



**HAL**  
open science

# Dynamic Modeling of Cellular Senescence Gene Regulatory Network

José Américo Nabuco Leva Ferreira de Freitas, Oliver Bischof

► **To cite this version:**

José Américo Nabuco Leva Ferreira de Freitas, Oliver Bischof. Dynamic Modeling of Cellular Senescence Gene Regulatory Network. SSRN Electronic Journal, 2022, 10.2139/ssrn.4199883. hal-03799379

**HAL Id: hal-03799379**

**<https://hal.science/hal-03799379v1>**

Submitted on 15 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Dynamic Modeling of Cellular Senescence Gene Regulatory Network

José Américo Nabuco Leva Ferreira Freitas<sup>1</sup> and Oliver Bischof<sup>2\*</sup>

<sup>1</sup> Sorbonne Université, CNRS, INSERM, Institut de Biologie Paris-Seine, Biological Adaptation and Ageing, B2A-IBPS, 75005, Paris, France.

<sup>2</sup> INSERM U955, Université Paris-Est Créteil (UPEC), FHU SENECA, 51 Av de Lattre de Tassigny, 94100 Créteil, France.

## Contact information

\*Correspondence: [oliver.bischof@cnr.fr](mailto:oliver.bischof@cnr.fr)

## Title page footnotes

<sup>2</sup>Lead contact

## Summary

Cellular senescence is a cell fate that prominently impacts physiological and pathophysiological processes. Diverse cellular stresses induce it, and dramatic gene expression changes accompany it. However, determining the interactions comprising the gene regulatory network (GRN) governing senescence remains a challenge. Recent advances in signal processing techniques provide opportunities to reconstruct GRNs. Here, we describe a GRN controlling senescence integrating time-series transcriptome and transcription factor depletion datasets. We infer a set of differential equations modeling the CS transcriptome using the “Sparse Identification of Nonlinear Dynamics” (SINDy) algorithm, discriminate genes with potential hidden regulators, validating the inferred GRN for time points not included in the training data. Our work is a proof of concept of a data-based method for GRN reconstruction, consolidating an iterative, powerful mathematical platform for more comprehensive senescence models that can be used to test hypotheses in silico and has the potential for future discoveries of clinical impact.

## Introduction

Senescence is a cell fate induced by diverse cellular stressors. It is characterized by a stable proliferative arrest, an inflammatory secretome (alias senescence-associated secretory phenotype, SASP), apoptosis resistance, and altered mitochondrial and lysosomal activity. Senescent cells accumulate with age in numerous tissues and are now considered a significant driver of age-related pathologies. Notably, the elimination of senescent cells improves organismal fitness. Conversely, senescent cells are also essential in physiological processes like embryonic development, wound healing, and tumor suppression. Consequently, there is great interest in deciphering senescence molecular networks and exploiting senescence targeting to extend health span (Gorgoulis et al., 2019; Martínez-Zamudio et al., 2017a, b).

The dynamic senescence cell fate specification process is regulated by regulatory gene networks (GRNs). A hierarchical transcription factor network defines the temporal order of specification events. (Martínez-Zamudio et al., 2020). Thus, gene expression modeling provides enormous opportunities to predict the behavior of cells under various external/internal senescence conditions. Due to recent advances in high-throughput data acquisition, several methods for GRN inference from transcriptomic data have been proposed, including information theory models, Boolean networks, Bayesian networks, and differential equations models (Delgado and Gómez-Vela, 2019). However, gene expression modeling still poses a fundamental challenge because the complexity of GRNs implies a high number of parameters to be inferred, and performing a sufficient number of experiments to infer a uniquely determined system of coefficients is generally not feasible.

A recent paradigm in signal processing, known as “Compressed Sensing (CoS),” revolutionized our understanding of the amount of high-quality data required to accurately recover a signal from measured data (Candès et al., 2006). In other words, one can recover the GRN from dynamic data with a lower number of experimentally measured time points, given that two assumptions are satisfied. First, the recovered signal must be sparse, i.e., genes must interact with only a subset of the GRN. Second, the matrix generated by aggregating the time course data must satisfy the “Restricted Isometry Property (RIP),” ensuring that a unique GRN can generate the experimental data. To fulfill these conditions, one can scale each time profile to zero-mean and unit variance profiles, as it has been shown that Gaussian matrices satisfy the RIP with high probability (Wang, 2016; Candès et al., 2006). Since GRNs are generally sparse (Broido and Clauset, 2019; Ramirez et al., 2017; Goode et al., 2016), the CoS paradigm can potentially be applied to

and significantly improve GRN inference. CoS theory also allows inferring potential hidden sources (HSs) from time-course data, i.e., the dynamics of regulators that were not experimentally measured but none-the-less impact gene expression (Wang et al., 2016; Chang et al., 2014; Shen et al., 2014; Su et al., 2014). Genes with HSs are highlighted when performing GRN recovery with distinct subsets of the measured data. The coefficients representing the effect of regulators on their transcription will present a high variance, even when considering the higher number of measured time points. Conversely, genes with coefficients converging to one finite value as the number of experimental data increases have dynamics that are suitably described by the regulators detected during GRN inference.

The “Sparse Identification of Non-linear Dynamic Systems (SINDy)” method explores the strength of CoS. It has been applied to generate sets of differential equations describing systems in distinct domains, including fluid dynamics, mechanical systems, enzyme kinetics, and metabolic networks (Kaheman et al., 2020; Hoffmann et al., 2019; Mangan et al., 2016; Brunton et al., 2016). Interestingly, this method also addresses nonlinear interactions between entities, accounting for collective interactions of genes when regulating a target, e.g., two proteins that bind together before binding to the chromatin. By explicitly representing the nonlinear terms governing the system dynamics, SINDy provides an interpretable and parsimonious model that can be examined to uncover experimental insights. Further to that, SINDy can be applied to distinct data subsets to identify possible HSs. However, SINDy and HS analysis have not been systematically used for gene expression modeling despite their demonstrated potential for modeling.

In this work, we integrate time-series experiments of cells undergoing oncogenic RAS-induced senescence (OIS) (Martínez-Zamudio et al., 2020) to build a system of differential equations that describes the GRN governing OIS using SINDy and high-performance computing. We validate the GRN generalization by comparing simulation results with experimental data acquired after the inhibition of two regulators of the CS phenotype, AP1-cJUN, and RELA, showing that model performance is dependent on the functions performed by each gene. Finally, we highlight the GRN genes whose dynamics can be inferred by hidden source analysis. Our novel modeling approach to the senescence transcriptome using a nonlinear system of equations provides a solid platform and stepping stone for unraveling senescent cells' gene regulatory interactions and vulnerabilities. The ultimate goal is to facilitate therapeutic decision-making for health benefits.

## Results

### Generation and integration of time-resolved senescence transcriptomes for mathematical modeling

Gene expression modeling still poses a fundamental challenge but provides enormous opportunities to predict the behavior of cells under various external/internal conditions. To mathematically describe the senescence transcriptome, we used time-resolved transcriptomes generated at six-time points (0, 24, 48, 72, 96, and 144 h) from cells undergoing oncogenic RAS-induced senescence (RAS-OIS) for two biological replicates and determined global gene expression levels as previously published (Martínez-Zamudio et al., 2020). Our dynamic sampling range encompasses the basal proliferative (0 h), hyperproliferative (24-48 h), establishment (72-96 h), and maintenance (144 h) stages of RAS-OIS, thus, highlighting sets of activated genes in this biological process, their expression changes, their order and their causal effects (Figure 1A). To empower our modeling approach further, we included transcriptomic data sets from RAS-OIS cells knocked down (KD) for the expression of cardinal transcription factors (TF) AP1-cJUN and RELA at 72h and 144h and ETS1 at 144h after RAS-induction by siRNAs collecting samples 48h after KD as previously described (Supplemental Figure S1A) (Martínez-Zamudio et al., 2020). Specifically, we used 72h AP1-cJUN and RELA KD and 144h ETS1 KD data sets for model training and the remaining two 144h KD datasets for AP1-cJUN and RELA for validation. The inclusion of TF KD datasets provides two benefits: first, the "transcriptome landscape" covered by the model training data is expanded, thus increasing the range of predictions it can perform (Casadiego et al., 2017). Second, it increases model resolution. From a data-driven perspective, genes with highly correlated expression profiles in one condition provide the same information to the model and cannot be individually distinguished. Therefore, by integrating disparate datasets, it is likely that the expression of highly correlated genes develops distinct responses, allowing the model to identify covariates that regulate each gene separately and refining the inferred gene regulatory network (GRN). Together, we used these data sets for model training to compute model coefficients and model validation, as outlined below.

Integration of disparate datasets naturally necessitates normalization to ensure that the effects being accounted for by the model are faithfully reflected in the biological process and not by technical variations inherent to high-throughput technologies. Therefore, as a first step, we normalized all datasets for technical variation using ComBat (Johnson et al., 2007). Next, we independently clustered each time course using WGCNA

(Langfelder and Horvath, 2008), identifying eleven to seventeen gene modules with highly correlated expression trajectories (Figure 1B and Supplemental Figures S1B-D). 5124 genes were differentially expressed ( $> 30\%$  in expression levels) compared to un-induced, proliferating cells at 0h across the four experimental data sets. We aggregated gene sets belonging to the same module for all training datasets, i.e., genes with expression profiles highly correlated in all experimental conditions. We tabulated the WGCNA module for each gene in each experimental condition and identified seven hundred and ninety-nine distinct transcription module combinations (Figure 1C). Each combination represents a unique temporal expression profile. We assigned each transcription module combination to a variable from which dynamics was inferred mathematically in this work. Three hundred seventy-five combinations are associated with one gene (Supplemental Figure S1E). Since AP1-cJUN, ETS1, and RELA are associated with variables corresponding to more than one gene, we assigned these TFs to an extra variable, resulting in eight hundred and two variables for our equation system.

Altogether, we identified distinct gene expression profiles and associated variables using time-resolved transcriptomes from RAS-OIS cells depleted for TFs AP1-cJUN, RELA, and ETS1 that form the foundational basis for the inference of a dynamic mathematical system for RAS-OIS transcriptome modeling.

### **Senescence transcriptome modeling using sparse identification of nonlinear dynamics (SINDy)**

The SINDy algorithm infers nonlinear dynamics from time-course data by performing a sparse linear regression in a matrix composed of nonlinear combinations of the input time courses (Brunton et al., 2016). This nonlinear paradigm is suitable to describe the collective activity of agents regulating the same target, e.g., TFs that functionally interact to modulate gene expression. The SINDy algorithm provides parsimonious models that leverage the "Compressed Sensing (CoS)" concept and avoids overfitting. In this algorithm, a matrix  $X$  consisting of time-course data is expanded into a nonlinear set of time profiles  $\Theta(X)$  (Figure 2A). Combined with the derivative of each dynamic profile, denoted  $\dot{X}$ , we inferred the sparse set of coefficients  $\Xi$ , which correspond to the inferred dynamic system's parameters, and, therefore, describe how each gene is regulated. To compute the derivative matrix  $\dot{X}$ , we interpolated each transcriptomic profile by 24, resulting in an interpolated matrix  $X$  with 284 rows (Supplemental Figure S2A). The interpolation factor value was chosen to obtain smooth, derivable time profiles while minimizing the number of rows in matrix  $X$  to reduce the computational resources required

for model inference. Matrix  $\Theta(X)$  contains one column for each variable and its corresponding interpolated time-course data and one column for each pairwise combination of variables, including the multiplication of their respective time profiles (Figure 2A). The inferred differential equations correspond to second-order polynomials, and the coefficients that describe the dynamics for each variable in the system correspond to columns of matrix  $\Xi$ . The latter is inferred independently for each variable by performing a sparse regression in the context of the CoS paradigm. We implemented the SINDy algorithm using the R statistical language (R Core Team, 2017) integrated with the Tensorflow library (Abadi et al., 2015) and run in a High-Performance Computing facility containing Graphics Processing Unit (GPU) clusters. The produced scripts are available as R Markdown files and can be executed by a Singularity container, including all the required libraries available at ZENODO (see section *Data and code availability*).

After identifying the coefficients for each of the eight hundred and two variables in our mathematical system (see Figure 1C), we determined model performance by simulating the obtained equations for the RAS-OIS time-course data set, depicting in a heatmap observed and simulated time-resolved expression profiles (Figure 2B) and in a corresponding histogram the correlation coefficients between expression profiles for each line (Figures 2C). This analysis demonstrated a high correlation between the experimental and simulated data with 429/802 (53 %) variables resulting in a correlation coefficient greater than 0.9 (Figure 2C). Next, we extended model performance to the AP1-cJUN (72h), RELA (72h), and ETS1 (144h) KD training datasets. In line with the RAS-OIS results, for the AP1-cJUN KD training set (Supplemental Figures S2B, C), we observed a significant congruency between experimental and simulation results, with 662/802 profiles (83 %) having a positive correlation coefficient greater than or equal to 0.9. Similarly, for RELA KD (Supplemental Figures S2D, E), we obtained 583/802 profiles (73%) having a positive correlation coefficient greater than or equal to 0.9. Finally, for ETS1 KD (Supplemental Figures S2F, G), we obtained 737/802 (92%) profiles having a positive correlation coefficient greater than or equal to 0.9.

To validate the inferred equations, we simulated the 144h AP1-cJUN KD (Figures 2D-E) and RELA KD transcriptome datasets (Supplemental Figures S2H-I). Figure 2D shows a heatmap for the experimental and simulated results of AP1-cJUN KD, their correlation for each variable, and the corresponding histogram with the correlation coefficient distribution (Figure 2E). Most profiles (472/802=59%) have a positive correlation coefficient greater than or equal to 0.9, while the minority (99/802=12%) have a negative correlation coefficient smaller than or equal to -0.9. For RELA KD, 298/802 profiles (37%)



exhibit a positive correlation coefficient greater than 0.9 and 193/802 (24%) a negative correlation coefficient smaller than -0.9 (S2H-I), suggesting RELA transcriptomic response might require more experimental data to be accurately predicted.

We conclude that our mathematical systems approach applying the SINDy algorithm can reliably model RAS-OIS gene expression changes and predict the response of TF perturbations at time points not included in the training dataset.

### **Hidden source analysis highlights genes with inferable dynamics**

Apart from allowing the reconstruction with a lower minimum on the required number of experimental time points, the CoS paradigm is also helpful in identifying nodes with hidden sources (HS), i.e., regulators whose activity is not included in the input data (Su et al., 2014). We used this strategy to estimate variables with retrievable dynamics given our available datasets in this work. The hidden source (HS) analysis consists of using SINDy for distinct subsets of the time-course data and assessing the variance of each coefficient for data subsets at different time points (Wang et al., 2016; Shen et al., 2014; Su et al., 2014). Coefficients showing increased variance for a high proportion of the time-course suggest that HSs regulate the variables they describe and that the available data lack necessary information for their modeling. By contrast, coefficients with decreasing variance as data quantity increases indicate that transcription dynamics are predicted (i.e., absence of HSs) even considering a subset of the available data.

To distinguish genes with and without HS and to investigate our model's reliability, we inferred it for four different ratios of time-points (i.e., 20 %, 40 %, 60 %, and 80 % of the 284 rows of matrix  $\Theta(X)$ ) and computed the average variance ( $\sigma_{\text{avg}}$ ) for all coefficients  $\xi_{ij}$  for each variable  $j$  in the model (Figure 3A). We repeated this process five times for each ratio of time points to maximize the  $\sigma_{\text{avg}}$  accuracy, considering the high computational power required to run each inference. To identify variables without HS accurately, we considered variables with decreasing  $\sigma_{\text{avg}}$  as variables coefficients, which presented a  $\sigma_{\text{avg}}$  up to 80 % of the  $\sigma_{\text{avg}}$  obtained with the immediate previous ratio of time-points. To identify variables with HS, we considered non-decreasing  $\sigma_{\text{avg}}$ , which presented a  $\sigma_{\text{avg}}$  higher than 80 % of the  $\sigma_{\text{avg}}$  obtained with the immediate previous ratio of time-points. In total, we got a decreasing  $\sigma_{\text{avg}}$  in ninety-eight variables, 20 of which we illustrated in Supplemental Figure S3A. For variables with non-decreasing  $\sigma_{\text{avg}}$ , we obtained seven hundred and four variables, 20 of which we depict in Supplemental Figure S3B. The 98 variables associated with no HS correspond to 1982 genes (Supplementary Table S1), while the 704 variables related to HS correspond to 3292. Despite the majority of model variables being putatively

associated with HS, the variables without HS are, on average, associated with a higher number of genes. We speculate that variables related to an increased number of genes group genes that share regulatory mechanisms.

To identify the biological functions of genes corresponding to variables associated with or without HS, we performed a pathway enrichment analysis using the Molecular Signatures Database (Figure 3B). We observed that many pathways enriched in genes associated with variables without HS are also enriched in genes potentially associated with HS, including E2F target genes and genes downregulated in ultra-violet response. These pathways are also enriched when considering all differentially expressed genes in cells undergoing RAS-OIS with or without TF perturbation. We also find that genes corresponding to variables without HS contain a higher ratio of MYC targets when compared to all other genes. MYC is a master TF controlling a broad set of biological processes, including protein synthesis, metabolic homeostasis, transcription, cell proliferation, and tumorigenesis (Chen et al., 2018; Dang, 2012). By contrast, genes related to variables with hidden regulators are enriched for pathways, including TNF $\alpha$  signaling via NF $\kappa$ B, epithelial-to-mesenchymal transition, mitotic spindle, and genes upregulated in KRAS signaling.

To determine the accuracy of the inferred equations relative to genes without HS, i.e., variables displaying decreasing  $\sigma_{\text{avg}}$ , we simulated the model for the AP1-cJUN and RELA KD transcriptome validation datasets at 144h. At each iteration of the equation solver, the next transcriptional state of genes without HS was computed by the equations inferred from the SINDy algorithm. The gene expression values of variables connected to HS were retrieved from the interpolated experimental datasets. The simulation results for initial conditions corresponding to AP1-cJUN inhibition 144h after RAS-OIS induction are shown in Figures 3C and 3D. Figure 3C consists of two heatmaps depicting the simulation and experimental time profiles for the 98 genes with decreasing  $\sigma_{\text{avg}}$ , along with their correlation for each line. At the same time, Figure 3D shows a histogram with the correlation distribution of all variables. Fifty-nine variables (60 %) correlated higher than or equal to 0.9 with the experimental time courses, while 11 variables (11%) displayed a correlation lower than or equal to -0.9. The heatmaps and histogram for RELA KD at 144h after RAS-OIS induction are shown in Supplemental Figures S3C-D, where 30-time profiles (31 %) correlated higher than or equal to 0.9 and 48 profiles (49 %) correlated lower than or equal to -0.9.

For AP1-cJUN KD, HS analyses yielded a higher proportion of high correlations between experimental and simulated profiles than the number of profiles resulting in

negative correlations with high amplitude (Figures 2E and 3D). However, for RELA inhibition, the number of variables resulting in negative correlations between experimental and simulated results exceeds the number of variables yielding high correlations (Figures S2I and S3D). We hypothesize that this difference is due to distinct properties of both TFs regarding their role in senescence-associated gene regulation, with AP1-cJUN being a master regulator and RELA an effector. Besides, RELA is an NF $\kappa$ B family member that regulates biological pathways enriched in genes with possible hidden sources, including TNF $\alpha$  signaling and epithelial-to-mesenchymal transition (Liu et al., 2017). Therefore, the reduced performance of the RELA inhibition simulation is likely a direct consequence of the increased  $\sigma_{\text{avg}}$  observed in the coefficients describing the dynamics of genes involved both in its regulation and in downstream responses.

Considering that our input matrix contains every pairwise combination of variables, the number of coefficients to be inferred is considerably high. This property assures that the resulting coefficients will be sparse for the library of functions that make up the matrix  $\Theta(X)$ . However, it also requires substantial computing power when inferring the model, mainly when this process is performed several times, as performed during the hidden source analysis. To expedite our analysis, we used “Graphic Processing Unit (GPU)” cores, which perform several linear operations in a massively parallel manner. Supplemental Figure S4 displays the results of a benchmark with artificial data performed with four different algorithms. Despite presenting a lower efficiency for small input sizes, GPUs are orders of magnitude more efficient than CPUs for input matrices with a few million entries, and the Broyden–Fletcher–Goldfarb–Shannon (BFGS) algorithm implementation in TensorFlow (Abadi et al., 2015) is the most efficient of the benchmarked tools.

Altogether, our analysis identifies the presence and absence of hidden sources that impact gene expression in cells undergoing RAS-OIS, highlighting cell proliferation-related genes that are highly co-expressed in all measured conditions.

## Discussion

A thorough understanding of gene-regulatory networks (GRNs) is critical to reveal what drives cell fate decisions forward, thus, setting the stage for informed, targeted interventions. Cellular senescence (CS) is a cell fate that plays a vital role in physiology and pathophysiology and is induced by cellular stress, including oncogenic stress. In this work, we present a novel and original computational approach combining differential equations, sparse identification of nonlinear dynamics (SINDy), and hidden-source analyses to model the process of oncogene-induced senescence based on time-resolved transcriptome data.

A significant finding is that SINDy-inferred coefficients accurately reproduce our training data demonstrating that it can infer stable differential equations from the available data. Remarkably, the obtained equations predicted the effect of AP1-cJUN silencing at later time points in CS that were not included in the training dataset. At the same time, this was only partially the case for RELA silencing. As previously published (Martinez-Zamudio et al., 2020), AP1-cJUN is a master regulator of the senescent phenotype, whereas RELA is a downstream effector TF of limited transcriptional impact. Because the simulation had a lower performance for RELA implies that RELA-mediated transcript output depends on hidden factors not only represented in the transcriptomic layer. The latter also underscores the need to accurately leverage diverse experimental data sets to describe TF function and transcriptional impact accurately. Given that TF knock-down data came from only two-time points, most profiles are monotonic, depending on the direction of the interpolation curve on the day of treatment. The latter explains why the correlations for the TF depletion experiments had high amplitudes and were either positive or negative.

To evaluate the influence of transcription regulatory events not present in our datasets, we applied a hidden source (HS) analysis (Wang et al., 2016; Shen et al., 2014; Su et al., 2014). This procedure measures the variance of coefficients when portions of data with distinct sizes are used to infer them. Thus, variables displaying an increasing variance with more data are subject to regulators not represented in the dataset. In the scope of this study, those hidden regulators might correspond to transcriptional regulators including post-translational TF modifications, histones modifications, or even co-factors of epigenetic modifiers (Mitsis et al., 2020). Notably, we observe that variables without HS are associated with gene sets enriched for several cell cycle signaling pathways indicating that much of the senescence arrest response is transcriptionally regulated. Comparing the

simulation results with the outcomes obtained from the model validation, we observe a slight increase in performance for AP1-cJUN silencing and, for RELA, a decrease in the number of positive correlations. Although the cause of this result is currently unknown to us, we noticed that variables associated with hidden sources contain genes enriched for NFkB signaling pathways, which belong to the same TF family as RELA (Liu et al., 2017). Overall, HS analysis highlights variables with higher confidence in their coefficient estimates. Thus, the ratio of variables without HS provides a quantitative metric to evaluate the performance of GRNs modeled in future approaches. Furthermore, this analysis demonstrates the independence of each inferred equation and highlights the set of genes to be prioritized for further refinement. In principle, genes associated with variables without HS should display similar coefficients across distinct studies or show new terms from genes that did not change in previous analyses.

CS differential equation models currently available by and large focus on specific signaling pathways, including cell cycle arrest (Mombach et al., 2014), mitochondrial dysfunction (Dalle Pezze et al., 2014; Passos et al., 2010), and inflammation (Mothes et al., 2015). Given the role of CS in cellular proliferation, several models also depict population size dynamics, implicitly encapsulating gene expression and protein activity into intermediary states (Galvis et al., 2019; Guimera et al., 2017; Schäuble et al., 2012). One caveat of these models is that they account only for a few dozen variables. By contrast, we present the first genome-scale differential equation system describing CS to our knowledge. A similar approach was performed with yeast transcriptomic data (Hackett et al., 2020), generating differential equations from time-course data sets collected from the perturbation of 200 TFs. Our solution innovates in considering the combinatorial TF activity that governs gene expression regulation compared to the latter approach.

We consider our approach a building block for breakthroughs in the Systems Biology of CS. Since SINDy processing is computationally intensive, we implemented it by running a sparse regression algorithm in GPUs, reducing execution time by around one thousandfold compared to regular CPUs. This reduction can be further enhanced with the arrival of quantum computing, as algorithms performing sparse regression for quantum computers have already been developed (Gyongyosi and Imre, 2019; Li et al., 2018). Furthermore, as such technology evolves, we will extend the HS analysis to predict surrogate regulators' activity from their measured downstream targets. Specifically, this procedure is implemented by computing the so-called "cancellation factors" (Su et al., 2014). Given that those factors are specific for each pair of nodes, their computation for the entire network is a task that is not feasible to date. However, as the proportion of

genes without HS in the GRN increases, the search space of HS becomes more limited, and the identification of their dynamics becomes less costly.

GRN research is a promising path in the discovery of novel therapeutic targets. Malod-Dognin et al. (2019) integrated protein-protein interaction (PPI), gene co-expression, and gene interaction networks. Although not differentially expressed in cancer cells, they observed several genes that displayed a rewiring of their local network compared to healthy tissue and were associated with lower patient survival. Kim et al. (2019) identified 130 factors in cancer tissue that link cell type, drug sensitivity, and gene expression, providing a rich resource to predict drug-repurposing efficacy. We and others recently identified pioneer TF AP1-cJUN as a master regulator of both the senescence (Martinez-Zamudio et al., 2020) and aging (Lee et al., 2021) phenotypes through the analysis of the collective interactions governing gene expression. Avelar et al. (2020) found 13 novel targets to induce CS by inspecting PPI and gene co-expression networks. These targets presented a high correlation with cancer and longevity-associated genes. Finally, dynamics analyses have demonstrated how network connectivity and repair rate can lead to organismal death (Kogan et al., 2015) or to reversing diseased to healthy phenotypes (Lee et al., 2021; Cho et al., 2017). Our work employs a recent breakthrough in signal processing to formalize gene transcription in CS and lays out a robust and extensible mathematical framework to infer gene regulatory mechanisms quantitatively. As our computation power and algorithms evolve, the generated models will facilitate precision therapies to restore organismal homeostasis and promote health span.

## Limitations of the study

To compute the expression rate at each time-point accurately, we interpolated experimental data. This could be a potential source for artifacts for genes with rapidly changing time profiles. To address this issue, we will infer expression rates from single-cell RNA sequencing data (La Manno et al., 2018) in the next iteration of the model. A sufficiently comprehensive dataset will allow us to define one specific time profile for each gene, eliminating the need to map multiple genes to a single model variable and increasing model resolution even further than our current solution. Another limitation of our model is that it scores only differentially expressed genes in at least one sample in the training dataset. By combining distinct experimental conditions, including assays where potential genes of interest present at least one activated and one inactivated observation, our approach will predict the impact of their perturbation. We will also extend our model structure to include other layers of biological information, as described in a process termed Reactive-SINDy (Hoffmann et al., 2019).

## Author contributions

J.A.N.L.F.F. and O.B. designed the study and conceptual ideas, and wrote the manuscript. J.A.N.L.F.F. retrieved data from open repositories, processed microarray data, and inferred and simulated the model. O.B. secured funding.

## Acknowledgments

J.A.N.L.F.d.F. was supported by La Ligue Nationale Contre le Cancer, Association Nationale de la Recherche (ANR) and INSERM–AGEMED. O.B. was supported by CNRS and INSERM–AGEMED. We thank past and present lab members, especially Bertrand Friguet, for support and valuable discussions. 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.

## Declaration of interests,

The authors declare no competing interests.

## Figure Titles and Legends

**Figure 1: Generation and integration of time-resolved senescence transcriptomes for mathematical modeling.** [A] Schematic overview for defining gene-regulatory network (GRN) of RAS-OIS in WI38 fibroblasts using time-resolved and siRNA-based transcription factor (TF) (RELA, AP1-cJUN, and ETS1) knock-down (KD) transcriptome data sets at the indicated time-points as previously published (Martinez-Zamudio et al., 2020). KD samples were collected 48 hrs after siRNA transfection.

[B] Heatmaps showing modules of temporally co-expressed genes specific for RAS-OIS time-course in WI38 fibroblasts defined using unsupervised WGCNA clustering. Data are expressed as row Z-scores.

[C] Heatmap showing the integration of all modules identified in the transcriptomes of RAS-OIS time-course and siRNA-mediated KD of RELA, AP1-cJUN, and ETS1 at indicated time-points for RAS-OIS time-course in WI38 fibroblasts. Each line depicts the cluster assigned to each gene in each experiment, and genes belonging to the same module in all data sets are grouped into one variable in the model.

**Figure 2: Non-linear sparse inference of the cellular senescence transcriptomic program.** [A] Flowchart of the SINDy algorithm (Brunton et al., 2016). See text for details. [B] Model simulation with the training dataset. Heatmaps depict model-observed (left) and -simulated (right) time-resolved RAS-OIS gene expression profiles at indicated time points for each model variable. Correlation coefficients between observed and simulated gene expression profiles are shown as a bar-plot for each line in the graph at the far right of heatmaps. Data are expressed as row Z-scores. [C] Histogram depicting Pearson coefficient correlation distribution between model-observed and -simulated gene expression profiles of RAS-OIS transcriptome time-course for each model variable. [D] Model simulation with validation dataset. Comparison of observed (left) and -predicted (right) RAS-OIS gene expression profiles at indicated time-point for control and AP1-cJUN knock-down. Correlation coefficients between observed and simulated gene expression profiles are shown as a bar-plot for each line in the graph at the far right of heatmaps. Data



are expressed as row Z-scores. **[E]** Histogram depicting Pearson coefficient correlation distribution between observed and predicted time gene-expression profiles for (D).

**Figure 3: Hidden source (HS) analysis highlights variables with predictable**

**dynamics. [A]** Flowchart of HS analysis. The model inference was repeated for subsets with distinct fractions of the interpolated time courses. Variables with higher average coefficient variance as input quantum increases are considered to be regulated by HSs.

**[B]** Molecular Signatures Database pathways enriched for all differentially expressed (DE) genes detected in the control and TF knock-down (KD) RAS-OIS transcriptome datasets for genes associated with variables without HSs and genes related to variables impacted by factors not present in our datasets.

**[C]** RAS-OIS gene expression profiles of the 98 observed (left) and predicted (right) variables without HSs when validating the model with AP1-cJUN KD at indicated time-point. The correlation for each variable is shown in the bar plot on the right.

**[D]** Pearson coefficient correlation distribution between observed and predicted gene expression profiles following AP1-cJUN knock-down in RAS-OIS cells for the 98 variables without HSs.

## **STAR Methods**

### **Microarray transcriptome data download, preprocessing, statistical analysis, and annotation.**

The raw Affymetrix HTA 2.0 data was downloaded from the Gene Expression Omnibus database (BioProject PRJNA439263, accession numbers GSE112084 and GSE143248; Martínez-Zamudio et al., 2020) and 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 hta20stranscriptcluster.db R package (MacDonald, 2017) and internal control probes were removed. Genes with differential expression lower than 30 % compared to uninduced cells were filtered.

### **Hierarchical clustering and identification of unique expression time profiles**

The training datasets, consisting of the Ras-induced senescence time course (RAS OIS), the inhibition of AP1-cJUN and RELA at 72h, and the inhibition of ETS1 at 144h after RAS activation, were aggregated as shown in Figure 1A. The genes in each dataset were clustered independently with the WGCNA tool (Langfelder and Horvath, 2008), where the median of its replicates represented each sample. The parameters minimum cluster size, deepSplit, and threshold for merging clusters were set to 100, 3, and 0.85. The soft threshold parameter was determined for each dataset separately, with the choice of the lowest value leading to a high Scale-free topology fit by applying the elbow method. Genes sharing the same modules in all time courses were aggregated, constituting one variable in the model. After RAS activation, genes were identified to change at least 30 % compared to replicating WI38 fibroblasts, and 799 distinct expression profiles were identified, with 353 associated with single genes. Since the profiles containing AP1-cJUN, RELA, and ETS1 also enclosed other genes, one profile for each gene was added as an isolated variable, constituting a model with 802 variables.

### **Model inference**

To satisfy the RIP, the time course samples were scaled to zero mean and unit variance (Wang et al., 2016; Candes et al., 2006). To compute the derivative of each variable time profile, each time course was smoothed by interpolating it by a factor of 24, and the respective derivatives were 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 defining one variable in the model (Figure S2A). 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 for 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  $X^{P2}$ . As described in Brunton et al. (2016), the columns in matrix  $X^{P2}$  consist of the pairwise multiplication of each column in  $X$ , therefore characterizing a second-order model that describes the transcriptome dynamics. 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 the Results section, CS gene expression dynamics is defined by the matrix  $\Xi$ , which columns  $\xi_i$  correspond to the coefficients representing the influence of each term in variable  $i$ . Formally,

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

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  $\Xi$ , the coefficients describing the impact of the current transcriptomic state in gene expression variation for each variable.

Assuming  $\Xi$  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):

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

where

$$\|\xi_i\|_1 = \sum_j |\xi_{j,i}|, \quad j \in [1, M] \quad (4.3)$$

is the L1 norm of each column  $\xi_i$  of  $\Xi$  and  $M$ , the number of rows in  $\Xi$ , i.e., the number of interpolated time points.

Given the high number of coefficients to be inferred, the optimization was run in a GPU-based computational environment containing NVIDIA R 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.

### **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 the previous section. This expanded array was multiplied by the inferred  $\Xi$  matrix, yielding the variation in gene expression given a specific state.

### **Model assessment**

To assess the performance of 802 variables, we built histograms depicting the correlation between a prediction and its corresponding expected profile as performed by Bonneau et al. (2007).

### **Hidden sources analysis**

To assess how many and which genes present a dynamics that can be reliably inferred from the available data, we performed a hidden source (HS) analysis as described by Wang et al. (2016), Shen et al. (2014), and Su et al. (2014). The matrix  $\Theta(X)$ , containing 284 rows, was randomly sampled five times for four distinct data points ratio  $R_m$  (20 %, 40 %, 60 %, and 80 %), and the inference procedure described in section "Model inference" was run independently for each sub-sampled matrix. For each value of  $R_m$ , the coefficient variance was computed, and the average variance ( $\sigma_{avg}$ ) for all coefficients describing the dynamics of a single variable was assessed.

Variables with decreasing  $\sigma_{avg}$  with respect to  $R_m$  were considered to be regulated by genes with expression levels accessible in our datasets and therefore, be described by deducible equations (Wang et al., 2016; Shen et al., 2014; Su et al., 2014). To ensure a strictly decreasing  $\sigma_{avg}$ , we only considered variables where  $\sigma_{avg}$  for a specific  $R_m$  value is 80 % lower than  $\sigma_{avg}$  for the immediate lower  $R_m$  value. To validate the performance of the predicted model when considering only the variables without HS, we simulated the model with initial conditions corresponding to the validation datasets, i.e., the KD of AP1-cJUN and RELA in senescent cells at 144h after RAS induction. The simulations were performed as described in section "Model simulation," where the expression levels corresponding to

variables with non-decreasing  $\sigma_{\text{avg}}$  were retrieved from the experimental time courses. The performance of each simulation was assessed as described in section “Model assessment.”

### Pathway enrichment analysis

Aiming to identify signaling pathways enriched for genes associated with variables with and without HSs, we performed an over-representation analysis for both groups using the `msigdb` (Dolgalev, 2020) and `clusterProfiler` (Wu et al., 2021) R packages, combined with the Molecular Signature Database (MsigDB) hallmark gene sets (Subramanian et al., 2005; Liberzon et al., 2015). This analysis was also repeated for the set of all genes with at least a 30 % change in expression during all time courses.

### Inference tools benchmark

To evaluate the performance of distinct optimization methods on different platforms, we used the example provided by the `R1magic` R package (Suzen, 2015). An array  $y$  is computed from the product of a Gaussian Matrix  $\Phi$  and a sparse array  $x_0$ . The execution time for each tool to recover  $x_0$  was registered for five distinct combinations of  $\Phi$  and  $y$ . The recovery was accomplished by solving the following linear program:

$$\begin{aligned} & \text{minimize } \|x_0\|_1 & (4.6) \\ & \text{subject to } y = \Phi^* x_0 \end{aligned}$$

Where  $\Phi$  is a matrix with  $m$  rows and  $n > m$  columns,  $y$  is an array with  $m$  elements, and  $x_0$  is an array with  $k$  nonnull elements and length  $n$ . This procedure was repeated for four distinct sets of values,  $m$ ,  $n$ , and  $k$ , summarized in Table 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 1: Inference tools benchmark parameters

The evaluated optimization implementations were (a) the L1 regularization implemented by the `R1magic` R package (Suzen, 2015), (b) a BFGS implementation for 20

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).

### Data and code availability

All data produced and software implemented are available on Zenodo (<https://doi.org/10.5281/zenodo.6521071>). Available data includes every file required to reproduce the results in the manuscript, including the generated time-course matrices; their nonlinear expansions; model coefficients considering the whole dataset and subsets used for HS analysis; and simulation results. The software implemented is available as both Rmarkdown files and html reports. We also share a Singularity container that runs model inference in a High-Performance Computing facility.

Previously published transcriptome data are hosted in the GEO website under accession codes GSE112084 and GSE143248.

### Supplemental information titles and legends

**Figure S1: Generation and integration of time-resolved senescence transcriptomes for mathematical modeling.** [A] Simplified transcription factor (TF) hierarchy network (adapted from Martínez-Zamudio et al., 2020). Arrows depict chromatin binding sequentiality. For example, TF AP1-cJUN precedes the chromatin binding of TFs ETS1 and RELA in cells undergoing RAS-OIS. [B-D] Heatmaps showing modules of temporally co-expressed genes specific for RAS-OIS time-course and [B] AP1-cJUN, [C] RELA, and [D] ETS1 TF knock-down at indicated time-points in WI38 fibroblasts defined using unsupervised WGCNA clustering. Only genes showing a fold-change greater than 30% were considered. Data are expressed as row Z-scores. [E] Distribution of the number of genes grouped into each model variable.

**Figure S2: Simulation of transcription factor knock-down training and validation transcriptome datasets.** [A] Interpolated temporal profiles computed for each variable.

The top four lines depict the transcriptomic module computed for each training dataset, containing one element for each unique module combination. The following 284 lines display the interpolated temporal profiles computed for each variable, consisting of the median for the scaled expression values of its associated genes. **[B, D, F]** Model simulation with training datasets. Heatmaps depicting model-observed (left) and -simulated (right) RAS-OIS gene expression profiles for each model variable for control and **[B]** AP1-cJUN, **[D]** RELA, and **[F]** ETS1 knock-down (KD) at indicated time-points. Correlation coefficients between observed and simulated gene expression profiles are shown as a bar-plot for each line in the graph at the far right of heatmaps. Data are expressed as row Z-scores. **[C, E, G]** Pearson coefficient correlation distribution for training datasets B, D, and F. **[H]** Model simulation with validation dataset. Comparison of observed (left) and -predicted (right) RAS-OIS gene expression profiles at indicated time-point for control and RELA knock-down. Correlation coefficients between observed and simulated gene expression profiles are shown as a bar-plot for each line in the graph at the far right of heatmaps. Data are expressed as row Z-scores. **[I]** Histogram depicting Pearson coefficient correlation distribution between observed and predicted time gene-expression profiles for (H).

**Figure S3: Model coefficients variance analysis as a function of input size and validation of Hidden source analysis.** **[A, B]** Average coefficient variance calculated from five different model inferences based on four different time points sampled from the interpolated experimental time courses. **[A]** 20 examples of variables with coefficients showing decreasing average variance as the amount of training data increases and **[B]** 20 examples with non-decreasing average coefficient variance as a function of input data points. **[C]** RAS-OIS transcriptomic profiles of observed (left) and predicted (right) variables without HSs when validating the inferred model with RELA knock-down (KD) at indicated time-point and the respective correlation computed for each variable. **[D]** Pearson coefficient correlation distribution between observed and predicted RAS-OIS gene expression profiles following RELA KD at indicated time-point for variables without HSs.

**Figure S4: Graphics Processing Unit (GPU) computing reduces model inference required time.** Running time of four different tools to solve the underdetermined system  $y = \Phi x$  using the compressed sensing paradigm as a function of the size of the input matrix  $\Phi$ .

**Table S1: Model variables metadata.** For each DE gene identified, we provide metadata including microarray probe ID, Entrez ID, and gene symbol, in addition to the index of the variable in the model, the transcriptional module for each training dataset, the association of the variable to HS and the transcriptional module the same probe was associated to in our previous study (Martinez-Zamudio et al., 2020).

## References

Gorgoulis, V., Adams, P. D., Alimonti, A., Bennett, D. C., Bischof, O., Bishop, C., Campisi, J., Collado, M., Evangelou, K., Ferbeyre, G., et al. (2019). Cellular senescence: defining a path forward. *Cell*, 179(4), 813-827. 10.1016/j.cell.2019.10.005.

Martínez-Zamudio, R. I., Robinson, L., Roux, P. F., and Bischof, O. (2017a). SnapShot: cellular senescence pathways. *Cell*, 170(4), 816-816. 10.1016/j.cell.2017.07.049.

Martínez-Zamudio, R. I., Robinson, L., Roux, P. F., and Bischof, O. (2017b). SnapShot: cellular senescence in pathophysiology. *Cell*, 170(5), 1044-1044. 10.1016/j.cell.2017.08.025.

Martínez-Zamudio, R. I., Roux, P. F., de Freitas, J. A. N., Robinson, L., Doré, G., Sun, B., Belenki, D., Milanovic, M., Herbig, U., Schmitt, C. A., Gil, J. and Bischof, O. (2020). AP-1 imprints a reversible transcriptional program of senescent cells. *Nature cell biology*, 22(7), 842-855. 10.1038/s41556-020-0529-5.

Delgado, F. M., and Gómez-Vela, F. (2019). Computational methods for Gene Regulatory Networks reconstruction and analysis: A review. *Artificial intelligence in medicine*, 95, 133-145. 10.1016/j.artmed.2018.10.006.

Candes, E. J., Romberg, J. K., and Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(8), 1207-1223. 10.1002/cpa.20124.



Wang, W. X., Lai, Y. C., and Grebogi, C. (2016). Data based identification and prediction of nonlinear and complex dynamical systems. *Physics Reports*, 644, 1-76.  
10.1016/j.physrep.2016.06.004.

Broido, A. D., and Clauset, A. (2019). Scale-free networks are rare. *Nature communications*, 10, 1017. 10.1038/s41467-019-08746-5.

Ramirez, R. N., El-Ali, N. C., Mager, M. A., Wyman, D., Conesa, A., and Mortazavi, A. (2017). Dynamic gene regulatory networks of human myeloid differentiation. *Cell systems*, 4(4), 416-429. 10.1016/j.cels.2017.03.005.

Goode, D. K., Obier, N., Vijayabaskar, M. S., Lie-A-Ling, M., Lilly, A. J., Hannah, R., Lichtinger, M., Batta, K., Florkowska, M., Patel, R., et al. (2016). Dynamic gene regulatory networks drive hematopoietic specification and differentiation. *Developmental cell*, 36(5), 572-587. 10.1016/j.devcel.2016.01.024.

Chang, Y. H., Gray, J. W., and Tomlin, C. J. (2014). Exact reconstruction of gene regulatory networks using compressive sensing. *BMC bioinformatics*, 15, 400. 10.1186/s12859-014-0400-4.

Shen, Z., Wang, W. X., Fan, Y., Di, Z., and Lai, Y. C. (2014). Reconstructing propagation networks with natural diversity and identifying hidden sources. *Nature communications*, 5, 4323. 10.1038/ncomms5323.

Su, R. Q., Lai, Y. C., Wang, X., and Do, Y. (2014). Uncovering hidden nodes in complex networks in the presence of noise. *Scientific reports*, 4, 3944. 10.1038/srep03944.

Kaheman, K., Kutz, J. N., and Brunton, S. L. (2020). SINDy-PI: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics. *Proceedings of the Royal Society A*, 476, 20200279. 10.1098/rspa.2020.0279.

Hoffmann, M., Fröhner, C., and Noé, F. (2019). Reactive SINDy: Discovering governing reactions from concentration data. *The Journal of chemical physics*, 150, 025101. 10.1063/1.5066099.

- Mangan, N. M., Brunton, S. L., Proctor, J. L., and Kutz, J. N. (2016). Inferring biological networks by sparse identification of nonlinear dynamics. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 2(1), 52-63. 10.1109/TMBMC.2016.2633265.
- Brunton, S. L., Proctor, J. L., and Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15), 3932-3937. 10.1073/pnas.1517384113.
- Casadiello, J., Nitzan, M., Hallerberg, S., and Timme, M. (2017). Model-free inference of direct network interactions from nonlinear collective dynamics. *Nature communications*, 8, 2192. 10.1038/s41467-017-02288-4.
- Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 8(1), 118–127. 10.1093/biostatistics/kxj037.
- Langfelder, P. and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*, 9, 559. 10.1186/1471-2105-9-559.
- R Core Team (2016). *R: A language and environment for statistical computing*, Vienna, Austria. <https://www.R-project.org/>.
- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: A system for large-scale machine learning. *IProceedings of the USENIX Conference on Operating Systems Design and Implementation*, 16,265-283. 10.5281/zenodo.4724125.
- Chen, H., Liu, H., and Qing, G. (2018). Targeting oncogenic Myc as a strategy for cancer treatment. *Signal transduction and targeted therapy*, 3, 5. 10.1038/s41392-018-0008-7.
- Dang, C. V. (2012). MYC on the path to cancer. *Cell*, 149(1), 22-35. 10.1016/j.cell.2012.03.003.
- Liu, T., Zhang, L., Joo, D., and Sun, S. C. (2017). NF- $\kappa$ B signaling in inflammation. *Signal transduction and targeted therapy*, 2, 17023. 10.1038/sigtrans.2017.23.

Mitsis, T., Efthimiadou, A., Bacopoulou, F., Vlachakis, D., Chrousos, G. P., and Eliopoulos, E. (2020). Transcription factors and evolution: an integral part of gene expression. *World Academy of Sciences Journal*, 2(1), 3-8. 10.3892/wasj.2020.32

Mombach, J. C., Bugs, C. A., and Chaouiya, C. (2014). Modelling the onset of senescence at the G1/S cell cycle checkpoint. *BMC genomics*, 15(7), 1-11. 10.1186/1471-2164-15-S7-S7

Dalle Pezze, P., Nelson, G., Otten, E. G., Korolchuk, V. I., Kirkwood, T. B., von Zglinicki, T., and Shanley, D. P. (2014). Dynamic modelling of pathways to cellular senescence reveals strategies for targeted interventions. *PLoS Comput Biol*, 10(8), e1003728. 10.1371/journal.pcbi.1003728.

Passos, J. F., Nelson, G., Wang, C., Richter, T., Simillion, C., Proctor, C. J., Miwa, S., Olijslagers, S., Hallinan, J., Wipat, A., (2010). Feedback between p21 and reactive oxygen production is necessary for cell senescence. *Molecular systems biology*, 6(1), 347. 10.1038/msb.2010.5.

Mothes, J., Ipenberg, I., Arslan, S. Ç., Benary, U., Scheidereit, C., and Wolf, J. (2020). A quantitative modular modeling approach reveals the effects of different A20 feedback implementations for the NF- $\kappa$ B signaling dynamics. *Frontiers in physiology*, 11, 896. 10.3389/fphys.2020.00896.

Galvis, D., Walsh, D., Harries, L. W., Latorre, E., and Rankin, J. (2019). A dynamical systems model for the measurement of cellular senescence. *Journal of the Royal Society Interface*, 16(159), 20190311. 10.1098/rsif.2019.0311.

Guimera, A. M., Welsh, C., Dalle Pezze, P., Fullard, N., Nelson, G., Roger, M. F., Stefan, A. P. and Shanley, D. P. (2017). Systems modeling aging: from single senescent cells to simple multi-cellular models. *Essays in biochemistry*, 61(3), 369. 10.1042/EBC20160087.

Schäuble, S., Klement, K., Marthandan, S., Münch, S., Heiland, I., Schuster, S., Hemmerich, P., and Diekmann, S. (2012). Quantitative model of cell cycle arrest and

cellular senescence in primary human fibroblasts. PLoS ONE, 7(8), e42150.  
10.1371/journal.pone.0042150.

Hackett, S. R., Baltz, E. A., Coram, M., Wrantik, B. J., Kim, G., Baker, A., Fan, M., Hendrickson, D. G., Berndl, M., and McIsaac, R. S. (2020). Learning causal networks using inducible transcription factors and transcriptome-wide time series. *Molecular systems biology*, 16(3), e9174. 10.15252/msb.20199174.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastrioti, M. E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. *Nature* 560, 494-498. 10.1038/s41586-018-0414-6.

Hoffmann, M., Fröhner, C., and Noé, F. (2019). Reactive SINDy: Discovering governing reactions from concentration data. *The Journal of chemical physics*, 150, 025101. 10.1063/1.5066099.

Gyongyosi, L., and Imre, S. (2019). Dense quantum measurement theory. *Scientific reports*, 9, 6755. 10.1038/s41598-019-43250-2.

Li, R. Y., Di Felice, R., Rohs, R., and Lidar, D. A. (2018). Quantum annealing versus classical machine learning applied to a simplified computational biology problem. *NPJ quantum information*, 4, 14. 10.1038/s41534-018-0060-8.

Malod-Dognin, N., Petschnigg, J., Windels, S. F., Povh, J., Hemingway, H., Ketteler, R., and Pržulj, N. (2019). Towards a data-integrated cell. *Nature communications*, 10, 805. 10.1038/s41467-019-08797-8

Kim, Y., Bismeyer, T., Zwart, W., Wessels, L. F., and Vis, D. J. (2019). Genomic data integration by WON-PARAFAC identifies interpretable factors for predicting drug-sensitivity in vivo. *Nature communications*, 10, 5034. 10.1038/s41467-019-13027-2.

Lee, H. Y., Jeon, Y., Kim, Y. K., Jang, J. Y., Cho, Y. S., Bhak, J., and Cho, K. H. (2021). Identifying molecular targets for reverse aging using integrated network analysis of transcriptomic and epigenomic changes during aging. *Scientific reports*, 11, 12317. 10.1038/s41598-021-91811-1.

Avelar, R. A., Ortega, J. G., Tacutu, R., Tyler, E. J., Bennett, D., Binetti, P., Budovsky, A., Chatsirisupachai, K., Johnson, E., Murray, A., et al. (2020). A multidimensional systems biology analysis of cellular senescence in aging and disease. *Genome Biology*, 21, 91. 10.1186/s13059-020-01990-9.

Kogan, V., Molodtsov, I., Menshikov, L. I., Reis, R. J. S., and Fedichev, P. (2015). Stability analysis of a model gene network links aging, stress resistance, and negligible senescence. *Scientific reports*, 5, 13589. 10.1038/srep13589

Cho, K. H., Lee, S., Kim, D., Shin, D., Joo, J. I., and Park, S. M. (2017). Cancer reversion, a renewed challenge in systems biology. *Current Opinion in Systems Biology*, 2, 49-58. 10.1016/j.coisb.2017.01.005.

Carvalho, B. S. and Irizarry, R. A. (2010). A framework for oligonucleotide microarray preprocessing. *Bioinformatics* 26(19), 2363–2367. 10.1093/bioinformatics/btq431.

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

Voss, T. C. and Hager, G. L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nature Reviews Genetics* 15, 69–81. 10.1038/nrg3623.

Garber, M., Yosef, N., Goren, A., Raychowdhury, R., Thielke, A., Guttman, M., Robinson, J., Minie, B., Chevrier, N., Itzhaki, Z., et al. (2012). A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals. *Molecular cell* 47(5), 810–822. 10.1016/j.molcel.2012.07.030.

Jones, E., Oliphant, T., Peterson, P., et al. (2001). SciPy: Open source scientific tools for Python. <http://www.scipy.org/>.

Soetaert, K., Petzoldt, T., and Setzer, R. W. (2010). Solving differential equations in R: package deSolve. *Journal of statistical software* 33, 1-25. 10.18637/jss.v033.i09.

Bonneau, R., Facciotti, M. T., Reiss, D. J., Schmid, A. K., Pan, M., Kaur, A., Thorsson, V., Shannon, P., Johnson, M. H., Bare, J. C., et al. (2007). A predictive model for transcriptional control of physiology in a free living cell. *Cell* 131(7), 1354–1365. 10.1016/j.cell.2007.10.053.

Dolgalev, I. (2020). msigdb: MSigDB gene sets for multiple organisms in a tidy data format. R package version, 7.1. <https://igordot.github.io/msigdb/>

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, 2(3), 100141. 10.1016/j.xinn.2021.100141.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43), 15545-15550. 10.1073/pnas.0506580102.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The molecular signatures database hallmark gene set collection. *Cell systems*, 1(6), 417-425. 10.1016/j.cels.2015.12.004.

Suzen, M. (2015). R1magic: Compressive Sampling: Sparse Signal Recovery Utilities. R package version 0.3.2. <https://CRAN.R-project.org/package=R1magic>.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature methods*, 17(3), 261-272. 10.1038/s41592-019-0686-2.