Université de Paris

École doctorale BIOSPC ED562
Institut Pasteur - Oncogenesis and Nuclear Organization / Senescence and Age-related Pathologies

Applying Systems Biology and Multi-layered Profiling Towards a Comprehensive Understanding of Cell Aging

José Américo Nabuco Leva Ferreira de Freitas

Thèse de doctorat de Biologie Cellulaire

Dirigée par **Oliver Bischof** et **Benno Schwikowski**

Présentée et soutenue publiquement à Paris le 25 mars 2020

Mendoza, Marco	CR1 - Genoscope	Président du jury
Kuperstein, Inna	Project Coordinator - Institut Curie	Rapporteur
Narita, Masashi	Group Leader - University of Cambridge	Rapporteur
Li, Han	G5 Group Leader – Institut Pasteur	Membre invité
Thomas-Vaslin, Véronique	Group Leader - Sorbonne Université	Membre invité
Bischof, Oliver	DR2 – Institut Pasteur	Directeur de thèse
Schwikowski, Benno	DR2 – Institut Pasteur	Co-directeur de thèse

Abstract

Cellular senescence (CS) is a cell fate characterized by a stable cell cycle arrest of dysfunctional cells. CS has an intricate role in physiology and pathophysiology. Senescent cells play a vital role in tumor suppression, embryonic development, and wound healing, but also in many age-related pathologies, including paradoxically, tumor development. My thesis work represents a comprehensive time-resolved analysis of the epigenome, transcriptome, and metabolome layers of cells undergoing CS and is divided into three subprojects.

First, I investigated the dynamics of transcription factor (TF) binding to enhancers in oncogene-induced senescence (OIS). TFs organize in a hierarchical network, with pioneers shaping the enhancer landscape by recruiting settlers and migrants to fine-tune gene expression. Specifically, I discovered that the AP1 family members precede the majority of other TFs, priming chromatin to initiate and coordinate the CS transcriptional response.

Second, I performed an extensive analysis of the metabolic changes associated with CS, integrating results from fibroblasts undergoing replicative, oncogene-induced, and DNA damage-induced senescence, in addition to a characterization of OIS in primary myoblasts. I identified several metabolites that accumulate or diminish in senescent cells, and those are associated with post-translational modifications, protein synthesis, lipid biosynthesis and oxidation, and energy production. In particular, alpha-ketoglutarate (aKG) and uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) act as substrates for chromatin modifiers, suggesting roles in gene regulation.

Third, I defined a mathematical model describing the transcriptional evolution of cells undergoing OIS. I generated this model using the Sparse Identification of Nonlinear Dynamics (SINDy) algorithm in a high-performance computing environment. I validated the model with transcriptome data derived from JUN and RELA depletion experiments. On inhibition of JUN, a member of the AP1 family, the model simulation behaved closer to senescent cells than on RELA, suggesting that TF rank in the chromatin binding hierarchy may determine the predictability of its transcriptional response.

Together, my integrative analysis provides a deeper understanding of CS and has the potential to reveal previously unknown vulnerabilities of senescent cells that may be exploited to treat cancer and age-related diseases, promoting a longer healthspan.

Keywords : cellular senescence, aging, epigenomics, transcriptomics, metabolomics, systems biology

Résumé

La sénescence cellulaire (SC) qui correspond à un destin cellulaire est caractérisée par un arrêt stable de son cycle dont les cellules présentent un dysfonctionnement. La SC joue un rôle complexe dans la physiologie et la physio-pathologie. Les cellules sénescentes sont extrêmement impliquées dans la suppression de tumeurs, le développement embryonnaire et la cicatrisation des plaies, mais aussi dans de nombreuses pathologies liées à l'âge, y compris, paradoxalement, dans le développement de tumeurs. Mon travail de thèse représente une analyse exhaustive, en temps différé, des différentes couches de l'épigénome, du transcriptome et du métabolome des cellules qui entrent en SC. Il est divisé en trois sous-projets.

Tout d'abord, j'ai étudié la dynamique de la liaison des facteurs de transcription (FT) aux activateurs dans la sénescence induite par oncogène (SIO). Les facteurs de transcription s'organisent en un réseau hiérarchisé, les "pioneers" façonnant le paysage des activateurs en recrutant des "settlers" et des "migrants" pour affiner la régulation de l'expression des gènes. Plus précisément, j'ai découvert que les membres de la famille des AP1 précèdent la majorité des autres FT, amorçant la chromatine afin d'initier et coordonner la réponse transcriptionnelle de la SC.

Ensuite, j'ai effectué une analyse approfondie des changements métaboliques associée à la SC, en intégrant les résultats de fibroblastes qui entrent en sénescence réplicative, induite par oncogènes et par dommages à l'ADN, en parallèle d'une caractérisation de la SIO dans les myoblastes primaires. J'ai identifié plusieurs métabolites qui s'accumulent ou diminuent dans les cellules sénescentes. Ceux-ci sont associés à des modifications post-traductionnelles, à la synthèse des protéines, à la biosynthèse et à l'oxydation des lipides, ainsi qu'à la production d'énergie. En particulier, l'alpha-cétoglutarate (aKG) et l'uridine diphosphate N-acétylglucosamine (UDP-GlcNAc) agissent comme des substrats pour les modificateurs de la chromatine, ce qui suggère un implication dans la régulation des gènes.

Troisièmement, j'ai défini un modèle mathématique décrivant l'évolution transcriptionnelle des cellules entrant en SIO. J'ai généré ce modèle en me servant de l'algorithme Sparse Identification of Nonlinear Dynamics (SINDy) dans un environnement de calcul haute performance. J'ai validé le modèle avec des données de transcriptomiques provenant des expériences d'inhibition de JUN et RELA. Lors de l'inhibition de JUN, un membre de la famille des AP1, la simulation du modèle s'est davantage rapprochée de la sénescence cellulaire que de RELA. Ce qui suggère que le rang des FT dans la hiérarchie de la liaison à la chromatine pourrait être déterminée par la prévisibilité de sa réponse transcriptionnelle.

Dans son ensemble, mon analyse intégrative permet une meilleure compréhension de la SC et possède le potentiel de révéler des vulnérabilités jusqu'alors inconnues des cellules sénescentes et qui pourraient être utilisées en vue de traiter le cancer et les maladies liées à l'âge, favorisant ainsi une plus longue et meilleure qualité de vie.

Mot-clés : sénescence cellulaire, vieillissement, épigénomique, métabolomique, transcriptomique, biologie des systèmes

Ao vovô José, que molda o homem que eu sou
À vovó Dalila, que me dá a energia para viver
que radiavam calorosamente ao decolar do voo
e hoje repousam após o anoitecer

Acknowledgments

This is the moment where I quote the already overused citation of Sir Isaac Newton: "If I have seen further, it is by standing on the shoulders of giants". This work is a natural consequence of all the amazing encounters I have been gifted with, and my role was mainly to piece together a story built by the collective knowledge of brilliant scientists.

I would like to thank my family, Clarissa N. L. F. de Freitas, Josneimar F. de Freitas, Marília N. L. F. de Freitas, Maddie N. Freitas and Hernâni N. L. F. de Freitas for the unconditional support in this one more phase of my life. I can't describe how glad I am to realize that we are all flying higher than three years ago, and I can't wait to see how our lives will unfold in the upcoming years. On the other side, this growth did not happen without pain, and I would like to use this space to immortalize the memory of Dalila Clara Nabuco Leva and José de Freitas, which laid down the foundations for our values with they beloved partners, Neusa Garcia Ferreira de Freitas and Américo Adauto Leva, and now lay on their eternal rest.

Thank you to the jury committee: Marco Mendoza, Inna Kuperstein, Masashi Narita, Han Li and Jana Wolf for dedicating your highly demanded attention to evaluating this work and your precious insights.

Thank you, Oliver Bischof, for welcoming this computer guy into your lab. I am still amazed by how life and technology allowed our journeys to cross and the outcome we produced out of this unlikely encounter. Working at your group was a pleasure and I will always be proud to have been able to interact with smart and hard working people on a daily basis. Our contribution to the field would not be possible without your visionary perspective on the importance of dynamics and a comprehensive profile of this agitated molecular dance that develops in aging cells.

Benno Schwikowski, thank you for engaging in this ambitious project with this overseas guy bringing a crazy idea. Your contributions were essential for the quality of this project, ensuring I addressed several technical aspects often overlooked by those who don't fully grasp the properties of the tools they handle. I am also grateful for your vital support when I needed the most.

Thanks to the everyone in Oliver's team: Lucas, Pierre-François, Nir, Greg, Lisa, Cláudia and Ricardo. You certainly have improved this work in our daily interactions and insightful observations. I also enjoyed a lot spending time with all of you personally, and I will always carry our memories with kindness.

Also, this project would not be possible without the contribution of

everyone in the ONO team. I was friendly welcomed by all of you straight on my first days in France and our relationship only grew since then. Our discussions also shaped my world perspectives and the directions of this work. Thank you, Anne Dejean, Florence, Samia, Coraline, Jacob, Alexandra, Jack, Pascal, Agnés, Cathy, Monique, Sophie, Elma, Thibault, Juan-Pablo, Victoria, Nacho, Tatiana, Ying and Yann. A special thanks to Ilan for going through all this manuscript, offering valuable insights, and for the great time spent together.

The AGEMED consortium also had a critical contribution to this work. Thank you Delphine Benarroch and Eric Gilson for cultivating the replicative senescent cells, Yara Bou Saada and Bertrand Friguet for the myoblasts, Ivan Nemazanyy and Mario Pende for the metabolomics data acquisition and technical advice. I would also like to thank Pauline Sechet, Katja Schürer, Cecilia Benedito, Nicolas Pecuchet, Sylvain Benito and Elisabeth Pecougambando for the rich discussions and insights on the mathematical model.

Thanks you to the members of the BCI department, for fruitful discussions in both scientific and social gatherings. Being immersed on the department has transformed my scientific worldview and allowed me to grasp a great number of concepts I could not familiar with in other environments. I also would like to thank Marc Lecuit, Guillaume Dumenil and the tutorat team, Karin Eiglmeier, Olivier Cassar and Catherine Dauga, for the institutional support and funding extension at the end of my contract.

I am also glad to have the opportunity to be part of the StaPa bureau and its whole community. The energy of these young researchers is inspiring and I feel I am more competent both on grasping biological concepts and managing scientific and social events. A special mention to YouPI members at Hellenic Pasteur Institute in Greece and to the Joint International Retreat organizers in Institut Pasteur in Italy, for all the scientific exchange between organizations and for showing me the borderless legacy of Louis Pasteur. Our efforts definitely brought the Pasteur community a little closer, and I am glad to know now I have friends all over Europe.

Thanks to my floormates at Fondation Deutsh de la Meurthe, the residents at Maison du Brésil (including the most wonderful floor) and all the Cité Universitaire community. This stimulating environment also allowed me to experience worldviews I could barely imagine of, including from regions of my own home country I have never been to. Thank you, Ana Gabriela Lima Bispo de Victa, for the amazing time together and support in rough times. I definitely became a better person by your side. A special thanks to the direction of Maison du Brésil and the members of the Comité des Residents Ablo Dianka.

And thanks to you that read everything until this point, having a glimpse of the rich experience life has gifted me during the last years.

You all made this time of my life a very special moment that will be remembered with tenderness and already makes me nostalgic. I wish you all a prosperous existence and a long healthspan!

Support

This project was financed by *La Ligue Nationale Contre le Cancer* and was only possible thanks to the generosity of its donors.

This work used the computational and storage services (TARS cluster) provided by the IT department at Institut Pasteur, Paris.

The authors acknowledge support by the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen, the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG.

Contents

Glossary **x**

List of Figures **xiv**

1 Introduction **1**
1.1 Cellular senescence 3
1.2 Thesis overview 5

2 A hierarchical transcription factor network governs cellular senescence gene expression **7**
2.1 Introduction 8
2.1.1 Senescence-associated growth arrest 9
2.1.2 Senescence-associated secretory phenotype 11
2.1.3 Transcription factor networks 13
2.2 Results 16
2.2.1 The TF hierarchy driving the senescence transcriptional program 16
2.3 Discussion 24
2.4 Conclusions and Perspectives 26
2.5 Materials and methods 29
2.5.1 Cell culture 29
2.5.2 Gene expression analysis 29
2.5.3 Histone modification ChIP-seq and chromatin state differential analysis 29
2.5.4 ATAC-seq 29
2.5.5 Transcription factor footprinting 30
2.5.6 Transcription factor chromatin binding properties . . . 30
2.5.7 TF chromatin binding hierarchy network 32
2.5.8 Proportion of incoming edges based on the classification of the TF source node 34
2.5.9 Transcription factor hierarchy networks overlap 34

3 Metabolic reprogramming in cellular senescence 35
3.1 Introduction . . . 36
3.1.1 Causes of cellular senescence . . . 36
3.1.2 The cellular senescence metabolome . . . 39
3.1.3 Metabolomics . . . 41
3.2 Results . . . 46
3.2.1 Biological discoveries . . . 46
3.2.2 Developed computational methods . . . 49
3.3 Discussion . . . 52
3.4 Conclusions and Perspectives . . . 56
3.5 Materials and methods . . . 57
3.5.1 Cell culture . . . 57
3.5.2 Liquid Chromatography and mass spectrometry . . . 57
3.5.3 Mass spectrometry data pre-processing . . . 57
3.5.4 Compounds annotation . . . 58
3.5.5 Statistical analysis . . . 58
3.5.6 Hierarchical clustering . . . 58
3.5.7 Integrative analysis . . . 59
3.5.8 Batch correction methods benchmark . . . 59

4 Modelling gene regulatory networks in oncogene-induced senescence 62
4.1 Introduction . . . 63
4.1.1 The epigenetic landscape governing cell fate . . . 63
4.1.2 Gene regulatory networks inference . . . 65
4.1.3 Compressed sensing . . . 66
4.2 Results . . . 68
4.2.1 Biological discoveries . . . 68
4.2.2 Developed computational methods . . . 80
4.3 Discussion . . . 82
4.4 Conclusions and Perspectives . . . 85
4.5 Materials and methods . . . 87
4.5.1 Cell culture, RNA extraction and microarrays . . . 87
4.5.2 Microarray transcriptome data preprocessing, statistical analysis and annotation . . . 87
4.5.3 Hierarchical clustering and identification of unique expression time profiles . . . 87
4.5.4 Model inference . . . 88
4.5.5 Model simulation . . . 89
4.5.6 Model assessment . . . 89
4.5.7 Canonical systems tool validation . . . 89

4.5.8 Hidden sources analysis 90
4.5.9 Inference tools benchmark 90

5 Conclusion **92**

Bibliography **94**

Appendix **126**
A.1 Contribution to publication 127

Glossary

αKG α-keto glutarate
γ-H2AX Phosporilated H2AX
σ_{avg} Average variance
C. elegans *Caenorhabditis elegans*
4-OHT tamoxifen
5LO 5-Lipoxygenase
AA Arachidonic acid
ACSL Acyl-CoA synthetase long chain
ADP Adenosine diphosphate
AKT Protein kinase B
AMP Adenosine monophosphate
AMPK AMP-activated protein kinase
ANOVA Analysis of variance
AP1 Activator protein 1
ARF Alternate open reading frame
ARP Age-related pathology
ATAC-seq Assay for transposase accessible chromatin
ATM Ataxia telangiectasia mutated
ATP Adenosine triphosphate
ATR Ataxia telangiectasia and Rad3 related
BC Batch correction
BE Batch effect
BFGS Broyden–Fletcher–Goldfarb–Shannon
bp Base pair
CCNA2 Cyclin A2
CCNE Cyclin E
CDK (-2, -4, -6) Cyclin-dependent kinases -2, -4, -6
CEBPβ CCAAT enhancer binding protein β
cGAS–STING Cyclic GMP–AMP synthase linked to stimulator of interferon genes
ChIP-seq Chromatin immunoprecipitation sequencing

CHK (-1, -2) Checkpoint kinases -1, -2
CoS Compressed sensing
CPT1 Carnitine palmitoyltransferase
CPU Central processing unit
CREB CAMP responsive element binding protein
CS Cellular senescence
CTCF CCCTC binding factor
CV Coefficient of variation
CXCL8 C-X-C motif chemokine ligand 8
DD DNA damage
DDR DNA damage response
DNA Deoxyribonucleic acid
DNA-SCARS DNA segments with chromatin alterations reinforcing senescence
dNTP Deoxyribonucleotides
DR Dietary restriction
DSB Double-strand break
E2F E2F transcription factor
EGR Early growth response
ELF1 E74 like ETS transcription factor 1
ELK4 ETS transcription factor ELK4
ErbB Erb-B2 receptor tyrosine kinase
ERK Extracellular signal-regulated kinase
ETS1 E twenty-six proto-oncogene 1
ETV6 ETS variant transcription factor 6
FAO Fatty acid oxydation
FAS Fatty acid synthesis
FBS Fethal bovine serum
FDR False discovery rate
FOXO Forkhead box O
FTRL Follow the regularized leader
G6P Glucose-6-phosphate
GABPα GA binding protein transcription factor subunit alpha
GATA4 Transcription factor GATA-4
GC Gas chromatography
GLUD1 (GDH1) Glutamate dehydrogenase 1
GLUT1 Glucose transporter 1
GPU Graphics processing unit
GRN Gene regulatory network
H2AX H2A histone family member X
H3K27Ac Histone 3 lysine 27 acetylation

H3K27me3 Histone 3 lysine 27 tri-methylation
H3K4me1 Histone 3 lysine 4 mono-methylation
H3K4me3 Histone 3 lysine 4 tri-methylation
H_2O_2 Hydrogen peroxide
HAT Histone acetyltraferase
HDAC Histone deacetylase
HDM2 Human homolog Of mouse double minute 2
HPC High performance computing
HPLC High performance liquid chromatography
Hsp72 Heat shock protein 72
HSPA9 (Mortalin) Heat shock 70kDa protein 9
IKK Inhibitor of nuclear factor kappa B kinase
IL (-1α, -1β, -6, -8) Interleukin -1α, -1β, -6, -8
JMJ3 Jumonji domain-containing protein 3
JMJD2D (KDM4D) Jumonji domain containing 2D
JUN Jun proto-oncogene
KD Knock-down
KLF Kruppel like factor
LC Liquid chromatography
MaBoSS Markovian Boolean stochastic simulator
MAFG MAF bZIP transcription factor G
MAPK Mitogen activated protein kinases
MDC1 Mediator of DNA damage checkpoint protein 1
MDM2 Mouse double minute 2
MEK Mitogen-activated protein kinase kinase 1
MiDAS Mitochondrial dysfunction-associated senescence
MRE11 Meiotic recombination 11 homolog 1
MS Mass spectrometry
mTOR Mamalian target of rapamycin
Myc Avian myelocytomatosis viral oncogene homolog
N1ICD Notch intracellular domain
NADH (NAD^+) Nicotinamide adenine dinucleotide
NAM Nicotinamide
NF-κB Nuclear factor κB
NFIL3 Nuclear factor interleukin 3 regulated protein
NLM Non-linear mapping
NMR Nuclear magnetic ressonance
Notch Notch receptor
O-GlcNAc O-linked N-acetylglucosamine
OGT O-Linked N-acetylglucosamine Transferase
OIS Oncogene-induced senescence

p16 (CDKN2A) Cyclin-dependent kinase inhibitor 2A
p21 (CDKN1A) Cyclin-dependent kinase inhibitor 1A
P38MAPK P38 mitogen-activated protein kinases
p53 Tumor protein 53
p63 Tumor protein 63
p73 Tumor protein 73
PARP1 Poly-ADP-ribose polymerase 1
PCA Principal component analysis
PcG Polycomb-group protein
PCG1α Peroxisome proliferator-activated receptor gamma coactivator-1-alpha
PI3K Phosphoinositide 3-kinase
PIQ Protein interaction quantification
PLS-DA Partial least squares - discriminant analysis
PTEN Phosphatase and tensin homolog
PTM Post-translational modification
QC Quality control
QN Quantile normalization
RAF Rapidly accelerated fibrosarcoma
RAS Rat sarcoma viral oncogene homolog
RB Retinoblastoma protein
RELA REL-associated protein A
RIP Restricted isometry property
RMA Robust multichip average
RNA Ribonucleic acid
RNA-seq RNA sequencing
ROS Reactive oxygen species
RRM2 Ribonucleotide reductase
RSD Relative standard deviation
SA-β-Gal Senescence-associated β-galactosidase
SAGA Senescence-associated growth arrest
SAHF Senescence-associated heterochromatin foci
SASP Senescence-associated secretory phenotype
SCC Strongly connected component
SEA Set enrichment analysis
SINDy Sparse identification of nonlinear systems
siRNA Small interferring RNA
SIRT Silent information regulator
SnC Senescent cells
SP Specificity protein
SSB Single-strand break

SUV39H1 Suppressor of variegation 3-9 homolog 1
TCA Tricarboxylic acid
TEF Thyrotroph embryonic factor
TET Ten-eleven transcription
TF Transcription factor
TGFβ Transforming growth factor beta 1
TNF Tumor necrosis factor
TNFα Tumor necrosis factor
TP53BP1 Tumor protein P53 binding protein 1
TR Transitive reduction
TRF2 Telomeric repeat binding factor 2
UDP-GlcNAc Uridine diphosphate N-acetylglucosamine
UPLC Ultra performance liquid chromatography
UVB Utra-violet radiation
VEGF Vascular endothelial growth factor
WGCNA Weighted gene correlation analysis
WI38 Wistar institute 38
YY1 Yin yang 1

List of Figures

1.1 CS integrates the hallmarks of aging 3

2.1 Transcriptional changes in CS 8
2.2 E2F inhibition by RB 10
2.3 Cell cycle arrest induced by the p53 and p16 pathways 11
2.4 SASP dynamic regulation in OIS 12
2.5 Pioneer TF chromatin binding 14
2.6 TF chromatin binding properties 17
2.7 Network inference validation 18
2.8 A hierarchical TF network 21
2.9 Edge overlap for networks relative to distinct transcriptomic profiles 23
2.10 SnCs TF perturbation transcriptional response 27

3.1 Molecular causes of CS 37
3.2 CS metabolic reprogramming 40
3.3 Metabolomics studies pipeline overview 42
3.4 Batch correction of MS data 43
3.5 Metabolite identification confidence levels 44
3.6 The CS metabolome is dynamic 47
3.7 Integrative visualization of CS metabolome 49
3.8 BC methods benchmark 51
3.9 RAS signaling pathways 55

4.1 Waddington epigenetic landscape 64
4.2 Artificial GRN 64
4.3 Computational techniques for network modelling 65
4.4 SINDy validation for known dynamic systems 69
4.5 Aggregation of genes co-expressed in all experimental conditions 71
4.6 Inferred model structure 72

4.7 Model simulation for OIS gene expression program following RAS induction 72
4.8 Model robustness to noise 73
4.9 Model simulation for TF KD relative to the training datasets . 75
4.10 Model validation for JUN and RELA KD at day 6 77
4.11 Hidden sources analysis and partial simulation 79
4.12 Optimization methods benchmark 81
4.13 Single-cell RNA velocity 86

CHAPTER 1

Introduction

Death is the only certainty of life. It has been contemplated by humanity since its early days, and inspired societies and individuals for millennia to find means to slow the aging process and prolong lifespan. The alchemist Ko Hung, who lived during the 3rd century, speculated that ingesting long-lasting substances would lead to a prolonged life and describes the "benefits" of consuming gold, jade and even mercury (Davis and Kuo-fu, 1941; Wallis et al., 2005). Gunpowder, which was later repurposed to the production of fireworks and weapons, was invented with the aim of extending lifespan of its consumers and its name means *fire medicine* (Andrade, 2016). In Europe, Middle Age alchemists aspired to synthesize the *Philosopher's Stone*, which was meant to lend immortality to its users (Augustyn et al., 2009).

Remarkably, McCay et al. (1935) showed that mice fed with a reduced calorie diet live longer than mice fed *ad libitum*, thus providing the first empiric evidence that lifespan can be extended in animals. Five decades later, Friedman and Johnson (1988) genetically modified the worm *Caenorhabditis elegans* (*C. elegans*) showing that mutations in the age-1 gene increased mean lifespan to 25 days, *i.e.*, 65 % more than the mean lifespan of 15 days observed in wild type worms. In 2008, Ayyadevara et al. discovered that second-generation homozygous age-1 mutants live between 145 and 190 days, a ten-fold increase in mean lifespan.

Further studies identified lifespan extending interventions not only in *C. elegans* but also in other invertebrates and vertebrates, including mammals (Mattison et al., 2017; Kenyon, 2010; Harrison et al., 2009; Grandison et al., 2009). Besides extending animal lifespan, these genetic interventions also delayed the onset of several age-related pathologies (ARPs), opening new possibilities for extending longevity in humans. However, it became clear from those studies that aging is a highly complex, multifactorial phenomenon that can be identified and have been categorized by nine molecular and cellular processes conserved during mammalian evolution, known as the "Hallmarks of aging": (1) genomic instability, (2) telomere attrition, (3) epigenetic alterations, (4) loss of proteostasis, (5) deregulated nutrient sensing, (6) mitochondrial dysfunction, (7) cellular senescence, (8) stem cell exhaustion, and (9) altered intercellular communication (López-Otín et al., 2013).

Cellular senescence (CS), a response to diverse forms of cellular stresses and characterized by a stable cell cycle arrest (SAGA) and a secretory phenotype (SASP), integrates all aging hallmarks at the molecular and cellular level and has been associated to several ARPs, including cataracts, sarcopenia, Alzheimer's disease and cancer (Gorgoulis et al., 2019; McHugh and Gil, 2018; Muñoz-Espín and Serrano, 2014). As represented in Figure 1.1, CS arises as consequence of the distinct forms of molecular and cellular damage characteristic of aging, causing a reduction in tissue regeneration and fitness,

and overall organismal health. Due to its central role in aging, a deep understanding of this process is essential to successfully develop senescence-oriented therapies to trat ARPs and decelerate aging.

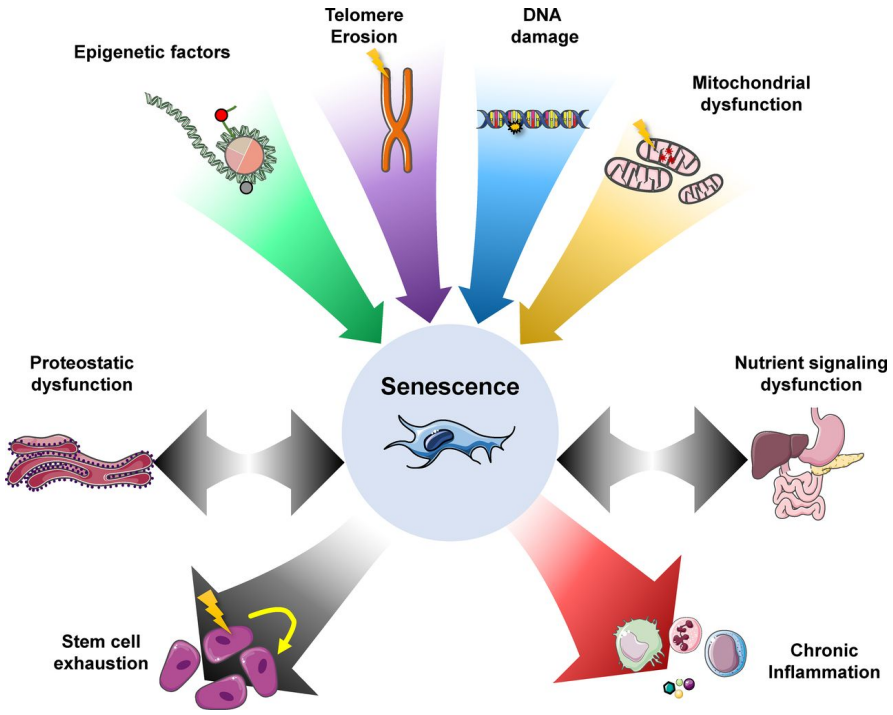


Figure 1.1: CS integrates the hallmarks of aging. CS is a response to cellular stress, including DNA damage, telomere erosion, epigenetic perturbation, as well as mitochondrial and proteostatic dysfunction. Ultimately, CS increases inflammation and stem cell exhaustion (McHugh and Gil, 2018). Reproduced from the article with permission - license number 1011829

1.1 Cellular senescence

CS is a damage response characterized by a stable cell cycle arrest, acting thus as potent tumor suppressor (McHugh and Gil, 2018; Soto-Gamez and Demaria, 2017; Muñoz-Espín and Serrano, 2014; López-Otín et al., 2013; Campisi and Fagagna, 2007). It was first described by Hayflick and Moorhead (1961) as a limitation in the maximum number of passages of primary fibroblasts in culture, in a process that was later explained by the loss of telomere integrity at each replication cycle (Allsopp et al., 1992). Aside from loss of telomere integrity, CS can be induced also by a variety of other stressors (McHugh

and Gil, 2018; Muñoz-Espín and Serrano, 2014; Campisi and Fagagna, 2007), such as non-telomeric DNA damage caused by high oxygen levels (Parrinello et al., 2003) or gamma radiation (Di Leonardo et al., 1994); hyperactivation of oncogenes (Serrano et al., 1997); decline in autophagic activity, leading to metabolic dysfunctions (García-Prat et al., 2016); and even the activation of pathways triggered by neighboring senescent cells, a process denominated *paracrine senescence* (Nelson et al., 2018; Hubackova et al., 2012; Acosta et al., 2013).

The senescence-associated growth arrest (SAGA) is mediated by the tumor suppressor retinoblastoma protein (RB), the activity of which is controlled by the tumor suppressors cyclin-dependent kinase inhibitor 2A (p16) and tumor protein 53 (p53) (Muñoz-Espín and Serrano, 2014; Rayess et al., 2012; Chicas et al., 2010; Serrano et al., 1997). Senescent cells (SnC) also release multiple proinflammatory molecules, including IL1, IL6, IL8, constituting the so-called senescence-associated secretory phenotype (SASP) (Nelson et al., 2018; Hubackova et al., 2012). The mechanisms regulating the SAGA and the SASP will be discussed in detail in section 2.1. Additional phenotypic changes associated with CS phenotype are apoptosis resistance, chromatin remodeling, reduced autophagy as well as transcriptional and metabolic reprogramming (McHugh and Gil, 2018; Parry and Narita, 2016; Wiley and Campisi, 2016; Muñoz-Espín and Serrano, 2014). However, CS response is as diverse as the nature of its triggers, and no universal bio-marker specific to SnC has been discovered to-date, making SnC detection a challenge, especially *in vivo*, requiring the measurement of multiple markers (Gorgoulis et al., 2019; Muñoz-Espín and Serrano, 2014).

SAGA and SASP are both beneficial and detrimental for tissue homeostasis. For example, the SAGA blocks proliferation of damaged cells, in particular pre-cancerous cells, and malignant transformation requires a by-pass of CS (Faget et al., 2019; Hoare and Narita, 2018). In addition, CS plays key roles in placental structure (Gal et al., 2019), embryonic development (Villiard et al., 2017; Storer et al., 2013) and tissue regeneration (Da Silva-Álvarez et al., 2019; Yun et al., 2015). Under normal physiological conditions, SnCs are eliminated by the immune system through SASP signaling. However, this immune-surveillance mechanism starts to malfunction as we age, leading to an accumulation of SnCs and a chronic , sterile inflammation, that can induce proliferation and metastasis of neighboring malignant cells (Faget et al., 2019; Hoare and Narita, 2018) and is associated with the onset of several ARPs, such as stroke (Fukazawa et al., 2007), cataracts (He et al., 2019), Alzheimer's disease (Bhat et al., 2012) and osteoarthritis (Martin et al., 2004).

Baker et al. (2011) showed that progeroid mice engineered with an inducible system of SnC clearance presented a delayed onset of several age-

related pathologies, such as cataracts, sarcopenia, osteoporosis, fat loss and cardiomyopathy. A more recent study from the same lab showed that clearance of senescent cells in old mice extends lifespan by up to 25 % (Baker et al., 2016). Those results formed the basis for the identification of so-called *senolytics*, molecules that are able to kill SnCs (Fuhrmann-Stroissnigg et al., 2017; Zhu et al., 2016; Zhu et al., 2015; Van Deursen, 2014). Indeed, the first clinical trials of senolytics in humans took place in the treatment of idiopathic pulmonary fibrosis, alleviating patient physical dysfunction, and osteoarthritis (Justice et al., 2019; Unity Biotechnology, 2019).

The recent advances in OMICS technologies allowed scientists to perform unbiased analysis, generating a large amount of data for one experiment and revealing a complexity higher than expected. For instance, a multi-omics analysis performed in our laboratory shows that CS is a process involving more than 4000 differentially expressed genes, being regulated by a changing epigenetic landscape shaped by the collective activity of, at least, 300 transcription factors (TFs) (Martínez-Zamudio et al., 2019). Despite the challenge that analyzing and integrating this data presents, it also provides an opportunity to explore the multitude and complexity of molecular interactions and identify a plethora of agents that can be therapeutically targeted, making this an exciting time for the field.

1.2 Thesis overview

This work presents an unprecedented and comprehensive analysis of the epigenome, transcriptome and metabolome of cells undergoing CS.

In chapter 2 I introduce the dynamic network of TFs binding to chromatin, regulating the gene expression program that drives CS. We observe that Activator protein 1 (AP1) family members to be already bound to the chromatin before CS stimulation and are thus situated at the top of the hierarchy preceding all other TFs. This suggests those TFs shape chromatin and prime the transcriptomic CS response. Additionally, TFs following AP1 constitute an intricate hierarchical network that differs even at enhancers of genes with similar expression profiles, showing that co-expressed genes can be regulated by distinct agents.

Chapter 3 contains a broad analysis of the CS metabolome. Specifically, I investigated differences and commonalities in metabolite levels in SnC induced by distinct inducers, namely, RAS, RAF, DNA damage (DD) and proliferative exhaustion. I identified alpha-keto glutarate (αKG), among others, as one metabolite differentially produced in SnC. αKG is a substrate for chromatin modification enzymes and a known onco-metabolite that modulates gene

expression and has the potential to be targeted in therapeutic settings.

In chapter 4, I describe the development of a mathematical model for oncogene-induced senescence (OIS) based on time-resolved transcriptome datasets acquired in CS cells, including the knock-down (KD) of key TFs for its regulation. The model contains 565 differential equations describing the behavior of the unique time profiles observed experimentally. In order to account for the high number of coefficients to be inferred, we employed a sparse approach running on a high performance computing (HPC) environment. Overall, the resulting simulations time profiles correlate with the training dataset, and the model is able to predict for conditions not included in the training data in some cases.

In conclusion, my integrative analysis provides a deeper understanding of CS mechanisms and highlights previously unknown potential vulnerabilities of senescent cells that may be exploited to treat cancer and ARPs, thus promoting a longer *healthspan* - the portion of our lives we are healthy, and not just alive.

CHAPTER 2

A hierarchical transcription factor network governs cellular senescence gene expression

2.1 Introduction

Cellular senescence (CS), as mentioned in Chapter 1, is a cell fate with a complex phenotype, characterized by a durable cell cycle arrest, release of inflammatory cytokines, chromatin remodeling, metabolic and transcriptomic reprogramming and apoptosis resistance (Gorgoulis et al., 2019; McHugh and Gil, 2018; Muñoz-Espín and Serrano, 2014; Campisi and Fagagna, 2007).

In order to characterize the dynamic transcriptomic and epigenomic changes shaping CS, our laboratory (Martínez-Zamudio et al., 2019) performed time course experiments in Wistar Institute 38 (WI38) fibroblasts undergoing RAS-OIS, collecting transcriptome, DNA accessibility and histone modification data over the course of six days after RAS OIS. Briefly, we found a dynamic epigenome with the temporal (in)activation of multiple enhancers, regulated by the collective activity of a variety of transcription factors (TFs). Gene expression changes followed seven distinct temporal profiles, as shown in Figure 2.1A.

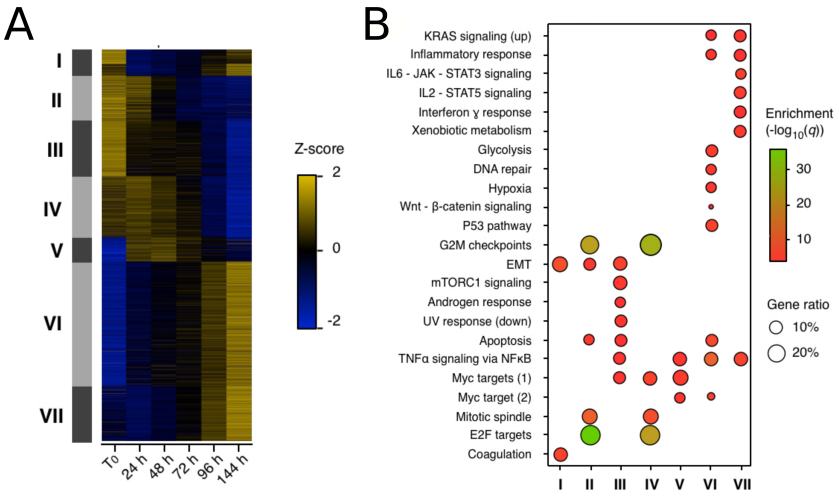


Figure 2.1: Transcriptional changes in CS (Martínez-Zamudio et al., 2019). [A] Heatmap depicting the dynamic expression level transitions during CS onset. Rows represent genes, clustered in seven co-expression modules, and columns represent time. [B] Pathway enrichment analysis for each of the observed gene modules.

To decipher the rules underlying gene expression changes executed by TF chromatin binding, I used *in silico* TF footprinting. These relationships constitute a hierarchical network, where-in TFs termed *pioneers*, able to interact with binding sites in closed chromatin, subsequently recruit so-called

settler and *migrant* TFs that will modulate gene expression as a response to senescence inducers. I generated one network for each of the seven identified modules of co-expressed genes, and here I discuss the results obtained for two gene modules II and VI, enriched for pathways associated with the most prominent phenotypes of SnCs: senescence-associated growth arrest (SAGA) and senescence-associated secretory phenotype (SASP) (Figure 2.1B).

2.1.1 Senescence-associated growth arrest

The cell cycle can be divided into four stages, G1, S, G2 and M in chronological order (Johnson and Walker, 1999). Phase M, which stands for *mitosis*, is the phase in which the one cell gives rise to two daughter cells. Their genetic material was synthesized at phase S, where the mother's cell DNA is replicated. The phases G1 and G2 are called *gap phases*, where the cell integrates growth factors signals and proceeds under favorable conditions to the S or M phases, respectively. The decisions to cease G1 or G2 progression are known as the *G1 and G2 checkpoints* and are regulated by CDK-cyclin complexes (Malumbres and Barbacid, 2009; Johnson and Walker, 1999). Cyclins, which were first thought to oscillate during distinct phases of the cell cycle and later discovered to respond to external signals such as growth factors (Johnson and Walker, 1999), bind to constitutively expressed Cyclin-dependent kinases (CDKs), activating signaling cascades that causes cell cycle progression.

At G1 phase, cyclin D becomes expressed in the presence of mitogenic signals (Malumbres and Barbacid, 2009), binding to CDK4 and CDK6. These complexes phosphorylate RB, a tumor suppressor that, in its hypophosphorylated form, is bound to the proto-oncogene E2F (Malumbres and Barbacid, 2009; Trimarchi and Lees, 2002; Johnson and Walker, 1999). When not bound to RB, E2F recruits the transcriptional machinery and activates the expression of a wide variety of genes that will regulate DNA replication and cell division (Figure 2.2). Hypophosphorylated RB can act as a transcriptional repressor by recruiting epigenetic regulators, such as histone deacetylases (HDACs) and the histone methyltransferase SUV39H1, to E2F target genes.

In SnCs, SAGA is mediated by cyclin-dependent kinase inhibitor 2A (p16) and tumor protein 53 (p53), which inhibit CDK activity and impair RB phosphorylation, thus preventing the cell from passing the G1 checkpoint, as discussed in details in section 3.1.1 (Gorgoulis et al., 2019; Muñoz-Espín and Serrano, 2014; Narita et al., 2003; Johnson and Walker, 1999).

The p16 gene is located at the INK4a/ARF locus, which contains several genes that control cell proliferation, is de-repressed as organisms age and frequently harbors mutations in several cancer types (López-Otín et al., 2013; Liu et al., 2009; Krishnamurthy et al., 2004). p16 directly binds to

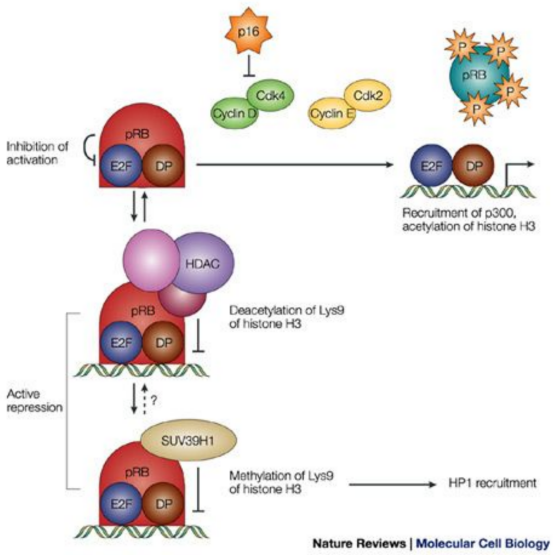


Figure 2.2: E2F inhibition by RB (Trimarchi and Lees, 2002). See text for details.

CDK4 and CDK6, keeping RB in a hypophosphorylated state, thus bound to and inactivating E2F (Sherr et al., 2016; Kim and Sharpless, 2006). The INK4a/ARF locus is silenced by H3K27me3 marks, maintained by polycomb-group proteins (PcG) (Rayess et al., 2012; Bracken et al., 2007). Upon CS, the demethylase Jumonji Domain-Containing Protein 3 (JMJ3) is expressed, activating p16 transcription (Agger et al., 2009; Barradas et al., 2009). Similarly to RB, p16 activation is essential for the establishment of senescence-associated heterochromatin focis (SAHFs), which have been shown to repress E2F targets, among other genes (Narita et al., 2003). Surprisingly, SAHF production involves a spatial chromatin reorganization and is not impacted when histone methylases, such as Jumonji Domain Containing 2D (JMJD2D), are inhibited (Parry and Narita, 2016; Chandra et al., 2012).

The p53 tumor suppressor gene is activated in response to DNA damage and therefore known as one of the main *guardians of the genome* (Lane, 1992). The mechanisms involved in its activation are described in section 3.1.1 and lead to the transcription of the CDK inhibitor p21 (Fagagna, 2008; Herbig et al., 2004). p21, encoded by the CDKN1A gene, forms a complex with RB and E2F, blocking CDK2 from phosphorylating RB and repressing E2F targets expression (Afshari et al., 1996; Dynlacht et al., 1994). As shown in Figure 2.3, p53 is inhibited by Human Homolog Of Mouse Double Minute 2 (HDM2) (Campisi and Fagagna, 2007; Lane, 1992). HDM2 inhibits p53 activity both by blocking its binding site to other proteins associated to

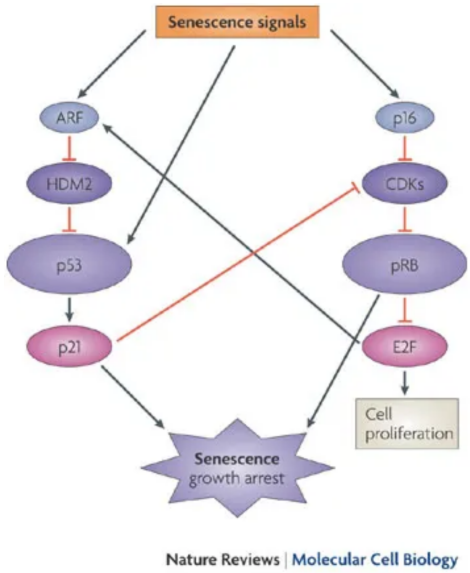


Figure 2.3: Cell cycle arrest induced by the p53 and p16 pathways (Campisi and Fagagna, 2007). See text for details.

transcription and by targeting it to ubiquitination and degradation (Juven-Gershon and Oren, 1999). This activity is blocked by the tumor suppressor ARF, with the formation of ARF/HDM2/p53 complexes. Interestingly, the ARF gene promoter contains E2F binding sites and is upregulated upon E2F overexpression (Dimri et al., 2000; Juven-Gershon and Oren, 1999), characterizing an incoherent feedforward loop regulating HDM2 and p53 activity. Moreover, p53 function is post-translationally modulated by histone acetyltransferases, being stimulated by p300 (Gu and Roeder, 1997) and inhibited by SIRT2 (Langley et al., 2002).

2.1.2 Senescence-associated secretory phenotype

The senescence-associated secretory phenotype (SASP) consists of a plethora of molecules secreted by SnCs, including cytokines, growth factors, pro-inflammatory molecules and lipids (Gorgoulis et al., 2019; Chan and Narita, 2019; Lopes-Paciencia et al., 2019; Coppé et al., 2010). The exact composition of the SASP is dependent on the senescence inducer, and can mediate multiple physiological effects at tissue and organism level, such as recruiting the immune system for clearance of pre-cancerous lesions (Chien et al., 2011), tissue regeneration and embryo patterning (Storer et al., 2013). However, it can also exert pathological responses, such as inflammation and tumorigenesis (Coppé et al., 2010). This process can also produce a by-stander effect by

inducing CS in neighboring cells (Nelson et al., 2018; Hubackova et al., 2012). The SASP is driven by several regulators, with Nuclear Factor κB (NF-κB) and CCAAT Enhancer Binding Protein β (CEBPβ) being the most described TFs modulating this phenomenon (Lopes-Paciencia et al., 2019; Ito et al., 2017; Muñoz-Espín and Serrano, 2014). A simplified model of their regulation is shown in Figure 2.4A.

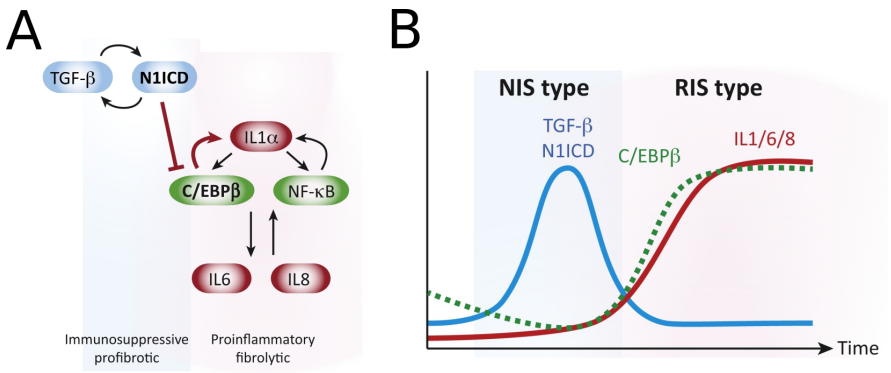


Figure 2.4: SASP dynamic regulation in OIS [A] Simplified network describing SASP regulation. [B] At the early phase of OIS, N1ICD is upregulated, inhibiting CEBPβ activity and promoting an immunosuppressive and profibrotic SASP. As N1ICD is downregulated, CEBPβ levels rise. CEBPβ activation increases the expression of interleukins and metalloproteases, leading to a proinflammatory and fibrolytic SASP. Adapted from (Ito et al., 2017).

NF-κB is a transcription factor family involved in inflammatory responses (Lopes-Paciencia et al., 2019; Liu et al., 2017). Its dynamics has been studied quantitatively and inspired several mathematical models (Benary and Wolf, 2019; Inoue et al., 2016; Lipniacki et al., 2004; Hoffmann et al., 2002). NF-κB activation implicates either a canonical or non-canonical pathway (Liu et al., 2017; Yılmaz et al., 2014; Ohanna et al., 2011). In the canonical pathway, which responds to a variety of stimuli including cytokines and Tumor Necrosis Factor (TNF) receptors, NF-κB is activated by the degradation of Inhibitor Of Nuclear Factor Kappa B Kinase (IKK), which is bound to NF-κB in the cytoplasm. When released from IKK, NF-κB family members can form homo- and heterodimers and translocate to the nucleus, activating the transcription of multiple genes involved in inflammatory response.

In CS, NF-κB is upregulated both by DNA damage response (DDR) (Kang et al., 2015) or by reactive oxygen species (ROS) (Nelson et al., 2018; Ohanna et al., 2011). DDR activates NF-κB by suppressing the degradation of Transcription Factor GATA-4 (GATA4), in a process mediated by IL1α (Kang et al., 2015). IL1α is also activated by mamalian target of Rapamycin

(mTOR), and rapamycin was shown to reduce NF-κB levels in human tumors (Laberge et al., 2015). ROS signaling upregulates NF-κB through P38 mitogen-activated protein kinases (P38MAPK) (Nelson et al., 2018), which also mediates p21 increase under Transforming Growth Factor Beta 1 (TGFβ) activation (Frippiat et al., 2002), constituting a mechanistic link between SASP and SAGA. Additionally, NF-κB is activated in the presence of cytosolic DNA, responding to the cGAS–STING pathway (Dou et al., 2017). When active, NF-κB binds to the chromatin and activates the expression of several interleukins, including Tumor necrosis factor alpha (TNFα), IL6, IL8, IL1β and IL1α, the latter characterizing a positive feedback loop that reinforces CS and promotes a DDR in bystander cells (Nelson et al., 2018; Dou et al., 2017; Kang et al., 2015; Hubackova et al., 2012).

CEBPβ cooperates with NF-κB in activating a variety of SASP components, including IL6, IL8, IL1β and IL1α (Lopes-Paciencia et al., 2019; Ito et al., 2017; Kuilman et al., 2008). In RAS-induced senescence, CEBPβ is activated at later CS stages, being initially inhibited by Notch signaling (Ito et al., 2017). As shown in Figure 2.4, early OIS produces a immunosuppressive SASP, regulated by Notch and TGFβ. As CS progresses, Notch is cleaved, derepressing CEBPβ and leading to a proinflammatory SASP (Ito et al., 2017; Hoare et al., 2016).

2.1.3 Transcription factor networks

The cell fate decision that characterizes CS is orchestrated by multiple TFs, as illustrated in sections 2.1.1 and 2.1.2. TF activity is ubiquitous in cellular biology, mediating processes including reproduction, development, metabolism, differentiation, stimulus response and disease in all known eukaryote organisms (Voss and Hager, 2014; Vaquerizas et al., 2009). TFs are able to bind to DNA due to the presence of specialized motifs, such as helix-turn-helix, zinc-fingers and zipper-type binding sites (Luscombe et al., 2000).

Non-coding DNA regions that modulate gene expression are called *cis-regulatory regions*, including, *e.g.*, *promoters* and *enhancers* (Schoenfelder and Fraser, 2019; Li et al., 2015; Wittkopp and Kalay, 2012). Promoters are usually located at the *transcription start site* of genes and correspond to the genomic location where transcription begins (Lenhard et al., 2012; Yuh et al., 1998). Therefore, gene expression levels are regulated by the affinity between the transcriptional machinery and the promoter, which is modulated by the TFs bound to those genomic positions (Yuh et al., 1998), among other factors. Enhancers regulate the transcriptional response by directly binding to their target genes, and they may be located hundreds of kilobase pairs from their targets (Schoenfelder and Fraser, 2019; Li et al., 2015;

Wittkopp and Kalay, 2012). Promoter-enhancer contacts are mediated by TFs and architectural proteins, including CTCF and YY1. Those cis-regulatory regions are characterized by specific histone modifications, where promoters correspond to regions enriched in H3K4me3 and enhancers, in H3K4me1 (Lenhard et al., 2012; Zentner et al., 2011). Additionally, active promoters and enhancers present H3K27Ac marks, while PcG-repressed chromatin is characterized by H3K27me3 (Schoenfelder and Fraser, 2019).

The principles underlying TF-mediated regulation of gene expression are still controversial and there are several theories describing how they access heterochromatin in order to activate compacted genes (Voss and Hager, 2014). The *pioneer proteins* model states that certain TFs have the ability to interact with closed chromatin and recruit chromatin modifiers and other TFs, as illustrated in Figure 2.5 (Mayran and Drouin, 2018; Zaret and Carroll, 2011; Voss and Hager, 2014). By shaping the epigenetic landscape, TF composition define cell identity, determining which genes are activated and the functions they perform (Mayran and Drouin, 2018; Zaret and Carroll, 2011; Vaquerizas et al., 2009). TFs can also respond to internal cues by post-translational modification (PTM), modulating their affinity both to chromatin and other TFs (Cossec et al., 2018).

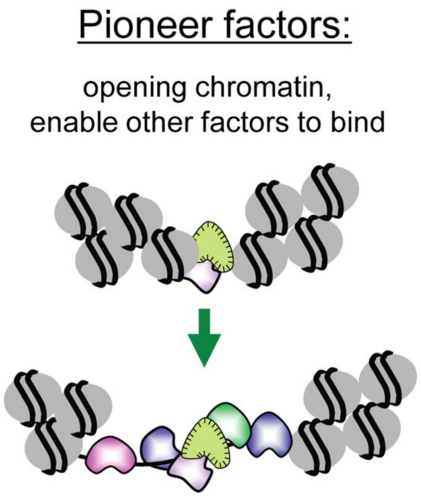


Figure 2.5: Pioneer TF chromatin binding Pioneer TFs bind to inaccessible chromatin regions, displacing nucleosomes and recruiting settler and migrant TFs. Adapted from (Zaret and Carroll, 2011).

The collective activity of TFs requires systemic approaches in order to determine transcriptional output as a function of chromatin occupancy. Jothi

et al. (2009) used available chromatin immunoprecipitation sequencing (ChIP-seq) yeast data to characterize its regulatory TF network. Goode et al. (2016) combined RNA sequencing (RNA-seq), histone marks and TF ChIP-seq data in order to determine how TFs impact on gene expression levels of hematopoietic differentiation. Wapinski et al. (2017) and Ramirez et al. (2017) used a similar approach, employing *in silico* approaches (Sherwood et al., 2014; Piper et al., 2013) to infer TF chromatin binding at cis-regulatory regions in fibroblasts trans-differentiating into neurons and myeloid differentiation, respectively. Garber et al. (2012) applied a distinct paradigm on TF networks, assessing TF chromatin binding over time in order to characterize pioneer TFs and the following recruited TFs during dendritic cells pathogen activation.

Currently, no comprehensive multidimensional profiling study has been performed for CS, a cell fate decision that defies our definition of cellular identity. In this chapter, I combine epigenome, DNA accessibility and gene expression time course data gathered by Martínez-Zamudio et al. (2019) in order to characterize the TF network controlling CS in a integrative manner.

2.2 Results

2.2.1 The TF hierarchy driving the senescence transcriptional program

To determine how TFs collectively control CS gene expression dynamics, I used *in silico* time-resolved TF footprinting performed in our laboratory (Martínez-Zamudio et al., 2019). Briefly, our laboratory collected time resolved datasets for transcriptome, epigenome (histone ChIP-seq) and chromatin accessibility (ATAC-seq) on human WI38 lung fibroblasts undergoing RAS-OIS. We identified seven modules of co-expressed genes corresponding to the CS program. Genes pertaining to each expression module were associated to their closest active enhancer, identified as genomic regions containing H3K4me1 and H3K27Ac. We used an *in silico* approach denominated Protein Interaction Quantification (PIQ) to infer TF chromatin binding based on DNA accessibility data (Sherwood et al., 2014). The correlation between chromatin accessibility and binding probability was used to classify each TF as *pioneer*, *settler* or *migrant* (Martínez-Zamudio et al., 2019; Sherwood et al., 2014).

With the purpose of characterizing the TF chromatin binding patterns, I calculated a set of metrics as performed by Garber et al. (2012) for 310 TFs retrieved from the JASPAR database (Khan et al., 2017). Those metrics consist of (a) total number of binding sites; (b) a dynamicity index, proportional to the number of new regions bound or left by a TF during the time course; (c) the ratio of regions already bound before senescence induction and (d) the percentage of enhancers in the total number of bound regions Garber et al. (2012). Figure 2.6A depicts a Principal Component Analysis (PCA) containing those covariates for each TF (Abdi and Williams, 2010). This plot shows that TFs with same classification share similar binding characteristics. Pioneers tend to bind to a higher number of regions, and most of those regions were already populated before RAS overexpression. The AP1 family members are the pioneers with highest percentage of binding instances in enhancers. Moreover, migrant TFs tend to be more specific, *i.e.*, bind to fewer regions and are more dynamic.

Figure 2.6B shows the dynamicity index and the number of bound regions calculated for each TF for each module of co-expressed genes. Each heatmap row corresponds to one TF, and TFs are ordered based on their ratio occupancy/dynamicity, with the ones having highest number of bound regions at the top and most dynamic at the bottom. The number of bound regions is approximately constant for all gene modules, while pioneers have a low

dynamicity index for all gene modules. Migrants, however, are more dynamic and display heterogeneity in their dynamicity index for distinct gene modules. In order to evaluate the enrichment of each TF classification in the list ranked by occupancy/dynamicity ratio, we performed a Set Enrichment Analysis (SEA) for pioneers (Figure 2.6C), settlers (Figure 2.6D) and migrants (Figure 2.6E). Pioneers and settlers are enriched at the top of the list, while migrants are enriched at its bottom.

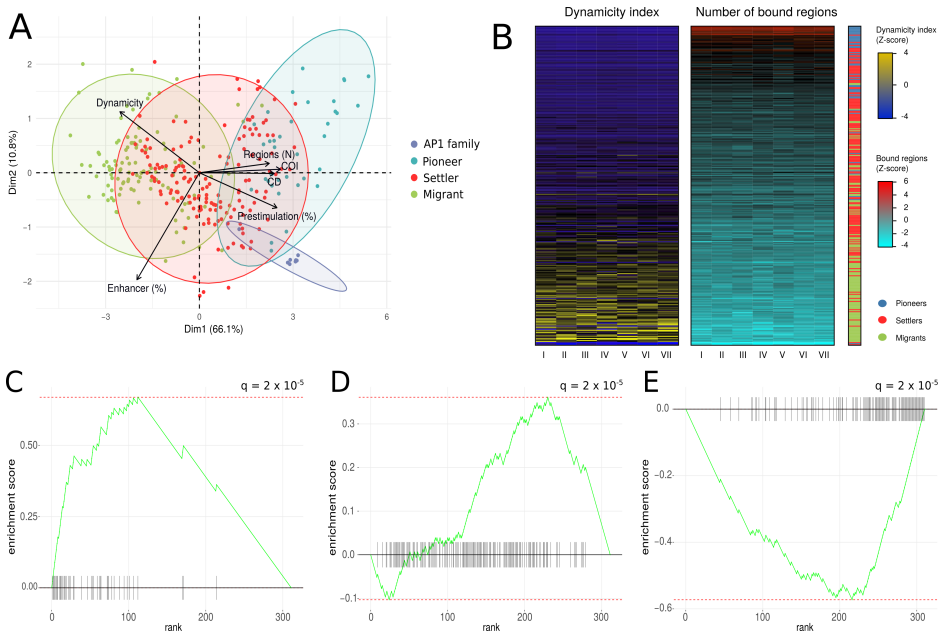


Figure 2.6: TF chromatin binding properties. [A] PCA depicting six TF chromatin binding characteristics. See text for details. [B] Heatmaps displaying TF dynamicity, chromatin coverage and classification, stratified by each gene expression module. Each line corresponds to a TF, and the dendrograms were ordered based on the ratio between the number of bound regions (coverage) and the dynamicity of each TF. [C-E] SEA performed on the TF list ordered by coverage-dynamicity ratio, depicting the enrichment for pioneers [C] and settlers [D] at the top of the list, and migrant TFs at the bottom [E].

Aiming to identify the underlying logic that governs TF hierarchy in driving gene expression, we built hierarchy networks as described by Garber et al. (2012). Briefly, an edge represents a precedence relationship between two TFs, where a directed edge from TF X to TF Y means that Y bound to a region where X also bound at the time point or before. Each edge is weighted by the ratio of regions bound by Y that were already bound by X. Figure 2.8A shows the network built on ChIP-seq datasets published by

Garber et al. (2012), and Figure 2.8B shows the original network from the article. All the edges detected by our method are present in the original network, and I retrieved 88,9 % of the edges from the original network.

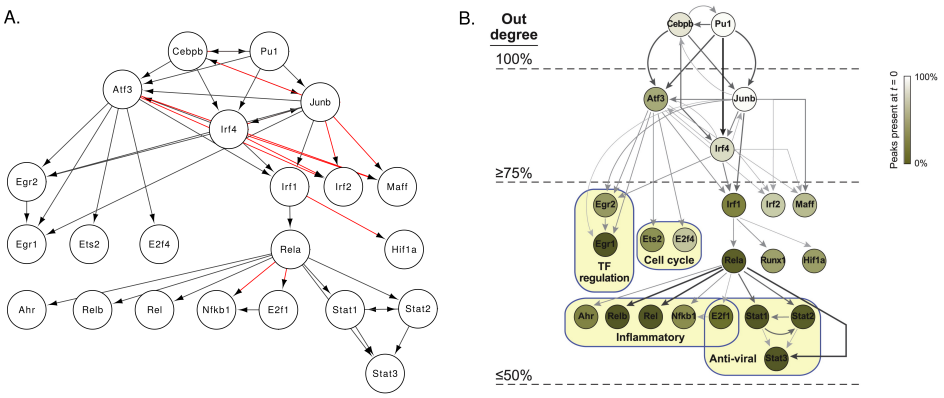


Figure 2.7: Network inference validation. [A] Validation network inferred from ChIP-seq data collected by Garber et al. (2012). [B] The original published network (reproduced from the article with permission - license number 4684440352557).

Figures 2.8A and D display the TF hierarchy network built for enhancers corresponding to genes in the transcriptional modules II and VI, respectively. As mentioned in section 2.1, module II is composed of several SAGA genes, downregulated after RAS stimulation, and module VI is enriched for SASP constituents. The AP1 family members are at the top of both networks, preceding all other TFs and binding to thousands of regions. The simplified networks (see section 2.5.7) contain 252 nodes and 474 edges or 250 nodes and 420 edges, respectively.

Figures 2.8B and E depict subnetworks focusing on the E2F and p53 families, which mediate the SAGA. Most TFs are present in both networks, such as the p53 family members tumor protein 63 (p63) and tumor protein 73 (p73), the cluster of TFs containing the E twenty-six Proto-Oncogene 1 (ETS1) family members, ETV6, ELF1, ELK4 and GABPα, and the Specificity Protein (SP) and Kruppel Like Factor (KLF) proteins following the Early Growth Response (EGR) family. However, the observed interactions between those TFs present some discrepancies between the distinct modules, such as the cobinding of p53 and MAF bZIP transcription factor G (MAFG) for genes at module VI and not at module II. In module VI, both E2F2 and 3 and E2F7 and 8 follow the AP1 family members when binding to the chromatin, while E2F2 and 3 seem to follow ETS1 in module II and E2F7 and 8 did not bind to the chromatin during CS onset.

Figure 2.8C and F portray the TF hierarchical subnetworks focusing on CEBPβ and NF-κB, which regulate SASP transcription. Despite both networks sharing most of their constituting TFs, the observed interactions are strikingly distinct. In module II, ETS1, MAFG and CEBPβ precede NFIL3 and TEF, while NF-κB family, which includes REL-associated protein A (RELA), follows only the AP1 family members. In module VI, NFIL3 also follows MAFG, while RELA binds to several positions already bound by ETS1. Additionally, NF-κB is preceded by EGR and SP factors.

As suggested by Figure 2.6, the pioneer TFs concentrate at the top of the network, while migrants are enriched at the bottom (Figures 2.8A-F). Figure 2.8G displays the percentage of incoming edges, stratified by the classification of the source TF, taken from all networks. Pioneer TFs mostly follow pioneers, while settler TFs follow both pioneers and settlers. Migrants succeed TFs from any classification, however the majority of the edges are originating from pioneers.

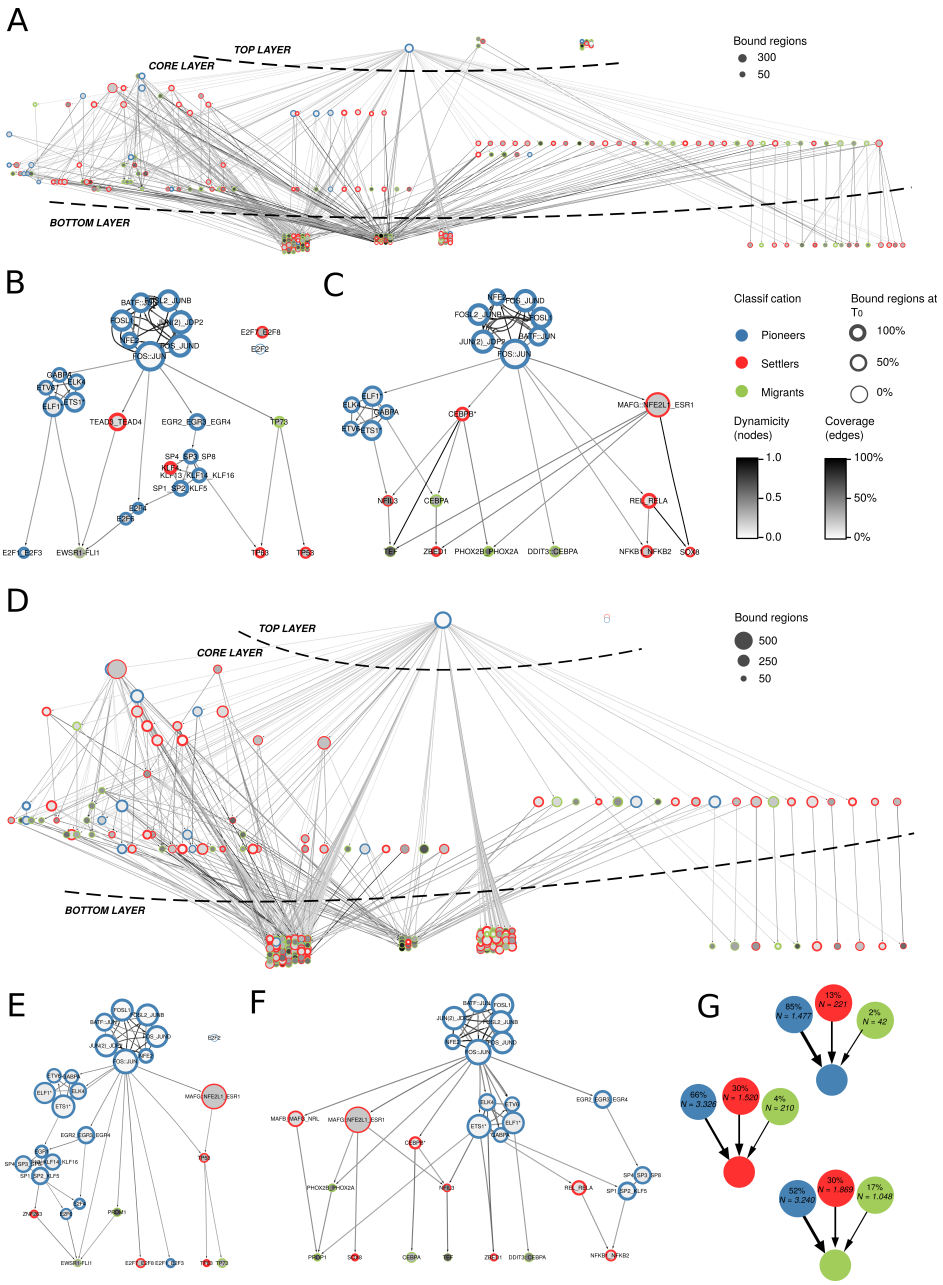

A
TOP LAYER
CORE LAYER
BOTTOM LAYER
Bound regions
300
50
B
C
Classif cation
Pioneers
Settlers
Migrants
Bound regions at T0
100%
50%
0%
Dynamicity (nodes)
1.0
0.5
0.0
Coverage (edges)
100%
50%
0%
D
TOP LAYER
CORE LAYER
BOTTOM LAYER
Bound regions
500
250
50
E
F
G
85% N = 1,477
13% N = 221
2% N = 42
66% N = 3,326
30% N = 1,520
4% N = 210
52% N = 3,243
30% N = 1,869
17% N = 1,048

Figure 2.8 ***(previous page)*****: A hierarchical TF network**. [A-F] TF chromatin binding hierarchy network. Nodes represent TFs, and an oriented edge from TF A to TF B means that at least 30 % of the regions bound by B were bound by A at the same time point or before. In order to simplify the visualization, we represent strongly connected components (SCC) as a single node and performed a transitive reduction (TR). Node color is based on the average dynamicity of the SCC members. Node border color indicates their classification as pioneer (blue), settler (red) or migrant (green). Node border thickness encodes the percentage of bound regions before RAS stimulation. Edge color was calculated accordingly to the relative coverage of the outgoing TF over the incoming TF. The network has three layers: top, core and bottom. Nodes in the top have no incoming edges and nodes in the bottom have no outgoing edges. The core layer comprises TFs that have both incoming and outgoing edges. [A] TF hierarchy network for genes in the cell cycle module. [B] Inset of panel A depicting a subset of the TFs that regulate SAGA. [C] Inset of panel A depicting a subset of the TFs that regulate SASP. [D] TF hierarchy network for genes in the SASP module. [E] Inset of panel D depicting a subset of the TFs that regulate SAGA. [F] Inset of panel D depicting a subset of the TFs that regulate SASP. [G] Ratio of incoming edges based on the classification of the TF source node. The relative and absolute number of edges corresponding to all seven modules are displayed inside the nodes, which are colored accordingly to TF classification as in previous panels. The thickness of links is proportional to the relative number of TF hierarchy edges connecting nodes with the corresponding classification.

Aiming to determine TF network properties that lead to distinct transcriptome dynamics, we compared the TF interactions observed for each gene module. Figure 2.9 shows a Chow-Ruskey diagram comprising all seven networks. A Chow-Ruskey diagram is a Venn diagram, where each region's area is proportional to the number of elements it contains (Chow and Ruskey, 2003). Each region refers to the set of edges of a TF network corresponding to a specific gene module. Panel A contains the number of interactions shared and specific for all networks. The following panels show edge overlaps for the (B) top, (C) middle and (D) bottom of the networks. The top of the networks contains a higher fraction of shared edges, while the majority of bottom edges are more specific. In conjunction with the heatmaps at Figure 2.6B, this data suggests that the migrants at the bottom of the network fine tune the gene expression.

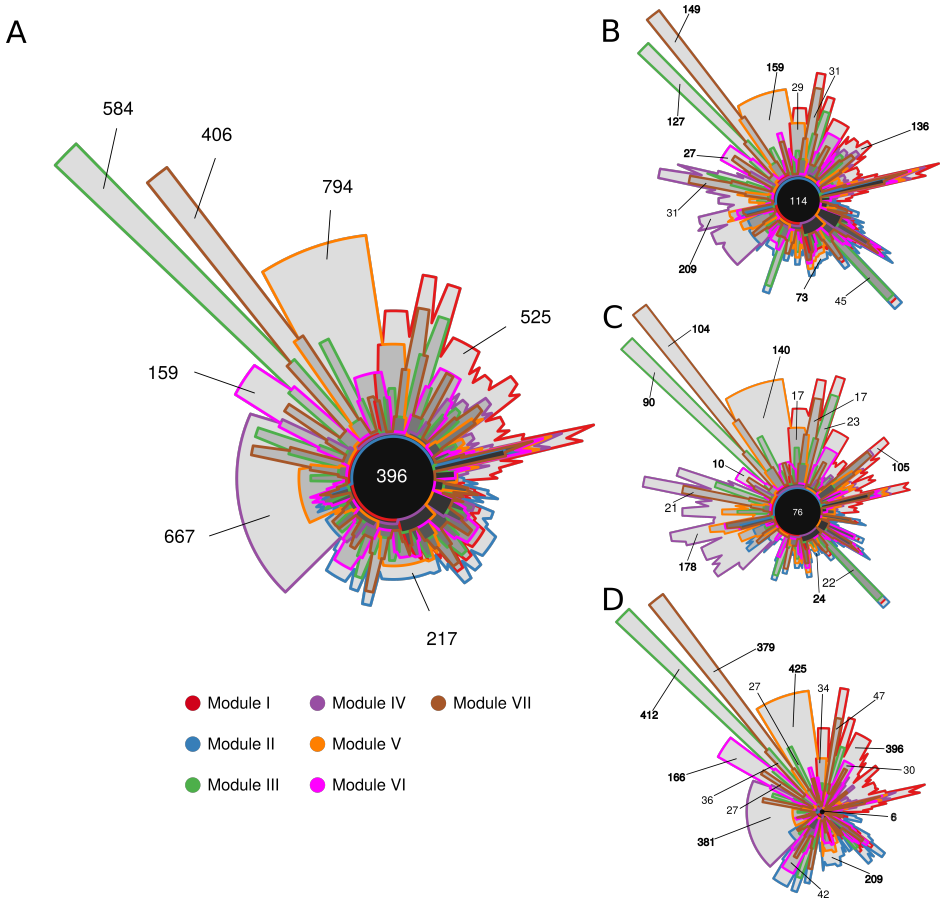


Figure 2.9: Edge overlap for networks relative to distinct transcriptomic profiles. [A] Euler diagram comprising the seven gene expression modules. Each region correspond to the set of edges belonging to the TF hierarchy networks for each gene module. An edge is shared between two modules if its source and target are connected in both networks. [B-D] Euler diagrams for edges with source at [B] the top of the network, [C] connecting nodes neither from top or the bottom of the network and [D] with target at the bottom of the network.

2.3 Discussion

By analyzing the temporal organization of TFs chromatin binding, I observed a complex set of interactions governing CS gene expression. As reported by Garber et al. (2012), the TF regulatory network is layered and hierarchically organized, with pioneer TFs preceding settlers and migrants (Figure 2.8G).

The PCA and the heatmap shown in Figures 2.6A and B depict a heterogeneous and dynamic TF activity, as described by Garber et al. (2012). Pioneer TFs bind to a high number of enhancers before CS stimulation by oncogenic RAS, suggesting a role in shaping and maintaining the fibroblast identity and pre-coding its ability to undergo senescence. Indeed, the AP1 family members, the most enriched TFs detected in enhancers of SnC, have been reported to be involved in fibroblasts activation during wound healing (Bergmann et al., 2018; Florin et al., 2004) and mediate androgen receptor signaling (Leach et al., 2017). The mechanisms through which pioneer TFs orchestrate gene expression include nucleosome displacement and recruitment of epigenetic modifiers (Mayran and Drouin, 2018; Radman-Livaja and Rando, 2010; Lascaris et al., 2000). Those modifications alter DNA accessibility, exposing binding sites for TFs termed *settlers* and *migrants* that will also regulate gene expression (Voss and Hager, 2014; Lomvardas and Thanos, 2002). As shown in Figure 2.6, settler and migrant TFs bind to fewer enhancers and exhibit a higher diversity accordingly to the transcriptional module they are bound to. Hence, I hypothesize that these TFs optimize transcriptional output in a gene unspecific fashion.

The AP1 family members are bound to chromatin before stimulation of WI38 fibroblasts into senescence, and the following TFs mostly bound to regions already populated by them (Figures 2.8A-F). This effect was observed in all transcriptional modules, suggesting that this family acts as a scaffold shaping the chromatin and recruits settler and migrant TFs. This finding provides an explanation for the diverse set of functions reported for this family, including cell proliferation, tumorigenesis, immune response, differentiation (Atsaves et al., 2019; Mirzaei et al., 2019; Karin et al., 1997) and, now, CS (Martínez-Zamudio et al., 2019), as their upregulation may expose binding sites for other TFs and their depletion might suppress TF recruitment to the chromatin.

I also observed a second tier of pioneer TFs preceding a considerable amount of TFs in the networks, consisting of the ETS1, and the EGR, SP and KLF families. As shown in Figures 2.8A-F, the ETS1 family members also bind to hundreds of cis-regulatory regions in all transcriptional modules. Additionally, ETS1 binds to some E2F targets in the cell cycle gene module

(module II, Figure 2.8B) and to RELA and NF-κB targets in the SASP-related gene module (module VI, Figure 2.8F). The EGR, SP and KLF families members precede E2F4 and 6 in both studied gene modules, with SP and KLF following both ETS1 and EGR (Figures 2.8B and E). However, these two classes of TF do not co-bind at the chromatin, meaning that they regulate distinct sets of genes sharing the same transcriptional profile. I observed the same pattern in module VI, where ETS1 is followed by E2F4 and 6 and NF-κB (Figures 2.8C and F).

MAFG is known to interact with the AP1 TF family (Kannan et al., 2012) and is classified as a *settler* TF due to a lower capacity of displacing nucleosomes and opening inaccessible chromatin (Martínez-Zamudio et al., 2019). MAFG binding activity follows AP1 family members and precedes other settler and migrant TFs (Figures 2.8A,C-F), including members of the p53 family and TFs following CEBPβ in module VI (Figures 2.8E,F).

TFs belonging to the same family, such as AP1, ETS1 or SP/KLF, tend to cluster together, forming dense sub-networks. Since edges represent a co-binding event at the same time point or later, this dense sub-networks mean those TFs consistently bind to the same regions of the genome at the same timepoint. From a topological perspective, they form a graph structure known as strongly connected component (SCC), where all nodes are reachable from each other (Tarjan, 1972).

Remarkably, I observed SASP-related TFs binding to genes in module II (Figure 2.8C) and cell cycle-related TFs binding to genes in module VI (Figure 2.8E). This finding shows that the same TF can produce a distinct effect on transcriptional output, yet the conditions determining this behavior still remain to be elucidated. Yuh et al. (1998) performed a series of elegantly designed experiments showing how distinct promoter segments of the sea urchin *Endo16* gene contribute to its expression, where TFs biding to some segments could act as amplifiers, switches or inhibitors, constituting an analog computing biochemical device. Therefore, genome sequence at cis-regulatory regions determines TF binding and its impact of gene expression. Additionally, this computation can be performed at other regulatory layers, including PTMs at the protein level. As mentioned before, RB hypophosphorylation prevents E2F chromatin binding and recruit epigenomic silencers at its target genes. Deacetylation of AP1 or NF-κB by SIRT1 also repress the expression of targets (Xie et al., 2013; Zhang et al., 2010; Yeung et al., 2004).

Since a similar set of TFs participate in the regulation of genes with distinct output, I compared their association in each network. The Euler diagram represented in Figure 2.9A reveals that most edges are either shared among the networks built for all seven identified gene modules or belonging to a single TF hierarchy network. As mentioned in the previous section, the

cause of this effect still needs to be addressed by further studies, with potential causes including the genome sequence of each enhancer recruiting a specific TF set or even PTMs that change TF affinity to its targets. Interestingly, the portion of interactions unique to each transcriptional module is higher at the bottom of the networks (Figure 2.9D), when compared to the TFs at the top of the network. This corroborates the hypothesis that late TFs fine-tune gene expression, producing the observed distinct temporal transcriptional patterns.

In order to validate the TF role in gene regulation, our laboratory conducted small interferring RNA (siRNA)-mediated experiments for three TFs at different levels of the network, namely Jun Proto-Oncogene (JUN), ETS1 and RELA, on SnC six days after induction, as explained in details in section 4.5.1 (Martínez-Zamudio et al., 2019). As shown in Figure 2.10A, targeting TFs at the top of the hierarchy disrupts a higher number of genes, while the effect of perturbing a TF at lower layers of the network produces smaller gene expression changes. The overlap of differentially expressed genes in each experiment suggests those TFs cooperate in regulating a subset of their targets. This result indicate that the generated hierarchy networks describe the TF interactions regulating CS transcription. Importantly, we could show that depletion of the AP1 family member JUN, at the top of the TF network, reverses the senescence clock (Martínez-Zamudio et al., 2019). As shown in Figure 2.10B, the transcriptional state of SnCs at day 6 after RAS-OIS induction submitted to JUN siRNA-mediated inhibition is similar to the state of cells 3 days after RAS over-expression.

2.4 Conclusions and Perspectives

In order to describe the molecular machinery controlling CS expression, I used time course epigenomic and transcriptomic data to depict co-binding TF interactions over time. My analysis revealed a layered hierarchical network, with TF activity increasing in dynamicity and specificity over time. Early binding activity is shared across gene modules and performed by AP1 *pioneer* TFs, able to displace nucleosomes and recruit *settlers* and *migrants*, that present a lower chromatin affinity and fine-tune gene expression (Knaap and Verrijzer, 2016; Voss and Hager, 2014).

Despite a comprehensive analysis of the TF landscape at chromatin, our analysis only covers 519 TFs currently registered at the JASPAR database (Khan et al., 2017). The number of TFs in humans is estimated between 1700 and 1900 distinct proteins (Vaquerizas et al., 2009). My developed methodology, associated to more comprehensive datasets, will be able to

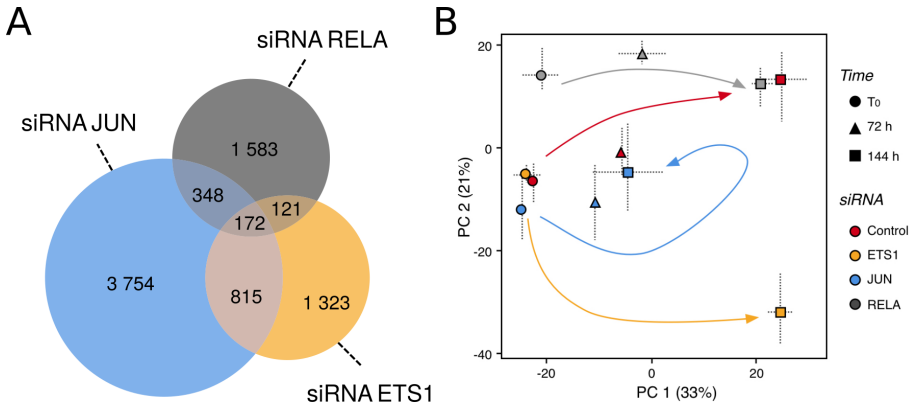


Figure 2.10: SnCs TF perturbation transcriptional response. [A] Venn diagram showing the number of target genes differentially expressed under TF inhibition by siRNA when compared to unperturbed SnC. [B] PCA depicting the transcriptome of cells undergoing CS (in red) and SnCs two days after siRNA-mediated TF inhibition of JUN, ETS1 and RELA. (Martínez-Zamudio et al., 2019).

provide a more comprehensive description of the hierarchical TF activity as public databases become more complete. Furthermore, it can also be applied in other studies where the order at which TF bind to the chromatin is relevant, such as embryonic development or cell differentiation.

This work describes the evolution of TF chromatin binding only at distal cis-regulatory regions, termed *enhancers*. We made this choice based on previous reports that gene expression changes correlate with a dynamic enhancer landscape, while promoter activity remains constant even in situations where transcriptional changes are observed (Goode et al., 2016; Tasdemir et al., 2016; Lara-Astiaso et al., 2014). Additionally, our time course epigenomic data also revealed a considerable increase in active enhancer regions during CS onset (Martínez-Zamudio et al., 2019). Still, no comprehensive analysis of the TF activity at promoter level has been performed in the context of CS as described for enhancers and may be a subject of future studies.

An interesting future application of this analysis involves the inference of the grammar underlying TF organization (Dunn, 2017; D'Ulizia et al., 2011). Martínez-Zamudio et al. (2019) determined the most frequent TF associations by applying a text-mining technique, integrating those results with the chromatin state changes in CS. Grammar inference consists of finding a set of rules that are able to derivate a sequence of tokens by the expansion of those rules (Chomsky, 1957). As we collect time course data on TF chromatin binding, we will be able to apply grammar inference techniques

in order to model how TF interact and the context dependency of those interactions. Furthermore, this knowledge may orient DNA sequence editing (Cong et al., 2013) therapies by targeting cis-regulatory regions and rewiring the gene regulatory network (GRN) logic, instead of the genome coding regions themselves (Sanz et al., 2017; Vaz-Drago et al., 2017; Diederichs et al., 2016).

This study provides a deeper understanding of how TFs govern CS gene expression, by highlighting the main factors shaping the chromatin in a time-dependent manner. Further studies on our proposed candidates will lead to new possibilities of senescence-modulating therapies in order to treat ARPs and cancer.

2.5 Materials and methods

2.5.1 Cell culture

WI38 fibroblasts were cultured as described by Martínez-Zamudio et al. (2019). Cells undergoing RAS-induced senescence were collected at six timepoints (0, 24, 48, 72, 96 and 144 h). Additionally, we harvested WI38 fibroblasts 0, 12, 24, 48, 72 and 96h after fethal bovine serum (FBS) withdrawal, leading to a quiescent state of reversible cell cycle arrest.

2.5.2 Gene expression analysis

RNA purification and analysis was performed as reported by Martínez-Zamudio et al. (2019). Briefly, the RNA from two replicates for each sample were purified and analyzed using Affymetrix Human Transcriptome Arrays 2.0. After preprocessing and normalization, genes with differential expression associated to both time and treatment (RAS induction or quiescence entry) were clustered using the Weighted Gene Correlation Analysis (WGCNA) tool (Langfelder and Horvath, 2008), resulting in seven modules of co-expressed genes.

2.5.3 Histone modification ChIP-seq and chromatin state differential analysis

Fibroblasts undergoing OIS after 0, 72 and 144h were collected for histone modification ChIP-seq and sequencing and data preprocessing was performed as previously described (Martínez-Zamudio et al., 2019). Two biological replicates were collected for each timepoint and chromatin immunoprecipitated using antibodies against H3K4me1, H3K4me3, H3K27Ac and H3K27me3 was sequenced. Preprocessed and normalized reads were used to classify genomic regions into distinct states, accordingly to the observed combination of histone modifications. Regions containing H3K4me1 were considered as putative enhancers, while regions containing H3K4me3 were classified as promoters. H3K27Ac is a mark of active cis-regulatory regions (enhancers or promoters), while H3K27me3 characterizes Polycomb repressed chromatin.

2.5.4 ATAC-seq

DNA accessibility data was collected, preprocessed and normalized for the six timepoints mentioned in section 2.5.1 of WI38 fibroblasts entering CS, as documented in Martínez-Zamudio et al. (2019).

2.5.5 Transcription factor footprinting

The binding of TFs to the chromatin was inferred *in silico* for a set of 310 TFs retrieved from the JASPAR database (Khan et al., 2017), using the PIQ tool (Sherwood et al., 2014) as performed by (Martínez-Zamudio et al., 2019).

2.5.6 Transcription factor chromatin binding properties

With the aim of characterizing the chromatin binding properties of each TF, I computed the dynamicity, the total number of bound regions, the fraction of bound regions at enhancers and the fraction of bound regions before stimulation.

Dynamicity

I quantified the dynamicity of a TF accordingly to the following expression:

$$d(A) = \frac{\sum_t \frac{n_t(A)}{TR_t}}{\sum_t \frac{t_t(A)}{TR_t}}$$

where $d(A)$ is the dynamicity of TF A; $n_t(A)$ is the number of regions bound by A for the first time at time point t; $t_t(A)$ is the number of regions bound by A at time point t and TR_t is the number of regions bound by any TF in time point t. The factor TR_t was added to the expression to account for differences in the amount of reads sequenced by the Assay for transposase accessible chromatin (ATAC-seq)-seq protocol and normalizes the number of regions bound by TF A based on the number of bound regions detected at its corresponding time point. Notice that, if all samples have the same amount of TF binding events, this expression is reduced to the quotient of the sum of the regions first bound at each time point by the sum of all regions bound by the TF at each time point.

By using this definition, the function $d(A)$ maps the activity of a TF to the interval $[\frac{1}{N_t}, 1]$, where N_t is the number of time points in the time course, and is higher as the TF binds to previously not bound regions or leaves already bound regions. In the case of a TF that, for every time point, leaves all its previous bound regions and binds to only regions not previously bound, the numerator will be identical to the denominator, leading to $d(A) = 1$. Alternatively, if a TF remains on the same regions it has bound at $t = 0$, then $\sum_t n_t = n_0$ and $\sum_t t_t = N_t * n_0$, resulting in $d(A) = \frac{1}{N_t}$. Additionally, one can observe that, if the same region is bound by TF A in different time points, it will contribute once to the numerator of the expression, while it will contribute to the denominator once for each time point it has been bound to.

Total number of bound regions

The number of bound regions was calculated by the following the expression:

$$R(A) = \sum_t \frac{n_t(A)}{TR_t} \times \frac{\sum_t TR_t}{N_t}$$

where $R(A)$ is the normalized number of bound regions by TF A during the time course and $n_t(A)$, TR_t and N_t are defined as in section 2.5.6. The first factor is a normalized sum of the regions bound by TF A, counting each region only once. The second factor scales the result by the mean of the number of regions bound by all TFs on each day.

TF percentage of binding at enhancers

The ratio of binding at enhancers, relative to all cis-regulatory regions, was assessed as:

$$P_E(A) = \frac{R_E(A)}{R_E(A)+R_P(A)}$$

where $P_E(A)$ is the percentage of bound regions in enhancers for TF A; $R_E(A)$ is the number of regions bound by TF A marked as enhancers and $R_P(A)$ is the number of regions bound by TF A marked as promoters.

TF prestimulation

For each TF, I computed the ratio of regions bound at Day 0, relative to the number of regions bound during the whole time course. I used the following definition for the prestimulation factor for each TF:

$$p(A) = \frac{\frac{n_{D0}(A)}{TR_{D0}}}{\sum_t \frac{n_t(A)}{TR_t}}$$

where $p(A)$ corresponds to the prestimulation of TF A and $n_t(A)$ and TR_t are defined as in section 2.5.6. The numerator of this expression corresponds to the normalized number of regions bound by TF A at t = D0, while the denominator is the normalized number of regions bound by TF A during the whole time course. Notice the denominator also corresponds to factor $R(A)$ in section 2.5.6 before scaling.

TF binding characteristics and transcriptional modules

In order to characterize the binding activity of each TF for the different gene modules, I performed a hierarchical clustering of the data relative to the dynamicity and the number of bound regions. The TFs were clustered based on the Euclidean distance of both parameters for each gene module as depicted in Figure 2.6, which was generated with the help of the ComplexHeatmap (Gu et al., 2016) and circlize (Gu et al., 2014) R packages. During the computation, I used the mean of the ratio dynamicity - number of bound regions to reorder the dendrogram branches, as implemented by the R package vegan (Oksanen et al., 2019).

I assessed the significance of pioneer (respectively, migrant) TF enrichment at the top (respectively, bottom) of the ranked clustered list by employing a SEA implemented in the package fgsea (Sergushichev, 2016).

2.5.7 TF chromatin binding hierarchy network

In order to assess the TF chromatin binding hierarchy, i.e., TFs required for the binding of a specific TF, we generated a network for each gene module depicting the precedence of TF chromatin binding. The algorithms mentioned were implemented in R (R Core Team, 2017) and all networks were visualized in CytoScape (Shannon et al., 2003).

Computing precedence relationships

The edges in the generated networks represent the precedence relationship of TFs: an oriented edge from TF A to TF B, represented as (A, B), means that A was present in at least 30 % of the cis-regulatory regions bound by B at the same instant of before (Garber et al., 2012). To account for the difference in the amount of reads sequenced for each sample in the ATAC-seq, we normalized the number of regions bound based on the first day they appeared. The weight of an edge from A to B is given by:

$$w_{A \to B} = \frac{\sum_t \frac{R_t(A,B)}{R_t}}{\sum_t \frac{R_t(B)}{R_t}}$$

where $R_t(B)$ stands for the number of regions first bound by TF B at time point t; $R_t(A, B)$, for the number of regions first bound by TF B at time point t that were bound by TF A at time point t or before; and R_t represents the total number of regions bound by any TF in time point t.

The networks were handled using the *igraph* R package (Csardi and Nepusz, 2006).

Network simplification

Aiming to analyze the hierarchical relationship of TFs and simplify the interpretation of the network, we performed two operations over each gene module network: Vertex Sort (Jothi et al., 2009) and transitive reduction (TR) (Aho et al., 1972).

Briefly, the vertex sort algorithm assigns two parameters for each node in the network: the distance, in edges, between the node and the bottom of the network; and the distance between the node and the top of the network. Combined, those parameters allow for the topological ordering of the network, which consists in listing its nodes such that nodes at the top precede downstream nodes. We then defined the 'top layer' as the set of nodes with lowest distance to the top of the network, i.e., nodes that have no incoming edges or nodes that assemble a SCC (Tarjan, 1972) with all upstream nodes. Analogously, the 'bottom layer' was defined as the set of nodes with lowest distance to the bottom of the network, i.e., nodes with no outgoing edges or that form a SCC with all downstream nodes. The 'core layer' comprises nodes that link top layer and bottom layer. We also provide a R implementation of Vertex Sort (Martínez-Zamudio et al., 2019).

The transitive reduction (TR), in turn, simplifies the network visualization by generating the network with the smallest number of edges that keeps the reachability of the original network (Aho et al., 1972). In other words, if there's at least one path between any two nodes in the original network, there will be a path between those nodes in the network's TR. The TR is performed by removing any edge (A, C) when the edges (A, B) and (B, C) exist.

Network visualization

In order to visualize the network, I exported the adjacency matrices in the R environment to CytoScape (Shannon et al., 2003) using the CyREST API (Ono et al., 2015). The networks' layout and style were automated with the help of packages RCy3 (Shannon et al., 2013), r2cytoscape (Pai et al., 2017) and RJSONIO (Lang, 2014).

Network validation

We validated our approach by comparing the network produced when the ChIP-seq data produced by (Garber et al., 2012) is applied to our software. Data was retrieved from Gene Expression Omnibus (GSE36099) and preprocessed as described in section 2.5.3. We computed the precedence relationships among TFs and generated the TF binding hierarchy networks for visualization. Atf4

and Runx1 were removed from the analysis because data was available for only one and two timepoints, respectively.

We compared the produced TF hierarchy network with the network shown in Garber et al. (2012) (Figure 2.8A and B) using two metrics: sensitivity and specificity. Sensitivity is calculated as the ratio of edges described in this study over the edge number sum for both networks. Specificity is defined as the ratio of the number of edges that were described to not exist in the network produced by our software over the number of edges described to not occur in any of both studies.

2.5.8 Proportion of incoming edges based on the classification of the TF source node

Aiming to assess the hierarchy of TFs accordingly to their chromatin dependence and chromatin opening index, we computed the number of edges connecting the sets of all TFs with a given classification for each gene module. We then divided those values by the number of edges that target TFs with a specific classification. Hence, the proportion of incoming edges based on TF classification is given by:

$$P_{C1 \to C2} = \frac{|W_{C1 \to C2}|}{\sum_K |W_{K \to C2}|}$$

where $P_{C1 \to C2}$ is the proportion of edges from nodes with classification C1 to nodes with classification C2; $W_{C1 \to C2}$ is the set of edges from nodes with classification C1 to nodes with classification C2; K can represent either *pioneer*, *settler* or *migrant* and $|\cdot|$ means the cardinality of a set, i.e., the number of elements it contains.

2.5.9 Transcription factor hierarchy networks overlap

To analyze the similarity between the networks for different transcriptional gene modules, I generated a 7-set Chow-Ruskey diagram (Chow and Ruskey, 2003), where each set contains the edges present in the TF hierarchy network relative to a gene module. Edges in two different networks are considered equal if they link nodes corresponding to the same TFs in their respective networks. The intersections of all possible combinations was computed with the use of the R package Vennerable (Swinton, 2019). The same procedure was repeated for specific subsets fo each network, comprising of edges at their top, core and bottom layers, as defined in section 2.5.7.

CHAPTER 3

Metabolic reprogramming in cellular senescence

3.1 Introduction

In chapter 2, I described how TFs associate and impact dynamic gene expression changes during CS. TFs activity is hierarchical, with pioneer TFs making closed chromatin accessible to settler and migrant TFs, which regulate cell fate decision. TF activity also responds to changes in the cell metabolic state, as their conformation can be affected by PTMs and influence their function. As mentioned in section 2.1.1, hypophosphorylated RB binds to chromatin, inhibiting E2F targets (Trimarchi and Lees, 2002). Low ATP levels induce p53 phosphorylation by AMP-activated protein kinase (AMPK), stabilizing p300 binding and the deposition of acetylation marks on the chromatin (Maclaine and Hupp, 2009; Jones et al., 2005). AP1 and NF-κB deacetylation by SIRT1 reduces their transcriptional activity (Xie et al., 2013).

The metabolic state of a cell also modulates gene expression by regulating the epigenome (Tatapudy et al., 2017; Tran et al., 2017; Knaap and Verrijzer, 2016). SIRT, which is activated by high NAD^+ levels and upregulated by physical activity, reduces histone acetylation levels, acting as a transcriptional repressor (Grabowska et al., 2017; Xie et al., 2013; Bori et al., 2012). The activity of the families of demethylases Jumonji and Ten-Eleven Transcription (TET), which reduce DNA and histone methylation, depends on metabolic cofactors such as αKG and uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) (Lewis and Hanover, 2014; Kooistra and Helin, 2012).

As mentioned in section 1.1, CS can arise as a response to a variety of stress stimuli, including proliferative exhaustion, DD and super-physiological activation of oncogenes (Gorgoulis et al., 2019; McHugh and Gil, 2018; Muñoz-Espín and Serrano, 2014; Campisi and Fagagna, 2007). Consequently, the phenotypic transformations occurring in SnC are stressor-dependent, presenting both common and unique response features.

3.1.1 Causes of cellular senescence

Similarly to the aging process itself, CS is a complex phenomenon, controlled by a variety of cross-talking pathways. As mentioned before, Allsopp et al. (1992) observed the first type of CS to be a consequence of telomere integrity loss during cell replication. In addition, several phenomena have been shown to induce CS, eliciting distinct responses depending on the nature of the inducer (McHugh and Gil, 2018; He and Sharpless, 2017; Muñoz-Espín and Serrano, 2014; Van Deursen, 2014; Campisi and Fagagna, 2007). CS induction is not restricted to stochastic variation in the cell and includes genetic programs such as wound healing and development, culminating with

the dephosphorylation of RB, as described in section 2.1.1 (Narita et al., 2003; Johnson and Walker, 1999; Helin, 1998). The known molecular mechanisms that mediate CS are summarized in Figure 3.1.

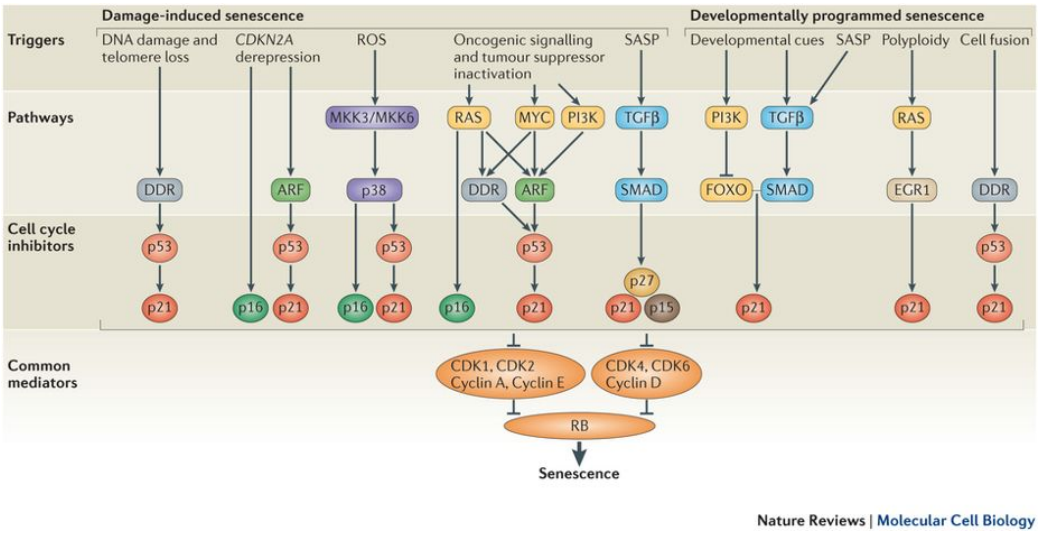


Figure 3.1: Molecular causes of CS (Muñoz-Espín and Serrano, 2014).

Replicative senescence

Telomeres constitute the chromosome extremities, consisting of 6 base pair (bp) sequence (TTAGGG) repetitions (Harley et al., 1990), with a single-strand 3' overhang in mammals (Schmutz and Lange, 2016; Palm and Lange, 2008). Telomere extremities are organized in a structure called t-loop, in which the 3' overhang invades the telomeric double-stranded DNA (Schmutz and Lange, 2016; Palm and Lange, 2008). T-loops are maintained by telomeric repeat binding factor 2 (TRF2), a subunit of the Shelterin complex, which protects telomeres from DD sensors (Schmutz and Lange, 2016; Palm and Lange, 2008). Telomere length can differ in different tissues, ranging from 3 kbp to 17 kbp, and is reduced at each replication cycle (Harley et al., 1990). Telomere shortening leads to a displacement of TRF2, exposing the telomere terminus to the DNA repair machinery (Schmutz and Lange, 2016; Palm and Lange, 2008; Fagagna, 2008). The PI3K-related kinases ATM and ATR respond primarily to double-strand break (DSB) and single-strand break (SSB), respectively (Palm and Lange, 2008), and both can phosphorylate the histone H2A histone family member X (H2AX) into γ-H2AX (Schmutz and Lange, 2016; Palm and Lange, 2008; Fagagna, 2008; Campisi and Fagagna, 2007). γ-H2AX recruits a variety of factors, including MDC1, MRE11 and TP53BP1 (Takai et al., 2003;

Paull et al., 2000), leading to the recruiting of ATM, characterizing a positive feedback response and the formation of structures called DNA segments with chromatin alterations reinforcing senescence (DNA-SCARS) (Klement and Goodarzi, 2014; Rodier et al., 2011; Fagagna, 2008). Additionally, ATM and ATR can phosphorylate CHK1 and 2, which phosphorylate p53 (Bartek and Lukas, 2003). p21 transcription (Fagagna, 2008; Herbig et al., 2004). As mentioned in section 2.1.1, p53/p21 activation arrests cell proliferation together with RB and E2F (Afshari et al., 1996; Dynlacht et al., 1994).

DNA Damage-induced senescence

DDR is not limited to telomeres and can be triggered by other agents, including drugs, hydrogen peroxide (H_2O_2) and irradiation (McHugh and Gil, 2018; Klement and Goodarzi, 2014; Herbig et al., 2004; Sedelnikova et al., 2004). Doxorubicin and etoposide are drugs used in cancer treatment, interfering with DNA integrity through several mechanisms, and exposure to moderate doses has been reported to cause CS (Bielak-Zmijewska et al., 2014; Roberson et al., 2005). Likewise, sublethal doses of H_2O_2 also damages chromatin structure and induces senescence through the p53/p21 pathway (De Magalhães et al., 2004; Poele et al., 2002). Supraphysiological oxygen levels induce CS in mouse embryonic fibroblasts through DNA damage (Parrinello et al., 2003). Rodier et al. (2011) induced the formation of DNA-SCARS by X-ray irradiation, while fibroblasts exit cell cycle and overexpress p53 and several CS markers when exposed to utra-violet radiation (UVB) (Debacq-Chainiaux et al., 2005) and gamma rays (Di Leonardo et al., 1994).

Oncogene-induced senescence

CS can also be induced by gene mutations that transform healthy cells into cancer cells (He and Sharpless, 2017; Muñoz-Espín and Serrano, 2014; Campisi and Fagagna, 2007). The genes subject to this phenomena are denominated *oncogenes*, a category with more than 50 known genes (Gorgoulis and Halazonetis, 2010). OIS was first discovered in cultured cells by Serrano et al. (1997), and later also observed *in vivo* in mice following oncogenic RAS activation (Sarkisian et al., 2007). RAS-OIS is followed by a short period of hyper-proliferation that triggers a DDR and is dependent on both ATM, p16 and p53 expression (Di Micco et al., 2006).

RAF is an oncogene directly downstream of RAS that also induces CS when over-expressed (Dhomen et al., 2009; Zhu et al., 1998). Its mutated version $BRAF^{V600E}$ (where a Valine at position 600 is replaced by glutamic acid) is associated with the formation of benign naevi (Michaloglou et al.,

2005) that can evolve into malignant tumors when associated to mutations in the p53 gene or CDKN2A locus (Dankort et al., 2007). Furthermore, Mitogen-Activated Protein Kinase Kinase 1 (MEK) and Extracellular Signal-Regulated Kinase (ERK), which act downstream of RAS and RAF, also induce CS, in addition to AKT, CCNE, Hsp72, Myc and PTEN (Liu et al., 2018b; Gorgoulis and Halazonetis, 2010; Gabai et al., 2009).

Stress and other senescence inducers

CS can arise as a response to reactive oxygen species (ROS) (Muñoz-Espín and Serrano, 2014; Campisi and Fagagna, 2007). In human fibroblasts, RAS-OIS in a hypoxic enviroment elicits a weaker p21 response, avoiding the onset of the senescent phenotype (Lee et al., 1999). p21 over-expression increases ROS and ceases proliferation of endometrial adenocarcinoma, and ROS inhibition rescues the tumor from CS (Macip et al., 2002). RAS expression and environmental stresses also activate the P38MAPK pathway during CS, which is prevented upon pharmacological inhibition (Borodkina et al., 2014; Debacq-Chainiaux et al., 2010). In human fibroblasts, RAS expression also promotes mitochondrial biogenesis and ROS production, which is only abrogated by inhibition of both p53 and RB (Moiseeva et al., 2009). Furthermore, inhibition of the mitochondrial chaperone Heat Shock 70kDa Protein 9 (HSPA9) or SIRT3/5 activates p53, promoting a senescent phenotype (Wiley et al., 2016; Yaguchi et al., 2007; Wadhwa et al., 2002).

3.1.2 The cellular senescence metabolome

SnCs undergo considerable metabolic reprogramming and are highly metabolically active, despite the suspension of their proliferating activity (Gorgoulis and Halazonetis, 2010; McHugh and Gil, 2018; Wiley and Campisi, 2016). Conversely, metabolic alterations also cause and drive the CS program (McHugh and Gil, 2018; Van Deursen, 2014). The main metabolic changes are depicted in Figure 3.2.

AMPK induces CS by p53 phosphorylation, as a response to high AMP/ATP and ADP/ATP levels, reducing ATP consumption and increasing mitochondrial biogenesis by activating PCG1α (Hardie et al., 2012; Jones et al., 2005). Overall, p53 antagonizes glycolysis by reducing glucose intake and inhibiting pyruvate production, while promoting fatty acid oxydation (FAO) (Wiley and Campisi, 2016). These effects are balanced by the transcriptional activity of its downstream effector RB (Takebayashi et al., 2015). Therefore, SnCs display a mostly glycolytic phenotype, producing lactate and maintaining low NAD^+/NADH ratios (Wiley and Campisi, 2016; Lee et al., 2012; Veer

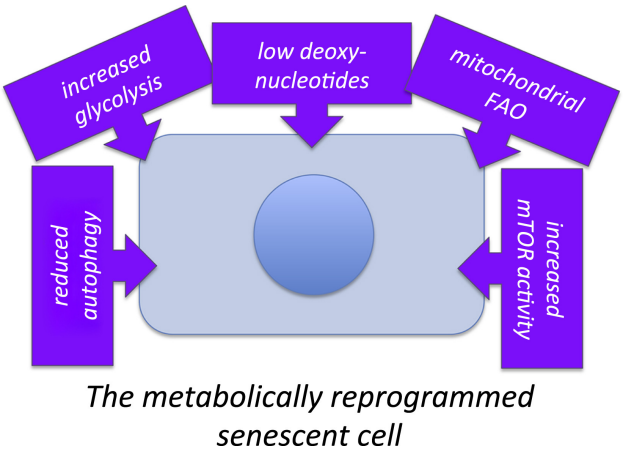


Figure 3.2: CS metabolic reprogramming (Wiley and Campisi, 2016).

et al., 2007). NAD^+ is both a co-factor for SIRTs and for Poly-ADP-ribose polymerase 1 (PARP1), which activate the SASP regulator NF-κB (Gorgoulis et al., 2019; Ohanna et al., 2011; Kim et al., 2005). Interestingly, the Mitochondrial dysfunction-associated senescence (MiDAS) doesn't secrete key inflammatory factors as other CS inducers, such as IL1β, IL6, IL8 and VEGF (Correia-Melo et al., 2016; Wiley and Campisi, 2016).

The SASP also relies on increased lysosome activity during CS (Herranz et al., 2015; Narita et al., 2011). Lysosomes are organelles that metabolize other damaged organelles and proteins, providing the building blocks for the secreted molecules (Hernandez-Segura et al., 2018; Wiley and Campisi, 2016; Narita et al., 2011). An increased lysosomal content can be verified by the accumulation of senescence-associated β-galactosidase (SA-β-Gal), one of the most used biomarkers for detecting SnC (Gorgoulis et al., 2019; Hernandez-Segura et al., 2018). Despite higher lysosomal number and size, SnCs present lower autophagy, impairing mitochondrial renewal and increasing ROS levels (Gorgoulis et al., 2019; Park et al., 2018). This reduced activity is attributed to a higher pH (6.0 instead of 4.5-5), suboptimal for degradation performance (Park et al., 2018; Lee et al., 2006). Additionally, mTOR is a major autophagy suppressor, and its inhibition by rapamycin reduces SASP factors levels, including IL6 and IL8 (Narita et al., 2011).

FAO is also affected in SnCs. Carnitine palmitoyltransferase (CPT1), the outer mitochondrial membrane enzyme that limits the rate of fatty acid synthesis (FAS), is inhibited in OIS, reducing oxygen consumption and impairing the expression of IL6 (Quijano et al., 2012). 5-Lipoxygenase (5LO), an enzyme that catalyzes ROS production from arachidonic acid (AA), promotes CS through a p53-mediated mechanism (Catalano et al., 2005).

Deoxyribonucleotides (dNTP)s are produced by ribonucleotide reductase (RRM2) from nucleotides and consist of the monomers that compose the DNA (Wiley and Campisi, 2016; Aird et al., 2013) . RRM2 is an E2F target (see section 2.1.1), and therefore its expression is decreased in CS, diminishing the availability of dNTP for DNA repair mechanisms. Additionally, its inhibition alone can induce CS, while its addition to SnCs can recover proliferation (Aird et al., 2013).

Notably, most metabolic changes associated with CS were also reported to modulate lifespan in several animal models when perturbed, mediating the aging benefits conferred by dietary restriction (DR) (Antikainen et al., 2017; Grabowska et al., 2017; Carmona and Michan, 2016; Moskalev et al., 2014; López-Otín et al., 2013). The anti-diabetes drug metformin binds to AMPK (Zhang et al., 2012) and is known to extend longevity of female mice by 20 % (Anisimov et al., 2008). In humans, metformin is used not only for the treatment of diabetes, but also cancer (Gallagher and LeRoith, 2011). Rapamycin, a known inhibitor of mTOR, also promoted increased longevity in mice (Zhang et al., 2013; Harrison et al., 2009) and is currently employed in the treatment of kidney diseases, tumours and as an enhancer of immune function in immunosuppressant therapies (Saxton and Sabatini, 2017; Johnson et al., 2013). In *C. elegans*, sir-2 overexpression results in 50 % lifespan increase (Tissenbaum and Guarente, 2001). Additionally, sir-2 is required for a 15 % longevity extension caused by a NAD^+-rich diet (Hashimoto et al., 2010).

3.1.3 Metabolomics

Metabolomics is a rapidly advancing field that produces a quantitative profile of many small molecules (up to 1kDa) present in a biological sample (Gorrochategui et al., 2016; Alonso et al., 2015). The metabolic state of a cell integrates environmental cues and is the primary sensor that cells use in fate decision, modulating the activity of epigenomic modifying enzymes and being the substract for TF PTMs (Tatapudy et al., 2017; Tran et al., 2017; Knaap and Verrijzer, 2016). Metabolite levels impact cell fate by modulating the epigenome and the activity of kinases, as mentioned previously (Tran et al., 2017; Knaap and Verrijzer, 2016). However, the metabolomic data analysis still remains a challenge to-date, and only a small fraction of the acquired data can be confidently associated with a known compound and be biologically interpreted (Schymanski et al., 2015). Figure 3.3 displays the basic steps in metabolomics data analysis.

The major techniques for acquisition of small molecular species abundances are nuclear magnetic ressonance (NMR) and Mass spectrometry (MS)

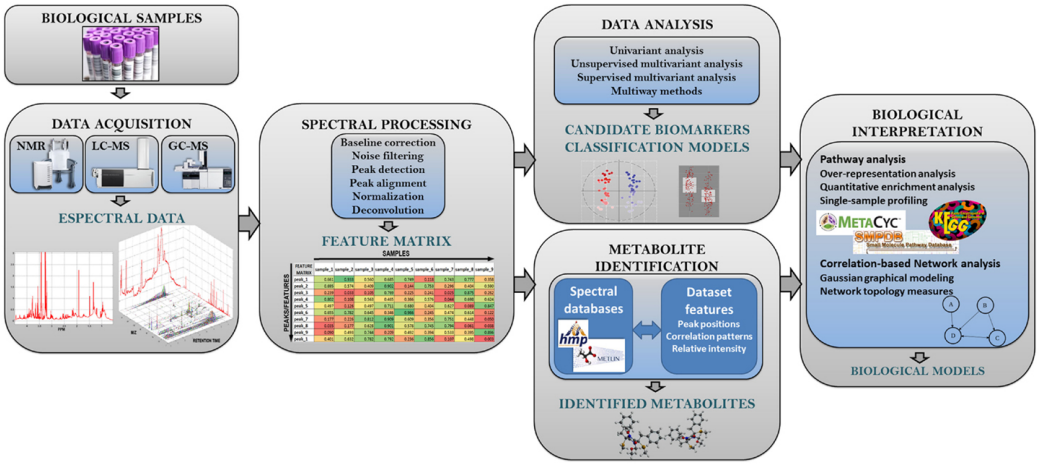


Figure 3.3: Metabolomics studies pipeline overview (Alonso et al., 2015).

(Alonso et al., 2015). NMR studies consist of stimulating samples with an electromagnetic field and measuring the radiation emitted by its protons in order to characterize its containing compounds (Bothwell and Griffin, 2011). This technique provides quantitative concentration of the metabolite abundances, at the cost of lower sensitivity when compared to MS. MS methods are based on launching charged ions into an electromagnetic field and inferring their masses from their displacement amplitude into the field, with heavier molecules moving slower than lighter molecules (Gorrochategui et al., 2016; Mackay et al., 2015). This approach yields a comprehensive spectrum of the sample, which is then de-convoluted and annotated into its constituents. Since several distinct compounds possess the same mass value, the sample is first injected in a chromatographic column, which will elute each metabolite at a distinct rate, thus eliminating a possible ambiguity in data analysis (Gorrochategui et al., 2016; Alonso et al., 2015; Mackay et al., 2015). Chromatographic separation techniques include gas chromatography (GC) or liquid chromatography (LC). While GC allows for a reproducible and sensitive acquisition, it only measures volatile compounds (Gorrochategui et al., 2016). Hence, high performance liquid chromatography (HPLC) and ultra performance liquid chromatography (UPLC) separation yield more comprehensive results (Mackay et al., 2015).

Data processing involves the identification of peaks in the spectral raw data, which are recorded in a feature matrix that lists the intensity of each identified peak in each sample (Gorrochategui et al., 2016; Alonso et al., 2015). The detected abundance of each compound depends on multiple

factors associated to its physico-chemical parameters and affinity with the solvent used at chromatographic separation phase (Kebarle and Verkerk, 2009). Even molecules with same concentration in a sample can produce distinct readouts (Cech and Enke, 2001). Additionally, the acquired results are sensitive to external influences, leading to technical variation that can be confounded with the biological effects from the applied treatment (Rusilowicz et al., 2016; Wehrens et al., 2016; Thévenot et al., 2015). Figure 3.4 illustrates how the measured abundance for a specific metabolite produced by *Arabidopsis thaliana* can drift as samples are injected into the spectrometer (Wehrens et al., 2016). In order to assess and correct the read values for those variations, quality control (QC) samples are injected at regular positions. QC samples can be composed of a set of representative metabolites that will act as a gold standard for later processing steps. In the case of untargeted studies, the QC samples can be produced as a mixture of the biological sample. In case of large-scale studies that make the acquisition in a single run impossible, those batch effects (BEs) can be corrected by the addition of common samples to the distinct batches and the use of computational tools, such as Combat (Thévenot et al., 2015; Johnson et al., 2007).

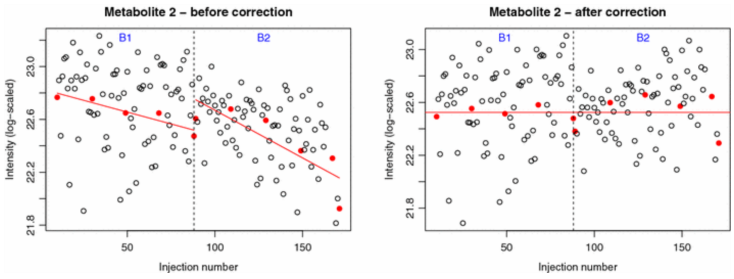


Figure 3.4: Batch correction of MS data. Intensity levels for an *Arabidopsis thaliana* unidentified metabolite measured in two distinct batches. Before correction, the detected levels drift to lower values as samples are injected into the mass spectrometer. The red lines indicate a linear fit of the intensities detected at quality control (QC) samples, which correspond to the same input and are represented as red dots. Additionally, batches B1 and B2 have distinct mean values for the same QCs. Batch correction (BC) techniques normalize both the signal drift and the batch averages based on the QCs, making the data acquired from distinct batches comparable (Wehrens et al., 2016)

LC-MS methods are suitable for both targeted and untargeted metabolome analysis (Gorrochategui et al., 2016; Alonso et al., 2015; Schymanski et al., 2015). In targeted analysis, a small set of metabolites previously characterized and annotated is studied. This strategy focus on a function or pathway of

interest and is more sensitive than untargeted analysis. The untargeted approach results in a more comprehensive and global description of the samples composition, generating a higher amount of data that includes uncharacterized metabolites (Gorrochategui et al., 2016; Alonso et al., 2015). The confidence in a given metabolite annotation is classified in five distinct levels, as depicted in Figure 3.5. Compounds with only known mass of interest are classified as *Level 5* annotation, while a determination of a molecular formula is labeled as identification of *Level 4* (Schymanski et al., 2015). A *Level 1* annotation consists of an identified molecule which spectra is similar to those of a purified metabolite.

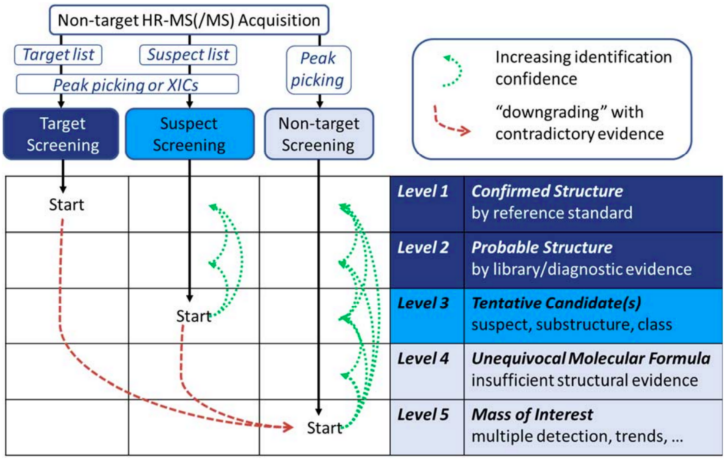


Figure 3.5: Metabolite identification confidence levels (Schymanski et al., 2015).

Finally, the tools for metabolomics data analysis and biological interpretation are common to most bioinformatics studies (Gorrochategui et al., 2016; Alonso et al., 2015). Univariate statistical methods, such as t test or Analysis of Variance (ANOVA), lead to the identification of compounds with a statistically significant abundance in distinct conditions of interest (Thévenot et al., 2015; Quijano et al., 2012). Furthermore, multivariate methods, including partial least squares - discriminant analysis (PLS-DA) and PCA, highlight the molecules and are ideal for studies comparing distinct conditions (Di Gangi et al., 2016; Gromski et al., 2015; Abdi and Williams, 2010). Signaling pathways associated to differentially accumulated metabolites are identified by enrichment analysis.

In this work, I integrated time-resolved metabolomic profiles derived from cells undergoing CS induced by oncogenic RAS and RAF, replicative senescence and DD (etoposide), adding quiescent fibroblasts as *bona fide* control. Among the metabolites that change the most in SnC were αKG and

UDP-GlcNAc, known cofactors for chromatin modifying enzymes that regulate gene expression (Kaochar and Tu, 2012). My comprehensive metabolome analysis provides a deeper understanding of CS and has the potential to lay bare previously unknown vulnerabilities of SnCs that may be exploited for therapeutic ends to promote healthspan.

3.2 Results

3.2.1 Biological discoveries

As many regulatory programs are triggered in response to environmental cues, cell fate decisions are strongly influenced by the metabolic state (Tran et al., 2017; Knaap and Verrijzer, 2016). For instance, the metabolite UDP-GlcNAc is known to be involved in histone modifications and influences the function of the TET family proteins, involved in DNA demethylation, however, the underlying mechanism is still debated (Lewis and Hanover, 2014). TET function also depends on the presence of the aforementioned αKG (Schvartzman et al., 2018). Acetyl-carnitine, which is produced in the mitochondria, is the source of acetyl groups for the activity of histone acetyltraferases (HATs) (Madiraju et al., 2009). Hence, the metabolome regulates the cell's epigenetic state and controls, along with the TF hierarchy, gene expression.

Aiming to generate a comprehensive description of metabolic changes associated with CS and integrate these changes with transcriptome and epigenome, we performed time-resolved metabolic profiling of WI38 fibroblasts undergoing replicative, oncogene-induced (RAS and RAF) and DNA damage induced senescence (etoposide). In addition, we used quiescent fibroblasts as control. Furthermore, we also collected the metabolome of myoblasts undergoing RAS-OIS to compare how different cell types respond to the same inducer. The time points collected for each dataset are represented in Figure 3.6A.

The data acquired for the 118 identified metabolites is shown in Figure 3.6B-G. Each dataset was clustered independently by using the WGCNA tool (Langfelder and Horvath, 2008) and the obtained modules were merged according to the trend of their average temporal profiles. In replicative senescence, I observed that the metabolites followed two distinct trends (Figure 3.6B). A group of 54 metabolites, including carnitine, αKG, glutamate, glutamine ad UDP-GlcNAc, is accumulated in senescent cells. This increase happens sharply at the end of the cells replicative life, as shown in the "Late accumulated" metabolite module. The second main trend in the metabolome of replicative senescence is characterized by an oscillatory behavior over the whole time course. This group is identified as "Oscillatory" in Figure 3.6B and contains metabolites such as lactate, pyruvate, nicotinamide (NAM), palmitoleic acid and AA. When submitted to DNA damage, most identified metabolites are accumulated, with the exception of NAM, aspartate and myristic acid (Figure 3.6C). Upon FBS withdrawal, fibroblasts enter in a quiescent state, characterized by a cell cycle arrest that can be reversed by

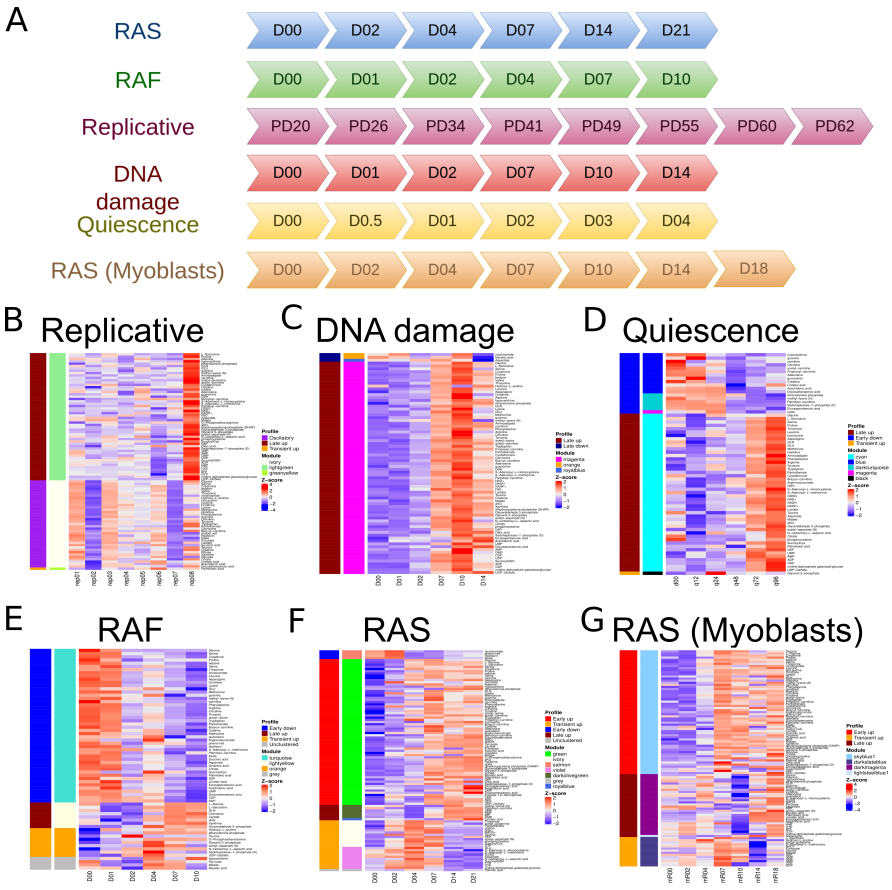


Figure 3.6: The CS metabolome is dynamic. [A] Experimental design. We collected time course data of WI38 cells undergoing CS due to RAS overexpression, RAF overexpression, DNA damage induced by addition of etoposide and replicative exhaustion. We also acquired the metabolome of cells entering quiescence. Additionally, we also performed a time course experiment on myoblasts undergoing RAS-OIS. [B-G] Heatmaps depicting the metabolome evolution in each experiment. [B] Replicative senescence. [C] DNA damage-induced CS. [D] WI38 entering quiescence. [E] RAF-OIS. [F] RAS-OIS. [G] Myoblast RAF-OIS.

serum addition. I observed that this transition is marked by a depletion of 18 identified metabolites, including carnitine, palmitoleic acid and AA, and the increase of 47 identified metabolites, containing αKG, UDP-GlcNAc and lactate (Figure 3.6D). Interestingly, the metabolome dynamics only stabilized after 72h, while we observed in our laboratory that the transcriptome dynamics reaches its regime state after 24h Martínez-Zamudio et al. (2019). In an OIS context, RAF overexpression, 48 identified metabolites were depleted, such as AA, NAM, palmitoleic acid and glutamate (Figure 3.6E). αKG and carnitine

increase over time, while UDP-GlcNAc peaks 4 days after stimulation. In RAS-induced senescence, 73 metabolites accumulated over time, including AA, UDP-GlcNAc, αKG and palmitoleic acid (Figure 3.6F). I also observed a transient accumulation of 22 identified metabolites, comprising palmitic acid. NAM is depleted over time. As fibroblasts, myoblasts undergoing RAS-induced senescence are characterized by a general increase in the identified metabolome (Figure 3.6G). AA, palmitoleic acid and myristic acid are depleted 10 days after stimulation, but their levels increase in late timepoints. UDP-GlcNAc, αKG, carnitine, palmitic acid and NAM are accumulated over time.

In order to integrate those results into a comprehensive picture of CS, I generated a riverplot (Weiner, 2017) with pairwise comparisons of each dataset, shown in Figure 3.7A. Most of the metabolites present in replicative SnCs are also accumulated upon DNA damage, with the exception of aspartate, which is decreased in the latter. This observation is in line with the fact that replicative senescent cells reach this stage due to telomere shortening, which also triggers DDR (Allsopp et al., 1992). When comparing RAS and RAF OIS, all metabolites accumulating upon RAF overexpression, along with the transiently accumulated UDP-GlcNAc, are accumulated in RAS-induced senescence. Conversely, all metabolites depleted upon RAS overexpression were depleted in RAF-induced senescence. Upon RAS induction, 61 identified metabolites followed the same trend in both cell types. However, all four metabolites depleted in fibroblasts in RAS-OIS were accumulated in myoblasts.

Figure 3.7B depicts the trend observed in each dataset for metabolites identified in all experiments. After correcting for batch effects using ComBat (Johnson et al., 2007), I performed a PCA on the dynamic profiles of those metabolites 3.7C, depicting the fibroblast dynamic trajectories upon CS induction.

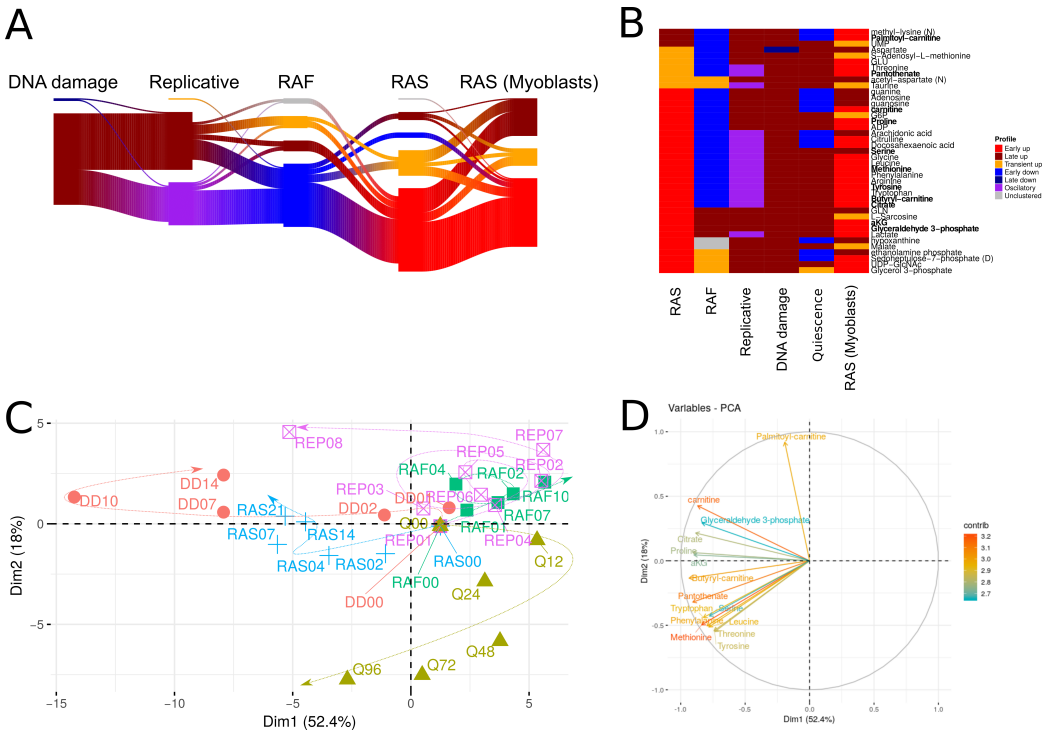


Figure 3.7: Integrative visualization of CS metabolome. [A] Riverplot depicting the intersections between dynamic modules in each dataset. Each dataset is represented by the aggregated clusters based on metabolite trends and curves connecting two adjacent datasets correspond to a shared metabolite among the corresponding clusters. [B] Heatmap showing 41 metabolites identified in all datasets and their corresponding trend in all observed datasets. [C] PCA depicting the temporal evolution of the metabolome for each perturbation. Arrows depict the time dimension, showing the metabolome dynamics for each experiment. [D] Loadings plot showing the top 15 metabolites with the highest contribution in the variance between the observed samples. Arrows direction represent the correlation between a metabolite and the samples in the PCA.

3.2.2 Developed computational methods

As described in section 3.1.3, MS data is subject to several sources of technical variation, potentially leading to an erroneous analysis of the results (Rusilowicz et al., 2016; Wehrens et al., 2016; Thévenot et al., 2015). This issue is addressed by the distribution of QC samples among the biological samples of interest. Those QC samples can be used as references to assess artifacts present in the data due to signal drift during a single acquisition or BEs due to data acquisition in multiple sequences. In order to generate a representative

reference, QC samples are usually prepared as a mixture of the samples of interest (Wehrens et al., 2016; Thévenot et al., 2015).

In this project, however, each time course corresponds to a distinct CS inducer, and using the QCs samples from each batch as references for normalization led to a higher similarity between uninduced fibroblasts and fully senescent cells than to uninduced fibroblasts from distinct datasets (Figure 3.8A). Therefore, I used the first sample of each experiment, which correspond to *biological replicates* acquired over time, as reference samples for batch correction (BC).

In order to validate this procedure, we collected another dataset to be used for BC validation. This dataset contains four samples from both *Quiescence* and *DD* experiments, comprising a set of technical replicates from both experiments acquired by the mass spectrometer at the same batch (Figure 3.8B). I have compared five BC strategies reported in the literature on the two mentioned datasets and used the *Validation* dataset to evaluate the efficacy of each method. Briefly, the strategies include: a) quantile normalization (QN), rescaling based on b) average of uninduced samples, c) average of QCs (Rusilowicz et al., 2016; Thévenot et al., 2015) or d) average of all samples in a batch (Rusilowicz et al., 2016) and e) an algorithm imported from transcriptomics analysis – ComBat (Johnson et al., 2007).

In order to evaluate the performance of each strategy, I compared the values obtained for three different metrics, namely relative standard deviation (RSD) Rusilowicz et al. (2016), repeatability (Wehrens et al., 2016) and the Battacharyya distance (Wehrens et al., 2016). The obtained results are shown in Figures 3.8C-E. From Figure 3.8C, ComBat preceded by QN results in the lowest RSD and highest *repeatability* (also represented in Figures 3.8D,E). Additionally, this approach has the second lowest Battacharyya distance, slightly higher than the value computed for the data normalized based on the uninduced samples average (Figures 3.8D,E).

Figures 3.8F-H show the computed values for each measured peak as a function of the metric computed for the ComBat approach. ComBat yields lower RSD (Figure 3.8F) and higher *repeatability* (Figure 3.8G) for most of the normalized peaks. The obtained Battacharyya distance for each sample is comparable when applying ComBat or any of the approaches consisting of using a set of samples average as the normalization reference (Figure 3.8H).

ComBat efficacy is illustrated in Figures 3.8I-J. In Figure 3.8I, samples corresponding to each batch are clustered together, while technical replicates present a high distance. After normalization, technical replicates are clustered together and samples collected in distinct batches follow parallel trajectories (Figure 3.8J).

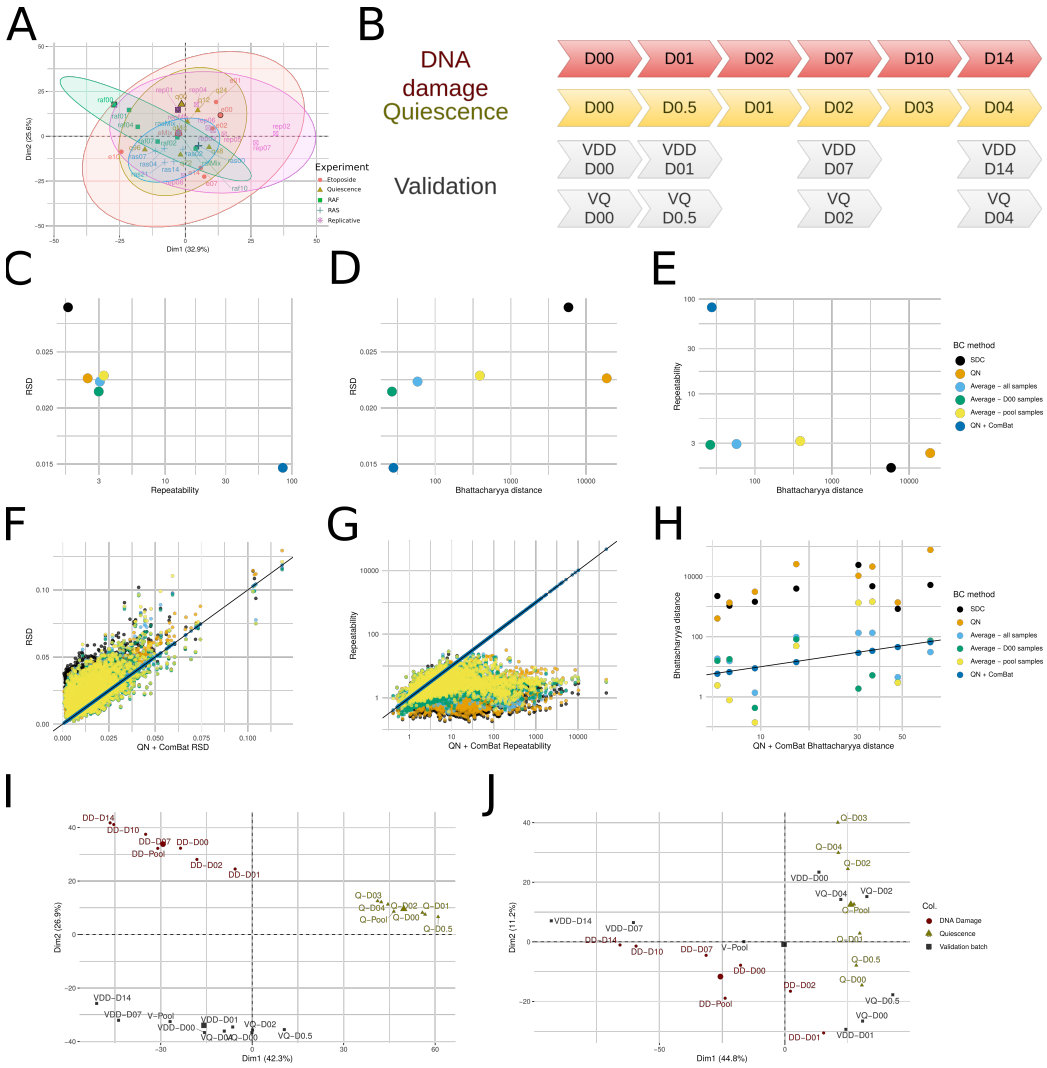


Figure 3.8: BC methods benchmark. [A] PCA depicting distant uninduced samples (marks with black borders) after BC using QC as reference samples. [B] Experimental design for BC validation. *DNA damage (DD)* and *Quiescence* samples correspond to the same samples shown in Figure 3.6. The *Validation* dataset comprises technical replicates of a subset of those samples. [C-E] Visualization of the average computed values for RSD, repeatability and Bhattacharyya distance for each approach. [F-H] Comparison between the obtained values for the ComBat approach following QN and the other approaches for each measured peak (or sample) based on three metrics: RSD [F], repeatability [G] and Bhattacharyya distance [H]. Black lines show the identity function. [I-J] PCA plots depicting the average of each sample used for BC validation before [I] and after [J] normalization by QN and ComBat.

3.3 Discussion

CS is a complex response, characterized by a metabolic reprogramming with unique features dependent on the stress source (Gorgoulis and Halazonetis, 2010; McHugh and Gil, 2018; Wiley and Campisi, 2016). Besides reacting to activated and inhibited signaling pathways, metabolites also regulate the epigenome and gene expression. The metabolome provides substrates and precursors for chromatin and histone modifiers, such as the demethylases Jumonji and TET and the deacetylases sirtuins (Tatapudy et al., 2017; Tran et al., 2017; Knaap and Verrijzer, 2016). Also, TF activity is regulated by PTMs induced by metabolic cues, including p53 phosphorylation, or AP1 and NF-κB deacetylation (Xie et al., 2013; Maclaine and Hupp, 2009). I performed a comprehensive time-resolved metabolome analysis on cells undergoing senescence under distinct types of stress, highlighting the production and accumulation of several metabolites.

My results show that αKG accumulates in all studied datasets. αKG is involved in a variety of physiological processes, including bone development, immune system homeostasis, protein and collagen synthesis (Wang et al., 2019; Wu et al., 2016). This molecule is an antioxidant intermediary of the tricarboxylic acid (TCA) cycle and can be converted to and from glutamate by Glutamate Dehydrogenase 1 (GLUD1) (Wang et al., 2019; Liu et al., 2018a). αKG is used as an alternative energy source by tumors, which develop a glucose-depleted micro-environment (Nguyen et al., 2019; Coloff et al., 2016). Additionally, it can activate NF-κB by directly binding to IKK, which enhances glucose uptake by increasing Glucose transporter 1 (GLUT1) expression (Wang et al., 2019). αKG also acts as a substrate for the families of demethylases TET and Jumonji (Tran et al., 2017; Knaap and Verrijzer, 2016), and a high αKG/succinate ratio is a required to maintain pluripotency (Tischler et al., 2019; TeSlaa et al., 2016; Carey et al., 2015). In SnCs, Agger et al. (2009) and Barradas et al. (2009) showed JMJ3 mediates the INK4a/ARF locus derepression, activating p16 expression (see section 2.1.1). Altogether, those facts imply that αKG is an essential agent for chromatin plasticity and cell fate decision, including CS, with the potential to be a target in ARPs and cancer treatment (Sica et al., 2019).

UDP-GlcNAc, which also accumulates over time in most datasets - except in RAF-OIS, where it is transiently upregulated and depleted after day 4 - can modulate gene expression as substrate for chromatin modifiers and TF PTM. This molecule is produced by the hexosamine biosynthetic pathway as a response to high glucose levels (Özcan et al., 2010). UDP-GlcNAc catalyzes the activity of O-Linked N-Acetylglucosamine Transferase (OGT), an enzyme

that adds O-linked N-acetylglucosamine (O-GlcNAc) as a PTM to histones, TFs and RNA polymerase II directly (Forma et al., 2014; Lewis and Hanover, 2014; Özcan et al., 2010). OGT has been shown to modify histones H2A, H2B and H3 when bound to TET2/3 (Forma et al., 2014; Lewis and Hanover, 2014). Interestingly, OGT does not affect TET demethylation activity. Furthermore, OGT loss impairs PcG repression in *Drosophila melanogaster* promoters (Forma et al., 2014). O-GlcNAc modification of p53 decreases its affinity to Mouse Double Minute 2 (MDM2) (see Figure 2.3), increasing its stability and target transcription (Özcan et al., 2010; Yang et al., 2006). The same pattern is observed in NF-κB dynamics, where its modification by O-GlcNAc inhibits IKK binding in response to hyperglycemia in vascular smooth muscle cells (Özcan et al., 2010; Yang et al., 2008). Additionally, O-GlcNAc-modified NF-κB accumulates in the nucleus and increase TNFα and IL6 transcription in mesangial cells (James et al., 2002). CEBPβ O-GlcNAc modification inhibits its phosphorylation, impairing its ability to bind target gene promoters (Li et al., 2009). phosphorylation inhibition by O-GlcNAc modification was also reported for CREB and Myc, as both modifications target serine and threonine residues (Özcan et al., 2010). Therefore, UDP-GlcNAc modulates gene expression levels by a variety of mechanisms as a response to glucose, being a promising target for further investigation in DR lifespan extension and the response of tumors to a malignant micro-environment.

Our results show that arachidonic acid (AA) is produced in RAS-OIS for both studied cell types (*i.e*, WI38 fibroblasts and myoblasts) and in DD-induced senescence, while being downregulated in RAF-OIS and quiescent cells. The production of AA by RAS-induced SnC was predicted by Sagini et al. (2018), which observed an increase in ACSL3/4 expression, which coordinate lipid biosynthesis and FAO. Here, I confirm that AA is produced as a response to RAS both in fibroblasts and myoblasts. In replicative senescence, AA follows an oscillatory dynamic and does not accumulate in SnCs, as previously reported (Raederstorff et al., 1995). Interestingly, Quijano et al. (2012) observed an accumulation of lipids in OIS, when compared to replicative senescence. AA levels decrease with age, and this fatty acid is the precursor of both pro- and anti-inflammatory metabolites (Das, 2018; Yarla et al., 2017; Lorenzini et al., 2001). AA stimulates NF-κB nuclear translocation and AA oxydative breakdown leads to a lower concentration of NF-κB in nuclear extracts (Camandola et al., 1996). Additionally, its inhibition by phytochemical agents prevents pancreatic cancer onset (Yarla et al., 2017). Therefore, AA has an important role in SASP regulation and modulation of mitochondrial ROS, which will be further characterized in future studies.

Carnitine is accumulated in replicative CS, DD, RAS-OIS for both fi-

broblasts and myoblasts, and depleted in quiescence and RAF-OIS. This metabolite is essential for FAO and is known as an anti-aging agent due to its antioxidant properties, reducing mitochondrial damage (Flanagan et al., 2010; Chang et al., 2002). Oral administration of carnitine reduces SASP levels in rat adipose tissue (Yang et al., 2019). In humans, a study conducted in hemodialysis patients ingesting carnitine for 12 weeks reported a decreased level of IL6 circulating in the blood, while IL1β levels remained unchanged (Shakeri et al., 2010). Carnitine has been reported to reduce liver inflammation in liver cancer through NF-κB inhibition (Jiang et al., 2016). Surprisingly, carnitine supplementation induces CS in gliobastoma and pancreatic cancer cells (Yang et al., 2019; Yamada et al., 2012). Overall, carnitine has the potential to treat malignant tumors by inhibiting the SASP and reducing their proliferative capacity.

When integrating the distinct CS inducers, I observe each dataset follows a distinct temporal profile, as represented by the unique trajectories followed in Figure 3.7C. Interestingly, fully SnCs for RAS-OIS, DD and replicative senescence are closer in the PCA plot than their earlier transiting counterparts, suggesting that metabolites accumulated in those samples constitute a *metabolic signature* for CS. As mentioned above and shown in Figure 3.7D, αKG and carnitine are included in this signature.

The responses elicited by RAS and RAF OIS are considerably distinct. The riverplot shown in Figure 3.7A and the heatmap in Figure 3.7B depict 24 metabolites accumulated under RAS overexpression and depleted in RAF activation. Interestingly, RAF is a direct phosphorylation target of RAS, mediating ERK activation (Liu et al., 2018b; Downward, 2003). As shown in Figure 3.9, RAS activates several signaling cascades besides the RAF/MEK/ERK pathway, including the PI3K and Forkhead box O (FOXO) pathways (Liu et al., 2018b; Van Deursen, 2014; Downward, 2003). Therefore, those alternative signaling pathways may explain the distinct timescale in CS onset and metabolic response for both inducers.

The metabolome of cells undergoing replicative senescence was organized in two major clusters, as shown in Figure 3.6B. The *Late up* cluster, containing αKG, UDP-GlcNAc and carnitine, presented lower levels until cell passage 60, and increased sharply at cell passage 62. This dynamic is coherent with the fact that replicative senescence is a response to telomere shortening, which triggers a DDR when telomeres reach a critical length (Afshari et al., 1996; Dynlacht et al., 1994). The second metabolite module observed in replicative senescence comprised molecules with an *Oscillatory* dynamics during cell proliferation, including the forementioned AA. In 3.7C, this response is illustrated by the *loops* present in the replicative senescence trajectory. Those loops show metabolites oscillating at a frequency between 15 and 20 division cycles.

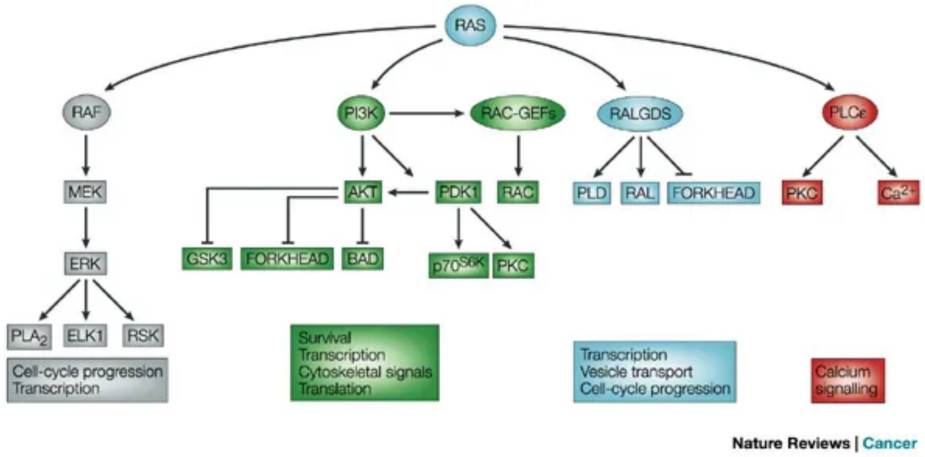


Figure 3.9: RAS signaling pathways. (Downward, 2003).

However, this result could be affected by a phenomena known as *aliasing*, where a high frequency signal is measured at sampling rate lower than its *Nyquist rate* (Unser, 2000; Shannon, 1949). The *Nyquist rate* corresponds to twice the frequency of a signal, and sampling rates below this limit introduce artifacts that block accurate reconstruction of the signal temporal profile from its measurements. Since our samples were collected with an interval between 5 to 8 population doublings, more observations at a higher resolution are needed in order to accurately state that metabolites comprising the *oscillatory* module oscillate accordingly to the cell cycle.

I observed that the global profile for RAS-OIS in fibroblasts and myoblasts is similar, with most identified metabolites being accumulated during CS onset. Specificities regarding cell types involve aspartate, glucose, threonine and panthotenate being transiently upregulated and NAM, folate and glycine depleted in senescent fibroblasts, while being all accumulated in senescent myoblasts. Furthermore, malate, G6P and sarcosine are accumulated in fibroblasts and transiently upregulated in myoblasts during RAS-OIS.

An additional consideration from large-scale metabolomics studies is the occurrence of instrumental biases that may introduce confounding effects in the variable levels and lead to inaccurate conclusions, named *batch effect (BE)*. As stated in section 3.1.3, this is performed with a set of QC samples with similar composition to the studied system. In this work, I used the samples before CS induction from each dataset as a reference for BC. This technique was validated by the acquisition of one extra batch containing a subset of the samples belonging to both the quiescent and DD-induced senescent cells. This approach resulted in a lower variation between corresponding samples

measured in distinct batches and provides a solution for comprehensive studies in which samples are produced over long periods of time. However, this technique is limited to batches containing overlapping samples, and could not be used to analyze fibroblasts and myoblasts quantitatively, where I only compared the dynamic trends followed by each identified molecule.

3.4 Conclusions and Perspectives

CS is a complex response to a plethora of stressors, including telomere erosion, DD, oncogene-activation and ROS production (Gorgoulis et al., 2019; McHugh and Gil, 2018; Muñoz-Espín and Serrano, 2014; Wiley and Campisi, 2016). The response elicited and biomarkers produced by SnCs at least in part is dependent on each inducer, varying in time required for a proliferation arrest, SASP composition and morphological features.

We analyzed an ensemble of time course data collected from cells undergoing different types of senescence and in two different cell-types. We identified metabolites such as αKG and UDP-GlcNAc to be accumulated in SnCs. They are known substrates for epigenome modifiers, Jumonji and TET enzymes, acting as potential targets for modulation of the transcriptome and cell fate (Tran et al., 2017; Knaap and Verrijzer, 2016).

Despite the comprehensiveness of this analysis, the identified metabolites correspond to less than 5 % of the measured peaks. As mentioned in section 3.1.3, metabolite identification is still an open problem in the field of metabolomics, where most data collected constitutes a *dark matter* of peaks that remain unindentified (Wishart et al., 2017; Silva et al., 2015). Consequently, the data produced in this study can be further explored as new inference techniques are developed for MS projects.

My results will be further integrated with transcriptome data collected for each CS inducer, providing a global description of the CS phenotype and the underlying logic governing its transition. My comprehensive analysis will provide insights for cancer suppression therapies through CS-based treatments and healthspan extension.

3.5 Materials and methods

3.5.1 Cell culture

WI38 fibroblasts undergoing RAS-, RAF-, replicative senescence and quiescence were cultured as described by Martínez-Zamudio et al. (2019). DD was induced with the addition of 20 μM Etoposide (Sigma-Aldrich # E1383-25MG) to the culture media. Samples corresponding to each batch were harvested at the time-points indicated in Figure 3.6.

Primary human myoblasts (SkMC) were seeded in cell culture dishes coated with type I collagen (Sigma-Aldrich #C 8919) and cultured in the proliferation medium (DMEM-high glucose (Sigma # D6429), 20% FBS (Life Technologies #10270106), 50μg/ml gentamicin (Life technologies # 15750037), 0,5% Ultroser G (PALL # 15950-017) at 37°C with 5% CO_2. All experiments were conducted between PD11 and PD29 to avoid replicative senescence, and myoblasts were passaged at a cell confluency not exceeding 50 % to avoid myogenic differentiation.

3.5.2 Liquid Chromatography and mass spectrometry

Datasets corresponding to each inducer and cell type were collected in separate batches as described in (Mackay et al., 2015). Briefly, 5μl of samples containing around 5 thousand cells were injected in randomized order in a QExactive Plus Orbitrap mass spectrometer (Thermo Fisher Scientific, Les Ulis, France) after a phase of HPLC separation using a SeQuant®ZIC®-pHILIC 5μm, 150 mm x 4.6 mm column. The scans were acquired in switching polarity. For each batch, 3 QC samples were prepared with equal aliquots of one replicate per sample and evenly injected in each experiment. For samples corresponding to RAF-induced senescence, the QC were injected in a randomized order. Additionally, 3 blank samples, consisting of water, were evenly injected in each batch.

3.5.3 Mass spectrometry data pre-processing

The files generated by the mass spectrometer were converted from Thermo Fisher *.RAW* format to the open data format *.mzXML* using the MSConvert software (Chambers et al., 2012), where ion scans related to positive and negative polarity were stored in independent files for each batch.

Each dataset was pre-processed independently using Bioconductor R packages. Peaks were identified, integrated, matched across samples and filled with the use of the XCMS R package (Smith et al., 2006), and the parameters

for each step were chosen with cross-validation performed by the IPO package (Libiseller et al., 2015). Peaks with an area smaller than the double of the observed area in blank samples were removed. Peaks not detected in any triplicate for a given sample were imputed with the minimum value detected in the batch (Wehrens et al., 2016).

Within-batch signal drift was corrected by fitting linear models to the QC values for each peak as a function of the sample injection order (as illustrated in Figure 3.4). The values observed for each peak were scaled to the ratio between the observed and expected QC intensity (Brunius et al., 2016; Thévenot et al., 2015), accordingly to the following expression:

$$X^*_{p,i} = X_{p,i} \frac{average(X_{p,j})_{j \in QC}}{k * a_p + b_p}$$

where $X^*_{p,i}$ corresponds to the corrected intensity of peak p for the k-th injected sample, denominated i; $X_{p,i}$ to its directly observed intensity; QC is the set of QC samples and a_p and b_p represent the linear and independent coefficients from the linear regression performed over the QC peak levels as a function of its injection order.

After normalization, peaks with a decreased coefficient of variation (CV) were removed.

3.5.4 Compounds annotation

Metabolites were identified by comparing the measured *mass-to-charge* and retention time values with an internal *in-house* database of compounds. After annotation, the tables corresponding to positive and negative polarity were merged into a single table for each acquisition batch.

3.5.5 Statistical analysis

For each batch, the data matrices were *log* transformed and I used ANOVA to determine the significance of metabolite levels. *p-values* were corrected using the false discovery rate (FDR) approach, and, since the obtained *q-values* were reasonable for $p \leq 0.05$, no *q-value* cutoff was applied. In total, 112 molecules were identified as differentially accumulated in at least one sample in at least one batch.

3.5.6 Hierarchical clustering

The compounds time profiles for each dataset were clustered independently using the WGCNA package (Langfelder and Horvath, 2008). This tool clusters

variables in a two-step process, where the first stage groups features based on their vicinity in a correlation network, and the second step merges clusters having *representatives* with high correlation. The representative of a cluster is computed as the median level observed for each sample of its members (Langfelder and Horvath, 2008). The *minimum cluster size* and *deepSplit* parameters from the first step were set, respectively to 50 and 3, resulting in a higher number of small clusters. The *soft threshold* parameter was determined for each batch separately, with the choice of the lowest value leading to a high *Scale free topology fit* by applying the elbow method. In the second step, the threshold for cluster merging was set to 0.80. For this project, I manually merged the modules obtained by WGCNA with similar temporal dynamics and designated the same nomenclature for the profiles from distinct batches.

3.5.7 Integrative analysis

Pairwise combinations between the merged profiles from distinct batches was performed by the *riverplot* R package (Weiner, 2017). The 40 metabolites identified in all batches were normalized using the ComBat tool (Johnson et al., 2007) and the PCA was generated using the R package *factoextra* (Kassambara and Mundt, 2017).

3.5.8 Batch correction methods benchmark

In order to quantify the batch effects (BEs) introduced in the data due to the comprehensive nature of the project, I used the samples corresponding to proliferating fibroblasts from each dataset as reference. I validated this approach in an experimental dataset consisting of technical replicates from a subset of the timecourses corresponding to cells undergoing DD-induced senescence and quiescence. The experimental design is depicted in Figure 3.8A.

Additionally, I evaluated five batch correction (BC) methods reported in the literature: quantile normalization (QN), implemented by the by the *oligo* R package (Carvalho and Irizarry, 2010); a BC using the QC samples form each batch as reference, as mentioned in section 3.1.3 and described at Rusilowicz et al. (2016) and Thévenot et al. (2015); a third BC approach based on the average of all samples in a given batch as a reference for normalization Rusilowicz et al. (2016); a strategy using samples corresponding to biological replicates in each batch as reference for BC (cells before CS or quiescence induction); and the ComBat tool, which infers the parameters of a linear model for BC using a Bayesian approach (Johnson et al., 2007). The approaches

consisting of using a set of samples average as the normalization reference follow the general form given by equation 3.1:

$$\begin{aligned} {X'}_{\mathrm{p,s,b}} &= X_{\mathrm{p,s,b}} \frac{R_{\mathrm{p}}}{C_{\mathrm{p,s,b}}} \\ \text{with } R_{\mathrm{p}} &= \underset{\forall i,j}{avg}(X_{\mathrm{p,i,j}}) \end{aligned} \tag{3.1}$$

Where ${X'}_{\mathrm{p,s,b}}$ and $X_{\mathrm{p,s,b}}$ are respectively the normalized and raw intesity of peak p at sample s in batch b; R_{p} is a scaling factor computed by the average of all detected values for a peak p in all samples in all batches and $C_{\mathrm{p,s,b}}$ is the correction factor computed on the set of reference samples. Its computation for each set of reference samples is given by the following equations:

$$\begin{aligned} \text{QC samples} &: C_{\mathrm{p,s,b}} = \underset{i \in QC(b)}{avg} (X_{\mathrm{p,i,b}}) \\ \text{uninduced samples} &: C_{\mathrm{p,s,b}} = \underset{i \in D00(b)}{avg} (X_{\mathrm{p,i,b}}) \\ \text{all batch samples} &: C_{\mathrm{p,s,b}} = \underset{\forall i \in b}{avg}(X_{\mathrm{p,i,b}}) \end{aligned} \tag{3.2}$$

Those methods were compared based on the values obtained by the computation of three metrics: relative standard deviation (RSD), repeatability and the Battacharyya distance.

The RSD consists of the ratio between the standard deviation (σ) and the average intensity values (μ) measured for each peak p. This value is computed for each sample s over all batches as determined by the following equation (Rusilowicz et al., 2016).

$$\mathrm{RSD} = \frac{\sigma_{\mathrm{p,s}}}{\mu_{\mathrm{p,s}}} \tag{3.3}$$

Repeatability measures the fraction of the variance between replicates of the same sample s over all batches (Wehrens et al., 2016). Its computation is performed for each measured peak p, dividing the variance *between* the averages of all replicates for sample s by the variance of the intensity observed in all replicates *within* the same sample, as shown in equation 3.4. High repeatability is attained by samples sparsely distributed, with replicates densely clustered. As the variance for replicates within a sample approaches (or surpasses) the variance between samples, this quantity decreases.

$$\mathrm{Repeatability} = \frac{{\sigma^2}_{\mathrm{between;p,s}}}{{\sigma^2}_{\mathrm{between;p,s}} + {\sigma^2}_{\mathrm{within;p,s}}} \approx \frac{{\sigma^2}_{\mathrm{biol;p,s}}}{{\sigma^2}_{\mathrm{total;p,s}}} \tag{3.4}$$

The Battacharyya distance D_{B} is an extension of the *Malahanobis distance*. The Malahanobis distance measures the distance between two sets of points,

normalized by their covariance. Therefore, tighter clusters will lead to a higher Malahanobis distance, for the same distance between their center of mass. The Battacharyya distance extends this concept with the introduction of a factor accounting for a distinct distribution in both sets. This metric was calculated using the *fpc* R package (Hennig, 2019) and is given by (Wehrens et al., 2016):

$$D_{\mathrm{B}} = \frac{1}{8}(\mu_{1;\mathrm{s}} - \mu_{2;\mathrm{s}})^{\mathrm{T}}\Sigma^{\text{-}1}(\mu_{1;\mathrm{s}} - \mu_{2;\mathrm{s}}) + \frac{1}{2}ln(\frac{det\Sigma_{\mathrm{s}}}{\sqrt{det\Sigma_{1;\mathrm{s}}det\Sigma_{2;\mathrm{s}}}}) \tag{3.5}$$

Where $\mu_{\mathrm{i;s}}$ corresponds to the center of mass of sample s for batch i, $\Sigma_{\mathrm{i;s}}$ is the covariance matrix for sample s replicates in batch i and Σ_{s} is the covariance matrix for sample s in all batches.

CHAPTER 4

Modelling gene regulatory networks in oncogene-induced senescence

4.1 Introduction

In the previous chapters I described the hierarchical TF network and the metabolic evolution that define CS. Together, these processes impact transcriptional changes, that reallocate cell resources and coordinate the comprehensive CS response (Gorgoulis et al., 2019; Hernandez-Segura et al., 2018; Campisi and Fagagna, 2007). In order to unravel the dynamic molecular interactions that implement this cell fate decision, I performed a quantitative approach.

4.1.1 The epigenetic landscape governing cell fate

Cell fate decision was first elaborated and represented by Waddington (1957) as the trajectory of a *ball* placed on a slightly rugged *landscape* (Figure 4.1). Each position of the ball corresponds to a specific transcriptional state, and the landscape shape determines the possible paths to be traversed, *i.e.*, the possible dynamical changes in gene expression given the cell's current state. Biologically, these possibilities are specified by the genome code, which encodes protein structure, promoter and enhancer nucleotide sequences, and nucleosome PTMs, that control DNA accessibility (Mojtahedi et al., 2016; Huang, 2012; Huang et al., 2005). Stable cell types observed in living organisms correspond to landscape valleys, so-called *attractors*, implemented by regulatory negative feedback loops (Aoki et al., 2019; Huang, 2012; Scheffer et al., 2012). External cues and internal damage reshape the cell's epigenetic landscape, creating and destroying attractors still not optimized by evolution and leading to a diseased state (Chen et al., 2012; Huang, 2012). Despite being characterized as a cell fate, the classification of CS as an attractor state is still debated (Gorgoulis et al., 2019; Martínez-Zamudio et al., 2019).

In dynamical system theory, the *Waddington landscape* is known as a *phase space* (Chang et al., 2017; Wang et al., 2016). Besides providing a graphical visualization from a trajectory, the phase space is associated to the structure of the system it describes. Figure 4.2A depicts an artificial GRN of two genes forming a negative feedback loop with an oscillatory dynamics depicted in Figure 4.2B. When gene x is expressed, it activates y expression (as represented by arrow d, an time interval I). Since y represses x (inhibitory link b), x levels diminish (time interval II). y degrades at a rate c (time interval III), decaying until x, which is produced at a rate a, is activated, starting a new cycle. In Figure 4.2C, the solid curve represents the points represented the same dynamic trajectory as function of x and y, where each axis corresponds to one variable of the system. The faded curves correspond to trajectories of the same system under distinct initial conditions,

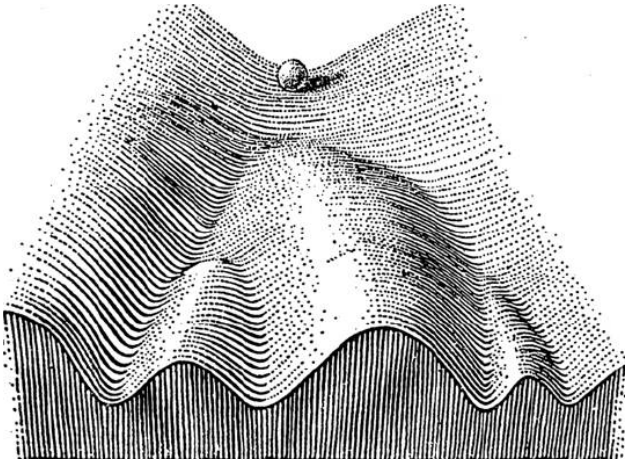


Figure 4.1: Waddington epigenetic landscape (Waddington, 1957)

which oscillates with distinct amplitudes. This representation constitutes the phase space for this artificial network, and it is analogous to the *Waddington epigenetic landscape* of a cell. This simplified GRN example is mathematically inspired in the Lotka-Volterra system, which was initially enunciated as a prey-predator model (Chauvet et al., 2002; Volterra, 1927).

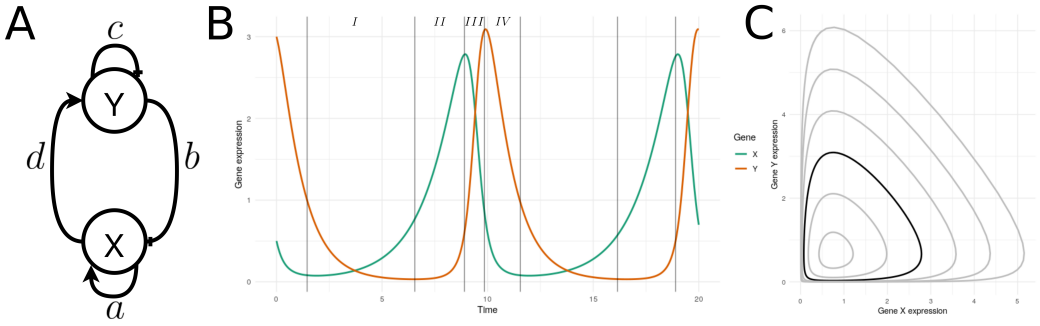


Figure 4.2: Artificial GRN. Inspired in the Lotka-Volterra prey-predator model (Chauvet et al., 2002).

Besides illustrating the correspondence between GRN structure and its dynamics, the Lotka-Volterra model also exemplifies the effect of non-linearities in the interactions between its variables, generating a behavior known as *mirage correlations* (Chang et al., 2017; Sugihara et al., 2012). During time interval II (Figure 4.2B), x and y are negatively correlated, as can be seen by the negative inclination of the corresponding segment in the phase space diagram (Figure 4.2C). However, sampling the time courses during time interval I or III leads to a very low correlation between the two molecular

species. This effect also applies to GRN inference from time course data, as the use of distinct sampling intervals would result in different interaction networks. Therefore, an accurate dynamical system identification requires a strategy that accounts for non-linearities in its regulation. This is especially true for gene expression regulation, which is often implemented by nested feedback loops (Inoue et al., 2016; Dimri et al., 2000; Juven-Gershon and Oren, 1999).

4.1.2 Gene regulatory networks inference

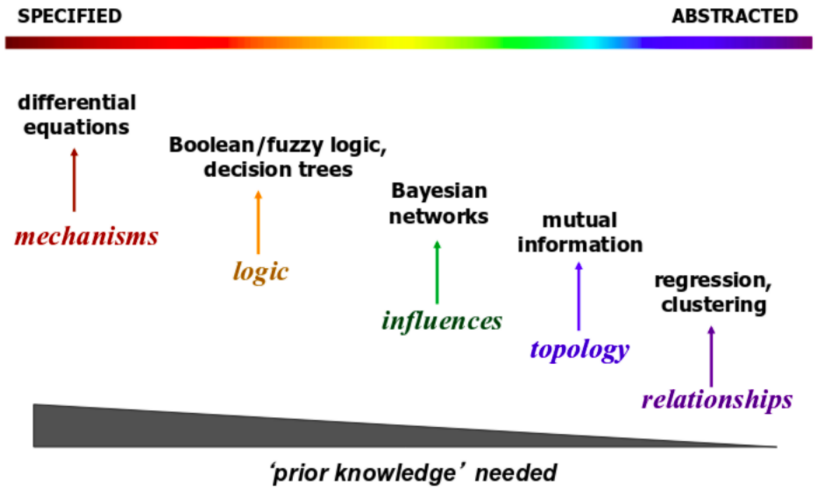


Figure 4.3: Computational techniques for network modelling (Liu and Lauffenburger, 2009)

There is a vast literature on GRN inference of real biological systems, which can exploit prior knowledge on its topology and available data, as shown in Figure 4.3 (Wang et al., 2016; Liu and Lauffenburger, 2009). Bonneau et al. (2007) integrated genome sequence, gene deletion and microarray data in order to feed a regression-based algorithm to generate the GRN of a member of the *Archaea* domain. Gong et al. (2015) used a Bayesian network method to infer the GRN of cardiac differentiation from TF and histone ChIP-seq and gene expression data. Kim et al. (2017) and Cho et al. (2016) built Boolean networks from online available signaling pathways in the context of colorectal cancer, which usually display mutations in RAS and p53 genes. Calzone et al. (2008) developed a network modeling the RB/E2F pathway by aggregating published results, a project that later developed into a comprehensive reconstruction of metabolic networks in cancer and can be

simulated by the Markovian Boolean Stochastic Simulator (MaBoSS) tool (Sompairac et al., 2019; Stoll et al., 2017; Bonnet et al., 2015).

In the context of differential equations, the TNFα and NF-κB signaling pathway have been modeled by several studies Mothes et al. (2015), Cho et al. (2003), and Hoffmann et al. (2002). Chen et al. (2009) developed a system of equations describing the effects of Erb-B2 Receptor Tyrosine Kinase (ErbB) signaling on Mitogen Activated Protein Kinases (MAPK) and PI3K cascades. Dalle Pezze et al. (2014) integrated mTOR signaling with ROS production and mitochondrial biogenesis. Galvis et al. (2019) modeled cell population growth under CS, along with markers including SA-β-Gal. Hackett et al. (2019) inferred a genome-scale network from a comprehensive collection of time course data under distinct TF perturbations.

4.1.3 Compressed sensing

A recently discovered paradigm, named compressed sensing (CoS), has been applied in the reconstruction of GRN due to its robustness under minimum data requirements (Wang et al., 2016; Chang et al., 2014; Candes et al., 2006). Compressed sensing (CoS) works by recovering a sparse signal x_0 with m components from a set y of n observations, with $n << m$ (Candès and Wakin, 2008; Candes et al., 2006). In other words, CoS provides the solution of interest x_0 for the under-determined linear system $y = Ax_0$, where matrix A has n rows and m columns, given x_0 contains less than $n/log(m/n)$ non-null elements. Additionally, the matrix A must obey the restricted isometry property (RIP), which guarantees that every sparse x_0 will lead to a distinct y when multiplied by A, assuring that recovery from y is indeed possible (Candes et al., 2006). Since it allows system structure inference from a reduced amount of measurements, this technique has been applied in several fields, including epidemiology, meteorology and fluid dynamics (Nitzan et al., 2017; Wang et al., 2016; Su et al., 2014).

In order to account for the non-linearities in the inferred interactions, Brunton et al. (2016) developed an approach that builds a set of differential equations based on time course data called "Sparse Identification of Nonlinear Systems (SINDy)". Briefly, this approach consists of a sparse regression on an expanded input matrix $\Theta(X)$ containing non-linear transformations computed from the genes time-series data X. Therefore, our goal is to find the coefficients ξ which satisfy $\dot{x} = \Theta(X)\xi$, where $\dot{x}$ represents the transcription time derivative of a given gene.

This chapter describes the construction of the GRN governing RAS-OIS, where I apply Sparse Identification of Nonlinear Systems (SINDy) to the time course data collected by Martínez-Zamudio et al. (2019) mentioned in

section 2.1 using a HPC environment. When simulating the model, 46 % of the variables displayed a correlation higher than 0.9 with the input data. I validated the model generalization capacity by comparing its output to datasets relative to the inhibition of two TFs, JUN and RELA, showing that the model performance is dependent on the functions performed by each gene. I will further refine the model with a more comprehensive dataset, aiming at providing an accurate description of CS that has as its ultimate goal the acceleration of therapies for ARPs and cancer.

4.2 Results

4.2.1 Biological discoveries

Aiming to develop a quantitative model for CS, I used time-resolved RAS-OIS and the siRNA-mediated TF depletion transcriptome datasets to design a system of differential equations that model RAS-OIS gene expression dynamics *in silico*. This system of equations will be used to:

1) simulate and predict future fibroblast phenotypic behavior in cells undergoing senescence, allowing for the validation of hypotheses in *in silico* experiments, which are more time and cost-effective (*e.g.*, we could perform several parallel simulations of two-week time courses in a few hours);

2) characterize the system by inspecting interactions described in the model that were not yet validated experimentally;

3) steer the system's state based on optimum control theory (Routh, 1877), which ultimately allows for the design of experimental procedures that could cause a senescent fibroblast to revert to a pre-senescent state or any other desired cellular state/fate. For instance, Meza et al. (2005) describe a variety of strategies to control the Lotka-Volterra system, mentioned in section 4.1.1, based on its mathematical description, and Uhlendorf et al. (2012) illustrate how this knowledge can be used experimentally to control a target gene.

Since genes can act collectively when regulating cell fate decisions (Voss and Hager, 2014), a model that describes their temporal profile is required to account for their shared interactions. This implies the presence of terms in the equations that depend on more than one variable, which are inherently non-linear. As mentioned in section 4.1.3, the SINDy algorithm infers non-linear dynamics from time course data by performing a linear regression in a matrix composed of non-linear combinations of the input time courses (Brunton et al., 2016). Figure 4.4A illustrates the approach for a system with three variables. The matrix $\Theta(X)$ is computed from time course data, where each time profile composes a matrix column, and those time profiles are multiplied in order to constitute the non-linear matrix columns. The time course derivatives are estimated from the dynamic data and are represented by the matrix $\dot{X}$. The coefficients that describe the system are inferred independently for each variable, by performing a sparse regression, in the context of the CoS paradigm. The columns in matrix Ξ correspond to the coefficients of each variable, and summarizes the system description.

I implemented the SINDy algorithm using the R statistical language (R Core Team, 2017). I validated the software using two known systems mentioned in section 4.1: the Lorentz Attractor and the Lotka-Volterra prey-predator model. The coefficients inferred for each model are represented in

Figures 4.4B and 4.4C, respectively, and the predicted dynamics simulated from the inferred structure are shown in Figures 4.4D and 4.4E.

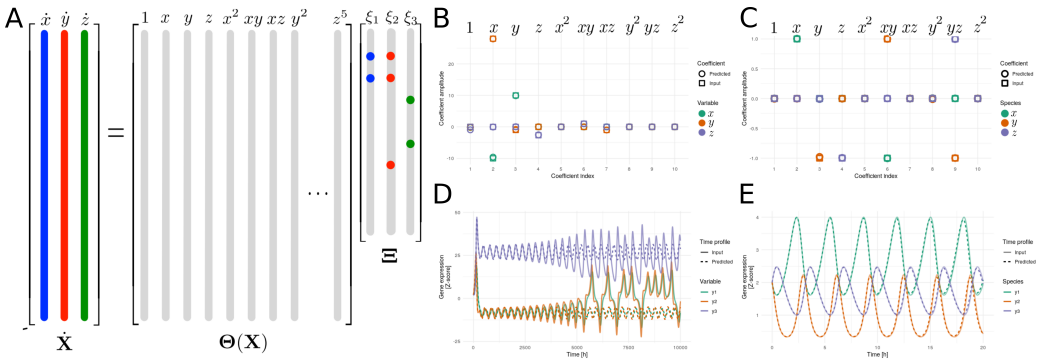


Figure 4.4: SINDy validation for known dynamic systems. [A] A visualization of the SINDy algorithm (Brunton et al., 2016) - Copyright 2016 National Academy of Sciences. See text for details. [B-C] Coefficients amplitude for the original (squares) and inferred models (circles) for the Lorentz [A] and Lotka-Volterra [B] systems. [D-E] Simulated dynamics for the original model, used for model training (shaded solid curves) and the predicted dynamics (dashed curves) simulated from the coefficients inferred by SINDy for the Lorentz [C] and Lotka-Volterra [D] systems.

As shown in Figure 4.4D, the inferred Lorentz model mirrors the beginning of the time course, and deviates from the input time course after around 2500 units of time, where the variable z oscillates with increasing amplitude, while the predicted model's oscillations are constant in amplitude. In Figure 4.4E, the inferred model accurately simulates the original time course.

After ensuring that my software is able to extract the patterns in a known dataset, I applied it to the experimental datasets collected from WI38 fibroblasts undergoing CS. Figure 4.5A shows the experimental design, in which cells were induced to enter in the senescent state by RASV12 overexpression under tamoxifen (4-OHT) treatment (Martínez-Zamudio et al., 2019; Puvvula et al., 2014). Additionally, we collected the transcriptome of cells undergoing senescence two days after inhibition of key transcription factors mentioned in section 2.2.1, namely JUN, RELA and ETS1. The model was inferred using the samples represented as arrows, and the square boxes are the samples used for validation.

Since genes with similar expression profiles in all datasets are indistinguishable from a data-driven perspective and it is impossible to deconvolute their individual contribution to the regulation of the same target, each time course was independently clustered using the WGCNA tool (Langfelder and Horvath,

2008), after being normalized for technical variations by ComBat (Johnson et al., 2007). Figures 4.5B-E depict the clustered time course expression data for each condition, and Figure 4.5F is a heatmap integrating the gene modules obtained for the four conditions used for model training.

In order to address the high number of coefficients in the model to be inferred, which grow exponentially with the chosen order of the model, I used the TensorFlow library to perform the sparse regression for each variable (Abadi et al., 2015). This package was developed for numerical computation in high performance computing evironments and supports execution in Graphics Processing Units (GPUs). Figure 4.6 illustrates the resulting coefficient sets for two genes responding to RAS overexpression. Figures 4.6A and C show the inferred coefficients for CDK6, which participates in cell cycle, and C-X-C Motif Chemokine Ligand 8 (CXCL8), a member of the SASP, respectively. On average, 84 % of the terms in each equation are associated to a null coefficient. Figures 4.6B and D depict the first-order terms for the same genes, corresponding to direct interaction from their regulators. The network's input and output degree distributions are displayed in Figures 4.6E and F, respectively, with an average input and output degree of 259 for first order terms.

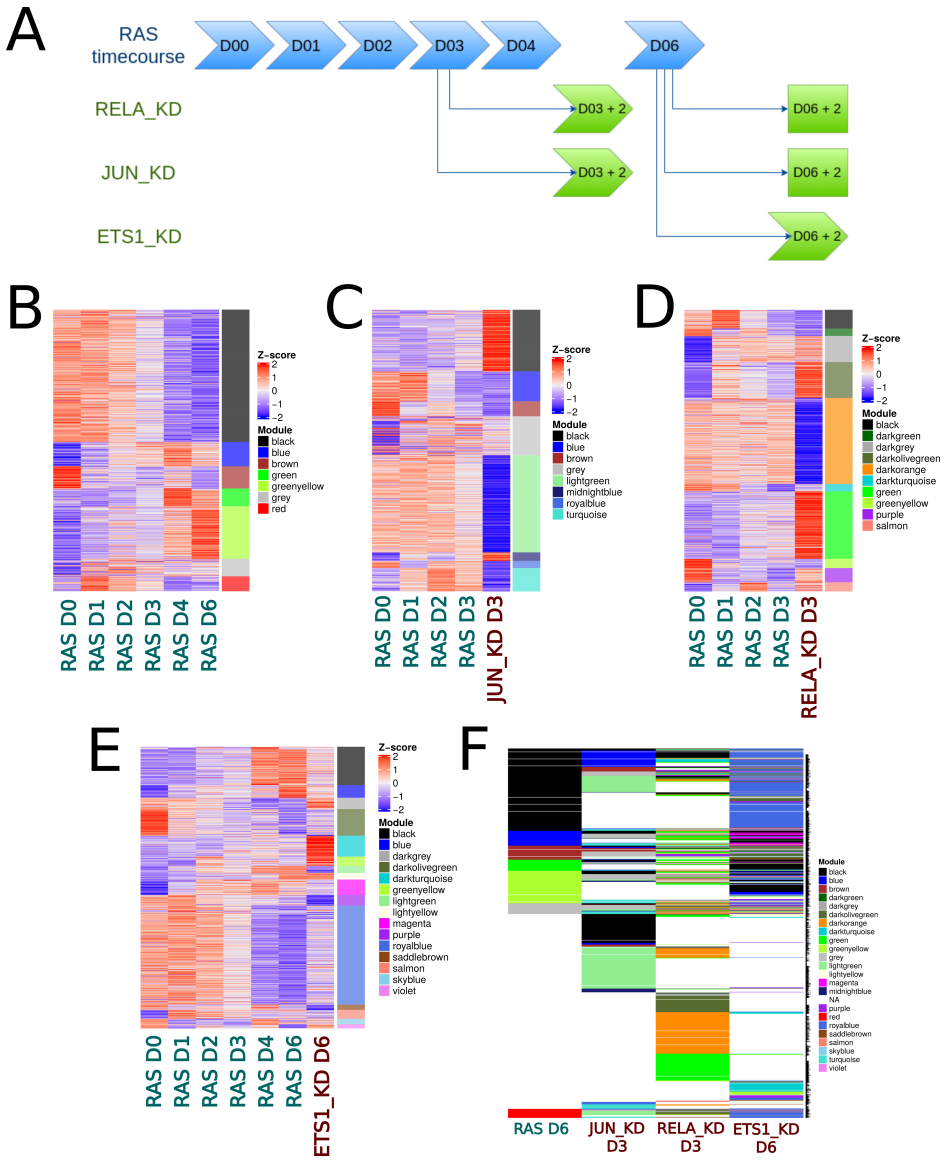


Figure 4.5: Aggregation of genes co-expressed in all experimental conditions. [A] Experimental design. We collected transcriptomic data of WI38 cells undergoing CS due to RAS overexpression for six days. We also performed KD of JUN and RELA at days 3 and 6 and a KD of ETS1 at day 6 using siRNA in cells undergoing OIS, collecting cells two days after treatment. [B-E] Heat maps depicting the data used for model construction. Each line shows the expression profiles for a specific gene over time. These dynamic profiles were clustered using WGCNA (Langfelder and Horvath, 2008). [B] RAS-OIS. [C] JUN KD at day 3. [D] RELA KD at day 3. [E] ETS1 KD at day 6. [F] Heatmap integrating the clustering depicted in panels B-E. Each line corresponds to one gene and 562 unique distinct profiles were identified

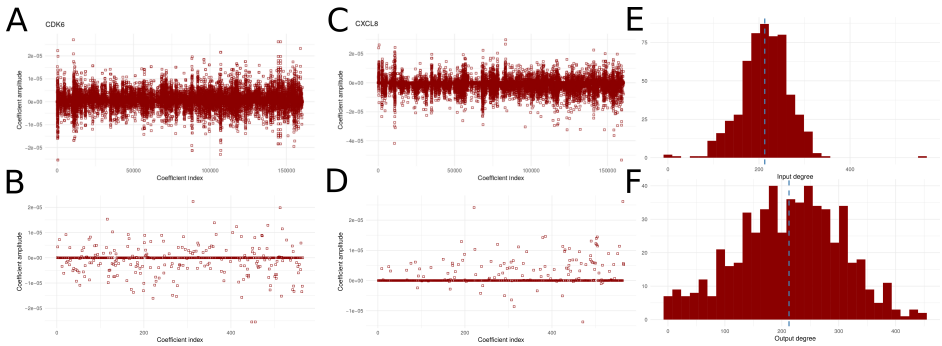


Figure 4.6: Inferred model structure. Coefficients inferred for the equations related to variables containing CDK6 and CXCL8. [A,C] All inferred coefficients for [A] CDK6 and [C] CXCL8, including second-order terms accounting for the interaction between two variables in target gene expression regulation. [B,D] Inferred first order terms, for [B] CDK6 and [D] CXCL8. [E-F] Node degree distribution for [E] incoming and [F] outgoing edges.

Aiming to assess if the model was able to reproduce the data used to generate it, I simulated the inferred coefficients with the deSolve R package (Soetaert et al., 2010). Figures 4.7A-D display the resulting temporal simulations for four genes: [A] CDK6, [B] Cyclin A2 (CCNA2), [C] IL1β and [D] CXCL8. Figure 4.7E depicts the distribution of correlations between the time profiles corresponding to input and prediction for all 565 variables (Bonneau et al., 2007). 260 profiles presented a correlation higher than 0.9.

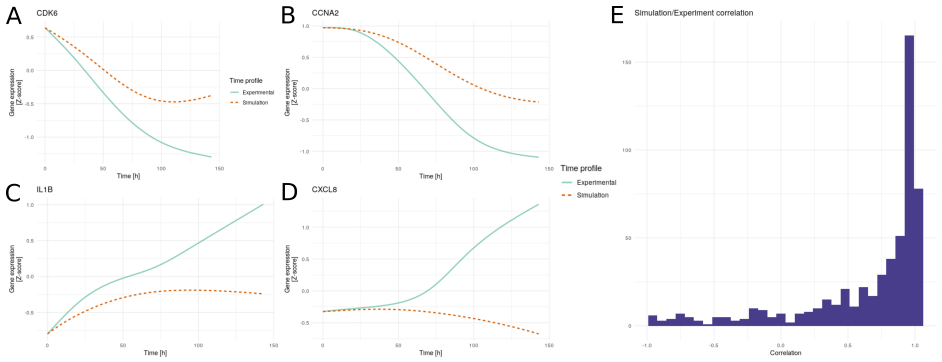


Figure 4.7: Model simulation for OIS gene expression program following RAS induction. Interpolated experimental data (solid light blue lines) and model simulation output (orange dashed lines) time course comparison for [A] CDK6 and [B] CCNA2, genes involved in cell replication, and [C] IL1β and [D] CXCL8, genes that encode cytokines secreted as part of the SASP. [E] Pearson correlation between input and simulation profiles for all 565 unique time profiles identified in our datasets. 260 profiles showed a correlation superior to 0.9.

After validating that the model is able to follow the trends of the experimental data, I investigated its response under the influence of noise (Figure 4.8). I ran five simulations with the addition of Gaussian noise to the system's initial condition. This procedure was repeated for noise with amplitude equivalent to 1 %, 5% and 10 % of the maximum initial condition level of all genes. Figures 4.8A, B and C display the results of these simulations for CDK6, CCNA2, IL1β and CXCL8. As the perturbation at the initial condition increases in amplitude, the simulated profiles keep similar trends. Therefore, the initial deviation is also kept during the whole simulated time frame, leading to a deviation in the model's final state with amplitude proportional to noise intensity.

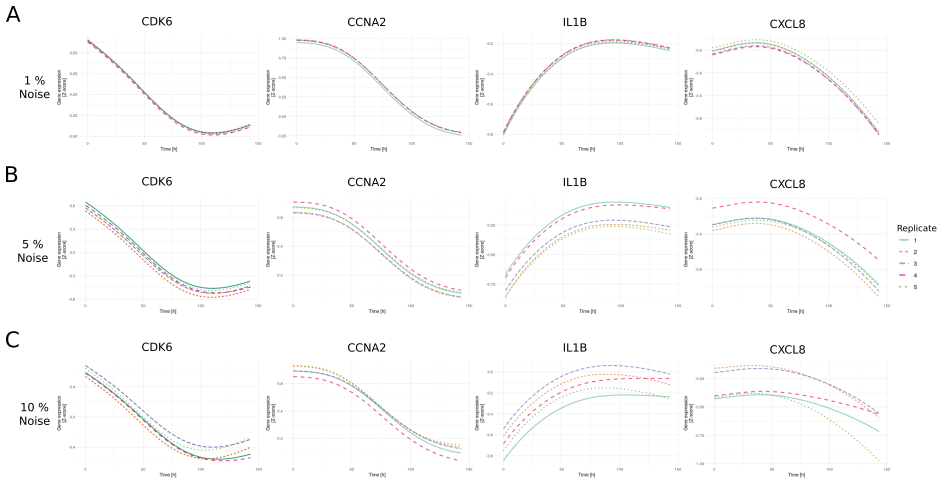


Figure 4.8: Model robustness to noise. Model simulation for initial conditions with three distinct noise intensities. Model simulation results for CDK6, CCNA2, IL1β and CXCL8 under the influence of noise with an intensity of [A] 1 %, [B] 5 % and [C] 10 % of the maximum scaled expression at D0. Simulations were run five times and curves with the same style for all variables correspond to profiles from the same run.

Furthermore, I also analyzed how the model can reproduce the training data regarding the experimentally KD of TFs. I simulated the model with initial conditions corresponding to the timepoint at which cells were treated. In order to implement TF KD *in silico*, I imputed the target TF expression level detected in the sample collected two days after treatment to the initial condition. For example, JUN KD at day 3 was simulated with an initial condition where I imputed the JUN expression level detected in the KD sample, collected two days after treatment, to the gene expression levels measured at day 3 after RAS induction.

Figure 4.9 depicts the results obtained by simulating JUN and RELA KD at day 3 and ETS1 KD at day 6 for CDK6, CCNA2, IL1β and CXCL8. Out of the 565 variables, the model achieved a correlation higher than 0.9 for 465, 434 and 505 profiles, respectively, for JUN and RELA KD at day 3 and ETS1 KD at day 6.

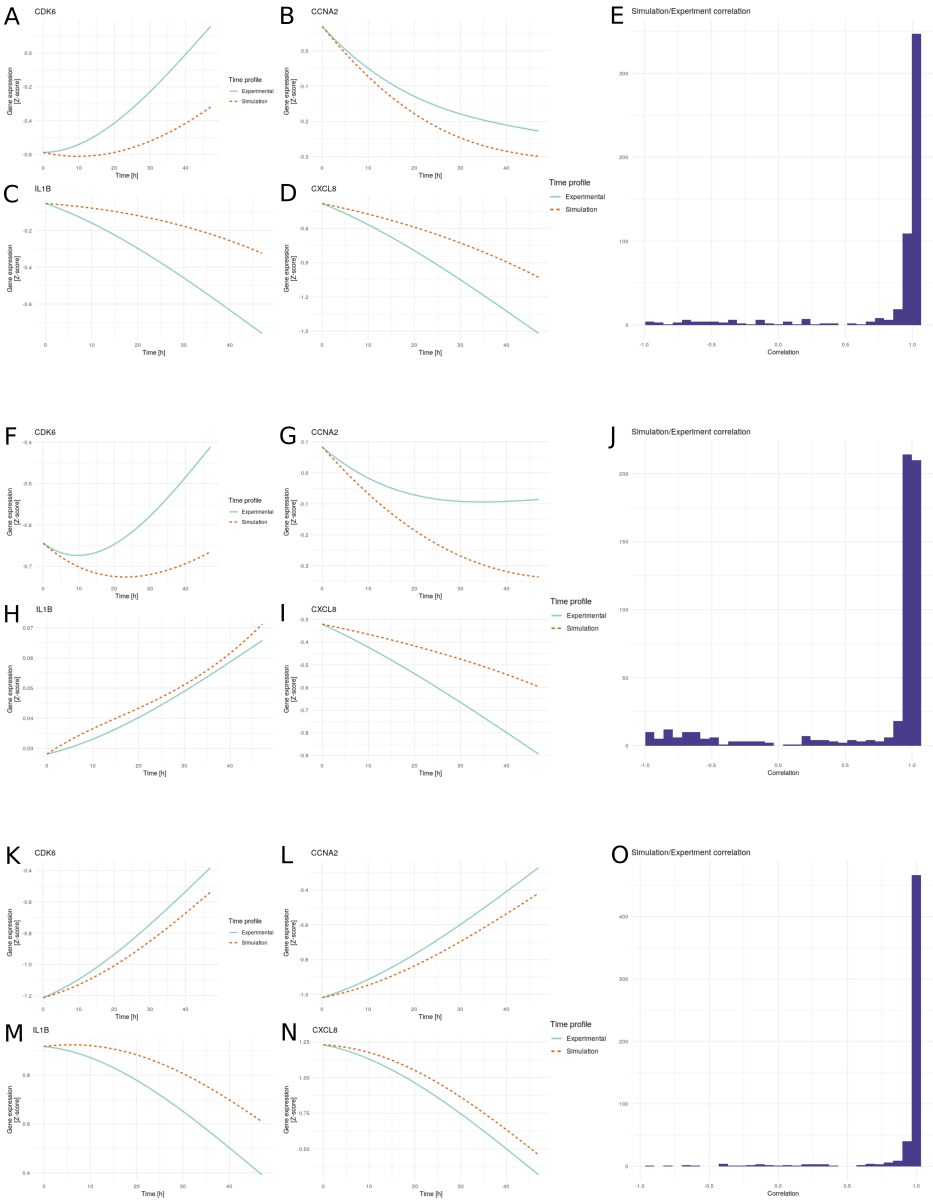


Figure 4.9: Model simulation for TF KD relative to the training datasets Expression profiles following [A-D] JUN and [F-I] RELA KD at day 3 and [K-N] ETS1 KD at day 6 for [A, F, K] CDK6, [B, G, L] CCNA2, [C, H, M] IL1β and [D, I, N] CXCL8. Interpolated experimental data is represented by solid light blue lines and model simulation output is depicted in orange dashed lines. [C] In silico KD implementation: the expression value detected in the KD sample was imputed to the previous time point. [E,J,O] Pearson correlation between input and simulation for [E] JUN and [J] RELA KD at day 3 and [O] ETS1 KD at day 6.

Aiming to validate the model's performance for datasets not used for model training, I evaluated simulations corresponding to the KD of JUN and RELA at day 6 (represented as square blocks in Figure 4.5A). Figure 4.10 depicts the results obtained by solving the equations with gene expression data collected from day 6 after RAS OIS as initial conditions, with imputation of detected values for JUN and RELA in their respective datasets. Since only one timepoint was available after TF inhibition, the correlation between simulated and experimental profiles displayed high absolute value, with an either positive or negative signal depending on the agreement between predicted and observed data upon expression trend for a given gene, as can be seen in the histograms in Figures 4.10E and 4.10J.

Under JUN inhibtion, the model obtained 325 (57 %) profiles with correlation higher than 0.9 and 104 (19 %) profiles with correlation lower than -0.9 when compared to the experimental data. The RELA KD simulation resulted in 234 (41 %) and 188 (33 %) profiles with correlation higher than 0.9 and lower than -0.9, respectively.

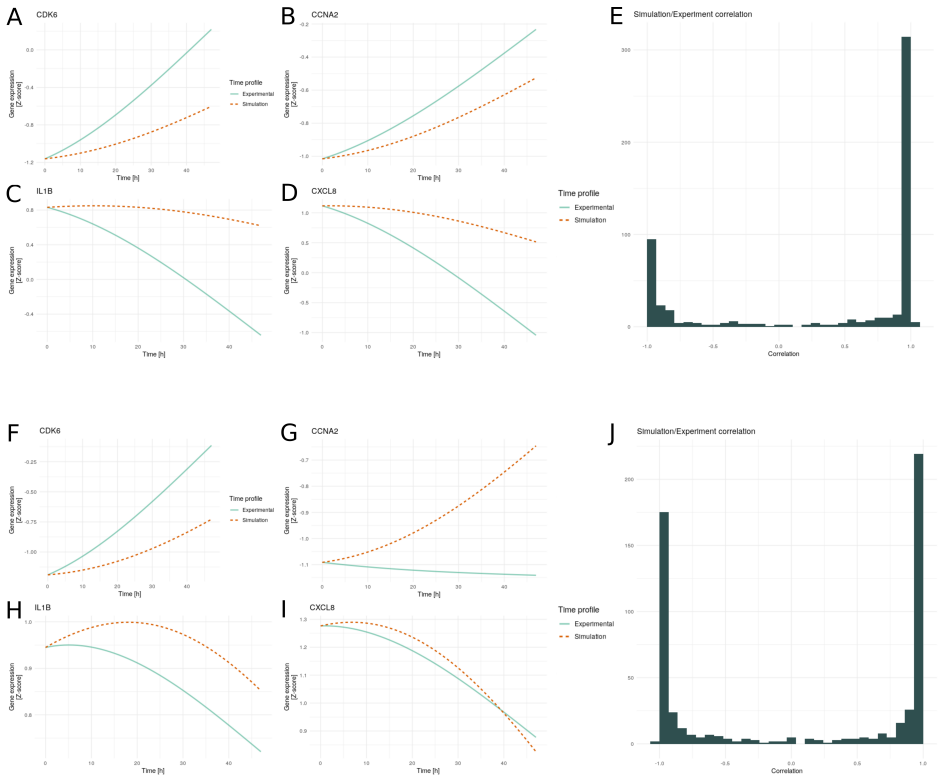


Figure 4.10: Model validation for JUN and RELA KD at day 6 Expression profiles for [A, F] CDK6, [B, G] CCNA2, [C, H] IL1β and [D, I] CXCL8. Interpolated experimental data is represented by solid light blue lines and model simulation output is depicted in orange dashed lines. [E,J] Pearson correlation between input and simulation for [E] JUN and [J] RELA KD at day 6.

Aiming to investigate deeper the model's reliability, I performed a hidden sources analysis in order to estimate which variables that can accurately be inferred from the experimental datasets. This analysis consists of inferring the model with different segments of the data and assessing the variance of each coefficient for segments with distinct sizes (Wang et al., 2016; Shen et al., 2014; Su et al., 2014). Coefficients with higher variance as the amount of data increases suggest that more data is needed to reliably estimate gene regulation. Conversely, a constant or decreasing variance suggest that even a fraction of the available data allow for dynamics prediction.

I run the model inference five times for four different number of data points ratio (20 %, 40 %, 60 % and 80 % of the 284 rows in the $\Theta(X)$ matrix) and calculated the average variance (σ_{avg}) of all coefficients for each model variable. 123 profiles displayed non-increasing σ_{avg} with an increase in the

number of data points used for inference, constituting around 22 % of the model size. Figure 4.11A depicts σ_{avg} for 20 profiles satisfying this condition, and Figure 4.11B illustrates the variables with increasing σ_{avg} as function of the data input size for variables.

I evaluated the model output when simulating the 123 variables with non-increasing σ_{avg}, for the initial conditions corresponding to the test datasets, namely JUN and RELA KD at day 6. Only the equations relative to the non-increasing σ_{avg} were solved, and gene expression levels corresponding to the variables with increasing σ_{avg} were retrieved from the interpolated experimental datasets. Figure 4.11 displays the resulting time profiles for CDK6 and CXCL8, corresponding to profiles with non-increasing σ_{avg}. The predicted results of JUN KD led to 67 (54 %) profiles with correlation higher than 0.9 when compared to experimental dataset, and 8 (7 %) profiles with correlation below -0.9. When inhibiting RELA in the model, 34 (28 %) and 54 (44 %) profiles presented correlation higher than 0.9 and lower than -0.9, respectively.

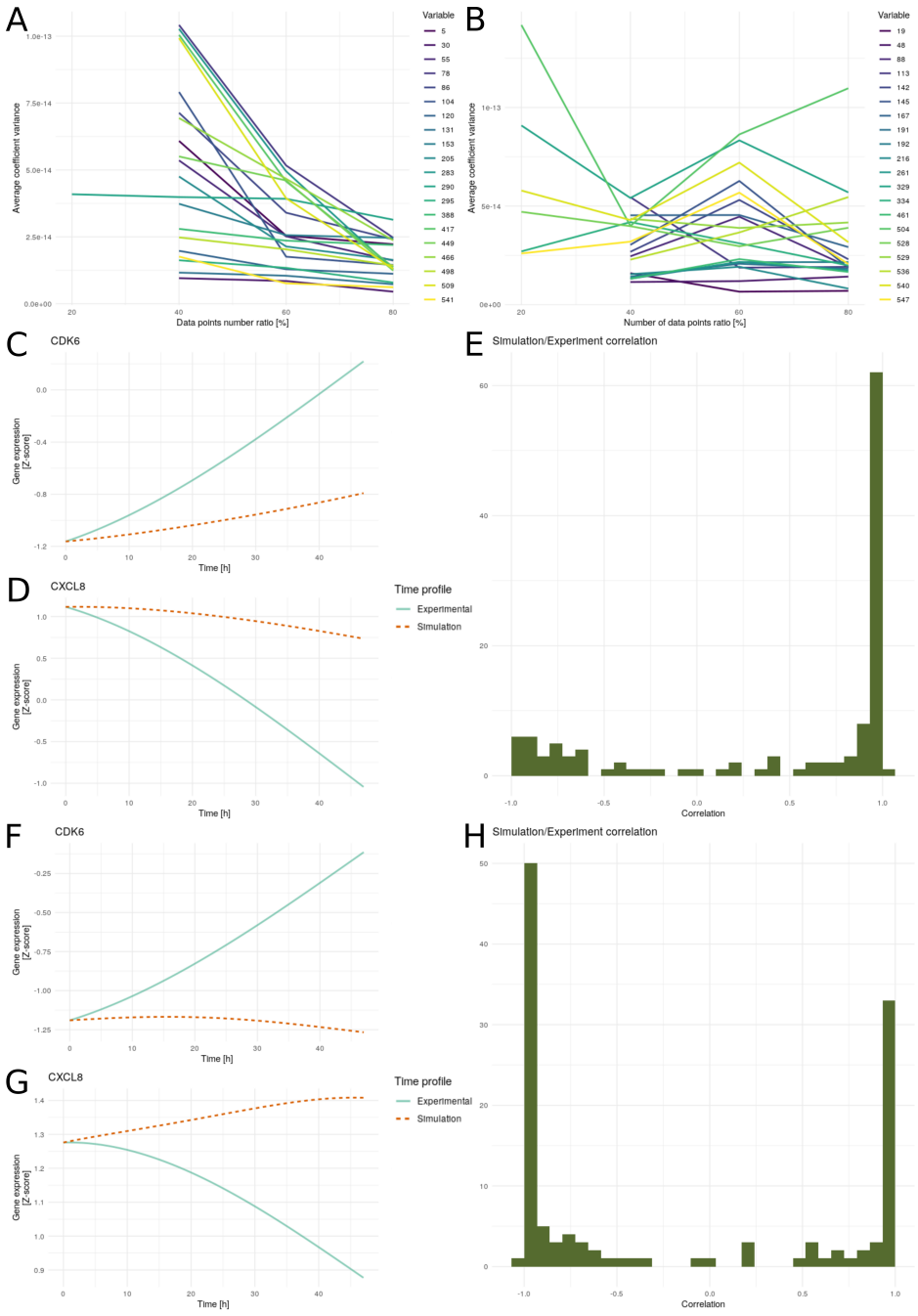


Figure 4.11: Hidden sources analysis and partial simulation. [A-B] Average coefficient variance calculated from five different inferences based on a fraction of the interpolated experimental time courses. [A] 20 variable profiles with decreasing variance and [B] 20 variable profiles with non-decreasing variance as the ratio of data points is increased. Expression profiles for [C, F] CDK6 and [F, G] CXCL8. Pearson correlation between input and simulation for the 123 variables with decreasing variance under [E] JUN and [J] RELA KD at day 6.

Altogether, the inferred model is able to follow the trends of CS onset under RAS overexpression and the response to TF KDs provided during the model training. The model is able to extend this knowledge to predict the results of the inhibition of a TF at the top of hierarchy (section 2.3), but fails to predict gene expression transitions due to the depletion of a TF at the bottom of the hierarchy network. I performed a hidden sources analysis to highlight genes which dynamics can be inferred by the experimental dataset, leading to an improvement in the results for JUN inhibition, but not RELA inhibition. The implemented software can be applied to a more diverse dataset and enhance the model in order to allow us to generate a systemic understanding of CS and efficiently explore its therapeutic potential.

4.2.2 Developed computational methods

The higher generality of SINDy relies on a larger number of coefficients to be inferred. Despite the reduction on the number of non-null coefficients in the solutions ξ_i due to the CoS sparsity assumption, the total number of parameters can increase exponentially with the order of the model to be inferred, considering $\Theta(X)$ is built as a set of polynomial libraries. Therefore, model determination is requires a high running time and considerable computational resources. Additionally, this step must be repeated for each variable in the model, described by an independent equation.

Aiming to find an approach able to address the constraints of this project in a timely manner, I performed a benchmark on GPU-based approaches. A Graphics Processing Unit (GPU) is a CPU specialized for graphics processing, which mainly consists of several linear operations repeated extensively. Therefore, GPUs perform a limited set of functions in a highly parallel manner. Since linear programming optimization methods rely on iteratively multiplying matrices, GPUs are a suitable tool for CoS applications.

The benchmark consists in evaluating the performance of CPU and GPU based tools in the recovery of a signal using the CoS paradigm. I generated a random Gaussian Matrix Φ and an array x_0, which were used to compute the array $y = \Phi x_0$. The task of each tool is to recover x_0 from Φ and y. This procedure was performed five times for four distinct sizes of Φ and x_0, as described in Table 4.1. The tools used for inference consist of the sparse recovery implemented by the *R1magic* R package (Suzen, 2015), using a Newton-type non-linear mapping (NLM); a CPUs implementation for the Broyden–Fletcher–Goldfarb–Shannon (BFGS) algorithm in the *R* language (R Core Team, 2017); and GPU-compatible implementations of the Follow the regularized leader (FTRL) and BFGS algorithms (Abadi et al., 2015). The results are shown in Figure 4.12.

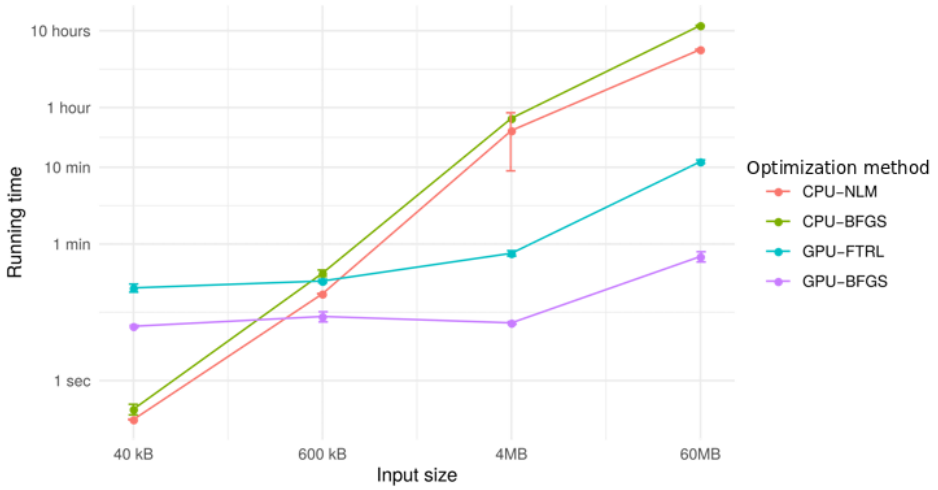


Figure 4.12: Optimization methods benchmark. Running time of for four distinct tools to solve the underdetermined system $y = \Phi x_0$ using the CoS paradigm as a function of the size of the input matrix Φ.

For small Φ, the CPU methods are more efficient due to the time required by the GPU processor to transfer the data to its memory. As the input size increases, GPU-based approaches quickly outperforms their CPU counterparts, with the BFGS implementation being more efficient than the FTRL algorithm in TensorFlow (Abadi et al., 2015). The matrix $\Theta(X)$ used to infer the GRN governing CS dynamics had approximately 400 MB in size, and the running time for the inference of each equation ranged from 10 min to 20 min.

4.3 Discussion

In order to characterize the GRN underlying gene expression changes in CS, I performed a data driven model inference. The model consists of a set of differential equations derived from time course data collected from cells subjected to RAS over-expression. We also measured the transcriptome of samples corresponding to TF KD at two distinct time-points after induction.

I identified 4660 genes to be differentially regulated in at least one sample of the training dataset. When analyzed separately, the genes in each experimental time course are grouped into 7 to 15 co-expression modules, as shown in Figure 4.5. From a data-driven perspective, all genes belonging to the same module, *i.e.*, with temporal profiles highly correlated, provide the same information and it is impossible to deconvolute their individual contribution to the regulation of a given target. Yet, this co-expression behavior changes with the perturbation performed in the cells (Figure 4.5F). I identified 562 unique gene profiles when considering all datasets simultaneously, and each module corresponds to one variable in the inferred model. 243 gene modules correspond to single genes, and the largest module contains 544 genes. To our knowledge, this is the model with the largest scale and highest granularity in the context of OIS (Guimera et al., 2017; Dalle Pezze et al., 2014; Mombach et al., 2014).

As illustrated by the examples of CDK6 and IL8 in Figure 4.6, the inferred model is sparse, where the equation for each variable has around 25 thousand non-zero terms, corresponding to 15 % of the total number of coefficients. The majority of those parameters correspond to non-linear terms in the model (Brunton et al., 2016), describing the collective contribution of two genes on their target expression. Therefore, the inferred GRN constitutes a *hypergraph*, a generalized graph where (hyper)edges can link more than two elements (Gallo et al., 1993). This structure has been used in order to describe protein-protein interaction networks (Patro and Kingsford, 2013) and gene expression regulation by protein complexes in cancer (Tian et al., 2009). Hypergraph visualization is still an open issue and is addressed accordingly to the specific needs of each application (Valdivia et al., 2019; Eschbach et al., 2006). When considering only first-order terms, *i.e.*, a term corresponding to a single variable, the average number of non-null coefficients per equation is 212 (Figures 4.6E,F), corresponding to a proportion of 37 % non-null terms (Figures 4.6B,D).

Figure 4.7 shows representative time profiles produced by the simulation of the inferred equations for initial conditions relative to RAS-OIS. The simulated profiles follow the experimentally measured gene expression levels

until the mark of 30h, followed by an increased divergence between the two temporal dynamics. As shown in Figure 4.7E, most variables keep the same trend as the gene module they correspond to, with 42 % of the time profiles resulting in a correlation higher than 0.9 when compared with the input. Additionally, I simulated the effect of noise at the initial conditions, as depicted in Figure 4.8. The model displays a similar overall tendency to the unperturbed simulation. Yet, this property also means that the initial variations are maintained during the whole timecourse, suggesting the absence of an attractor that leads the dynamics of the model. This phenomenon can be visualized by using the Waddington's landscape representation (Figure 4.1). The initial perturbation can be represented as a second ball released at the side of the represented ball from the top of the landscape, with the distance between the balls proportional to the magnitude of the noise. The results depicted in Figure 4.8 are analogous to two balls following parallel trajectories, arriving at final positions with the same distance as their initial offset.

In addition to RAS-OIS, I simulated the model using initial conditions corresponding to the KD of TFs used as part of the training data (Figure 4.5A, arrow-shaped boxes). These simulations corresponded to a 48h interval and are shown in Figure 4.9. A considerable portion of the simulated profiles presented the same trend as the experimentally collected data.

I validated the model's generalization ability by simulating TF KD at later timepoints in CS onset. As depicted in Figure 4.10, the correlations between the measured and inferred profiles tended to be of high amplitude. This is due to the fact that most genes followed a monotone dynamic, either strictly increasing or strictly decreasing over time. Therefore, equations predicting the same trends as the observed ones led to a high positive correlation, and variables with predicted trends opposite as the measured values resulted in a highly negative correlation. IL1β under RELA inhibition illustrates one exception to this rule, where the inferred dynamic peaks right before 20h and decreases until the end of the simulation (Figure 4.10H). Overall, the validation performance is inferior to the training performance, meaning that some genes follow a distinct dynamic at later CS timepoints, and that this behavior is not encoded in early timepoints. Interestingly, JUN KD yielded a higher percentage of positive correlations than RELA KD. As shown in chapter 2, JUN is a TF at the top of the hierarchy of TFs binding at the chromatin, and is required for RELA regulation of SASP genes. This suggests that the role performed by the TF and its influence in cell state impacts its predictability, where the inhibition of a driver of cell fate leading to more reliable results. Hackett et al. (2019) modelled the yeast GRN based on TF KD timecourse data using a sparse approach. They collected on average 8

timepoints after the inhibition of 201 TFs, yielding a model that explains 43 % of the variability in the data.

Aiming to achieve higher model performance, I investigated the influence of hidden sources on gene expression dynamics, as implemented by Su et al. (2014) and described in detail by Wang et al. (2016). Briefly, equations with increasing coefficient variance as the amount of data used for inference increases are considered to require more observations to be reliably estimated. I inferred the same matrix Ξ on five distinct subsets of the interpolated data $\Theta(X)$ for four distinct percentages of the data (Figure 4.11). 22 % of the distinct time profiles displayed non-increasing variance with higher amounts of data, suggesting that their dynamic is entirely predicted by the observed values, at least in RAS-OIS. Simulating only those variables yielded an improvement on performance when using initial conditions corresponding to JUN KD, while having a higher proportion of negative correlation values for RELA KD. Those results corroborate the hypothesis that, as a regulator of CS, JUN drives transcriptional output and yields more reliable results, while RELA influence on transcription might be dependent on other properties of the cell. Interestingly, this assay illustrates the independence of each equation during inference. Datasets generated in the future can be used both to improve the false predictions with further inference cycles and to validate equations with robust predictions.

The inference of coefficients relative to second-order terms, corresponding to pairwise interactions that regulate transcriptional output, highly increases the number of parameters to be identified. The number of columns in the expanded input matrix $\Theta(X)$ is proportional to the square of the number of variables used in the inference. Therefore, the procedure of finding the set of coefficients that model gene expression dynamics is computationally intensive. I addressed this issue by performing model inference in a HPC environment, leveraging the power of GPU processors. Compared to regular CPUs, GPUs have a simpler architecture per core and are composed of hundreds to a few thousand cores, being suitable for the execution for simple computations repeated several times. Since the inference of matrix Ξ consists of sequential large matrix multiplications, GRN inference can be highly sped up by the use of this technology. The recent popularization of machine learning techniques and its applications fostered the development of software libraries integrating GPU functionalities (Paszke et al., 2019; Abadi et al., 2015), which facilitate the implementation of tools by removing the need to learn architecture-specific programming languages. Also, those libraries can be integrated into virtual containers, *e.g.*, *Sigularity* containers (Kurtzer et al., 2017), that can be transferred between laboratories and computing facilities, therefore fostering reproducibility. Furthermore, the software I implemented can be applied to

larger datasets in order to generate more refined GRN for distinct contexts, including cell differentiation and cancer development.

4.4 Conclusions and Perspectives

I used the CoS paradigm to generate a model describing the regulatory interactions that govern CS. Due to the large number of variables, I performed the inference in a GPU-based environment, decreasing the time required to determine the model's parameters. Validation with the inhibition of two distinct TFs yielded more reliable results for a TF with the ability to open inaccessible chromatin and recruit other TFs in order to regulate transcriptional changes.

CoS has revolutionized the field of signal processing, with applications in image processing, fluid dynamics and meteorological studies Wang et al. (2016), Brunton et al. (2016), Chang et al. (2014), and Candes et al. (2006). The innovation of this concept lies in the possibility of acquiring and storing a signal with less measurements. Until the beginning of this century, it was believed that one needs to collect samples at a rate at least twice the highest frequency in the measured in order to reconstruct it properly, known as the *Nyquist rate* (Unser, 2000; Shannon, 1949). Although this limit is still true for systems with a fixed sampling rate, Candes et al. (2006) showed that, given the data can be sparsely represented in any domain, we can recover the signal from fewer measurements corresponding to linear combinations of the data points.

In the context of Systems Biology, this paradigm also has the potential to reduce the minimum amount of data required in order to build a GRN. In general, biological networks tend to follow a *scale-free* distribution, where few nodes concentrate a high number of connections, and most nodes connect to few neighbors (Broido and Clauset, 2019; Ramirez et al., 2017; Wang et al., 2016; Chang et al., 2014). Therefore, most edges in the networks are null, characterizing scale-free networks as sparse networks and satisfying the CoS sparseness requirement (Aslan et al., 2016; Del Genio et al., 2011). Importantly, this approach may be combined with other layers of data measurable by recently developed technologies, such as ATAC-seq and histone ChIP-seq, which provide epigenome information that impact transcriptional output (Buenrostro et al., 2015; O'Geen et al., 2011). La Manno et al. (2018) developed a computational approach able to infer the rate of RNA synthesis from single-cell RNA-seq data (Figure 4.13), which can be combined with the absolute transcriptional levels to perform a GRN inference using SINDy on thousands of data points from a single experimental assay.

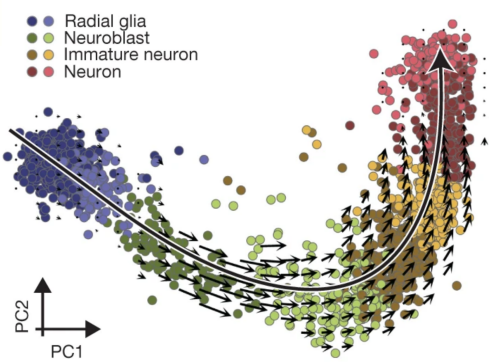


Figure 4.13: Single-cell RNA velocity. Transcriptional evolution of embryonic neural development. Colors correspond to cell types and intermediate cell states. Arrows depict the derivative of gene expression computed from spliced/unspliced RNA ratio (La Manno et al., 2018).

Additionally, CoS approaches have been extended to distinct frameworks that can be used in the context of GRN inference. Hackett et al. (2019) generated a GRN based on univariate sigmoidal functions using yeast transcriptomic data. In a more general framework, Mangan et al. (2016) extended SINDy by using an implicit formulation that allows a dynamic system description with higher order derivatives and rational functions. Kaiser et al. (2018) combined the SINDy paradigm with an optimal control approach, pre-selecting variables that will be used to steer the system future's states.

GRN reconstruction will benefit from quantum computing developments in future studies. In this work, GPU computing decreased inference running time by diminishing the time required for each iteration of the optimization. By exploring state superposition, quantum computers are able to perform several iterations simultaneously, and quantum sparse regression approaches have already been conceived (Gyongyosi and Imre, 2019; Li et al., 2018). Despite the low number of *qubits* in state-of-the-art quantum computers, the field is rapidly growing (Arute et al., 2019). Furthermore, this application can be combined with tools such as the *biglasso* R package, which allows for sparse regression with datasets larger than the primary memory of a computer (Zeng and Breheny, 2017).

GRN inference is quickly evolving, and the development of high-throughput techniques and computational resources have the potential to accelerate this evolution at increasing rates. Our proposed predictive modeling approach will provide a deeper understanding of cellular senescence and has the potential to unravel unknown vulnerabilities of SnCs that may be exploited to promote *healthspan*.

4.5 Materials and methods

4.5.1 Cell culture, RNA extraction and microarrays

The data used in this project includes the time course transcriptome of cells undergoing RAS-induced senescence as described in section 2.5.1. Additionally, we collected gene expression data from senescent cells 48h after siRNA transfection targeting JUN, ETS1 and RELA, as reported by Martínez-Zamudio et al. (2019). JUN and RELA siRNA transfection was performed 3 and 6 days after RAS activation, and ETS1 was inhibited in SnCs 6 days after induction (Figure 4.5A). For the assessment of BE, each KD sample was accompanied of a control sample treated with non-targeting siRNA (Martínez-Zamudio et al., 2019).

4.5.2 Microarray transcriptome data preprocessing, statistical analysis and annotation

The raw Affymetrix HTA 2.0 data was pre-processed using Bioconductor R packages. All samples were normalized using the robust multichip average (RMA) tool implemented by the *oligo* R package (Carvalho and Irizarry, 2010) and batch effects were removed using the Combat tool (Johnson et al., 2007). Affymetrix probes were annotated using the *hta20sttranscriptcluster.db* R package (MacDonald, 2017) and internal control probes were removed. Genes with differential expression lower than 30 % compared to uninduced cells were filtered.

4.5.3 Hierarchical clustering and identification of unique expression time profiles

The training datasets, consisting of the RAS-induced senescence time course, the inhibition of JUN and RELA at D3 and the inhibition of ETS1 at D6, were aggregated as shown in Figure 4.5. The genes in each dataset were clustered independently with the WGCNA tool (Langfelder and Horvath, 2008), where each sample was represented by the median of its replicates. The parameters *minimum cluster size*, *deepSplit* and *threshold for merging clusters* were set, respectively to 100, 3 and 0.85. The *soft threshold* parameter was determined for each dataset separately. See section 3.5.6 for details.

Genes sharing the same modules in all time courses were aggregated, constituting one variable in the model. In total, 4660 genes were identified to change at least 30 % compared to replicating WI38 fibroblasts, and 562

distinct expression profiles were identified, with 243 profiles being associated to single genes. Since the profiles containing JUN, RELA and ETS1 also enclosed other genes, one profile for each gene was added as an isolated variable, constituting a model with 565 variables.

4.5.4 Model inference

In order to satisfy the RIP, the time course samples were scaled to zero mean an unit variance (Wang et al., 2016). Each time course was interpolated by a factor of 24 and the derivative for each variable time profile was numerically computed using the secant method. The time-profiles for each dataset were concatenated into a single matrix X, with rows representing the normalized expression levels for a given time-point and columns representing one variable in the model. The time derivatives for each dataset were also concatenated in a single array $\dot{x}_i$ per variable i, such that each element matches the corresponding row in X.

Aiming to account the collective interaction of TFs in regulating gene transcription (Voss and Hager, 2014; Garber et al., 2012), a library matrix $\Theta(X)$ was computed by columnwise concatenating the matrix X to a second matrix X^{P_2}. As described in Brunton et al. (2016), the columns in matrix X^{P_2} consists of the pairwise multiplication of each column in X, therefore characterizing a second-order model that describes the transcriptome dynamics. In order to avoid redundancy during the inference, only the time points corresponding to samples collected after siRNA addition were used for the KD time courses.

As stated in section 4.1, CS gene expression dynamics is defined by the matrix Ξ, which columns ξ_i correspond to the coefficients representing the influence of each term in variable i. Formally,

$$\dot{X} = \Theta(X)\Xi \tag{4.1}$$

Where $\dot{X}$ represents the concatenated derivatives of each time course for each variable, $\Theta(X)$, the concatenated time courses for each variable and the respective second-order terms, and Ξ, the coefficients describing the impact of the current transcriptomic state in gene expression variation for each variable. Assuming Ξ is sparse and $\Theta(X)$ obeys the RIP, we can find the coefficients by solving the following optimization problem for each variable (Wang et al., 2016):

$$\begin{aligned} &\text{minimize } ||\xi_i||_1 \\ &\text{subject to } \dot{x}_i = \Theta(X)\xi_i \end{aligned} \tag{4.2}$$

where

$$||\xi_i||_1 = \sum_{j=1}^{M} \xi_{j,i} \tag{4.3}$$

is the L1 norm of each column ξ_i of Ξ.

Given the high number of coefficients to be inferred, the optimization was run in a GPU-based computational environment containing NVIDIA®accelerators Tesla K80, Tesla P100 and Tesla M40, with memory ranging from 12 GB to 24 GB. The size of matrix $\Theta(X)$ is approximately 420 MB. The optimization was performed using the SciPy (Jones et al., 2001) implementation of the BFGS algorithm.

4.5.5 Model simulation

All model simulations were run with the R package *deSolve* (Soetaert et al., 2010). At each solver iteration, the model's current state was concatenated with another array consisting of its own pairwise multiplied elements, analogous to the computation of the $\Theta(X)$ matrix described in section 4.5.4. This expanded array was multiplied by the inferred Ξ matrix, yielding the variation in gene expression given a specific state.

4.5.6 Model assessment

In order to assess the performance of 565 variables, I built histograms depicting the correlation between a prediction and its corresponding expected profile as performed by Bonneau et al. (2007).

4.5.7 Canonical systems tool validation

In order to validate the inference implementation, I applied it to two canonical systems: a 3D Lotka-Volterra prey-predator model (Chauvet et al., 2002) and the chaotic Lorentz attractor (Brunton et al., 2016).

The Lotka-Volterra model is described by the following equations:

$$\begin{aligned} \dot{x} &= ax - by \\ \dot{y} &= cx - dy - ez \\ \dot{z} &= fy - gz \end{aligned} \tag{4.4}$$

And the Lorentz attractor:

$$\begin{aligned} \dot{x} &= \sigma(y - x) \\ \dot{y} &= x(\rho - z) - y \\ \dot{z} &= xy - \beta z \end{aligned} \tag{4.5}$$

In both cases, the equations were simulated as described in section 4.5.5 and the produced time courses were used as input for my R implementation of SINDy (see section 4.5.4). The resulting set of coefficients was also simulated as previously described.

4.5.8 Hidden sources analysis

For the purpose of assessing how many and which genes present a dynamics that can be reliably inferred from the available data, I performed a hidden analysis as described by Wang et al. (2016), Shen et al. (2014), and Su et al. (2014). The matrix $\Theta(X)$ (see section 4.5.4), containing 284 rows, was randomly sampled five times for four distinct number of data points ratio R_m (20 %, 40 %, 60 % and 80 %) and the inference procedure described in section 4.5.4 was run independently for each sub-sampled matrix.

For each value of R_m, the coefficient variance was computed, and the σ_{avg} for all coefficients describing the dynamics of a single variable was assessed. Variables with non-decreasing σ_{avg} with respect to R_m were considered to require more data in order to have their dynamics reliably determined.

With the aim of validating the performance of the predicted model when considering only the deducible equations, I simulated the model with initial conditions corresponding to the validation datasets, *i.e.*, the KD of JUN and RELA in senescent cells at D6 after RAS induction. Equations deducible from our datasets correspond to equations with non-increasing coefficient variance as more data is used for its inference Wang et al. (2016), Shen et al. (2014), and Su et al. (2014). The simulations were performed as described in section 4.5.5, and the expression levels corresponding to variables with non-decreasing σ_{avg} were added from the experimental time courses. The performance of each simulation was assessed as described in section 4.5.6.

4.5.9 Inference tools benchmark

In order to evaluate the performance of distinct optimization methods on different platforms, I used the example provided by the *R1magic* R package (Suzen, 2015). An array y is computed from the product of a Gaussian Matrix Φ and a sparse array x_0 and the execution time for each tool to recover x_0

was registered for five distinct combinations of Φ and y. The recovery was accomplished by solving the following linear program:

$$\begin{aligned} &\text{minimize } ||x_0||_1 \\ &\text{subject to } y = \Phi x_0 \end{aligned} \tag{4.6}$$

Where Φ is a matrix with m rows and n columns, y is an array with m elements and x_0 is an array with k non null elements and length n. This procedure was repeated for four distinct sets of values m, n and k, summarized in Table 4.1.

Tests	n	m	k	Memory
1-5	100	40	5	40kB
6-10	400	160	20	600kB
11-15	1000	400	50	4 MB
16-20	4000	1600	200	60 MB

Table 4.1: Table to test captions and labels

The evaluated optimization implementations were (a) the L1 regularization implemented by the *R1magic* R package (Suzen, 2015), (b) a BFGS implementation for CPUs (R Core Team, 2017), and GPU-compatible implementations of the (c) FTRL (Abadi et al., 2015) and (d) BFGS algorithms (Virtanen et al., 2019).

CHAPTER 5

Conclusion

Senescent cells arise as a consequence of molecular damage and accumulate in our organism over time (Gorgoulis et al., 2019; Karin et al., 2019; McHugh and Gil, 2018; López-Otín et al., 2013). Despite its role in tumor suppression and tissue regeneration, chronic presence of SnC cause tissue damage and the onset of several ARPs. Aiming to comprehensively describe the unfolding of CS, I integrated transcriptional, epigenetic and metabolic time-resolved datasets.

I identified the AP1 family members as pioneering TFs preceding the binding of the majority of other TFs on the chromatin, suggesting that they shape the epigenetic landscape governing CS transcriptional changes. TF chromatin binding is organized in a hierarchical network, where pioneers bind to a high number of regions and are followed by settler and migrants, characterized by an activity restricted to fewer targets. The TF networks corresponding to distinct modules of co-expressed genes present unique set of interactions, with a higher overlap at their top and increased specificity at the bottom. Furthermore, even genes with highly correlated transcriptional profiles were bound to distinct TF combinations, suggesting that co-expression does not necessarily imply co-regulation.

CS is characterized by a profound metabolic shift depending on the nature of the damage inflicted on the cell. With the use of a high-throughput approach, I identified a diverse set of metabolites accumulating in SnCs, notably αKG and UDP-GlcNAc acting as substrates for chromatin modifiers, suggesting an important role in the transcriptomic and epigenetic changes associated with CS (Schvartzman et al., 2018; Lewis and Hanover, 2014). UDP-GlcNAc also acts as a precursor to a PTM that controls TF activity in response to glucose (Özcan et al., 2010; Li et al., 2009) AA, an essential precursor in FAO and lipid biosynthesis (Das, 2018), is accumulated during in cells undergoing RAS-OIS, and not in cells senescent due to replicative exhaustion, as suggested by previous studies and confirmed by our observations (Sagini

et al., 2018; Quijano et al., 2012; Raederstorff et al., 1995). Additionally, AA can modulate the activity of TFs regulating the SASP (Camandola et al., 1996). Carnitine is an antioxidant molecule that reduces inflammation levels (Jiang et al., 2016; Shakeri et al., 2010), while inducing CS in malignant cells (Yang et al., 2019; Yamada et al., 2012), being a potential target for clinical applications. Overall, the metabolic response caused by distinct inducers is diverse, suggesting that CS operates in unique ways accordingly to the damage source.

Furthermore, I generated a mathematical model that replicates the transcriptional changes characterizing RAS-OIS. This model consists of a set of second-order polynomial differential equations inferred using the CoS paradigm in a HPC environment. Validation with TF KD datasets yielded better outcome for JUN inhibition when compared to RELA depletion, suggesting that the effects of perturbing TFs at the top of the chromatin binding hierarchy are more predictable. This model will be further refined in future work with additional datasets in order to increase its performance and generate hypotheses that will be experimentally validated.

This study represents a comprehensive description of CS, integrating the distinct layers of regulation that characterize this cell fate. It comprises an in-depth analysis of transcriptome and epigenome data in order to describe the collective TF activity modulating gene expression; a bioinformatics pipeline for mass spectrometry data; and a systems biology approach to mathematically characterize the underlying GRN regulating dynamic changes in transcription. Our study highlights numerous components crucial for CS onset and maintenance, and has the potential to unravel hitherto underappreciated vulnerabilities of SnCs that may be exploited for therapeutic ends to promote healthspan.

Bibliography

Abadi, Martín, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. URL: `https://www.tensorflow.org/`.

Abdi, Hervé and Lynne J Williams (2010). "Principal component analysis". In: *Wiley interdisciplinary reviews: computational statistics* 2.4, pp. 433–459.

Acosta, Juan Carlos, Ana Banito, Torsten Wuestefeld, Athena Georgilis, Peggy Janich, Jennifer P Morton, Dimitris Athineos, Tae-Won Kang, Felix Lasitschka, Mindaugas Andrulis, et al. (2013). "A complex secretory program orchestrated by the inflammasome controls paracrine senescence". In: *Nature cell biology* 15.8, p. 978.

Afshari, CA, MA Nichols, Y Xiong, and M Mudryj (1996). "A role for a p21-E2F interaction during senescence arrest of normal human fibroblasts." In: *Cell growth & differentiation: the molecular biology journal of the American Association for Cancer Research* 7.8, pp. 979–988.

Agger, Karl, Paul AC Cloos, Lise Rudkjær, Kristine Williams, Gitte Andersen, Jesper Christensen, and Kristian Helin (2009). "The H3K27me3 demethylase JMJD3 contributes to the activation of the INK4A–ARF locus in response to oncogene-and stress-induced senescence". In: *Genes & development* 23.10, pp. 1171–1176.

Aho, Alfred V., Michael R Garey, and Jeffrey D. Ullman (1972). "The transitive reduction of a directed graph". In: *SIAM Journal on Computing* 1.2, pp. 131–137.

Aird, Katherine M, Gao Zhang, Hua Li, Zhigang Tu, Benjamin G Bitler, Azat Garipov, Hong Wu, Zhi Wei, Stephan N Wagner, Meenhard Herlyn, et al. (2013). "Suppression of nucleotide metabolism underlies the establishment and maintenance of oncogene-induced senescence". In: *Cell reports* 3.4, pp. 1252–1265.

Allsopp, Richard C, Homayoun Vaziri, Christopher Patterson, Samuel Goldstein, Edward V Younglai, A Bruce Futcher, Carol W Greider, and Calvin B Harley (1992). "Telomere length predicts replicative capacity of human fibroblasts". In: *Proceedings of the National Academy of Sciences* 89.21, pp. 10114–10118.

Alonso, Arnald, Sara Marsal, and Antonio Julià (2015). "Analytical methods in untargeted metabolomics: state of the art in 2015". In: *Frontiers in bioengineering and biotechnology* 3, p. 23.

Andrade, T. (2016). *The Gunpowder Age: China, Military Innovation, and the Rise of the West in World History*. pg. 30. Princeton University Press. ISBN: 9781400874446. URL: https://books.google.fr/books?id=1jRJCgAAQBAJ.

Anisimov, Vladimir N, Lev M Berstein, Peter A Egormin, Tatiana S Piskunova, Irina G Popovich, Mark A Zabezhinski, Margarita L Tyndyk, Maria V Yurova, Irina G Kovalenko, Tatiana E Poroshina, et al. (2008). "Metformin slows down aging and extends life span of female SHR mice". In: *Cell cycle* 7.17, pp. 2769–2773.

Antikainen, Henri, Monica Driscoll, Gal Haspel, and Radek Dobrowolski (2017). "TOR-mediated regulation of metabolism in aging". In: *Aging cell* 16.6, pp. 1219–1233.

Aoki, Stephanie K, Gabriele Lillacci, Ankit Gupta, Armin Baumschlager, David Schweingruber, and Mustafa Khammash (2019). "A universal biomolecular integral feedback controller for robust perfect adaptation". In: *Nature*, p. 1.

Arute, Frank, Kunal Arya, Ryan Babbush, Dave Bacon, Joseph C Bardin, Rami Barends, Rupak Biswas, Sergio Boixo, Fernando GSL Brandao, David A Buell, et al. (2019). "Quantum supremacy using a programmable superconducting processor". In: *Nature* 574.7779, pp. 505–510.

Aslan, Melih S, Xue-Wen Chen, and Hong Cheng (2016). "Analyzing and learning sparse and scale-free networks using Gaussian graphical models". In: *International Journal of Data Science and Analytics* 1.2, pp. 99–109.

Atsaves, Vasileios, Vasiliki Leventaki, George Z Rassidakis, and Francois X Claret (2019). "AP-1 Transcription Factors as Regulators of Immune Responses in Cancer". In: *Cancers* 11.7, p. 1037.

Augustyn, Adam, Patricia Bauer, Brian Duignan, Alison Eldridge, Erik Gregersen, Amy McKenna, Melissa Petruzzello, John P. Rafferty, Kara

Ray Michael abd Rogers, Amy Tikkanen, Jeff Wallenfeldt, Adam Zeidan, and Alicja Zelazko (2009). *Philosopher's stone*. URL: https://www.britannica.com/topic/philosophers-stone.

Ayyadevara, Srinivas, Ramani Alla, John J Thaden, and Robert J Shmookler Reis (2008). "Remarkable longevity and stress resistance of nematode PI3K-null mutants". In: *Aging cell* 7.1, pp. 13–22.

Baker, Darren J, Tobias Wijshake, Tamar Tchkonia, Nathan K LeBrasseur, Bennett G Childs, Bart Van De Sluis, James L Kirkland, and Jan M van Deursen (2011). "Clearance of p16 Ink4a-positive senescent cells delays ageing-associated disorders". In: *Nature* 479.7372, p. 232.

Baker, Darren J, Bennett G Childs, Matej Durik, Melinde E Wijers, Cynthia J Sieben, Jian Zhong, Rachel A Saltness, Karthik B Jeganathan, Grace Casaclang Verzosa, Abdulmohammad Pezeshki, et al. (2016). "Naturally occurring p16 Ink4a-positive cells shorten healthy lifespan". In: *Nature* 530.7589, p. 184.

Barradas, Marta, Emma Anderton, Juan Carlos Acosta, SiDe Li, Ana Banito, Marc Rodriguez-Niedenführ, Goedele Maertens, Michaela Banck, Ming-Ming Zhou, Martin J Walsh, et al. (2009). "Histone demethylase JMJD3 contributes to epigenetic control of INK4a/ARF by oncogenic RAS". In: *Genes & development* 23.10, pp. 1177–1182.

Bartek, Jiri and Jiri Lukas (2003). "Chk1 and Chk2 kinases in checkpoint control and cancer". In: *Cancer cell* 3.5, pp. 421–429.

Benary, Uwe and Jana Wolf (2019). "Controlling Nuclear NF-κB Dynamics by β-TrCP—Insights from a Computational Model". In: *Biomedicines* 7.2, p. 40.

Bergmann, C, B Merlevede, C Beyer, L Hallenberger, A Brandt, C Dees, S Pötter, A Bozec, G Schett, and J Distler (2018). *FRI0400 The ap1 transcription factor cjun amplifies hedgehog-induced fibroblast activation and tissue fibrosis*.

Bhat, Rekha, Elizabeth P Crowe, Alessandro Bitto, Michelle Moh, Christos D Katsetos, Fernando U Garcia, Frederick Bradley Johnson, John Q Trojanowski, Christian Sell, and Claudio Torres (2012). "Astrocyte senescence as a component of Alzheimer's disease". In: *PloS one* 7.9, e45069.

Bielak-Zmijewska, Anna, Maciej Wnuk, Dorota Przybylska, Wioleta Grabowska, Anna Lewinska, Olga Alster, Zbigniew Korwek, Anna Cmoch, Aleksander Myszka, Slawomir Pikula, et al. (2014). "A comparison of replicative senescence and doxorubicin-induced premature senescence of vascular smooth muscle cells isolated from human aorta". In: *Biogerontology* 15.1, pp. 47–64.

Bonneau, Richard, Marc T Facciotti, David J Reiss, Amy K Schmid, Min Pan, Amardeep Kaur, Vesteinn Thorsson, Paul Shannon, Michael H Johnson,

J Christopher Bare, et al. (2007). "A predictive model for transcriptional control of physiology in a free living cell". In: *Cell* 131.7, pp. 1354–1365.

Bonnet, Eric, Eric Viara, Inna Kuperstein, Laurence Calzone, David PA Cohen, Emmanuel Barillot, and Andrei Zinovyev (2015). "NaviCell Web Service for network-based data visualization". In: *Nucleic acids research* 43.W1, W560–W565.

Bori, Zoltan, Zhongfu Zhao, Erika Koltai, Ioannis G Fatouros, Athanasios Z Jamurtas, Ioannis I Douroudos, Gerasimos Terzis, Athanasios Chatzinikolaou, Apostolos Sovatzidis, Dimitrios Draganidis, et al. (2012). "The effects of aging, physical training, and a single bout of exercise on mitochondrial protein expression in human skeletal muscle". In: *Experimental gerontology* 47.6, pp. 417–424.

Borodkina, Aleksandra, Alla Shatrova, Polina Abushik, Nikolay Nikolsky, and Elena Burova (2014). "Interaction between ROS dependent DNA damage, mitochondria and p38 MAPK underlies senescence of human adult stem cells". In: *Aging (Albany NY)* 6.6, p. 481.

Bothwell, John HF and Julian L Griffin (2011). "An introduction to biological nuclear magnetic resonance spectroscopy". In: *Biological Reviews* 86.2, pp. 493–510.

Bracken, Adrian P, Daniela Kleine-Kohlbrecher, Nikolaj Dietrich, Diego Pasini, Gaetano Gargiulo, Chantal Beekman, Kim Theilgaard-Mönch, Saverio Minucci, Bo T Porse, Jean-Christophe Marine, et al. (2007). "The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells". In: *Genes & development* 21.5, pp. 525–530.

Broido, Anna D and Aaron Clauset (2019). "Scale-free networks are rare". In: *Nature communications* 10.1, pp. 1–10.

Brunius, Carl, Lin Shi, and Rikard Landberg (2016). "Large-scale untargeted LC-MS metabolomics data correction using between-batch feature alignment and cluster-based within-batch signal intensity drift correction". In: *Metabolomics* 12.11, p. 173.

Brunton, Steven L, Joshua L Proctor, and J Nathan Kutz (2016). "Discovering governing equations from data by sparse identification of nonlinear dynamical systems". In: *Proceedings of the National Academy of Sciences* 113.15, pp. 3932–3937.

Buenrostro, Jason D, Beijing Wu, Howard Y Chang, and William J Greenleaf (2015). "ATAC-seq: a method for assaying chromatin accessibility genome-wide". In: *Current protocols in molecular biology* 109.1, pp. 21–29.

Calzone, Laurence, Amélie Gelay, Andrei Zinovyev, François Radvanyi, and Emmanuel Barillot (2008). "A comprehensive modular map of molecular interactions in RB/E2F pathway". In: *Molecular systems biology* 4.1.

Camandola, Simonetta, Gabriella Leonarduzzi, Tiziana Musso, Luigi Varesio, Rita Carini, Antonella Scavazza, Elena Chiarpotto, Patrick A Baeuerle, and Giuseppe Poli (1996). "Nuclear factor kB is activated by arachidonic acid but not by eicosapentaenoic acid". In: *Biochemical and biophysical research communications* 229.2, pp. 643–647.

Campisi, Judith and Fabrizio d'Adda di Fagagna (2007). "Cellular senescence: when bad things happen to good cells". In: *Nature reviews Molecular cell biology* 8.9, p. 729.

Candès, Emmanuel J and Michael B Wakin (2008). "An introduction to compressive sampling". In: *IEEE signal processing magazine* 25.2, pp. 21–30.

Candes, Emmanuel J, Justin K Romberg, and Terence Tao (2006). "Stable signal recovery from incomplete and inaccurate measurements". In: *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences* 59.8, pp. 1207–1223.

Carey, Bryce W, Lydia WS Finley, Justin R Cross, C David Allis, and Craig B Thompson (2015). "Intracellular α-ketoglutarate maintains the pluripotency of embryonic stem cells". In: *Nature* 518.7539, p. 413.

Carmona, Juan José and Shaday Michan (2016). "Biology of healthy aging and longevity". In: *Revista de investigacion clinica* 68.1, pp. 7–16.

Carvalho, Benilton S and Rafael A Irizarry (2010). "A framework for oligonucleotide microarray preprocessing". In: *Bioinformatics* 26.19, pp. 2363–2367.

Catalano, Alfonso, Sabrina Rodilossi, Paola Caprari, Vincenzo Coppola, and Antonio Procopio (2005). "5-Lipoxygenase regulates senescence-like growth arrest by promoting ROS-dependent p53 activation". In: *The EMBO journal* 24.1, pp. 170–179.

Cech, Nadja B and Christie G Enke (2001). "Practical implications of some recent studies in electrospray ionization fundamentals". In: *Mass spectrometry reviews* 20.6, pp. 362–387.

Chambers, Matthew C, Brendan Maclean, Robert Burke, Dario Amodei, Daniel L Ruderman, Steffen Neumann, Laurent Gatto, Bernd Fischer, Brian Pratt, Jarrett Egertson, et al. (2012). "A cross-platform toolkit for mass spectrometry and proteomics". In: *Nature biotechnology* 30.10, p. 918.

Chan, Adelyne Sue Li and Masashi Narita (2019). "Short-term gain, long-term pain: the senescence life cycle and cancer". In: *Genes & development* 33.3-4, pp. 127–143.

Chandra, Tamir, Kristina Kirschner, Jean-Yves Thuret, Benjamin D Pope, Tyrone Ryba, Scott Newman, Kashif Ahmed, Shamith A Samarajiwa, Rafik Salama, Thomas Carroll, et al. (2012). "Independence of repressive

histone marks and chromatin compaction during senescent heterochromatic layer formation". In: *Molecular cell* 47.2, pp. 203–214.

Chang, BaoJun, Manabu Nishikawa, Eisuke Sato, Kozo Utsumi, and Masayasu Inoue (2002). "L-Carnitine inhibits cisplatin-induced injury of the kidney and small intestine". In: *Archives of Biochemistry and Biophysics* 405.1, pp. 55–64.

Chang, Chun-Wei, Masayuki Ushio, and Chih-hao Hsieh (2017). "Empirical dynamic modeling for beginners". In: *Ecological research* 32.6, pp. 785–796.

Chang, Young Hwan, Joe W Gray, and Claire J Tomlin (2014). "Exact reconstruction of gene regulatory networks using compressive sensing". In: *BMC bioinformatics* 15.1, p. 400.

Chauvet, Erica, Joseph E Paullet, Joseph P Previte, and Zac Walls (2002). "A Lotka-Volterra three-species food chain". In: *Mathematics magazine* 75.4, pp. 243–255.

Chen, Luonan, Rui Liu, Zhi-Ping Liu, Meiyi Li, and Kazuyuki Aihara (2012). "Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers". In: *Scientific reports* 2, p. 342.

Chen, William W, Birgit Schoeberl, Paul J Jasper, Mario Niepel, Ulrik B Nielsen, Douglas A Lauffenburger, and Peter K Sorger (2009). "Input–output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data". In: *Molecular systems biology* 5.1.

Chicas, Agustin, Xiaowo Wang, Chaolin Zhang, Mila McCurrach, Zhen Zhao, Ozlem Mert, Ross A Dickins, Masashi Narita, Michael Zhang, and Scott W Lowe (2010). "Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence". In: *Cancer cell* 17.4, pp. 376–387.

Chien, Yuchen, Claudio Scuoppo, Xiaowo Wang, Xueping Fang, Brian Balgley, Jessica E Bolden, Prem Premsrirut, Weijun Luo, Agustin Chicas, Cheng S Lee, et al. (2011). "Control of the senescence-associated secretory phenotype by NF-κB promotes senescence and enhances chemosensitivity". In: *Genes & development* 25.20, pp. 2125–2136.

Cho, Kwang-Hyun, Sung-Young Shin, Hyeon-Woo Lee, and Olaf Wolkenhauer (2003). "Investigations into the analysis and modeling of the TNFα-mediated NF-κB-signaling pathway". In: *Genome Research* 13.11, pp. 2413–2422.

Cho, Sung-Hwan, Sang-Min Park, Ho-Sung Lee, Hwang-Yeol Lee, and Kwang-Hyun Cho (2016). "Attractor landscape analysis of colorectal tumorigenesis and its reversion". In: *BMC systems biology* 10.1, p. 96.

Chomsky, Noam (1957). "Syntactic Structures (The Hague: Mouton, 1957)". In: *Review of Verbal Behavior by BF Skinner, Language* 35, pp. 26–58.

Chow, Stirling and Frank Ruskey (2003). "Drawing area-proportional Venn and Euler diagrams". In: *International Symposium on Graph Drawing*. Springer, pp. 466–477.

Coloff, Jonathan L, J Patrick Murphy, Craig R Braun, Isaac S Harris, Laura M Shelton, Kenjiro Kami, Steven P Gygi, Laura M Selfors, and Joan S Brugge (2016). "Differential glutamate metabolism in proliferating and quiescent mammary epithelial cells". In: *Cell metabolism* 23.5, pp. 867–880.

Cong, Le, F Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D Hsu, Xuebing Wu, Wenyan Jiang, Luciano A Marraffini, et al. (2013). "Multiplex genome engineering using CRISPR/Cas systems". In: *Science* 339.6121, pp. 819–823.

Coppé, Jean-Philippe, Pierre-Yves Desprez, Ana Krtolica, and Judith Campisi (2010). "The senescence-associated secretory phenotype: the dark side of tumor suppression". In: *Annual Review of Pathology: Mechanisms of Disease* 5, pp. 99–118.

Correia-Melo, Clara, Francisco DM Marques, Rhys Anderson, Graeme Hewitt, Rachael Hewitt, John Cole, Bernadette M Carroll, Satomi Miwa, Jodie Birch, Alina Merz, et al. (2016). "Mitochondria are required for pro-ageing features of the senescent phenotype". In: *The EMBO journal* 35.7, pp. 724–742.

Cossec, Jack-Christophe, Ilan Theurillat, Claudia Chica, Sabela Búa Aguín, Xavier Gaume, Alexandra Andrieux, Ane Iturbide, Gregory Jouvion, Han Li, Guillaume Bossis, et al. (2018). "SUMO safeguards somatic and pluripotent cell identities by enforcing distinct chromatin states". In: *Cell stem cell* 23.5, pp. 742–757.

Csardi, Gabor and Tamas Nepusz (2006). "The igraph software package for complex network research". In: *InterJournal* Complex Systems, p. 1695. URL: `http://igraph.org`.

Da Silva-Álvarez, Sabela, Jorge Guerra-Varela, Daniel Sobrido-Cameán, Ana Quelle, Antón Barreiro-Iglesias, Laura Sánchez, and Manuel Collado (2019). "Cell senescence contributes to tissue regeneration in zebrafish". In: *Aging cell*.

Dalle Pezze, Piero, Glyn Nelson, Elsje G Otten, Viktor I Korolchuk, Thomas BL Kirkwood, Thomas von Zglinicki, and Daryl P Shanley (2014). "Dynamic modelling of pathways to cellular senescence reveals strategies for targeted interventions". In: *PLoS computational biology* 10.8, e1003728.

Dankort, David, Elena Filenova, Manuel Collado, Manuel Serrano, Kirk Jones, and Martin McMahon (2007). "A new mouse model to explore the initiation, progression, and therapy of BRAFV600E-induced lung tumors". In: *Genes & development* 21.4, pp. 379–384.

Das, Undurti N (2018). "Ageing: is there a role for arachidonic acid and other bioactive lipids? A review". In: *Journal of advanced research* 11, pp. 67–79.

Davis, Tenney L. and Ch'ên Kuo-fu (1941). "The Inner Chapters of Pao-'P'u-tzŭ". In: *Proceedings of the American Academy of Arts and Sciences* 74.10, pp. 297–325. ISSN: 01999818. URL: http://www.jstor.org/stable/20023410.

De Magalhães, João Pedro, Florence Chainiaux, Françoise De Longueville, Véronique Mainfroid, Valérie Migeot, Laurence Marcq, José Remacle, Michel Salmon, and Olivier Toussaint (2004). "Gene expression and regulation in H2O2-induced premature senescence of human foreskin fibroblasts expressing or not telomerase". In: *Experimental gerontology* 39.9, pp. 1379–1389.

Debacq-Chainiaux, F, C Borlon, T Pascal, V Royer, F Eliaers, N Ninane, G Carrard, B Friguet, F de Longueville, S Boffe, J Remacle, and Toussaint O (2005). "Repeated exposure of human skin fibroblasts to UVB at subcytotoxic level triggers premature senescence through the TGF-beta1 signaling pathway". In: *J. Cell Sci* 118, pp. 743–758.

Debacq-Chainiaux, Florence, Emmanuelle Boilan, Jérémie Dedessus Le Moutier, Geoffroy Weemaels, and Olivier Toussaint (2010). "P38 MAPK in the senescence of human and murine fibroblasts". In: *Protein Metabolism and Homeostasis in Aging*. Springer, pp. 126–137.

Del Genio, Charo I, Thilo Gross, and Kevin E Bassler (2011). "All scale-free networks are sparse". In: *Physical review letters* 107.17, p. 178701.

Dhomen, Nathalie, Jorge S Reis-Filho, Silvy da Rocha Dias, Robert Hayward, Kay Savage, Veronique Delmas, Lionel Larue, Catrin Pritchard, and Richard Marais (2009). "Oncogenic Braf induces melanocyte senescence and melanoma in mice". In: *Cancer cell* 15.4, pp. 294–303.

Di Gangi, Iole Maria, Tommaso Mazza, Andrea Fontana, Massimiliano Copetti, Caterina Fusilli, Antonio Ippolito, Fulvio Mattivi, Anna Latiano, Angelo Andriulli, Urska Vrhovsek, et al. (2016). "Metabolomic profile in pancreatic cancer patients: a consensus-based approach to identify highly discriminating metabolites". In: *Oncotarget* 7.5, p. 5815.

Di Leonardo, Aldo, Steven P Linke, Kris Clarkin, and Geoffrey M Wahl (1994). "DNA damage triggers a prolonged p53-dependent G1 arrest and long-term induction of Cip1 in normal human fibroblasts." In: *Genes & development* 8.21, pp. 2540–2551.

Di Micco, Raffaella, Marzia Fumagalli, Angelo Cicalese, Sara Piccinin, Patrizia Gasparini, Chiara Luise, Catherine Schurra, Paolo Giovanni Nuciforo, Aaron Bensimon, Roberta Maestro, et al. (2006). "Oncogene-induced

senescence is a DNA damage response triggered by DNA hyper-replication". In: *Nature* 444.7119, p. 638.

Diederichs, Sven, Lorenz Bartsch, Julia C Berkmann, Karin Fröse, Jana Heitmann, Caroline Hoppe, Deetje Iggena, Danny Jazmati, Philipp Karschnia, Miriam Linsenmeier, et al. (2016). "The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations". In: *EMBO molecular medicine* 8.5, pp. 442–457.

Dimri, Goberdhan P, Koji Itahana, Meileen Acosta, and Judith Campisi (2000). "Regulation of a senescence checkpoint response by the E2F1 transcription factor and p14ARF tumor suppressor". In: *Molecular and cellular biology* 20.1, pp. 273–285.

Dou, Zhixun, Kanad Ghosh, Maria Grazia Vizioli, Jiajun Zhu, Payel Sen, Kirk J Wangensteen, Johayra Simithy, Yemin Lan, Yanping Lin, Zhuo Zhou, et al. (2017). "Cytoplasmic chromatin triggers inflammation in senescence and cancer". In: *Nature* 550.7676, p. 402.

Downward, Julian (2003). "Targeting RAS signalling pathways in cancer therapy". In: *Nature Reviews Cancer* 3.1, p. 11.

Dunn, Jonathan (2017). "Computational learning of construction grammars". In: *Language and Cognition* 9.2, pp. 254–292.

Dynlacht, Brian David, Osvaldo Flores, Jacqueline A Lees, and Ed Harlow (1994). "Differential regulation of E2F transactivation by cyclin/cdk2 complexes." In: *Genes & development* 8.15, pp. 1772–1786.

D'Ulizia, Arianna, Fernando Ferri, and Patrizia Grifoni (2011). "A survey of grammatical inference methods for natural language learning". In: *Artificial Intelligence Review* 36.1, pp. 1–27.

Eschbach, Thomas, Wolfgang Günther, and Bernd Becker (2006). "Orthogonal Hypergraph Drawing for Improved Visibility." In: *Journal of Graph Algorithms and Applications* 10.2, pp. 141–157.

Fagagna, Fabrizio d'Adda di (2008). "Living on a break: cellular senescence as a DNA-damage response". In: *Nature Reviews Cancer* 8.7, p. 512.

Faget, Douglas V, Qihao Ren, and Sheila A Stewart (2019). "Unmasking senescence: context-dependent effects of SASP in cancer". In: *Nature Reviews Cancer* 19.8, pp. 439–453.

Flanagan, Judith L, Peter A Simmons, Joseph Vehige, Mark DP Willcox, and Qian Garrett (2010). "Role of carnitine in disease". In: *Nutrition & metabolism* 7.1, p. 30.

Florin, Lore, Lars Hummerich, Bernd Thilo Dittrich, Felix Kokocinski, Gunnar Wrobel, Sabine Gack, Marina Schorpp-Kistner, Sabine Werner, Meinhard Hahn, Peter Lichter, et al. (2004). "Identification of novel AP-1 target genes

in fibroblasts regulated during cutaneous wound healing". In: *Oncogene* 23.42, p. 7005.

Forma, Ewa, Paweł Jóźwiak, Magdalena Bryś, and Anna Krześlak (2014). "The potential role of O-GlcNAc modification in cancer epigenetics". In: *Cellular & molecular biology letters* 19.3, p. 438.

Friedman, David B and Thomas E Johnson (1988). "A mutation in the age-1 gene in Caenorhabditis elegans lengthens life and reduces hermaphrodite fertility." In: *Genetics* 118.1, pp. 75–86.

Frippiat, Christophe, Janique Dewelle, Jos Remacle, and Olivier Toussaint (2002). "Signal transduction in H2O2-induced senescence-like phenotype in human diploid fibroblasts". In: *Free Radical Biology and Medicine* 33.10, pp. 1334–1346.

Fuhrmann-Stroissnigg, Heike, Yuan Yuan Ling, Jing Zhao, Sara J McGowan, Yi Zhu, Robert W Brooks, Diego Grassi, Siobhan Q Gregg, Jennifer L Stripay, Akaitz Dorronsoro, et al. (2017). "Identification of HSP90 inhibitors as a novel class of senolytics". In: *Nature communications* 8.1, p. 422.

Fukazawa, Ryuji, Ei Ikegam, Miki Watanabe, Miharu Hajikano, Mitsuhiro Kamisago, Yasuhiro Katsube, Hitoshi Yamauchi, Masami Ochi, and Shunichi Ogawa (2007). "Coronary artery aneurysm induced by Kawasaki disease in children show features typical senescence". In: *Circulation Journal* 71.5, pp. 709–715.

Gabai, Vladimir L, Julia A Yaglom, Todd Waldman, and Michael Y Sherman (2009). "Heat shock protein Hsp72 controls oncogene-induced senescence pathways in cancer cells". In: *Molecular and cellular biology* 29.2, pp. 559–569.

Gal, Hilah, Marina Lysenko, Sima Stroganov, Ezra Vadai, Sameh A Youssef, Keren Tzadikevitch-Geffen, Ron Rotkopf, Tal Biron-Shental, Alain de Bruin, Michal Neeman, et al. (2019). "Molecular pathways of senescence regulate placental structure and function". In: *The EMBO journal* 38.18.

Gallagher, Emily Jane and Derek LeRoith (2011). "Diabetes, cancer, and metformin: connections of metabolism and cell proliferation". In: *Annals of the New York Academy of Sciences* 1243.1, pp. 54–68.

Gallo, Giorgio, Giustino Longo, Stefano Pallottino, and Sang Nguyen (1993). "Directed hypergraphs and applications". In: *Discrete applied mathematics* 42.2-3, pp. 177–201.

Galvis, Daniel, Darren Walsh, Lorna W Harries, Eva Latorre, and James Rankin (2019). "A dynamical systems model for the measurement of cellular senescence". In: *Journal of the Royal Society Interface* 16.159, p. 20190311.

Garber, Manuel, Nir Yosef, Alon Goren, Raktima Raychowdhury, Anne Thielke, Mitchell Guttman, James Robinson, Brian Minie, Nicolas Chevrier, Zohar Itzhaki, et al. (2012). "A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals". In: *Molecular cell* 47.5, pp. 810–822.

García-Prat, Laura, Marta Martínez-Vicente, Eusebio Perdiguero, Laura Ortet, Javier Rodríguez-Ubreva, Elena Rebollo, Vanessa Ruiz-Bonilla, Susana Gutarra, Esteban Ballestar, Antonio L Serrano, et al. (2016). "Autophagy maintains stemness by preventing senescence". In: *Nature* 529.7584, p. 37.

Gong, Wuming, Naoko Koyano-Nakagawa, Tongbin Li, and Daniel J Garry (2015). "Inferring dynamic gene regulatory networks in cardiac differentiation through the integration of multi-dimensional data". In: *BMC bioinformatics* 16.1, p. 74.

Goode, Debbie K, Nadine Obier, MS Vijayabaskar, Michael Lie-A-Ling, Andrew J Lilly, Rebecca Hannah, Monika Lichtinger, Kiran Batta, Magdalena Florkowska, Rahima Patel, et al. (2016). "Dynamic gene regulatory networks drive hematopoietic specification and differentiation". In: *Developmental cell* 36.5, pp. 572–587.

Gorgoulis, Vassilis, Peter D Adams, Andrea Alimonti, Dorothy C Bennett, Oliver Bischof, Cleo Bishop, Judith Campisi, Manuel Collado, Konstantinos Evangelou, Gerardo Ferbeyre, et al. (2019). "Cellular senescence: defining a path forward". In: *Cell* 179.4, pp. 813–827.

Gorgoulis, Vassilis G and Thanos D Halazonetis (2010). "Oncogene-induced senescence: the bright and dark side of the response". In: *Current opinion in cell biology* 22.6, pp. 816–827.

Gorrochategui, Eva, Joaquim Jaumot, Sílvia Lacorte, and Romà Tauler (2016). "Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: overview and workflow". In: *TrAC Trends in Analytical Chemistry* 82, pp. 425–442.

Grabowska, Wioleta, Ewa Sikora, and Anna Bielak-Zmijewska (2017). "Sirtuins, a promising target in slowing down the ageing process". In: *Biogerontology* 18.4, pp. 447–476.

Grandison, Richard C, Matthew DW Piper, and Linda Partridge (2009). "Amino-acid imbalance explains extension of lifespan by dietary restriction in Drosophila". In: *Nature* 462.7276, p. 1061.

Gromski, Piotr S, Howbeer Muhamadali, David I Ellis, Yun Xu, Elon Correa, Michael L Turner, and Royston Goodacre (2015). "A tutorial review: Metabolomics and partial least squares-discriminant analysis–a marriage of convenience or a shotgun wedding". In: *Analytica chimica acta* 879, pp. 10–23.

Gu, Wei and Robert G Roeder (1997). "Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain". In: *Cell* 90.4, pp. 595–606.

Gu, Zuguang, Lei Gu, Roland Eils, Matthias Schlesner, and Benedikt Brors (2014). "circlize implements and enhances circular visualization in R". In: *Bioinformatics* 30 (19), pp. 2811–2812.

Gu, Zuguang, Roland Eils, and Matthias Schlesner (2016). "Complex heatmaps reveal patterns and correlations in multidimensional genomic data". In: *Bioinformatics*.

Guimera, Alvaro Martinez, Ciaran Welsh, Piero Dalle Pezze, Nicola Fullard, Glyn Nelson, Mathilde F Roger, Stefan A Przyborski, and Daryl P Shanley (2017). "Systems modelling ageing: from single senescent cells to simple multi-cellular models". In: *Essays in biochemistry* 61.3, pp. 369–377.

Gyongyosi, Laszlo and Sandor Imre (2019). "Dense Quantum Measurement Theory". In: *Scientific reports* 9.1, p. 6755.

Hackett, Sean R, Edward A Baltz, Marc Coram, Bernd J Wranik, Griffin Kim, Adam Baker, Minjie Fan, David G Hendrickson, Marc Berndl, and R Scott McIsaac (2019). "Time-resolved genome-scale profiling reveals a causal expression network". In: *bioRxiv*, p. 619577.

Hardie, D Grahame, Fiona A Ross, and Simon A Hawley (2012). "AMPK: a nutrient and energy sensor that maintains energy homeostasis". In: *Nature reviews Molecular cell biology* 13.4, p. 251.

Harley, Calvin B, A Bruce Futcher, and Carol W Greider (1990). "Telomeres shorten during ageing of human fibroblasts". In: *Nature* 345.6274, p. 458.

Harrison, David E, Randy Strong, Zelton Dave Sharp, James F Nelson, Clinton M Astle, Kevin Flurkey, Nancy L Nadon, J Erby Wilkinson, Krystyna Frenkel, Christy S Carter, et al. (2009). "Rapamycin fed late in life extends lifespan in genetically heterogeneous mice". In: *nature* 460.7253, p. 392.

Hashimoto, Teppei, Makoto Horikawa, Toshihisa Nomura, and Kazuichi Sakamoto (2010). "Nicotinamide adenine dinucleotide extends the lifespan of Caenorhabditis elegans mediated by sir-2.1 and daf-16". In: *Biogerontology* 11.1, p. 31.

Hayflick, Leonard and Paul S Moorhead (1961). "The serial cultivation of human diploid cell strains". In: *Experimental cell research* 25.3, pp. 585–621.

He, Qing, Yuhao Gao, Tongxing Wang, Lujun Zhou, Wenxia Zhou, and Zengqiang Yuan (2019). "Deficiency of Yes-associated protein induces cataract in mice". In: *Aging and disease* 10.2, p. 293.

He, Shenghui and Norman E Sharpless (2017). "Senescence in health and disease". In: *Cell* 169.6, pp. 1000–1011.

Helin, Kristian (1998). "Regulation of cell proliferation by the E2F transcription factors". In: *Current opinion in genetics & development* 8.1, pp. 28–35.

Hennig, Christian (2019). *fpc: Flexible Procedures for Clustering*. R package version 2.2-3. URL: https://CRAN.R-project.org/package=fpc.

Herbig, Utz, Wendy A Jobling, Benjamin PC Chen, David J Chen, and John M Sedivy (2004). "Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21CIP1, but not p16INK4a". In: *Molecular cell* 14.4, pp. 501–513.

Hernandez-Segura, Alejandra, Jamil Nehme, and Marco Demaria (2018). "Hallmarks of cellular senescence". In: *Trends in cell biology* 28.6, pp. 436–453.

Herranz, Nicolás, Suchira Gallage, Massimiliano Mellone, Torsten Wuestefeld, Sabrina Klotz, Christopher J Hanley, Selina Raguz, Juan Carlos Acosta, Andrew J Innes, Ana Banito, et al. (2015). "mTOR regulates MAPKAPK2 translation to control the senescence-associated secretory phenotype". In: *Nature cell biology* 17.9, p. 1205.

Hoare, Matthew and Masashi Narita (2018). "The power behind the throne: senescence and the hallmarks of cancer". In: *Annual Review of Cancer Biology* 2, pp. 175–194.

Hoare, Matthew, Yoko Ito, Tae-Won Kang, Michael P Weekes, Nicholas J Matheson, Daniel A Patten, Shishir Shetty, Aled J Parry, Suraj Menon, Rafik Salama, et al. (2016). "NOTCH1 mediates a switch between two distinct secretomes during senescence". In: *Nature cell biology* 18.9, p. 979.

Hoffmann, Alexander, Andre Levchenko, Martin L Scott, and David Baltimore (2002). "The IκB-NF-κB signaling module: temporal control and selective gene activation". In: *Science* 298.5596, pp. 1241–1245.

Huang, Sui (2012). "The molecular and mathematical basis of Waddington's epigenetic landscape: A framework for post-Darwinian biology?" In: *Bioessays* 34.2, pp. 149–157.

Huang, Sui, Gabriel Eichler, Yaneer Bar-Yam, and Donald E Ingber (2005). "Cell fates as high-dimensional attractor states of a complex gene regulatory network". In: *Physical review letters* 94.12, p. 128701.

Hubackova, Sona, Katerina Krejcikova, Jiri Bartek, and Zdenek Hodny (2012). "IL1-and TGFβ-Nox4 signaling, oxidative stress and DNA damage response are shared features of replicative, oncogene-induced, and drug-induced paracrine 'bystander senescence'". In: *Aging (Albany NY)* 4.12, p. 932.

Inoue, Kentaro, Hisaaki Shinohara, Marcelo Behar, Noriko Yumoto, Gouhei Tanaka, Alexander Hoffmann, Kazuyuki Aihara, and Mariko Okada-Hatakeyama (2016). "Oscillation dynamics underlie functional switching

of NF-κB for B-cell activation". In: *NPJ Systems Biology and Applications* 2, p. 16024.

Ito, Yoko, Matthew Hoare, and Masashi Narita (2017). "Spatial and temporal control of senescence". In: *Trends in cell biology* 27.11, pp. 820–832.

James, Leighton R, Damu Tang, Alistair Ingram, Hao Ly, Kerri Thai, Lu Cai, and James W Scholey (2002). "Flux through the Hexosamine pathway is a determinant of nuclear factor κB–dependent promoter activation". In: *Diabetes* 51.4, pp. 1146–1156.

Jiang, Fang, Zongqi Zhang, Yi Zhang, Jianping Wu, Li Yu, and Su Liu (2016). "L-carnitine ameliorates the liver inflammatory response by regulating carnitine palmitoyltransferase I-dependent PPARγ signaling". In: *Molecular medicine reports* 13.2, pp. 1320–1328.

Johnson, David G and CL Walker (1999). "Cyclins and cell cycle checkpoints". In: *Annual review of pharmacology and toxicology* 39.1, pp. 295–312.

Johnson, Simon C, Peter S Rabinovitch, and Matt Kaeberlein (2013). "mTOR is a key modulator of ageing and age-related disease". In: *Nature* 493.7432, p. 338.

Johnson, W Evan, Cheng Li, and Ariel Rabinovic (2007). "Adjusting batch effects in microarray expression data using empirical Bayes methods". In: *Biostatistics* 8.1, pp. 118–127.

Jones, Eric, Travis Oliphant, Pearu Peterson, et al. (2001). *SciPy: Open source scientific tools for Python.* URL: `http://www.scipy.org/`.

Jones, Russell G, David R Plas, Sara Kubek, Monica Buzzai, James Mu, Yang Xu, Morris J Birnbaum, and Craig B Thompson (2005). "AMP-activated protein kinase induces a p53-dependent metabolic checkpoint". In: *Molecular cell* 18.3, pp. 283–293.

Jothi, Raja, S Balaji, Arthur Wuster, Joshua A Grochow, Jörg Gsponer, Teresa M Przytycka, L Aravind, and M Madan Babu (2009). "Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture". In: *Molecular systems biology* 5.1, p. 294.

Justice, Jamie N, Anoop M Nambiar, Tamar Tchkonia, Nathan K LeBrasseur, Rodolfo Pascual, Shahrukh K Hashmi, Larissa Prata, Michal M Masternak, Stephen B Kritchevsky, Nicolas Musi, et al. (2019). "Senolytics in idiopathic pulmonary fibrosis: Results from a first-in-human, open-label, pilot study". In: *EBioMedicine* 40, pp. 554–563.

Juven-Gershon, Tamar and Moshe Oren (1999). "Mdm2: the ups and downs". In: *Molecular medicine* 5.2, pp. 71–83.

Kaiser, Eurika, J Nathan Kutz, and Steven L Brunton (2018). "Sparse identification of nonlinear dynamics for model predictive control in the low-data limit". In: *Proceedings of the Royal Society A* 474.2219, p. 20180335.

Kang, Chanhee, Qikai Xu, Timothy D Martin, Mamie Z Li, Marco Demaria, Liviu Aron, Tao Lu, Bruce A Yankner, Judith Campisi, and Stephen J Elledge (2015). "The DNA damage response induces inflammation and senescence by inhibiting autophagy of GATA4". In: *Science* 349.6255, aaa5612.

Kannan, Meenakshi B, Vera Solovieva, and Volker Blank (2012). "The small MAF transcription factors MAFF, MAFG and MAFK: current knowledge and perspectives". In: *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1823.10, pp. 1841–1846.

Kaochar, Salma and Benjamin P Tu (2012). "Gatekeepers of chromatin: small metabolites elicit big changes in gene expression". In: *Trends in biochemical sciences* 37.11, pp. 477–483.

Karin, Michael, Zheng-gang Liu, and Ebrahim Zandi (1997). "AP-1 function and regulation". In: *Current opinion in cell biology* 9.2, pp. 240–246.

Karin, Omer, Amit Agrawal, Ziv Porat, Valery Krizhanovsky, and Uri Alon (2019). "Senescent cell turnover slows with age providing an explanation for the Gompertz law". In: *Nature Communications* 10.1, pp. 1–9.

Kassambara, Alboukadel and Fabian Mundt (2017). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R package version 1.0.5. URL: `https://CRAN.R-project.org/package=factoextra`.

Kebarle, Paul and Udo H Verkerk (2009). "Electrospray: from ions in solution to ions in the gas phase, what we know now". In: *Mass spectrometry reviews* 28.6, pp. 898–917.

Kenyon, Cynthia J (2010). "The genetics of ageing". In: *Nature* 464.7288, p. 504.

Khan, Aziz, Oriol Fornes, Arnaud Stigliani, Marius Gheorghe, Jaime A Castro-Mondragon, Robin van der Lee, Adrien Bessy, Jeanne Cheneby, Shubhada R Kulkarni, Ge Tan, et al. (2017). "JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework". In: *Nucleic acids research* 46.D1, pp. D260–D266.

Kim, Mi Young, Tong Zhang, and W Lee Kraus (2005). "Poly (ADP-ribosyl) ation by PARP-1:PAR-laying'NAD+ into a nuclear signal". In: *Genes & development* 19.17, pp. 1951–1967.

Kim, William Y and Norman E Sharpless (2006). "The regulation of INK4/ARF in cancer and aging". In: *Cell* 127.2, pp. 265–275.

Kim, Yunseong, Sea Choi, Dongkwan Shin, and Kwang-Hyun Cho (2017). "Quantitative evaluation and reversion analysis of the attractor landscapes of an intracellular regulatory network for colorectal cancer". In: *BMC systems biology* 11.1, p. 45.

Klement, Karolin and Aaron A Goodarzi (2014). "DNA double strand break responses and chromatin alterations within the aging cell". In: *Experimental cell research* 329.1, pp. 42–52.

Knaap, Jan A van der and C Peter Verrijzer (2016). "Undercover: gene control by metabolites and metabolic enzymes". In: *Genes & development* 30.21, pp. 2345–2369.

Kooistra, Susanne Marije and Kristian Helin (2012). "Molecular mechanisms and potential functions of histone demethylases". In: *Nature reviews Molecular cell biology* 13.5, p. 297.

Krishnamurthy, Janakiraman, Chad Torrice, Matthew R Ramsey, Grigoriy I Kovalev, Khalid Al-Regaiey, Lishan Su, and Norman E Sharpless (2004). "Ink4a/Arf expression is a biomarker of aging". In: *The Journal of clinical investigation* 114.9, pp. 1299–1307.

Kuilman, Thomas, Chrysiis Michaloglou, Liesbeth CW Vredeveld, Sirith Douma, Remco van Doorn, Christophe J Desmet, Lucien A Aarden, Wolter J Mooi, and Daniel S Peeper (2008). "Oncogene-induced senescence relayed by an interleukin-dependent inflammatory network". In: *Cell* 133.6, pp. 1019–1031.

Kurtzer, Gregory M, Vanessa Sochat, and Michael W Bauer (2017). "Singularity: Scientific containers for mobility of compute". In: *PloS one* 12.5.

La Manno, Gioele, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E Kastriti, Peter Lönnerberg, Alessandro Furlan, et al. (2018). "RNA velocity of single cells". In: *Nature* 560.7719, p. 494.

Laberge, Remi-Martin, Yu Sun, Arturo V Orjalo, Christopher K Patil, Adam Freund, Lili Zhou, Samuel C Curran, Albert R Davalos, Kathleen A Wilson-Edell, Su Liu, et al. (2015). "MTOR regulates the pro-tumorigenic senescence-associated secretory phenotype by promoting IL1A translation". In: *Nature cell biology* 17.8, p. 1049.

Lane, David P (1992). "Cancer. p53, guardian of the genome". In: *Nature* 358, pp. 15–16.

Lang, Duncan Temple (2014). *RJSONIO: Serialize R objects to JSON, JavaScript Object Notation*. R package version 1.3-0. URL: `https://CRAN.R-project.org/package=RJSONIO`.

Langfelder, Peter and Steve Horvath (2008). "WGCNA: an R package for weighted correlation network analysis". In: *BMC bioinformatics* 9.1, p. 559.

Langley, Emma, Mark Pearson, Mario Faretta, Uta-Maria Bauer, Roy A Frye, Saverio Minucci, Pier Giuseppe Pelicci, and Tony Kouzarides (2002). "Human SIR2 deacetylates p53 and antagonizes PML/p53-induced cellular senescence". In: *The EMBO journal* 21.10, pp. 2383–2396.

Lara-Astiaso, David, Assaf Weiner, Erika Lorenzo-Vivas, Irina Zaretsky, Diego Adhemar Jaitin, Eyal David, Hadas Keren-Shaul, Alexander Mildner, Deborah Winter, Steffen Jung, et al. (2014). "Chromatin state dynamics during blood formation". In: *science* 345.6199, pp. 943–949.

Lascaris, Romeo F, Ellen de Groot, Peter-Bram't Hoen, Willem H Mager, and Rudi J Planta (2000). "Different roles for abf1p and a T-rich promoter element in nucleosome organization of the yeast RPS28A gene". In: *Nucleic acids research* 28.6, pp. 1390–1396.

Leach, Damien A, Vasilios Panagopoulos, Claire Nash, Charlotte Bevan, Axel A Thomson, Luke A Selth, and Grant Buchanan (2017). "Cell-lineage specificity and role of AP-1 in the prostate fibroblast androgen receptor cistrome". In: *Molecular and cellular endocrinology* 439, pp. 261–272.

Lee, Andrew C, Brett E Fenster, Hideki Ito, Kazuyo Takeda, Nancy S Bae, Tazuko Hirai, Zu-Xi Yu, Victor J Ferrans, Bruce H Howard, and Toren Finkel (1999). "Ras proteins induce senescence by altering the intracellular levels of reactive oxygen species". In: *Journal of Biological Chemistry* 274.12, pp. 7936–7940.

Lee, Bo Yun, Jung A Han, Jun Sub Im, Amelia Morrone, Kimberly Johung, Edward C Goodwin, Wim J Kleijer, Daniel DiMaio, and Eun Seong Hwang (2006). "Senescence-associated β-galactosidase is lysosomal β-galactosidase". In: *Aging cell* 5.2, pp. 187–195.

Lee, Seung-Min, So Hee Dho, Sung-Kyu Ju, Jin-Soo Maeng, Jeong-Yoon Kim, and Ki-Sun Kwon (2012). "Cytosolic malate dehydrogenase regulates senescence in human fibroblasts". In: *Biogerontology* 13.5, pp. 525–536.

Lenhard, Boris, Albin Sandelin, and Piero Carninci (2012). "Metazoan promoters: emerging characteristics and insights into transcriptional regulation". In: *Nature Reviews Genetics* 13.4, pp. 233–245.

Lewis, Brian A and John A Hanover (2014). "O-GlcNAc and the epigenetic regulation of gene expression". In: *Journal of Biological Chemistry* 289.50, pp. 34440–34448.

Li, Richard Y, Rosa Di Felice, Remo Rohs, and Daniel A Lidar (2018). "Quantum annealing versus classical machine learning applied to a simplified computational biology problem". In: *NPJ quantum information* 4.1, p. 14.

Li, Xi, Henrik Molina, Haiyan Huang, You-you Zhang, Mei Liu, Shu-wen Qian, Chad Slawson, Wagner B Dias, Akhilesh Pandey, Gerald W Hart, et al. (2009). "O-Linked N-Acetylglucosamine Modification on CCAAT Enhancer-binding Protein β ROLE DURING ADIPOCYTE DIFFERENTIATION". In: *Journal of Biological Chemistry* 284.29, pp. 19248–19254.

Li, Yifeng, Chih-yu Chen, Alice M Kaye, and Wyeth W Wasserman (2015). "The identification of cis-regulatory elements: A review from a machine learning perspective". In: *Biosystems* 138, pp. 6–17.

Libiseller, Gunnar, Michaela Dvorzak, Ulrike Kleb, Edgar Gander, Tobias Eisenberg, Frank Madeo, Steffen Neumann, Gert Trausinger, Frank Sinner, Thomas Pieber, et al. (2015). "IPO: a tool for automated optimization of XCMS parameters". In: *BMC bioinformatics* 16.1, p. 118.

Lipniacki, Tomasz, Pawel Paszek, Allan R Brasier, Bruce Luxon, and Marek Kimmel (2004). "Mathematical model of NF-κB regulatory module". In: *Journal of theoretical biology* 228.2, pp. 195–215.

Liu, Edison T and Douglas A Lauffenburger (2009). *Systems biomedicine: concepts and perspectives*. Academic press.

Liu, Shaojuan, Liuqin He, and Kang Yao (2018a). "The antioxidative function of alpha-ketoglutarate and its applications". In: *BioMed research international* 2018.

Liu, Ting, Lingyun Zhang, Donghyun Joo, and Shao-Cong Sun (2017). "NF-κB signaling in inflammation". In: *Signal transduction and targeted therapy* 2, p. 17023.

Liu, Xue-ling, Jian Ding, and Ling-hua Meng (2018b). "Oncogene-induced senescence: a double edged sword in cancer". In: *Acta Pharmacologica Sinica* 39.10, p. 1553.

Liu, Yan, Hanna K Sanoff, Hyunsoon Cho, Christin E Burd, Chad Torrice, Joseph G Ibrahim, Nancy E Thomas, and Norman E Sharpless (2009). "Expression of p16INK4a in peripheral blood T-cells is a biomarker of human aging". In: *Aging cell* 8.4, pp. 439–448.

Lomvardas, Stavros and Dimitris Thanos (2002). "Modifying gene expression programs by altering core promoter chromatin architecture". In: *Cell* 110.2, pp. 261–271.

Lopes-Paciencia, Stéphane, Emmanuelle Saint-Germain, Marie-Camille Rowell, Ana Fernández Ruiz, Paloma Kalegari, and Gerardo Ferbeyre (2019). "The senescence-associated secretory phenotype and its regulation". In: *Cytokine* 117, pp. 15–22.

López-Otín, Carlos, Maria A Blasco, Linda Partridge, Manuel Serrano, and Guido Kroemer (2013). "The hallmarks of aging". In: *Cell* 153.6, pp. 1194–1217.

Lorenzini, A, S Hrelia, A Bordoni, P Biagi, L Frisoni, T Marinucci, and VJ Cristofalo (2001). "Is increased arachidonic acid release a cause or a consequence of replicative senescence?" In: *Experimental gerontology* 36.1, pp. 65–78.

Luscombe, Nicholas M, Susan E Austin, Helen M Berman, and Janet M Thornton (2000). "An overview of the structures of protein-DNA complexes". In: *Genome biology* 1.1, reviews001–1.

MacDonald, James W. (2017). *pd.hta.2.0: Platform Design Info for Affymetrix HTA-2.0*. R package version 3.12.2.

Macip, Salvador, Makoto Igarashi, Li Fang, Angus Chen, Zhen-Qiang Pan, Sam W Lee, and Stuart A Aaronson (2002). "Inhibition of p21-mediated ROS accumulation can rescue p21-induced senescence". In: *The EMBO journal* 21.9, pp. 2180–2188.

Mackay, Gillian M, Liang Zheng, Niels JF Van Den Broek, and Eyal Gottlieb (2015). "Analysis of cell metabolism using LC-MS and isotope tracers". In: *Methods in enzymology*. Vol. 561. Elsevier, pp. 171–196.

Maclaine, Nicola J and Ted R Hupp (2009). "The regulation of p53 by phosphorylation: a model for how distinct signals integrate into the p53 pathway". In: *Aging* 1.5, p. 490.

Madiraju, Padma, Shri V Pande, Marc Prentki, and SR Murthy Madiraju (2009). "Mitochondrial acetylcarnitine provides acetyl groups for nuclear histone acetylation". In: *Epigenetics* 4.6, pp. 399–403.

Malumbres, Marcos and Mariano Barbacid (2009). "Cell cycle, CDKs and cancer: a changing paradigm". In: *Nature reviews cancer* 9.3, p. 153.

Mangan, Niall M, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz (2016). "Inferring biological networks by sparse identification of nonlinear dynamics". In: *IEEE Transactions on Molecular, Biological and Multi-Scale Communications* 2.1, pp. 52–63.

Martin, James A, Thomas D Brown, Anneliese D Heiner, and Joseph A Buckwalter (2004). "Chondrocyte senescence, joint loading and osteoarthritis." In: *Clinical Orthopaedics and Related Research (1976-2007)* 427, S96–S103.

Martínez-Zamudio, Ricardo Iván, Pierre-François Roux, José Américo NLF de Freitas, Lucas Robinson, Gregory Doré, Bin Sun, Jesús Gil, Utz Herbig, and Oliver Bischof (2019). "AP-1 Imprints a Reversible Transcriptional Program of Senescent Cells". In: *bioRxiv*, p. 633594.

Mattison, Julie A, Ricki J Colman, T Mark Beasley, David B Allison, Joseph W Kemnitz, George S Roth, Donald K Ingram, Richard Weindruch, Rafael De Cabo, and Rozalyn M Anderson (2017). "Caloric restriction improves health and survival of rhesus monkeys". In: *Nature communications* 8, p. 14063.

Mayran, Alexandre and Jacques Drouin (2018). "Pioneer transcription factors shape the epigenetic landscape". In: *Journal of Biological Chemistry* 293.36, pp. 13795–13804.

McCay, Carl M, Mary F Crowell, and Lewis A Maynard (1935). "The effect of retarded growth upon the length of life span and upon the ultimate body size: one figure". In: *The journal of Nutrition* 10.1, pp. 63–79.

McHugh, Domhnall and Jesús Gil (2018). "Senescence and aging: causes, consequences, and therapeutic avenues". In: *J Cell Biol* 217.1, pp. 65–77.

Meza, Magno Enrique Mendoza, Amit Bhaya, and Eugenius Kaszkurewicz (2005). "Controller design techniques for the Lotka-Volterra nonlinear system". In: *Sba: Controle & Automação Sociedade Brasileira de Automatica* 16.2, pp. 124–135.

Michaloglou, Chrysiis, Liesbeth CW Vredeveld, Maria S Soengas, Christophe Denoyelle, Thomas Kuilman, Chantal MAM Van Der Horst, Donné M Majoor, Jerry W Shay, Wolter J Mooi, and Daniel S Peeper (2005). "BRAF E600-associated senescence-like cell cycle arrest of human naevi". In: *Nature* 436.7051, p. 720.

Mirzaei, Habibollah, Nastaran Khodadad, Chiman Karami, Roya Pirmoradi, and Sayyad Khanizadeh (2019). "The AP-1 pathway; A key regulator of cellular transformation modulated by oncogenic viruses". In: *Reviews in Medical Virology*.

Moiseeva, Olga, Véronique Bourdeau, Antoine Roux, Xavier Deschênes-Simard, and Gerardo Ferbeyre (2009). "Mitochondrial dysfunction contributes to oncogene-induced senescence". In: *Molecular and cellular biology* 29.16, pp. 4495–4507.

Mojtahedi, Mitra, Alexander Skupin, Joseph Zhou, Ivan G Castaño, Rebecca YY Leong-Quong, Hannah Chang, Kalliopi Trachana, Alessandro Giuliani, and Sui Huang (2016). "Cell fate decision as high-dimensional critical state transition". In: *PLoS biology* 14.12, e2000640.

Mombach, José CM, Cristhian A Bugs, and Claudine Chaouiya (2014). "Modelling the onset of senescence at the G1/S cell cycle checkpoint". In: *BMC genomics* 15.7, S7.

Moskalev, Alexey, Alexander Aliper, Zeljka Smit-McBride, Anton Buzdin, and Alex Zhavoronkov (2014). "Genetics and epigenetics of aging and longevity". In: *Cell Cycle* 13.7, pp. 1063–1077.

Mothes, Janina, Dorothea Busse, Bente Kofahl, and Jana Wolf (2015). "Sources of dynamic variability in NF-κB signal transduction: A mechanistic model". In: *BioEssays* 37.4, pp. 452–462.

Muñoz-Espín, Daniel and Manuel Serrano (2014). "Cellular senescence: from physiology to pathology". In: *Nature reviews Molecular cell biology* 15.7, p. 482.

Narita, Masako, Andrew RJ Young, Satoko Arakawa, Shamith A Samarajiwa, Takayuki Nakashima, Sei Yoshida, Sungki Hong, Lorraine S Berry, Stefanie Reichelt, Manuela Ferreira, et al. (2011). "Spatial coupling of mTOR and

autophagy augments secretory phenotypes". In: *Science* 332.6032, pp. 966–970.

Narita, Masashi, Sabrina Nuñez, Edith Heard, Masako Narita, Athena W Lin, Stephen A Hearn, David L Spector, Gregory J Hannon, and Scott W Lowe (2003). "Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence". In: *Cell* 113.6, pp. 703–716.

Nelson, Glyn, Olena Kucheryavenko, James Wordsworth, and Thomas von Zglinicki (2018). "The senescent bystander effect is caused by ROS-activated NF-κB signalling". In: *Mechanisms of ageing and development* 170, pp. 30–36.

Nguyen, Tu, Brian James Kirsch, Ryoichi Asaka, Karim Nabi, Addison Quinones, Jessica Tan, Marjorie Justine Antonio, Felipe Camelo, Ting Li, Stephanie Nguyen, et al. (2019). "Uncovering the role of N-acetyl-aspartyl-glutamate as a glutamate reservoir in cancer". In: *Cell reports* 27.2, pp. 491–501.

Nitzan, Mor, Jose Casadiego, and Marc Timme (2017). "Revealing physical interaction networks from statistics of collective dynamics". In: *Science advances* 3.2, e1600396.

Ohanna, Mickaël, Sandy Giuliano, Caroline Bonet, Véronique Imbert, Véronique Hofman, Joséphine Zangari, Karine Bille, Caroline Robert, Brigitte Bressac-de Paillerets, Paul Hofman, et al. (2011). "Senescent cells develop a PARP-1 and nuclear factor-κB-associated secretome (PNAS)". In: *Genes & development* 25.12, pp. 1245–1261.

Oksanen, Jari, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, and Helene Wagner (2019). *vegan: Community Ecology Package*. R package version 2.5-6. URL: `https://CRAN.R-project.org/package=vegan`.

Ono, Keiichiro, Tanja Muetze, Georgi Kolishovski, Paul Shannon, and Barry Demchak (2015). "CyREST: Turbocharging Cytoscape Access for External Tools via a RESTful API". In: *F1000Research* 4.478. DOI: `10.12688/f1000research.6767.1`.

Özcan, Sabire, Sreenath S Andrali, and Jamie EL Cantrell (2010). "Modulation of transcription factor function by O-GlcNAc modification". In: *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1799.5-6, pp. 353–364.

O'Geen, Henriette, Lorigail Echipare, and Peggy J Farnham (2011). "Using ChIP-seq technology to generate high-resolution profiles of histone modifications". In: *Epigenetics Protocols*. Springer, pp. 265–286.

Pai, Shraddha, Ruth Isserlin, Alex Pico, and Kei Ono (2017). *r2cytoscape: Simple wrapper functions to use CyREST in R*. R package version 0.0.3.

Palm, Wilhelm and Titia de Lange (2008). "How shelterin protects mammalian telomeres". In: *Annual review of genetics* 42, pp. 301–334.

Park, Joon Tae, Young-Sam Lee, Kyung A Cho, and Sang Chul Park (2018). "Adjustment of the lysosomal-mitochondrial axis for control of cellular senescence". In: *Ageing research reviews* 47, pp. 176–182.

Parrinello, Simona, Enrique Samper, Ana Krtolica, Joshua Goldstein, Simon Melov, and Judith Campisi (2003). "Oxygen sensitivity severely limits the replicative lifespan of murine fibroblasts". In: *Nature cell biology* 5.8, p. 741.

Parry, Aled John and Masashi Narita (2016). "Old cells, new tricks: chromatin structure in senescence". In: *Mammalian Genome* 27.7-8, pp. 320–331.

Paszke, Adam, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala (2019). "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach, H. Larochelle, A. Beygelzimer, F. d Alché-Buc, E. Fox, and R. Garnett. Curran Associates, Inc., pp. 8024–8035. URL: `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

Patro, Rob and Carl Kingsford (2013). "Predicting protein interactions via parsimonious network history inference". In: *Bioinformatics* 29.13, pp. i237–i246.

Paull, Tanya T, Emmy P Rogakou, Vikky Yamazaki, Cordula U Kirchgessner, Martin Gellert, and William M Bonner (2000). "A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage". In: *Current Biology* 10.15, pp. 886–895.

Piper, Jason, Markus C Elze, Pierre Cauchy, Peter N Cockerill, Constanze Bonifer, and Sascha Ott (2013). "Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data". In: *Nucleic acids research* 41.21, e201–e201.

Poele, Robert H te, Andrei L Okorokov, Lesley Jardine, Jeffrey Cummings, and Simon P Joel (2002). "DNA damage is able to induce senescence in tumor cells in vitro and in vivo". In: *Cancer research* 62.6, pp. 1876–1883.

Puvvula, Pavan Kumar, Rohini Devi Desetty, Pascal Pineau, Agnés Marchio, Anne Moon, Anne Dejean, and Oliver Bischof (2014). "Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit". In: *Nature communications* 5, p. 5323.

Quijano, Celia, Liu Cao, Maria M Fergusson, Hector Romero, Jie Liu, Sarah Gutkind, Ilsa I Rovira, Robert P Mohney, Edward D Karoly, and Toren Finkel (2012). "Oncogene-induced senescence results in marked metabolic and bioenergetic alterations". In: *Cell Cycle* 11.7, pp. 1383–1392.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.R-project.org/`.

Radman-Livaja, Marta and Oliver J Rando (2010). "Nucleosome positioning: how is it established, and why does it matter?" In: *Developmental biology* 339.2, pp. 258–266.

Raederstorff, D, V Loechleiter, and U Moser (1995). "Polyunsaturated fatty acid metabolism of human skin fibroblasts during cellular aging." In: *International journal for vitamin and nutrition research. Internationale Zeitschrift fur Vitamin-und Ernahrungsforschung. Journal international de vitaminologie et de nutrition* 65.1, pp. 51–55.

Ramirez, Ricardo N, Nicole C El-Ali, Mikayla Anne Mager, Dana Wyman, Ana Conesa, and Ali Mortazavi (2017). "Dynamic gene regulatory networks of human myeloid differentiation". In: *Cell systems* 4.4, pp. 416–429.

Rayess, Hani, Marilene B Wang, and Eri S Srivatsan (2012). "Cellular senescence and tumor suppressor gene p16". In: *International journal of cancer* 130.8, pp. 1715–1725.

Roberson, Rachel S, Steven J Kussick, Eric Vallieres, Szu-Yu J Chen, and Daniel Y Wu (2005). "Escape from therapy-induced accelerated cellular senescence in p53-null lung cancer cells and in human lung cancers". In: *Cancer research* 65.7, pp. 2795–2803.

Rodier, Francis, Denise P Muñoz, Robert Teachenor, Victoria Chu, Oanh Le, Dipa Bhaumik, Jean-Philippe Coppé, Eric Campeau, Christian M Beauséjour, Sahn-Ho Kim, et al. (2011). "DNA-SCARS: distinct nuclear structures that sustain damage-induced senescence growth arrest and inflammatory cytokine secretion". In: *J Cell Sci* 124.1, pp. 68–81.

Routh, Edward John (1877). *A treatise on the stability of a given state of motion: particularly steady motion*. Macmillan and Company.

Rusilowicz, Martin, Michael Dickinson, Adrian Charlton, Simon O'Keefe, and Julie Wilson (2016). "A batch correction method for liquid chromatography–mass spectrometry data that does not depend on quality control samples". In: *Metabolomics* 12.3, p. 56.

Sagini, Krizia, Lorena Urbanelli, Eva Costanzi, Nico Mitro, Donatella Caruso, Carla Emiliani, and Sandra Buratta (2018). "Oncogenic H-Ras Expression Induces Fatty Acid Profile Changes in Human Fibroblasts and Extracellular Vesicles". In: *International journal of molecular sciences* 19.11, p. 3515.

Sanz, David J, Jennifer A Hollywood, Martina F Scallan, and Patrick T Harrison (2017). "Cas9/gRNA targeted excision of cystic fibrosis-causing deep-intronic splicing mutations restores normal splicing of CFTR mRNA". In: *PLoS One* 12.9, e0184009.

Sarkisian, Christopher J, Blaine A Keister, Douglas B Stairs, Robert B Boxer, Susan E Moody, and Lewis A Chodosh (2007). "Dose-dependent oncogene-induced senescence in vivo and its evasion during mammary tumorigenesis". In: *Nature cell biology* 9.5, p. 493.

Saxton, Robert A and David M Sabatini (2017). "mTOR signaling in growth, metabolism, and disease". In: *Cell* 168.6, pp. 960–976.

Scheffer, Marten, Stephen R Carpenter, Timothy M Lenton, Jordi Bascompte, William Brock, Vasilis Dakos, Johan Van de Koppel, Ingrid A Van de Leemput, Simon A Levin, Egbert H Van Nes, et al. (2012). "Anticipating critical transitions". In: *science* 338.6105, pp. 344–348.

Schmutz, Isabelle and Titia de Lange (2016). "Shelterin". In: *Current Biology* 26.10, R397–R399.

Schoenfelder, Stefan and Peter Fraser (2019). "Long-range enhancer–promoter contacts in gene expression control". In: *Nature Reviews Genetics*, p. 1.

Schvartzman, Juan Manuel, Craig B Thompson, and Lydia WS Finley (2018). "Metabolic regulation of chromatin modifications and gene expression". In: *J Cell Biol* 217.7, pp. 2247–2259.

Schymanski, Emma L, Heinz P Singer, Jaroslav Slobodnik, Ildiko M Ipolyi, Peter Oswald, Martin Krauss, Tobias Schulze, Peter Haglund, Thomas Letzel, Sylvia Grosse, et al. (2015). "Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis". In: *Analytical and bioanalytical chemistry* 407.21, pp. 6237–6255.

Sedelnikova, Olga A, Izumi Horikawa, Drazen B Zimonjic, Nicholas C Popescu, William M Bonner, and J Carl Barrett (2004). "Senescing human cells and ageing mice accumulate DNA lesions with unrepairable double-strand breaks". In: *Nature cell biology* 6.2, p. 168.

Sergushichev, Alexey (2016). "An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation". In: *bioRxiv*. DOI: 10.1101/060012. URL: http://biorxiv.org/content/early/2016/06/20/060012.

Serrano, Manuel, Athena W Lin, Mila E McCurrach, David Beach, and Scott W Lowe (1997). "Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a". In: *Cell* 88.5, pp. 593–602.

Shakeri, Azam, Hadi Tabibi, and Mehdi Hedayati (2010). "Effects of l-carnitine supplement on serum inflammatory cytokines, C-reactive protein, lipopro-

tein (a), and oxidative stress in hemodialysis patients with Lp (a) hyperlipoproteinemia". In: *Hemodialysis international* 14.4, pp. 498–504.

Shannon, Claude Elwood (1949). "Communication in the presence of noise". In: *Proceedings of the IRE* 37.1, pp. 10–21.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks". In: *Genome research* 13.11, pp. 2498–2504.

Shannon, Paul T, Mark Grimes, Burak Kutlu, Jan J. Bot, and David J. Galas (2013). "RCytoscape: tools for exploratory network analysis". In: *BMC Bioinformatics* 14.1, pp. 217–315. ISSN: 1471-2105. DOI: `http://dx.doi.org/10.1186/1471-2105-14-217`.

Shen, Zhesi, Wen-Xu Wang, Ying Fan, Zengru Di, and Ying-Cheng Lai (2014). "Reconstructing propagation networks with natural diversity and identifying hidden sources". In: *Nature communications* 5, p. 4323.

Sherr, Charles J, David Beach, and Geoffrey I Shapiro (2016). "Targeting CDK4 and CDK6: from discovery to therapy". In: *Cancer discovery* 6.4, pp. 353–367.

Sherwood, Richard I, Tatsunori Hashimoto, Charles W O'donnell, Sophia Lewis, Amira A Barkal, John Peter Van Hoff, Vivek Karun, Tommi Jaakkola, and David K Gifford (2014). "Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape". In: *Nature biotechnology* 32.2, p. 171.

Sica, Valentina, Jose Manuel Bravo-San Pedro, Valentina Izzo, Jonathan Pol, Sandra Pierredon, David Enot, Sylvere Durand, Noélie Bossut, Alexis Chery, Sylvie Souquere, et al. (2019). "Lethal Poisoning of Cancer Cells by Respiratory Chain Inhibition plus Dimethyl α-Ketoglutarate". In: *Cell reports* 27.3, pp. 820–834.

Silva, Ricardo R da, Pieter C Dorrestein, and Robert A Quinn (2015). "Illuminating the dark matter in metabolomics". In: *Proceedings of the National Academy of Sciences* 112.41, pp. 12549–12550.

Smith, Colin A, Elizabeth J Want, Grace O'Maille, Ruben Abagyan, and Gary Siuzdak (2006). "XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification". In: *Analytical chemistry* 78.3, pp. 779–787.

Soetaert, Karline ER, Thomas Petzoldt, and R Woodrow Setzer (2010). "Solving differential equations in R: package deSolve". In: *Journal of statistical software* 33.

Sompairac, Nicolas, Jennifer Modamio, Emmanuel Barillot, Ronan MT Fleming, Andrei Zinovyev, and Inna Kuperstein (2019). "Metabolic and sig-

nalling network maps integration: application to cross-talk studies and omics data analysis in cancer". In: *BMC bioinformatics* 20.4, p. 140.

Soto-Gamez, Abel and Marco Demaria (2017). "Therapeutic interventions for aging: the case of cellular senescence". In: *Drug Discovery Today* 22.5, pp. 786–795.

Stoll, Gautier, Barthélémy Caron, Eric Viara, Aurélien Dugourd, Andrei Zinovyev, Aurélien Naldi, Guido Kroemer, Emmanuel Barillot, and Laurence Calzone (2017). "MaBoSS 2.0: an environment for stochastic Boolean modeling". In: *Bioinformatics* 33.14, pp. 2226–2228.

Storer, Mekayla, Alba Mas, Alexandre Robert-Moreno, Matteo Pecoraro, M Carmen Ortells, Valeria Di Giacomo, Reut Yosef, Noam Pilpel, Valery Krizhanovsky, James Sharpe, et al. (2013). "Senescence is a developmental mechanism that contributes to embryonic growth and patterning". In: *Cell* 155.5, pp. 1119–1130.

Su, Ri-Qi, Ying-Cheng Lai, Xiao Wang, and Younghae Do (2014). "Uncovering hidden nodes in complex networks in the presence of noise". In: *Scientific reports* 4, p. 3944.

Sugihara, George, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch (2012). "Detecting causality in complex ecosystems". In: *science* 338.6106, pp. 496–500.

Suzen, Mehmet (2015). *R1magic: Compressive Sampling: Sparse Signal Recovery Utilities*. R package version 0.3.2. URL: `https://CRAN.R-project.org/package=R1magic`.

Swinton, Jonathan (2019). *Vennerable: Venn and Euler area-proportional diagrams*. R package version 3.1.0.9000. URL: `https://github.com/js229/Vennerable`.

Takai, Hiroyuki, Agata Smogorzewska, and Titia de Lange (2003). "DNA damage foci at dysfunctional telomeres". In: *Current biology* 13.17, pp. 1549–1556.

Takebayashi, Shin-ichiro, Hiroshi Tanaka, Shinjiro Hino, Yuko Nakatsu, Tomoka Igata, Akihisa Sakamoto, Masashi Narita, and Mitsuyoshi Nakao (2015). "Retinoblastoma protein promotes oxidative phosphorylation through upregulation of glycolytic genes in oncogene-induced senescent cells". In: *Aging Cell* 14.4, pp. 689–697.

Tarjan, Robert (1972). "Depth-first search and linear graph algorithms". In: *SIAM journal on computing* 1.2, pp. 146–160.

Tasdemir, Nilgun, Ana Banito, Jae-Seok Roe, Direna Alonso-Curbelo, Matthew Camiolo, Darjus F Tschaharganeh, Chun-Hao Huang, Ozlem Aksoy, Jessica E Bolden, Chi-Chao Chen, et al. (2016). "BRD4 connects enhancer remodeling to senescence immune surveillance". In: *Cancer discovery* 6.6, pp. 612–629.

Tatapudy, Sumitra, Francesca Aloisio, Diane Barber, and Todd Nystul (2017). "Cell fate decisions: emerging roles for metabolic signals and cell morphology". In: *EMBO reports* 18.12, pp. 2105–2118.

TeSlaa, Tara, Andrea C Chaikovsky, Inna Lipchina, Sandra L Escobar, Konrad Hochedlinger, Jing Huang, Thomas G Graeber, Daniel Braas, and Michael A Teitell (2016). "α-Ketoglutarate accelerates the initial differentiation of primed human pluripotent stem cells". In: *Cell metabolism* 24.3, pp. 485–493.

Thévenot, Etienne A, Aurélie Roux, Ying Xu, Eric Ezan, and Christophe Junot (2015). "Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses". In: *Journal of proteome research* 14.8, pp. 3322–3335.

Tian, Ze, TaeHyun Hwang, and Rui Kuang (2009). "A hypergraph-based learning algorithm for classifying gene expression and arrayCGH data with prior knowledge". In: *Bioinformatics* 25.21, pp. 2831–2838.

Tischler, Julia, Wolfram H Gruhn, John Reid, Edward Allgeyer, Florian Buettner, Carsten Marr, Fabian Theis, Ben D Simons, Lorenz Wernisch, and M Azim Surani (2019). "Metabolic regulation of pluripotency and germ cell fate through α-ketoglutarate". In: *The EMBO journal* 38.1.

Tissenbaum, Heidi A and Leonard Guarente (2001). "Increased dosage of a sir-2 gene extends lifespan in Caenorhabditis elegans". In: *Nature* 410.6825, p. 227.

Tran, Thai Q, Xazmin H Lowman, and Mei Kong (2017). "Molecular pathways: metabolic control of histone methylation and gene expression in cancer". In: *Clinical Cancer Research* 23.15, pp. 4004–4009.

Trimarchi, Jeffrey M and Jacqueline A Lees (2002). "Transcription: sibling rivalry in the E2F family". In: *Nature reviews Molecular cell biology* 3.1, p. 11.

Uhlendorf, Jannis, Agnès Miermont, Thierry Delaveau, Gilles Charvin, François Fages, Samuel Bottani, Gregory Batt, and Pascal Hersen (2012). "Long-term model predictive control of gene expression at the population and single-cell levels". In: *Proceedings of the National Academy of Sciences* 109.35, pp. 14271–14276.

Unity Biotechnology (2019). *A Safety and Tolerability Study of UBX0101 in Patients With Osteoarthritis of the Knee. Identification No. NCT03513016.* URL: `https://ClinicalTrials.gov/show/NCT03513016`.

Unser, Michael (2000). "Sampling—50 years after Shannon". In: *Proceedings of the IEEE* 88.ARTICLE, pp. 569–587.

Valdivia, Paola, Paolo Buono, Catherine Plaisant, Nicole Dufournaud, and Jean-Daniel Fekete (2019). "Analyzing Dynamic Hypergraphs with Parallel

Aggregated Ordered Hypergraph Visualization". In: *IEEE transactions on visualization and computer graphics*.

Van Deursen, Jan M (2014). "The role of senescent cells in ageing". In: *Nature* 509.7501, p. 439.

Vaquerizas, Juan M, Sarah K Kummerfeld, Sarah A Teichmann, and Nicholas M Luscombe (2009). "A census of human transcription factors: function, expression and evolution". In: *Nature Reviews Genetics* 10.4, p. 252.

Vaz-Drago, Rita, Noélia Custódio, and Maria Carmo-Fonseca (2017). "Deep intronic mutations and human disease". In: *Human genetics* 136.9, pp. 1093–1111.

Veer, Eric van der, Cynthia Ho, Caroline O'Neil, Nicole Barbosa, Robert Scott, Sean P Cregan, and J Geoffrey Pickering (2007). "Extension of human cell lifespan by nicotinamide phosphoribosyltransferase". In: *Journal of Biological Chemistry* 282.15, pp. 10841–10845.

Villiard, Éric, Jean-François Denis, Faranak Sadat Hashemi, Sebastian Igelmann, Gerardo Ferbeyre, and Stéphane Roy (2017). "Senescence gives insights into the morphogenetic evolution of anamniotes". In: *Biology open* 6.6, pp. 891–896.

Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake Vand erPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1. 0 Contributors (2019). "SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python". In: *arXiv e-prints*, arXiv:1907.10121. URL: https://ui.adsabs.harvard.edu/abs/2019arXiv190710121V.

Volterra, Vito (1927). *Variazioni e fluttuazioni del numero d'individui in specie animali conviventi*.

Voss, Ty C and Gordon L Hager (2014). "Dynamic regulation of transcriptional states by chromatin and transcription factors". In: *Nature Reviews Genetics* 15.2, pp. 69–81.

Waddington, CH (1957). *The strategy of the genes: a discussion of some aspects of theoretical biology*. Tech. rep. URL: https://archive.org/details/in.ernet.dli.2015.547782.

Wadhwa, Renu, Tomoko Yaguchi, Md K Hasan, Youji Mitsui, Roger R Reddel, and Sunil C Kaul (2002). "Hsp70 family member, mot-2/mthsp70/GRP75, binds to the cytoplasmic sequestration domain of the p53 protein". In: *Experimental cell research* 274.2, pp. 246–253.

Wallis, T.F.G.S.J.L.F., T.F. Glick, T. Glick, S.J. Livesey, and F. Wallis (2005). *Medieval Science, Technology, and Medicine: An Encyclopedia*. Encyclopedias of the Middle Ages / Encyclopedias of the Middle Ages. pg. 20. Routledge. ISBN: 9780415969307. URL: https://books.google.fr/books?id=SaJlbWK_-FcC.

Wang, Wen-Xu, Ying-Cheng Lai, and Celso Grebogi (2016). "Data based identification and prediction of nonlinear and complex dynamical systems". In: *Physics Reports* 644, pp. 1–76.

Wang, Xiongjun, Ruilong Liu, Xiujuan Qu, Hua Yu, Huiying Chu, Yajuan Zhang, Wencheng Zhu, Xueyuan Wu, Hong Gao, Bangbao Tao, et al. (2019). "α-Ketoglutarate-Activated NF-κB Signaling Promotes Compensatory Glucose Uptake and Brain Tumor Development". In: *Molecular cell* 76.1, pp. 148–162.

Wapinski, Orly L, Qian Yi Lee, Albert C Chen, Rui Li, M Ryan Corces, Cheen Euong Ang, Barbara Treutlein, Chaomei Xiang, Valérie Baubet, Fabian Patrik Suchy, et al. (2017). "Rapid chromatin switch in the direct reprogramming of fibroblasts to neurons". In: *Cell reports* 20.13, pp. 3236–3247.

Wehrens, Ron, Jos A Hageman, Fred van Eeuwijk, Rik Kooke, Pádraic J Flood, Erik Wijnker, Joost JB Keurentjes, Arjen Lommen, Henriëtte DLM van Eekelen, Robert D Hall, et al. (2016). "Improved batch correction in untargeted MS-based metabolomics". In: *Metabolomics* 12.5, p. 88.

Weiner, January (2017). *riverplot: Sankey or Ribbon Plots*. R package version 0.6. URL: https://CRAN.R-project.org/package=riverplot.

Wiley, Christopher D and Judith Campisi (2016). "From ancient pathways to aging cells—connecting metabolism and cellular senescence". In: *Cell Metabolism* 23.6, pp. 1013–1021.

Wiley, Christopher D, Michael C Velarde, Pacome Lecot, Su Liu, Ethan A Sarnoski, Adam Freund, Kotaro Shirakawa, Hyung W Lim, Sonnet S Davis, Arvind Ramanathan, et al. (2016). "Mitochondrial dysfunction induces senescence with a distinct secretory phenotype". In: *Cell metabolism* 23.2, pp. 303–314.

Wishart, David S, Yannick Djoumbou Feunang, Ana Marcu, An Chi Guo, Kevin Liang, Rosa Vázquez-Fresno, Tanvir Sajed, Daniel Johnson, Carin Li, Naama Karu, et al. (2017). "HMDB 4.0: the human metabolome database for 2018". In: *Nucleic acids research* 46.D1, pp. D608–D617.

Wittkopp, Patricia J and Gizem Kalay (2012). "Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence". In: *Nature Reviews Genetics* 13.1, pp. 59–69.

Wu, Nan, Mingyao Yang, Uma Gaur, Huailiang Xu, Yongfang Yao, and Diyan Li (2016). "Alpha-ketoglutarate: physiological functions and applications". In: *Biomolecules & therapeutics* 24.1, p. 1.

Xie, Jun, Xiaoming Zhang, and Li Zhang (2013). "Negative regulation of inflammation by SIRT1". In: *Pharmacological Research* 67.1, pp. 60–67.

Yaguchi, Tomoko, Satoshi Aida, Sunil C Kaul, and Renu Wadhwa (2007). "Involvement of mortalin in cellular senescence from the perspective of its mitochondrial import, chaperone, and oxidative stress management functions". In: *Annals of the New York Academy of Sciences* 1100.1, pp. 306–311.

Yamada, Shuichi, Ryosuke Matsuda, Fumihiko Nishimura, Ichiro Nakagawa, Yasushi Motoyama, Young-Su Park, Mitsutoshi Nakamura, Hiroyuki Nakase, Yukiteru Ouji, and Masahide Yoshikawa (2012). "Carnitine-induced senescence in glioblastoma cells". In: *Experimental and therapeutic medicine* 4.1, pp. 21–25.

Yang, Lian-wei, Ming Song, Yu-lin Li, Ya-peng Liu, Chun Liu, Lu Han, Zhihao Wang, Wei Zhang, Yan-qiu Xing, and Ming Zhong (2019). "l-Carnitine inhibits the senescence-associated secretory phenotype of aging adipose tissue by JNK/p53 pathway". In: *Biogerontology* 20.2, pp. 203–211.

Yang, Won Ho, Ji Eun Kim, Hyung Wook Nam, Jung Won Ju, Hoe Suk Kim, Yu Sam Kim, and Jin Won Cho (2006). "Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability". In: *Nature cell biology* 8.10, p. 1074.

Yang, Won Ho, Sang Yoon Park, Hyung Wook Nam, Jeong Gu Kang, Eun Seok Kang, Yu Sam Kim, Hyun Chul Lee, Kwan Soo Kim, Jin Won Cho, et al. (2008). "NFκB activation is associated with its O-GlcNAcylation state under hyperglycemic conditions". In: *Proceedings of the National Academy of Sciences* 105.45, pp. 17345–17350.

Yarla, Nagendra Sastry, Olga Sukocheva, Ilaria Peluso, Swathi Putta, Pallaval Veera Bramhachari, Rajesh Yadala, Dinesh K Tiwari, Srinivas Jagarlamudi, Luciana Scotti, Marcus T Scotti, et al. (2017). "Targeting Arachidonic Acid Pathway-Associated NF-κB in Pancreatic Cancer". In: *Role of Transcription Factors in Gastrointestinal Malignancies*. Springer, pp. 403–411.

Yeung, Fan, Jamie E Hoberg, Catherine S Ramsey, Michael D Keller, David R Jones, Roy A Frye, and Marty W Mayo (2004). "Modulation of NF-κB-dependent transcription and cell survival by the SIRT1 deacetylase". In: *The EMBO journal* 23.12, pp. 2369–2380.

Yılmaz, Zekiye Buket, Bente Kofahl, Patrick Beaudette, Katharina Baum, Inbal Ipenberg, Falk Weih, Jana Wolf, Gunnar Dittmar, and Claus Scheidereit (2014). "Quantitative dissection and modeling of the NF-κB p100-

p105 module reveals interdependent precursor proteolysis". In: *Cell reports* 9.5, pp. 1756–1769.

Yuh, Chiou-Hwa, Hamid Bolouri, and Eric H Davidson (1998). "Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene". In: *Science* 279.5358, pp. 1896–1902.

Yun, Maximina H, Hongorzul Davaapil, and Jeremy P Brockes (2015). "Recurrent turnover of senescent cells during regeneration of a complex structure". In: *Elife* 4, e05505.

Zaret, Kenneth S and Jason S Carroll (2011). "Pioneer transcription factors: establishing competence for gene expression". In: *Genes & development* 25.21, pp. 2227–2241.

Zeng, Yaohui and Patrick Breheny (2017). "The biglasso package: a memory- and computation-efficient solver for lasso model fitting with big data in R". In: *arXiv preprint arXiv:1701.05936*.

Zentner, Gabriel E, Paul J Tesar, and Peter C Scacheri (2011). "Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions". In: *Genome research* 21.8, pp. 1273–1283.

Zhang, Ran, Hou-Zao Chen, Jin-Jing Liu, Yu-Yan Jia, Zhu-Qin Zhang, Rui-Feng Yang, Yuan Zhang, Jing Xu, Yu-Sheng Wei, De-Pei Liu, et al. (2010). "SIRT1 suppresses activator protein-1 transcriptional activity and cyclooxygenase-2 expression in macrophages". In: *Journal of Biological Chemistry* 285.10, pp. 7097–7110.

Zhang, Yaya, Yongjun Wang, Chuanen Bao, Yingyi Xu, Huili Shen, Junjie Chen, Jianghua Yan, and Yuqiang Chen (2012). "Metformin interacts with AMPK through binding to γ subunit". In: *Molecular and cellular biochemistry* 368.1-2, pp. 69–76.

Zhang, Yiqiang, Alex Bokov, John Gelfond, Vanessa Soto, Yuji Ikeno, Gene Hubbard, Vivian Diaz, Lauren Sloane, Keith Maslin, Stephen Treaster, et al. (2013). "Rapamycin extends life and health in C57BL/6 mice". In: *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences* 69.2, pp. 119–130.

Zhu, Jiyue, Douglas Woods, Martin McMahon, and J Michael Bishop (1998). "Senescence of human fibroblasts induced by oncogenic Raf". In: *Genes & development* 12.19, pp. 2997–3007.

Zhu, Yi, Tamara Tchkonia, Tamar Pirtskhalava, Adam C Gower, Husheng Ding, Nino Giorgadze, Allyson K Palmer, Yuji Ikeno, Gene B Hubbard, Marc Lenburg, et al. (2015). "The Achilles' heel of senescent cells: from transcriptome to senolytic drugs". In: *Aging cell* 14.4, pp. 644–658.

Zhu, Yi, Tamara Tchkonia, Heike Fuhrmann-Stroissnigg, Haiming M Dai, Yuanyuan Y Ling, Michael B Stout, Tamar Pirtskhalava, Nino Giorgadze, Kurt O Johnson, Cory B Giles, et al. (2016). "Identification of a novel

senolytic agent, navitoclax, targeting the Bcl-2 family of anti-apoptotic factors". In: *Aging cell* 15.3, pp. 428–435.

Appendix

A.1 Contribution to publication

AP-1 Imprints a Reversible Transcriptional Program of Senescent Cells

Ricardo Iván Martínez-Zamudio[1,3,11], Pierre-François Roux[1,3,10], José Américo N L F de Freitas[1,2], Lucas Robinson[1,2], Gregory Doré[1], Bin Sun[4,5], Dimitri Belenki[6,7], Kolja Schleich[6], Maja Milanovic[6], Utz Herbig[11] Clemens A. Schmitt[6-9], Jesús Gil[4,5], and Oliver Bischof[1,‡]

[1] Institut Pasteur, 75015 Paris, France

[2] Université de Paris, Sorbonne Paris Cité, Paris, France

[3] These authors contributed equally to this work

[4] MRC London Institute of Medical Sciences (LMS), Du Cane Road, London, W12 0NN, UK

[5] Institute of Clinical Sciences (ICS), Faculty of Medicine, Imperial College London, Du Cane Road, London W12 0NN, UK

[6] Charité - University Medical Center, Department of Hematology, Oncology and Tumor Immunology, Virchow Campus, and Molekulares Krebsforschungszentrum, Augustenburger Platz 1, 13353 Berlin, Germany

[7] Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association, Robert-Rössle-Straße 10, 13125 Berlin, Germany

[8] Deutsches Konsortium für Translationale Krebsforschung (German Cancer Consortium), Partner site Berlin, Germany

[9] Kepler University Hospital, Department of Hematology and Oncology, Johannes Kepler University, Krankenhausstraße 9, 4020 Linz, Austria

[10] Johnson & Johnson, Upstream Skin Research, 92130 Issy-les-Moulineaux, France

[11] Department of Microbiology, Biochemistry & Molecular Genetics, Rutgers Biomedical & Health Sciences, Rutgers University, 205 South Orange Avenue, Newark, NJ 07103, USA

[‡] Lead contact

*Correspondence: oliver.bischof@pasteur.fr

SUMMARY

Senescent cells play important physiological- and pathophysiological roles in tumor suppression, tissue regeneration, and aging. While select genetic and epigenetic elements crucial for senescence induction were identified, the dynamics, underlying epigenetic mechanisms, and regulatory networks defining senescence competence, induction and maintenance remain poorly understood, precluding a deliberate therapeutic manipulation of these dynamic processes. Here, we show, using dynamic analyses of transcriptome and epigenome profiles, that the epigenetic state of enhancers predetermines their sequential activation during senescence. We demonstrate that activator protein 1 (AP-1) ‘imprints’ the senescence enhancer landscape effectively regulating transcriptional activities pertinent to the timely execution of the senescence program. We define and validate a hierarchical transcription factor (TF) network model and demonstrate its effectiveness for the design of senescence reprogramming experiments. Together, our findings define the dynamic nature and organizational principles of gene-regulatory elements driving the senescence program and reveal promising inroads for therapeutic manipulation of senescent cells.

INTRODUCTION

Cellular senescence plays beneficial roles during embryonic development, wound healing, and tumor suppression. Paradoxically, it is also considered a significant contributor to aging and age-related diseases including cancer and degenerative pathologies[1]. As such, research on therapeutic strategies exploiting senescence targeting (*e.g.*, senolytics, senomorphics or pro-senescence cancer therapies) to improve healthspan has gained enormous momentum in recent years[2].

Cellular senescence is a cell fate that stably arrests proliferation of damaged and dysfunctional cells as a complex stress response. The most prominent inducers of senescence are hyper-activated oncogenes (oncogene-induced senescence, OIS) and therapeutic interventions to induce senescence in cancerous cells (therapy-induced senescence, TIS)[3]. The senescence arrest is accompanied by widespread changes in gene expression, including a senescence-associated secretory phenotype (SASP) – the expression and secretion of inflammatory cytokines, growth factors, proteases, and other molecules such as stemness factors[4], which exert both cell-autonomous and cell-nonautonomous effects[5]. Importantly, although activation of the senescence program can pre-empt the initiation of cancer, the long-term effects of the SASP make the local tissue environment more vulnerable to the spread of cancer and other age-related diseases. Therefore, therapeutic interventions aimed at limiting SASP production are of relevance for cancer and many age-related diseases[5-7].

The knowledge on epigenetic mechanisms underlying senescence has only recently increased revealing significant contributions of select transcription factors (TFs), chromatin modifiers and structural components, as well as non-coding RNAs to the senescent phenotype[8-13]. A major limitation of such studies, however, was their restriction

to start-end-point comparisons, ignoring the dynamic nature of the senescence fate transition. Consequently, critical gene-regulatory aspects of the execution and maintenance of the senescence state remain poorly understood. Therefore, an integrative, temporally resolved, multidimensional profiling approach is required to establish essential regulatory principles that govern this key biological decision-making process. Such knowledge would be instrumental both for identifying stage-specific senescence regulators and urgently needed specific biomarkers as well as control points in TF and gene regulatory networks, which would pave the way for a deliberate therapeutic manipulation of the senescence cell fate.

Enhancers are key genomic regions that drive cell fate transitions. The enhancer landscape is established during development by the concerted action of TF networks and chromatin modifiers[14]. The details on how this information converges in *cis* remain unclear, and we still lack valid organizational principles that explain the function of mammalian TF networks. In mammalian cells, enhancer elements are broadly divided into two major categories -active and poised. While active enhancers are characterized by the simultaneous presence of H3K4me1 together with H3K27ac and are associated with actively transcribed genes, only H3K4me1 marks poised enhancers, and their target genes are generally not expressed[15]. A subset of enhancers may also be activated *de novo* from genomic areas devoid of any TF binding and histone modifications. These latent or nascent enhancers serve an adaptive role in mediating stronger and faster gene expression upon cycles of repeated stimulation[16, 17]. Recent evidence showed a role for enhancer remodeling in driving senescence-associated gene expression[12, 13, 18]. However, it is currently unknown which enhancer elements, epigenetic marks or TFs render cells competent to respond to senescence-inducing signals with precise timing and

output. A thorough understanding of how senescence competence is established, realized and what defines it would allow for the prediction of a positive senescence engagement for example in pro-senescence cancer therapies[19].

Pioneer TFs are critical in establishing new cell fate competence by granting long-term chromatin access to non-pioneer factors and are also crucial determinants of cell identity through their opening and licensing of the enhancer landscape[20, 21]. We can now reliably deduce pioneer and non-pioneer TF activity from chromatin accessibility data allowing for the hierarchization of TF function whereby pioneer TFs sit atop a TF binding hierarchy, recruiting non-pioneers such as settler and migrant TFs to gene-regulatory regions for optimal transcriptional output[22]. The pioneer TFs bestowing senescence potential have not been identified to date. However, their identification might be a pre-requisite for reprogramming or manipulation of senescent cells for future therapeutic benefit as was shown successfully for the reprogramming to adopt full stem cell identity[23].

In this study, we examined the possibility that the epigenetic state of enhancers could determine senescence cell fate. We explored this question by generating time-resolved transcriptomes and comprehensive epigenome profiles during oncogenic RAS-induced senescence and validated central findings in different cell biology and disease models of senescence, both mouse and human. Through integrative analysis and further functional validation, we revealed novel and unexpected links between enhancer chromatin, TF recruitment, and senescence potential and defined the organizational principles of the TF network that drive the senescence program. Together, this allowed us to manipulate the senescence phenotype with potential therapeutic implications. Specifically, we show that the senescence program is predominantly encoded at the enhancer level and that the enhancer landscape is dynamically reshaped at each step of

the senescence transition. Remarkably, we find that this process is pre-determined before senescence induction, and AP-1 acts as a pioneer TF that ‘premarks’ prospective senescence enhancers to direct and localize the recruitment of other transcription factors into a hierarchical TF network that drives the senescence transcriptional program after induction. We also uncover a class of enhancers that retain an epigenomic memory after their inactivation during the senescence transition. These “remnant” enhancers lack traditional enhancer histone-modification marks but are instead “remembered” by AP-1 TF bookmarking for eventual future re-activation. Functional perturbation of prospective senescence enhancers and AP-1 validated and underscored the importance of these entities for the timely execution of the senescence gene expression program and allowed for the precise reprogramming and reversal of the senescence cell fate. Importantly, we reveal that AP1 plays a critical role in therapy-induced senescence (TIS) of colorectal cancer cells and a mouse model of B cell lymphoma and show that an AP1-regulated senescence-associated gene signature can stratify mouse lymphomas into treatment-responsive and recurring, treatment-resistant lymphomas in mice and humans.

RESULTS

We employed time-series experiments on normal, human diploid lung fibroblasts (HDF, strain WI38) undergoing oncogene-induced senescence (OIS) using a tamoxifen-inducible ER:RAS^{V12} expression system[24] (RAS-OIS) in two biological replicates. We determined global gene expression profiles by microarrays and mapped the full set of accessible chromatin sites by ATAC-seq[25] at 6-time points (0, 24, 48, 72, 96 and 144 h, 2 biological replicates). Cells intended for ChIP-seq were crosslinked at 3-time-points (0, 72 and 144h, 2 biological replicates) and used for profiling histone modifications including H3K4me1 (putative enhancers), H3K4me3 (promoters), H3K27ac (active enhancers and promoters) and H3K27me3 (polycomb repressed chromatin). From accessible chromatin regions determined by ATAC-seq we deduced TF binding dynamics and hierarchies (Figure 1a). For comparison, we included cells undergoing quiescence (Q time points: 0, 12, 24, 48, 72 and 96h) by serum withdrawal for up to 96h. Unlike senescence arrested cells, quiescence arrested cells can be triggered to re-enter the cell cycle upon serum addition. We also validated our approach in two additional cell biology models of senescence: i) oncogenic RAF-induced senescence (RAF-OIS time points: 0, 12, 24, 48, 72, and 96 h for transcriptome and ATAC-seq analysis; 0, 48 and 96h for H3K4me1 and H3K27ac ChIP-seq), ii) replicative senescence (RS time points: 0, 144, 264, 432, 624, 792, 1008 and 2112 h for transcriptome and ATAC-seq analysis; 0, 264, 1008 and 2112h for H3K4me1 and H3K27ac ChIP-seq analysis), and in RAS-OIS of GM21 skin fibroblasts (time points: 0, 192 and 336 h for transcriptome and ATAC-seq analysis). Quiescence and senescence were verified using senescence-associated ß-Gal (SABG) staining, population doublings, EdU incorporation and RT-qPCR-based expression profiling of genes pertinent for senecence (Supplementary Figures 1a-q)

Multi-state establishment of the senescence transcriptional program

To identify and visualize dynamic gene expression patterns across the entire Q and RAS-OIS time-courses, we employed an unsupervised, self-organizing map (SOM) machine learning technique[26] (Figure 1b) and multidimensional scaling (MDS) (Figure 1c) to our transcriptome data sets. Remarkably, serum-deprived fibroblasts rapidly established a Q-specific gene expression program within 24 h after serum deprivation, which changed only marginally within the remainder of the time-course (Figure 1b, left column and Figure 1c), and mainly involves only up-regulated (Figure 1b, top right corner, red) and down-regulated (Figure 1b, bottom left corner, blue) genes. By contrast, fibroblasts undergoing RAS-OIS displayed dynamic gene expression trajectories that evolved steadily, both for up- (red) and down-regulated metagenes (blue) (Figures 1b, right column and Figure 1c), which was corroborated by the expression profiles of a selection of senescence-associated genes (Supplementary Figure S1e). Thus, RAS-OIS, unlike Q, is highly dynamic and does not gyrate towards a stable transcriptome end state. To substantiate this further we calculated the diversity and specialization of transcriptomes and gene specificity[27] (Figure 1d). These analyses demonstrated that RAS-OIS cells exhibit a temporally evolving increase in transcriptional diversity, whereas Q cells exhibit a temporally evolving, specific gene expression program. We conclude that the RAS-OIS cell fate is a succession of cell states rather than a fixed cell fate with a defined end-point, which is the current tenet. The transcriptional diversity may provide a fertile soil for the eventual escape of pre-cancerous senescent cells as previously shown[4, 28].

To further delineate the evolution of the RAS-OIS gene expression program, we performed dynamic differential gene expression analysis on the Q and OIS datasets[29]. A total of 4,986 genes (corresponding to 2,931 up-regulated and 2,055 down-regulated

genes) were differentially regulated in at least one-time point (with a minimal leading $\log_2$ fold-change of 1.2; $q=5*10^{-4}$) and partitioned into seven (I-VII) gene expression modules with distinct functional overrepresentation profiles in line with the senescence phenotype (Figures 1e-f and Supplementary Figure 1u). The highly reproducible gene expression dynamics and modularity during RAS-OIS transition suggest a high degree of preprogramming of this succession of cell states, which we confirmed in additional cell biology models of senescence, i.e., WI-38 lung fibroblasts undergoing RAF-OIS and RS as well as in GM21 skin fibroblasts undergoing RAS-OIS (Supplementary Figures 1r-t).

Cell-fate decisions are typically associated with stable changes in gene expression that shift the regulatory system from one steady state to the next[30]. In line with this, we found that in RAS-OIS of WI38 fibroblasts proliferation-promoting genes of modules II and IV (e.g., E2F targets and G2M checkpoint such as *CDKN1B*, *CCNB2* and *CDK1*) became increasingly repressed (*i.e.* senescence arrest), while pro-senescent SASP genes of modules VI and VII (e.g., *IL1ß*, *IL6* and *IFNGR2*) became persistently induced (Figures 1e-f and Supplementary Figure 1o). Apoptosis-related genes of module III (e.g., *COL5A2, CCN2* and *IGFBP3*) were repressed very early on in the time-course (within the first 24-48 hours during RAS-OIS induction; Figure 1f and Supplementary Figure 1u) indicating that the commitment to undergo senescence is a very early event made at the expense of apoptosis. Finally, we identified a set of genes in modules I and V involved in the epithelial-to-mesenchymal transition (EMT) and TNF signaling (e.g., *MMP3, COL4A2, MRTO4*) that would have gone unnoticed in a traditional start-end-point analysis because they follow a sharp transition. In these modules, transcript levels spiked-down (module I; e.g., *MMP3* and *COL4A2*) or up (module V; e.g., *MTRO4* and *C1QBP*) within 24 h following RAS-OIS induction, before transitioning to a new steady state, similar to the

original levels (Figure 1e and Supplementary Figure 1u). These expression patterns support the notion that genes of module V play a role early in the transition to RAS-OIS, while genes in module I play a role in the maintenance of the proliferative fibroblast state.

Altogether, our investigation of transcriptome dynamics in different cell biology models of senescence defined a modular organization and transcriptional diversity of the senescence gene expression program, providing a framework to unravel the gene-regulatory code underlying the senescence process.

A dynamic enhancer program shapes the senescence transcriptome

Senescence cell fate involves a global remodeling of chromatin and specifically, the enhancer landscape[12, 13]. An unanswered question, however, is how TFs and epigenetic modifications cooperatively shape a transcriptionally permissive enhancer landscape prior to gene activation to endow the cell with senescence potential.

To provide mechanistic insight into this question, we first comprehensively mapped genomic regulatory elements (*i.e.* putative enhancers, promoters and polycomb-repressed chromatin) during the transition of proliferating WI38 fibroblasts to RAS-OIS, profiling genome-wide histone modifications by ChIP-seq and transposon-accessible chromatin by ATAC-seq. To capture and quantify chromatin state dynamics we used ChromstaR (see Materials and Methods), which identified a total of sixteen chromatin states in RAS-OIS (Supplementary Figure 2a). The majority of the genome (≈80%) was, irrespective of the time-point, either devoid of any of the histone modifications analyzed (≈62%) or polycomb-repressed (≈18%). The fraction of the genome represented by active and accessible chromatin states (*i.e.*, enhancers and promoters) was comparably lower

(≈20% combined). Chromatin state transitions occurred most prominently at enhancers, while promoters were only mildly affected (Figures 2a,b and Supplementary Figure 2a, see insets indicated by arrows) congruent with previous results[12]. Unexpectedly, we found that many enhancers were activated *de novo* (*i.e.,* acquisition of H3K4me1 and H3K27ac) from unmarked chromatin at the T0-72 h and 72 h-144 h intervals, followed by the more stereotypical enhancer activation from a poised state (H3K4me1^{+} plus H3K27ac acquisition) and enhancer poising from the unmarked and polycomb-repressed state at the T0-72 h interval (acquisition of H3K4me1) (Figures 2a,b). Dynamics of sequential enhancer activation was preserved in WI38 fibroblasts undergoing RAF-OIS and RS (Supplementary Figures 2b,c). We conclude that the regulatory landscape of senescence is largely predetermined by sequential enhancer activation from *de novo* and poised enhancers implying the existence of a prior imprint of past cell fate decisions.

The chronology of enhancer activation was highly concordant with the temporal expression pattern of the nearest genes, indicating that most of these elements indeed function as *bona fide* enhancers (Supplementary Figure 2d). In line with this, correspondence analysis (CA) (Supplementary Figure 2e) revealed a strong correlation between gene expression modules (Figure 1e) and chromatin state transitions (Figure 2a). For example, globally up-regulated transcriptomic modules V, VI, VII projected proximally to chromatin state transitions involving enhancer activation congruent with the installation of the SASP. By contrast, dynamic enhancer inactivation associated with repressed transcriptomic modules (II, III, IV) congruent with installation of the senescence arrest. Finally, the oscillatory expression of genes in the module I associated with an equally oscillatory activation of its closest enhancers. Therefore, dynamic remodeling of

the enhancer landscape reflects and defines the modular and dynamic nature of the RAS-OIS gene expression program.

We next addressed the question of which TFs are key drivers for the dynamic enhancer remodeling driving the senescence transcriptome. To this end, we first intersected ATAC-seq peaks with the identified enhancer coordinates (Figure 2a,b) and performed a motif over-representation test. This analysis identified AP-1 super-family members (cJUN, FOS, FOSL1, FOSL2, BATF) as well as AP-1-associated TFs ATF3 and ETS1 as the most enriched motifs at any given time-point, thus, hinting at a putative chromatin priming and pioneer function for these TFs (Supplementary Figure 2f). Because AP-1 TFs are essential and inducible downstream effectors for the RAS signaling pathway in cellular transformation[31] the possibility remains that the observed enrichment of AP-1 TFs at enhancers is strictly dependent on oncogenic RAS signaling *per se* and not a reflection of a specific pioneering role in the enhancer landscape independent of RAS signaling. We therefore compared ATAC-seq peaks for TF binding sites in WI38 lung fibroblasts undergoing RAS-OIS, RAF-OIS, RS (which is driven by loss of telomere integrity), GM21 skin fibroblasts undergoing RAS-OIS and growth factor-deprived (and thus RAS signaling-muted) quiescent WI-38 fibroblasts (Supplementary Figures 2g-k). In all cases, the AP-1 motif ascended as the predominant motif enriched, thus, corroborating the notion that AP-1 TFs act as universal pioneers imprinting the global as well as senescence-associated enhancer landscape.

To elaborate this further, we analyzed our time-resolved RAS-OIS ATAC-seq data sets obtained in WI38 fibroblasts by adapting the “Protein Interaction Quantitation (PIQ)” algorithm, which was developed initially for DNAse-seq-based digital TF footprinting[22]. Importantly, PIQ allows for the functional hierarchization of TFs into pioneers, settlers, and

migrants - whereby pioneer TFs bind nucleosome-compacted chromatin to initiate chromatin remodeling and to enable subsequent binding of non-pioneers (*i.e.*, settler and migrant TFs). PIQ segregated TFs into pioneers (*e.g.*, AP-1 TF family members), settlers (*e.g.*, NFY and RELA subunit of NF-κB) and migrants (*e.g.*, TF RAR family members and SREBF1) (Figure 2c). We confirmed this TF hierarchization by inspecting a selection of individual TF footprints for their adjacent nucleosomal positioning (Supplementary Figure 2l-n). AP-1 family member FOSL1, for example, bound to its cognate binding site despite the presence of strongly positioned flanking nucleosomes, as would be expected from a pioneer TF (Supplementary Figure 2l), while RELA binding required partial nucleosome displacement/chromatin opening, as would be expected for a settler TF (Supplementary Figure 2m), and SREBF1 bound to its cognate site in a near-nucleosome free context, as would be expected for a migrant TF (Supplementary Figure 2n). Importantly, there was a high correspondence between PIQ predictions and TF ChIP-seq profiling as exemplified for AP-1-members FOSL2 and cJUN, which we used as surrogate marks for bound AP-1 (which is typically a complex of JUN-JUN or JUN-FOS family member dimers), and RELA (Supplementary Figure 2o).

To decode additional TF properties critical for shaping the dynamic RAS-OIS enhancer landscape, we applied a principal component analysis (PCA) considering several metrics describing TF binding characteristics (Figure 2d). This analysis revealed two key features: First, pioneer TFs bind statically, extensively, and most importantly before RAS-OIS induction (*i.e.*, pre-stimulation) along the genome, while settler and migrant TFs bind more dynamically (“Dynamicity” in Figure 2d), far less frequently (“Windows” in Figure 2d), and on average less often before OIS induction (*i.e.* pre-

stimulation) along the genome. Second, and in line with the proposed pioneering activity of AP-1 TFs, the latter clearly stand out amongst other pioneer TFs (highlighted by black circle in Figure 2d) because they bind exclusively and extensively to enhancers prior to RAS-OIS induction whereas most of the remaining pioneer TFs tend to accumulate away from them.

In summary, we identify *de novo* enhancer activation and AP-1 as novel and key elements that pioneer and shape a transcriptionally permissive enhancer landscape in senescence.

AP-1 pioneer TF bookmarking of senescence enhancer landscape foreshadows the senescence transcriptional program

Given our unexpected finding that most of the enhancer activation occurred *de novo* out of unmarked chromatin territories, *i.e.*, devoid of enhancer-related histone modifications H3K4me1 and H3K27ac and ending in an active H3K4me1^{+}/ H3K27ac^{+} enhancer state at 144h, and that AP-1 TFs act as pioneers to shape the senescence enhancer landscape, we explored a possible role of AP-1 as a general bookmarking agent for future and past enhancer activity during RAS-OIS in WI38 fibroblasts. Quantification of enhancer mark dynamics (Figure 3a and Supplementary Figures 3a-c) unveiled that for windows shifting from the "unmarked" state at T_0 to an "active enhancer" state (H3K4me1^{+} / H3K27ac^{+}) at either 72 h or 144 h, i.e. "*de novo* enhancers", there is both a gradual increase in H3K4me1 and H3K27ac levels from initial levels (T_0) similar to steady-state unmarked regions but different from poised enhancers, to final levels (144 h) indistinguishable from constitutive enhancers (Figure 3a and Supplementary Figures 3a,b). By contrast, for windows shifting

from an “active enhancer” state at T_0 to an “unmarked” state at either 72 h or 144 h, that we refer to as “remnant enhancers”, there is a progressive decrease both in H3K4me1 and H3K27ac levels from initial levels indistinguishable from constitutive enhancers to final levels similar to unmarked regions and distinct from poised enhancers (Figure 3a and Supplementary Figures 3a,c). The dynamic behavior of each enhancer class on average associated with the expression profile of nearby genes, with constitutive enhancers displaying constant gene expression, *de novo* enhancers increasing and remnant enhancers decreasing gene expression (Supplementary Figure 3d).

To directly show the functional role of *de novo* enhancers we used a CRISPR interference (CRISPRi) approach[32, 33]. Expression of 4 different gRNA targeting the dCas9-KRAB transcriptional repressor to *de novo* enhancers in the *IL1α /IL1ß* locus (g7, -14, -15, and -61) significantly reduced the expression of *IL1ß* when analyzed 8 days after oncogenic RAS induction (Figure 3b). Interestingly, *IL1α* expression was only mildly reduced by the two gRNAs (g61 and g7) adjacent to the *IL1ß* promoter (Figure 3b). While similar results were observed 14 days after oncogenic RAS induction (Supplementary Figure 3e), a control gRNA (g54) targeting a genomic region just downstream of the *IL1α /IL1β* locus did not affect either expression, while control gRNA guides g2 and g48, targeting sequences in-between two *de novo* enhancers, had—moderate effects (Supplementary Figure 3f). Together, we render ample evidence that *de novo* and remnant enhancers are novel senescence-associated *cis*-regulatory modules that define the senescence transcriptional program.

We next determined whether TFs bookmark *de novo* enhancers for future activation and also, whether TFs bookmark remnant enhancers after their inactivation as part of a molecular memory. Indeed, as shown in Figure 3c, we found that AP-1 is the

predominant TF bookmarking *de novo* and remnant enhancers. Importantly, and highlighting the importance of AP-1 in bookmarking *de novo* enhancers for future activation, gRNAs chosen for CRISPRi were either overlapping with AP-1 binding sites (g14, g15 and g61) or in close proximity (g7), *i.e.* ~125 bp outside of it (Figure 3b). Because CRISPRi can control repression over a length of two nucleosomes (~300bp)[34], it is highly probable that g7 also affects this AP-1 binding site. Moreover, a control gRNA (g2) targeting a non-enhancer AP-1 site (Supplementary Figure 3f) did not affect *IL1* expression strongly suggesting that only enhancer-positioned AP-1 sites are functional. Finally, we validated the importance of AP-1 TFs for *de novo* and remnant enhancer bookmarking by examining their positioning also in cells undergoing replicative senescence, which demonstrated that AP-1 TFs here also play a leading role for bookmarking (Supplementary Figure 3g). We conclude that AP-1 bookmarking of *de novo* and remnant enhancers is independent of oncogenic RAS signaling and a novel and cardinal feature that reflects future and past transcriptional activities in senescence.

While performing this analysis, we noticed that only 2,480 out of 3,334 *de novo* enhancers were TF bookmarked, while the remainder (n=854) lacked any detectable TF binding activity (Figure 3d). Thus, *de novo* enhancers can be further divided into two subclasses: 1) “TF bookmarked *de novo* enhancers” and 2) “TF virgin *de novo* enhancers” that are reminiscent to previously described latent enhancers[16, 35] expanding the senescence enhancer landscape. Next, we considered the chromatin state environment of the two *de novo* enhancer classes to further characterize them (Supplementary Figure 3h). While a chromatin state environment already rich in constitutive enhancers surrounded bookmarked *de novo* enhancers at T_0 (*i.e.*, pre-OIS stimulation; left top and bottom plots), a chromatin state environment poor in constitutive enhancer elements

surrounded virgin *de novo* enhancers at T_0 (right top and bottom plots). Both AP-1 bookmarked and virgin *de novo* enhancers became progressively activated and expanded upon RAS-OIS induction. Given that AP-1 premarked *de novo* enhancers operate within pre-existing, active enhancer-rich *cis*-regulatory regions and virgin *de novo* enhancers in poor ones, we hypothesized that this might impact absolute gene expression levels and kinetics upon enhancer activation. Indeed, we observed that the nearest genes associated with bookmarked *de novo* enhancers were already expressed at higher basal levels (as were genes proximal to poised enhancers) and reached significantly higher absolute expression levels with faster kinetics after RAS-OIS induction. In contrast, virgin *de novo* enhancers showed only low-to-background basal expression levels and reached comparatively lower absolute expression levels with slower kinetics after RAS-OIS induction (Figure 3e). These results argue that TF bookmarking of *de novo* enhancers, similar to traditional enhancer poising[36], is a chromatin-priming event that impacts gene expression kinetics and absolute output. Contrary to latent enhancers, our newly identified virgin enhancers do not serve an adaptive role in mediating stronger and faster gene expression upon restimulation as observed in macrophages[16], but instead serve as novel enhancer elements for *de novo* gene expression. Finally, we plotted leading gene expression fold-changes against the number of *de novo* enhancers in a given prospective senescence enhancer region. Remarkably, we discovered that a single *de novo* enhancer element of 100 bp can substantially activate the expression of its nearest gene and that there exists a positive correlation between the number of *de novo* enhancer elements and the expression increase of their nearest genes (Supplementary Figure 3i).

Altogether, our results provide compelling evidence that *de novo* and remnant enhancers play a critical role for ensuring that genes pertinent for senescence are

expressed at the correct time and the correct level and highlight the importance of AP-1 bookmarking for epigenetic memorization of past and future enhancer activity to define the senescence transcriptional program.

A hierarchical TF network defines the senescence transcriptional program

The combinatorial and dynamic binding of TFs to enhancers and their organization into TF networks are central to the spatiotemporal specificity of gene expression and a key determinant in cell fate decisions[37]. TF networks are frequently corrupted in disease and thus, a detailed understanding on TF networks has important implications for developing and improving new therapeutic strategies[38]. Currently, a TF network regulating senescence is not available, which precludes a deliberate therapeutic manipulation of the senescence phenotype. Importantly, TF networks deduced *in silico* from the integration of time-resolved multidimensional, genome-wide datasets improve the accuracy and predictive power of such networks.

To elucidate the combinatorial and dynamic binding of TFs to enhancers and their organization into TF networks, we first computed co-occurring pairs of TFs in enhancers in WI38 fibroblasts undergoing RAS-OIS, RAF-OIS, and RS (Figures 4a-c, Supplementary Figure 4a and Supplementary data: see under Code availability in Material and Methods). For RAS-OIS, we also applied a topic machine learning approach that dissects the complexity of combinatorial binding of many TFs into compact and easily interpretable regulatory modules or TF "lexicons" that form the thematic structures driving the RAS-OIS gene expression program (Figure 4d)[39, 40]. These analyses illustrated two key points. First, as shown in the co-binding matrices of Figures 4a-c and heatmap of Figure 4d, AP-1 pioneer TFs interact genome-wide with most of the remaining non-pioneer TFs (*i.e.*,

settlers and migrant TFs), have the highest total number of binding sites (Figure 4d, grey curve) and contribute to virtually all of the 54 TF lexicons (Figure 4d, green curve) with lexicon 22 being the most frequently represented lexicon genome-wide (Figure 4d, orange curve and Supplementary Figure 4c). Our interactive heatmap of Figure 4d (Supplementary data: see under Code availability in Material and Methods) provides a valuable resource for generating new hypotheses to functionally dissect TF interactions in cells undergoing RAS-OIS. Second, TF lexicon usage associates with specific chromatin states (Supplementary Figure 4b). For example, lexicons 21 and 22 are exclusively used for enhancers holding most of the AP-1 binding instances, while lexicon 50 is strongly related to polycomb repressor complex (PRC)-repressed regions and lexicons 44 and 52 predominantly associate with promoters (see Supplementary Figure 4c for specific examples). Interestingly among the most prominent TFs in lexicon 50 are the known PRC-interacting transcriptional co-repressor complex REST and insulator CTCF[41, 42]. The latter implies that these proteins may recruit PRC to silence or structure genomic regions, an intriguing possibility that deserves further investigation. Moreover, the promoter-centric lexicon 52 holds many E2F TFs, which is in line with a primary role of E2Fs at promoters[43].

Next, we developed an algorithm, based on our temporal TF co-binding information and a previously published TF network[44], to visualize the hierarchical structure of the senescence TF network. In Figure 4e we show a representative example of the TF network of SASP gene module VI. The network has a three-layered architecture: i) a top layer defined exclusively by the AP-1 family of pioneer TFs ii) a core layer composed mostly of other pioneer and settler TFs, and iii) a bottom layer characterized by settler and migrant TFs (Figure 4e and Cytoscape interactive maps in Supplementary Data: see

under Code availability in Material and Methods). The core layer itself separates into a multi-level and single-level core, depending on the complexity of TF connectivity to the top and bottom layers (Figure 4e). Remarkably, the organizational logic of the TF network is highly similar, if not identical, for all gene expression modules despite high transcription factor diversity in the core and bottom layers (Supplementary data: see under Code availability in Material and Methods). The TF network topology for RAS-OIS is congruent with the biochemical and dynamic properties of each TF category (*i.e.*, pioneer, settler or migrant) in each layer of the network. As the interactions flow from the top to the bottom layer, there is an increasing dynamicity and number of TFs and a decreasing number of bound regions (Supplementary Figures 4d,e). Ranking the dynamicity index and the number of bound regions for all TFs in each network confirmed the hierarchical principles of their organization, with a common core of highly connected TFs from the top and core layers shared across all networks (Supplementary Figure 4f, black circle at center). Variability in the composition of the most dynamic TFs of the core and bottom layers defines the gene expression module specificity for each network and its corresponding specialized transcriptional output (Supplementary Figures 4g-i). Thus, TF network topology imposes and constrains the position of a given TF in the network and thus, its gene-regulatory contribution. Our data also revealed unanticipated plasticity in transcription factor binding leading to similar gene expression, thus, refuting the simple rule that co-expression behooves co-regulation[45].

Our hierarchical TF network model for RAS-OIS enhancers predicted that the number of direct target genes regulated by a given TF is a function of its position in the TF network hierarchy. To test this prediction, we performed transient RNA interference (siRNA) experiments targeting AP-1-cJUN (top layer), ETS1 (multi-level core layer) and

RELA (single-level core layer) using two independent siRNAs per TF in fully senescent RAS-OIS cells (144 h), determined the global transcriptome profiles and compared them to the transcriptomes of cells transfected with a non-targeting siRNA (siCTRL) (Figure 4f and Supplementary Figures 4j-l). Consistent with the TF network hierarchy, silencing of AP-1-cJUN affected the most substantial number direct gene targets (n=5,089), followed by ETS1 (n=2,431) and RELA (n=2,224), thus, confirming the master regulatory role of AP-1 pioneer TFs at enhancers and in the execution of the RAS-OIS gene expression program. Specifically, 172 genes were co-regulated by the three TFs, while 987 were co-regulated by cJUN and ETS1, 520 by JUN and RELA, and 293 by ETS and RELA (Figure 4F). Correspondence analysis (CA) revealed that perturbing the function of AP-1-cJUN, ETS1 or RELA could separate faithfully ($p = 1.8 \times 10^{-149}$) up-regulated (V-VII) from down-regulated gene expression modules (I-IV) (Figure 4g), which aligns both with the CA for chromatin states (see Figures 2a,b) and the differential impact of the TFs on RAS-OIS-associated enhancer activation as predicted in the TF network analysis (Figure 4e and Supplementary data: see under Code availability in Material and Methods).

We conclude that the senescence response is encoded by a universal three-layered TF network architecture and relies strongly on the exploitation of an enhancer landscape implemented by AP-1 pioneer TFs to choreograph the OIS transcriptional program *via* local, diverse and dynamic interactions with settler and migrant TFs.

Hierarchy Matters: Functional Perturbation of AP-1 pioneer TF, but no other TF, reverts the senescence clock

Pioneer TFs have been identified as important drivers of cell fate changes during adaptive and cellular reprogramming as well as in cells undergoing malignant transformation[46, 47].

As such, they represent attractive targets to manipulate cell fate for diverse research and therapeutic purposes[20].

The identification of AP-1-cJUN as a principal pioneer TF in fibroblasts undergoing RAS-OIS raised the possibility that perturbing its function could considerably change the transcriptional trajectory of the OIS cell fate, while perturbation of other TFs should not. To test this hypothesis, we depleted AP-1-cJUN, ETS1 and RELA at T_0, 72 h and 144 h following oncogenic RAS expression using two independent siRNAs per TF, and compared global gene expression profiles with siCTRL treated cells at identical time-points. Knockdown efficiency for each siRNA was verified by RT-qPCR (see Supplementary Figures S4j-l). Capturing their transcriptional trajectories using PCA illustrated that functional perturbation of ETS1 and RELA shifted trajectories along the second principal component (PC2, which captures siRNA-related variability) at any given time-point compared to the control time course, but it did not affect the timely execution of the RAS-OIS gene expression program, since there is not shift along the first principal component (PC1, which captures time-related variability). By contrast, perturbing AP-1-cJUN function shifted trajectories both along PC1 and PC2 and effectively reverted the RAS-OIS transcriptional trajectory to a profile closely related to that of siCTRL-treated fibroblasts at 72 h after RAS-OIS induction. Silencing AP-1-cJUN expression at 72 h also pushed the transcriptional profile closer to control-treated cells at ~~day T0~~ T_0 (Figure 5A, blue arrow). Functional overrepresentation analyses of the target genes (direct and/or indirect) of each TF further supported the siJUN-mediated reversion of the RAS-OIS transcriptional trajectory demonstrating that depletion of AP-1-cJUN led to repression of the inflammatory response (*i.e.*, the SASP) and a partial reactivation of pro-proliferation

genes (*i.e.,* E2F, G2M and mitotic spindle targets) (Figure 5b and Supplementary Figures 5a-c). A complete exit of senescence is not expected here, however, as AP-1 is absolutely required for proliferation[48, 49].

To quantify and visualize the temporal overlaps in differentially expressed genes between siJUN and siCTRL-treated cells we used an UpSet plot (Supplementary Figures 5d) and expression heatmaps (Figure 5c and Supplementary Figures 5e-g). Congruent with a resetting of the senescence clock, a significant number of pro-proliferation E2F target genes (14%; *e.g. CCNB2* or *CDCA8*) were up-regulated (Supplementary Figure 5e), and NF-κB-regulated SASP target genes (*e.g. IL1ß* or *IL6*) were down-regulated (60%) (Supplementary Figure 5F) and a subset of p53 target genes (33%) was dysregulated by cJUN knockdown, indicating a functional interaction between AP-1 and p53, which we confirmed by interrogating our TF networks (Supplementary Figures 5g-i and Supplementary Table S7). cJUN-depleted RAS-OIS cells also shared a similar expression profile for a subset of genes (27%) of the Notch-1 intracellular domain (NC1ID)-induced senescence (NIS) transcriptional signature[50] that develops within the first 72-96 h of RAS-OIS (Supplementary Figure 5j). Altogether, these data identify AP-1 as a master regulator and molecular “time-keeper” of senescence. Our detailed description of the layered TF network architecture will facilitate targeted disruption of TFs to manipulate specific features of the senescence phenotype for future therapeutic benefit.

Functional role of AP-1 in therapy-induced senescence

Senescence exerts both cell-autonomous and non-cell- autonomous effects on (pre-) neoplastic lesions through a durable cell-cycle arrest, its secretome (SASP), stemness reprogramming and anti-apoptotic effects, all of which affect tumor aggressiveness and

clinical outcome after anti-cancer therapy[4, 51]. To corroborate our findings of AP-1 as a pioneering, master regulator of the senescence-associated gene expression program in fibroblasts undergoing RAS-OIS and, more specifically, the expression of SASP, stemness, and apoptosis mediators, we asked whether it may also play a decisive role during (chemo)therapy-induced senescence (TIS) *in vitro* and *in vivo*.

To this end, we first induced TIS by exposure to the chemotherapeutic agent Adriamycin (ADR) in two colorectal cancer (CRC) cell lines, HCT116 and SW480, overexpressing either a non-phosphorylatable, dominant-negative isoform of cJUN (cJUN4A) or empty vector control. Expression of cJUN4A had no measurable effect on senescence inducibility as indicated by a robust increase in SABG staining (Figures 6a,b) and a decrease in proliferation as measured by incorporation of BrdU (Supplementary Figures 6a,b). However, it significantly blunted TIS-induced transcriptional upregulation of cJUN in both CRC cell lines, which is consistent with the role of cJUN driving its expression[52] (Supplementary Figures 6c,d; compare empty/ADR and cJUN4A/ADR). Next, we measured the expression of selected AP-1-dependent SASP factors, stemness- and apoptosis-related genes, as well as E2F target genes that we had identified as being differentially expressed in fibroblasts undergoing OIS (see Figure 4f and Tables S7 and S8). Indeed, this analysis revealed dramatic repression of SASP (*e.g.*, *IL6, IL1ß,* and *MMP10*), stemness (*e.g., LIF*, *ABCG2*, and *CD44*) and anti-apoptotic (*e.g., BCL2A1*) target genes only in cJUN4A-expressing, but not in empty vector control cells. These results, therefore, indicate that AP-1 is critical for regulating the expression of these genes not only in human fibroblasts undergoing OIS-RAS but also in human colorectal cancer cell lines undergoing TIS (Figures 6c,d). Of note, E2F -target genes (e.g., CCNB2 and CDCA8) remained repressed in cJUN4A-expressing CRC cell lines, which seems to

indicate cell-type dependent differences when compared to our findings in RAS-OIS of fibroblasts.

To extend our findings to a primary tumor of different origin, we assessed the role of AP-1 in a well-established Eμ-myc transgenic mouse B-cell lymphoma model[4]. Consistent with the results in CRC cancer cell lines, overexpression of cJUN4A in primary murine B-cell lymphomas (stably expressing Bcl2 to block apoptosis) did not affect TIS establishment in response to ADR treatment, as documented by a robust increase in SABG-positive cells (Figure 6e). By contrast, it actively repressed the expression of selected AP-1 target genes similar to what we had observed in CRC cell lines undergoing ADR-TIS (Figure 6f). These results thus extend the critical role for AP-1 in the transcriptional regulation of the senescence-associated gene expression program from robust fibroblast models of senescence to disease-centered demonstrations of therapy-induced senescence in different cancer settings of human and murine origin.

Given these *in vitro* results, we hypothesized AP-1-dependent senescence to impinge on long-term outcome after anti-cancer therapy *in vivo*. To this end, we performed Gene Set Enrichment Analyses (GSEA), using an AP-1 senescence gene signature (see Table S9 and Material and Methods for details), first in a patient-reminiscent, primary Eμ-myc lymphoma-based clinical trial-like mouse cohort exposed to Cyclophosphamide (CTX) *in vivo*. The AP-1 senescence gene signature was significantly enriched (FDR = 0,018) at diagnosis (*i.e.*, before any drug encounter) in lymphomas that initially responded to CTX treatment before eventually relapsing (designated “relapse-prone [RP]”), discriminating them clearly from the same set of lymphomas subsequently presenting as full-blown resistance to repetitive administrations of CTX (designated “resistant [RES]”) (Figure 6g). Consequently, we then asked whether a humanized version of this AP-1

senescence signature would also be enriched for in human diffuse large B-cell lymphoma (DLBCL) material obtained at diagnosis from patients achieving lasting tumor control ("tumor-free") in response to R-CHOP induction therapy (*i.e.*, CD20 antibody Rituximab plus CTX/ADR/Vincristine/Prednisone standard care chemotherapy for DLBCL). Remarkably, two publicly available independent datasets (GSE3131253 and GSE9858854), comprising data on lymphoma transcriptomes at diagnosis and the clinical courses of DLBCL patients, exhibited a highly significant enrichment (FDR=0,0018 and 0,0093) for the AP-1 senescence gene signature in long-term tumor-free DLBCL patients compared to those that relapsed after R-CHOP therapy (Figures 6h,i). Thus, these results designate a critical role for AP-1 in mediating TIS *in vivo* and are consistent with the previous demonstration that the ability of lymphomas to undergo TIS improves disease outcome[53].

Collectively, our data emphasize the physiological importance of the AP-1-governed senescence-associated gene expression program, highlight its contribution to the long-term outcome after anti-cancer therapy *in vivo*, and warrant further investigations on modulating AP-1 activity as a promising therapeutic target for senescence therapies.

DISCUSSION

Exploiting senescence targeting for treating age-related diseases and cancer requires a detailed knowledge of the transcriptional, epigenetic, and signaling mechanisms defining the basis and realization of the senescence program, which is currently missing. To fill this critical gap in our knowledge, we used a dynamic, multidimensional approach at high-resolution to define the gene-regulatory code driving the senescence cell fate.

A central finding of our study is that the senescence program is defined and driven by a predetermined enhancer landscape that is sequentially (in)activated during the senescence process. AP-1 is instrumental for this predetermination by imprinting a prospective senescence enhancer landscape that, in the absence of traditional enhancer histone-modification marks, foreshadows future transcriptional activation. This is a surprising discovery given that AP-1 transcriptional activation has been traditionally linked to growth-factor and MAPK signaling[54]. There is, however, now accumulating evidence that AP-1 also plays an essential role as a pioneering factor for establishing cell type-specific enhancers and cellular identities[55, 56]. In line with its role in pioneering and bookmarking enhancers, we show that AP-1 is also recruited *de novo* as a first line TF to "virgin" enhancers and that it serves as a molecular memory for enhancers that become inactivated during the senescence fate transition that we termed "remnant" enhancers. Based on these findings we propose a model by which "enhancer recycling" of AP-1-bookmarked future and past enhancer activities is an evolutionary conserved mechanism that allows for modular and flexible, yet, efficient transcriptional responses to incoming signals. We stipulate that the senescence program is preserved through AP-1 binding to enhancer chromatin as part of epigenetic memory of the cell's developmental (stress) history bypassing histone modification-dependent bookmarking to store genomic

information. Further, given the pristine specificity of the newly identified prospective and remnant enhancers they can be used as urgently needed specific, rather than associated, senescence biomarkers and to predict a cell's potential to undergo senescence. This latter carries also important translational implications for identifying cancers that would respond positively to pro-senescence therapy. A natural question that arises from our data is whether the senescence program is universal to all inducing stimuli and cell types or if multiple senescence programs exist. Based on the data presented here and work in progress, we predict that the organizational principles of the senescence program we defined here hold for all cell types and inducers. Additional time-resolved studies of various inducers in different cell types are required and currently ongoing to answer this question definitively.

Another key finding is the reversibility of senescence by an informed intervention on network topology that we delineated in this study. Indeed, silencing the function of a single TF sitting atop the TF network hierarchy, AP-1, is sufficient to revert the "senescence clock". We thus define after the "telomere clock" a second, "epigenomic-enhancer clock" regulating the senescence process. Why does functional AP-1 perturbation not lead to complete senescence exit? Based on published[49] and our own results we surmise that AP-1 depletion does not lead to a full cell cycle re-entry and proliferation, because AP-1 plays important roles for proliferation. Thus, AP-1 confines cells to their existing proliferative state and therefore may be viewed as a 'locking device' that restricts cells to their current state. However, we provide compelling evidence that AP-1 is critical for the expression of the senescence-associated secretory phenotype (SASP) genes both in different cell biology models of senescence and in an in vivo model of therapy-induced senescence (Figures 5, 6 and Supplementary Figures 5 and 6). Importantly, we

demonstrate that an AP-1 senescence gene signature positively correlates with disease outcome after TIS in lymphomas both in mouse and humans. These findings therefore emphasize the importance of AP-1 in endowing cancer cells with the ability to undergo TIS *in vivo*, which acts as a desirable effector principle for long-term outcome to anti-cancer therapy[53]. However, continued expression and secretion of SASP and stemness factors could drive reprogramming and escape from the senescence arrest[4], resulting in relapse. Therefore, while targeting AP-1 using selective inhibitors (*e.g.* improved T-5224 derivatives) would be desirable to control effectively the detrimental effects of the SASP in promoting cancer[57-59], choosing the moment of intervention will require careful evaluation of disease progression during anti-cancer therapy given AP-1 critical role in mediating TIS. In summary, we believe that AP-1 is a prime target for therapeutic modulation of the senescence phenotype *in vivo*.

By determining the layered architecture/organizational principles of the TF network that orchestrate(s) the transition to OIS, we revealed the plasticity and stability of the senescent phenotype. We show that a highly flexible, combinatorial TF interactome establishes the senescence program, which is in line with the TF network dynamics during hematopoietic and stem cell differentiation[60, 61]. In addition, we demonstrate that targeted engineering of specific nodes at different layers of the TF network disrupts gene expression with a corresponding magnitude, suggesting a path for the manipulation of the senescent phenotype *in vivo*. Pharmacological inhibition of TFs (see above for AP-1), signal transduction molecules, such as kinases or acetylases that converge in the activation of TFs, could represent a viable approach for manipulating the senescent phenotype *in vivo*[62]. Alternatively, small molecules that prevent TF-TF combinatorial interactions could also be envisioned[63].

In conclusion, the present work emphasizes the advantages of, and indeed the need for, integrating time-resolved genome-wide profiles to describe and interrogate the transition to senescence. This approach generates detailed knowledge necessary to develop strategies for manipulating/engineering the senescent cell fate (and other cell fate transitions) *in vivo* for research and therapeutic purposes. Overall, our study provides a comprehensive resource for the generation of novel hypotheses regarding senescence regulation, offers important mechanistic, regulatory insights that could translate to the study of other cell fate transitions and provide new inroads for the diagnosis and manipulation of the senescence state in age-related diseases and cancer.

MATERIAL AND METHODS

Cell culture

WI-38 fibroblasts (purchased from ECCAC) were cultured in Dulbecco's Modified Eagle's medium (DMEM) containing 10% FBS and 1X Primocin (Invivogen) at 37ºC and 3% oxygen. WI-38-ER:RASV12 fibroblasts were generated by retroviral transduction as previously described[10]. Senescence was induced by addition of 400 nM 4-hydroxytamoxifen (4-OHT) to the culture medium and samples were collected and processed at the time points indicated in the main text. GM21- skin fibroblasts were cultured as WI38 fibroblasts. GM21 skin fibroblasts constitutively expressing RASV12 or empty vector control were generated by retroviral transduction as previously described[10]. Doxycycline-inducible oncogenic bRAFV600E retroviral construct was a gift from Carl Mann (CEA, Gif-sur-Yvette, France). Oncogenic RAF-induced senescence (RAF-OIS) was induced in WI-38 fibroblasts with 100ng/ml doxycycline, and cells were collected and processed correspondingly at the time points indicated in the text. Replicative senescent cells were generated by proliferative exhaustion under 21% O_2 and were subsequently collected and processed at the indicated times in the main text. For the induction of quiescence, WI-38 fibroblasts were cultured in DMEM containing 0.2% FBS for up to 4 consecutive days and samples were collected and processed as described in the main text.

ATAC-seq

The transposition reaction and library construction were performed as previously described[25]. Briefly, 50,000 cells from each time point of the senescence time course (2 biological replicates) were collected, washed in 1X in PBS and centrifuged at 500 x *g* at

4ºC for 5 min. Nuclei were extracted by incubation of cells in Nuclear Extraction Buffer (NEB) containing 10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% IGEPAL CA-630 and immediately centrifuging at 500 x *g* at 4ºC for 5 min. The supernatant was carefully removed by pipetting, and the transposition was performed by resuspension of nuclei in 50 µL of Transposition Mix containing 1X TD Buffer (Illumina) and 2.5 µL Tn5 (Illumina) for 30 min at 37ºC. DNA was extracted using the QIAGEN MinElute kit. Libraries were produced by PCR amplification (12-14 cycles) of tagmented DNA using the NEB Next High-Fidelity 2x PCR Master Mix (New England Biolabs). Library quality was assessed in an Agilent Bioanalyzer 2100. Paired-end sequencing was performed in an Illumina Hiseq 2500. Typically, 30-50 million reads per library are required for downstream analyses.

Histone modification and transcription factor ChIP-seq

WI-38-ER:RASV12 fibroblasts were treated with 400 nM 4-OHT for 0, 72 and 144 hours, doxycycline-inducible bRAFV600E-expressing WI-38 fibroblasts were treated with 100 ng/ml doxyxcline for 0, 48 and 96 hours, replicative senescent cells (0, 264, 1008, 2112 h) were generated as described, and 10^7 cells (per time point, minimum two biological replicates) were fixed in 1% formaldehyde for 15 min, quenched in 2M glycine for additional 5 min and pelleted by centrifugation at 2,000 rpm, 4ºC for 4 min. For histone modification ChIP-seq, nuclei were extracted in Extraction Buffer 2 (0.25 M sucrose, 10 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 1% Triton X-100 and proteinase inhibitor cocktail) on ice for 10 min followed by centrifugation at 3,000 x *g* at 4ºC for 10 min. The supernatant was removed and nuclei were resuspended in Nuclei Lysis Buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS and proteinase inhibitor cocktail). Sonication was performed

using a Diagenode Picoruptor until the desired average fragment size (100-500 bp) was obtained. Soluble chromatin was obtained by centrifugation at 11,500 rpm for 10 min at 4ºC and chromatin was diluted 10-fold. Immunoprecipitation was performed overnight at 4ºC with rotation using 1-2 x 10^6 cell equivalents per immunoprecipitation using antibodies (5 µg) against H3K4me1 (Abcam), H3K27ac (Abcam), H3K4me3 (Millipore, only used for RAS-OIS), H3K27me3 (Millipore, only used for RAS-OIS). Subsequently, 30 µL of Ultralink Resin (Thermo Fisher Scientific) was added and allowed to tumble for 4h at 4ºC. The resin was pelleted by centrifugation and washed three times in low salt buffer (150 mM NaCl, 0.1% SDS, 1% Triton X-100, 20 mM EDTA, 20 mM Tris-HCl pH 8.0), one time in high salt buffer (500 mM NaCl, 0.1% SDS, 1% Triton X-100, 20 mM EDTA, 20 mM Tris-HCl pH 8.0), two times in lithium chloride buffer (250 mM LiCl, 1% IGEPAL CA-630, 15 sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.0) and two times in TE buffer (10 mM Tris-HCl, 1 mM EDTA). For transcription factor ChIP-seq, fibroblasts were treated as described above except that chromatin was isolated using the enzymatic SimpleChIP kit (Cell Signaling) according to the manufacturer's instructions, obtaining chromatin with an average fragment length of 4-5 nucleosomes. Immunoprecipitation was performed overnight at 4ºC with rotation using 6-10 x 10^6 cell equivalents per immunoprecipitation using antibodies (5 µg) against cJUN, FOSL2 and RELA (Santa Cruz Biotechnologies) and processed as described above. Washed beads were resuspended in elution buffer (10 mM Tris-Cl pH 8.0, 5 mM EDTA, 300 mM NaCl, 0.5% SDS) treated with RNAse H (30 min, 37 ºC) and Proteinase K (2 h, 37ºC), 1 µL glycogen (20 mg/mL, Ambion) was added, and decrosslinked overnight at 65 ºC. For histone modifications, DNA was recovered by mixing the decrosslinked supernatant with 2.2X SPRI beads followed by 4 min incubation at RT. The SPRI beads were washed twice in 80% ethanol, allowed to dry, and DNA was

eluted by in 35 µL 10 mM Tris-Cl pH 8.0. For transcription factors, DNA was eluted by phenol/chloroform extraction (2X) followed by ethanol precipitation overnight at -20ºC. The DNA pellet was washed with 70% ethanol, allowed to dry, and DNA was resuspended in 35 µL 10 mM Tris-Cl pH 8.0. Histone modification libraries were constructed using the NextFlex ChIP-seq kit (Bioo Scientific) according to the manufacturer's instructions. Libraries were amplified for 12 cycles. Transcription factor libraries were constructed using a modified protocol from the Accel-NGS 2S Plus DNA Library Kit (#21024), where we performed DNA extraction at each step using 25:24:1 phenol:chloroform:isoamyl alcohol followed by overnight ethanol precipitation of DNA at each step of the protocol. Additionally, we enriched for small DNA fragments using AMPure-XP beads (Beckman-Coulter (#A63881). Libraries were then resuspended in 20 µL of low EDTA-TE buffer. Libraries were quality controlled in an Agilent Technologies 4200 Tapestation (G2991-90001) and quantified using the Invitrogen Qubit DS DNA HS Assay kit (Q32854). Libraries were sequenced using an Illumina High-Seq 2500. Typically, 30-50 million reads were required for downstream analyses.

RNA and microarrays

RNA from each time point from the different senescence models and quiescence time series, as well RAS-OIS cells treated with control and ETS1-, RELA- and JUN-targeting siRNAs (2 biological replicates) was purified using the QIAGEN RNeasy Plus kit according to the manufacturer's instructions. 100 ng RNA per sample was analyzed using Affymetrix Human Transcriptome Arrays 2.0, according to the manufacturer's instructions.

EdU staining and senescence-associated beta galactosidase activity (SABG)

Representative samples from the senescent and quiescent time series were evaluated for EdU incorporation using the Click-iT EdU Alexa Fluor Imaging Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. SABG activity was assessed as previously described[64]. Cells were imaged in a Zeiss confocal fluorescence microscope and images analyzed using the ZEN suite.

RNA interference

Small interfering RNAs (20 μM) targeting JUN (Dharmacon), ETS1 (QIAGEN) and RELA (QIAGEN) as well as non-targeting controls were transfected into WI-38-ER:RAS^{V12} using siIMPORTER reagent (Millipore) according to the manufacturer's instructions (2 biological replicates per transcription factor per time course experiment). Transfections were performed in triplicate wells and cells from each siRNA treatment were pooled for RNA purification. RAS-OIS was induced with 400 nM 4-OHT concomitantly with the addition of DMEM containing 20% FBS 4 hours after transfection and incubated overnight. Sixteen hours after transfection, cells were replenished with new media containing 10% FBS and 400 nM 4-OHT, and RNA was isolated at indicated time points and analyzed in Affymetrix Human Transcriptome Arrays 2.0.

Expression microarray pre-processing

Raw Affymetrix HTA 2.0 array intensity data were analyzed using open-source Bioconductor packages on R. The quiescence and the RAS-OIS time series data were normalized together (2 conditions, 2 biological replicates per condition, 6 time points per replicates) using the robust multi-array average normalization approach implemented in

the *oligo* package. Internal control probe sets were removed and average expression deciles over time-points were independently defined for each treatment. Probes whose average expression was lower than the 4th expression decile in both conditions were removed for subsequent analyses. To remove sources of variation and account for batch effects, data were finally corrected with the *sva* package. To recover as much annotation information as possible, we combined Affymetrix HTA 2.0 annotations provided by Affymetrix and Ensembl through the packages *hta20sttranscriptcluster.db* and *biomaRt*. Principal component analysis and bi-clustering based on Pearson's correlation and Ward's aggregation criterion were used to confirm consistency between biological replicates and experimental conditions at each step of the pre-processing.

Self-organizing maps (SOM)

Normalized log-scaled and filtered expression values were processed using the unsupervised machine learning method implemented in *oposSOM*[26] to train a self-organizing map. This algorithm applies a neural network algorithm to project high dimensional data onto a two-dimensional visualization space. In this application, we used a two-dimensional grid of size 60 x 60 metagenes of rectangular topology. The SOM portraits were then plotted using a logarithmic fold-change scale. To define modules of co-expressed meta-genes, we used a clustering approach relying on distance matrix and implemented in *oposSOM*. Briefly, clusters of gene expression were determined based on the patterns of the distance map which visualizes the mean Euclidean distance of each SOM unit to its adjacent neighbors. This clustering algorithm – referred to as D-clustering – finds the SOM units referring to local maxima of their mean distance with respect to their neighbors. These pixels form halos edging the relevant clusters in the respective distance

map and enable robust determination of feature clusters in the SOM. We finally performed a gene set over-representation analysis in each cluster considering the Molecular Signature Database (MSigDB) hallmark gene sets using a right-tail modified Fisher's exact test and the hypergeometric distribution to provide *p*-value.

Correlation and multidimensional analyses

To highlight differences in expression profiles between quiescence and RAS-OIS through time, we used multi-dimensional scaling plot representing leading fold change, which is defined as the root-mean-square average of the log-fold-changes for the genes best distinguishing each pair of samples. To quantify the evolution of transcriptomic variability and noise through time, we looked at the gene expression density distributions for all possible pairs of treated *vs* T_0 transcriptomes. Distributions were estimated using kernel density estimation of all genes' expression in the i^{th} T_0 transcriptome and the j^{th} treated transcriptome. We also computed Pearson's correlation for each of these combinations. The Pearson's correlation between two transcriptomes, X and Y containing n gene expressions, is obtained by $R(X,Y) = \sum_{i=1}^{n}(x_i - \mu_X)\,(y_i - \mu_Y)/(\sigma_X\sigma_Y)$, where x_i and y_i are the i^{th} observation in the vectors X and Y respectively, μ_X and μ_Y the average values of each transcriptome, and σ_X and σ_Y, the corresponding standard deviations.

Information theory – derived metrics

To evaluate transcriptome diversity and specialization, we used an approach based on information theory as described in [27].

Gene expression time series analysis

Normalized log-scaled and filtered expression data related to the quiescence and the OIS time series were further considered for differential analysis with *limma*[65]. To define an RAS-OIS specific transcriptomic signature, we proceeded in three steps, each relying on linear mixed model cubic B-splines, as nonlinear response patterns are commonly encountered in time course biological data. For each probe, and each treatment the expression was modeled as follow:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \sum_{k=0}^{K-1} \gamma_k (x - \xi_k)^3 + \varepsilon$$

$$\text{with } (x - \xi_k) = \begin{cases} 1 if x \leq \xi \\ x - \xi if x > \xi \end{cases}$$

where $\beta_0$is the average probe expression over all samples in a given condition, β_{1-3} the model coefficients, K the number of knots, ξ_k the k^{th} knot and ε the error term. First, we defined probes responding over time to RASV12 induction. Second, we considered all together the quiescence and the RAS-OIS time series, as well as the interaction between time and treatment, and defined probes responding to one or the other treatment over time, as well as probes responding differently between the two treatments at any time point. We finally defined the set of probes responding consistently to both treatment and time and removed these probes from the global set of probes responding to RASV12 induction defined at the first step. Moderated *F*-statistics that combine the empirical Bayes moderated *t*-statistics for all contrasts into an overall test of significance for each probe were used to assess the significance of the observed expression changes. At any step of this workflow, *p*-values were corrected for multiple testing using the FDR approach for a stringent significance level of 0.005. For validation purposes, we compressed the RAS-OIS time-series to achieve a volcano plot representation. To this end, we computed the

maximal absolute $\log_2$ fold change in expression in the RAS-OIS time series considering T_0 as the reference and selected up and down regulated probes using an absolute $\log_2$ fold change cutoff at 1.2 and a corrected *p*-value cutoff of 0.005. We then build a scatter-plot plotting the $\log_{10}$ significance versus $\log_2$ fold-change on the y and x axes, respectively. Probes responding consistently to both ER: RASV12 induction and quiescence were finally over-plotted.

Gene expression unsupervised clustering

Probes constitutive of the RAS-OIS specific transcriptomic signature were clustered using the weighted gene correlated network analysis approach implemented in the *WGCNA* R package[66]. Standard WGCNA parameters were used for the analysis, with the exceptions of soft-thresholding power, which was defined using methods described by and set at 18. The 7 co-expressed probe clusters identified were further functionally characterized using gene set over-representation tests. The same approach as previously described for the SOM-defined clusters was used.

Histone modification ChIP-seq data processing

Reads were cleaned and trimmed using *fastq-mcf* from the *ea-utils* suite v1.1.2 to remove adapters, low quality bases and reads, and discard reads shorter than 25 bp after filtering. Reads were then aligned to the human reference genome (hg19) with *bowtie* v1.1.1 using best matches parameters (bowtie -v 2 -m 1 --best --strata). Alignment files were further processed with *samtools* v1.2 and *PicardTools* v1.130 to flag PCR and optical duplicates and remove alignments located in Encode blacklisted regions. Fragment size was estimated *in silico* for each library using *spp* v1.10.1. Genome-wide consistency between

replicates was checked using custom R scripts. Enriched regions were identified for each replicate independently with *MACS* v2.1.0 with non-IPed genomic DNA as a control (macs2 callpeak --nomodel --shiftsize --shift-control --gsize hs -p 1e-1). These relaxed peak lists were then processed through the irreproducible discovery rate (IDR) pipeline[67] to generate an optimal and reproducible set of peaks for each histone modification and each time point.

ATAC-seq data processing

Paired-ends reads were cropped to 100bp with *trimmomatic* v0.36[68] and cleaned using *cutadapt* v1.8.3[69] to remove Nextera adapters, low quality bases and reads, and discard reads shorter than 25 bp after filtering. Fragments were then aligned to the human reference genome (hg19) using *bowtie2* v2.2.3 discarding inconsistent pairs and considering a maximum insert size of 2kb (bowtie2 -N 0 --no-mixed --no-discordant --minins 30 --maxins 2000). Alignment files were further processed with *samtools* v1.2 and *PicardTools* v1.130 to flag PCR and optical duplicates and remove alignments located in Encode blacklisted regions. Accessible regions were identified using *MACS2* v2.1.0 without control (macs2 callpeak --gsize hs -p 1e-3). These relaxed peak lists were then processed through the irreproducible discovery rate (IDR) pipeline to generate an optimal and reproducible set of peaks for each time point.

Normalized ATAC-seq and ChIP-seq signal tracks

After verifying the consistency between biological replicates, time points and data type using *deepTools*[67], alignments related to biological replicates for a given assay and a given time point were combined. We then binned the genome in 200bp non-overlapping

windows and generated genome-wide read count matrices for each assay independently. These matrices were finally quantile normalized with custom R script and further used to generate genome-wide signal tracts.

Histone modification ChIP-seq and ATAC-seq differential analysis

After assessing library saturation using *preseqR*, alignment and peak data were imported and pre-processed in R using the *DiffBind* package[68]. Briefly, for a given histone modification type, we first defined the global reproducible peak set as the union of each time-specific reproducible peak sets defined previously. We then counted the number of reads mapping inside each of these intervals at each time point and for each replicate. The raw count matrix was then normalized for sequencing depth using a non-linear full quantile normalization as implemented in the *EDASeq* package[70]. To remove sources of unwanted variation and consider batch effects, data were finally corrected with the *RUVSeq*[71] package considering 2 surrogate variables. Differential analyses for count data were performed using *edgeR*[72] considering time and batch in the design matrix, by fitting a negative binomial generalized log-linear model to the read counts for each peak. Peaks were finally annotated using *ChIPpeakAnno* considering annotations provided by Ensembl v86.

Chromatin state differential analysis

To quantify and define combinatorial chromatin state dynamics in space and time, we analyzed histone modification combinations with the *chromstaR* package[73]. Briefly, after partitioning the genome into 100bp non-overlapping bins and counting the number of reads mapping into each bin at each time point and for each histone modification, this

algorithm relies on a univariate Hidden Markov Model (HMM) with two hidden states (unmodified, modified). This HMM is used to fit the parameters of the two-component mixture of zero-inflated negative binomial distribution considered to model read counts for every ChIP-seq experiments. A multivariate HMM is then used to assign every bin in the genome to one of the multivariate components considering $2^{(3 \text{ time points x } 4 \text{ histone modifications})}$ possible states. To limit computational burden and focus on accurate differences, the analysis was run in differential mode with a 100bp resolution (*i.e.* smaller than a single nucleosome), such that every mark is first analyzed separately with all conditions combined while the full combinatorial state dynamics is rebuilt by combining the differential calls obtained for the four marks. We finally filtered out differential calls not overlapping with any histone modification and ATAC-seq reproducible peaks. To properly associate histone modification combinations with biologically meaningful mnemonics, we made an extensive comparison between the binning we obtained in WI38 fibroblasts undergoing RAS-OIS and IMR90 fetal lung fibroblasts chromatin states described in the scope of the Epigenomic Roadmap consortium. To test for association between changes in chromatin states through time and gene expression modules we ran a correspondence analysis. Briefly, genomic loci experiencing changes in chromatin states through time were first associated to the nearest gene. We then specifically focused on loci associated to genes belonging to any expression module and built a two-way contingency table summarizing the number of transition in states (considering all possible combinations) occurring in each expression module, further used as an input for a correspondence analysis using *FactoMineR*[74]. The significance of association between the two qualitative variables (transition in state and module) was assed using a χ^2 test. Results of the CA were visualized using a row-metric-preserving contribution asymmetric biplot and filtering

for the top contributing and well-projected (squared cosine > 0.5) changes in chromatin states.

Motif enrichment analysis in active enhancers

For each time point independently, we defined the set of active enhancers as the overlap between H3K4me1, H3K27ac and ATAC-seq reproducible peaks using *bedtools*[75]. We then ran 3 independent motif enrichment analyses with *homer* v4.9 using default parameters.

Transcription factor footprinting

All transcription factor Position-Weight Matrices (PWM) representing eukaryote transcription factors were downloaded from the JASPAR database and used as an input for PIQ[22] to predict transcription factor binding sites from the genome sequence on down-sampled ATAC-seq alignments. For each motif, we retained only binding sites that were within the reproducible ATAC-seq peaks and passed the default purity cut-off (70%). We then computed pairwise PWM similarities based on Pearson's correlation, and clustered together PWMs sharing more than 90% similarity, defining a set of 310 non-redundant and distinct PWMs. The Pearson's correlation between two PWM P^1 and P^2 of length l was defined as:

$$r(P^1, P^2) = \frac{1}{l} \times \sum_{i=1}^{l} \frac{\sum_{b \in \{A,C,G,T\}} \left(P^1_{i,b} - 0.25\right)\left(P^2_{i,b} - 0.25\right)}{\sqrt{\sum_{b \in \{A,C,G,T\}} \left(P^1_{i,b} - 0.25\right)^2 \times \sum_{b \in \{A,C,G,T\}} \left(P^2_{i,b} - 0.25\right)^2}}$$

We further combined the bound instances identified with PIQ according to the PWM clustering.

Transcription factor metrics

For each transcription factor, we computed the chromatin-opening index (COI), the motif dependence and the chromatin dependence (CD) following the approach described in [22].

Validation of PIQ predictions through ChIP-seq

To compare PIQ prediction with RELA, JUN and FOSL2 ChIP-seq data, we first used the approach suggested in [22], computing how many of the total ChIP-seq peaks are overlapping with any potential factor motif (since ChIP-Seq peaks can result from co-factor binding, and methods such as digital genomic footprinting are factor agnostic). We then used a more sophisticated approach aiming at correlating the ChIP-seq signal intensity with the bound / unbound status at PWM matches. For a given transcription factor (cJUN, FOSL2 or RELA,), we first considered all the PWM matches located inside ATAC-seq reproducible peaks, we selected all the PWM matches assigned with a purity score > 0.7 (the threshold used to define “bound” instances), and then randomly selected 3 times more PWM matches assigned to a purity score < 0.7 (considered as “unbound” instances) to obtain a global set containing 25% / 75% of bound / unbound instances for each TF. The selected regions were extended up to 2kb (1kb in each direction, from the middle of the match), and the 2kb intervals were binned in one hundred 20bp windows. We computed the normalized ChIP-seq and ATAC-seq signal inside each bin. The windows were finally ranked according to the summed ChIP-seq signal in the 10 most central bins (200bp). We finally run a set enrichment analysis with the *fgsea* package to assess whether bound / unbound PWM matches were enriched / depleted along this ranking and computed the enrichment score (ES, positive when bound instances are enriched for

highest ChIP-seq signals, negative when unbound instances are depleted for highest ChIP-seq signals) and *p*-values which revealed the strength of the correlation. We performed 1,000 permutations to obtain *p*-values.

Transcription factor co-binding

For every cluster of PWM and time-point independently, we first removed all the bound instances identified outside enhancers. The remaining bound instances for all PWM were then combined for every time point using GEM *regulatory module discovery*[40] setting at 500 bp the minimal distance for merging nearby TF bound instances into co-binding regions and at 3 the minimum number of TF bound instances in a co-binding region.

Global pairwise co-binding heatmap. At this step, we obtained a set of contingency matrices M_{mt} of dimension $n_{mt} \times j$ with *i* the number of co-binding regions for the transcriptomic module *m* at the time point *t* and *j* = 310 PWM clusters, for each time point and each transcriptomic module. We then generated module- and time- specific normalized pairwise co-binding matrices C_{mt} by computing the normalized cross-product of matrices M_{mt} defined as:

$$C_{mt} = \frac{M_{mt} \times {}^{t}M_{mt}}{\sum_t \sum_m \sum_j a_{tmj}} \times 10^6$$

with a_{tmj} the number of bound instances for the PWM clusters *j*, in transcriptomic module *m*, at the time point *t*. To get a global picture of pairwise co-binding, we summed these matrices and tested for each combination of PWM clusters A and B whether the overlap between bound instances for A and B was significant using a hyper-geometric test defined as:

$$p(Q,M,n,k) = \sum_{m=k}^{min\{k,B\}} \frac{\binom{M}{m}\binom{Q-M}{n-m}}{\binom{Q}{m}}$$

where *Q* is the overall number of regions in the universe, *M* is the number of regions bound by A, *n* is the number of regions bound by B, and *k* the total number of regions bound by A and B. The *p*-values were further corrected for multiple testing using the Bonferroni strategy. We finally clustered the co-binding occurrence matrix using Ward's aggregation criterion and projected corresponding corrected *q*-values on this clustering.

Pair-wise co-binding circos plots. To generate the co-binding circos plots, we used the global time- and, module-specific pair-wise normalized co-binding matrix C_{mt} described above, after a logarithmic transformation. For each time-point and module independently, we selected the top 500 interactions based on their occurrence *N*. The images were generated using the *Circos* suite[76].

Identification of TF regulatory modules

We used the data-sets generated using GEM regulatory module discovery described above. We applied a Hierarchical Dirichlet Process topic model which automatically determines the number of topics from the data, with the hyperparameter for the topic Dirichlet distribution set at 0.1 (encoding the assumption that most of the topics contains a few TFs) and the maximum number of iterations set at 2000. The lexicon usage for each time point and each transcriptomic was explored using a multiple factor analysis (MFA) with the R package *FactoMineR,* and lexicons were further selected based on their goodness of representation on the 3 first components (squared cosine > 0.5).

TF properties

With the aim of characterizing the binding properties of each TF, we computed the dynamicity, the total number of bound regions, the fraction of bound regions in enhancers and the fraction of bound regions before stimulation.

Dynamicity. We quantified the dynamicity of a TF accordingly to the following expression:

$$d(A) = \frac{\sum \frac{n_t(A)}{TR_t}}{\sum \frac{t_t(A)}{TR_t}}$$

where $d(A)$ is the dynamicity of TF A; $n_t(A)$ is the number of regions bound by A for the first time at time point t ; $t_t(A)$ is the number of regions bound by A at time point t and TR_t is the number of regions bound by any TF in time point t. The factor TR_t was added to the expression to account for differences in the number of reads sequenced by the ATAC-seq protocol and normalizes the number of regions bound by TF A based on the number of bound regions detected at its corresponding time point. Notice that, if all samples have the same amount of TF binding events, this expression is reduced to the quotient of the sum of the regions first bound at each time point by the sum of all regions bound by the TF at each time point. By using this definition, the function $d(A)$ maps the activity of a TF to the interval $\left[\frac{1}{N_t}, 1\right]$, where N_t is the number of time points in the timecourse and is higher as the TF binds to previously not bound regions or leaves already bound regions. In the case of a TF that, for every time point, leaves all its previous bound regions and binds to only regions not previously bound, the numerator will be identical to the denominator, leading to $d(A) = 1$. Alternatively, if a TF remains on the same regions it has bound at $t = 0$, then $\sum n_t = n_0$ and $\sum t_t = N_t * n_0$, resulting in $d(A) = \frac{1}{N_t}$. One can

observe that, if the same region is bound by TF A in different time points, it will contribute once to the numerator of the expression, while it will contribute to the denominator once for each time point it has been bound to.

Total number of bound regions. The number of bound regions was calculated by the following the expression:

$$R(A) = \frac{\sum \frac{n_t(A)}{TR_t} \times \sum TR_t}{N_t}$$

where $R(A)$ is the normalized number of bound regions by TF A during the timecourse and $n_t(A)$, TR_t and N_t are defined as above. The first factor is a normalized sum of the regions bound by TF A, counting each region only once. The second factor scales the result by the mean of the number of regions bound by all TFs on each day.

TF percentage of binding at enhancers. The ratio of binding at enhancers, relative to all cis regulatory regions, was assessed by:

$$P_E(A) = \frac{R_E(A)}{R_E(A) + R_P(A)}$$

where $P_E(A)$ is the percentage of bound regions in enhancers for TF A, $R_E(A)$ is the number of regions bound by TF A marked as enhancers and $R_P(A)$ is the number of regions bound by TF A marked as promoters.

TF prestimulation binding. For each TF, we computed the ratio of regions bound at T_0, relative to the number of regions bound during the whole timecourse. We used the following definition for the prestimulation binding factor for each TF:

$$p(A) = \frac{\frac{n_{D0}(A)}{TR_{D0}}}{\sum \frac{n_t(A)}{TR_t}}$$

where $p(A)$ corresponds to the prestimulation binding of TF A and $n_t(A)$ and TR_t are defined as above. The numerator of this expression corresponds to the normalized number of regions bound by TF A at t = T_0, while the denominator is the normalized number of regions bound by TF A during the whole timecourse. Notice the denominator also corresponds to factor $R(A)$ before scaling.

Hierarchical transcription factor network

In order to assess the TF chromatin binding hierarchy, i.e. TFs required for the binding of a specific TF, we generated a network for each gene module depicting the precedence of TF chromatin binding. The algorithms mentioned were implemented in R and all networks were visualized in CytoScape[77].

Computing precedence relationships. The edges in the generated networks represent the precedence relationship of TFs: an oriented edge from TF A to TF B, represented as (A, B), means that A was present in at least 30 % of the cis-regulatory regions bound by B at the same instant or before[44]. To account for the difference in the number of reads sequenced for each sample in the ATAC-seq, we normalized the number of regions bound based on the first day they appeared. The weight of an edge from A to B is given by:

$$w_{A\to B} = \frac{\sum \frac{R_t(A,B)}{R_t}}{\sum \frac{R_t(B)}{R_t}}$$

where $R_t(B)$ stands for the number of regions first bound by TF B at time point t ; $R_t(A,B)$, for the number of regions first bound by TF B at time point t that were bound by TF A at time point t or before; and R_t represents the total number of regions bound by any TF in time point t. In order to handle the networks, we used the *igraph* package[78].

Network simplification. Aiming to analyze the hierarchical relationship of TFs and simplify the interpretation of the network, we performed two operations over each gene module network: Vertex Sort and transitive reduction (TR)[79]. Briefly, the vertex sort algorithm assigns two parameters for each node in the network: the distance, in edges, between the node and the bottom of the network; and the distance between the node and the top of the network. Combined, those parameters allow for the topological ordering of the network, which consists in listing its nodes such that nodes at the top precede downstream nodes. We then defined the 'top layer' as the set of nodes with lowest distance to the top of the network, i.e., nodes that have no incoming edges or nodes that assemble a strongly connected component (SCC) with all upstream nodes. Analogously, the 'bottom layer' was defined as the set of nodes with lowest distance to the bottom of the network, i.e., nodes with no outgoing edges or that form a SCC with all downstream nodes. The 'core layer' comprises nodes that link top layer and bottom layer. Nodes in the core layer that are exactly one edge from both top and bottom layers constitute the 'single-level core layer', while nodes that link top and bottom through paths composed of more than one edge form the 'multi-level core layer'. The result of this procedure for each gene module can be seen in Figure 4 and supplementary data. The TR, in turn, simplifies the network visualization by generating the network with the smallest number of edges that keeps the reachability of the original network.

Network validation. We validated our approach by comparing the network produced when applying our method to the ChIP-seq data produced by [44] with the network they obtained. Transcription factor ChIP-seq peak files were retrieved from Gene Expression Omnibus (GSE36099, 23 TFs, and 4 time points; note that RUNX1 and ATF4 were discarded from the analysis since one and three time points, respectively, were missing on GEO for those

TFs) and preprocessed as previously described to generated time resolved co-binding matrices, further used as an input for our networking algorithm. We computed the precedence relationships among TFs and generated the TF binding hierarchy networks for visualization. We compared the produced TF hierarchy network with the network shown in Figure S4 and in [44] using two metrics: sensitivity and specificity. Sensitivity is calculated as the ratio of edges described in this study over the edge number sum for both networks. Specificity is defined as the ratio of the number of edges that were described to not exist in the network produced by our software over the number of edges described to not occur in any of both studies.

Proportion of incoming edges based on the classification of the TF source node. Aiming to assess the hierarchy of TFs accordingly to their chromatin dependence and chromatin-opening index, we computed the number of edges connecting the sets of all TFs with a given classification for each gene module. We then divided those values by the number of edges that target TFs with a specific classification. Hence, the proportion of incoming edges based on TF classification is given by:

$$P_{C1\to C2} = \frac{|W_{C1\to C2}|}{\sum|W_{K\to C2}|}$$

where $P_{C1\to C2}$ is the proportion of edges from nodes with classification C1 to nodes with classification C2; $W_{C1\to C2}$ is the set of edges from nodes with classification C1 to nodes with classification C2; K can represent either pioneer, settler or migrant and $|\cdot|$ means the cardinality of a set, *i.e.* the number of elements it contains.

We assessed the classification precedence significance for TF interaction with a hypergeometric test. We consider the sample space as all possible oriented edges in a

network with the same number of nodes for each classification as the hierarchy network for a given transcriptional module. Formally:

$$p_{C1\to C2}(E, E_{C1\to C2}, W, W_{C1\to C2}) = \sum_{x=|W_{C1\to C2}|}^{|W|} \frac{\binom{E_{C1\to C2}}{x}\binom{E-E_{C1\to C2}}{|W|-x}}{\binom{E}{|W|}}$$

Where $E$is the number of edges on the sample space network, i.e., a fully connected network with the same number of nodes as the TF hierarchy network for a given transcriptional module (excluding self-loops),$E_{C1\to C2}$is the number of edges from TFs with classification $C1$to TFs with classification $C2$ in the sample space network, $|W|$is the number of edges on the TF hierarchy network for a given transcriptional module and $|W_{C1\to C2}|$is the number of edges in the same network connecting TFs with classification $C1$to TFs with classification $C2$.

Network visualization. In order to visualize the network, we exported the adjacency matrices in the R environment to CytoScape using the CyREST API[80]. The networks' layout and style were automated with the help of packages RCy3[81] and RJSONIO.

Network mining

With the purpose of identifying key TFs in the transition to the senescent phenotype, we analyzed the TF binding characteristics with their relative location in the chromatin binding hierarchy networks for each gene module. The figures illustrating this analysis were generated with the help of the ggplot2 R package.

TF classification. For each network relative to a transcriptional gene module, the number of TF classified as either pioneer, settler or migrant was calculated for each layer, with the subdivision of the core layer as 'multi-level' and 'single-level' (see "*Network simplification*"). The overrepresentation of TFs with a specific classification in a given layer

was evaluated by using a hypergeometric test. We calculated the *p*-value given by:

$$p(K, N, n, k) = \sum \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

where *K* is the number of TFs with a certain classification in the whole network, *N* is the number of TFs in the network; *n* is the number of TFs that belong to a specific layer and *k* is the number of TFs that belong to the same layer and have the referred classification. The *p*-values were corrected for multiple testing with FDR and a corrected $p = 0.05$ was considered an indicative of enrichment for that specific classification in the corresponding layer.

TF dynamicity. For each network relative to a transcriptional gene module, we compared the distribution of the dynamicity of TFs belonging to a certain layer with the distribution of the dynamicity of TFs belonging to the rest of the network. We used the dynamicity index defined previously for each TF, considering only the regions marked as enhancers belonging only to the gene module relative to the network. For each layer in the network, we applied the Kolmogorov-Smirnov test to compare the TF dynamicity distribution for the chosen layer with the dynamicity distribution relative to the TFs belonging to three other layers in the respective network. To account for multiple hypothesis testing, we also performed an FDR correction, considering values of $p = 0.05$ as an indicative of statistical significance.

TF number of binding regions. We performed the same analysis as described in the previous section ("*TF dynamicity*") for the number of bound regions defined in section "*Total number of bound regions*", instead of the dynamicity index.

TF binding characteristics and transcriptional modules. In order to characterize the binding

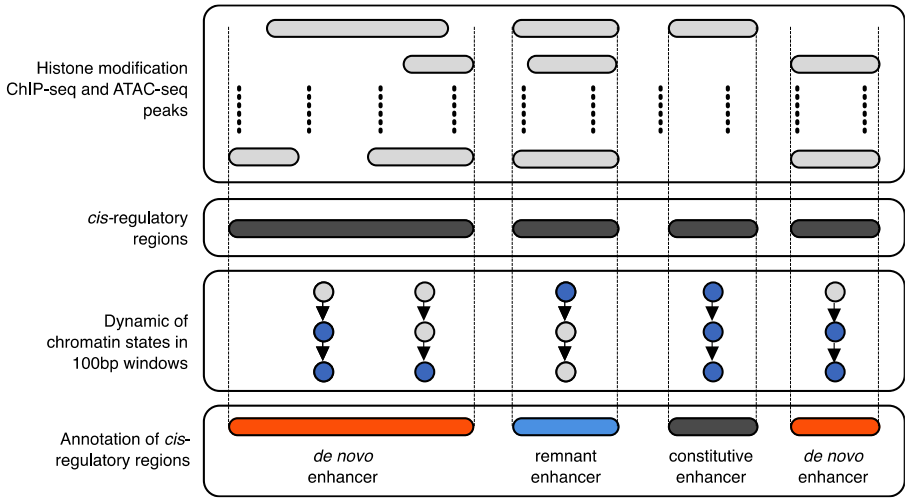
activity of each TF for the different gene modules, we ranked them accordingly to their dynamicity and their number of bound regions. Both parameters for each gene module are shown in Supplementary Figure 4E, which was generated with the *ComplexHeatmap*[82] and *circlize*[83] R packages. We used the mean of the ratio dynamicity - number of bound regions to order the TFs. We assessed the significance of pioneer (respectively, migrant) TF enrichment at the top (respectively, bottom) of the ranked clustered list by employing a set enrichment analysis implemented in the package fgsea.

TF chromatin binding hierarchy networks overlap. To analyze the similarity between the networks for different transcriptional gene modules, we generated a 7-set Euler diagram, where each set contains the edges present in the TF hierarchy network relative to a gene module. Edges in two different networks are considered equal if they link nodes corresponding to the same TFs in their respective networks. We used the package Vennerable to compute the intersections of all possible network combinations and to create the Euler diagram in Supplementary Figures 4F-I. In this figure, the area of each region is proportional to the number of edges shared by the networks corresponding to the sets that contain the referred region and was calculated using the Chow-Ruskey algorithm[84]. A Euler diagram is similar to a Venn diagram, with the difference that the area of a region representing a set is proportional to the number of elements in the set.

Analysis of *de novo* and remnant enhancers

To track combinatorial chromatin state dynamics in space and time, we integrated histone modification ChIP-seq signals at a sub-nucleosomal resolution considering non-overlapping 100bp windows genome-wide using chromstaR (see above), which converts quantitative ChIP-seq data to qualitative chromatin states. For subsequent analysis, since

these 100bp windows can be either isolated or organized in stretches experiencing consistent changes in states, we summarized the information at a higher level, and linked them with the histone modification peaks identified using the more classical ChIP-seq and ATAC-seq peak-calling approach (see flowchart). Briefly, after merging all the peaks identified for all the time-points, for all the histone modification and for the ATAC-seq data sets defining *cis*-regulatory regions, we determined the overlap between "poised enhancers"-, "*de novo* enhancers"-, "remnant enhancer" or "constitutive enhancers"-flagged 100bp windows. When an overlap was found, the entire *cis*-regulatory regions were annotated according to the 100bp window it is overlapping with. This operation rendered a list of annotated *cis*-regulatory regions with *de novo*, constitutive, poised or remnant enhancer elements. We finally considered the center +/- 10kb of these elements.



CRISPR interference (CRISPRi)

hU6-gRNA-hUbc-dCas9-KRAB plasmid was a kind gift from Charles Gerbach (Addgene 71236). gRNA cloning was as published [34]. Briefly, the plasmid was digested with BsmBI and dephosphorylated before ligation with phosphorylated oligo pairs. The gRNA

sequences were listed in the Table 1. The plasmid was then transfected in HEK293T cells, together with packaging plasmids psPAX2 and pMD2.G. 24 hours after fresh medium was added and the medium containing lentivirus was collected and filtered subsequently. Cells were infected for 3 hours. 3 days post infection, cells were passaged and selected with puromycin and used for analyses.

Table 1. gRNA sequences

gRNA	Sequence	
ctrl	caccgGTATTACTGATATTGGTGGG	aaacCCCACCAATATCAGTAATACc
2	caccgAGATGAGGTGTTGCGTGTCT	aaacAGACACGCAACACCTCATCTc
7	caccgTCTGCTCATTGGGGATCGGA	aaacTCCGATCCCCAATGAGCAGAc
14	caccgAAGGCGAAGAAGACTGACTC	aaacGAGTCAGTCTTCTTCGCCTTc
15	caccgCAATGAAATGACTCCCTCTC	aaacGAGAGGGAGTCATTTCATTGc
48	caccgGGAGAACAGTCGCATGAACA	aaacTGTTCATGCGACTGTTCTCCc
54	caccgTTCCAGGGAGTCACCTGTCC	aaacGGACAGGTGACTCCCTGGAAc
61	caccgTTGAAGCAGCACTAGTATCC	aaacGGATACTAGTGCTGCTTCAAc

Immunofluorescence staining and imaging of CRISPR-modified cells

Immunofluorescence staining was performed as previously published[51]. Cells grown in 96-well plates were fixed with 4% PFA and permeabilised with 0.2% Triton-X in PBS. After blocking, the cells were incubated with primary antibody for 1 hour, and then Alexa Fluor secondary antibody for 30 min. Nuclei were counterstained with DAPI. The antibodies were listed in Table 2. The imaging was carried out by IN Cell Analyzer 2000 (GE

Healthcare) with the 20x objective and the quantification was processed using IN Cell Investigator 2.7.3 software.

Table 2. Antibodies for CRISPRi studies

Antibody	Dilution	Cat. No.
IL1α	1:100	R&D MAB200
IL1β	1:100	R&D MAB201

Quantitative RT-qPCR

RNA was extracted with TRIzol (Ambion) and RNAeasy Mini Kit (Qiagen) according to the manufacturer's protocol. Reverse transcription was carried out with SuperScript II RT kit (Invitrogen). Samples were analysed with SYBR Green PCR Master Mix (Applied Biosystems) in CFX96^{TM} Real-Time PCR Detection system (Bio-Rad). Ribosomal protein S14 (*RPS14*) was used as the housekeeping gene. Primers used in CRISPR interference experiments are listed below in Table 3.

Table 3. qPCR primer sequence for CRISPRi studies

Gene	Sequence	
RPS14	CTGCGAGTGCTGTCAGAGG	TCACCGCCCTACACATCAAACT
IL1α	AGTGCTGCTGAAGGAGATGCCTGA	CCCCTGCCAAGCACACCCAGTA
IL1β	GGAGATTCGTAGCTGGATGC	AGCTGATGGCCCTAAACAGA

For RAS-OIS, RAF-OIS and RS gene expression profiling in WI38 fibroblasts, Qiagen Quantitect primers were used using *GAPDH* expression as the housekeeping gene.

For RT-qPCR analyses of colorectal cancer cell lines and lymphomas RNA was transcribed into cDNA using SuperScript II reverse transcriptase (Invitrogen) and oligodT primers. RT-qPCR was performed using Taqman Gene Expression Master Mix and Taqman assays (both Thermo Fisher Scientific) listed in Table S8 on a StepOnePlus cycler (Applied Biosystems). GAPDH was used as a housekeeping gene.

Mouse strains and lymphoma generation

All animal protocols used in this study were approved by the governmental review board (Landesamt Berlin), and conform to the appropriate regulatory standards. We generated Eμ-myc transgenic lymphomas with or without defined genetic defects in the Suv39h1 locus and with or without retroviral Bcl2 overexpression as described previously1-3. Eμ-Myc;Bcl2 lymphomas were further stably transduced with either cJun4A (murine cJun with non-phosphorylatable JNK target sites S63A; S73A; T91A; T93A) cDNA, subcloned into MSCV-IRES-GFP, or empty vector control.

Human colorectal cancer cell lines

Colorectal cancer (CRC) cell lines HCT116 (kindly provided by A. Relogio) and SW480 (DSMZ; ACC-313) were transduced either with cJun4A cDNA, subcloned into MSCV-puro, or empty vector control. HCT116 were cultured in DMEM (gibco), SW480 in RPMI-1640 (gibco), supplemented with 10% FBS (Sigma) and 1% Penicillin/Streptomycin (Biochrom).

Therapy-induced senescence

For the induction of therapy-induced senescence (TIS), Adriamycin (ADR), a topoisomerase II inhibitor widely used in the clinic to treat lymphomas and other malignancies, was added once at a concentration of 0.05 µg/ml for Eμ-Myc;Bcl2 lymphomas for a duration of three days and 0.1 µg/ml for CRC cell lines for a duration of five days. Senescence was assessed two days after drug removal by SABG activity and standard cell cycle analysis using 5-bromo-2′-deoxyuridine/propidium iodide (BrdU/PI)-based flow cytometric measurement as described previously[64, 85].

In vivo lymphoma drug treatment

Individual lymphomas were propagated in up to two strain-matched, non-transgenic, fully immune-competent 6-8-week-old wild-type mice each via tail vein injection of 106 viable cells. Recipient mice were treated with a single intraperitoneal dose of cyclophosphamide (CTX, Sigma, 300 mg/kg body weight) when their lymphadenopathy became well-palpable (i.e. about 8-10 mm in diameter). Treatment responses were monitored by inspection and lymph-node palpation at least twice a week for a maximum of a 100-day observation period, and documented as described previously[86].

Gene expression profiling and data availability for mouse lymphomas

RNA was isolated from lymphoma cells using the RNAeasy Mini kit (Qiagen) and hybridized to Affymetrix Mouse Gene 1.0 ST or Genome 430 2.0 microarrays according to the manufacturer's instructions. Arrays were hybridized, washed and scanned by standard Affymetrix protocols. The mouse model-derived raw microarray data – from our previously published control;Bcl2, Suv39h1-;Bcl2 and Suv39h1-;Bcl2 transduced with 4OHT-inducible Suv39h1 (Suv39h1:ER;Bcl2) lymphomas 2,3 – were deposited at the

Gene Expression Omnibus (GEO) repository of the National Center for Biotechnology Information under accession number GSE134753. Data from our clinical-trial like model were deposited under accession number GSE134751. For assessing long-term outcome after in vivo-treatments, seven or more tumor-bearing animals per arm were used. Survival analysis was done using the survival package in R. Differential gene expression analysis was performed using limma and Empirical Bayes statistics. In order to focus on single genes, probesets without annotations were removed and probesets collapsed to the gene level using the probeset with highest statistical difference between senescent and non-senescent groups by an unpaired t-test prior to the analysis. P-values were corrected for multiple testing using the Benjamini-Hochberg method to control for false discovery rate.

Gene set enrichment analysis (GSEA)

"TIS_up_siJUN_down" gene list, referred to as AP1 senescence gene signature, was generated by intersecting genes down-regulated by siRNA-mediated cJUN depletion in RAS-OIS fibroblasts (see Table S7 and Figure 4F) and genes specific for therapy induced senescence (TIS) in Eμ-myc lymphomas (defined as differentially expressed genes in Adriamycin-treated TIS-competent lymphomas (GSE31099)[86], but not in equally treated TIS-incompetent lymphomas (GSE44355)[86], fold-change >2.0; adjusted p-val < 0.01). The resulting list of 50 genes (Table S9) was used to perform a gene set enrichment analysis (GSEA)[87] for three independent transcriptome settings: the transcriptome of native Eμ-myc lymphomas with known clinical outcome (GSE134751) and two diffuse large B cell lymphoma (DLBCL) patient cohorts (GSE31312 and GSE98588)[88, 89]. For the mouse transcriptome the enrichment for the gene list was compared between therapy-

naïve, initially therapy-sensitive, but destined to fail lymphomas (“Relapse Prone = RP”) and their matched relapses (“Resistant = RES”) after single Cyclophosphamide treatment (300 mg/kg, i.p.). DLBCL patient samples (all profiled at diagnosis) were classified into tumor-free and progressive disease categories based on disease status at last follow-up after standard (R)-CHOP treatment (CD20 antibody Rituximab plus CTX/ADR/Vincristin/Prednisone). GSEA was performed using the R package clusterProfiler. Probe sets were collapsed to the gene level using the correlation-based approach[90] where correlation of probe sets representing the same gene was computed to decide whether to average probe sets ($c > 0.2$) or to use the probe set with highest average expression across samples ($c \leq 0.2$). Probe sets without known annotation were removed. The signal-to-noise ratio $(\mu A - \mu B)/(\sigma A + \sigma B)$ (μ = mean, σ = standard deviation) was used as ranking metric and statistics based on gene set permutations. FDR q values 0.05 were considered significant.

Data availability

All transcriptome data are hosted on Gene Expression Omnibus (BioProject PRJNA439263, accession numbers GSE112084 and GSE143248). ATAC-seq, and ChIP-seq data (histone modification and transcription factor) are hosted on SRA (BioProject PRJNA439280).

Code availability for reproducible science

Interactive maps, circos plots, workflows, scripts and software developed to pre-process raw data, perform statistical analyses as well as data mining and integration are available as .html, and R Markdown files provided in Supplementary data hosted on Zenodo

(https://zenodo.org, DOI: 10.5281/zenodo.1493872). This archive collapses all the material (including processed data) required to reproduce figures presented in the manuscript.

ACKNOWLEDGMENTS

We thank all members, in particular Nir Rozenblum, of the O.B. laboratory for fruitful discussions and suggestions through the course of this work. We would like to thank the Transcriptome and Epigenome facility of Institut Pasteur. We thank Claudia Chica for expert advice on ChIP-seq data processing. We thank Ido Amit and Deborah Winter for valuable discussion and technical support. We thank Benno Schwikowski for key insights and technical advice. We also thank Lars Zender, Eric Gilson, and Hinrich Gronemeyer for valuable intellectual input. R.I.M-Z. was supported by La Ligue Nationale Contre le Cancer and is a Mexican National Scientific and Technology Council (CONACYT) and Mexican National Researchers System (SNI) fellow. Lucas Robinson was supported by the Pasteur - Paris University (PPU) International PhD Program and by the Fondation pour la Recherche Médicale (FRM). J.A.N.L.F.F. was supported by La Ligue Nationale Contre le Cancer. J.G. was supported from Medical Research Council (MRC; MC_U120085810) and a grant from Worlwide Cancer Research (WCR; 18-0215). O.B was supported by the Pasteur-Weizmann Foundation, ANR-BMFT, Fondation ARC pour la recherche sur le Cancer, La Ligue Nationale Contre le Cancer, INSERM-AGEMED. Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number R01CA136533. The content is

solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. O.B. is a CNRS Research Director DR2.

AUTHOR CONTRIBUTIONS

R.I.M-Z, P.-F.R and O.B conceived of the study and conceptual ideas. R.I.M-Z, P.-F.R and O.B. planned, designed experiments, interpreted data and wrote the manuscript. All authors discussed the results and contributed to the final manuscript. R.I.M-Z generated the cell culture systems, performed ChIP-seq, ATAC-seq and RNAi experiments, analyzed data and prepared figures. P-F.R performed computational analyses, designed bioinformatics pipelines and prepared figures. L.R. performed senescence characterization studies and performed ChIP-seq experiments. J.A.N.L.F.F. generated TF networks. G.D. generated Affymetrix microarray data. B.S. and J.G. performed CRISPRi experiments. M.M., D.B., K.S., and C.A.S. performed and analyzed in vitro and in vivo TIS studies and performed GSEA. U.H. supported the study. O.B. supervised, managed and obtained funding for the study.

DECLARATION OF INTERESTS

J.G owns equity and has acted as a consultant for Unity Biotechnology and Geras Bio. Unity Biotechnology funded research on senolytics in J.G. laboratory. J.G. is a named inventor in an MRC patent related to senolytic therapies (PCT/GB2018/051437). All of these links are not directed related to the results presented in this paper.

REFERENCES

1. Martinez-Zamudio, R.I., Robinson, L., Roux, P.F. & Bischof, O. SnapShot: Cellular Senescence in Pathophysiology. *Cell* **170**, 1044-1044 e1041 (2017).
2. Kirkland, J.L. & Tchkonia, T. Cellular Senescence: A Translational Perspective. *EBioMedicine* **21**, 21-28 (2017).
3. Martinez-Zamudio, R.I., Robinson, L., Roux, P.F. & Bischof, O. SnapShot: Cellular Senescence Pathways. *Cell* **170**, 816-816 e811 (2017).
4. Milanovic, M. *et al.* Senescence-associated reprogramming promotes cancer stemness. *Nature* **553**, 96-100 (2018).
5. Coppe, J.P., Desprez, P.Y., Krtolica, A. & Campisi, J. The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu Rev Pathol* **5**, 99-118 (2010).
6. Franceschi, C. & Campisi, J. Chronic inflammation (inflammaging) and its potential contribution to age-associated diseases. *J Gerontol A Biol Sci Med Sci* **69 Suppl 1**, S4-9 (2014).
7. Schosserer, M., Grillari, J. & Breitenbach, M. The Dual Role of Cellular Senescence in Developing Tumors and Their Response to Cancer Therapy. *Front Oncol* **7**, 278 (2017).
8. Benhamed, M., Herbig, U., Ye, T., Dejean, A. & Bischof, O. Senescence is an endogenous trigger for microRNA-directed transcriptional gene silencing in human cells. *Nat Cell Biol* **14**, 266-275 (2012).
9. Shah, P.P. *et al.* Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. *Genes Dev* **27**, 1787-1799 (2013).
10. Puvvula, P.K. *et al.* Long noncoding RNA PANDA and scaffold-attachment-factor SAFA control senescence entry and exit. *Nat Commun* **5**, 5323 (2014).
11. Rai, T.S. *et al.* HIRA orchestrates a dynamic chromatin landscape in senescence and is required for suppression of neoplasia. *Genes Dev* **28**, 2712-2725 (2014).
12. Tasdemir, N. *et al.* BRD4 Connects Enhancer Remodeling to Senescence Immune Surveillance. *Cancer Discov* **6**, 612-629 (2016).
13. Sen, P. *et al.* Histone Acetyltransferase p300 Induces De Novo Super-Enhancers to Drive Cellular Senescence. *Mol Cell* **73**, 684-698 e688 (2019).
14. Heinz, S., Romanoski, C.E., Benner, C. & Glass, C.K. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* **16**, 144-154 (2015).
15. Creyghton, M.P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936 (2010).
16. Ostuni, R. *et al.* Latent enhancers activated by stimulation in differentiated cells. *Cell* **152**, 157-171 (2013).
17. van Oevelen, C. *et al.* C/EBPalpha Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. *Stem Cell Reports* **5**, 232-247 (2015).
18. Huggins, C.J. *et al.* C/EBPgamma suppresses senescence and inflammatory gene expression by heterodimerizing with C/EBPbeta. *Mol Cell Biol* **33**, 3242-3258 (2013).
19. Nardella, C., Clohessy, J.G., Alimonti, A. & Pandolfi, P.P. Pro-senescence therapy for cancer treatment. *Nat Rev Cancer* **11**, 503-511 (2011).
20. Drouin, J. Minireview: pioneer transcription factors in cell fate specification. *Mol Endocrinol* **28**, 989-998 (2014).
21. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555-568 (2015).
22. Sherwood, R.I. *et al.* Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat Biotechnol* **32**, 171-178 (2014).

23. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).
24. Martin, N. *et al.* Physical and functional interaction between PML and TBX2 in the establishment of cellular senescence. *EMBO J* **31**, 95-109 (2012).
25. 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**, 1213-1218 (2013).
26. Loffler-Wirth, H., Kalcher, M. & Binder, H. oposSOM: R-package for high-dimensional portraying of genome-wide expression landscapes on bioconductor. *Bioinformatics* **31**, 3225-3227 (2015).
27. Martinez, O. & Reyes-Valdes, M.H. Defining diversity, specialization, and gene specificity in transcriptomes through information theory. *Proc Natl Acad Sci U S A* **105**, 9709-9714 (2008).
28. Patel, P.L., Suram, A., Mirani, N., Bischof, O. & Herbig, U. Derepression of hTERT gene expression promotes escape from oncogene-induced cellular senescence. *Proc Natl Acad Sci U S A* **113**, E5024-5033 (2016).
29. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, Article17 (2005).
30. Yosef, N. & Regev, A. Impulse control: temporal dynamics in gene transcription. *Cell* **144**, 886-896 (2011).
31. Deng, T. & Karin, M. c-Fos transcriptional activity stimulated by H-Ras-activated protein kinase distinct from JNK and ERK. *Nature* **371**, 171-175 (1994).
32. Thakore, P.I. *et al.* Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* **12**, 1143-1149 (2015).
33. Klann, T.S. *et al.* CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol* **35**, 561-568 (2017).
34. Gilbert, L.A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647-661 (2014).
35. Kaikkonen, M.U. *et al.* Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol Cell* **51**, 310-325 (2013).
36. Wang, A. *et al.* Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* **16**, 386-399 (2015).
37. Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M. & Furlong, E.E. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**, 65-70 (2009).
38. Wilkinson, A.C., Nakauchi, H. & Gottgens, B. Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity. *Cell Syst* **5**, 319-331 (2017).
39. Neph, S. *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83-90 (2012).
40. Guo, Y. & Gifford, D.K. Modular combinatorial binding among human trans-acting factors reveals direct and indirect factor binding. *BMC Genomics* **18**, 45 (2017).
41. Ren, X. & Kerppola, T.K. REST interacts with Cbx proteins and regulates polycomb repressive complex 1 occupancy at RE1 elements. *Mol Cell Biol* **31**, 2100-2110 (2011).
42. Li, T. *et al.* CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. *Mol Cell Biol* **28**, 6473-6482 (2008).
43. Weinmann, A.S., Bartley, S.M., Zhang, T., Zhang, M.Q. & Farnham, P.J. Use of chromatin immunoprecipitation to clone novel E2F target promoters. *Mol Cell Biol* **21**, 6820-6832 (2001).
44. Garber, M. *et al.* A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Mol Cell* **47**, 810-822 (2012).

45. Novershtern, N. *et al.* Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296-309 (2011).
46. Zaret, K.S. & Carroll, J.S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* **25**, 2227-2241 (2011).
47. Jozwik, K.M. & Carroll, J.S. Pioneer factors in hormone-dependent cancers. *Nat Rev Cancer* **12**, 381-385 (2012).
48. Wisdom, R., Johnson, R.S. & Moore, C. c-Jun regulates cell cycle progression and apoptosis by distinct mechanisms. *EMBO J* **18**, 188-197 (1999).
49. Weitzman, J.B., Fiette, L., Matsuo, K. & Yaniv, M. JunD protects cells from p53-dependent senescence and apoptosis. *Mol Cell* **6**, 1109-1119 (2000).
50. Hoare, M. *et al.* NOTCH1 mediates a switch between two distinct secretomes during senescence. *Nat Cell Biol* **18**, 979-992 (2016).
51. Georgilis, A. *et al.* PTBP1-Mediated Alternative Splicing Regulates the Inflammatory Secretome and the Pro-tumorigenic Effects of Senescent Cells. *Cancer Cell* **34**, 85-102 e109 (2018).
52. Angel, P., Hattori, K., Smeal, T. & Karin, M. The jun proto-oncogene is positively autoregulated by its product, Jun/AP-1. *Cell* **55**, 875-885 (1988).
53. Schmitt, C.A. *et al.* A senescence program controlled by p53 and p16INK4a contributes to the outcome of cancer therapy. *Cell* **109**, 335-346 (2002).
54. Duronio, R.J. & Xiong, Y. Signaling pathways that control cell proliferation. *Cold Spring Harb Perspect Biol* **5**, a008904 (2013).
55. Biddie, S.C. *et al.* Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol Cell* **43**, 145-155 (2011).
56. Vierbuchen, T. *et al.* AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. *Mol Cell* **68**, 1067-1082 e1012 (2017).
57. Krtolica, A., Parrinello, S., Lockett, S., Desprez, P.Y. & Campisi, J. Senescent fibroblasts promote epithelial cell growth and tumorigenesis: a link between cancer and aging. *Proc Natl Acad Sci U S A* **98**, 12072-12077 (2001).
58. van Deursen, J.M. The role of senescent cells in ageing. *Nature* **509**, 439-446 (2014).
59. Eggert, T. *et al.* Distinct Functions of Senescence-Associated Immune Responses in Liver Tumor Surveillance and Tumor Progression. *Cancer Cell* **30**, 533-547 (2016).
60. Tsankov, A.M. *et al.* Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**, 344-349 (2015).
61. Goode, D.K. *et al.* Dynamic Gene Regulatory Networks Drive Hematopoietic Specification and Differentiation. *Dev Cell* **36**, 572-587 (2016).
62. Xu, M. *et al.* Senolytics improve physical function and increase lifespan in old age. *Nat Med* **24**, 1246-1256 (2018).
63. Overman, J. *et al.* Pharmacological targeting of the transcription factor SOX18 delays breast cancer in mice. *Elife* **6** (2017).
64. Itahana, K., Campisi, J. & Dimri, G.P. Methods to detect biomarkers of cellular senescence: the senescence-associated beta-galactosidase assay. *Methods Mol Biol* **371**, 21-31 (2007).
65. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
66. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
67. Landt, S.G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res* **22**, 1813-1831 (2012).
68. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

69. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011* **17**, 3 (2011).
70. Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. GC-content normalization for RNA-Seq data. *BMC Bioinformatics* **12**, 480 (2011).
71. Risso, D., Ngai, J., Speed, T.P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* **32**, 896-902 (2014).
72. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).
73. Taudt, A., Nguyen, M. A., Heinig, M., Johannes, F. & Colome-Tatche, M. chromstaR: Tracking combinatorial chromatin state dynamics in space and time. *bioRxiv* (2016).
74. Lê, S., Josse, J. & Husson, F. FactoMineR: An R Package for Multivariate Analysis. *2008* **25**, 18 (2008).
75. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
76. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645 (2009).
77. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).
78. Gábor. C., T.N. (2006).
79. Aho, A.V., Garey, M.R. & Ullman, J.D. The Transitive Reduction of a Directed Graph. *SIAM Journal on Computing* **1**, 131-137 (1972).
80. Ono, K., Muetze, T., Kolishovski, G., Shannon, P. & Demchak, B. CyREST: Turbocharging Cytoscape Access for External Tools via a RESTful API. *F1000Res* **4**, 478 (2015).
81. Shannon, P.T., Grimes, M., Kutlu, B., Bot, J.J. & Galas, D.J. RCytoscape: tools for exploratory network analysis. *BMC Bioinformatics* **14**, 217 (2013).
82. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).
83. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811-2812 (2014).
84. Chow, S. & Ruskey, F. 466-477 (Springer Berlin Heidelberg, Berlin, Heidelberg; 2004).
85. Reimann, M. *et al.* Tumor stroma-derived TGF-beta limits myc-driven lymphomagenesis via Suv39h1-dependent senescence. *Cancer Cell* **17**, 262-272 (2010).
86. Dorr, J.R. *et al.* Synthetic lethal metabolic targeting of cellular senescence in cancer therapy. *Nature* **501**, 421-425 (2013).
87. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
88. Chapuy, B. *et al.* Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med* **24**, 679-690 (2018).
89. Visco, C. *et al.* Comprehensive gene expression profiling and immunohistochemical studies support application of immunophenotypic algorithm for molecular subtype classification in diffuse large B-cell lymphoma: a report from the International DLBCL Rituximab-CHOP Consortium Program Study. *Leukemia* **26**, 2103-2113 (2012).
90. Monti, S. *et al.* Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* **105**, 1851-1861 (2005).

FIGURE 1:

Multi-state establishment of the senescence transcriptional program

(a) Schematic overview for defining the gene-regulatory code of RAS-OIS in WI38 fibroblasts using time-resolved, high-throughput transcriptome (microarray) and epigenome (ChIP-seq, and ATAC-seq) data sets. Quiescence and all additional models of senescence followed the same scheme and were performed in biological replicates (n=2) at the indicated times.

(b) Self-organizing maps (SOMs) of gene expression profiles in WI38 fibroblasts for quiescence and RAS-OIS time-series experiments as logarithmic fold-change. Red spots mark overexpression, blue spots underexpression.

(c) Multidimensional scaling (MDS) analysis scatter plot visualizing the level of similarity/dissimilarity between normalized quiescence and RAS-OIS time-series transcriptomes in WI38 fibroblasts. Distances between samples represent leading logarithmic fold-changes defined as the root-mean-squared average of the logarithmic fold-changes for the genes best distinguishing each pair of samples.

(d) Scatter plot depicting the evolution of transcriptome diversity (H_j) *vs.* transcriptome specialization (σ_j) in cells undergoing quiescence or RAS-OIS in WI38 fibroblasts. For each time-point and treatment, the average H_j and σ_j values across biological replicates are given. T_0 is start of time-course.

(e) Heatmap showing seven modules (I-VII) of temporally co-expressed genes specific for RAS-OIS in WI38 fibroblasts defined using an unsupervised WGCNA clustering approach. Data are expressed as row Z-scores. Representative genes are depicted for each module.
(f) Functional over-representation map depicting Molecular Signatures Database (MSigDB) hallmark gene sets associated to each transcriptomic cluster. Dots are color-coded according to the FDR corrected *p*-value based on the hypergeometric distribution. Size is proportional to the percentage of genes in the gene set belonging to the cluster.

FIGURE 2:

A dynamic enhancer program shapes the senescence transcriptome

(a) Arc plot visualizing dynamic chromatin state transitions for the indicated intervals during RAS-OIS in WI38 fibroblasts. Edge width is proportional to the number of transitions.

(b) Histogram showing the total number of windows of the top 15 chromatin states transitions during RAS-OIS in WI38 fibroblasts. Chromatin state transitions corresponding to *de novo* enhancer activation are highlighted as white bars.

(c) Chromatin dependence (CD) versus chromatin opening index (COI) are plotted for high-confidence TF sequence motifs used in our study (see Materials and Methods for details) during RAS-OIS in WI38 fibroblasts. Pioneer, settler and migrant TFs as defined by their COI and CD property are color-coded and select members of each TF class are listed. Same color code is used in all figures.

(d) Biplot for principal component analysis performed with select TF binding parameters during RAS-OIS in WI38 fibroblasts: dynamicity, total number of bound windows (N), percentage of binding at enhancers, pioneer index (referred to as the number of bound

windows pre-stimulation), chromatin opening index (COI) and chromatin dependence (CD). The plot depicts the projections of the TFs and the loading of the different covariates for the first two principal components which explain 76.9% of the total inertia. The ellipses delineate the 95% confidence intervals for AP-1 pioneers (blue with black outline), non-AP-1 pioneers (blue), settlers (red), and migrants (green).

FIGURE 3:

AP-1 pioneer TF bookmarking of senescence enhancer landscape foreshadows the senescence transcriptional program

(a) Distribution of fold-change in normalized enhancer marks H3K27ac and H3K4me1 ChIP-seq signals over input in the “unmarked”-, “constitutive”-, “poised”-, “*de novo”*-, and “remnant enhancers”-flagged genomic bins at indicated time-points during RAS-OIS in WI38 fibroblasts (see Material and Methods for details). The cartoon at the top illustrates the temporal rules used to flag genomic bins. Bottom specifies the genomic coverage in mega bases (Mb) for each category and the corresponding number of enhancers.

(b) WI38-ER:RASV12 fibroblasts were super-infected with dCas9-KRAB and individual guides (g14, 15, g61 and g7) targeting two *de novo* enhancers. Cells were pharmacologically selected and induced into RAS-OIS by 4-OHT. 8 days after RAS-OIS induction cells were stained by indirect immunofluorescence for IL1ß or analyzed by RT-qPCR for the *IL1α* or *IL1ß* expression. WI38-ER:RASV12 treated with 4-OHT or DMSO served as positive and negative controls. Data represent mean ± SD (n=3). *p<0.05, ***p<0.001. Comparison with ctrl 4-OHT, one-way ANOVA (Dunnett’s test). Scale bar, 100 µm.

(c) Rank plot depicting the summed occurrences for TF binding in *de novo* enhancers before RAS-OIS induction (left) and remnant enhancers after RAS-OIS induction (144 h) (right) in WI38 fibroblasts. Top ten TFs are indicated.

(d) Distribution of total number (N) of TFs bound per enhancer for constitutive enhancers (dark grey), TF pre-marked *de novo* enhancers (orange) and TF virgin *de novo* enhancers (red).

(e) Average absolute expression level (log_2 scale) kinetics for genes associated with: poised (blue), TF pre-marked *de* novo (orange), and TF virgin *de novo* enhancers (red). Dots depict the average absolute expression level, and bars depict the standard error of the mean. Inset histogram illustrates the average leading log_2 fold-change in expression (+/- standard error of the mean) for genes associated with constitutive (black), poised (light blue), TF pre-marked *de novo* (orange) and TF virgin *de novo* (red), and remnant enhancers (dark blue).

FIGURE 4:

A hierarchical TF network defines the senescence transcriptional program

(a-c) Genome-wide transcription factor co-binding occurrence matrix summed across all time-points in WI-38 fibroblasts undergoing RAS-OIS (a), RAF-OIS (b) and RS (c) (left, shades from blue to yellow, in log_{10} scale). Overlap significance was calculated by a hypergeometric test (right, shades from blue to red, in $-log_{10}$ scale). The co-binding occurrence matrix was clustered using Ward's aggregation criterion and the corresponding corrected *q*-values were projected on this clustering. The graphs on the left and bottom show the density in pioneer, migrant and settler TFs along each axis of the matrix. AP-1 members are indicated.

(d) Heatmap describing the association between individual TFs (row) and TF lexicons (columns). Four boxed out insets provide detailed information on TF composition of lexicons. A comprehensive, high-resolution and interactive heatmap is shown in Supplementary Data (see under Code availability in Material and Methods). The right bar plot shows the total number of binding sites for each TF. The top bar plot shows the total number of regions for each regulatory module. The bottom bar plot shows the average proportion of AP-1 binding sites inside each regulatory module.

(e) Graphical representation of the hierarchical TF network for transcriptomic module VI. Nodes (circles) represent TFs and an oriented edge (line) connecting TFs A and B means that at least 30 % of the regions bound by B were also bound by A at the same time point or before. In order to simplify the visualization, we represent strongly connected components (SCCs) as a single node and performed a transitive reduction (TR). Node color is based on the average dynamicity of the SCC members. Node border color indicates their classification as pioneer (blue), settler (red) or migrant (green). Node border thickness encodes the percentage of bound regions before RAS stimulation. Edge color was calculated accordingly to the relative coverage of the outgoing TF over the incoming TF. The network has three layers: top, core and bottom. Nodes in the top have no incoming edges and nodes in the bottom have no outgoing edges. The core layer comprises TFs that have both incoming and outgoing edges. Interactive Cytoscape graphs are accessible as Supplementary data (see under Code availability in Material and Methods).

(f) Venn diagram showing the specificities and overlaps in differentially expressed direct target genes upon siRNA-mediated AP-1-c*JUN, ETS1,* and *RELA* depletion in RAS-OIS WI38 fibroblasts at 144h (fully senescent cells). Genes are considered as direct targets of

a given TF when PIQ predicts that the TF bound to an enhancer associates to this gene (see Materials and Methods for details). Promoters were excluded from the analysis.

(g) Asymmetric biplot for correspondence analysis between transcriptomic clusters and the number of up-, down-, up-or-down- or nonregulated (stable) genes upon siRNA-mediated AP-1-c*JUN*, *ETS1* or *RELA* depletion. The *p*-value reflects the strength of the association as assessed with a χ^2 test.

FIGURE 5:

Hierarchy Matters: Functional Perturbation of AP-1 pioneer TF, but no other TF, reverts the senescence clock

(a) Principal component analysis (PCA) on transcriptomes obtained from siRNA-mediated depletion of *AP-1-cJUN*, *ETS1* or *RELA* at indicated timepoints of the RAS-OIS time-course. Horizontal and vertical bars show minimal and maximal coordinates for each siRNA and time-point on principal components one (PC1, horizontal axis) and two (PC2, vertical axis).

(b) Functional overrepresentation map showing Molecular Signature Database (MSigDB) hallmark pathways associated to “All”, “Direct Target” and “Indirect Target” genes differentially expressed after siRNA-mediated *AP-1-cJUN*, *ETS1* or *RELA* depletion. Genes are considered as direct targets when a PIQ prediction for the given TF is falling inside an enhancer associated to this specific gene. Promoters are excluded from the analysis. The size of dots is proportional to the $-\log_{10}$ *q*-value based on the hypergeometric distribution obtained when testing for over-representation, and their color denote whether the term is enriched for an up or down-regulated gene list.

(c) Heatmap comparing gene expression profiles of siRNA-Control-treated (siCTRL) cells at indicated time-points of OIS and siRNA-cJUN treated senescent RAS-OIS WI38 fibroblasts at 144h. Data are expressed as row Z-scores.

FIGURE 6:

Functional role of AP1 in therapy-induced senescence

(a-b) SABG staining of HCT116 and SW480 colorectal cancer cells overexpressing cJUN4A or an empty vector as control and treated with Adriamycin (100 ng/ml) to undergo TIS or left untreated. Inset, mean percentage of SABG positive cells ± SDEV. A representative image of 3 independent experiments is shown. Scale bars, 100 µm.

(c-d) Heatmap of transcript levels for select AP-1 target genes (see Table S8) as determined by RT-qPCR in HCT116 (c), SW480 (d) cells under the conditions indicated in Figures 6a,b. One representative heatmap from 2 independent experiments is shown. Data are expressed as row Z-scores.

(e) SABG staining in primary lymphomas from Eμ-*myc* transgenic mice overexpressing cJUN4A or an empty vector as control and treated with Adriamycin (50 ng/ml, 5 days) to undergo TIS or left untreated. Inset, mean percentage of SABG positive cells ± SDEV. A representative image of 3 independent lymphomas is shown. Scale bars, 50µm.

(f) Heatmap of transcript levels for select AP-1 target genes (see Table S8) as determined by RT-qPCR under the conditions indicated in Figure 6e. Data are expressed as row Z-scores.

(g-i) GSEA analysis showing normalized enrichment score (NES) plots and FDRs for the AP-1 senescence gene signature enrichment in the transcriptomes (GSE134751) of

therapy-naïve, initially therapy-sensitive, but destined to fail lymphomas (Relapse Prone = RP) and their matched relapses (Resistant = RES) after single Cyclophosphamide (CTX) treatment (300 mg/kg, i.p.) (g), and transcriptomes (GSE31312 and GSE98588) of DLBCL patients (all profiled at diagnosis) classified into tumor-free and progressive disease categories (*i.e.,* tumor) based on disease status at last follow-up after standard (R)-CHOP treatment (CD20 antibody Rituximab plus CTX/ADR/Vincristin/Prednisone) (h-i).

Figure S1:

Multi-state establishment of the senescence transcriptional program

(a-b) Representative DAPI, EdU, SABG indirect fluorescence and phase contrast (from left to right) microscopy images of WI38 fibroblasts undergoing RAS-OIS or quiescence at indicated time-points. Insets, mean percentage of SABG positive cells ± SDEV and proliferative capacity expressed as percent EdU-positive staining cells ± SDEV. Scale bar, 100µm.

(c-q) Growth curves, EdU incorporation and representative RT-qPCR expression profiles of selected target genes for RAS-OIS (c-e), RAF-OIS (f-h), RS (i-k), RAS-OIS in GM21 skin fibroblasts (l-n) and quiescence (o-q) at indicated time-points. For l-m, average +/- standard error of the mean of three independent infections is shown.

(r-t) Heatmaps for temporally co-expressed differentially regulated genes and associated functional over-representation of MSigDB hallmark gene sets for each module described for RAF-OIS (r) and replicative senescence (RS) (s) in WI-38 fibroblasts, and RAS-OIS in GM21 skin fibroblasts (t) at the indicated time points in hours (h).

(u) Boxplots depicting the gene expression profiles for each of the RAS-OIS transcriptomic modules in WI38 fibroblasts. Note the sharp transitions of modules I and V. Data are expressed as row Z-score.

Figure S2:

A dynamic enhancer program shapes the senescence transcriptome

(a) Histogram showing the percentage of genome covered by each chromatin state at indicated time. Bottom table assigns histone modification combinations (grey: presence, white: absence) to biologically meaningful mnemonics. Venn diagrams highlight the specificities and overlaps in chromatin states associated with active (left) and poised enhancers (right) at indicated time-points.

(b-c) Arc plots visualizing dynamic chromatin state transitions for RAF-OIS and replicative senescence at the indicated intervals. Edge width is proportional to the number of transitions.

(d) Boxplots showing the distribution of relative gene expression (row Z-score) through time for genes associated to regions undergoing different chromatin state changes. The pictogram at the top of each graph describes the class of chromatin state change considered.

(e) Asymmetric biplot for correspondence analysis between changes in chromatin states and gene expression modules. The *p*-value reflects the strength of the association a assessed using a χ^2 test. Only the top 20 contributing and best projected (squared cosine > 0.5) chromatin state changes are shown.

(f-k) Most enriched sequence motifs in (f) active enhancers for RAS-OIS in WI38 fibroblasts and (g-k) ATAC-seq peaks at each time point for (g) RAS-OIS (WI38), (h) RAF-

OIS (WI38, (i) replicative senescence (RS)(WI38), (j) RAS-OIS (GM21 skin fibroblasts) and (k) quiescence (WI38) time courses. Motif logos are shown on left of the histogram. Black, dotted boxes highlight the core motif for AP-1 transcription factor family members. Note that the transcriptional repressor BACH shares this motif.

(l-n) ATAC-seq (grey lines for forward, black lines for reverse reads) and nucleosome (red line) density for (l) AP-1 FOSL1 (pioneer), (m) RELA (settler), and (n) SREBF1 (migrant).

(o) Comparison between PIQ predictions and RELA (left), AP-1-JUN (middle) and AP-1-FOSL2 (right) ChIP-seq. The two density heatmaps at the center of each panel illustrate ChIP-seq (left) and ATAC-seq (right) signals computed in 10bp non-overlapping windows at selected bound- (25%) and unbound- (75%) predicted PWM hits ± 1kb ranked according to the ChIP-seq signal in the most central 100bp. The stack histogram on the left shows the distribution of bound (red) and unbound (green) PWM hits as defined by PIQ along the ranking. The curves on the right depict the evolution of the enrichment score (ES) along the ranking as defined with a set enrichment analyses (SEA) comparing the ChIP-seq signal and the bound (red) and unbound (green) status of the PWM hit. For each SEA, we performed 1 000 permutations and provide the associated Benjamini–Hochberg adjusted *p*-value and ES score.

Figure S3:

AP-1 pioneer TF bookmarking of senescence enhancer landscape foreshadows the senescence transcriptional program

(a) Density heatmaps of normalized H3K27ac and H3K4me1 ChIP-seq signals computed in 10bp non-overlapping windows at enhancers +/- 10kb grouped by enhancer status (constitutive, *de novo* or remnant) at indicated time-points after RAS-OIS induction.

(b-c) Representative genome browser screenshots of normalized H3K4me1 (pink), H3K27ac (orange), H3K4me3 (blue) and H3K27me3 (green) ChIP-seq and ATAC-seq (light grey) profiles and chromatin states at **(b)** *IL1ß* and **(c)** *CDC6* gene loci. Red boxes single-out **(b)** IL1ß *de novo* and **(c)** CDC6 remnant enhancers.

(d) Boxplots depicting the distribution of relative gene expression (row Z-score) through time for genes associated with constitutive (left), *de novo* (middle) and remnant (left) enhancer windows.

(e) RAS-OIS cells at day 14 infected with dCas9-KRAB and individual guides (g14, g15, g61, and g7) and analyzed by RT-qPCR for the expression of *IL1α* or *IL1β* as described in Figure 3B. Data represent mean ± SD (n=3). *p<0.05, ***p<0.001. Comparison with ctrl 4OHT, one-way ANOVA (Dunnett's test).

(f) RAS-OIS cells were infected with dCas9-KRAB and individual guides (g2, g48 and g54) for non-enhancer regions (outside *de novo* enhancers) as described in Figure 3B. 8 or 14 days after infection, cells were stained for IL1α or IL1β by indirect immunofluorescence and percentage positive cells were quantified (n=3 for 8 days and n=2 for 14 days). Data represent mean ± SD. *p<0.05, **p<0.01, ***p<0.001. Comparison with ctrl 4OHT, one-way ANOVA (Dunnett's test).

(g) Rank plot depicting the summed occurrences for TFs binding in proliferating cells (T_0) in *de novo* enhancers (left) and after replicative senescence in remnant enhancers (right). Top ten TFs are highlighted.

(h) Metaprofiles showing the density in "active enhancer"-flagged genomic bins (top) and "constitutive enhancer"-flagged genomic bins (bottom) in the vicinity (+/- 50kb) of TF bookmarked *de novo* (left) and TF virgin *de novo* enhancers (right). The density in "active enhancer"-flagged genomic bins is provided for the indicated time points.

(i) Boxplot showing the correlation between absolute leading $\log_2$ expression fold-change and the number of genomic bins flagged as "*de novo"* enhancers per enhancer. ***: *p*-value < 10^{-3}, Student's *t*-test considering regions with 0 "*de novo"* enhancers bins as a control.

Figure S4:

A hierarchical TF network defines the senescence transcriptional program

(a) Representative Circos plots summarizing pairwise transcription factor co-binding at enhancers for down-regulated transcriptomic module (IV, top) and up-regulated transcriptomic module (VI, bottom) at indicated time-points. Co-interactions involving AP-1 are shown in black. Selected examples of gained (green) and lost (orange) interactions are highlighted. Pioneer TFs blue, settler TFs red, migrant TFs green. See also dynamic Circos plot movies in Supplementary Data (see under Code availability in Material and Methods).

(b) Heatmap showing the overlap between TF lexicons (rows) and chromatin states, ChIP-seq and ATAC-seq peaks (columns). The dendrograms were computed by applying hierarchical clustering on the fraction matrix with Pearson's correlation and average linkage.

(c) Representative genome browser screenshots for lexicons 22 and 50 as described in Figure 4E. Chromatin states are color-coded as in Figure 2. Transcription factor binding instances constituting each lexicon are highlighted in the inset.

(d) Ratio of incoming edges based on the classification of the TF source node. The relative and absolute number of edges corresponding to all seven modules are displayed inside the nodes, which are colored accordingly to TF classification as in previous panels. The

thickness of links is proportional to the relative number of TF hierarchy edges connecting nodes with the corresponding classification.

(e) Number of bound regions and dynamicity index for each TF (rows) across all gene modules (columns). The left heatmap depicts the dynamicity index scaled by column. The middle heatmap depicts the square root of number of bound regions scaled by column. The right single-column heatmap illustrates TF classification.

(f) Chow-Ruskey diagram showing specificities and overlaps of TF interactions in each gene module. Each set corresponds to the TF-TF network edges identified for a given transcriptomic module. The global area of each set is proportional to the number of edges in its respective transcriptomic module and was calculated with the Chow- Ruskey algorithm.

(g-i) Chow-Ruskey diagrams for edges **(g)** originating only from TFs at the top of hierarchy, **(h)** connecting only TFs at the core layer or **(i)** reaching only TFs at the bottom. Note that edges at the top of the hierarchy are shared among the gene modules while edges towards the bottom of the hierarchy are module-specific.

(j-l) Knockdown efficiency for siRNAs against cJUN (j), ETS1 (k) and RELA (l) was assessed by RT-qPCR for each time-point assessed relative to non-targeting siRNA (siC).

Figure S5:

Hierarchy Matters: Functional Perturbation of AP-1 pioneer TF, but no other TF, reverts the senescence clock

(a-c) Volcano plots depicting the $-\log_{10}$ *p*-value as a function of the $\log_2$ fold-change in gene expression defined by a differential analysis conducted with *limma* to highlight the effect of siRNA-mediated (a) AP-1-c*JUN*, (b) *ETS*1 and (c) *RELA* depletion in senescent

RAS-OIS cells at day 6 (144h). Blue dots in respective plots indicate probes corresponding to AP-1-c*JUN*, *ETS*1 and *RELA*. Black outlined dots highlight direct targets of AP-1-c*JUN*, *ETS*1 and *RELA*.

(d) Upset plot depicting specificities and overlaps in differentially expressed genes between siRNA-Control and siRNA-JUN silenced OIS fibroblasts at indicated time-points. The yellow dots highlight gene sets specific to a single comparison set, while green dots highlight gene sets find in two different pair-wise comparison.

(e-h) Venn diagrams (top) and heatmaps (bottom) depicting the overlap between genes belonging to (e) E2F-, (f) NFκB-, (g) p53-target and (h) N1ICD-induced senescence (NIS) gene signatures in RAS-OIS fibroblasts treated with si-JUN. Venn diagrams show the overlap of up-regulated genes after siRNA-mediated AP-1-c*JUN* knock-down for upregulated E2F- (*i.e.* pro-proliferation genes), p53 targets, NIS- (*i.e.* early SASP genes), and downregulated NFκB target genes (*i.e.* late SASP genes) RAS-OIS cells at day 6 (144h). Bottom heatmaps show the comparison of gene expression profiles of siRNA-Control (siCtrl) and siRNA-cJUN treated cells undergoing RAS-OIS at indicated time-points. Data are expressed as row Z-score. E2F targets and NFκB targets were defined according to Molecular Signature Database (MSigDB).

(i-j) Network representation of the interaction between AP1 TFs and p53 family TFs at enhancers of genes in gene modules II (i) and VI (j) as described in Figure 1E. P53 is highlighted by arrows.

Figure S6

Functional role of AP1 in therapy-induced senescence

(a-b) Cell cycle analysis using BrdU incorporation and propidium iodide staining flow cytometry for HCT116 (a) and SW480 (b) CRC cell lines under experimental conditions described in Figure 6a,b. Insets, percentage of cells in respective cell cycle phase.

(c-d) cJUN gene expression determined by RT-qPCR in HCT116 (c) and SW480 (d) CRC cell lines underexperimental conditions described in 6a,b. Mean of 3 independent experiments.

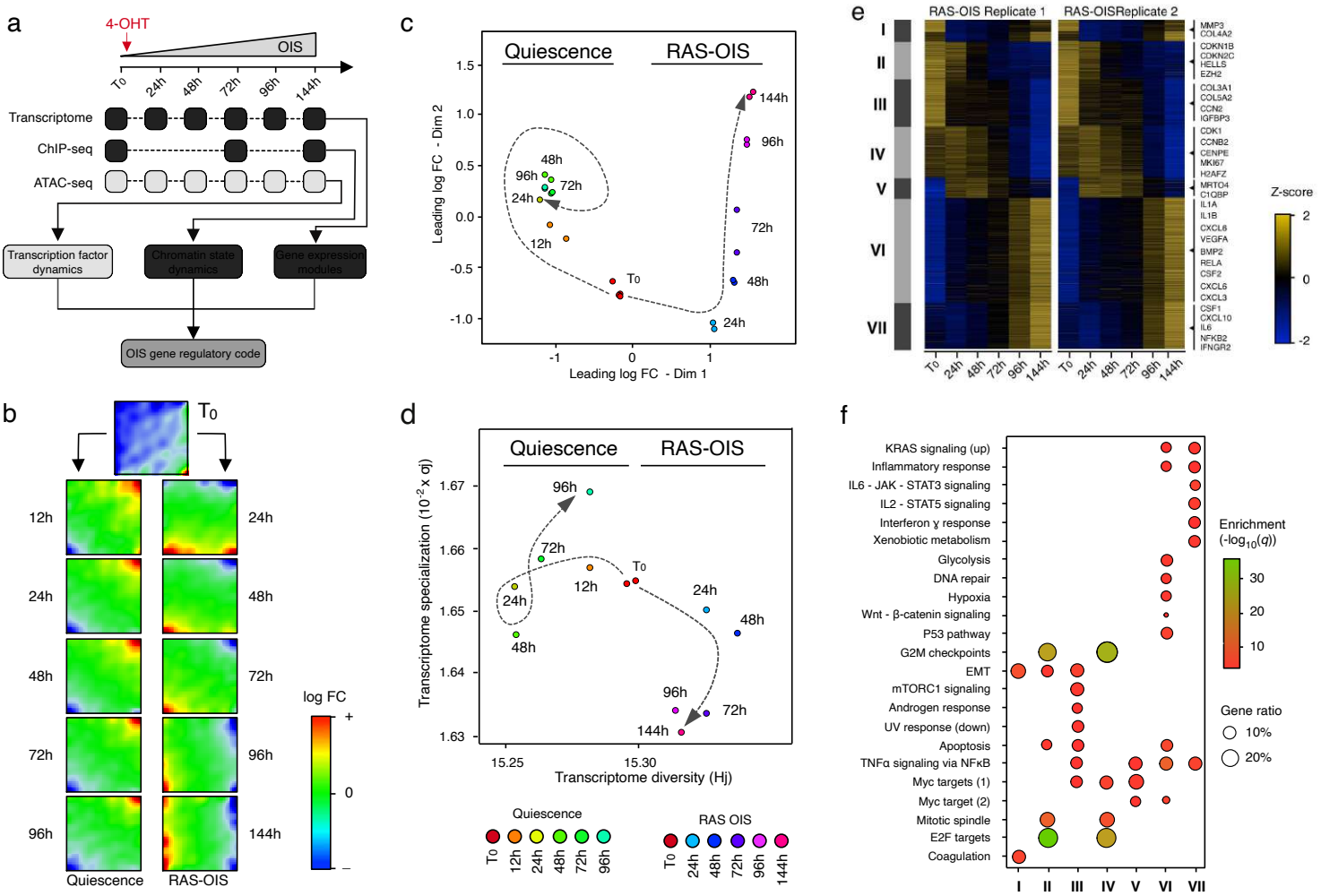

a
4-OHT
OIS
T0
24h
48h
72h
96h
144h
Transcriptome
ChIP-seq
ATAC-seq
Transcription factor dynamics
Chromatin state dynamics
Gene expression modules
OIS gene regulatory code
b
T0
12h
24h
48h
72h
96h
144h
log FC
+
0
–
Quiescence
RAS-OIS
c
Quiescence
RAS-OIS
Leading log FC - Dim 2
Leading log FC - Dim 1
d
Quiescence
RAS-OIS
Transcriptome specialization (10-2 x σj)
Transcriptome diversity (Hj)
Quiescence
RAS OIS
e
RAS-OIS Replicate 1
RAS-OISReplicate 2
I
II
III
IV
V
VI
VII
MMP3
COL4A2
CDKN1B
CDKN2C
HELLS
EZH2
COL3A1
COL5A2
CCN2
IGFBP3
CDK1
CCNB2
CENPE
MKI67
H2AFZ
MRTO4
C1QBP
IL1A
IL1B
CXCL6
VEGFA
BMP2
RELA
CSF2
CXCL6
CXCL3
CSF1
CXCL10
IL6
NFKB2
IFNGR2
Z-score
2
0
-2
f
KRAS signaling (up)
Inflammatory response
IL6 - JAK - STAT3 signaling
IL2 - STAT5 signaling
Interferon γ response
Xenobiotic metabolism
Glycolysis
DNA repair
Hypoxia
Wnt - β-catenin signaling
P53 pathway
G2M checkpoints
EMT
mTORC1 signaling
Androgen response
UV response (down)
Apoptosis
TNFα signaling via NFκB
Myc targets (1)
Myc target (2)
Mitotic spindle
E2F targets
Coagulation
Enrichment (-log10(q))
30
20
10
Gene ratio
10%
20%

Figure 1

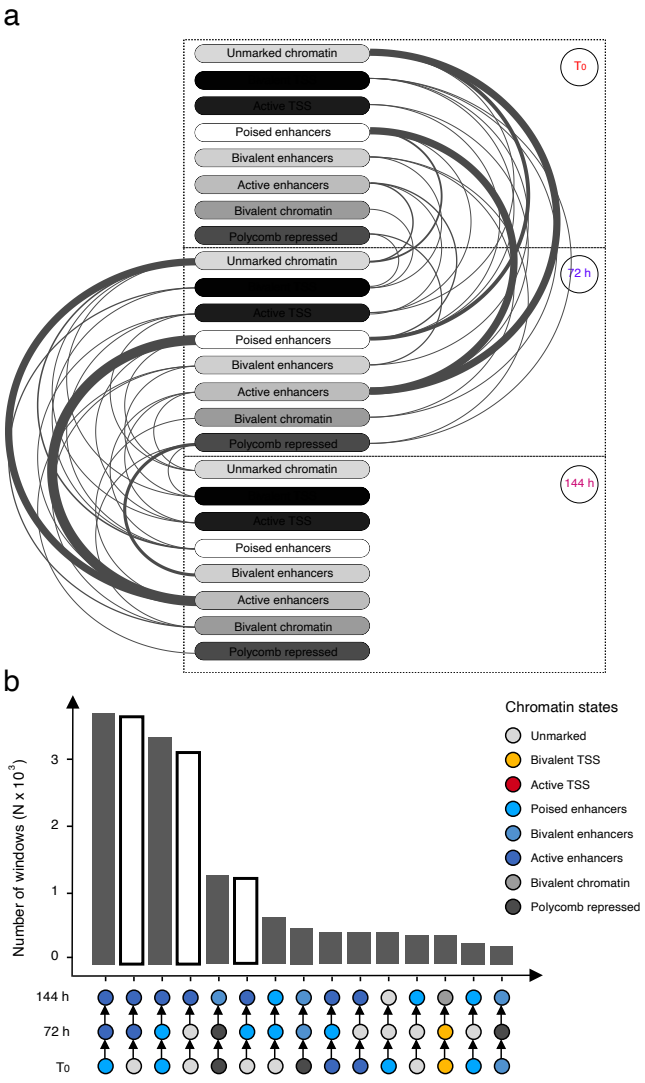

a
Unmarked chromatin
Active TSS
Poised enhancers
Bivalent enhancers
Active enhancers
Bivalent chromatin
Polycomb repressed
T0
72 h
144 h
b
Number of windows (N x 10^3)
Chromatin states
Unmarked
Bivalent TSS
Active TSS
Poised enhancers
Bivalent enhancers
Active enhancers
Bivalent chromatin
Polycomb repressed
144 h
72 h
T0
c
Migrants
- HOX
- CEBP
- POU
- SOX
- TBX
- RAR
- TCF
- SREBF1
Pioneers
- AP1
- E2F
- EGR
- ELF
Settlers
- FOX
- RXR
- STAT
- NFY
- RELA
Chromatin opening index
Chromatin dependence
d
AP1 family
Pioneers
Settlers
Migrants
Dynamicity
Windows
COI
CD
Prestimulation
Enhancers
Dim 2 (10.8%)
Dim 1 (66.1%)

Figure 2

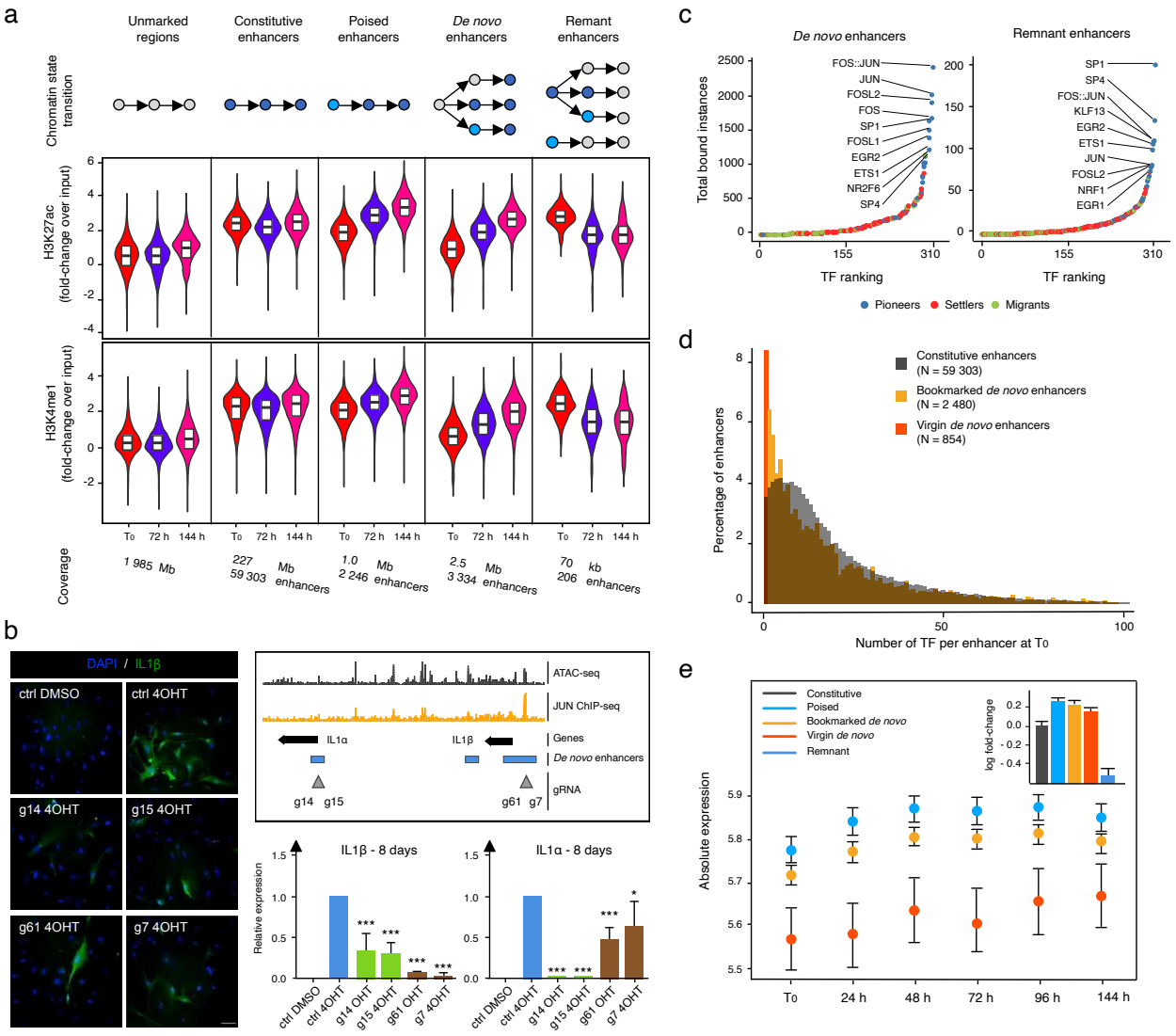

a
Unmarked regions
Constitutive enhancers
Poised enhancers
De novo enhancers
Remant enhancers
Chromatin state transition
H3K27ac (fold-change over input)
H3K4me1 (fold-change over input)
T0 72 h 144 h
Coverage
1 985 Mb
227 59 303 Mb enhancers
1.0 2 246 Mb enhancers
2.5 3 334 Mb enhancers
70 206 kb enhancers
b
DAPI / IL1β
ctrl DMSO
ctrl 4OHT
g14 4OHT
g15 4OHT
g61 4OHT
g7 4OHT
ATAC-seq
JUN ChIP-seq
IL1α
IL1β
Genes
De novo enhancers
gRNA
g14 g15
g61 g7
IL1β - 8 days
IL1α - 8 days
Relative expression
ctrl DMSO
ctrl 4OHT
g14 OHT
g15 4OHT
g61 OHT
g7 4OHT
c
De novo enhancers
Remnant enhancers
Total bound instances
TF ranking
FOS::JUN
JUN
FOSL2
FOS
SP1
FOSL1
EGR2
ETS1
NR2F6
SP4
SP1
SP4
FOS::JUN
KLF13
EGR2
ETS1
JUN
FOSL2
NRF1
EGR1
Pioneers
Settlers
Migrants
d
Constitutive enhancers (N = 59 303)
Bookmarked de novo enhancers (N = 2 480)
Virgin de novo enhancers (N = 854)
Percentage of enhancers
Number of TF per enhancer at T0
e
Constitutive
Poised
Bookmarked de novo
Virgin de novo
Remnant
log fold-change
Absolute expression
T0
24 h
48 h
72 h
96 h
144 h

Figure 3

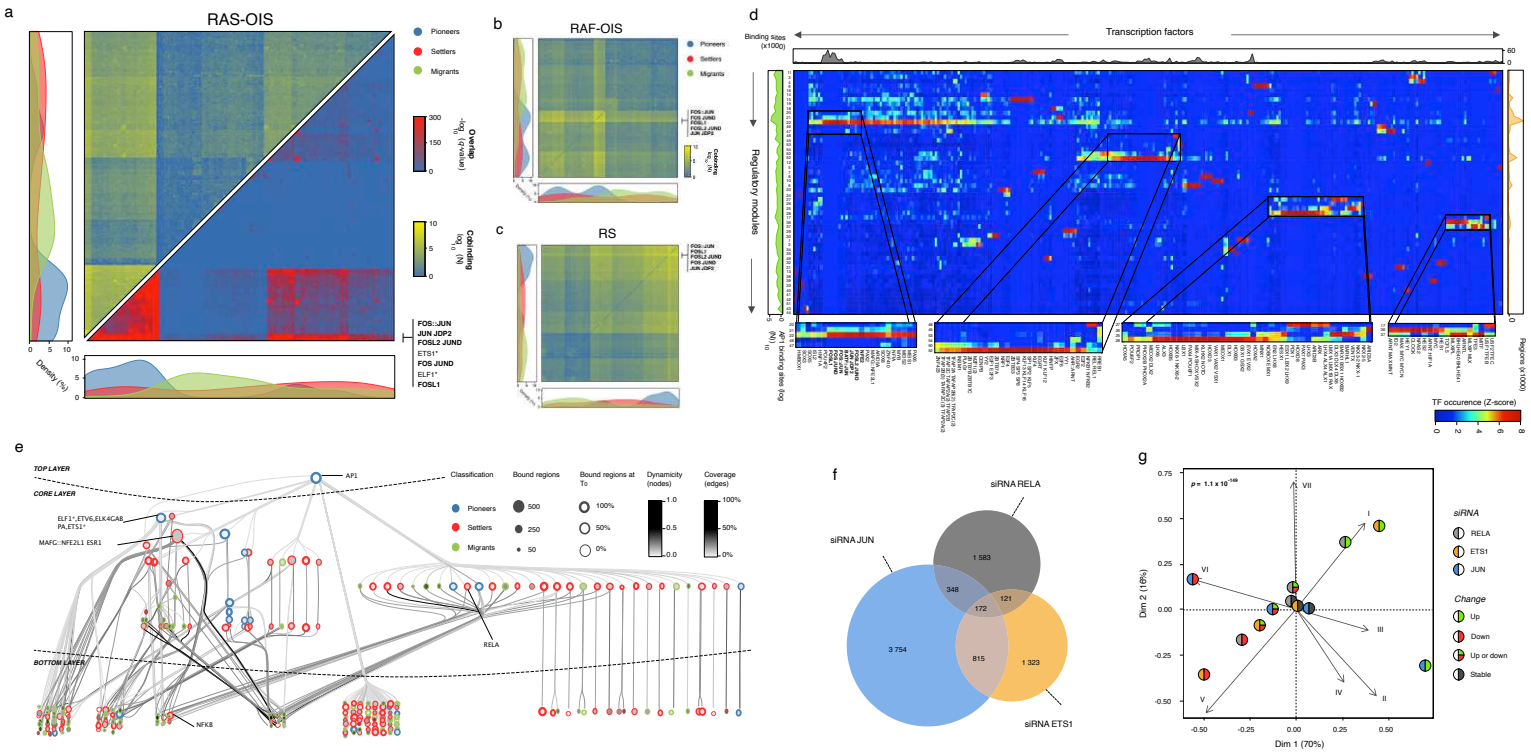

a
RAS-OIS
Pioneers
Settlers
Migrants
Overlap
Colocalizing
FOS::JUN
JUN JDP2
FOSL2 JUND
ETS1*
FOS JUND
ELF1*
FOSL1
Density (%)
b
RAF-OIS
c
RS
d
Transcription factors
Binding sites (x100)
Regulatory modules
Regions (x100)
TF occurence (Z-score)
e
TOP LAYER
CORE LAYER
BOTTOM LAYER
AP1
ELF1*,ETV6,ELK4CAB
PA,ETS1*
MAFG::NFE2L1 ESR1
RELA
NFKB
Classification
Bound regions
Bound regions at To
Dynamicity (nodes)
Coverage (edges)
f
siRNA RELA
siRNA JUN
siRNA ETS1
1 583
348
121
172
3 754
815
1 323
g
Dim 1 (70%)
Dim 2 (18%)
siRNA
RELA
ETS1
JUN
Change
Up
Down
Up or down
Stable

Figure 4

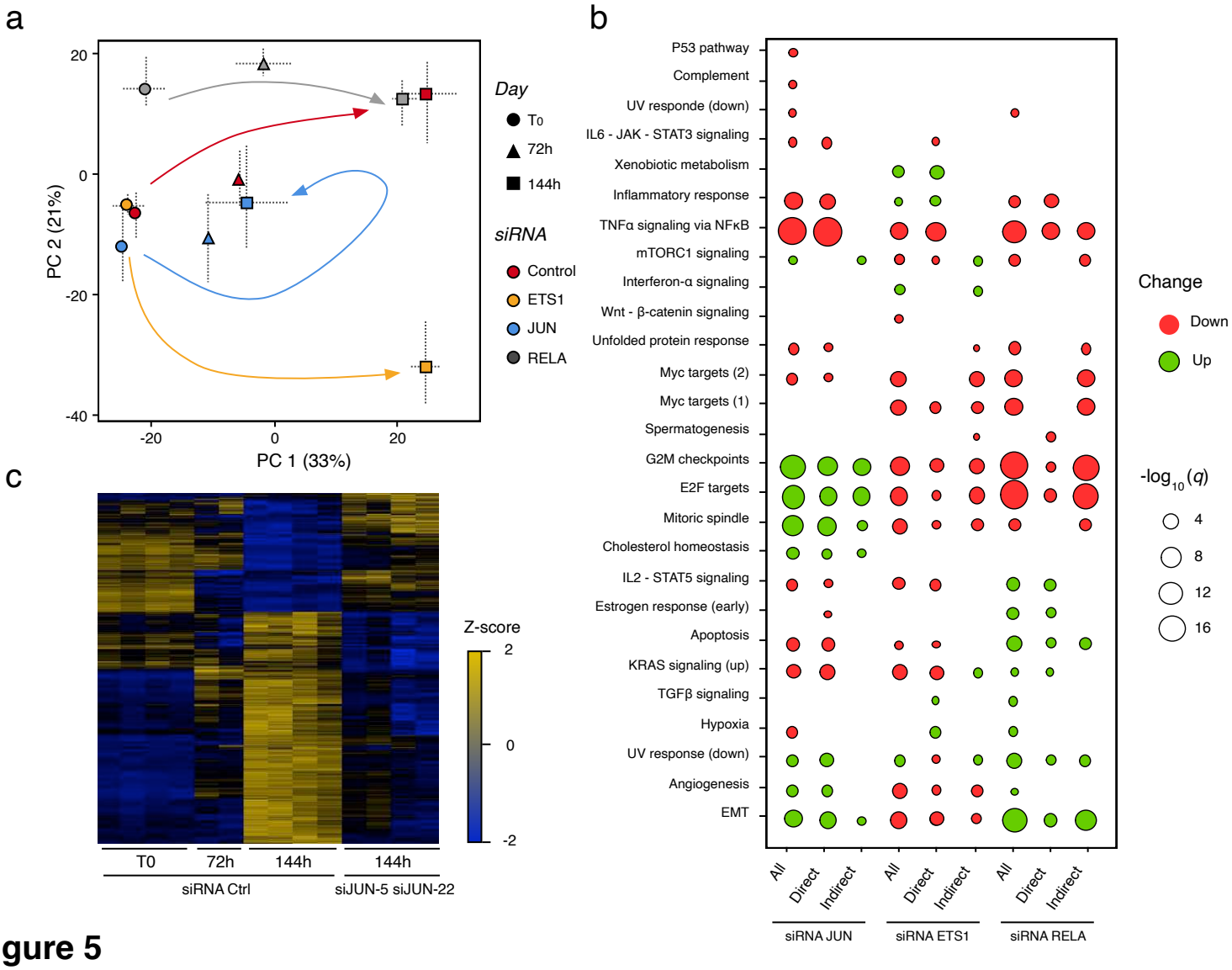

a
PC 2 (21%)
PC 1 (33%)
Day
T0
72h
144h
siRNA
Control
ETS1
JUN
RELA
b
P53 pathway
Complement
UV responde (down)
IL6 - JAK - STAT3 signaling
Xenobiotic metabolism
Inflammatory response
TNFα signaling via NFκB
mTORC1 signaling
Interferon-α signaling
Wnt - β-catenin signaling
Unfolded protein response
Myc targets (2)
Myc targets (1)
Spermatogenesis
G2M checkpoints
E2F targets
Mitoric spindle
Cholesterol homeostasis
IL2 - STAT5 signaling
Estrogen response (early)
Apoptosis
KRAS signaling (up)
TGFβ signaling
Hypoxia
UV response (down)
Angiogenesis
EMT
All
Direct
Indirect
siRNA JUN
siRNA ETS1
siRNA RELA
Change
Down
Up
-log10(q)
4
8
12
16
c
Z-score
T0
72h
144h
siRNA Ctrl
siJUN-5 siJUN-22

Figure 5

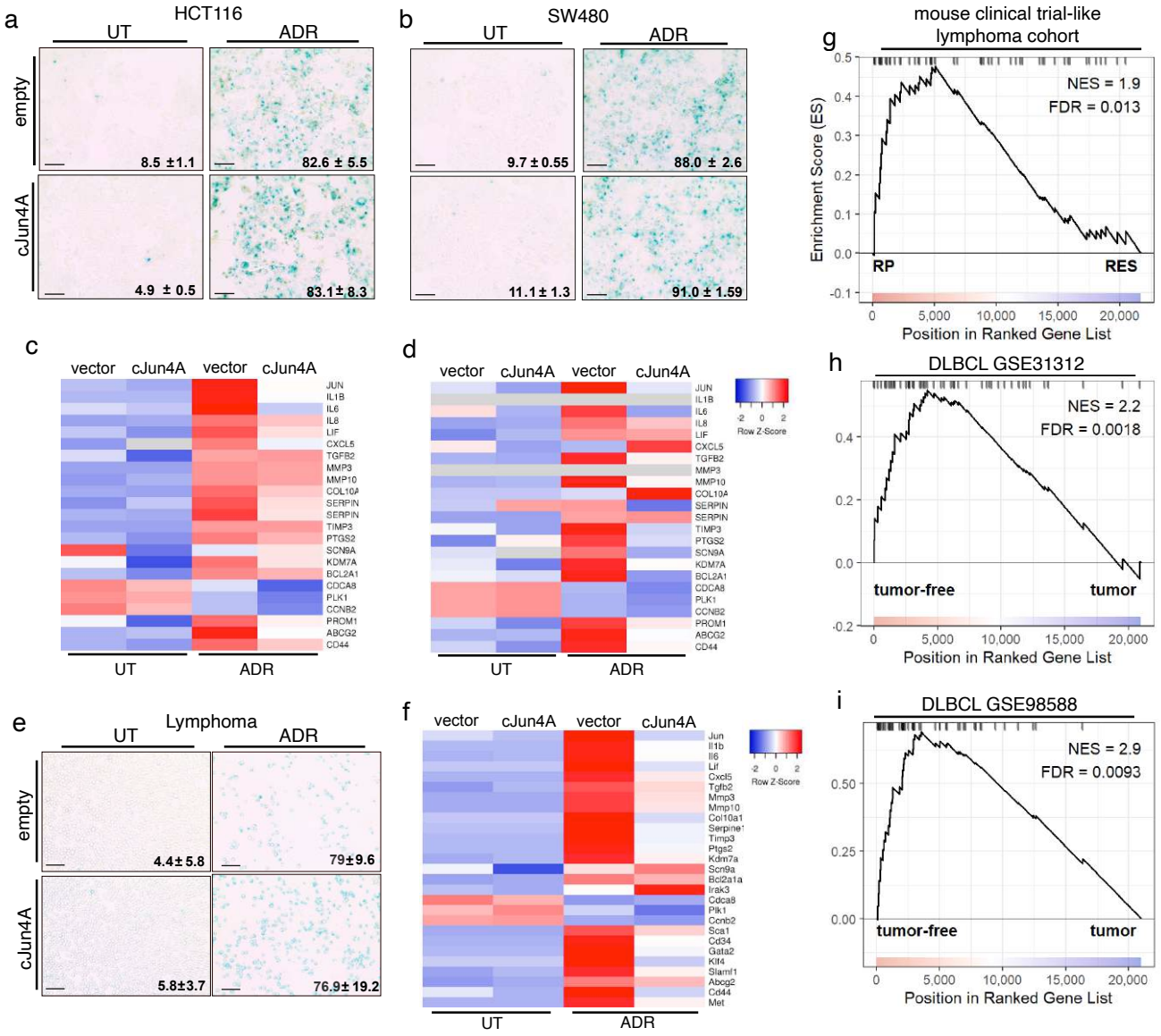
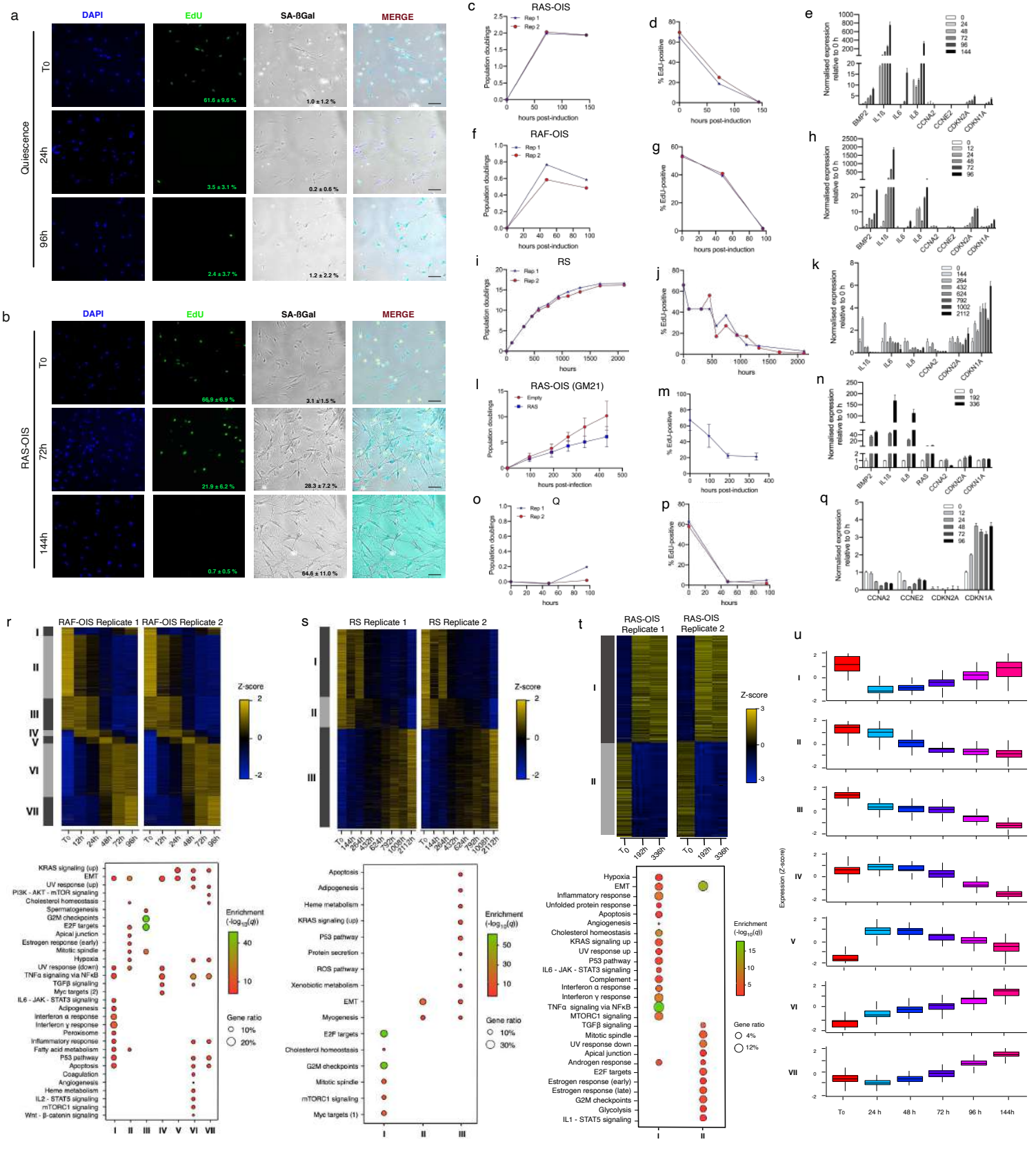

a
HCT116
UT
ADR
empty
cJun4A
8.5 ±1.1
82.6 ± 5.5
4.9 ± 0.5
83.1 ± 8.3
b
SW480
UT
ADR
9.7 ± 0.55
88.0 ± 2.6
11.1 ± 1.3
91.0 ± 1.59
g
mouse clinical trial-like lymphoma cohort
NES = 1.9
FDR = 0.013
Enrichment Score (ES)
RP
RES
Position in Ranked Gene List
c
vector
cJun4A
vector
cJun4A
JUN
IL1B
IL6
IL8
LIF
CXCL5
TGFB2
MMP3
MMP10
SERPIN
SERPIN
TIMP3
PTGS2
SCN9A
KDM7A
BCL2A1
CDCA8
PLK1
CCNB2
PROM1
ABCG2
CD44
UT
ADR
d
vector
cJun4A
vector
cJun4A
Row Z-Score
UT
ADR
h
DLBCL GSE31312
NES = 2.2
FDR = 0.0018
tumor-free
tumor
Position in Ranked Gene List
e
Lymphoma
UT
ADR
empty
cJun4A
4.4± 5.8
79±9.6
5.8±3.7
76.9± 19.2
f
vector
cJun4A
vector
cJun4A
Jun
Il1b
Il6
Lif
Cxcl5
Tgfb2
Mmp3
Mmp10
Col10a1
Serpine1
Timp3
Ptgs2
Kdm7a
Scn9a
Bcl2a1a
Irak3
Cdca8
Plk1
Ccnb2
Sca1
Cd34
Gata2
Klf4
Slamf1
Abcg2
Cd44
Met
Row Z-Score
UT
ADR
i
DLBCL GSE98588
NES = 2.9
FDR = 0.0093
tumor-free
tumor
Position in Ranked Gene List

Figure 6


a
DAPI
EdU
SA-ßGal
MERGE
Quiescence
To
24h
96h
61.6 ± 9.6 %
1.0 ± 1.2 %
3.5 ± 3.1 %
0.2 ± 0.6 %
2.4 ± 3.7 %
1.2 ± 2.2 %
b
RAS-OIS
72h
144h
66.9 ± 6.9 %
3.1 ± 1.5 %
21.9 ± 6.2 %
28.3 ± 7.2 %
0.7 ± 0.5 %
64.6 ± 11.0 %
c
RAS-OIS
Rep 1
Rep 2
Population doublings
hours post-induction
d
% EdU-positive
e
Normalised expression relative to 0 h
BMP2
IL1B
IL6
IL8
CCNA2
CCNE2
CDKN2A
CDKN1A
f
RAF-OIS
g
h
i
RS
hours
j
k
l
RAS-OIS (GM21)
Empty
RAS
hours post-infection
m
n
o
Q
p
q
r
RAF-OIS Replicate 1
RAF-OIS Replicate 2
Z-score
KRAS signaling (up)
EMT
UV response (up)
PI3K - AKT - mTOR signaling
Cholesterol homeostasis
Spermatogenesis
G2M checkpoints
E2F targets
Apical junction
Estrogen response (early)
Mitotic spindle
Hypoxia
UV response (down)
TNFα signaling via NFκB
TGFβ signaling
Myc targets (2)
IL6 - JAK - STAT3 signaling
Adipogenesis
Interferon α response
Interferon γ response
Peroxisome
Inflammatory response
Fatty acid metabolism
P53 pathway
Apoptosis
Coagulation
Angiogenesis
Heme metabolism
IL2 - STAT5 signaling
mTORC1 signaling
Wnt - β-catenin signaling
Enrichment (-log10(q))
Gene ratio
s
RS Replicate 1
RS Replicate 2
Apoptosis
Adipogenesis
Heme metabolism
KRAS signaling (up)
P53 pathway
Protein secretion
ROS pathway
Xenobiotic metabolism
EMT
Myogenesis
E2F targets
Cholesterol homeostasis
G2M checkpoints
Mitotic spindle
mTORC1 signaling
Myc targets (1)
t
RAS-OIS Replicate 1
RAS-OIS Replicate 2
Hypoxia
EMT
Inflammatory response
Unfolded protein response
Apoptosis
Angiogenesis
Cholesterol homeostasis
KRAS signaling up
UV response up
P53 pathway
IL6 - JAK - STAT3 signaling
Complement
Interferon α response
Interferon γ response
TNFα signaling via NFκB
MTORC1 signaling
TGFβ signaling
Mitotic spindle
UV response down
Apical junction
Androgen response
E2F targets
Estrogen response (early)
Estrogen response (late)
G2M checkpoints
Glycolysis
IL1 - STAT5 signaling
u
Expression (Z-score)
To
24 h
48 h
72 h
96 h
144h

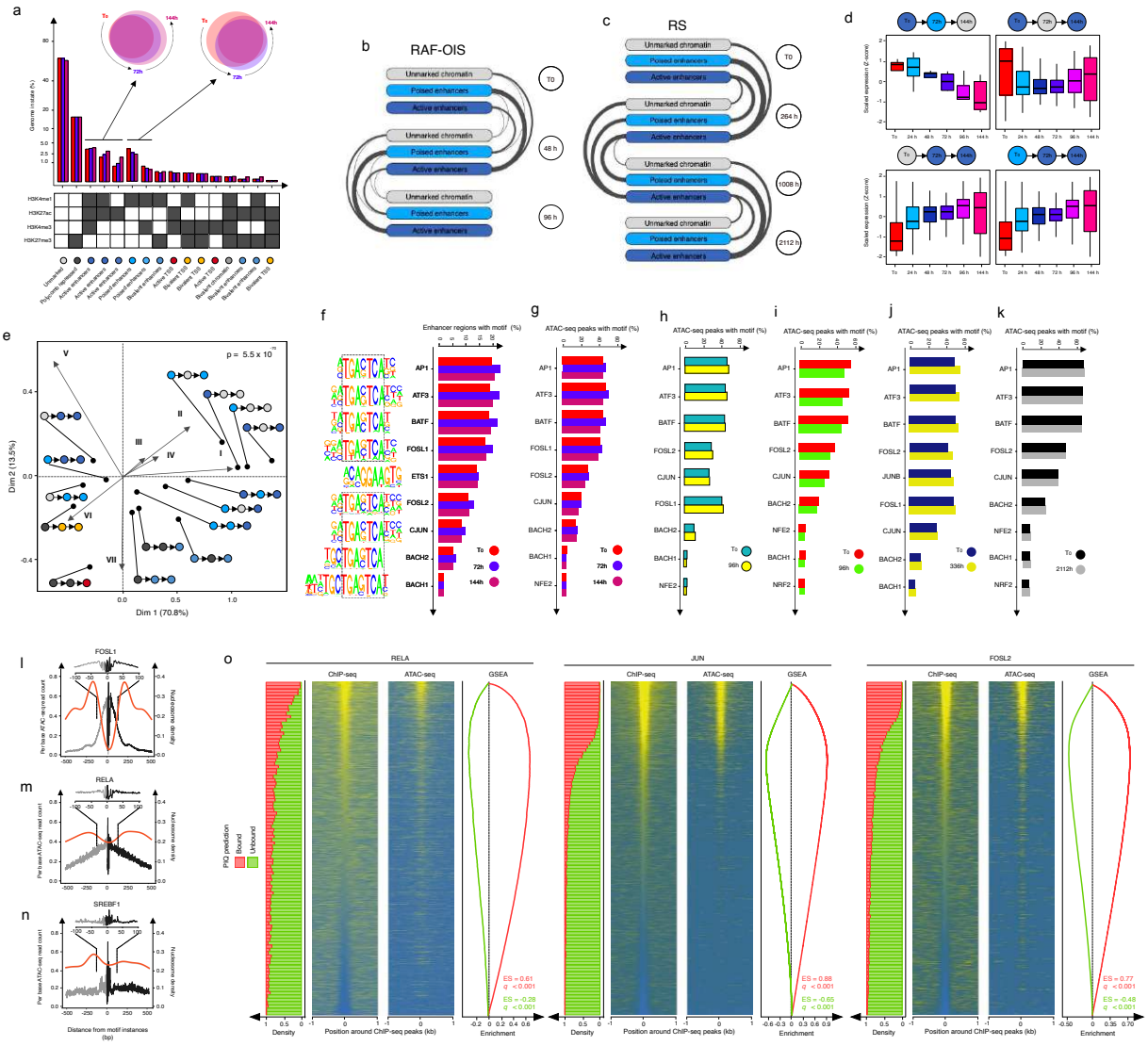

a
b
RAF-OIS
c
RS
d
e
f
g
h
i
j
k
l
m
n
o
RELA
JUN
FOSL2
ChIP-seq
ATAC-seq
GSEA
Density
Enrichment

Figure S2

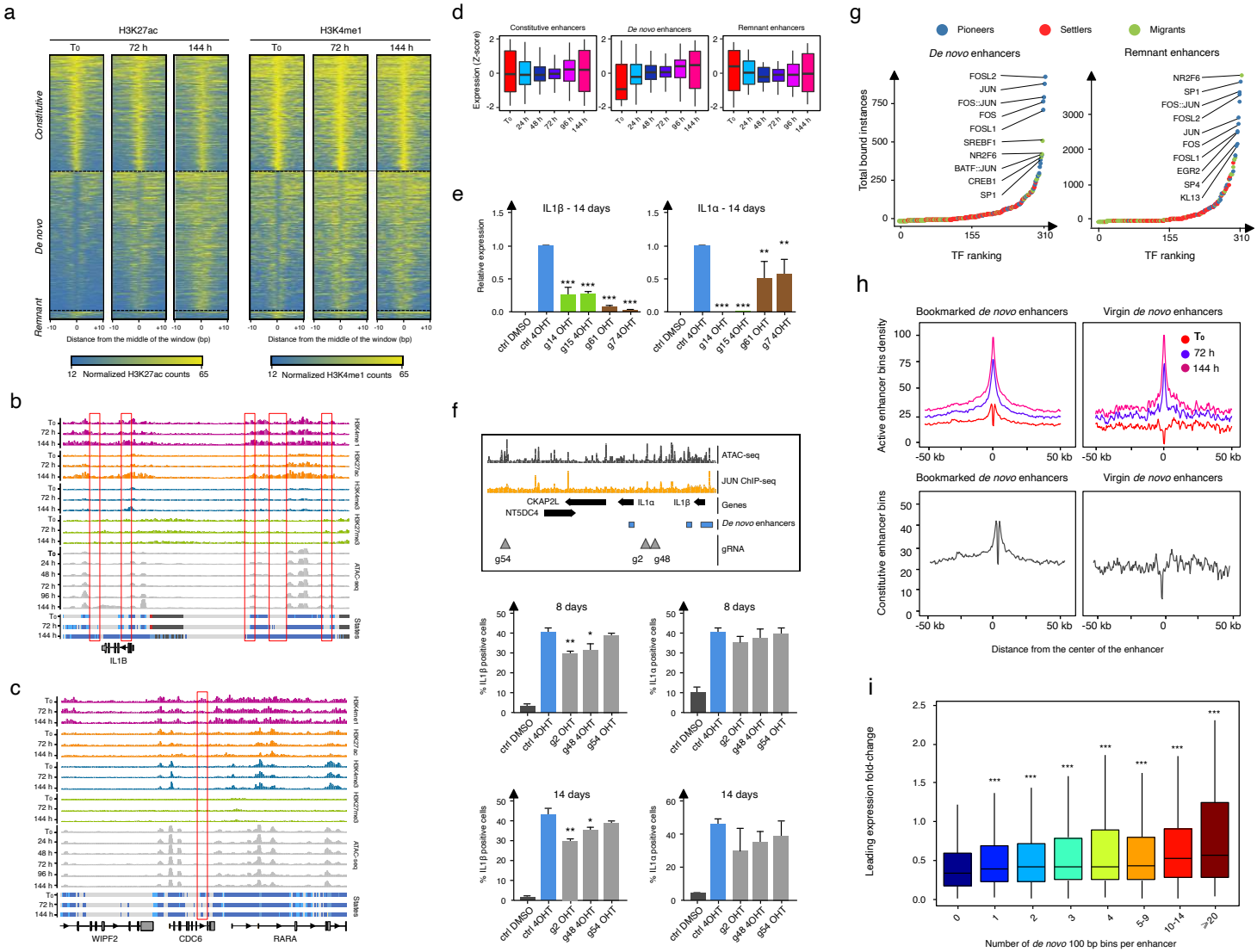

a
H3K27ac
H3K4me1
T0
72 h
144 h
Constitutive
De novo
Remnant
Distance from the middle of the window (bp)
Normalized H3K27ac counts
Normalized H3K4me1 counts
b
IL1B
c
WIPF2
CDC6
RARA
d
Constitutive enhancers
De novo enhancers
Remnant enhancers
Expression (Z-score)
e
IL1β - 14 days
IL1α - 14 days
Relative expression
ctrl DMSO
ctrl 4OHT
g14 OHT
g15 4OHT
g61 OHT
g7 4OHT
f
ATAC-seq
JUN ChIP-seq
CKAP2L
IL1α
IL1β
NT5DC4
Genes
De novo enhancers
gRNA
g54
g2
g48
8 days
14 days
% IL1β positive cells
% IL1α positive cells
g2 OHT
g48 4OHT
g54 OHT
g
Pioneers
Settlers
Migrants
De novo enhancers
Remnant enhancers
Total bound instances
TF ranking
FOSL2
JUN
FOS::JUN
FOS
FOSL1
SREBF1
NR2F6
BATF::JUN
CREB1
SP1
NR2F6
SP1
FOS::JUN
FOSL2
JUN
FOS
FOSL1
EGR2
SP4
KL13
h
Bookmarked de novo enhancers
Virgin de novo enhancers
Active enhancer bins density
Constitutive enhancer bins
Distance from the center of the enhancer
i
Leading expression fold-change
Number of de novo 100 bp bins per enhancer

Figure S3

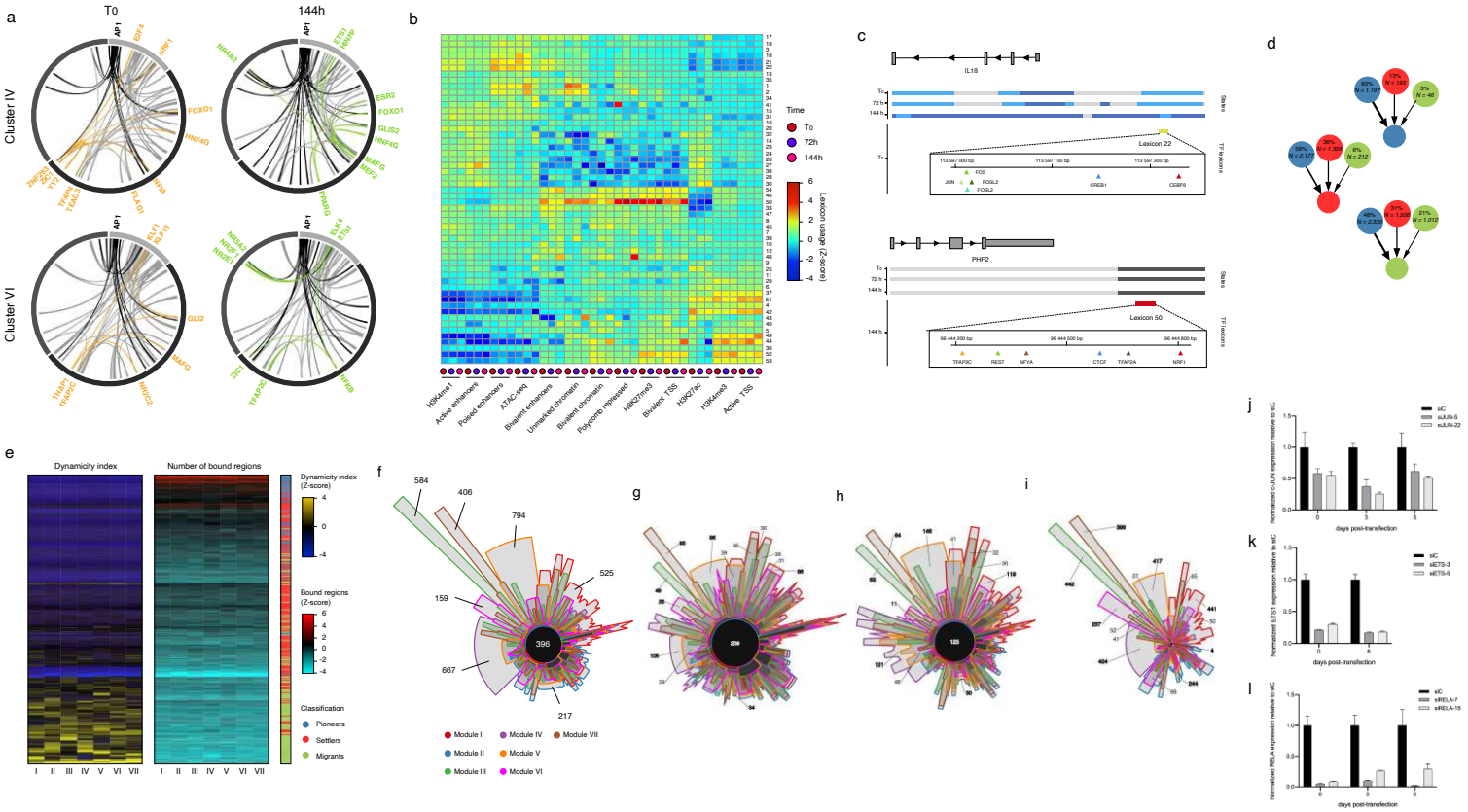

a
T0
144h
Cluster IV
Cluster VI
b
Time
T0
72h
144h
Lexicon usage (Z-score)
H3K4me1
Active enhancers
Poised enhancers
ATAC-seq
Bivalent enhancers
Unmarked chromatin
Bivalent chromatin
Polycomb repressed
H3K27me3
Bivalent TSS
H3K27ac
H3K4me3
Active TSS
c
IL18
Lexicon 22
PHF2
Lexicon 50
d
e
Dynamicity index
Number of bound regions
Dynamicity index (Z-score)
Bound regions (Z-score)
Classification
Pioneers
Settlers
Migrants
f
584
406
794
525
159
396
667
217
Module I
Module II
Module III
Module IV
Module V
Module VI
Module VII
g
h
i
j
Normalized c-JUN expression relative to siC
siC
siJUN-5
siJUN-22
days post-transfection
k
Normalized ETS1 expression relative to siC
siC
siETS-3
siETS-5
days post-transfection
l
Normalized RELA expression relative to siC
siC
siRELA-7
siRELA-15
days post-transfection

Figure S4

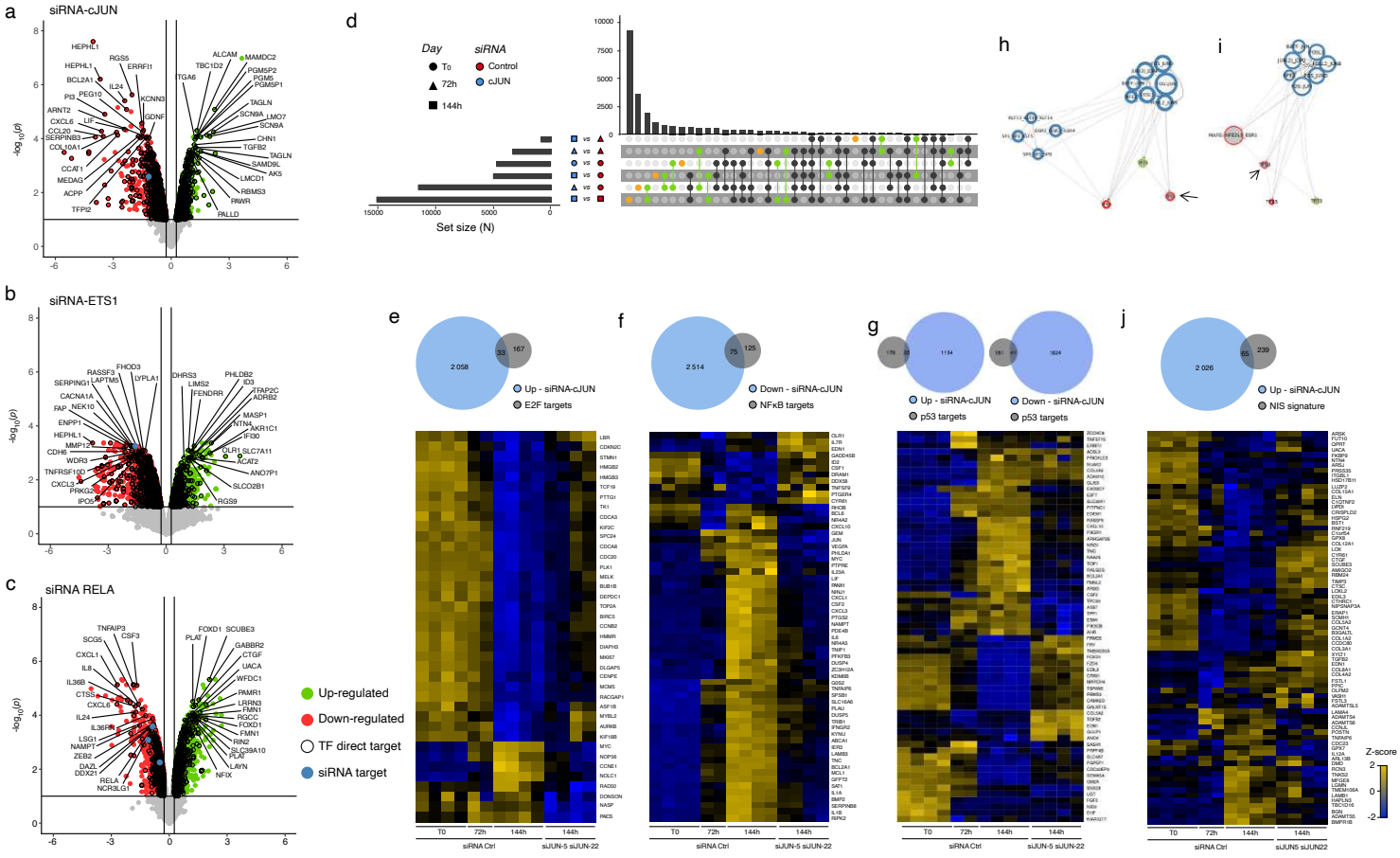

a
siRNA-cJUN
b
siRNA-ETS1
c
siRNA RELA
Up-regulated
Down-regulated
TF direct target
siRNA target
d
Day
siRNA
To
72h
144h
Control
cJUN
Set size (N)
h
i
e
2 058
33
167
Up - siRNA-cJUN
E2F targets
f
2 514
75
125
Down - siRNA-cJUN
NFκB targets
g
Up - siRNA-cJUN
p53 targets
Down - siRNA-cJUN
p53 targets
j
2 026
65
239
Up - siRNA-cJUN
NIS signature
Z-score
siRNA Ctrl
siJUN-5 siJUN-22
siJUN5 siJUN22

Figure S5

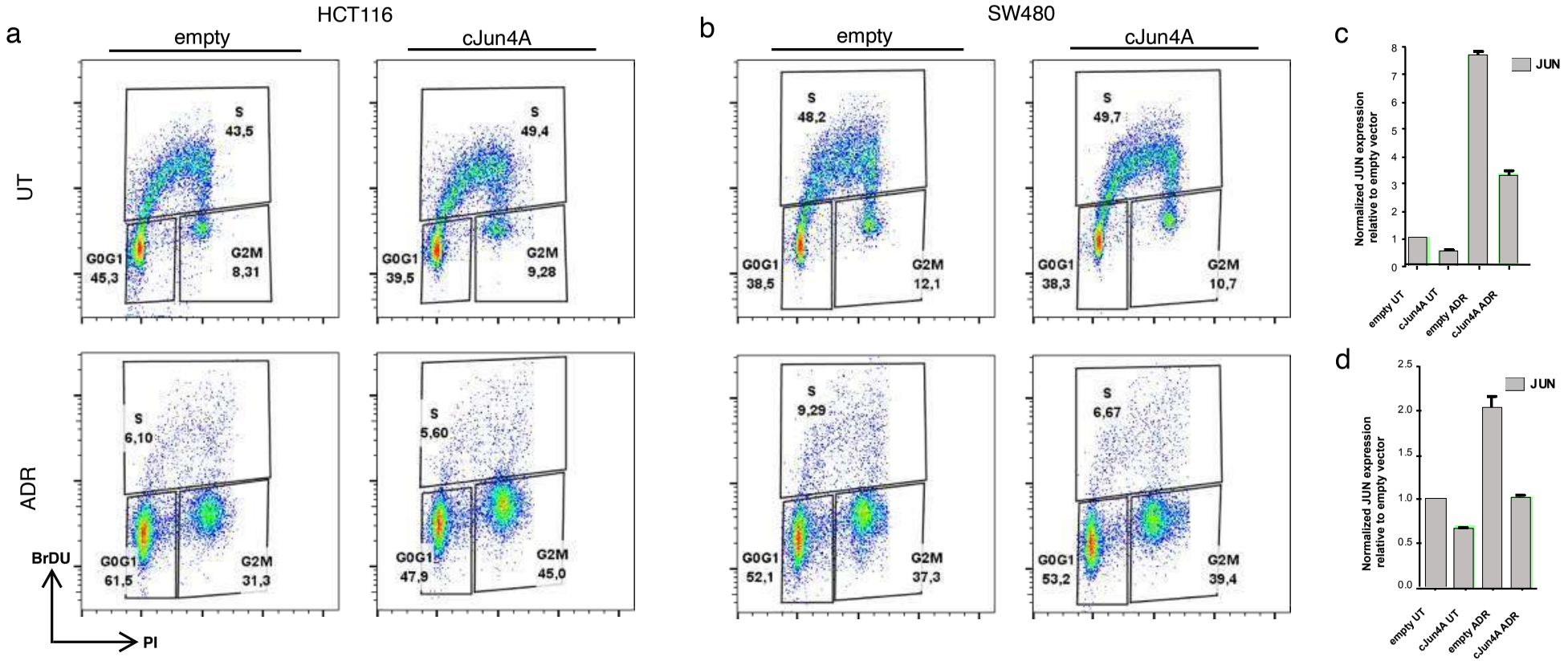

a
HCT116
empty
cJun4A
UT
ADR
BrDU
PI
S
43,5
G0G1
45,3
G2M
8,31
S
49,4
G0G1
39,5
G2M
9,28
S
6,10
G0G1
61,5
G2M
31,3
S
5,60
G0G1
47,9
G2M
45,0
b
SW480
empty
cJun4A
S
48,2
G0G1
38,5
G2M
12,1
S
49,7
G0G1
38,3
G2M
10,7
S
9,29
G0G1
52,1
G2M
37,3
S
6,67
G0G1
53,2
G2M
39,4
c
JUN
Normalized JUN expression relative to empty vector
empty UT
cJun4A UT
empty ADR
cJun4A ADR
d
JUN
Normalized JUN expression relative to empty vector
empty UT
cJun4A UT
empty ADR
cJun4A ADR

Figure S6