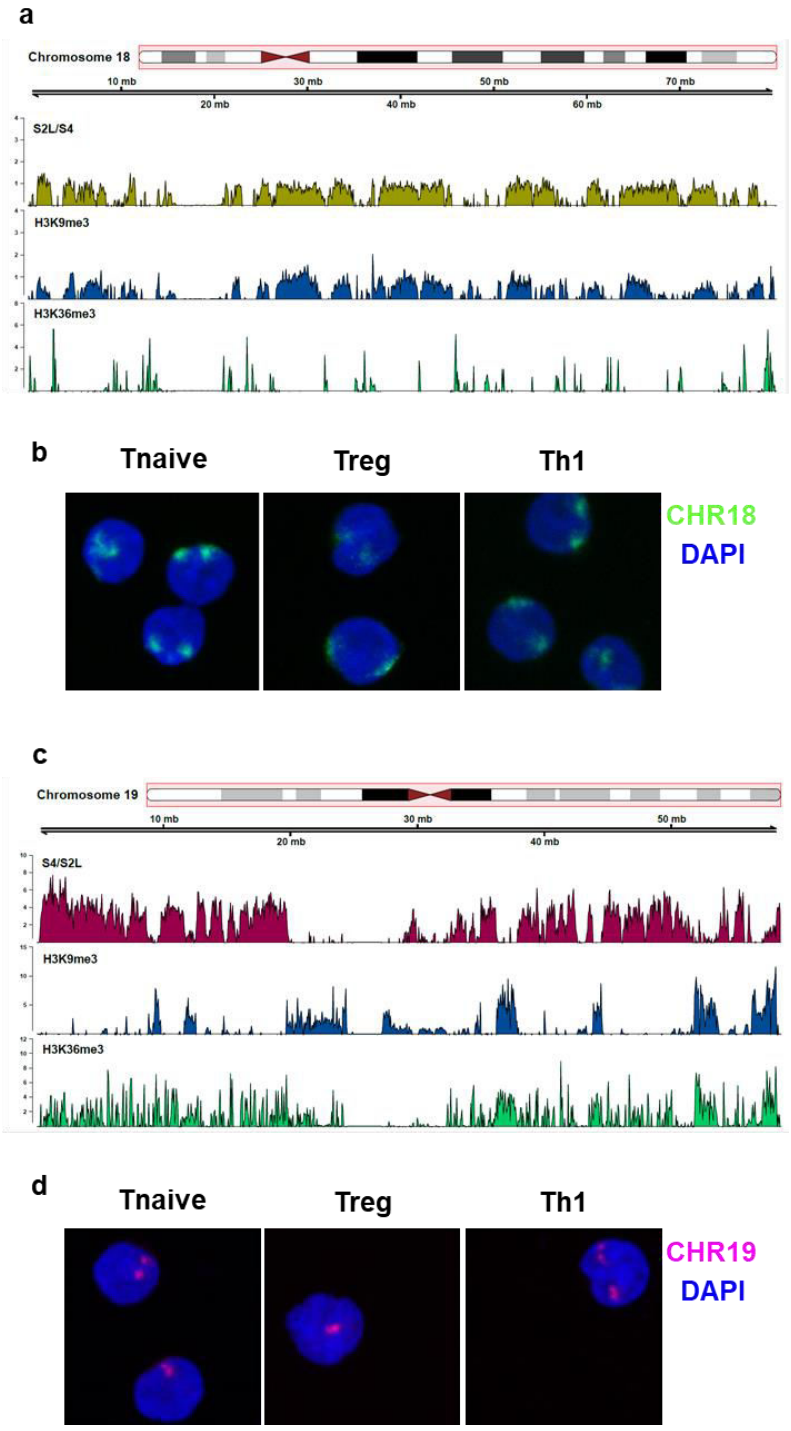


Figure 13 continued on the next page.

Figure 13 continued.

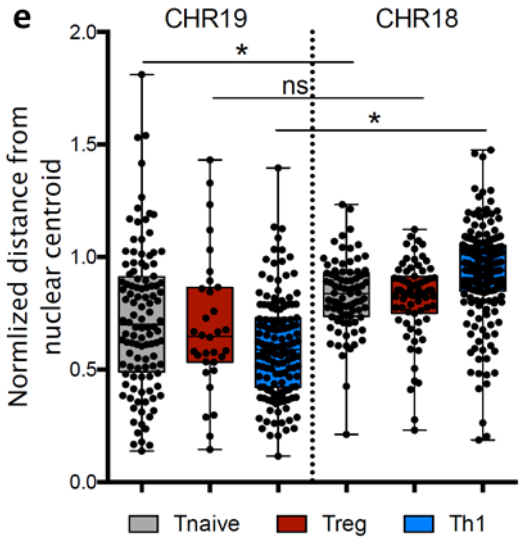
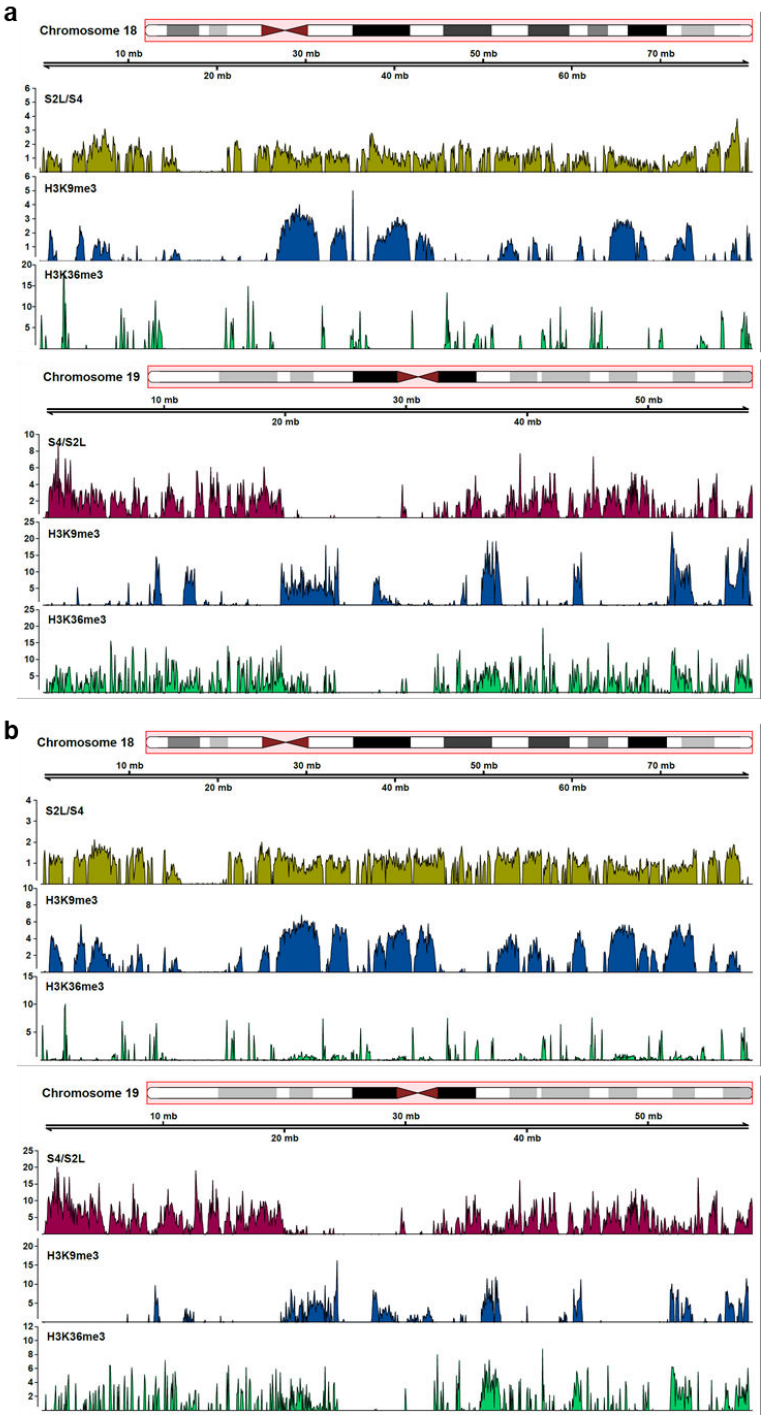


Figure 13 – Chromosome 18 and 19 territories and their HMs are globally conserved in lymphocytes. a, Visualization of SAMMY-seq enrichment signal in the soluble fraction (S2LvsS4) along chromosome 18 in a Tnaive (donor 3) sample. The smoothed differential signal is calculated with SPP and regions of signal enrichment over the reference are marked in yellow-green. ChIP-seq track for H3K9me3 (blue) and H3K36me3 (green) in the same region are also shown. **b**, Representative confocal microscope images of Tnaive, Treg and Th1 nuclei stained for chromosome 18 (green). **c**, Visualization of SAMMY-seq enrichment signal in insoluble fraction (S4vsS2L) along chromosome 18 in a Tnaive (donor 3) sample. The smoothed differential signal is calculated with SPP and regions of signal enrichment over the reference are marked in purple. ChIP-seq track for H3K9me3 (blue) and H3K36me3 (green) in the same region are also shown. **d**, Representative confocal microscope images of Tnaive, Treg and Th1 nuclei stained for chromosome 19 (coral). **e**, Measure of the nuclear distribution of chromosomes 18 and 19. Distance of chromosome centroid from nuclear centroid was normalized on the mean nuclear radio, with 0 representing coincidence with nuclear centroid and 1 representing coincidence with nuclear membrane. Results shown derive from measures on at least 35 chromosome for each condition. Values higher than 1 derive from irregular nuclear shape. Statistics by analysis of variance (ANOVA) with multiple comparisons. * $P < 0.05$



Supplementary Figure 12 – Chromosome 18 and 19 solubility and HMs are conserved among the three CD4$^+$ T lymphocytes subsets

analysed. a, b Visualization of SAMMY-seq enrichment signal in the soluble fraction (S2LvsS4,yellow-green) along chromosome 18 and in the insoluble fraction (S4vsS2L, purple) along chromosome 19 in a Treg sample (donor 2) (a) and in a Th1 sample (donor 3) (b). The smoothed differential signal is calculated with SPP. ChIP-seq tracks for H3K9me3 (blue) and H3K36me3 (green) in the same region are also shown.

To gain an additional point of view, we also tested by immunofluorescent staining the general nuclear distribution of heterochromatin and euchromatin-associated HMs in Tnaive cells. The constitutive heterochromatin H3K9me3-mark and the active promoters H3K4me3-mark displayed the expected pattern, with H3K9me3 forming a ring at the nuclear periphery (Fig. 14a) and H3K4me3 localizing in the nuclear interior (Fig. 14b). On the other hand, the active transcription mark H3K36me3 revealed a peculiar accumulation along the nuclear membrane (Fig. 14c).

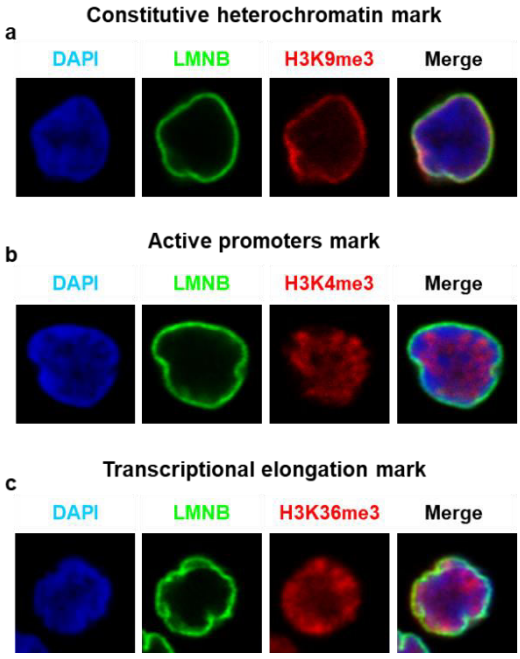


Figure 14 – Histone marks intranuclear pattern reveals atypical euchromatin distribution. a, b, c Representative confocal microscopy

images of Tnaive samples stained for H3K9me3 (a), H3K4me3 (b) or HeK36me3 (c) (red) in combination with Lamin B (green) and DAPI (blue).

All together these data suggest that fraction enrichment does not directly reflect nuclear positioning and that lymphocytes have a cell-specific genome organization, which may be associated with their biological functions.

4.8 $CD4^+$ T lymphocyte subpopulations share most of their chromatin organization

Comparing SAMMY-domains detected in the different T lymphocyte subsets, we noticed that part of them is subset-specific, while a higher portion is shared between all the subsets (Fig. 15a and b). Accordingly, a great portion of the genes contained in soluble (S2LvsS4 or S3vsS4) and insoluble (S4vsS2L or S4vsS3) SAMMY-domains are common to the three subsets analysed (Fig. 16).

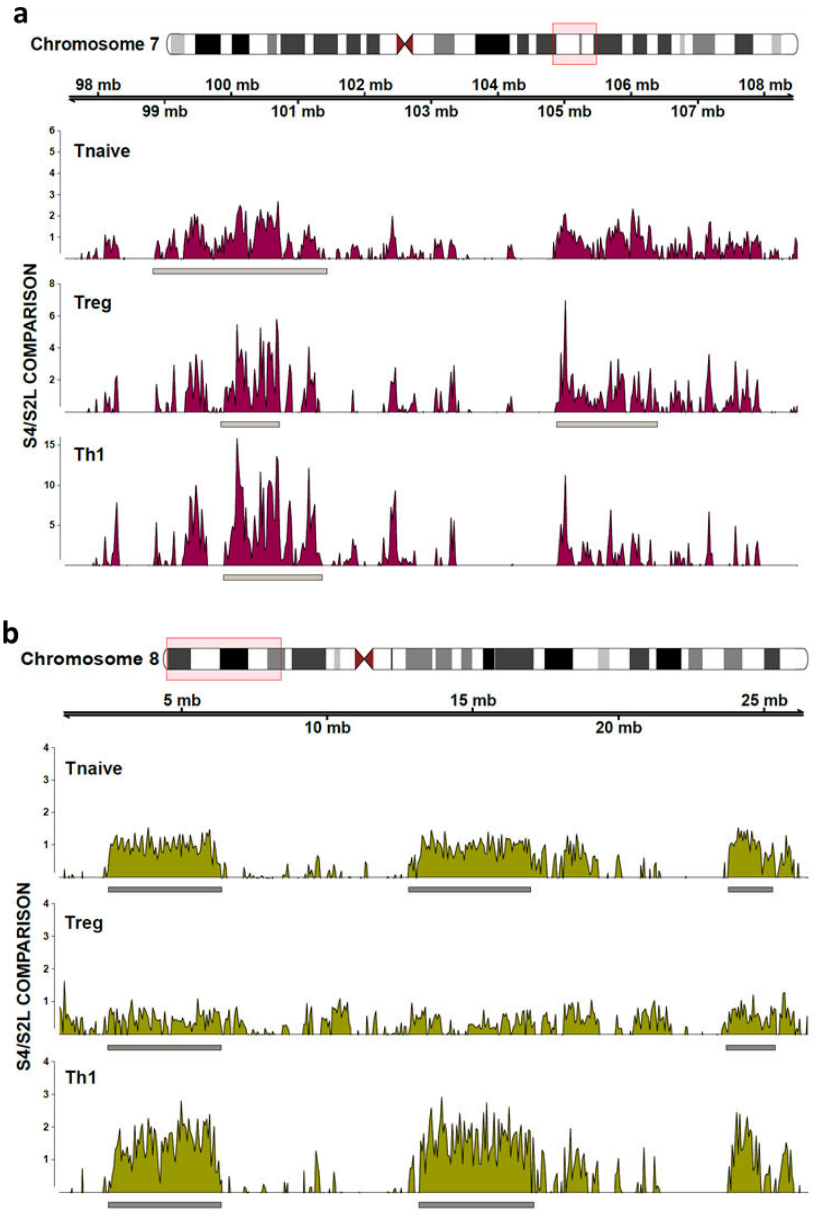


Figure 15 – T CD4$^+$ lymphocytes share the majority of soluble and insoluble domains. a, Visualization of SAMMY-seq enrichment signal in the insoluble fraction (S4vsS2L, purple) in a Tnaive (Donor 2), Treg (donor 2) and Th1 (donor 3) sample along a representative region (11Mb on chr7:97,500,000-108,500,000. The intersection of significantly enriched regions (SAMMY-domains) called by EDD in all donor analysed is reported as grey boxes under the enrichment signal track. **b**, Visualization

of SAMMY-seq enrichment signal in the soluble fraction (S2LvsS4, yellow-green) in a Tnaive (Donor 3), Treg (donor 3) and Th1 (donor 2) sample along a representative region (26Mb on chr8:800,000-26,500,000) The intersection of significantly enriched regions (SAMMY-domains) called by EDD in all donor analysed is reported as grey boxes under the enrichment signal track.

In particular, 57% of the insoluble genes are shared between the three subsets (Fig. 16a), with the majority of them being also commonly active (data not shown). Similarly, 42% of the soluble genes are shared between the subsets (Fig. 16b), with the majority of them being also commonly inactive (data not shown). The similarity in chromatin conformation and gene expression of Tnaive, Th1 and Treg is expected, due to their common origin and their resting state. However, these data also suggest that soluble chromatin could play a role in T cell identity determination. In fact, soluble fractions show a higher number of genes not commonly compartmentalized among the three populations (Fig. 16b).

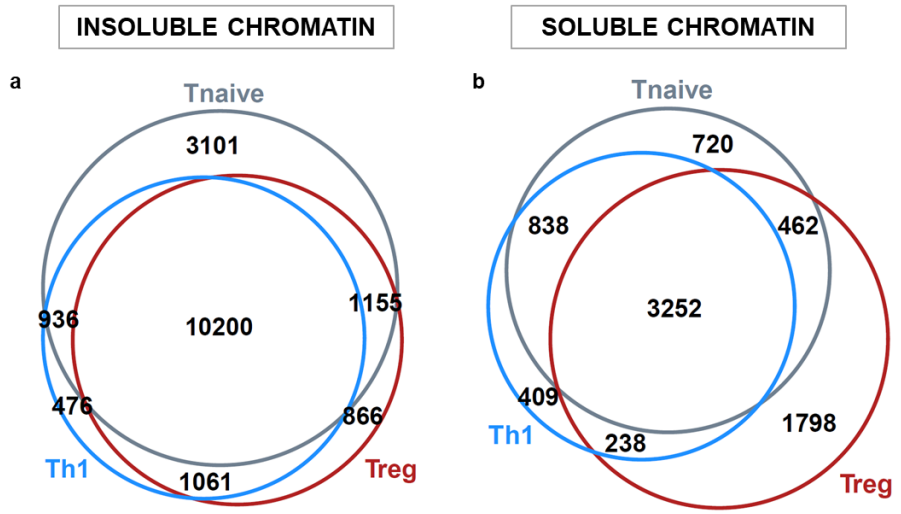


Figure 16 – T CD4⁺ lymphocytes share the majority of soluble and insoluble genes. a, b Intersection of genes in insoluble (S4vsS3, S4vsS2L) (a) or soluble (S2LvsS4, S3vsS4) (b) domains in Tnaive, Treg and Th1. In each subset, the composition of the groups of soluble and

insoluble domains was calculated as the union of genes annotated in the two soluble or the two insoluble SAMMY-domains.

4.9 Lymphocytes chromatin compartmentalization discriminates gene activation and function

To address the function of genes in the soluble and insoluble SAMMY-compartments, we performed gene ontology (GO) analysis.

In the active genes of the insoluble compartment, representing the vast majority of the total active genes, we found, for all the subsets, an overrepresentation of pathways involved in basic cellular metabolic processes (Fig. 17, down-left), confirming that lymphocytes functional euchromatin[21] preferentially resides in the insoluble fraction. On the other hand, active genes in the soluble fraction are primarily involved in cell-cell adhesion and cell morphogenesis (Fig. 17, up-left), with the exception of Tnaive cells where we did not found any GO enrichment. This suggests that chromatin solubility per se plays a role in clustering active pathways. The inactive genes, characterized by a lower percentage of ID uniquely mapping in GO (Fig. 17, up-right and down-right), are equally divided between soluble and insoluble fraction. As expected, given the transcriptional repressed state of these genes, we found in both soluble and insoluble fraction an over-representation of several pathways unrelated to lymphocytes functions (Fig. 17). In Th1, however, we also noticed an over-representation of pathways related to adaptive immune response in the insoluble fraction (Fig. 17a, down-right), and an enrichment of pathways involved in chemokines reception (GPCR pathways) in the soluble compartment (Fig. 17a, up-right). This suggests a possible different availability for transcription of specific immune genes

before and after TCR stimulation. Further analysis of GO clustering or Gene Set Enrichment Analysis (GSEA), will elucidate the role of chromatin compartmentalization in different cell subsets.

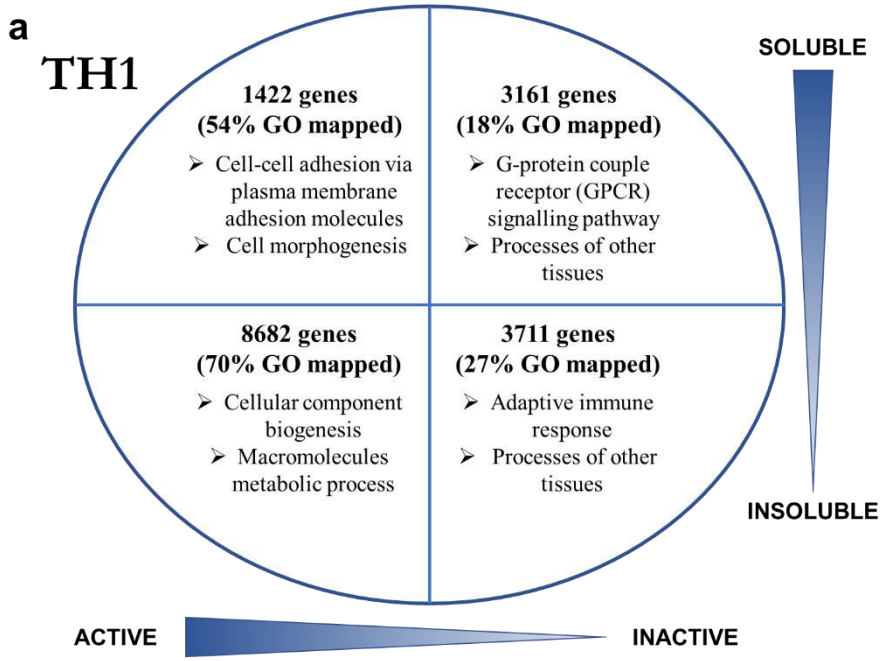
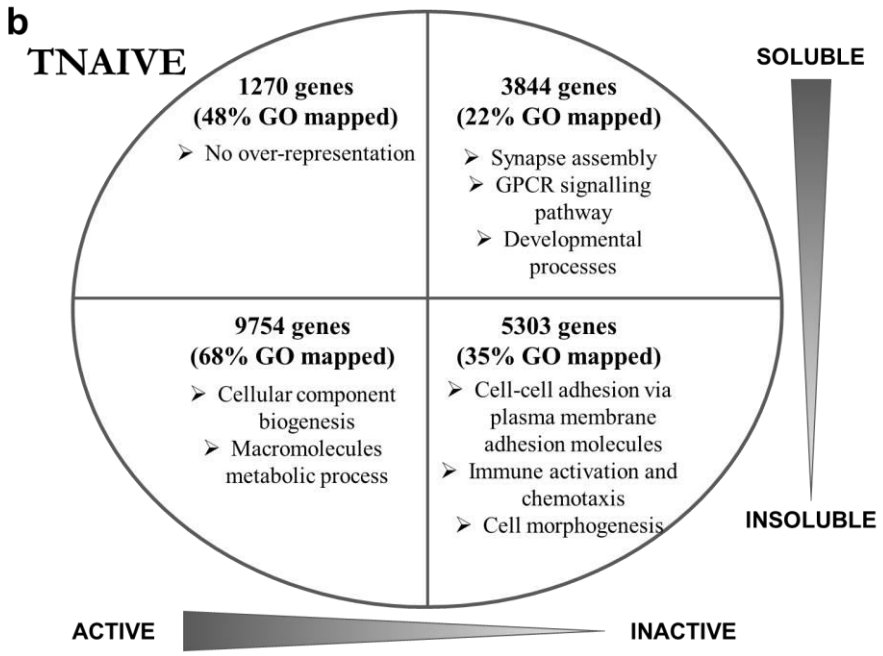


Figure 17 continued on the next page.

Figure 17 continued.



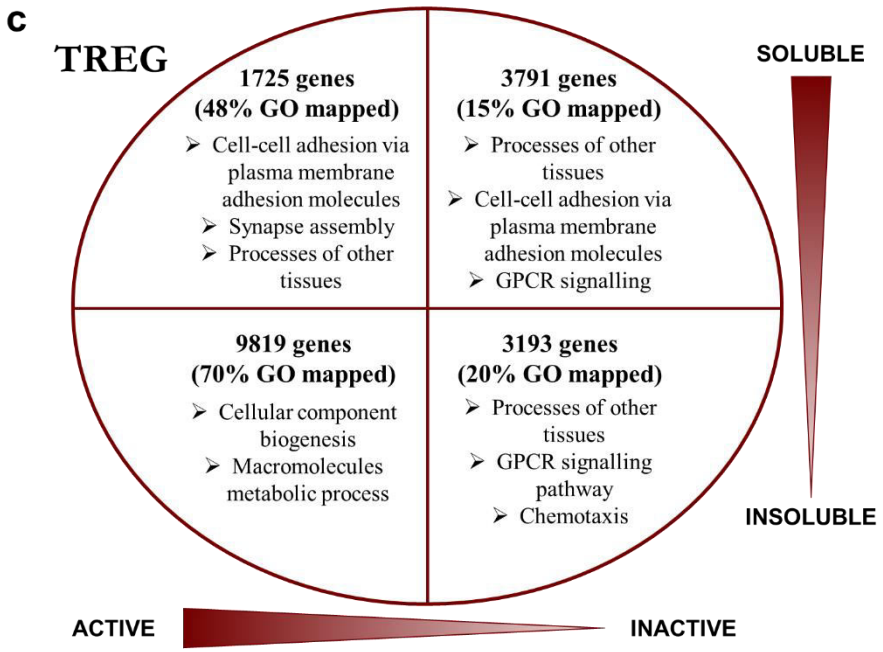


Fig. 17 – Functional characterization of gene activity and compartmentalization. a, b, c Exemplifying categories of GO terms

over-represented in the active and the inactive genes, based on their chromatin environment in Th1 (a), Tnaive (b) and Treg (c).

Materials and methods

Cell cultures

Primary fibroblast cell lines were cultured in High Glucose Dulbecco's Modified Eagle's Medium (DMEM) supplemented with glutamax, 15% fetal bovine serum (FBS) and 1% Penicillin-Streptomycin. AG07095 (CTRL002) human dermal fibroblasts were obtained from the Coriell Institute. Preputial fibroblast strain #2294 (CTRL004) was a generous gift from the Laboratory of Molecular and Cell Biology, Istituto Dermopatico dell'Immacolata (IDI)-IRCCS (Rome, Italy), while control dermal fibroblast CTRL013 was kindly provided by the Italian Laminopathies Network.

C2C12 murine myoblasts were cultured in High Glucose DMEM supplemented with 10% FBS, 1% Pen-Strept and 1% L-glutamine.

For activation experiments, human primary $CD4^+$ T lymphocytes were cultured in Roswell Park Memorial Institute (RPMI) 1640 supplemented with 10% FBS, 1% Pen-Strept, 1% L-glutamine, 1% non-essential aminoacids (NEAA), 1% sodium-pyruvate and 200 U/mL IL-2 (Miltenyi Biotec, 130-097-748) at 37°C, 5% CO_2. They were incubated with a 1:1 ratio of CD3/CD28 beads (Gibco, 11131D) for the indicated time.

Isolation of human primary T lymphocytes

Buffy coats of healthy donors were obtained from Fondazione I.R.C.C.S. Ca' Granda Ospedale Maggiore Policlinico (Milan, Italy). The use of peripheral blood mononuclear cells (PBMCs) from healthy donors for

research purposes was approved by the ethical committee of Fondazione I.R.C.C.S. Ca' Granda Ospedale Maggiore Policlinico and informed consent was obtained from subjects.

PBMCs were isolated from buffy coat by Ficoll-hypaque density gradient centrifugation. $CD4^+$ T lymphocytes were enriched by negative selection using Human $CD4^+$ T Cell Isolation Kit (Miltenyi Biotec, 130-096-533). Primary lymphocyte subsets were then purified >95% by fluorescence-activated cell sorting (FACS) using different combinations of surface markers. $CD4^+$ T Naïve (Tnaive): $CD4^+CD45RO^-$, $CD4^+$ T Helper 1 (Th1): $CD4^+CD45RO^+CXCR3^+$, $CD4^+$ T regulatory (Treg): $CD4^+CD25^+CD127^-$.

Sorting antibodies: CD4 VioGreen™ (Miltenyi Biotec, 130-096-900 or 130-113-230), CD127 PE (Miltenyi Biotec, 130-094-889 or 130-113-414), CD25 VioBright™FITC (Miltenyi Biotec, 130-104-323 or 130-113-283), CD45RO APC (Miltenyi Biotec, 130-095-460 or 130-113-556), CXCR3 Pe-Cy™5 (BD Pharmigen™, 561731)

Immunofluorescence

In order to preserve the integrity of chromatin architecture, enriched $CD4^+$ T lymphocytes were fixed in Roswell Park Memorial Institute (RPMI) 1640 supplemented with 10% FBS using 1% formaldehyde (Sigma, F8775) solution (5mM HepesKOH pH7.5, 10mM NaCl, 0.1mM EDTA, 0.05mM EGTA, 1.1% formaldehyde) for 9 minutes at RT. Formaldehyde was then quenched with 125 mM Glycine (Sigma, 8898) for 5 minutes at RT and removed with two washings in cold 20%FBS/PBS. After FACS isolation, T lymphocytes were placed at density of 100.000/mL on coverslip previously coated with poly-L-lysine solution (Sigma, P8920). After 30 minutes of sedimentation at RT,

coverslips were fixed with 4% PFA(Sigma, P6148)/PBS for 7 minutes at RT. Cells were then washed in 0,05% Triton X-100 (Sigma, T8787)/PBS twice rapidly and three times for 5 minutes in agitation. They were permeabilized with 0,5% Triton X-100/PBS for 10 minutes, rinsed in PBS and incubated overnight at RT in 20% Glycerol (Sigma, G5516)/PBS. On the next day, coverslip were freezed on dry ice (30 seconds), thawed gradually at RT and immerse in 20% Glycerol/PBS for four times. They were washed once in 0,5% Triton X-100/PBS for 5 minutes and twice in 0,05% Triton X-100/PBS for 5 minutes. Coverslips were then incubated in 0.1M HCl (Sigma, H1758)/H_2O for 15 minutes and rinsed in PBS. Blocking of aspecific binding sites was obtained through 30 minutes incubation at RT in 4% BSA (Sigma, A7030)/PBS. Primary antibodies against modified histones were diluted in blocking solution and incubated on coverslips overnight at 4°C in a humidified chamber. On the next day, coverslips were washed three times in 0.05%Triton/DPBS for 5 minutes at room temperature. Then they were incubated in a humidified chamber at RT for 2 hours with primary antibody against lamins diluted in blocking solution. After washings in 0.05% Triton X-100/PBS, they were incubated in the dark in a humidified chamber at RT for 1 hour with secondary antibodies diluted in 1%BSA/PBS. After washings in 0.05% Triton X-100/PBS, coverslips were incubated with 0.1% DAPI/PBS for 3 minutes at RT in the dark and then washed again in PBS. Finally, they were mounted on microscope slides using ProLong™ antifade mountant (Invitrogen, P36930) and images were acquired through a Leica TCS SP5 Confocal Laser Scanning Microscope.

Primary antibodies: anti-H3K9me3 (Abcam, ab8898) 1:500 (2 ng/ul), anti-H3K4me3 (Millipore, 07-473) 1:150 (6.7ng/uL), anti-H3K36me3 (Abcam, ab9050) 1:150 (6.7 ng/uL), anti-Lamin B (Santacruz, sc-6216)

1:100 (2 ng/uL), anti-Lamin A/C (Santacruz, sc-6215) 1:100 (2 ng/uL), rabbit IgG isotype control (Thermo Fisher Scientific, 02-6102) (6.7 ng/uL), goat IgG isotype control (Invitrogen, 02-6202) (2ng/ul)
Secondary antibodies: Alexa Fluor™ 647-conjugated Donkey Anti-Rabbit IgG (Invitrogen, A31573) 1:1000, Alexa Fluor™ 488-conjugated Donkey Anti-Goat IgG (Jackson ImmunoResearch, 705-545-003) 1:500

Chromosome painting

Cells were treated according to the immunofluorescence protocol until the four freeze and thaw cycles and the following washes. Then coverslips were incubated with 2.5μL RNase Cocktail™ (Invitrogen, AM2286) in 150 μL of PBS for 20 minutes at 37°C. Later they were rinsed in PBS, incubated in 0.1M HCl for 15 minutes at RT, washed twice in PBS for 5 minute in agitation at RT, rinsed in 2X saline sodium citrate (SSC) and incubated overnight at RT in 50% formamide (Sigma, 47670) pH 7.0/2X SSC. On the next day, fluorescent probe for the desired chromosome was denaturated in a thermomixer at 70°C, 300 rpm for 5 minutes (2.5uL of probe for each 12mm round coverslip were used, diluted to 5ul in a solution of 20% dextran sulfate (DXS)/4X SSC). Coverslips were accurately dry, put on a drop of probe and sealed. Cells and probe were denaturated in the dark on a heated plate at 75°C for 4 minutes, then transferred to a humidified chamber and incubate at 37°C overnight. On the next day coverslips were washed three times in 2X SSC at 37°C, 90 rpm for 5 minutes in agitation, then three times in 0.1X SSC at 60°C, 90 rpm for 5 minutes in agitation, twice in 0.2% Tween/4X SSC at 37°C, 90 rpm for 5 minutes in agitation and finally once in PBS at RT for 5 minutes in agitation. They were incubated in 0.2% DAPI/PBS for 5 minutes at RT, washed twice in PBS for 5 minutes in agitation at RT and mounted on a

microscope slide using ProLong™ antifade mountant. Images were acquired through a Leica TCS SP5 Confocal Laser Scanning Microscope. Image analysis was performed using plug-in GA3 (for segmentation pipeline) and plug-in 3D Richardson-Lucy (deconvolution algorithm, 20 interactions) on a Nikon Nis Elements (version 5.11.01). Distances were normalized on the radius derived from equalized diameter.

Probes: XCP18 green (MetaSystem, D-0318-050-FI), XCP19 orange (MetaSystem, D-0319-050-OR)

Chromatin fractionation

Chromatin fractionation on CTRL004, C2C12 murine myoblasts and T lymphocytes was performed with minor adaptations to the protocol described in (Sebestyen et al.). 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 $MgCl_2$; 1X Protease Inhibitor Cocktail (Roche, 04693116001); 1 mM PMSF (Sigma, 93482)) supplemented with 1 mM DTT and 0,5% Triton X-100. After 10 minutes on a wheel at 4°C, the cytoskeletal structures were separated from soluble proteins by 3 minutes of 900g centrifugation at 4°C. Supernatants were conserved as S1 fraction and pellets were washed for 10 minutes on the wheel at 4°C with an additional volume of the same buffer. Pellets were then suspended in cold CSK buffer and incubated for 1 hour at 37°C with 25U of RNase–free DNase I (Invitrogen, AM2222). To stop digestion, ammonium sulphate was added to the samples to a final concentration of 250 mM and incubated for 5 minutes on ice. Samples were pelleted by 3 minutes of 900g centrifugation at 4°C. Supernatants were conserved as S2 fraction, while pellets were washed for 10 minutes on the wheel at 4°C with an additional volume of cold CSK buffer and collected by 3 minutes

3000g centrifugation at 4°C. Pellets were then solubilized in CSK buffer with increased (2M) NaCl concentration, incubated 10 minutes on the wheel at 4°C and centrifuged for 3 minutes at 2300g at 4°C. Supernatants were conserved as S3 fraction, while pellets were washed twice for 10 minutes on the wheel at 4°C with an additional volume of high salt CSK buffer. Finally, after 3 minutes of 3000g centrifugation at 4°C, pellets were solubilized in highly denaturating 8M urea for 10 minutes and conserved as S4 fraction. Fractions were stored at – 80°C until DNA extraction or protein analysis.

DNA extraction, library preparation and sequencing

For DNA content analysis of each fraction 33% of S1 volume, 50% of S2 volume, total S3 and S4 volumes were diluted in TE buffer (10mM TrisHCl pH 8.0, 1 mM EDTA) and incubated with 40µg of RNAse A (Roche, 10109142001) at 37°C for 90 minutes, followed by 40µg of Proteinase K (Invitrogen, AM2548) at 55°C for 150 minutes. DNA was then isolated through double phenol/chloroform (Sigma, 77617) extraction and precipitated in 3 volumes of cold ethanol, 0.3M sodium acetate and 20ug glycogen (Ambion AM9510) for 1 hour on dry ice or overnight at -20°C. Dry pellets were suspended in 50 µl (S2) or 15 ul (S1, S3 and S4) of nuclease-free water and incubated at 4°C overnight. On the next day, S2 was additionally purified using PCR DNA Purification Kit (Qiagen, 28106) and suspend in 50 µL of nuclease-free water. DNA fragments in S2 were separated using AMPure XP paramagnetic beads (Beckman Coulter, A63880) with the ratio 0,90/0,95. Smaller fragments were conserved as S2S fraction and larger fragments as S2L fraction. Both were suspended in 20 ul of nuclease-free water and then brought to 15ul using a centrifugal vacuum concentrator. S1, S2L, S3 and S4 fractions were

sonicated in a Covaris M220 focused-ultrasonicator using screw cap microTUBEs (Covaris, 004078) (water bath 20°C, peak power 30.0, duty factor 20.0, cycles/burst 50, duration: 125 seconds for S1, S2L and S3, 150 seconds for S4). Fractions were quantified using Qubit 2.0 fluorometer with Qubit dsDNA HS or BR Assay Kits (Invitrogen, Q32854 or Q32850) and run on Agilent 2100 Bioanalyzer using High Sensitivity DNA Kit (Agilent, 5067-4626). DNA fragments should form a smear, peaking between 150-200 bp. Libraries were created from each sample using the automation instrument Biomek FX (Beckman Coulter), then qualitatively and quantitatively checked on Bioanalyser 2100. Libraries with distinct adapter indexes were multiplexed and, after cluster generation on FlowCell, were sequenced for 50 bases at the IEO Genomic Unit in Milan. Samples from donor 1, 2 and 3 were sequenced in the single read mode on a HiSeq 2000 sequencing system from Illumina, samples from donor 4 were sequenced in the paired-end mode on a NovaSeq 6000 sequencing system from Illumina. A sequencing depth of at least 30 million reads was obtained for each sample.

Western blot

Supernatants from each chromatin extraction step were quantified using Qubit 2.0 fluorometer with Qubit Protein Assay Kit (Invitrogen, Q33212), run on 4-12% Bis-tris Plus acrylamide gels (Invitrogen, NW04125), immunoblotted and revealed with SuperSignal West Dura (Thermo Scientific, 34076).

Primary antibodies: anti-Bmi1 (Millipore, 05-637) 1:1000, anti-alpha-Tubulin (Sigma, T5168)1:1000, anti-H3 (Abcam, ab1791) 1:5000, anti-H3K36me3 (Abcam, ab9050) 1:1000, anti-H3K9me3 (Abcam, ab8898)

1:1000, anti-H3K27me3 (Millipore, 07-449) 1:1000, anti-Lamin B (Santacruz, sc-6216) 1:5000, anti-Lamin A/C (Santacruz, sc-6215) 1:2000 Secondary antibodies: Anti-Mouse IgG-Peroxidase (Sigma, A9044), Anti-Rabbit IgG-Peroxidase (Sigma, A9169), Anti-Goat IgG-Peroxidase (Sigma, A5420)

ChIP-seq

$CD4^+$ Th1 lymphocytes isolated from buffy coats of healthy donors were fixed in RPMI 1640 supplemented with 10% FBS using 1% formaldehyde solution for 9 minutes at RT. Formaldehyde was then quenched with addition of 125 mM Glycine for 5 minutes at RT and removed with two washings in cold 20%FBS/PBS. Pellets of 2 million cells were flash frozen in a dry ice-ethanol bath and stored at -80°C until sonication, when they were suspended in 130 ul of sonication buffer (10 mM TrisHCl pH 8.0, 2 mM EDTA, 0.25% SDS) and sonicated in a Covaris M220 focused-ultrasonicator using snap cap microTUBEs (Covaris, 520045) (water bath 7°C, peak power 75.0, duty factor 10.0, cycles/burst 250, duration 1020 seconds). Fragmentation of chromatin to about 150 bps was checked on Agilent 2100 Bioanalyzer. Equilibration buffer (10mM Tris pH8, 233mM NaCl, 1.66% Triton X-100, 0.166% DOC, 1mM EDTA, 1X PMSF, 1X protease inhibitors) was added to every sample to bring solution to IP buffer composition (10 mM TrisHCl pH 8.0, 140 mM NaCl, 1 mM EDTA, 0.1% SDS, 0.1% DOC, 1% Triton X-100 1X PMSF, 1X protease inhibitors). Samples were then spun at 14000g for 10 minutes at 4°C to pellet insoluble material. Supernatants were quantified using Nanodrop 1000 spectrophotometer and samples were bring to desired volume for IP (300 ul) using IP buffer. 1% of total chromatin was preserved at 4°C as input sample. Remaining chromatin was incubated with 2ug of the

antibody of interest overnight on a rotating wheel at 4°C. Protein G beads (Life Technology, 1004D) were washed twice in 0.1% BSA/IP buffer, resuspended in the same buffer and incubated on the wheel overnight at 4°C. On the next day, 15ul of protein G beads were added to each sample and incubated on the rotating wheel at 4°C for 2 hours. Beads were then washed twice with IP buffer, twice with high-salt IP buffer (500 mM NaCl instead of 140 mM), twice with RIPA-LiCL buffer (10 mM TrisHCl pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% DOC, 0.5% NP-40, 1X PMSF, 1X protease inhibitors) and once with 10 mM TrisHCl pH 8.0. Crosslinking was reversed by incubating the beads at 65°C overnight in Elution buffer (10 mM TrisHCl pH 8.0, 5 mM EDTA, 300mM, 0.4% SDS). Input chromatin was also decrosslinked in elution buffer overnight at 65°C. On the next day, all samples were diluted with one volume of TE 10:1, treated with 5 ul of RNase cocktail (Ambion, AM2286) at 37°C for 2 hours at 37°C and then with 5 ul of Proteinase K (Thermo Scientific, EO0491) at 55°C for 2 hours. Finally magnetic beads were removed, DNA was isolated through phenol-chloroform extraction and precipitated. Pellets were suspended in 31ul of nuclease-free water and quantified using Qubit 2.0 fluorometer with Qubit dsDNA HS Assay Kits (Invitrogen, Q32854). Before library preparation, IPs were validated using SYBR select master mix (Invitrogen, 4472908) on a StepOnePlus Real-Time PCR System (Thermo Fisher Scientific).

Primers: chr11-FW: GCTGTGAATAAGCAGTGACC, chr11-REV: CAGCAATGGACAGATGTGAAG (region amplified, chr11:61240408-61240540), CD4-FW: TGATCTCAGCCTCTCGTTCC, CD4-REV: CTTCGAGACCTTTGCCTCC (chr12:6817103-6817233).

Libraries from IPs and input were prepared using NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB, E7645L). They were

quantified using Qubit 2.0 fluorometer with Qubit dsDNA HS Assay Kit (Invitrogen, Q32854) and run on Agilent 2100 Bioanalyzer using High Sensitivity DNA Kit (Agilent, 5067-4626). Libraries with distinct adapter indexes were multiplexed and, after cluster generation on FlowCell, were sequenced for 50 bases in the paired-end mode on a NovaSeq 6000 sequencing system from Illumina at the IEO Genomic Unit in Milan. A sequencing depth of at least 30 million reads was obtained for each sample.

Antibodies: H3K9me3 (Abcam, ab8898), H3K36me3 (Abcam, ab9050), H3K27me3 (Millipore, 07-449).

SAMMY-seq sequencing read analysis

Sequencing reads were quality controlled with fastqc (version 0.11.3; http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and trimmed using Trimmomatic in single end mode (version 0.32; illuminaclip: 2:30:10, leading: 3, trailing: 3, slidingwindow: 4:15, minlen: 35). Remaining reads were aligned to the Ensembl hg38 reference human genome (release 87 for lymphocytes, release 92 for C2C12) using bwa aln (version 0.7.12) with options -n 2 -k 2 and saved the results in sam format with bwa samse. The sam files were converted to bam and name sorted with samtools (version 1.3.1). PCR duplicates were marked using the biobambam2 toolset (version 2.0.54). PCR duplicates, qcfail, multimapping and unmapped reads were discarded with samtools. The bam files were converted to bedgraph using bedtools (version 2.24.0) with scaling factor to normalize for library size; then bedgraph were converted to bigWig using UCSC's bedgraphToBigWig tool for reads distribution visualization. For paired-end samples (all cell types from the fourth donor) only reads from the first mate were used for the analyses. For the Tnaive S4 fraction of the second donor and for the S2L fraction of the third

donor a second sequencing run was made to reach a minimum of 20 million of uniquely mapped reads.

ChIP-seq sequencing read analysis

First mate sequencing reads of Th1 ChIP-seq samples were quality checked, trimmed, aligned, filtered and converted using the same tools as for the SAMMY-seq reads. H3K9me3 domains were called using Enriched Domain Detector (EDD) with --write-log-ratios and --write-bin-scores options and also excluding blacklisted genomic regions containing telomeric, centromeric, and certain heterochromatic regions. The --bin-size and --gap-penalty options were instead automatically estimated by the tool. The required_fraction_of_informative_bins parameter was changed to 0.98. The SPP package (version 1.14) was used to calculate the genome wide ChIP-seq signal with the input as control. The bam files were imported into the R (version 3.0.1) statistical environment, and informative reads were selected with the get.binding.characteristics and select.informative.tags functions, and anomalous positions with extremely high number of reads were removed using the remove.local.tag.anomalies function; then the differential signal, smoothed by a Gaussian kernel, is calculated using the get.smoothed.tag.density function with the default bandwidth parameter and tag.shift = 0.

The quality of ChIP-seq was confirmed by plotting their distribution along the hg38 gene annotation using deepTools suite (computeMatrix and plotProfile functions).

Literature data processing

ChIP-seq histone mark datasets for Tnaive and Treg lymphocytes were downloaded from Roadmap Epigenomics (H3K9me3, H3K36me3, H3K27me3 and H3K4me3). Sequencing reads were quality checked, trimmed, aligned, filtered and converted using the same tools as for the SAMMY-seq reads. The genome wide ChIP-seq signal for each HM was calculated using the SPP package as for Th1 ChIP-seq samples.

ChIP-seq histone mark datasets for fibroblasts were downloaded from the same source. Redacted alignments in bam format were filtered and converted using the same tools as for the SAMMY-seq reads. The genome wide ChIP-seq signal for each HM was calculated using the SPP package as above.

The C2C12 H3K9me3 ChIP-seq sample was taken from[22] and the C2C12 Lamin B DamID sample was taken from[23]. These samples were analyzed as lymphocytes ChIP-seq data starting from sequencing reads.

The C2C12 H3K36me3 and H3K4me3 ChIP-seq IP over INPUT fold enrichment (bigwig) files were downloaded from GEO using the GEO accession IDs GSM918417 and GSM918415. The coordinates that originally refer to the genome assembly mm9 have been converted to the mm10 assembly using the liftOver tool (https://genome-store.ucsc.edu) using as "map.chain" the file mm9ToMm10.over.chain and as "unMapped" the list of unalignable regions for mm10 downloaded from UCSC.

RNA-seq sequencing read analysis

Poly-A RNA-seq of Tnaive, Treg and Th1 lymphocytes was retrieved from ArrayExpress (https://www.ebi.ac.uk/arrayexpress/; E-MTAB-2319; [16]). Sequencing reads from 5 replicates for each cell type were quality

controlled with fastqc (version 0.11.2) and trimmed using Trimmomatic in paired-end mode (version 0.33; illuminaclip: 2:30:10, leading: 3, trailing: 3, slidingwindow: 4:15, minlen: 50). Transcript level quantification was done with Kallisto (version 0.44.0) to estimate per transcipt TPM (Transcripts Per Million) values. Transcript level TPM values were summarized to gene level TPM values using a custom R script. The Ensembl fasta (release 87) for cDNA e ncRNA from Ensembl ftp was merged and used to build the transcriptome index. Kallisto was run the --gtf option to provide the Ensembl annotation (release 87).

Starting from gene level TPM values, genes were classified in expressed and inactive categories using the zFPKM R package (version 1.4.1). Gene were defined as expressed if the gene level TPM zeta-score is higher than 3 in at least 3 over 5 replicates for each cell type. After that expressed genes were further classified in quartiles using quantile R function.

SAMMY-domain calling, signal calculation, overlap analysis and visualization

Significantly enriched regions were selected using EDD by means of relative comparisons of SAMMY-seq fractions within each sample. Less accessible fractions were compared to more accessible ones (S4 vs S2L and S4 vs S3 comparisons) and the conversely (S2L vs S4, S3 vs S4 and S2S vs S4), with --write-log-ratios and --write-bin-scores options and also excluding blacklisted genomic regions containing telomeric, centromeric, and certain heterochromatic regions63. The --bin-size and --gap-penalty options were instead automatically estimated by the tool. The required_fraction_of_informative_bins parameter was changed to 0.98.

The genome wide differential signal for all comparisons was calculated using the SPP package as for ChIP-seq data. The bedtools (version 2.24.0;

jaccard command) was used to calculate the Jaccard Index for the pairwise overlap between replicates.

The Gviz Bioconductor package (version 1.26.5) was used to visualize SAMMY-seq read coverages, differential signals and domains and ChIP-seq enrichment signals. The ggplot2 package (version 3.2.1) was used to plot Jaccard index summaries.

Data correlation and visualization

Genome-wide Spearman correlations between lymphocytes SAMMY-seq samples and all of the public or in-house ChIP-seq marks was conducted using deepTools suite (version 3.0.1). When using single fractions the multiBamSummary function was used to make the 100kb genome binning then the plotCorrelation function was used to calculate the pairwise correlation values which were plot with ggplot2 R package. Whereas for SAMMY-seq comparison the multiBigWigSummary function was used to make the 100kb genome binning then the plotCorrelation function was used to calculate and visualized the pairwise correlation values with the relative clustering. Genome-wide Spearman correlations between C2C12 SAMMY-seq samples and all of the public ChIP-seq marks was conducted using deepTools suite as for lymphocytes.

The correlation of SAMMY-seq replicates on the intersection of called domains was tested by PCA using deepTools suite. The multiBigWigSummary function was used to make the 100kb genome binning then the plotPCA function was used to calculate the principal components which were plot with ggplot2 R package.

The correlation between HMs signal and SAMMY-seq comparison was also confirmed by plotting their distribution using deepTools suite. The computeMatrix function was used to make the 10kb genome binning then

the plotProfile function was used to visualize the mean enrichment signal of each HM over the intersection of SAMMY-domains.

Gene annotation and visualization

The gene annotation of SAMMY-domains intersection between replicates was made using ChIPpeakAnno R package (version 3.16.1) retaining overlapping genes. The overlap of genes annotated to different SAMMY-seq comparisons or to different cell types was visualized with area-weighted Venn diagrams by means of the Vennerable R package (version 3.0; https://github.com/js229/Vennerable/).

References to chapter 4

1. He, D., Nickerson, J. A. & Penman, S. Core filaments of the nuclear matrix. *J. Cell Biol.* **110**, 569–580 (1990).
2. Marasca, F., Marullo, F. & Lanzuolo, C. Determination of Polycomb Group of Protein Compartmentalization Through Chromatin Fractionation Procedure. in *Polycomb Group Proteins: Methods and Protocols* 167–180 (Springer Science, 2016).
3. Cesarini, E. *et al.* Lamin A/C sustains PcG protein architecture, maintaining transcriptional repression at target genes. *J. Cell Biol.* **211**, 533–551 (2015).
4. Sebestyen, E. *et al.* Early Polycomb-target deregulations in Hutchinson-Gilford Progeria Syndrome revealed by heterochromatin analysis. *BioRxiv 799668* (2019).
5. Lund, E., Oldenburg, A. R. & Collas, P. Enriched domain detector: A program for detection of wide genomic enrichment domains robust against local variations. *Nucleic Acids Res.* **42**, (2014).
6. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–329 (2015).
7. McCord, R. P. *et al.* lamin A / C interactions in Hutchinson-Gilford progeria syndrome. *Cold spring Harb. Lab. Press* 260–269 (2013).
8. Sadaie, M. *et al.* Redistribution of the Lamin B1 genomic binding profile affects rearrangement of heterochromatic doma.pdf. *Genes Dev.* **27**, 1800–1808 (2013).
9. Smith, D. K., Yang, J., Liu, M. L. & Zhang, C. L. Small Molecules Modulate Chromatin Accessibility to Promote NEUROG2-Mediated Fibroblast-to-Neuron Reprogramming. *Stem Cell Reports* **7**, 955–969 (2016).
10. Chen, D. S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. *Nature* **541**, 321–330 (2017).
11. Chen, L., Qiao, D., Wang, J., Tian, G. & Wang, M. Cancer

immunotherapy with lymphocytes genetically engineered with T cell receptors for solid cancers. *Immunol. Lett.* **216**, 51–62 (2019).

12. Guo, X. *et al.* Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* **24**, (2018).
13. Zhang, L. *et al.* Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**, 268–272 (2018).
14. Zheng, C. *et al.* Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell* **169**, 1342-1356.e16 (2017).
15. De Simone, M. *et al.* Transcriptional Landscape of Human Tissue Lymphocytes Unveils Uniqueness of Tumor-Infiltrating T Regulatory Cells. *Immunity* **45**, 1135–1147 (2016).
16. Ranzani, V. *et al.* LincRNA landscape in human lymphocytes highlights regulation of T cell differentiation by linc-MAF-4. *Nat. Immunol.* **16**, 318–325 (2015).
17. Gonzalez-Granado, J. M. *et al.* Nuclear Envelope Lamin-A Couples Actin Dynamics with Immunological Synapse Architecture and T Cell Activation. *Sci. Signal.* **7**, ra37–ra37 (2014).
18. Hart, T., Komori, H. K., LaMere, S., Podshivalova, K. & Salomon, D. R. Finding the active genes in deep RNA-seq gene expression studies. *BMC Genomics* **14**, (2013).
19. Solovei, I. *et al.* Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution. *Cell* **137**, 356–368 (2009).
20. Tanabe, H. *et al.* Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc. Natl. Acad. Sci.* **99**, 4424–4429 (2002).
21. Ricciardi, S. *et al.* The Translational Machinery of Human CD4+ T Cells Is Poised for Activation and Controls the Switch from Quiescence to Metabolic Remodeling. *Cell Metab.* **28**, 961 (2018).
22. Singh, K. *et al.* A KAP1 phosphorylation switch controls MyoD function during skeletal muscle differentiation. *Genes Dev.* **29**, 513–525

(2015).

23. Wu, F., Olson, B. G. & Yao, J. DamID-seq: Genome-wide mapping of protein-DNA interactions by high throughput sequencing of adenine-methylated DNA fragments. *J. Vis. Exp.* **2016**, 1–11 (2016).

CHAPTER 5.
FINAL CONSIDERATIONS

5.1. Summary and conclusions

Chromatin organization inside the nucleus of eukaryotic cells fulfil a key role in the shaping of gene expression[1,2]. Thus, the first step into potentially solving altered cell phenotypes resides in the deep understanding of the mechanisms that regulate chromatin twist and turn inside the nucleus. Our research group is particularly interested in the keepers of the correct genome folding in the nuclear space, as well as in the modifiers acting to adapt chromatin structure during dynamic cell processes. Studying muscle stem cells (MuSCs), whose differentiation to mature myocytes needs the sequential epigenetic repression of specific transcriptional programmes[3], we demonstrated that Lamin A/C is a crucial chromatin organizer and a keeper of cell identity. In fact, it preserves MuSCs identity and their regenerative capacity by impeding the spreading of repressive histone marks along the chromatin fibre. In the absence of Lamin A/C, Polycomb proteins, transcriptional repressors involved in the regulation of developmental genes, diffuse from their targets over the flanking genomic regions. This causes the loss of higher-order structures and the de-repression of bivalent genes, key players in stemness maintenance and cell identity specification. As a consequence, a portion of MuSCs gains co-expression of muscular and adipogenic markers and loses self-renewal ability, leading to muscular dystrophy. These data show the importance of three-dimensional chromatin organization at specific genomic loci for the implementation of a proper gene expression programme. In another work, we developed the high-throughput sequencing-based technology of SAMMY-seq (under patenting, application n. 18200482.0–1118) with the aim of expanding our perspective on chromatin organization to a genome-wide level. With

SAMMY-seq, in a single experimental protocol on living cells, we are able to split chromatin into three compartments of decreasing solubility. Results obtained on healthy skin primary fibroblasts show that lamina-associated heterochromatic regions (LADs)[4] are reproducibly mapped as the less soluble portions of chromatin (insoluble SAMMY-domains). In contrast, fibroblasts at early stages of Hutchinson-Gilford Progeria Syndrome (HGPS) reveal large-scale alterations in the insoluble regions. In particular, early-stage HGPS cells lose various SAMMY-domains with respect to control fibroblasts, even without a corresponding loss of H3K9me3 or transcriptional deregulation of the genes residing in these regions. However, as for MuSCs lacking Lamin A/C, in early-stage HGPS cells we found anomalies in the genomic distribution of H3K27me3 mark, with functional effects on PcG-regulated bivalent genes. These aberrancies can be at the origin of HGPS senescent phenotype. All together, these data indicate that SAMMY-seq captures a chromatin remodeling upstream of the epigenetic alterations seen in the advanced HGPS phenotype[5].

SAMMY-seq represents probably the easiest method for the identification of lamina-associated heterochromatic regions at a genome-wide level in primary cells. In fact, SAMMY-seq does not require the expression of exogenous genes, it is independent from the chemical modification of chromatin and does not rely on antibodies. Improvements introduced in SAMMY-seq protocol, with the separation of small (S2S) and large (S2L) DNA fragments originating from S2, further enhanced the potential of this technology. The increased specificity of the signal enrichment detected in S2 subfractions, in fact, allowed the simultaneous identification of soluble and insoluble, both highly reproducible, domains. In fibroblasts and myoblasts, these chromatin domains have a high correlation with

euchromatin and heterochromatin, respectively. On the other hand, SAMMY-seq on quiescent $CD4^+$ T lymphocytes extracted from the peripheral blood of healthy donors evidenced an inversion in the SAMMYdomains-HMs correlation. Soluble fractions correlate with heterochromatin marks, while insoluble fractions with active transcription marks. Recent findings described the inverted nuclear localization of euchromatin and heterochromatin in the nuclei of rod photoreceptors in the retina of nocturnal mammals[6]. Hypothesizing this could explain the unconventional solubility pattern of $CD4^+$ T lymphocytes chromatin, we performed FISH and immunofluorescence experiments. The results exclude the inversion, indicating a classical peripheral positioning for constitutive heterochromatin. However, a portion of euchromatin also localizes along the nuclear lamina, suggesting an uncoupling between solubility and transcription in lymphocytes, with a non-univocal correspondence between either of them and nuclear positioning. To address at the functional level the role of soluble and insoluble conformation, we performed some gene ontology (GO) analysis. Active genes, highly enriched in the insoluble fraction, show overrepresentation of loci involved in basic metabolic processes[7] confirming that functional euchromatin preferentially resides in the insoluble fraction. In the inactive gene pool, equally divided between soluble and insoluble fractions, we observed overrepresentation of pathways implicated in T cells activation, proliferation and cytokine secretion, together with pathways unrelated to lymphocytes functions. The residence of part of the T cell specific genes in the soluble rather than insoluble compartment could indicate an higher predisposition for rapid activation upon TCR stimulation. A detailed differential analysis of pathways enriched in one or the other chromatin compartment will possibly elucidate a time-dependent activation dynamic.

Preliminary analysis on the domains enriched in S2SvsS2L show a higher gene density compared to the other soluble domains (Fig. 1a) and a much higher percentage of active genes (Fig. 1b). Overall, genes in these domains show little overlap with previously identified domains (data not shown) and GO analysis on Th1 reveals a clear enrichment of genes related to T cell activation. This first evidence supports the hypothesis of a potential solubility-controlled timeline of gene activation.

Since S2S fraction is extremely variable in Tnaive and Treg cells, we are now planning to relieve the DNase step of SAMMY-seq, to evaluate a possible increase in the recovery of activation-related genes in their soluble compartment, further validating our hypothesis.

a

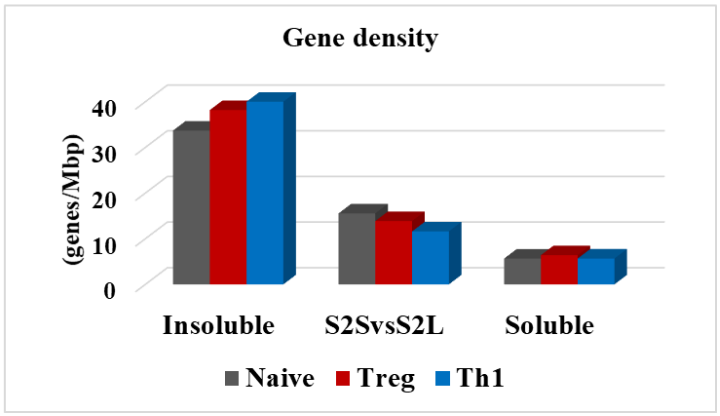


Figure 1 continued on the next page.

Figure 1 continued.

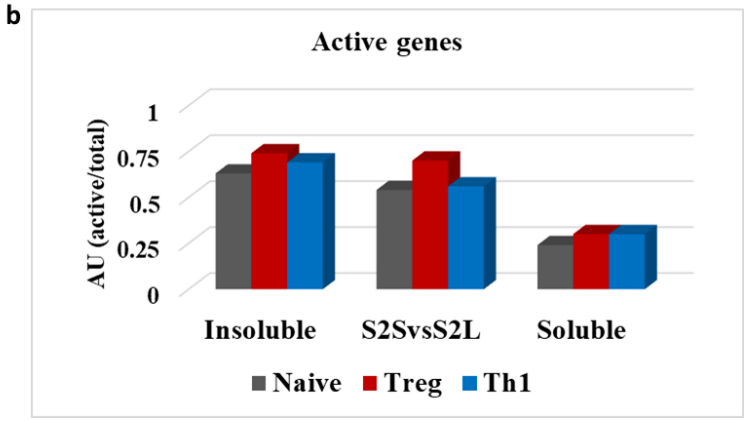


Figure 1 – Preliminary characterization of S2SvsS2L compartment. **a,** Average number of genes for each megabase of genome in insoluble (union of S4vsS2L and S4vsS3), soluble (union of S2LvsS4 and S3vsS4) and S2SvsS2L SAMMY-domains in Tnaive, Treg and Th1 cells. **b,** Proportion of active genes in insoluble, soluble and S2SvsS2L domains in Tnaive, Treg and Th1 cells.

5.2. Future perspectives

The unconventional nuclear architecture observed in $CD4^+$ T lymphocytes stimulates key questions: 1-Which is the link between chromatin solubility and genome function? 2-Why lymphocytes present this unconventional chromatin architecture?

The emergence in Th1 cells of two groups of repressed genes related to T cell activation, one in the more soluble fraction (S2SvsS2L) and the other one in the less soluble fraction (S4vsS2L and S4vsS3), suggests they could have different transcriptional dynamics during the course of activation. We speculate that genes in the more soluble compartment may be involved in the sensing of the environment and in the first reactions to TCR engagement while the genes in the insoluble compartment could be involved in the consolidation of the active state and in the specialized

downstream functions. At present we are pursuing detailed GO analysis aiming at verifying this hypothesis.

Furthermore, to trace changes in the solubility of genes identified in the quiescent state, we are working on SAMMY-seq analysis of *in vitro* activated Tnaive and Th1 cells. Immunofluorescent assays spanning early lymphocytes activation steps already reported changes in H3K36me3 nuclear pattern (Fig. 2), suggesting a fast remodelling of active chromatin during this process.

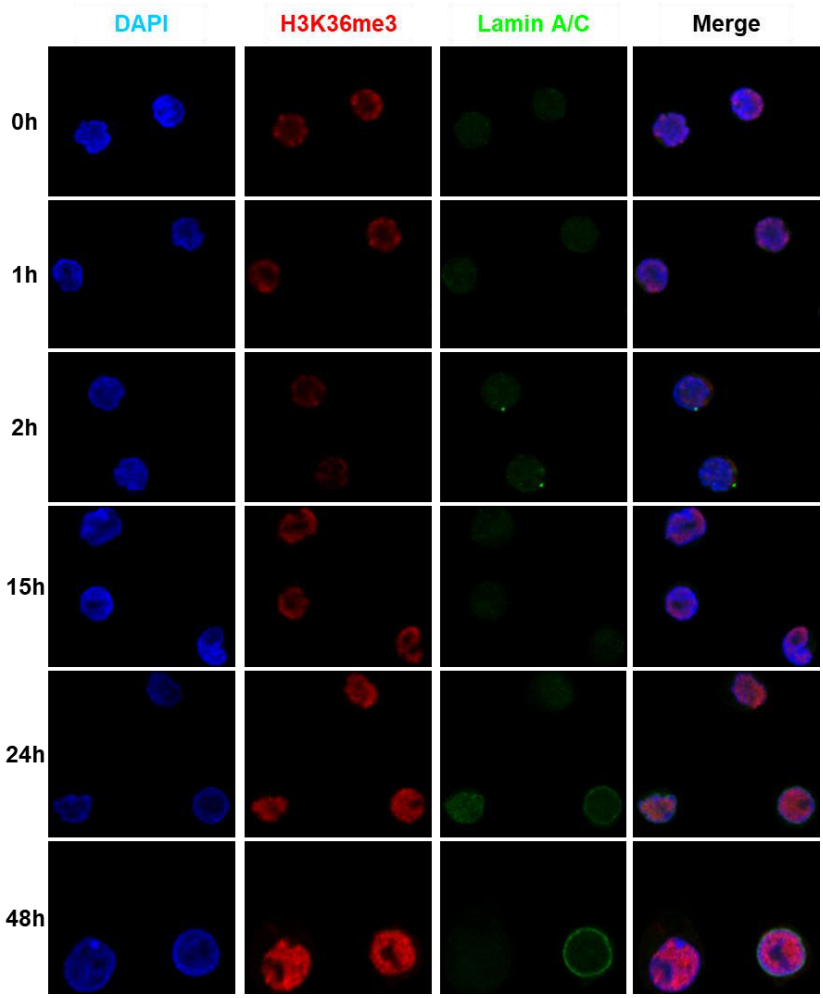


Figure 2 – Euchromatin reorganization in CD4$^+$ Naive T lymphocytes during activation. Representative confocal images of

$CD4^+$ Naive T lymphocytes before (0h) and during (2h, 15h, 24h and 48h) *in vitro* stimulation with CD3/C28 magnetic beads. For each time point DAPI (blue), H3K36me3 (red) and Lamin A/C (green) staining are shown.

Another hypothesis, not mutually exclusive with the above, associates lymphocytes unconventional chromatin organization with their capability to bear extremely high levels of environmental mechanical stress. Lymphocytes, in fact, patrol the organism by continuously migrating through blood and lymph. Furthermore, to extravasate towards sites of inflammation, they need to squeeze between one cell and the other in the endothelial wall, infiltrating then into tissues with various degrees of stiffness. Mechanical stress could cause serious damage to lymphocytes nuclei and their DNA[8]. Therefore, the incorporation of euchromatin in an insoluble environment could help its protection from breaks and consequent mutations. The analysis of mutagenesis rate in lymphocytes euchromatin and heterochromatin could validate this hypothesis.

To continuously increase the potential of SAMMY-seq technology, we also recently adapted it for the recovery of RNA from each chromatin fraction. Results obtained on myoblast samples indicate a clear differential distribution of RNA classes between the fractions (Fig. 3). For example, mRNAs are predominant in S1, while long non coding (lnc) RNAs increase towards the more insoluble chromatin compartments (data not shown).

For some well characterized lncRNAs we were also able to associate the distribution across fractions with the described function[9–11] (Fig. 4).

With this final advance, we propose that SAMMY-seq technology is an easy, flexible and competitive tool for the detailed characterization of chromatin conformation in any type of cells. The simultaneous analysis of

protein, DNA and RNA enrichment in distinct chromatin fractions of cultured or primary cells will in fact provide an accurate snapshot of their chromatin organization. This information, matched with phenotypic characterization, could finally pave the way for a more complete understanding of the effects of chromatin conformation on cell identity.

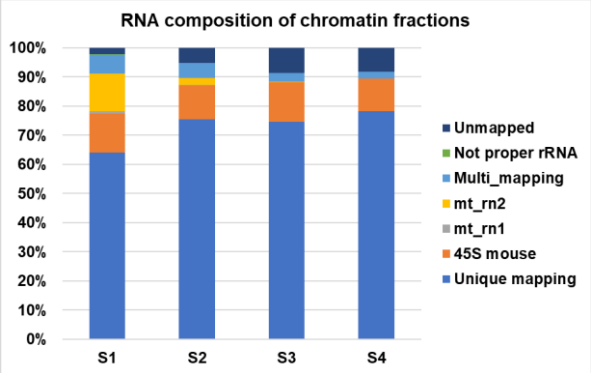


Figure 3 – RNA composition of murine myoblasts SAMMY-seq fractions. Percentage of uniquely mapped (light-blue), 45S (orange), mitochondrial (grey and yellow), multimapping (cyan), not proper rRNA (green) and unmapped (dark-blue) RNA with respect to the total RNA retrieved from every SAMMY-seq fraction of a C2C12 sample.

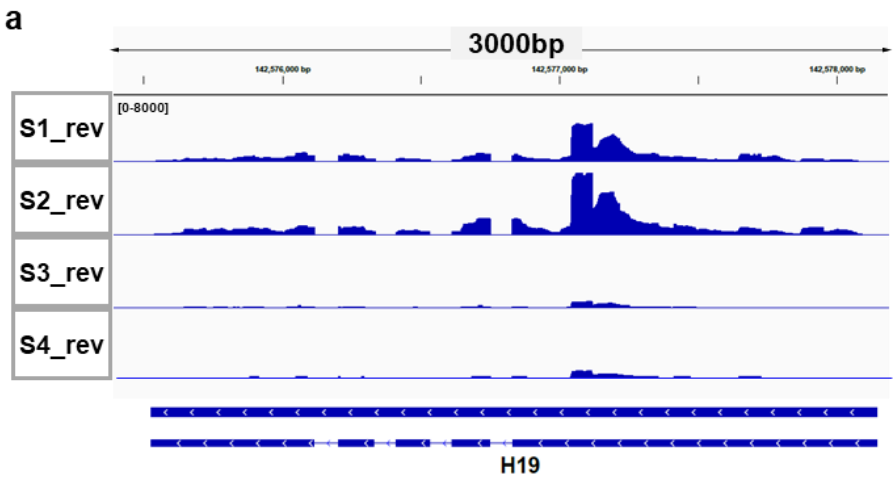


Figure 4 continued on the next page.

Figure 4 continued.

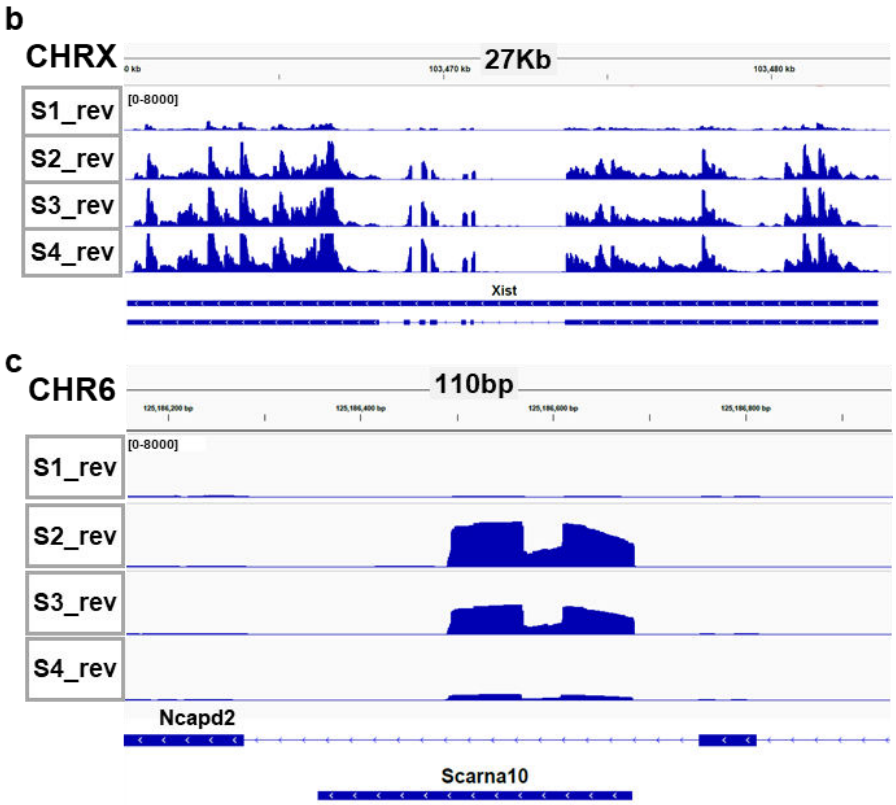


Figure 4 – LncRNAs distribution between SAMMY-fractions. a, b, c, Distribution of SAMMY-seq reads in S1, S2, S3 and S4 of a C2C12 sample along the region of H19 (chr7:142,575,500-142,578,200) (a), the region of Xist (chrX:102,460-103,490) (b) and the region of scarna10 (chr6:125,186,050-125,187,150) (c). Library size normalized read counts are shown.

References to chapter 5

1. Sartorelli, V. & Puri, P. L. Shaping Gene Expression by Landscaping Chromatin Architecture: Lessons from a Master. *Mol. Cell* **71**, 375–388 (2018).
2. Sewitz, S. A., Fahmi, Z. & Lipkow, K. Higher order assembly: folding the chromosome. *Curr. Opin. Struct. Biol.* **42**, 162–168 (2017).
3. Robinson, D. & Dilworth, F. J. Epigenetic regulation of adult myogenesis. *Curr. Top. Dev. Biol.* 235–285 (2018).
4. van Steensel, B. & Belmont, A. S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780–791 (2017).
5. Gonzalo, S., Kreienkamp, R. & Askjaer, P. Hutchinson-Gilford Progeria Syndrome: A premature aging disease caused by LMNA gene mutations. *Ageing Res. Rev.* **33**, 18–29 (2017).
6. Solovei, I. *et al.* Nuclear Architecture of Rod Photoreceptor Cells Adapts to Vision in Mammalian Evolution. *Cell* **137**, 356–368 (2009).
7. Ricciardi, S. *et al.* The Translational Machinery of Human CD4+ T Cells Is Poised for Activation and Controls the Switch from Quiescence to Metabolic Remodeling. *Cell Metab.* **28**, 961 (2018).
8. Bennett, R. R. *et al.* Elastic-Fluid Model for DNA Damage and Mutation from Nuclear Fluid Segregation Due to Cell Migration. *Biophys. J.* **112**, 2271–2279 (2017).
9. Lecerf, C., Le Bourhis, X. & Adriaenssens, E. The long non-coding RNA H19: an active player with multiple facets to sustain the hallmarks of cancer. *Cell. Mol. Life Sci.* **76**, 4673–4687 (2019).
10. Loda, A. & Heard, E. Xist RNA in action: Past, present, and future. *PLOS Genet.* **15**, e1008333 (2019).
11. Zhang, K. *et al.* SCARNA10, a nuclear-retained long non-coding RNA, promotes liver fibrosis and serves as a potential biomarker. *Theranostics* **9**,

3622–3638 (2019).

PUBLICATIONS AND PATENTS

- Bianchi A., Manti P.G., Lucini F., & Lanzuolo, C. (2018). *Mechanotransduction, nuclear architecture and epigenetics in emery dreifuss muscular dystrophy: Tous pour un, un pour tous*. Nucleus, 9(1), 321-335.

- Lucini F., Bianchi A., Lanzuolo C. *Formaldehyde-mediated snapshot of nuclear architecture.* Methods Mol Biol., *in press*

- Sebestyén E., Marullo F., Lucini F., Bianchi A., Petrini C, Valsoni S., Olivieri I., Antonelli L., Gregoretti F., Oliva G., Ferrari F. and Lanzuolo C. *Early Polycomb-target deregulations in Hutchinson-Gilford Progeria Syndrome revealed by heterochromatin analysis.* BioRxiv 799668 *[Preprint]*. October 10, 2019. doi: 10.1101/799668

- Bianchi A., Mozzetta C., Lucini F., Valsoni S., Pegoli G., Rosti V., Petrini C., Cortesi A., Antonelli L., Gregoretti F., Oliva G., De Bardi M., Rizzi R., Bodega B., Pasini D., Ferrari F., Bearzi C.and Lanzuolo C. *Polycomb dysfunctional transcriptional repression contributes to Lamin A/C dependent muscular dystrophy. Submitted*

- Livi C. M., Tagliaferri I., Pal K., Sebestyén E., Bianchi A., Lucini F., Lanzuolo C. and Ferrari F. *A ChIC solution for ChIP-seq quality assessment. Submitted*

- Lanzuolo C., Marullo F., Lucini F., Ferrari F., Sebestyén E. *Method for sequential analysis of macromolecules.* Application n. 18200482.0–1118. *Under patenting*