MYC dosage compensation is mediated by miRNA-transcription factor interactions in aneuploid cancer

Acón, ManSai[1,2]. Geiß, Carsten[7]. Torres-Calvo, Jorge[1,2]. Oviedo, Guillermo[1,2]. Arias-Arias, Jorge L[1]. Vásquez-Vargas, Gloriana[1]. Oses-Vargas, Yendry[1]. Guevara-Coto, José[3]. Segura-Castillo, Andrés[6]. Siles-Canales, Francisco[4,5]. Quirós-Barrantes, Steve[1,5]. Régnier-Vigouroux, Anne[7]. Mendes, Pedro[8]. Mora-Rodríguez, Rodrigo[1,2,5,*].

* Corresponding author.

1. Lab of Tumor Chemosensitivity (LQT), Research Center for Tropical Diseases (CIET), Faculty of Microbiology, University of Costa Rica, 11501-2060 San José, Costa Rica.
2. Master Program on Bioinformatics and Systems Biology, Posgraduate Program SEP, University of Costa Rica, 11501-2060 San José, Costa Rica.
3. School of Computer Sciences and Informatics (ECCI), University of Costa Rica, San Jose Costa Rica, 11501-2060 San José, Costa Rica.
4. Pattern Recognition and Intelligent Systems Laboratory, Department of Electrical Engineering, Universidad de Costa Rica, 11501-2060 San José, Costa Rica.
5. DC Lab, Lab of Surgery and Cancer, University of Costa Rica, 11501-2060 San José, Costa Rica.
6. Laboratorio de Investigación e Innovación Tecnológica, Universidad Estatal a Distancia (UNED), 474-2050 San José, Costa Rica
7. Institute for Developmental Biology and Neurobiology, Johannes Gutenberg University, 55128 Mainz, Germany
8. Center for Cell Analysis and Modeling and Department of Cell Biology, University of Connecticut School of Medicine, Farmington, 06030 Connecticut, USA

Running title: "Modeling cancer gene dosage compensation"

Abstract— Cancer complexity is consequence of genomic instability leading to aneuploidy. We hypothesize that dosage compensation of critical genes arise from systems-level properties for cancer cells to withstand the negative effects of aneuploidy. We developed a computational platform to identify a network of miRNAs and transcription factors interacting with candidate dosage-compensated genes using NCI-60 multi-omic data. We next constructed a mathematical model where the property of dosage compensation emerged for *MYC* and *STAT3* and was dependent on the kinetic parameters of their feedback and feed-forward interactions with four miRNAs. We developed a genetic tug-of-war approach by overexpressing an exogenous *MYC* sequence to experimentally validate *MYC* dosage compensation circuits as demonstrated by the over-expression of the three microRNAs involved and the respective down-regulation of endogenous *MYC*. In addition, *MYC* overexpression or inhibition of its compensating miRNAs led to dosage-dependent cytotoxicity in *MYC*-amplified colon cancer cells. The study of TCGA breast cancer patient data indicated that *MYC* dosage compensation could lead to lower patient survival, highlighting the potential of targeting gene dosage compensation to prevent aneuploid cancer progression.

(BioNetUCR, available here: https://cloud.prislab.org/s/gt2W2jfZQx3E3Jm).

Keywords—miRNAs/gene dosage compensation/ cancer/NCI60/systems biology

Introduction

Aneuploidy is a genomic alteration characterized by gains and losses of whole or partial chromosomes (Hanahan & Weinberg, 2011). This is lethal for normal cells and entire organisms, representing the most frequent cause of abortions and mental retardation in humans (Hardy & Hardy, 2015), (Kojima & Cimini, 2019), (Sheltzer & Amon, 2011). Aneuploidy is also a hallmark of cancer: 90% of solid tumors and 75% of hematologic tumors present aneuploidy, which was considered by early reports as the driving force of cancer genomic evolution and seems to have a role in oncogenesis because: i) it appears before malignant transformation, ii) there are clonal genetic alterations that compromise the fidelity of chromosomal segregation and iii) there are inherited mutations in checkpoints leading to aneuploidy and predisposing to cancer (Sheltzer & Amon, 2011). However, aneuploidy is also the major source of autocatalytic genomic instability, where the cells with the highest aneuploidy have also the highest instability of their genomes (Duesberg, Rausch, Rasnick, & Hehlmann, 1998). This leads to many defects at the cellular level: decreases cell proliferation and viability, increases proteotoxic stress, metabolic requirements, lactate production and induces recombination defects. The most likely explanation for all those negative effects of aneuploidy is given by the alteration of the gene dosage: gains or losses of entire chromosomes immediately alter the doses of hundreds of genes in the cell, leading to unbalanced load of critical proteins, altering energetic requirements and protein homeostasis (Sheltzer & Amon, 2011).

It is therefore unknown how cancer cells deal with so much aneuploidy whereas normal cells are very sensitive. A hint about the nature of this mechanisms developed by cancer cells to minimize the negative effects of aneuploidy is provided by the fact that hyperdiploidy (higher than normal chromosome number) is more frequently observed than hypodiploidy (lower than normal chromosome number) (Weaver & Cleveland, 2006)(Cimini, 2008). Thus, in most cases in which aneuploidy occurs, cells will possess a number of chromosomes that is above the norm. A possible explanation is given by the hypothesis of gene dosage compensation, a mechanism by which the expression of certain genes is modulated to compensate for differences in gene dosage when extra chromosomes are present due to aneuploidy (Kojima & Cimini, 2019)

Gene dosage compensation has been described very early for other organisms to compensate the negative effects of aneuploidy (Devlin, Holm, & Grigliatti, 1982). Indeed, the concept of gene dosage compensation, or gene dosage balance, is a widespread phenomenon that was discovered in the early days of genetics and there is accumulating evidence that it has an effect on gene expression, quantitative traits, aneuploid syndromes, population dynamics of copy number variants and differential evolutionary fate of genes after partial or whole-genome duplication (for review see (Birchler & Veitia, 2012)).

There are several potential mechanisms of dosage compensation at the many levels of the path that leads from a gene to a final product. At the protein synthesis level, the effects of gene dosage compensation are hypothesized to result from stoichiometric differences among members of macromolecular complexes, the interactome, and signaling pathways (Birchler & Veitia, 2012; Veitia, Bottani, & Birchler, 2008). Gene dosage compensation represents a compensatory mechanism that may ameliorate the imbalanced gene expression and restore protein homeostasis in aneuploid cells. The multiple consequences of aneuploidy are mainly regulated at the protein level due to a side-effect of protein folding defects and increased protein degradation by proteosome and autophagy (Donnelly & Storchová, 2014). An approach to identify genes with dosage compensation by increasing the copy number of individual genes using the genetic tug-of-war technique showed that approximately 10% of the genome shows gene dosage compensation, and consists predominantly of subunits of multi-protein complexes, which are regulated in a stoichiometry-dependent manner (Ishikawa, Makanae, Iwasaki, Ingolia, & Moriya, 2017), although this approach was designed to identify compensated genes at the protein level only. In aneuploid cancer it has been shown that messenger RNA (mRNA) levels generally correlate well with an increased DNA copy number (gene dosage) but these changes are not reflected at the protein level for several genes (Stingele et al., 2012). Indeed, gene dosage compensation in cancer cells has been recently (Brennan et al., 2019) demonstrated at the protein level, whereby protein aggregation mediates the stoichiometry of protein complexes in aneuploid cells. In that study they show that excess subunits are either degraded or aggregated and that protein aggregation is nearly as effective as protein degradation at lowering the level of functional proteins (Brennan et al., 2019) .

The regulation of gene transcription might be another effective mechanism to compensate the gene dosage changes in aneuploid cells, as it maintains the stoichiometry and preserves the energy required for transcription, translation, and eventual degradation of the extra proteins (Donnelly & Storchová, 2014). At this level, several experimental models have been used to investigate the effects of aneuploidy on gene expression with contradicting findings. Many studies have found that acquisition of an extra chromosome results in a proportional increase in the expression of genes carried on that chromosome. Other studies have reported that, in addition to an increase in the expression of genes carried by the extra chromosome, aneuploidy can also perturb the expression levels of genes on other chromosomes. A number of studies have also reported some kind of dosage compensation, by which the genes on the aneuploid chromosome are not expressed at the expected levels based on gene dosage but this has been reported to occur for variable numbers of genes depending on context. Moreover, the degree of compensation may vary and it may occur simultaneously with effects on other chromosomes (for review see (Kojima & Cimini, 2019)).

miRNAs are small endogenous RNA molecules that bind mRNAs and repress gene expression (Fabian, Sonenberg, & Filipowicz, 2010). miRNAs are one of the most predominantly represented non-coding RNA (ncRNA) groups in clinical research. A typical miRNA is processed from a long primary RNA sequence to a short mature functional transcript around 22 nucleotides in length. A common characteristic of a miRNA is its ability

to pleiotropically target potentially hundreds or even thousands of genes (Hanna, Hossain, & Kocerha, 2019) and their target genes can also be regulated by several miRNAs (Ritchie, Rasko, & Flamant, 2013). Indeed, current estimates point to the human genome containing 1917 annotated hairpin precursors, and 2654 mature sequences of miRNAs (Kozomara, Birgaoanu, & Griffiths-Jones, 2019), estimated to directly regulate >60% of human mRNAs (Kim et al., 2016). In consequence, there is a good possibility that miRNA-transcription factor interactions may regulate the expression of genes amplified or deleted in cancer. Therefore, we hypothesize that gene dosage compensation in cancer can be mediated, at least in part, by the emerging properties of complex miRNA-TF networks, controlling the expression of genes that have altered copy number.

In the present work we report the transcriptional gene dosage compensation of *MYC*, *STAT3*, *STAT5A* and *STAT5B* mediated by an emergent property of feedback and feedforward loops. The interactions forming these network motifs have a specific set of kinetic parameters enabling compensation. We experimentally validated the compensation circuits for *MYC* and determined that this mechanism can be blocked by depleting key miRNAs, affecting in a larger extent those cells with higher copy numbers of those compensated genes. The study of *MYC* dosage compensation in breast cancer patients showed heterogeneity regarding to which regulatory motifs mediate this regulation. It also suggests that *MYC* dosage compensation could lead to lower patient survival, highlighting the potential of targeting gene dosage compensation against aneuploid cancer.

We propose the existence of a regulatory network mediated by miRNAs that compensates for gene dosage changes in aneuploid cancer cells. We suggest that the manipulation of specific nodes of this miRNA-based regulatory network could block gene dosage compensation, representing a novel type of specific targeting of aneuploid cancer cells.

Results

Candidate genes with low tolerance to variation in their RNA expression despite high copy number variation are present across the cancer genomes of the NCI-60 panel

To identify genes under putative dosage compensation, we compared copy number, gene expression and proteomic data of all genes in the NCI60 panel, a collection of 59 cell lines of different tumor types (Shoemaker, 2006). We considered input data including high resolution data of Copy Number Variation (CNV) of the NCI-60 Cancer Cell lines from 4 different platforms (Bussey et al., 2006), the Gene Transcript (RNA) Average Intensities of 5 Platforms (Gmeiner, Reinhold, & Pommier, 2010), and the protein levels (Protein) of a global proteome analysis of the NCI-60 cell line panel (Gholami et al., 2013). Figure 1A-left shows the variation of the absolute values of DNA copy number, RNA expression and Protein expression. Next, we calculated the average RNA or protein expression of the diploid cell

lines (copy number between -0.25 and 0.25) for each gene to obtain the average diploid expression for every gene. Afterwards, we normalized all expression data against this diploid average expression for each gene and calculated log2 transformed values. Using this approach, we could directly and simultaneously compare the changes in DNA Copy Number, RNA and Protein expression across the NCI60 panel to look for genes with high variations in DNA copy numbers but a low tolerance to variation in RNA or protein expression as they could represent candidate targets under dosage compensation (Figure 1A). We plotted the Standard Deviation (SD) of the DNA, RNA and Protein values across the 59 cell lines and observed that several dots separated from the main gene population as they present high SD of DNA levels (Figure 1B).

To obtain a good separation of candidate compensated genes from the main cluster, we implemented a Gaussian mixture model (GMM) with these data (Figure 1C). After classification, we could observe one group of genes with high SD of DNA copies and relatively low SD of RNA and/or protein expression values (Figure 1D). This cluster showed the expected behavior with high amplitudes in DNA values but very low variation in RNA or protein expression (Figure 1D and Figure 1E). For instance, *MYC* presents high frequency of amplifications in the NCI60 panel without the corresponding increase in RNA levels (Figure 1F left). A similar behavior is observed for *Rab5C* for both RNA and protein levels despite its copy number variation (Figure 1F middle), compared to a gene such as *Sav1* where copy number and RNA expression are well correlated (Figure 1F right). Next, we discarded those genes with orthologues in X/Y chromosomes since they cannot be differentiated using microarray techniques, and obtained a list of 56 gene candidates with putative dosage compensation at the transcript level. In some cases there was also protein data confirming this behavior but there was no candidate with putative compensation at the protein level only. Moreover, we reduced our list of candidates by discarding genes with high DNA variation due mostly to deletion (more than 6 cell lines with compensated deletions) and kept only genes with at least 6 cell lines with compensated amplifications. For example, *MYC* presents 27 compensated amplifications within the NCI60 with very similar expression levels even with 8 genetic copies. This approach reduced our list to 21 candidate genes with dosage amplifications (Expanded View Figure 1) potentially compensated at the transcript level across the NCI60 panel.

We next aimed to correlate the behavior of the copy number variations among these candidate genes to their corresponding chromosomal locations across the NCI60 panel. The clustergram in Figure 1G shows different clusters suggesting a similar behavior of copy number variations for these candidate genes. As expected, many of these genes are correlated by their genomic locations, especially gene clusters belonging to the chromosomal bands 3q26, 6p25, 8q24, 9p21-22, 10q22-23, 11q22-24, 17q11 and 17q21-23. These data suggests the existence of genes with low tolerance to variation in their RNA expression despite high copy number variation across the NCI60 panel, partially associated by common chromosomal locations.

A mathematical model of gene dosage compensation mediated by a network of miRNA-transcription factor interactions for the NCI60 panel

We hypothesized, at this point, that the inhibition of a dosage compensation mechanism would release the brake imposed on the expression of all these additional copies leading to significant overexpression of these genes. If the dosage compensation of these genes was favored during cancer evolution to restrain the expression of key lethal genes, its overexpression could potentially lead to death of the cell lines with these specific amplifications. Therefore, we focused our next steps on the description of possible mechanisms involved in dosage compensation of amplified genes identified in the NCI60 panel.

We asked whether dosage compensation mechanism could be mediated by systems-level properties arising from a complex regulatory network of gene expression. Since there is expression data available of miRNAs (Blower et al., 2007) and transcription factors (TF) (Gmeiner et al., 2010) for the NCI60 panel, we performed a correlation analysis of the copy numbers of target genes with the expression levels of miRNAs and transcription factors in order to identify possible regulators responding to the dosage of our candidate genes (Expanded View Figure 2). There are both miRNAs and TF with positive or negative correlation to the candidate copy numbers (Expanded View Figure 2A), although most of those correlations do not correspond to the direct reported interactions (Expanded View Figure 2B). Using all those interactions we constructed a network of putative and validated regulatory interactions connecting all candidate genes (Expanded View Figure 2C). However, from the construction of this network topology, it is easy to infer that this network could sense changes in copy number of target genes only if they also have a TF-function (red colored interactions, Expanded View figure 2C). Only a gene candidate with TF function will trigger a signal that can be propagated throughout the network and back if this gene has a regulatory loop. This loop would compensate gene expression in response to changes in gene dosage.

The regulatory motifs with potential systems-level properties to mediate dosage compensation are widely present within this putative network (Expanded View section). Nevertheless, a gene-dosage response mediated by one or several of these regulatory loops will depend on the strength of the interactions (i.e. their kinetic parameter values). Due to the high complexity of this miRNA/TF regulatory network, we envisaged the construction of a large scale mathematical model in order to gain insight into a possible mechanism of gene dosage compensation mediated by any of the many regulatory loops identified in this network. A quantitative systems-level approach for the calculation of the goodness of fit of a mathematical model describing the interactions enabled us to assess the feasibility of a proposed network topology for gene dosage compensation.

First, we reduced the network complexity to include only the NCI60 target genes with TF-

function as they are the only genes for which a copy number variation can be sensed (*MYC*, *Znf217, STAT3, Foxc1, Foxf2, Pgr, STAT5A, STAT5B, Pgr*). This resulted in a network of 517 nodes and 44016 arcs of putative and validated interactions. Since we collected about 65,000 experimentally validated interactions out of HTRI, Pazar, Transmir and Mirtarbase, we decided to include only experimentally validated interactions to further reduce model complexity. This led to a network of 78 nodes and 578 arcs of experimentally validated interactions. After this simplification, several target genes became dead ends, since no validated interactions are reported influencing other nodes of the network. Therefore, we removed them, further reducing the model to include only 4 target genes with TF-function and validated interactions influencing other nodes of the network: *MYC*, *STAT3*, *STAT5A* and *STAT5B*. This led to a network of 65 nodes and 506 arcs. The nodes include 45 TFs and 20 miRNAs. The arcs formed 16 feedback loops, 28 coherent feedforward loops and 45 incoherent feedforward loops (Figure 2A).

Second, we built an automated platform to generate ordinary differential equation (ODE) mathematical models for COPASI based on our candidate network topology (BioNetUCR, see materials and methods). To do so, we introduced ODE expressions for the schematic representations of the modeled interactions for the target gene, miRNAs and associated TFs (Figure 2B) and translated them into differential equations. The biochemical model includes 65 species, 130 reactions and 305 parameters. Reactions were represented using simple kinetic functions. Most follow mass action kinetics, except for transcription of genes that are induced or repressed by transcription factors, where linear activation or inhibition terms are added, as well as the degradation of RNA species that are affected by miRNAs, where linear inhibition terms were included. Gene copy numbers are also included as part of the mass action rate constant. The model parameters were estimated to fit the model to the data using the Parameter Estimation function of COPASI with the Hooke-Jeeves algorithm of optimization (Hooke & Jeeves, 1961) under the assumption that each cell line of the NCI-60 panel represents a different steady state of this model depending only on the copy numbers of the genes in the network. We increased the weight of the objective function of the expression of the 4 target genes in order to favor the search of model parameter values leading to dosage compensation. Model refinement was performed by interrupting the fitting process after several days without any significant progress. At these periods of time, we calculated the goodness-of-fit for each species and performed sensitivity analysis using the metabolic control analysis framework of COPASI (Figure 2F). We calculated the p-values (t-Student to compare simulated with experimental data) for the goodness-of-fit for the first resulting model, which presents a good fit for most species except for miR106a ($p = 0.012$). After removing miR106a, the resulting model achieved a lower objective function value (better fit) but miR145 had the worse goodness-of-fit ($p=0.075$) without significant control on the target genes. We decided to remove also miR145 from the model and continued the parameter estimation to obtain an optimized model with a lower objective function (figure 2C). For this last fit, there was no statistical difference of the values of the simulated species compared to the experimental data ($p>0.1$, figure 2D), obtaining thereby a mathematical model describing the experimental data of our target candidate and their associated miRNAs and TFs.

Furthermore, we asked whether the resulting model is able to compensate the gene dosage of our candidate target genes in response to an increase in gene copy numbers. In order to ascertain this, we performed a parameter scan using COPASI for the copy number of the 4 target genes individually. As shown in figure 3E, increasing the copy number of *MYC* from 1 to 5 leads only to an increase of 1.95 fold in *MYC* expression. For *STAT3*, an increase from 1 to 5 in copy number leads to an increase of 3.19 fold in *STAT3* expression. This was not the case for *STAT5A* and *STAT5B*, which copy number increases were not compensated in the model. We next performed a metabolic control analysis using COPASI to identify candidate miRNAs or TFs with high positive or negative regulatory control on the expression level of the 4 target genes. We plotted the 6 miRNAs with the highest negative control and the 6 TFs with the highest positive control (figure 2F). The results showed that *MYC* levels are mainly controlled by an interplay between miR-17, miR20a, miR19a and *BRCA1* and *NF-KB1*. *STAT3* levels are mainly controlled by miR-17, miR-20a, miR19a and *BRCA1*, *NF-KB1*, *Ets1* and *ATF1*. Although *STAT5A* and *STAT5B* were not compensated upon changes in their copy numbers, they also present control in their expression levels by some of the same miRNAs and TFs, suggesting that the mechanism of *MYC* and *STAT3* compensation could also impact the levels of *STAT5A* and *STAT5B*. These observations demonstrate that our mathematical model is able to describe a theoretical network of validated interactions between miRNAs and TFs with systems-level properties mediating direct dosage compensation of *MYC* and *STAT3*, with a possible impact on *STAT5A* and *STAT5B*. If this model is correct, the inhibition of some miRNAs negatively controlling the expression levels of the target genes could potentially block gene dosage compensation.

A minimal model of MYC and STAT3 dosage compensation leads to the identification of a putative mechanism of gene-dosage compensation in aneuploid cancer

We next asked which are the minimal components of the network to identify the exact regulatory loops mediating gene dosage compensation and to design strategies to interfere with this mechanism *in silico*. First, we sought to identify the essential species involved in the feedback or feedforward loops controlling gene dosage compensation. We performed parameter scans changing the copy number of the compensated genes (*MYC* and *STAT3*) and observed the behavior of the concentrations of the miRNAs and TFs of the model. For *MYC* copy number scan, the *MYC* expression showed a compensated increase, observed also for several TFs and miRNAs (Figure 3A left), including those identified by metabolic control analysis (miR-19a, miR-20a, miR-17, Figure 2F) but others, such as *STAT3* and miR21, decreased with increasing *MYC* copy numbers (Figure 3A left). For *STAT3*, increasing gene copy numbers led to a compensated increase of *STAT3*, several miRNAs including miR21, miR17, miR221, miR19a and miR20a and other TFs but also a decrease in *STAT5A*, *Ets1* and *Pparg* (Figure 3A right).

To assess the role in dosage compensation of these miRNAs and TFs responding the changes in *MYC* and *STAT3* copy numbers, we performed single *in silico* inhibitions of the

corresponding nodes by setting their interaction (inhibition/activation) parameters to 0. The inhibitions of single species altered the steady state levels of expression of *MYC* and *STAT3* for the basal conditions (copy number of 1) but this effect is not shown due to the normalization of the expression, to better appreciate the effect on gene dosage compensation (figure 3B). The parameter scans under these conditions showed that *MYC* and *STAT3* compensations are completely abolished when their respective interaction parameters are set to 0, confirming that their TF function is essential for the network to sense the changes in copy number (*MYC*(-) and *STAT3*(-) conditions, figure 3B). Also, *MYC* compensation is slightly altered when *STAT3* is inhibited (*STAT3*(-)) and *STAT3* compensation is completely abolished when *MYC* is inhibited (*MYC*(-)), suggesting an interplay in the mechanisms of gene dosage compensation of these two genes. In contrast, the single inhibition for all remaining TFs and miRNAs had little or no effect on *MYC* and *STAT3* dosage compensation, suggesting that this mechanism could be redundantly mediated by more than one miRNA or TF.

Therefore, we performed another *in silico* experiment changing the copy number of *MYC* and *STAT3* while all miRNAs or TFs were inhibited, except by *MYC* and *STAT3* themselves (Figure 3C). The results showed that the inhibition of all the other TFs (TF (-)) had no effect on the dosage compensation of *MYC* and *STAT3*. However, the inhibition of all miRNAs (MiR(-)) completely abolished the dosage compensation for both genes and led to a huge increase in their expression. We then proceeded to restore some miRNAs individually back into the MiR inhibited model, for example (MiR(-) + miR17) to simulate the effect of the active miR17 alone. Surprisingly, more than one miRNA was able to restore dosage compensation for both *MYC* and *STAT3*, confirming thereby that this mechanism is redundant. Indeed, miR-17, miR-19a and miR20a were individually able to compensate *MYC* dosage and miR17, miR20a and miR21 were able restore *STAT3* dosage compensation (Figure 3C). Additionally, to confirm the role of these redundant miRNAs on gene dosage compensation in silico, we performed their triple inhibition in the complete model. Indeed, the triple inhibition (miR17, miR19a and miR20a (-)) was able to block *MYC* dosage compensation (figure 3C left), whereas (miR17, miR20a, miR21 (-)) abolished *STAT3* dosage compensation (Figure 3C right), confirming that *MYC* and *STAT3* dosage compensation is mediated by a redundant mechanism.

Based on these results of redundant compensation and the interplay between *MYC* and *STAT3*, we proposed a minimal network topology of gene dosage compensation for *MYC* and *STAT3* (Figure 3D). After reconstructing the interactions from the original network, the minimal network states that *MYC* is compensated by 3 redundant negative feedback loops formed with miR17, miR19a and miR20a. The compensation of *STAT3* is mediated by 1 feedback loop with miR21 and 2 feedforward loops: (*STAT3*-*MYC*-miR17-*STAT3*) and (*STAT3*-*MYC*-miR20a-*STAT3*). From this minimal network topology, we constructed a minimal mathematical model of gene dosage compensation for *MYC* and *STAT3*. After fitting this minimal model to the experimental data and repeating the parameter scan experiments for gene copy numbers of *MYC* and *STAT3*, we reconstituted the same dosage compensation

mechanism observed in the original large model of interactions (Figure 3E).

These results indicate that we identified *in silico* a minimal model of gene dosage compensation for *MYC* and *STAT3* mediated by their interactions with miR17, miR19a, miR20a and miR21.

A genetic tug-of-war approach enabled the experimental demonstration of MYC dosage compensation

In order to experimentally demonstrate that *MYC* is dosage-compensated in cancer cells, we designed a tug-of-war approach inspired on the approach of Ishikawa et al for the identification of post-traslational dosage compensation (Ishikawa et al., 2017). In this report, the authors inserted a TAP tag next to the endogenous target gene to be evaluated. When the protein is over-expressed by transfecting a plasmid with an exogenous version of the gene, they observed downregulation of the endogenous protein to demonstrate dosage compensation at the translational level. To adapt this concept for the identification of gene dosage compensation at the transcriptional level, we designed an experimental approach to overexpress an exogeneous version of *MYC* (exo-*MYC*) with a differential codon optimization compared to the human endogenous *MYC* (endo-*MYC*). In addition, we removed the 3'UTR segment of the exogenous *MYC* to avoid any miRNA regulation on this transcript. Although both transcripts will be translated into the same protein, they can be differentiated at the transcript level by RT-qPCR. Under the pressure of the dosage compensation circuits, we hypothesize that the exogenous *MYC* overexpression will lead to the down-regulation of endogenous *MYC*.

First, we transfected NCH82 glioblastoma cells with either a control plasmid containing GFP (GFP-Ctrl) or a plasmid containing GFP and the exogenous *MYC* (GFP-(exo)*MYC*) (Figure 4A). We validated the ability of our sets of primers and probes to differentiate between the endogenous and exogenous Myc transcripts: An expected band of 404 bp for endogenous Myc was observed in the control, GFP-Ctrl and (exo)*MYC*-GFP transfected cells when detected with the corresponding primers and probe designed for endogenous *MYC* transcript. In contrast, an expected band of 216 base pairs was observed for the (exo)*MYC* transfected cells only when detected with the primers and probe designed for the exogenous *MYC* (Figure EV 4A).

The efficiency of transient transfection is usually low, resulting in a subpopulation of transfected cells hindering the interpretation of any quantitative assay to monitor gene expression out of whole-population lysates. Therefore, we proceeded to monitor the GFP expression and cell death by flow cytometry (Figure 4B) and fluorescence microscopy (Figure EV4B). We noticed that GFP-Ctrl transfected cells conserved a similar intensity in their fluorescence over time as shown by the fluorescence microscopy images. In contrast,

(exo)*MYC*-GFP transfected cells started with a dim GFP fluorescence that did not increase over time. In fact, the number of transfected cells decreased over time and some of these cells showed a rounded morphology, suggesting that exogenous myc overexpression is cytotoxic (Figure EV 4B). To quantify and confirm these results, we performed flow cytometry experiments to monitor GFP fluorescence and cell death, using propidium iodide (PI) staining. The condition of GFP-Ctrl transfected cells displayed 20% GFP positive cells including 2% of dead cells at 24 h post-transfection. This amount decreased to 16% with 4% of dead cells at 72 h post transfection. The condition of (GFP-(exo)*MYC)* transfected cells showed a significant lower efficiency of transfection (15% of GFP positive cells after 24 h of transfection). This value decreased to 9% at 72 h post transfection with a high proportion of dead cells (more than 50% of the transfected cells), supporting our suggestion that *MYC* overexpression is cytotoxic (Figure 4B).

In order to assess if *MYC* is dosage-compensated at the transcriptional level in the subpopulation of transfected NCH82 cells, we performed cell sorting of the viable (PI negative) GFP positive population of cells 24 h, 48 h and 72 h post-transfection. We extracted RNA from the sorted subpopulations to confirm expression of exogenous *MYC* by RT-qPCR (Figure EV 4C). We observed a strong increase of exogenous *MYC* 24 h and especially 48 h post-transfection, but no difference compared to the GFP negative subpopulation 72 h post-transfection. To quantify the effects of exogenous *MYC* on the expression of endogenous *MYC,* we compared *MYC* expression in the PI-/GFP+ population of GFP-Ctrl and GFP-(exo)*MYC* transfected cells (Figure 4C). Overexpression of *Myc* led to a significant reduction of approximately 30% of endogenous *MYC* expression at 24 h and 48 h post-transfection. 72 h after transfection we could not detect a significant difference in (exo)*MYC* expression (Figure EV 4C), neither in endogenous *MYC* expression, probably caused by an exo-*MYC* induced cytotoxicity, resulting in the time-dependent death of *MYC* overexpressing cells. These findings demonstrate the existence of regulatory circuits modulating the expression of *MYC* through an increase in gene-dosage, indicating that *MYC* dosage compensation is an active process in NCH82 cancer cells. To further study the nature of these regulatory circuits, we evaluated the expression of the 3 miRNAs predicted by the minimal model of *Myc* dosage compensation. As expected from the model, miR-17, miR-19a and miR-20a did increase upon (exo)*MYC* overexpression, confirming thereby the identity of the regulatory circuits involved in *MYC* dosage compensation (Figure 4D).

The blockade of MYC gene dosage compensation induces MYC dosage-dependent cytotoxicity

Since we observed that *MYC* overexpression induces NCH82 cytotoxicity, we asked whether the blockade of its dosage compensation also induces cell death and if the gene dosage plays a role in cell sensitivity to this blockade. However, since the minimal model indicates that *MYC* dosage compensation is redundantly regulated by miR17, miR19a and miR20a, we had to explore *in silico* the properties to identify the optimal strategies to achieve an inhibition

leading to *MYC* overexpression. For that purpose, we used the minimal model of *MYC* compensation to perform a triple parameter scan on the miRNA synthesis parameters (ks) to systematically reduce the amounts of the 3 miRNAs and identify steady states leading to *MYC* overexpression (Figure 5A). In addition, we aimed to evaluate the effect of *MYC* dosage on *MYC* overexpression under the inhibition of gene dosage compensation. Therefore, the scan was repeated under 3 conditions of different *MYC* copy numbers (2, 4 and 7) and the effect on the *MYC* concentration was observed. Although there are some slight alterations under a complete depletion of the 3 miRNAs for the diploid condition (CN = 2), the increase in *MYC* expression is much higher with increasing copy numbers, and less miRNA depletion is required to trigger the increase in *MYC* expression. The results indicate that the interruption of *MYC* dosage compensation requires the blockade of the 3 miRNAs, although less inhibition of miR17 is required compared to miR19a and miR20a. Moreover, the *MYC* overexpression was easier to achieve with higher *MYC* copy numbers. These results indicate that the perturbation of the mechanism of dosage compensation induces a gene dosage-dependent increase in *MYC* expression and suggests that *MYC*-amplified cells will be more sensitive to the blockade of *MYC* dosage compensation (Figure 5A).

A sudden increase in *MYC* expression in a therapeutic context has the potential to induce cytotoxicity, suggesting that the blockade of *MYC* dosage compensation has a promising therapeutic range that increases with the extent of dosage amplification. In order to confirm this hypothesis experimentally, we chose 3 colon cancer cell lines of the NCI60 panel with varying copy numbers of *MYC*, the HCT15 (copy number of 2), the HT29 (copy number of 4) and the SW60 (copy number of 7). We tested the effect of blocking *MYC* dosage compensation by transfecting the 3 cell lines with increasing concentrations (from 10 to 30 pM) of a mixture of anti-miR17, anti-miR19a and anti-miR20a and the respective concentrations of a negative RNA control. The cells were incubated and monitored by live microscopy for 72 h in the presence of Hoechst to stain all nuclei and Propidium Iodide to stain dead cells. After acquisition, images were analyzed using an image analysis pipeline programmed in CellProfiler to quantify the numbers of live and dead cells over time (Figure 5B). The qualitative observation of the images indicated no difference in cell growth and cell death between the anti-miRs and control transfected HCT15 cells, whereas the difference between those two conditions increased slightly for the HT29 and much more for the SW60 cells (Figure 5C).

The quantification of cell death for several replicates showed that cell death increased specially in SW60 cells treated with 30 pM of the anti-miR mixture compared to the negative control (Figure 5C). These results indicate that cells with higher *MYC* copy numbers are more sensitive to the blockade of gene dosage compensation by the perturbation of miR17, miR19a and miR20a, suggesting that aneuploid cancer cell lines are more sensitive to blockade of gene dosage compensation.

The kinetic parameters of the TF-miRNA interactions determine the capability of a putative network motif for gene dosage compensation

Since we were able to identify a minimal model of gene-dosage compensation for *MYC* and *STAT3* mediated by their interactions with 4 miRNAs, we next asked whether we could identify the properties of these network motifs conferring gene dosage-compensation. First, we designed a rationale for the identification of motifs mediating gene dosage-compensation based on our findings with *MYC* and *STAT3* (Figure 6A). To evaluate this rationale we constructed individual networks of our remaining gene candidates using the experimentally validated interactions included in BioNetUCR (*ZNF217, FOXC1, FOXF2, PGR*, *STAT5A*, *STAT5B*). Next, we looked for regulatory motifs of one, two and three nodes returning to the gene candidates. Only pgr, *STAT5A* and *STAT5B* had regulatory motifs returning to them after one, two or three nodes (not shown). Using these 3 networks, we built corresponding COPASI models and fitted them to the NCI60 dataset and evaluated the ability of the model to compensate by a parameter scan (Figure 6B). Afterwards, we performed in silico experiments by inhibiting individual or groups of model species (such as those experiments performed to identify the minimal model of *MYC-STAT3*) in order to identify the minimal models for *STAT5A* and *STAT5B* dosage compensation (unpublished observations). We identified a feedforward loop for the dosage compensation of *STAT5A* and 3 feedforward loops for the dosage compensation of *STAT5B* (Figure 6C). Indeed, those minimal models recapitulated the behavior of dosage compensation as expected (Figure 6D).

To study the kinetic properties of the regulatory loops of *MYC*, *STAT3*, *STAT5A* and *STAT5B* postulated hereby as dosage compensation circuits, we explored the kinetic parameters describing the interactions between miRNAs and TFs by performing combined parameter scans in COPASI using the minimal model of *MYC-STAT3*. Since the compensation in that model is redundant, we inhibited miR19A and miR20a so that we can evaluate only the behavior of the single negative feedback loop between *MYC* and miR17. Next, we monitored the concentration of *MYC* for two conditions: a *MYC* copy number of 1 and a *MYC* copy number of 3. Under these two conditions we scanned the effect of different values of the parameters and found that 3 parameters were best to separate the parameter landscape into two regions, a non-compensation space and a compensation space of parameters: the ks (synthesis rate of the miRNA), ka (activation parameter of the TF) and kr (repression parameter of the miRNA). After calculating the ratio of *MYC* concentrations of those 2 conditions, we obtained a three dimensional landscape of gene dosage compensation with a color-coded heatmap of the fold increase in *MYC* expression upon a 3 fold increase in copy number (Figure 6E). The blue region corresponds to the dosage-compensation space of parameters whereas the red region represents the space of parameters leading to a linear increase of gene expression as a function of copy number (no compensation). Moreover, when we located the positional values of the parameters corresponding to the compensating motifs, they all appeared in the blue region of compensation (Figure 6E) except by the interactions between *STAT3* and miR21 that appeared in the yellow transition zone. These results confirmed that the simple identification of network motifs such as feedback or

feedforward loops is not enough to identify systems-level properties such as gene-dosage compensation. This behavior is indeed dependent on the data-driven identification of the magnitude of the parameter values accounting for the emergence of this property of gene dosage compensation.

Breast Cancer patients with MYC dosage compensation have lower survival than patients with a non-compensated configuration

In order to further validate the importance of gene dosage compensation in cancer and its potential as therapeutic target, we set out to investigate the emergence of dosage compensation in the breast cancer dataset of the Tumor Cell Genome Atlas (TCGA). First, we tried to fit the minimal model of *MYC-STAT3* dosage compensation to all of the data using COPASI but the objective function was very high (not shown). We noticed a huge heterogeneity in the objective function values of the individual patient cases, suggesting indeed a heterogeneous configuration in this network. Therefore, we used the model as a tool to characterize the patient cases by fitting the model to the data of each individual experiment (case) and obtained a set of parameters for each patient. Next, we performed individual parameter scans to further characterize the cases in function of their response to copy number increases for each of the candidate genes for which a minimal compensation model was postulated (*MYC*, *STAT3*, *STAT5A* and sta5b). Therefore, we increased the gene copy number from 1 to 5 and observed the increase in gene expression: a sub-linear increase denotes the capability of the respective network configuration to achieve dosage compensation. We observed a large heterogeneity in *MYC* and *STAT5A* dosage compensation, with only a fraction of compensation-able cases for *STAT5B* and no compensation at all for *STAT3* in the models for these breast cancer patients (Figure 7A).

We aimed to classify the patients according to their gene amplification and compensation capability of their fitted models. The amplification in a gene suggests that the tumor is progressing in its transformation and raises the question of whether this amplification is compensated or not. We therefore plotted the real gene copy numbers against the simulated gene expression at a copy number of 5 to assess the ability of the network configuration for compensation. The lower the simulated gene expression, the higher the capability of the corresponding network configuration to perform compensation. Very few cases of amplification were observed for *STAT5A* and *STAT5B*, whereas many amplifications were observed for *MYC* with a separation between compensated and non-compensated cases that became more evident with higher copy numbers of *MYC* (Figure 7B). We therefore separated the cases in three groups: “No amplification” (CNV lower or equal to 1.5), and for those with higher amplification a group with “low compensation” (*MYC* expression higher than 3.7) and a group with “high compensation” (*MYC* expression lower than 3.7) (figure 7B) and plotted their survival curves implementing a Kaplan-Meier plot in Matlab (ecdf function, MathWorks) to compare the survival probabilities among the groups (Figure 7C). The significance in their differences was assessed by a Cox proportional hazards regression

(coxphfit, Matlab). We observed no significant difference between the "No amplification" group and the amplified cases with "low compensation" ($p = 0.1$). In contrast, we observed a significant decrease in the survival probability of the patients with "high compensation" compared to those with "low compensation" ($p = 0.009$) and compared to the cases of "No amplification" ($p = 0.007$). This result suggests that *MYC* dosage compensation enables cancer progression towards malignancy and it could indeed represent an important therapeutic target to prevent this progression.

Finally, to further characterize the nature of *MYC* dosage compensation in these models of breast cancer patients, we defined the new ranges of the 3 parameters previously identified to best partition the parameter landscape into the non-compensated and compensated spaces (Figure 6E). Using these ranges we performed the corresponding triple parameter scan and could identify again both regions of compensation behavior for the breast cancer cases (Figure 7D). Using the values of the fitted parameters for each patient, we used the N-D nearest point search algorithm (dsearchhn, Matlab, Mathworks) to obtain the compensation value for each patient for that particular miRNA-mediated feedback motif. We observed that miR-17 and miR-20a mediate *MYC* dosage compensation only in a small fraction of cases whereas most cases of breast cancer patients are compensated by the feedback motif regulated by miR-19. This suggests that miR-19 inhibition would represent a potential therapeutic target to block *MYC* dosage compensation in breast cancer.

Altogether, our results presented here a rationale for the identification of motifs mediating gene dosage-compensation. We developed a computational platform for that purpose, including bioinformatics and dynamical modeling tools (BioNetUCR, available here: https://cloud.prislab.org/s/gt2W2jfZQx3E3Jm). In addition, we presented a genetic tug-of-war approach to experimentally validate those compensation circuits. Besides, we evaluated the therapeutic potential of MYC dosage compensation to target aneuploid cancer.

Discussion

During carcinogenesis, aneuploidy autocatalizes genomic instability of transforming cancer cells leading to many unstable karyotypes and cell death due to error catastrophe (Solé & Deisboeck, 2004). However, in rare occasions, a specific combination of various alterations is met that overcomes those error thresholds, leading to malignant cells that are able to survive aneuploidy and genomic instability. This bottleneck in cancer evolution represents a gate to evolve malignant karyotypes leading to drug-resistance and metastasis (L. Li et al., 2009). Within that chaos of chromosomal instability, some conserved patterns in the karyotypic configurations suggest the presence of a stable mechanism, which function has to be maintained to ensure survival: i) there are specific aneusomies at different stages of cell transformation (Fabarius, Li, Yerganian, Hehlmann, & Duesberg, 2008), ii) there are clonal karyotypes evolving during the cell passages (Fabarius, Willer, Yerganian, Hehlmann, & Duesberg, 2002), iii) the cancer causing karyotypes have a chromosomal equilibrium

between destabilizing aneuploidy and the stabilizing selection for oncogenic function (L. Li et al., 2009), and iv) a large scale study revealed 2 distinct pathways to aneuploidy where the cells gain or lose chromosomes to restore the balance of their altered proteins and maintain viability (Ozery-Flato, Linhart, Trakhtenbrot, Izraeli, & Shamir, 2011). A possible explanation for this stability is given by the hypothesis of gene dosage compensation, a mechanism by which the expression of certain genes is modulated to compensate for differences in gene dosage when extra chromosomes are present (Kojima & Cimini, 2019).

In the present work we studied the mechanism of gene dosage compensation in aneuploid tumoral cells with the aid of models constructed by a biocomputational platform. We could identify *in silico* a behavior of dosage-compensation for *MYC*, *STAT3*, *STAT5A* and *STAT5B* mediated by their interactions with miRNAs and other transcription factors with a determined set of kinetic parameters. We experimentally validated the compensation circuits for *MYC* mediated by 3 feed-back loops with miR-17, miR-19a and miR-20a. The overexpression of *MYC* or the depletion of those miRNAs induced a MYC dosage-dependent cytotoxicity. We also found that *MYC* dosage compensation in TCGA breast cancer patients is heterogeneous and those patients with an active compensation presented lower survival probabilities compared to patients without dosage compensation.

The identification of dosage-compensated candidate genes is not trivial. We postulated a list of such candidates by analyzing genomic, transcriptomic and proteomic data of the NCI60 cell panel. This includes a collection of 59 cell lines of 7 types of cancer which are fully characterized at different omic levels. This led to the identification of a cluster of 56 candidate genes using our criteria of low tolerance in variation of their expression regardless of a high variation in their copy numbers (Figure 1). Our list of candidates includes many cancer-important genes, specially 7 highly connected transcription factors (Table 1), some of them clustered in common chromosomal locations. A similar criterium of a CNV-buffering score was proposed for aneuploid yeast strains as the dosage-compensated genes had higher variation in copy number but a constrained expression level (Hose et al., 2015).

Previous studies showed dosage compensation in different extent. For certain experimental models, this phenomenon appears to be broad affecting a large fraction of the aneuploid gene content. For example, a report on aneuploid wild yeast isolates showed gene-dosage compensation in 10-30% of amplified genes compared to isogenic or closely related euploid strains. They also predicted that dosage compensation occurs at genes that are most toxic when overexpressed and that their expression may also be under greater evolutionary constraint (Hose et al., 2015). In contrast, some studies indicate that only a handful of genes are compensated. For example, a previous report of the insertion of an additional chromosome 5 revealed that most proteins coded on the extra chromosomes are more abundant than proteins from diploid chromosomes. They also found that the messenger RNA (mRNA) levels generally correlate well with an increased DNA copy number in aneuploid cancers indicating that there is no general efficient mechanism for 'gene dosage compensation'. However, some specific proteins are maintained at diploid levels, specially those corresponding to kinases and ribosomal subunits (Stingele et al., 2012). Thus, there is controversy in the identification of genes under dosage compensation. This is not straightforward since the correlation between gene copy number alterations and gene

expression could be altered by many different factors. Therefore, novel criteria are required to unravel the degree of transcriptional adaptation to copy number alterations in a genome-wide fashion enabling thereby the identification of other genes under dosage compensation.

An additional question is whether dosage compensation is operational at the protein or at the transcriptional level. Gene dosage compensation was mostly reported at the protein level, whereby the stoichiometry of protein complexes in aneuploid cells determines that excess subunits are either degraded or aggregated (Brennan et al., 2019). Others reported that approximately 10% of the genome shows gene dosage compensation at the protein level, using a genetic tug-of-war technique (Ishikawa et al., 2017), although this approach was not designed to identify compensation at the transcriptional level.

Dosage compensation by the regulation of gene transcription might be an upstream and more effective mechanism, as it maintains the stoichiometry and preserves the energy required for transcription, translation, and eventual degradation of the extra proteins. Our approach was therefore intended to identify compensated candidates at both levels but we observed that all our candidates presented that behavior both at the transcript level pointing to a transcriptional dosage compensation that propagates to the protein level. However, the protein data set of the NCI60 is limited, probably precluding the identification of candidates compensated at the protein level only. To our knowledge, dosage-compensation at the transcriptional level was previously reported only for wild yeast isolates by comparing them to isogenic euploid strains (Hose et al., 2015) and the only previous evidence in mammalian cells was provided for some few genes upon the insertion of an additional chromosome 5 (Stingele et al., 2012) and for the inactivation of sex chromosomes (see below).

The immediate following question was about the mechanism of such dosage compensation at the transcriptional level. Three major mechanisms have been studied for sex chromosomes: inactivation of one of the X chromosomes in female embryos in mammals (Heard, Clerc, & Avner, 1997); doubling of transcription of genes on the single male X chromosome in *D. melanogaster* (Lucchesi & Kuroda, 2015) and halving the expression of genes on the two X chromosomes in hermaphrodites of *Caenorhabditis elegans* (Barbara J. Meyer, 2005) (B. J. Meyer, McDonel, Csankovszki, & Ralston, 2004). Moreover, several studies report that an extra chromosome appears to lead to a common epigenetic signature, characterized by global DNA hypomethylation of gene-poor regions and DNA hypermethylation of promoter regions (for review see (Kojima & Cimini, 2019)). Referring to specific genes, other authors propose that gene dosage compensation is most likely due to feedback control enabling rapid karyotypic evolution in yeast (Hose et al., 2015) and buffering of mRNA levels of amplified or deleted chromosomal regions in naturally occurring yeast strains (Kvitek, Will, & Gasch, 2008) as well as lager brewing yeast (Bond, Neal, Donnelly, & James, 2004). Moreover, it has been shown using synthetic transcriptional and post-transcriptional incoherent feedfoward loops that a gene product adapts to changes in DNA template abundance, supporting a previously hypothesized endogenous role in gene dosage compensation for such motifs (Bleris et al., 2011; Shimoga, White, Li, Sontag, & Bleris, 2014).

In endogenous transcription networks, interactions of miRNAs and transcription factors have been reported to assemble those kind of complex motifs including negative feedback loops, positive feedback loops, coherent feedforward loops, incoherent feedforward loops, miRNA

clusters and target hubs leading to non-linear, systems-level properties such as bistability, ultrasensitivity and oscillations (Lai et al., 2013; Vera et al., 2013), driving our attention towards a potential role of miRNA networks regulating gene dosage compensation. Indeed, miRNA networks are robust regulators of gene expression upon environmental changes (Herranz & Cohen, 2010) and they show adaptation to gene dosage through the formation of regulatory circuits with transcription factors (Bleris et al., 2011). Since the miRNA-TF networks are very complex, the gene dosage compensation could arise as an emergent property of the system meaning that a differential expression analysis of gene targets or miRNAs would be insufficient to identify the underlying mechanism of gene dosage compensation. Therefore, it is very hard to identify single miRNA-target interactions with relevant biological function. Classical approaches start with the identification of dysregulated miRNAs related to disease and extensive molecular and cellular biology work to validate the gene targets related to the phenotype of interest, which is inefficient because each miRNA can alter the expression of hundreds of genes by only 1.5 to 4 fold (Vera, Lai, Schmitz, & Wolkenhauer, 2013) and it is the cooperative effect of miRNA networks that makes them robust regulators (Herranz & Cohen, 2010; Matsuo, Eno, Im, Rosenshein, & Sood, 2010) . Thus, the identification of critical miRNAs involved in gene-dosage compensation networks requires the analysis of a mathematical model of those interactions by means of sensitivity analysis combined with predictive simulations to suggest key biochemical processes to become potential therapeutic targets (Lai et al., 2013; Vera et al., 2013).

We developed BioNetUCR, a computational platform to construct large scale models of miRNAs and TF interactions with the reverse approach, starting with a list of genes of interest that could be determined by differential expression analysis or customized by the researchers. Afterwards, a network topology is constructed with all the reported interactions of those target genes with miRNAs and TFs. However, the network topology is not enough to identify the most robust targets to control the phenotype of interest. Therefore, our approach goes beyond identifying a network and establishes a full dynamical system based on ODEs, which is calibrated with existing experimental data. Finally, the sensitivity analysis and the *in silico* experiments enabled us to identify the most robust target nodes to regulate the phenotype of interest, to determine the mechanism behind that phenotype and the interventions on those target nodes. This results in a smaller model with a higher explanatory power than the original (large) network.

Using our platform, we seek to identify a gene dosage compensation mechanism mediated by the emerging properties of complex miRNA-TF regulatory networks. With the aid of our biocomputational platform we studied the gene dosage compensation mechanism in the NCI60 panel of cancer cell lines. Using our candidate input genes, we determined a miRNA/ TF interaction network that connects all candidate genes but its structural analysis revealed that only those with transcription factor function can be subjected to gene dosage compensation. After model simplification, we defined a large-scale network of interactions including sources for both putative and experimentally-validated interactions. Using this network, we constructed and fitted an ODE mathematical model, which was not able to reproduce any behavior of dosage compensation. Despite of the initial set-backs, the turning point toward the modeling of the gene dosage compensation was the insight of the sensor

loop: the group of interactions starting in a transcription factor and returning after a defined number of interactions with other species including miRNAs and other TFs. These includes feedback and feedforward loops but also more complex interactions. The sensor loop itself also helps the network to adapt to fluctuations of the genes by repressing its expression. Once we looked for sensor loops in the regulatory networks and used only experimentally validated interactions to simplify our model, we easily modeled this mechanism and validated the hypothesis for the genes *MYC* and *STAT3* (Figure 2).

A deeper analysis of the mathematical model for *MYC* and *STAT3* enabled us to understand the underlying mechanism of gene dosage compensation by the reduction of model complexity into a minimal model recapitulating the same behavior of dosage compensation for both genes (Figure 3). First, a sensitivity analysis in COPASI enabled us to identify the main TFs and miRNAs regulating the concentrations (but not the compensation) of *MYC* and *STAT3*. Second, a parameter scan varying the copy number of *MYC* and *STAT3* enabled us to confirm which species increase or decrease together with increasing copy numbers for both genes. Afterwards, the same experiment was repeated along with the single inhibition of those species but the compensation behavior was unaltered. This immediately indicated that the mechanism of compensation is redundant, prompting us to inhibit groups of molecules instead of a single one. The inhibition of all miRNAs indeed abolished gene dosage compensation but it was restored by the single reactivation of 3 miRNAs for each gene. Moreover, we could confirm that the only way to abolish gene dosage compensation was to simultaneously inhibit those 3 corresponding miRNAs for each gene. After reconstructing the interactions from the original network, the minimal network states that *MYC* is compensated by 3 redundant negative feedback loops formed with miR17, miR19a and miR20a. The compensation of *STAT3* is mediated by 1 feedback loop with miR21 and 2 feedforward loops: (*STAT3*-*MYC*-miR17-*STAT3*) and (*STAT3*-*MYC*-miR20a-*STAT3*)(Figure 3). Actually, an important characteristic in biological networks like the miRNA-TF networks is its capability to adapt to fluctuating concentrations of its biomolecular components. In fact, coherent and incoherent feed forward loops (FFL) and feedback loops (FBL) help the network to adapt from such fluctuations (Carignano, Mukherjee, Singh, & Seelig, 2019; Osella, Bosia, Corá, & Caselle, 2011).

The biocomputational platform developed hereby was an accelerator of discovery for this work. By constructing models in minutes and automating several processes the platform shortened the time required to complete the construction of mathematical models. The compilation of around 450,000 experimentally validated regulatory interactions had a huge impact on the project. The amount of interactions allowed the platform to build interesting models with enough interactions. And by not using putative interactions, the models were cleaner and smaller. Furthermore, the biology- inspired *in silico* experiments enabled us to propose a minimal model of gene dosage compensation for *MYC* and *STAT3* out of experimentally-validated interactions. Due to the redundancy in gene dosage compensation, it would be unfeasible to identify such mechanism by single or even double inhibitions using a functional genomics approach without the guidance of a systems biology approach.

Nevertheless, our model has been simplified in several ways. From the initial 56 candidate genes, our current model offers an explanation for the gene dosage compensation of *MYC* and

STAT3 only. This is clearly a limitation due to the small amount of cases of the NCI60 panel representing only 59 cases of aberrant genomes. Further work with larger data sets such as the CCLE and TCGA could lead to the identification of further candidate genes and the availability of more data to obtain more accurate models of gene dosage compensation. In addition, we faced a limitation due to paramount complexity of the mathematical models, which prevent us from using putative (not experimentally validated) interactions to reduce model complexity. Despite of these limitations, and based on the experience with the minimal model for *MYC* and *STAT3*, we built single gene models for other two compensated candidates including only experimentally validated interactions. Hereby, we filtered for only those arcs establishing feed-back or feed-forward loops obtaining smaller models suitable for a faster parameter estimation. Among those, we could provide an explanation for the gene dosage compensation of *STAT5A* and *STAT5B* and reconstruct a minimal model for their corresponding underlying mechanisms of compensation (Figure 6). When plotting the kinetic parameters describing the main interactions of the compensating network motifs for all those genes within a three dimensional landscape of gene dosage compensation, we observed that they are all located in the compensation region. Therefore, we critically established that the network topology alone is not enough, as with different parameter values the model would lose gene dosage compensation, strengthening the need for full dynamic models.

Although our proposed minimal models of gene dosage compensation were based on experimentally-validated interactions that were previously reported and the interaction parameters were fitted to experimental data, it was necessary to confirm the ability of those regulatory circuits to perform gene dosage compensation. Therefore, we undertook the experimental validation of the *MYC* dosage compensation circuits by designing a novel tug-of-war approach to identify gene dosage compensation at the transcriptional level. We got inspired by a brilliant analogous approach to identify gene dosage compensation at the protein level. They used a TAP tag next to an endogenous target gene to evaluate its response to the overexpression of an exogenous protein that was over-expressed and they were able to differentiate both exogenous and endogenous versions at the protein level. If dosage compensation circuits were present, the overexpression of the exogenous protein puts pressure on the regulatory circuits leading to the down-regulation of the endogenous protein (Ishikawa et al., 2017). We developed a similar approach to identify gene-dosage compensation at the transcriptional level based on the same principles: 1. To differentiate the exogenous and endogenous transcripts of *MYC*, we overexpressed an exogeneous version of *MYC* ((exo)*MYC*) with a differential codon optimization compared to the human endogenous *MYC* ((endo)*MYC*. 2. We removed the 3'UTR segment of the exogenous *MYC* to avoid any miRNA regulation on this transcript so that it can exert a higher pressure on the *MYC* regulatory circuits. 3. Although both transcripts will be translated into the same protein and exert the same regulatory activities, they can be differentiated at the transcript level by RT-qPCR using specific primers and probes. Under the pressure of the dosage compensation circuits, we observed that the exogenous *MYC* overexpression leads to the down-regulation of endogenous *MYC*, confirming thereby the ability of these circuits to perform gene dosage compensation. In addition, we confirmed the overexpression of the 3 corresponding miRNAs involved in those regulatory circuits responding to (exo)*MYC* expression.

Not all of dosage-altered genes are critical nor require compensation to maintain cell viability. Therefore, we asked about the potential to target *MYC* dosage compensation against aneuploid cancer since its blockade would lead to an increase in *MYC* expression. The gene *MYC* (*MYC* proto-oncogene, bHLH transcription factor), aka c-*MYC*, is a well known and studied transcription factor involved in proliferation, cell growth, cell differentiation and apoptosis. It is estimated that *MYC* regulates 15% of the genes and it is also an oncogene, usually overexpressed in many kinds of cancer (Dang, 1999) and active in 70% of human cancer, but it is also related to apoptosis (Prendergast, 1999). Interestingly, the dysregulation of *MYC* may lead to cancer but may also lead to a cell suicide (Nilsson & Cleveland, 2003) , it is reported to have a dual function from oncogene to tumor suppressor in leukemia (Uribesalgo, Benitah, & Di Croce, 2012) and that dysregulated *MYC* triggers rapid apoptosis (Nilsson & Cleveland, 2003). In fact, we observed that (exo)*MYC* expression leads to cell death in glioblastoma cells. These findings support an important prediction of the study of gene dosage compensation in aneuploid yeast, that the phenomenon of dosage compensation occurs at genes that are most toxic when overexpressed (Hose et al., 2015). These results also suggest that the inhibition of the miRNAs compensating *MYC* induces cell death.

The miR17, miR19a and miR20a are actually all co-regulated as the miRNA cluster miR-17-92. This is considered an oncogenic cluster and is actually called OncomiR-1 since it is overexpressed in several types of cancer (for review see (Fuziwara & Kimura, 2015)). Several reports showed that these miRNAs form important network motifs with *MYC* in B-cell lymphoma (Mihailovich et al., 2015), with E2F/*MYC* (Y. Li, Li, Zhang, & Chen, 2011) and with *STAT3* in retinoblastoma (Jo et al., 2014). Interestingly, the *in silico* inhibition of those 3 miRNAs involved in *MYC* compensation circuits led to a dosage-dependent increase in *MYC* expression and the *in vitro* inhibition of those miRNAs in an experimental model of colon cancer led to increasing levels of cytotoxicity as *MYC* copy number increases (Figure 5), indicating that dosage-altered cancers are more fragile to blockade of gene-dosage compensation. The identification of miRNAs controlling cancer robustness has a huge therapeutic potential since miRNAs are becoming more attractive targets for therapy, as shown for the first time by miRNA-122 against hepatitis C infection and hepatic cancer (Lindow & Kauppinen, 2012). Nowadays, at least 7 miRNA mimics or inhibitors are being tested in clinical trials thanks to the development of chemical modifications to increase their stability, to improve targeting to disease sites or the transport by several options of delivery systems (Rupaimoole & Slack, 2017).

Nevertheless, we also hypothesized that cancer patients could have differential configurations in the circuits postulated hereby as dosage-compensation mechanisms. Using the personalized mathematical models for each patient, we observed heterogeneity in the capability of their individual models to achieve gene-dosage compensation *in silico*, especially for *MYC* and *STAT5A*. No dosage compensation capability was observed for *STAT3* and only few cases were observed for *STAT5B*. This suggests that the configuration of gene dosage compensation circuits may vary across patients and probably tumor types. For instance, *MYC* compensation in the models of these breast cancer patients is mostly mediated by miR-19a and only few cases were compensated by the feedback loops with miR-17 or miR-20a. The future determination of the molecular determinants of compensation could lead

to the identification of valuable biomarkers to direct precision medicine strategies.

The response of the corresponding network configuration also shed light on the importance of gene dosage compensation in cancer. There were very few cases of amplification for *STAT5A* and *STAT5B* to evaluate the effect of gene dosage-compensation on patient survival but *MYC* amplifications were very frequent. Interestingly, the higher the extent of this amplification the higher the separation of cases in terms of *MYC* compensation leading to a clear separation of 2 groups with a significant difference in patient survival. These findings suggest that *MYC* dosage compensation contributes to a state of pro-tumoral stability enabling cancer cells to further evolve into more malignant phenotypes.

The further analysis of the current models of gene dosage compensation or even other genes could reveal novel specific targets against cancer. Thus, we suggest that cancer has a robust Achilles-Heel due to an increased sensitivity to perturbations in these circuits, which is not necessarily reflected as differences in miRNA expression levels but at systems-level properties. The identification of control points blocking the dosage compensation could lead to the overexpression of these genes and others under their influence in a context of fragility for the cancer cell. This strategy is promising inasmuch the overexpression of these important transcription factors seem to be more sensitive to blockade of gene dosage compensation when their copy numbers are more amplified.

In conclusion, the present work led to the construction of a complex mathematical model of gene dosage compensation and formulated model-driven hypothesis for the identification of novel targets against aneuploid cancer. In addition, the computational platform built with this project has other potential applications to understand miRNA-mediated gene regulation and perform simulations of the systems-level effects of perturbations in miRNA networks related to disease. Furthermore, we presented a rationale to identify dosage compensation circuits and a genetic tug-of-war approach to experimentally validate those regulatory motifs.

Future work is required to confirm the effect of gene dosage compensation on patient survival and to identify other compensated genes to direct personalized precision therapies against cancer. Indeed, the *in silico* evaluation of gene dosage compensation in larger data sets could lead to the identification of novel targets and the elucidation of a core network of essential genes able to maintain cancer stability, enabling thereby the variation in all the other genes during cancer evolution. Altogether, the current results could contribute to the identification of that stability core of essential genes, which manipulation of specific nodes has the potential to become a novel approach to specifically target aneuploid cancer cells.

Materials and Methods

Data sources

Data was gathered from several sources. The primary sources were from experiments on the NCI60 panel: gene copy number (Bussey et al., 2006), RNA gene expression (Shankavaram et al., 2007) and protein expression (Gholami et al., 2013). MicroRNA related data was downloaded from Mirtarbase (Hsu et al., 2014) and MiRBase (Kozomara et al., 2019; Kozomara & Griffiths-Jones, 2014). For background knowledge on gene regulation we relied on several sources: Transmir (Wang, Lu, Qiu, & Cui, 2010), Pazar (Portales-Casamar et al., 2009), TRED (Transcriptional Regulatory Element Database) (Jiang, Xuan, Zhao, & Zhang, 2007), CircuitsDB (Friard, Re, Taverna, De Bortoli, & Corá, 2010) .

Gene classification using Gaussian Mixture Models

In order to classify genes according to their behavior, we developed a computer algorithm based on the Gaussian Mixture Model functions in MATLAB. An increasing number of components (ki) of the GMM model is added sequentially and the GMM training is performed for several iterations searching for the best fit to the experimental data. The resulting GMM is used to classify the cells of the original data set. A MANOVA test is applied to the resulting clusters to evaluate the statistical significance of adding another component to the GMM. This is done until the new component adds no further significance.

Construction of the Regulatory Network

We are interested in the regulatory network of miRNAs and TFs in the proximity of directly affecting the genes selected by the GMM. In order to build this network, we relied heavily on the interaction database gathered previously. A regulatory interaction in this database is an arc which starting point is the regulator and the ending point is the target. For many interactions, the database tells whether the regulation is an activation or repression. When the database does not provide this information, if the regulator of the interaction is a TF we assume that the regulation is an activation, and if the regulator is a miRNA we assume that it is a repression.

We use a directed graph to represent this network. We first identify the nodes of this graph which are composed by all the genes selected by the GMM as well as all the direct regulators and regulated nodes of these genes. To identify the regulators, we search in the interaction database for arcs ending in one of the selected genes. The direct regulators are all the starting nodes of each of the arcs found. Similarly, we search for arcs starting in one of the selected genes and the end node of the arc is a directly regulated gene. Hence, we have successfully identified all the nodes of the graph. Each node is a selected gene, a miRNA or TF, all of

them biomolecular species relevant in the gene dosage compensation phenomenon. We then build the list of arcs of the graph. From the list of nodes or species identified, for each pair of species, if there is an arc in the interaction database linking these two species, we add this arc in the list of arcs of the graph. The lists of species and arcs identified becomes the graph representing the regulatory network affecting the selected genes of interest.

In order to examine the target/miRNA/TF network for the presence of regulatory motifs with systems-level properties, we searched for positive and negative feedback loops (between miRNAs and TFs), coherent feedforward loops and incoherent feedforward loops.

Ordinary differential equation modeling of miRNA-TF networks

We now construct a system of ODEs from the graph representation of the regulation network. Since the gene dosage compensation phenomenon is gene expression related, we are interested in constructing a kinetic model from the graph. For each species of the graph, we write an equation computing the concentration or mRNA expression of this specie. The list of all these equations defines a system of ODEs modeling the metabolic behavior of the regulation network. We have three different types of equations, one for each type of species: candidate gene, TF and miRNA. The schematic representations of these equations are shown in (Figure 2B).

For a candidate gene, the mRNA expression equation has a synthesis part and a degradation part. The expression of each gene equals to the synthesis part subtracted by the degradation part.

The synthesis equation of the candidate gene is the product of three factors: a) the gene copy number, b) the synthesis rate of the gene, a parameter to be fitted by the Parameter Estimation process, and c) the total regulatory effect exercised by TFs on the candidate gene which in the graph is represented by all the incoming arcs of the candidate gene where the arc's regulator is a TF. The total regulatory effect factor is usually >1 since most TFs are activators but it could be <1 when the regulating TFs are mainly suppressors.

The degradation equation of a candidate gene is also the product of three factors: a) the mRNA concentration of the gene, b) the degradation rate of the gene, another parameter to be estimated, and c) the total regulatory effect of miRNAs on the gene. Because the degradation equation is a product and the total repression effect is one of the factors, the value of the repression effect is usually >1 since miRNAs are suppressors, thus exercising an accelerating effect on the degradation.

For modeling the total regulatory effect, we define the following parameters, also to be estimated later against experimental data. For a TF we define an activation rate that represents the rate at which the TF activates o intensifies the expression of the TF's targets. Likewise we define a repression rate representing the rate at which the TF suppresses or silences its targets when the TF has a repressing role. For a miRNA we also define a repression rate since miRNA can only repress its targets. The single regulatory effect of one TF or one miRNA over all of its targets is the product of the concentration of the TF or miRNA multiply by the corresponding rate. For instance, the concentration of *MYC* multiplied by *MYC*'s activation rate or the concentration of mir19 by mir19's repression rate. Given that the concentration of a regulator, the activation rate and the repression rate are all zero or positive, the single effect of one regulator is also zero or positive. Following this definition, we have only one repression rate for a miRNA, therefore a given miRNA will exercise the same rate of repression over each of its targets.

We model the total repression effect of miRNAs in the degradation equation of a candidate gene as the sum of 1 plus the single effect of each of the miRNAs repressing that gene. This value is always greater or equal than 1. When there are no miRNA repressing the gene, this factor would be 1. These values model the expected function of the total repression effect in the degradation equation for the presence of repressing miRNAs would accelerate the degradation of the gene.

We define the total regulation effect of TFs in the synthesis equation of a candidate gene as a ratio between the total activation effect of the TFs divided by the total repression effect of the TFs. The total activation effect is the sum of 1 plus the single effect of each of the TFs activating the gene. Likewise, the total repression effect is the sum of 1 plus the single effect of each TFs repressing the gene. When there are no TFs regulating the gene, this ratio is 1. When there are TFs activating or repressing the gene, if the total activation effect is greater than the total repression effect then the ratio is greater than 1, hence accelerating the synthesis of the gene. If there are more repression effect than activation effect, the ratio is less than 1 slowing down the synthesis. If both effects are equal, the ratio is 1 exercising no effect on the synthesis. These ratio's values model the expected behavior of the interplay between the activating and repressing TFs in the synthesis equation of a gene.

$$Synthesis\ gene = CN_{gene} \cdot KS_{gene} \cdot \frac{1+KA_{TFA1} \cdot [TFA1]+...}{1+KR_{TFR1} \cdot [TFR1]+...}$$

$$Degradation\ gene\ RNA = KD_{geneRNA} \cdot [GeneRNA] \cdot (1 + KR_{miRNA1} \cdot [miRNA1] + ...)$$

$$Synthesis\ TF = CN_{TF} \cdot KS_{TF} \cdot \frac{1+KA_{TFA1} \cdot [TFA1]+...}{1+KR_{TFR1} \cdot [TFR1]+...+KR_{miRNA1} \cdot [miRNA1]+...}$$

$$Degradation\ TF = KD_{TF} \cdot [TF]$$

$$Synthesis\ miRNA = CN_{miRNA} \cdot KS_{miRNA} \cdot \frac{1+KA_{TFA1} \cdot [TFA1]+...}{1+KR_{TFR1} \cdot [TFR1]+...}$$

$$Degradation\ miRNA = KD_{miRNA} \cdot [miRNA]$$

The mRNA expression equations for a TF or miRNA are slightly different from the equation of a candidate gene. These two equations also have a synthesis part and a degradation part but instead of having the repressing effect exercising an accelerator effect in the degradation part, they have the repressing effect in the synthesis part applying a decelerating effect in the synthesis part.

The degradation equation for a TF or miRNA is the product of only two factors: a) the mRNA concentration of the TF or miRNA, and b) the degradation rate of TF or miRNA. The synthesis equation for a TF or miRNA is the product of three factors: a) the copy number of the TF or miRNA, b) the synthesis rate of the TF or miRNA, and c) the total regulatory effect exercised by others TFs and miRNAs on this TF or miRNA which in the graph is represented by all the incoming arcs of this TF or miRNA where the arc's regulator is a TF or miRNA.

While the total activation effect depends only on activating TFs, now the total repressing effect involves not only repressing TFs but also repressing miRNAs. Whether there are more total activating effect or more total repressing effect, the total regulatory effect would be >1 or <1, or exactly 1 if both effects are the same. Although this equation applies to TFs and miRNAs, in practice the synthesis equation for miRNA do not involve repressing miRNA effects since miRNAs do not repress other miRNAs.

From the three types of equations defined, we detect three different types of elements: copy numbers, rates and concentrations. There is a copy number parameter for each species of the graph. There are four different types of rates o parameters to be fitted against experimental data: synthesis, degradation, activation and repression rates. A candidate gene has only a synthesis and degradation rate. A TF has all four types of rates. And a miRNA has all but an activation rate since miRNAs only repress. Finally, there is also one concentration for each species of the graph.

The concentrations of the regulators that are part of the total regulation effect factor of the equation play an important role in the modeling of the regulatory effects. For an activator TF, as its concentration increases, so increases the activation effect and the concentration of the

TF's regulated gene. In the same way, as the concentration of a miRNA increases, so increases the repression effect causing a decline in the concentration of the miRNA's regulated gene, either by slowing down its synthesis or by accelerating its degradation. The concentrations also play an important role in the modeling of common network motifs such as feedback loops, feedforward loops or other types of loops. In a loop like A->B->C->A, the concentration of A is part of the regulatory factor of B's equation, the concentration of B is part of C's equation, and C's concentration is part of A's equation. The interplay of these three equations models the dynamics of the loop in the graph.

Having defined the equations of the species and identified the different elements of these equations, translating the graph of the regulatory network into an SMBL model (Hucka et al., 2003) is straightforward. An SBML model follows the XML standard (W3C/XML, 2008) and have several sections for the different elements of the SBML model.

Model parameter estimation

The SBML model is imported into COPASI, an application for analyzing biochemical networks like the miRNA-TF regulatory network. The platform also add 59 sets of experimental data to this model, one set for each cell line of NCI60. We then run a *Parameter Estimation* task in COPASI. The task adjusts the parameters of the model to fit the RNA expression of each species of the model as close as possible to the corresponding RNA expression of 59 experimental data-sets.

To assess that the RNA expression of the model is *close enough* to the RNA expression of the experimental data: a) we build a model for each cell line by using the fitted model and changing the gene copy number and initial concentration values to the values of the corresponding cell line; b) we run a *Time Course* task for each model and gather the RNA expression for each species in each of the cell lines; c) we use a dependent *t*-test for paired samples with $\alpha=0.05$ to compare the model's RNA expression of the 60 cell lines of a species with the corresponding experimental RNA expression. The pairing is on the cell lines. Species with significant differences between their experimental and model's RNA expression are discarded from the model and a new Parameter Estimation task is run for the modified model until the difference between the experimental and the simulated data is not significant (T-student, $p > 0.05$). The presumption is that the experimental data do not explain the discarded species and they therefore do not belong to the model.

Assessment of gene dosage compensation behavior

To evaluate if a gene is compensated in the fitted model, we run a *Parameter Scan* in COPASI, varying the copy number of the gene from 1 to 5 and computing the respective

RNA expression. A non-compensated gene has a near linear fold increase in RNA expression level as its corresponding copy number increases linearly. We consider a gene compensated if the increase in RNA expression follows a sublinear increase as the copy number increases linearly.

Anti-miRNA transfections and cytotoxicity assays.

Optimized densities of colon cancer cells lines were seeded in 96 well plates (Greiner Bio-One µclear) to evaluate the cytotoxicity of the miRNA inhibitors for 72 h: 40 000 SW-620 cells, 20 000 HCT-15 and 40 000 HT-29 cells were plated per well in 100 µL of RPMI medium supplemented with 10% of fetal serum (GIBCO), 24 h before the start of the experiment and incubated in 5% CO_2 and 37°C. Next, cell nuclei were labeled using a Hoechst 33342 (Invitrogen™ H3570, 1.25 µg/mL final concentration) for 10 minutes before washing. Afterwards, the cells were transfected with the mixture of 3 antimiRs: i. *mir*Vana® miRNA inhibitor, hsa-miR-17-5p - assay ID: MH12412 (Invitrogen 4464084), ii. *Mir*Vana® miRNA inhibitor, hsa-miR-19a-3p – assay ID: MH10649 (Invitrogen 4464084) and iii. *mir*Vana® miRNA inhibitor, hsa-miR-20a-5p – assay ID: MH10057, (Invitrogen 4464084) or the negative control (*mir*Vana® miRNA inhibitor, Negative Control #1, Invitrogen 4464076) using Lipofectamine™ RNAiMAX Transfection Reagent (Invitrogen 13778075) and following the Lipofectamine® RNAiMAX Transfection Protocol of the manufacturer at final concentrations of 10, 20 and 30 pmol/well. Briefly, a 1:1 mixture of lipofectamine and miRNA mixture was prepared, incubated 5 min at room temperature and 10 µL were added to each well. The cells were incubated for 6 h at 37°C with 5% CO_2 before replacing the media. Propidium iodide (PI, Invitrogen™ P3566, final concentration of 10 µg/mL) to stain dead cells in a further 72 h incubation inside the imaging chamber of a Cytation 5 Imaging System (Biotek) with an imaging frequency of every 12 hours. Control wells of unstransfected cells were included for each cell line. Images in the red (PI, dead) and blue (Hoechst, all nuclei of all cells) channels were taken. With this information, live cell percentage was determined using the Cell Profiler image analysis software, through an image analysis pipeline designed to count total and dead cells (see supplementary pipeline file attached). The percentage of cytotoxicity was calculated dividing PI negative cells by the total amount of cells in each well.

Molecular design and cloning

A construct encoding the moxGFP gene (Costantini et al., 2015) followed by 2A sequences from Porcine Teschovirus-1 (P2A) and Thosea asigna virus (T2A) was commercially synthetized with human/mouse codon optimization (GenScript, Piscataway, NJ) and inserted into the NcoI and XbaI sites of the transposon-based plasmid pSBbi-Pur (a gift from Eric Kowarz, Addgene plasmid #60523) (Kowarz, Löscher, & Marschalek, 2015), to produce the vector pSBbi-moxGFP-P2A-T2A-Puro. Next, nucleotides 409-1728 of the c-*MYC* cDNA (GenBank reference sequence NM_002467.6) were commercially synthetized with human/mouse codon optimization (GenScript) and inserted into the SmaI and XbaI sites of

the plasmid pSBbi-moxGFP-P2A-T2A-Puro, to generate the polycistronic vector pSBbi-moxGFP-P2A-T2A-c*MYC*-Puro.

Culture of human glioblastoma cell lines

The human primary glioblastoma cell line NCH82 was generated at the Department of Neurosurgery, Heidelberg University Hospital (Heidelberg, Germany) (Karcher et al., 2006). Cells were cultured in cDMEM [DMEM (Sigma-Aldrich, St. Louis, MO, USA), 10% heat-inactivated FCS (Sigma-Aldrich), 2 mM L-Glutamine (Gibco Invitrogen, Carlsbad, California, USA), 50 µg/ml Gentamicin (Gibco Invitrogen)] at 37°C and 5% CO_2.

Plasmid transfection

One day prior transfection, 100,000 NCH82 cells/well were seeded in cDMEM in 24 well tissue culture vessels (Greiner Bio-One, Frickenhausen, Germany) and incubated for 24 h (37°C, 5% CO_2). At the day of transfection, old medium was replaced by 500 µl fresh cDMEM; 1 µg plasmid and 3 µl Lipofectamine™ 2000 Transfection Reagent (ThermoFisher Scientific, Waltham, MA, USA) were mixed in 100 µl Opti-MEM™ I Reduced Serum Medium (Gibco Invitrogen) and incubated for 5 min at RT. Afterwards transfection mix was added to the wells and cells were incubated for 24 -72 h. Cell growth and GFP expression were documented using a Leica DM IL LED (Leica Microsystems GmbH, Wetzlar, Germany).

Flow Cytometry and Fluorescent Activated Cell Sorting (FACS)

Prior to analysis, cells were detached using 0.25% Trypsin-EDTA (Life Technologies, Carlsbad, California, USA), centrifuged (400xg, 5 min, 4°C) and washed with ice-cold PBS (Gibco) and stained for 15 min in PBS containing 1µg/ml propidium iodide (PI; Carl Roth, Karlsruhe, Germany). Samples for flow cytometry were analyzed using an Attune NxT Flow Cytometer (ThermoFisher Scientific). For FACS, samples of three technical replicates were pooled and analyzed using a BD FACSAria™ III Cell Sorter (BD Biosciences, San Jose, CA, USA) at the Flow Cytometry Core Facility of the Institute of Molecular Biology, Mainz. PI positive cells were discarded whereas PI negative cells were divided in a GFP positive and a GFP negative population. Cells were collected in QIAzol Lysis Reagent (Qiagen, Hilden, Germany) and stored at -80°C until total RNA extraction.

Total RNA isolation, cDNA transcription and gene expression profiling

After FACS, total RNA isolation was performed on the sorted cells using the miRNeasy Mini Kit (Qiagen) according to manufacturer's instructions. RNA concentration and quality were determined using a Nanodrop 2200 (ThermoFisher Scientific). For cDNA transcription of mRNA, the FastGene Scriptase II - Ready Mix (NIPPON Genetics Europe, Düren, Germany) was used as described in the manufacturer's protocol. For miRNA transcription, 10X M-

MuLV-buffer (New England Biolabs [NEB], Ipswich, MA, USA), 1 mM adenosine triphosphate (Life Technologies), 1 mM dNTP-Mix (NEB), 10 U/µl moloney murine leukaemia virus reverse transcriptase (NEB), 0.1 U/µl poly-A polymerase (NEB), 0.05 pM cel-mir-39 as spike-in control and 0.5 µM of reverse transcriptase primers (Table 2) were added to the extracted total RNA. The reverse transcription protocol consisted of 30 min at 37°C, followed by 60 min at 42°C and a final incubation for 5 min at 65°C.

Quantitative PCR (qPCR) analysis was performed using 96-well qPCR plates (Sarstedt, Nümbrecht, Germany) and a StepOnePlus System (Applied Biosystems, Foster City, CA, USA). For gene expression analysis, 10 µl of qPCRBIO Probe Mix (NIPPON Genetics Europe) were mixed with 5 ng cDNA, 1 µl of the appropriate primer (Table 2) and 4 µl H_2O. For miRNA analysis, 5 µl iTaq Universal SYBR Green Supermix (Bio-Rad, Feldkirchen, Germany) were mixed with 0.56 ng cDNA and 0.5 mM of the appropriate primer (Table 2). Relative quantification (RQ) of gene or miRNA expression was determined using the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen, 2001) *GAPDH* and *SDHA* were used as reference genes for mRNA analysis. SNORD110 was used to normalize miRNA results. Correctness of amplicons was verified using melt-curve analysis (miRNA; Unpublished Observations) and agarose gel electrophoresis (Figure EV4).

Statistical analysis was performed using GraphPad Prism software. T-tests were performed to analyze means of samples with a significance defined by an α of 0.05. The resulting p-values were corrected for multiple comparisons using the Holm-Sidak method.

Acknowledgments

We thank the initial contribution of Dr. Michael Kesling for his advises on the project formulation and the data analysis. This project was funded by the grant FEES-CONARE (Costa Rica) and Vicerrectoría de Investigación, University of Costa Rica and Universidad Estatal a Distancia (UNED). Carsten Geiß was supported by the FAZIT-STIFTUNG. Support by the IMB Flow Cytometry Core Facility (Mainz, Germany) is gratefully acknowledged. PM thanks the NIH for supporting the development, use and dissemination of COPASI (grants GM080219 and GM137787)

Author contributions

Conflict of interest

The authors declare no conflict of interest.

Tables

Table 1. Candidate genes under dosage compensation.

Gen ID	Location	Description	Transcription factor	Forward loops	Type of regulation RNA	Type of regulation Protein
MYC	8q24.21	v-myc avian myelocytomatosis viral oncogene homolog	yes	461	yes	no data
ZNF217	20q13.2	zinc finger protein 217	yes	451	yes	yes
MTSS1	8p22	metastasis suppressor 1	no	102	yes	no data
SEMA3D	7q21.11	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3D	no	40	yes	yes
TRIM37	17q23.2	tripartite motif containing 37	no	54	yes	no data
STAT3	17q21.31	signal transducer and activator of transcription 3 (acute-phase response factor)	yes	577	yes	yes
KCNH4	17q21.2	potassium voltage-gated channel, subfamily H (eag-related), member 4	no	9	yes	no data
RAB5C	17q21.2	RAB5C, member RAS oncogene family	no	99	yes	yes
FOXC1	6p25	forkhead box C1	yes	41	yes	no data
BIRC2	11q22	baculoviral IAP repeat containing 2	no	2	yes	yes
CUL5	11q22.3	cullin 5	no	179	yes	yes
ATM	11q22-q23	ATM serine/threonine kinase	no	2	yes	yes
NPAT	11q22-q23	nuclear protein, ataxia-telangiectasia locus	no	76	yes	yes
MMP12	11q22.3	matrix metallopeptidase 12 (macrophage elastase)	no	2	yes	no data
DCUN1D5	11q22.3	DCN1, defective in cullin neddylation 1, domain containing 5	no	31	yes	yes
PGR	11q22-q23	progesterone receptor	yes	1	yes	yes
PDCD10	3q26.1	programmed cell death 10	no	139	yes	yes
ATP1B2	17p13.1	ATPase, Na+/K+ transporting, beta 2 polypeptide	no	17	yes	no data
ANKFY1	17p13.3	ankyrin repeat and FYVE domain containing 1	no	147	yes	yes

Table 2: Primers and their respective sequence used for quantitative PCR. In case commercial primers were used assay ID (ThermoFisher scientific) is shown.

Primer description	**Sequence (5'-3') or Assay ID**
reverse transcriptase primer CT	CAGGTCCAGTTTTTTTTTTTTTTTCT
reverse transcriptase primer CA	CAGGTCCAGTTTTTTTTTTTTTTTCA
reverse transcriptase primer GC	CAGGTCCAGTTTTTTTTTTTTTTTGC
GAPDH	Hs02758991_g1
SDHA	Hs00188166_m1
Exogeneous *MYC*	Fw: GGCCGTGACACCTTTCAGCC Rv: TCGGACACCAGCTTTGCTGC Probe: ACAGAGCTGCTGGGCGGCGA
Endogenous *MYC*	Fw: CCTCCTCACAGCCCACTGGT Rv: CGTCGTTTCCGCAACAAGTCC Probe: CGCAGCGCCTCCCTCCACTCG
Cel-miR-39	Fw: CAGTCACCGGGTGTAAATC Rv: GTCCAGTTTTTTTTTTTTTTTCAAGC
SNORD110	Fw: ACGCAATCACTGATGTCTC Rv: AGTTTTTTTTTTTTTTTGCTCAGA
hsa-miR-17-5p	Fw: GCAAAGTGCTTACAGTGC Rv: TCCAGTTTTTTTTTTTTTTTCTACC
hsa-miR-19a-3p	Fw: GCAGTGTGCAAATCTATGC Rv: GGTCCAGTTTTTTTTTTTTTTTCAG
hsa-miR-20a-5p	Fw: GCAGTAAAGTGCTTATAGTGC Rv: GTCCAGTTTTTTTTTTTTTTTCTACC

FIGURE LEGENDS

Figure 1. Identification of candidate genes under dosage compensation in the NCI60 panel. A. Input data of gene copy number (DNA), gene expression (RNA) and protein levels (protein) of the NCI60 panel. The absolute values are shown on the left panel and the right panel corresponds to the log2 values normalized to the averaged RNA and protein of the diploid cell lines for the respective gene (Normalized to diploid). B. Standard deviations (SD) of the DNA, RNA and Protein levels for each gene across the 59 cell lines of the NCI60 panel. C. Gaussian Mixture Model to identify a cluster of subpopulation of genes with high SD DNA and low SD RNA and/or low SD Protein (white arrow). D. Gene Clustering according to the model in C (left). The cyan cluster contains candidate genes under dosage compensation, characterized by high SD DNA, low SD Protein (middle) and low SD RNA (right). E. Absolute and Normalized values of selected candidate genes under dosage compensation. F. Examples of candidate genes under dosage compensation (*MYC* and rab5c). G. Clusters of candidate genes according to their copy number variations across the 59 cell lines of the NCI60 panel.

Figure EV1. DNA, RNA and protein levels for all candidate genes.

Figure 2. A mathematical model driven by NCI-60 data leads to a phenotype of gene dosage compensation for *MYC* and *STAT3*. A. Network of miRNA-TF interactions for *MYC*, *STAT3*, *STAT5A* and *STAT5B*. **B.** Schematic representations of the modeled interactions for the target gene, miRNAs and associated TFs. **C.** The results of the parameter estimation task to fit the ODE mathematical model to the experimental data upon different modifications led to a reduction in the objective function (difference between modeled and experimental data). **D.** The graphical comparison of modeled and experimental data shows no significant differences. **E.** Results of parameter scan on the values of copy number for the candidate target genes showing gene dosage compensation for *MYC* and *STAT3*. **F.** The sensitivity analysis points to some candidate regulatory molecules controlling the concentrations of *MYC* and *STAT3*.

Figure EV2. Putative miRNA-transcription factor interaction network linking all candidate target genes. A. Correlation coefficients of copy number variations of the candidate genes and miRNA/transcription factor expression across the NCI60 panel. **B.** Correlation coefficients of copy number variations of the candidate genes and miRNA/transcription factor expression across the NCI60 panel, only for those putative interactions present on the network in C. **C.** Network of putative interactions among target genes, miRNAs and transcription factors (red indicates repression and blue activation) including different types of regulatory loops formed by target genes, miRNAs and TFs.

Figure 3. Biology-inspired experiments with the mathematical model led to the identification of a minimal model of gene dosage compensation for *MYC* and *STAT3*. A. A parameter scan on *MYC* and *STAT3* copy numbers identifies which molecules increase or decrease together with varying copy numbers of those two genes. **B.** The single inhibition of those varying molecules has no effect on the gene dosage compensation of *MYC* and *STAT3*. **C.** The inhibition of groups of molecules and the restoration of single miRNAs indicate a redundant mechanism for the gene dosage compensation of *MYC* and *STAT3*. **D.** A minimal model of gene dosage compensation for *MYC* and *STAT3* is hypothesized based on the results of these experiments. **E.** The fitting and parameter scan demonstrates that this minimal model recapitulates the dosage compensation of *MYC* and *STAT3*.

Figure 4. Overexpression of exogeneous *MYC* decreases endogenous *MYC* and increases expression of the predicted miRNA in human glioblastoma NCH82 cells. A. Vector map of the construct pSBbi-moxGFP-P2A-T2A-c*MYC*-Puro for the expression of (exo)*MYC* with a deletion of the 3'UTR region to avoid miRNA regulation to induce induce miRNA overexpression and the subsequent down-regulation of endogenous *MYC*. **B.** Upper part: Representative dot plots showing PI and GFP intensities of transfected NCH82 cells 24, 48 and 72 h after transfection. Left panel: cells transfected with the control GFP-plasmid (GFP-Ctrl); right panel: cells transfected with the (exo)*MYC* plasmid (GFP-(exo)*MYC*). Quadrants from upper left to lower left: Q1: PI+/GFP-; Q2: PI+/GFP+; Q3: PI-/GFP+; Q4: PI-/GFP-. Lower part: Statistical analysis of flow cytometric measurements of the PI+/GFP+ and PI-/GFP+ populations of transfected cells. Marked with * are statistically significant changes of total GFP+ cells in the GFP-(exo)*MYC* transfected cells compared to GFP-Ctrl transfected cells. (n = 6); **C.** Quantitative RT-PCR detection of *MYC* expression in the PI-/GFP+ population of GFP-(exo)*MYC* transfected cells relative to the PI-/GFP+ population of GFP-Ctrl transfected cells. (n = 3); **D.** Quantitative RT-PCR detection of hsa-miR-17, hsa-miR-19a and hsa-miR-20a in the PI-/GFP+ population of GFP-(exo)*MYC* transfected cells relative to the PI-/GFP+ population of GFP-Ctrl transfected cells. (n = 3); t-tests were performed to analyze means of samples with a significance defined by an α of 0.05. The resulting p-values were corrected for multiple comparisons using the Holm-Sidak method.

Figure EV4 – Characterization of the human GB cells overexpressing exogeneous *MYC*. A. Analysis on a 2% agarose gel of amplicons after quantitative RT-PCR. Expected product sizes were 216 bp (exogeneous *MYC*) and 404 bp (endogenous *MYC*). Lane 1: 100 bp ladder; lanes 2-4: untransfected, GFP-Ctrl transfected and GFP-(exo)*MYC* transfected NCH82 cDNAs analyzed using exogenous *MYC* primers; lanes 5-7: untransfected, GFP-Ctrl transfected and GFP-(exo)*MYC* transfected NCH82 cDNAs analyzed using endogenous *MYC* primers. **B.** Representative microscopic images of transfected NCH82 cells 24, 48 and 72 h after transfection. Left panel: cells transfected with the GFP-Ctrl plasmid; right panel: cells transfected with the GFP-(exo)*MYC* plasmid. Scale bars: 200 µm. **C.** Quantitative RT-PCR detection of (exo)*MYC* expression in the *MYC* positive population relative to the *MYC* negative population after sorting of GFP-(exo)*MYC* transfected cells. (n = 3); t-tests were performed to analyze means of samples with a significance defined by an α of 0.05. The

resulting p-values were corrected for multiple comparisons using the Holm-Sidak method.

Figure 5. The inhibition of *MYC* dosage compensation induces dosage-dependent cytotoxity. A. The in silico simulations for the dependence of *MYC* concentration on the amounts of the 3 compensating miRNAs suggest that *MYC* dosage compensation in cancer with *MYC* amplification is more sensitive to the inhibition of those miRNAs. **B.** The experimental inhibition of those 3 miRNAs in colon cancer cells with 3 different copy numbers of *MYC* suggest a higher sensitivity to the inhibition of gene dosage compensation for colon cancer cell lines with *MYC* amplification (fluorescent+brightfield microscopy including the result of the image analysis pipeline in CellProfiler). **C.** The quantification of dead cells confirms a significant difference in cell death in cells treated with antimiRs compared to cells treated with the negative control RNA for HT29 and SW-620 cells (n=3, the differences becomes significant after 36 hours of incubation for HT29 cells and after 24 hours for the SW-620 cells with a $p<0.01$).

Figure 6. The gene dosage compensation depends on the kinetic parameters of the interactions of basic regulatory network motifs. A. Rationale of a biocomputational pipeline for the identification of motifs mediating gene dosage compensation. **B.** Evaluation of gene dosage compensation for the networks containing feedback and feedforward loops for pgr, *STAT5A* and *STAT5B*. **C.** The postulated network topologies of the minimal models of gene dosage compensation include 1 feedforward loop for *STAT5A* and 3 feedforward loops for *STAT5B*. **D.** These minimal models recapitulate the gene dosage compensation for *STAT5A* and *STAT5B*. **E.** A three dimensional landscape of gene dosage compensation dependent on the values of ks (synthesis rate of the miRNA), ka (activation parameter of the TF) and kr (repression parameter of the miRNA) demonstrates that the gene dosage compensation mediated by some basic regulatory network motifs depends on the values of these kinetic parameters. Individual cases of regulatory motifs are located within this regulatory landscape.

Figure 7. Breast Cancer patients show heterogeneity in gene dosage compensation leading to differences in patient survival. A. Assessment of gene dosage compensation by parameter scans of the gene copy numbers of *MYC*, *STAT3*, *STAT5A* and *STAT5B* for each individual patient-tailored model. **B.** Patient case separation by the amplifications and compensation behavior for *MYC*, *STAT5A* and *STAT5B*. **C.** Survival probabilities for the patient cases divided according to the amplification and compensation behavior of *MYC*. **D.** Compensation landscape of model parameters for the models of the TCGA breast cancer patients. **E.** The interpolation of each individual patient model within the compensation landscape indicates that miR-19a mediates *MYC* dosage compensation in most breast cancer patients.

SUPPLEMENTARY SECTION

A miRNA/TF interaction network connects all candidate genes but its structural analysis reveals that only those with transcription factor function can be subjected to gene dosage compensation

We performed a correlation analysis of the copy numbers of target genes with the expression levels of miRNAs and transcription factors in order to identify possible regulators responding to the dosage of our candidate genes. We calculated the correlation coefficients between the Z-scores of copy number variations of the candidate genes and the Z-scores of miRNA/TF expression data across the NCI60 panel. As depicted in figure 2A, there are both miRNAs and TF with positive or negative correlation to the candidate copy numbers (figure 1A), suggesting that many miRNAs and TFs are potentially involved in the regulation of the candidate expression in response to their gene dosage.

We asked whether this potential regulation is exerted by direct or indirect interactions. To evaluate whether there are any reported or predicted connections between these candidate genes and miRNAs/TF, we searched for putative and validated interactions between our candidate genes and miRNA/transcription factors in several databases: Mirtarbase, MiRBase, Pazar, TRED and the study of Neph et al (Neph *et al*, 2012). We also included the database of Human Transcriptional Regulation interaction (HTRI, http://www.lbbc.ibb.unesp.br/htri), which is a repository of about 50.000 interactions between transcription factors and target genes, including the experimental techniques and references used to validate this interactions (Bovolenta et al., 2012). After filtering the correlation matrix of figure 1A by these reported interactions, we obtained the correlation coefficients for direct gene-miRNA and gene-TF interactions as shown in figure 2SB. Within those remaining correlations, we found that the highest positive correlation with miRNAs is only 0.34 for [(ZNF217 copy number) to (miR26b expression)] and the highest anti-correlation with TFs is -0.63 for [(*STAT3* copy number) to (HOXB13 expression)] followed by -0.5 for [(*MYC* copy number) to (*STAT3* expression)]. These results suggest that most correlations and anti-correlations between copy number and miRNA/TF expression should be due to indirect interactions; it is unlikely that the direct interactions alone would be responsible for gene dosage compensation of their corresponding candidate genes given those fairly low correlations.

A mechanism of gene dosage compensation will require a way to “sense” the changes in gene copy number and a type of feedback to exert the compensation. Therefore, we asked if those many miRNAs and TFs may form regulatory feedback loops to potentially compensate the dosage of our candidate genes. First, we constructed a network of putative and validated regulatory interactions based on the information available on the miRNA-target and TF-target databases cited above. In addition, we included the interactions between miRNAs and TFs using the information available at Transmir (Tong *et al*, 2019) and CircuitsDB 2 (Friard *et al*,

2010b) . In order to construct the network topology we wrote an algorithm that takes as input data the list of all interactions I (experimental or putative) and the list of gene targets. The output obtained includes a list of nodes N, a list of arcs A and a list of regulators R. To build R, the algorithm considers each arc X to Y, where X is a miRNA or a TF and Y is one of the target genes, including thereby all the regulators X of the target genes in R and all the arcs X to Y in A. Also, if the target gene is a TF, it is included in R. Next, the algorithm will add the arcs between the regulators in R. For each pair of regulators in R (r[i], r[j]), an arc is added to A if r[i] to r[j] exists in the list of interactions I. This generated a network with 612 nodes and 47122 interactions, which connected all 21 candidate target genes with 428 TFs and 163 miRNAs (Figure 2B).

In order to examine the target/miRNA/TF network for the presence of regulatory interactions, we searched for motifs with systems-level properties including positive and negative feedback loops (between miRNAs and TFs), coherent feedforward loops and incoherent feedforward loops (Vera *et al*, 2013). In order to quantify these interactions we wrote an algorithm to search for cycles (materials and methods).

We identified a total number of 5054 putative regulatory motifs. For example, miR-15A participates in a coherent feedforward loop inhibiting both SEMAD3 and BRCA1, a positive regulator of BRCA1. Also, BRCA1 forms an incoherent feedforward loop, activating the transcription of SEMAD3 but also miR195, a negative regulator of SEMAD3. In addition, miR15A forms a negative feedback loop with BRCA1 (Figure 2C). In total, this network includes 156 negative feedback loops between TFs and miRNAs, 2337 coherent feedforward loops and 2561 incoherent feedforward loops formed by the interaction of TFs and miRNAs with the target genes. The analysis of this network topology indicates that several putative regulatory loops link all the candidate target genes.

The regulatory motifs with potential systems-level properties to mediate dosage compensation are widely present within this putative network. However, from the construction of this network topology is easy to infer that this network could sense changes in copy number of target genes only if they also have a TF-function (red colored interactions, figure 2C). Otherwise, the remaining target genes are dead ends of the network, having only input arcs but no outputs. Among the 21 candidate target genes, only 7 of them correspond to transcription factors: *MYC*, foxc1, foxf2, *STAT3*, *STAT5A*, *STAT5B* and znf217. This structural analysis of the network topology indicates that for the remaining genes, the current network of interactions does not include a direct regulatory loop able to explain a potential dosage compensation. However, the gene dosage compensation of candidates with TF function might also impact the expression levels of other genes.

For the candidate genes with TF function, the identification of many potential regulatory loops offers a putative explanation of gene dosage compensation. Indeed, *MYC*, znf217 and *STAT3* are the most connected genes in the network, indicating that they participate in

multiple regulatory loops. These results showed that a putative network of target-miRNA-TF interactions connects all the compensated genes but only those with TF function might be able to trigger a signal and get a compensatory response from that network. Due to the complexity of this regulatory network, a quantitative systems-level approach is required to identify regulatory loops mediating gene dosage compensation.

References

Avalle, L., Camporeale, A., Camperi, A., & Poli, V. (2017). STAT3 in cancer: A double edged sword. *Cytokine*. https://doi.org/10.1016/j.cyto.2017.03.018

Birchler, J. a, & Veitia, R. a. (2012). Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(37), 14746–14753. https://doi.org/10.1073/pnas.1207726109

Bleris, L., Xie, Z., Glass, D., Adadey, A., Sontag, E., & Benenson, Y. (2011). Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Molecular Systems Biology*, *7*(519), 1–12. https://doi.org/10.1038/msb.2011.49

Blower, P. E., Verducci, J. S., Lin, S., Zhou, J., Chung, J.-H., Dai, Z., … Sadee, W. (2007). MicroRNA expression profiles for the NCI-60 cancer cell panel. *Molecular Cancer Therapeutics*, *6*(5), 1483–1491. https://doi.org/10.1158/1535-7163.MCT-07-0009

Bond, U., Neal, C., Donnelly, D., & James, T. C. (2004). Aneuploidy and copy number breakpoints in the genome of lager yeasts mapped by microarray hybridisation. *Current Genetics*, *45*(6), 360–370. https://doi.org/10.1007/s00294-004-0504-x

Brennan, C. M., Vaites, L. P., Wells, J. N., Santaguida, S., Paulo, J. A., Storchova, Z., … Amon, A. (2019). Protein aggregation mediates stoichiometry of protein complexes in aneuploid cells. *Genes & Development*, *33*(15–16), 1031–1047. https://doi.org/10.1101/gad.327494.119

Bussey, K. J., Chin, K., Lababidi, S., Reimers, M., Reinhold, W. C., Kuo, W.-L., … Weinstein, J. N. (2006). Integrating data on DNA copy number with gene expression levels and drug sensitivities in the NCI-60 cell line panel. *Molecular Cancer Therapeutics*, *5*(4), 853–867. https://doi.org/10.1158/1535-7163.MCT-05-0155

Carignano, A., Mukherjee, S., Singh, A., & Seelig, G. (2019). Extrinsic Noise Suppression in Micro RNA Mediated Incoherent Feedforward Loops. *Proceedings of the IEEE Conference on Decision and Control*. https://doi.org/10.1109/CDC.2018.8619371

Cimini, D. (2008). Merotelic kinetochore orientation, aneuploidy, and cancer. *Biochimica et Biophysica Acta - Reviews on Cancer*. https://doi.org/10.1016/j.bbcan.2008.05.003

Costantini, L. M., Baloban, M., Markwardt, M. L., Rizzo, M., Guo, F., Verkhusha, V. V., & Snapp, E. L. (2015). A palette of fluorescent proteins optimized for diverse cellular environments. *Nature Communications*. https://doi.org/10.1038/ncomms8670

Dang, C. V. (1999). c-Myc Target Genes Involved in Cell Growth, Apoptosis, and Metabolism. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.19.1.1

Devlin, R. H., Holm, D. G., & Grigliatti, T. a. (1982). Autosomal dosage compensation Drosophila melanogaster strains trisomic for the left arm of chromosome 2. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(4), 1200–1204. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=345929&tool=pmcentrez&rendertype=abstract

Donnelly, N., & Storchová, Z. (2014). Dynamic karyotype, dynamic proteome: buffering the effects of aneuploidy. *Biochimica et Biophysica Acta*, *1843*(2), 473–481. https://doi.org/10.1016/j.bbamcr.2013.11.017

Duesberg, P., Rausch, C., Rasnick, D., & Hehlmann, R. (1998). Genetic instability of cancer cells is proportional to their degree of aneuploidy. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(23), 13692–13697. https://doi.org/10.1073/pnas.95.23.13692

Fabarius, A., Li, R., Yerganian, G., Hehlmann, R., & Duesberg, P. (2008). Specific clones of spontaneously evolving karyotypes generate individuality of cancers. *Cancer Genetics and Cytogenetics*, *180*(2), 89–99. https://doi.org/10.1016/j.cancergencyto.2007.10.006

Fabarius, A., Willer, A., Yerganian, G., Hehlmann, R., & Duesberg, P. (2002). Specific aneusomies in Chinese hamster cells at different stages of neoplastic transformation, initiated by nitrosomethylurea. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(10), 6778–6783. https://doi.org/10.1073/pnas.251670699

Fabian, M. R., Sonenberg, N., & Filipowicz, W. (2010). Regulation of mRNA translation and stability by microRNAs. *Annual Review of Biochemistry*, *79*, 351–379. https://doi.org/10.1146/annurev-biochem-060308-103103

Friard, O., Re, A., Taverna, D., De Bortoli, M., & Corá, D. (2010). CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. *BMC Bioinformatics*, *11*, 435. https://doi.org/10.1186/1471-2105-11-435

Fuziwara, C. S., & Kimura, E. T. (2015). Insights into regulation of the miR-17-92 cluster of miRNAs in cancer. *Frontiers in Medicine*, *2*(SEP), 13–17. https://doi.org/10.3389/fmed.2015.00064

Gholami, A. M., Hahne, H., Wu, Z., Auer, F., Meng, C., Wilhelm, M., & Kuster, B. (2013). Global proteome analysis of the NCI-60 cell line panel. *Cell Reports*, *4*(3), 609–620. https://doi.org/10.1016/j.celrep.2013.07.018

Gmeiner, W. H., Reinhold, W. C., & Pommier, Y. (2010). Genome-wide mRNA and

microRNA profiling of the NCI 60 cell-line screen and comparison of FdUMP[10] with fluorouracil, floxuridine, and topoisomerase 1 poisons. *Molecular Cancer Therapeutics*, *9*(12), 3105–3114. https://doi.org/10.1158/1535-7163.MCT-10-0674

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell*, *144*(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013

Hanna, J., Hossain, G. S., & Kocerha, J. (2019). The potential for microRNA therapeutics and clinical research. *Frontiers in Genetics*, *10*(MAY). https://doi.org/10.3389/fgene.2019.00478

Hardy, K., & Hardy, P. J. (2015). 1(st) trimester miscarriage: four decades of study. *Translational Pediatrics*. https://doi.org/10.3978/j.issn.2224-4336.2015.03.05

Heard, E., Clerc, P., & Avner, P. (1997). X-chromosome inactivation in mammals. *Annual Review of Genetics*. https://doi.org/10.1146/annurev.genet.31.1.571

Herranz, H., & Cohen, S. M. (2010). MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems. *Genes & Development*, *24*(13), 1339–1344. https://doi.org/10.1101/gad.1937010

Hooke, R., & Jeeves, T. A. (1961). "Direct Search" Solution of Numerical and Statistical Problems. *Journal of the ACM (JACM)*. https://doi.org/10.1145/321062.321069

Hose, J., Yong, C. M., Sardi, M., Wang, Z., Newton, M. A., & Gasch, A. P. (2015). Dosage compensation can buffer copynumber variation in wild yeast. *ELife*. https://doi.org/10.7554/eLife.05462

Hsu, S. Da, Tseng, Y. T., Shrestha, S., Lin, Y. L., Khaleel, A., Chou, C. H., … Huang, H. Da. (2014). MiRTarBase update 2014: An information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Research*, *42*(D1). https://doi.org/10.1093/nar/gkt1266

Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., … Wang, J. (2003). The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btg015

Ishikawa, K., Makanae, K., Iwasaki, S., Ingolia, N. T., & Moriya, H. (2017). Post-Translational Dosage Compensation Buffers Genetic Perturbations to Stoichiometry of Protein Complexes. *PLoS Genetics*, *13*(1). https://doi.org/10.1371/journal.pgen.1006554

Jiang, C., Xuan, Z., Zhao, F., & Zhang, M. Q. (2007). TRED: A transcriptional regulatory element database, new entries and other development. *Nucleic Acids Research*, *35*(SUPPL. 1). https://doi.org/10.1093/nar/gkl1041

Jo, D. H., Kim, J. H., Cho, C. S., Cho, Y. L., Jun, H. O., Yu, Y. S., … Kim, J. H. (2014). STAT3 inhibition suppresses proliferation of retinoblastomathrough down-regulation of positive feedback loop of STAT3/miR-17-92 clusters. *Oncotarget*.

https://doi.org/10.18632/oncotarget.2546

Johnston, P. A., & Grandis, J. R. (2011). STAT3 signaling: Anticancer strategies and challenges. *Molecular Interventions*. https://doi.org/10.1124/mi.11.1.4

Karcher, S., Steiner, H. H., Ahmadi, R., Zoubaa, S., Vasvari, G., Bauer, H., … Herold-Mende, C. (2006). Different angiogenic phenotypes in primary and secondary glioblastomas. *International Journal of Cancer*. https://doi.org/10.1002/ijc.21648

Kim, D., Sung, Y. M., Park, J., Kim, S., Kim, J., Park, J., … Baek, D. (2016). General rules for functional microRNA targeting. *Nature Genetics*, *48*(12), 1517–1526. https://doi.org/10.1038/ng.3694

Kojima, S., & Cimini, D. (2019). Aneuploidy and gene expression: Is there dosage compensation? *Epigenomics*, Vol. 11, pp. 1827–1837. https://doi.org/10.2217/epi-2019-0135

Kowarz, E., Löscher, D., & Marschalek, R. (2015). Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. *Biotechnology Journal*. https://doi.org/10.1002/biot.201400821

Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2019). MiRBase: From microRNA sequences to function. *Nucleic Acids Research*, *47*(D1), D155–D162. https://doi.org/10.1093/nar/gky1141

Kozomara, A., & Griffiths-Jones, S. (2014). MiRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, *42*(D1). https://doi.org/10.1093/nar/gkt1181

Kvitek, D. J., Will, J. L., & Gasch, A. P. (2008). Variations in stress sensitivity and genomic expression in diverse S. cerevisiae isolates. *PLoS Genetics*, *4*(10), 31–35. https://doi.org/10.1371/journal.pgen.1000223

Lai, X., Bhattacharya, A., Schmitz, U., Kunz, M., Vera, J., & Wolkenhauer, O. (2013). A systems' biology approach to study microrna-mediated gene regulatory networks. *BioMed Research International*, *2013*(Ii). https://doi.org/10.1155/2013/703849

Levy, D. E., & Lee, C. K. (2002). What does Stat3 do? *Journal of Clinical Investigation*. https://doi.org/10.1172/JCI0215650

Li, L., McCormack, A. a, Nicholson, J. M., Fabarius, A., Hehlmann, R., Sachs, R. K., & Duesberg, P. H. (2009). Cancer-causing karyotypes: chromosomal equilibria between destabilizing aneuploidy and stabilizing selection for oncogenic function. *Cancer Genetics and Cytogenetics*, *188*(1), 1–25. https://doi.org/10.1016/j.cancergencyto.2008.08.016

Li, Y., Li, Y., Zhang, H., & Chen, Y. (2011). MicroRNA-mediated positive feedback loop and optimized bistable switch in a cancer network Involving miR-17-92. *PloS One*, *6*(10), e26302. https://doi.org/10.1371/journal.pone.0026302

Lindow, M., & Kauppinen, S. (2012). Discovering the first microRNA-targeted drug. *The Journal of Cell Biology*, *199*(3), 407–412. https://doi.org/10.1083/jcb.201208082

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods*. https://doi.org/10.1006/meth.2001.1262

Lucchesi, J. C., & Kuroda, M. I. (2015). Dosage compensation in drosophila. *Cold Spring Harbor Perspectives in Biology*. https://doi.org/10.1101/cshperspect.a019398

Matsuo, K., Eno, M. L., Im, D. D., Rosenshein, N. B., & Sood, A. K. (2010). Gynecologic Oncology Clinical relevance of extent of extreme drug resistance in epithelial ovarian carcinoma. *Gynecologic Oncology*, *116*(1), 61–65. https://doi.org/10.1016/j.ygyno.2009.09.018

Meyer, B. J., McDonel, P., Csankovszki, G., & Ralston, E. (2004). Sex and X-chromosome-wide repression in Caenorhabditis elegans. *Cold Spring Harbor Symposia on Quantitative Biology*. https://doi.org/10.1101/sqb.2004.69.71

Meyer, Barbara J. (2005). X-Chromosome dosage compensation. *WormBook : The Online Review of C. Elegans Biology*. https://doi.org/10.1895/wormbook.1.8.1

Mihailovich, M., Bremang, M., Spadotto, V., Musiani, D., Vitale, E., Varano, G., … Bonaldi, T. (2015). MiR-17-92 fine-tunes MYC expression and function to ensure optimal B cell lymphoma growth. *Nature Communications*. https://doi.org/10.1038/ncomms9725

Nilsson, J. A., & Cleveland, J. L. (2003). Myc pathways provoking cell suicide and cancer. *Oncogene*. https://doi.org/10.1038/sj.onc.1207261

Osella, M., Bosia, C., Corá, D., & Caselle, M. (2011). The role of incoherent microRNA-mediated feedforward loops in noise buffering. *PLoS Computational Biology*. https://doi.org/10.1371/journal.pcbi.1001101

Ozery-Flato, M., Linhart, C., Trakhtenbrot, L., Izraeli, S., & Shamir, R. (2011). Large-scale analysis of chromosomal aberrations in cancer karyotypes reveals two distinct paths to aneuploidy. *Genome Biology*, *12*(6), R61. https://doi.org/10.1186/gb-2011-12-6-r61

Portales-Casamar, E., Arenillas, D., Lim, J., Swanson, M. I., Jiang, S., McCallum, A., … Wasserman, W. W. (2009). The PAZAR database of gene regulatory information coupled to the ORCA toolkit for the study of regulatory sequences. *Nucleic Acids Research*, *37*(SUPPL. 1). https://doi.org/10.1093/nar/gkn783

Prendergast, G. C. (1999). Mechanisms of apoptosis by c-Myc. *Oncogene*. https://doi.org/10.1038/sj.onc.1202727

Ritchie, W., Rasko, J. E. J., & Flamant, S. (2013). MicroRNA target prediction and validation. *Advances in Experimental Medicine and Biology*, *774*, 39–53. https://doi.org/10.1007/978-94-007-5590-1_3

Rupaimoole, R., & Slack, F. J. (2017). MicroRNA therapeutics: Towards a new era for the management of cancer and other diseases. *Nature Reviews Drug Discovery*, *16*(3), 203–221. https://doi.org/10.1038/nrd.2016.246

Shankavaram, U. T., Reinhold, W. C., Nishizuka, S., Major, S., Morita, D., Chary, K. K., … Weinstein, J. N. (2007). Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. *Molecular Cancer Therapeutics*, *6*(3), 820–832. https://doi.org/10.1158/1535-7163.MCT-06-0650

Sheltzer, J. M., & Amon, A. (2011). The aneuploidy paradox: costs and benefits of an incorrect karyotype. *Trends in Genetics*, *27*(11), 446–453. https://doi.org/10.1016/j.tig.2011.07.003

Shimoga, V., White, J. T., Li, Y., Sontag, E., & Bleris, L. (2014). Synthetic mammalian transgene negative autoregulation. *Molecular Systems Biology*, *9*(1), 670–670. https://doi.org/10.1038/msb.2013.27

Shoemaker, R. H. (2006). The NCI60 human tumour cell line anticancer drug screen. *Nature Reviews. Cancer*, *6*(10), 813–823. https://doi.org/10.1038/nrc1951

Solé, R. V, & Deisboeck, T. S. (2004). An error catastrophe in cancer? *Journal of Theoretical Biology*, *228*(1), 47–54. https://doi.org/10.1016/j.jtbi.2003.08.018

Stingele, S., Stoehr, G., Peplowska, K., Cox, J., Mann, M., & Storchova, Z. (2012). Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. *Molecular Systems Biology*, *8*(608), 608. https://doi.org/10.1038/msb.2012.40

Uribesalgo, I., Benitah, S. A., & Di Croce, L. (2012). From oncogene to tumor suppressor: The dual role of Myc in leukemia. *Cell Cycle*. https://doi.org/10.4161/cc.19883

Veitia, R. A., Bottani, S., & Birchler, J. A. (2008). *Cellular reactions to gene dosage imbalance : genomic , transcriptomic and proteomic effects*. (June). https://doi.org/10.1016/j.tig.2008.05.005

Vera, J., Lai, X., Schmitz, U., & Wolkenhauer, O. (2013). MicroRNA-regulated networks: the perfect storm for classical molecular biology, the ideal scenario for systems biology. *Advances in Experimental Medicine and Biology*, *774*, 55–76. https://doi.org/10.1007/978-94-007-5590-1_4

Wang, J., Lu, M., Qiu, C., & Cui, Q. (2010). TransmiR: a transcription factor-microRNA regulation database. *Nucleic Acids Research*, *38*(Database issue), D119-22. https://doi.org/10.1093/nar/gkp803

Weaver, B. A., & Cleveland, D. W. (2006). Does aneuploidy cause cancer? *Current Opinion in Cell Biology*. https://doi.org/10.1016/j.ceb.2006.10.002

Yue, P., & Turkson, J. (2009). Targeting STAT3 in cancer: How successful are we? *Expert Opinion on Investigational Drugs*. https://doi.org/10.1517/13543780802565791

Figure 1. Identification of candidate genes under dosage compensation in the NCI60 panel. A. Input data of gene copy number (DNA), gene expression (RNA) and protein levels (protein) of the NCI60 panel. The absolute values are shown on the left panel and the right panel corresponds to the log2 values normalized to the averaged RNA and protein of the diploid cell lines for the respective gene (Normalized to diploid). B. Standard deviations (SD) of the DNA, RNA and Protein levels for each gene across the 59 cell lines of the NCI60 panel. C. Gaussian Mixture Model to identify a cluster of subpopulation of genes with high SD DNA and low SD RNA and/or low SD Protein (white arrow). D. Gene Clustering according to the model in C (left). The cyan cluster contains candidate genes under dosage compensation, characterized by high SD DNA, low SD Protein (middle) and low SD RNA (right). E. Absolute and Normalized values of selected candidate genes under dosage compensation. F. Examples of candidate genes under dosage compensation (*MYC* and rab5c). G. Clusters of candidate genes according to their copy number variations across the 59 cell lines of the NCI60 panel.

Figure EV1. DNA, RNA and protein levels for all candidate genes.

Figure 2. A mathematical model driven by NCI-60 data leads to a phenotype of gene dosage compensation for *MYC* and *STAT3*. A. Network of miRNA-TF interactions for *MYC*, *STAT3*, *STAT5A* and *STAT5B*. **B.** Schematic representations of the modeled interactions for the target gene, miRNAs and associated TFs. **C.** The results of the parameter estimation task to fit the ODE mathematical model to the experimental data upon different modifications led to a reduction in the objective function (difference between modeled and experimental data). **D.** The graphical comparison of modeled and experimental data shows no significant differences. **E.** Results of parameter scan on the values of copy number for the candidate target genes showing gene dosage compensation for *MYC* and *STAT3*. **F.** The sensitivity analysis points to some candidate regulatory molecules controlling the concentrations of *MYC* and *STAT3*.

Figure EV2. Putative miRNA-transcription factor interaction network linking all candidate target genes. A. Correlation coefficients of copy number variations of the candidate genes and miRNA/transcription factor expression across the NCI60 panel. **B.** Correlation coefficients of copy number variations of the candidate genes and miRNA/transcription factor expression across the NCI60 panel, only for those putative interactions present on the network in C. **C.** Network of putative interactions among target genes, miRNAs and transcription factors (red indicates repression and blue activation) including different types of regulatory loops formed by target genes, miRNAs and TFs.

Figure 3. Biology-inspired experiments with the mathematical model led to the identification of a minimal model of gene dosage compensation for *MYC* and *STAT3*. A. A parameter scan on *MYC* and *STAT3* copy numbers identifies which molecules increase or decrease together with varying copy numbers of those two genes. **B.** The single inhibition of those varying molecules has no effect on the gene dosage compensation of *MYC* and *STAT3*. **C.** The inhibition of groups of molecules and the restoration of single miRNAs indicate a redundant mechanism for the gene dosage compensation of *MYC* and *STAT3*. **D.** A minimal model of gene dosage compensation for *MYC* and *STAT3* is hypothesized based on the results of these experiments. **E.** The fitting and parameter scan demonstrates that this minimal model recapitulates the dosage compensation of *MYC* and *STAT3*.

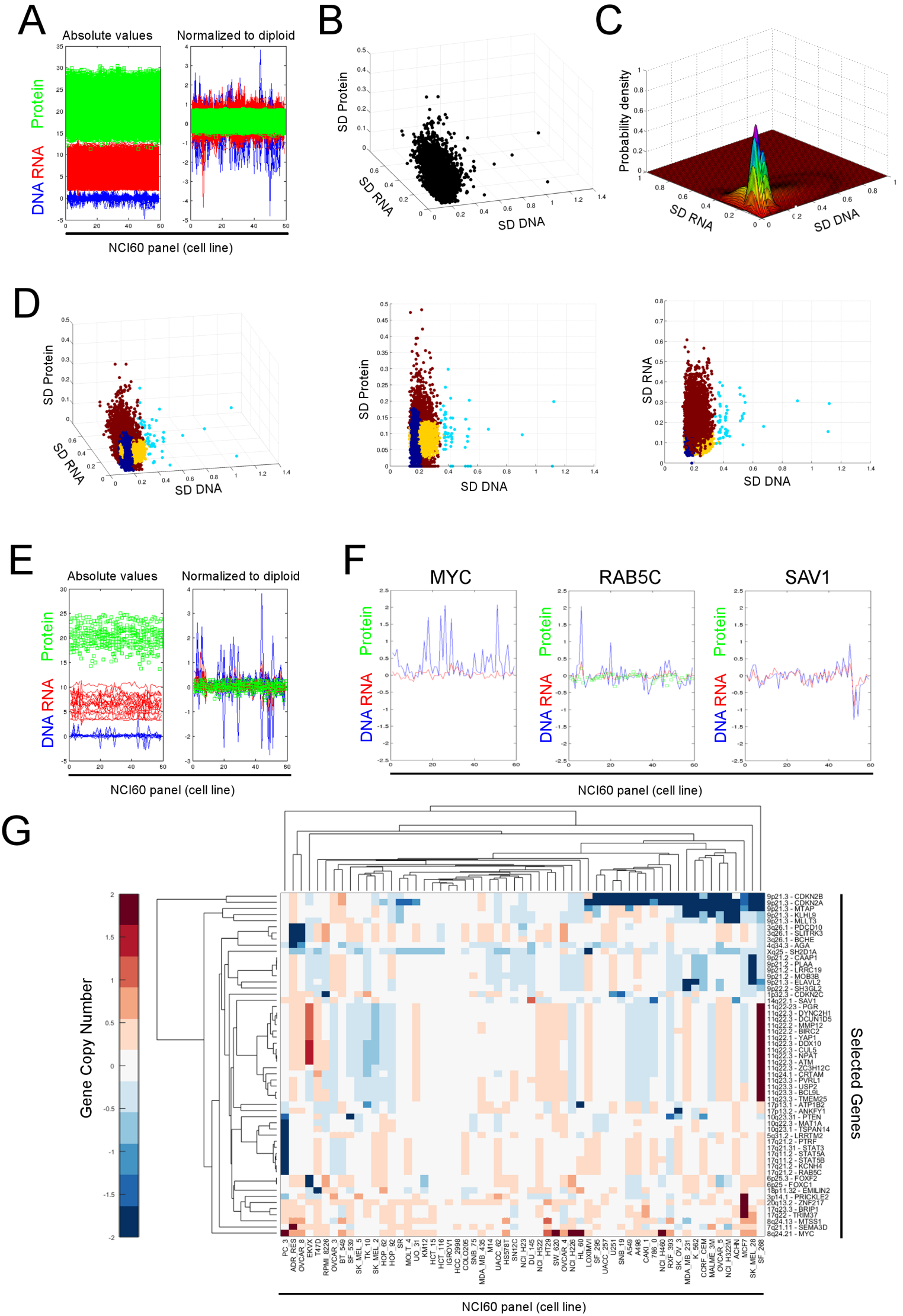
Figure 4. Overexpression of exogeneous *MYC* decreases endogenous *MYC* and increases expression of the predicted miRNA in human glioblastoma NCH82 cells. A. Vector map of the construct pSBbi-moxGFP-P2A-T2A-c*MYC*-Puro for the expression of (exo)*MYC* with a deletion of the 3'UTR region to avoid miRNA regulation to induce induce miRNA overexpression and the subsequent down-regulation of endogenous *MYC*. **B.** Upper part: Representative dot plots showing PI and GFP intensities of transfected NCH82 cells 24, 48 and 72 h after transfection. Left panel: cells transfected with the control GFP-plasmid (GFP-Ctrl); right panel: cells transfected with the (exo)*MYC* plasmid (GFP-(exo)*MYC*). Quadrants from upper left to lower left: Q1: PI+/GFP-; Q2: PI+/GFP+; Q3: PI-/GFP+; Q4: PI-/GFP-. Lower part: Statistical analysis of flow cytometric measurements of the PI+/ GFP+ and PI-/GFP+ populations of transfected cells. Marked with * are statistically significant changes of total GFP+ cells in the GFP-(exo)*MYC* transfected cells compared to GFP-Ctrl transfected cells. (n = 6); **C.** Quantitative RT-PCR detection of *MYC* expression in the PI-/GFP+ population of GFP-(exo)*MYC* transfected cells relative to the PI-/GFP+ population of GFP-Ctrl transfected cells. (n = 3); **D.** Quantitative RT-PCR detection of hsa-miR-17, hsa-miR-19a and hsa-miR-20a in the PI-/GFP+ population of GFP-(exo)*MYC* transfected cells relative to the PI-/GFP+ population of GFP-Ctrl transfected cells. (n = 3); t-tests were performed to analyze means of samples with a significance defined by an α of 0.05. The resulting p-values were corrected for multiple comparisons using the Holm-Sidak method.

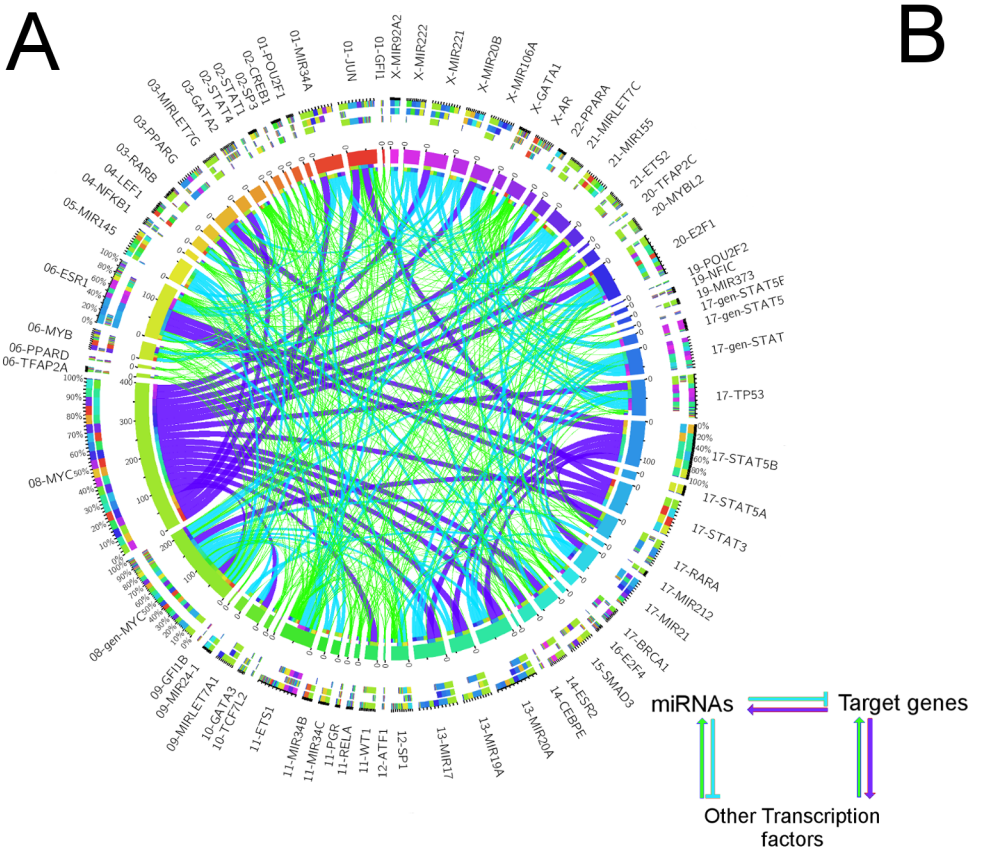
Figure EV4 – Characterization of the human GB cells overexpressing exogeneous *MYC*. A. Analysis on a 2% agarose gel of amplicons after quantitative RT-PCR. Expected product sizes were 216 bp (exogeneous *MYC*) and 404 bp (endogenous *MYC*). Lane 1: 100 bp ladder; lanes 2-4: untransfected, GFP-Ctrl transfected and GFP-(exo)*MYC* transfected NCH82 cDNAs analyzed using exogenous *MYC* primers; lanes 5-7: untransfected, GFP-Ctrl transfected and GFP-(exo)*MYC* transfected NCH82 cDNAs analyzed using endogenous *MYC* primers. **B.** Representative microscopic images of transfected NCH82 cells 24, 48 and 72 h after transfection. Left panel: cells transfected with the GFP-Ctrl plasmid; right panel: cells transfected with the GFP-(exo)*MYC* plasmid. Scale bars: 200 µm. **C.** Quantitative RT-PCR detection of (exo)*MYC* expression in the *MYC* positive population relative to the *MYC* negative population after sorting of GFP-(exo)*MYC* transfected cells. (n = 3); t-tests were performed to analyze means of samples with a significance defined by an α of 0.05. The resulting p-values were corrected for multiple comparisons using the Holm-Sidak method.

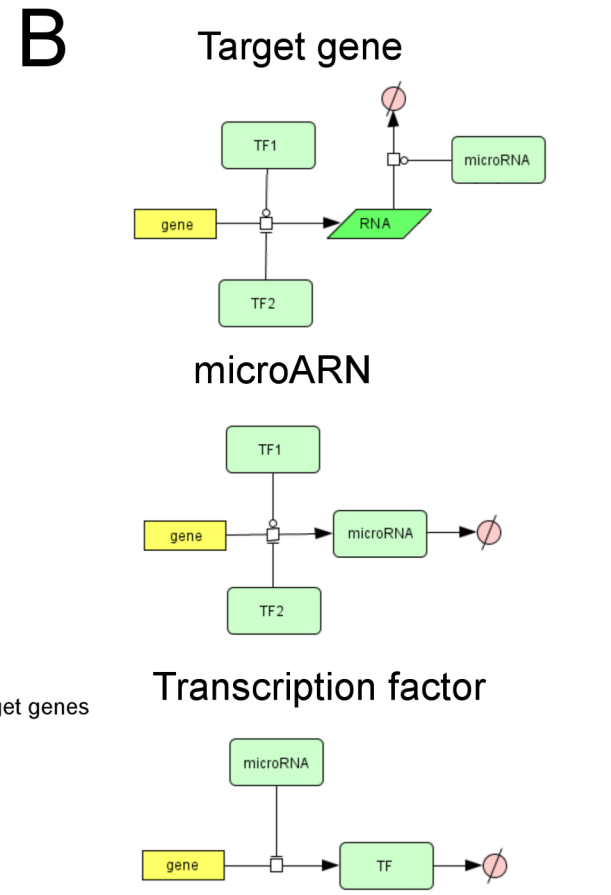
Figure 5. The inhibition of *MYC* dosage compensation induces dosage-dependent cytotoxity. A. The in silico simulations for the dependence of *MYC* concentration on the amounts of the 3 compensating miRNAs suggest that *MYC* dosage compensation in cancer with *MYC* amplification is more sensitive to the inhibition of those miRNAs. **B.** The experimental inhibition of those 3 miRNAs in colon cancer cells with 3 different copy numbers of *MYC* suggest a higher sensitivity to the inhibition of gene dosage compensation for colon cancer cell lines with *MYC* amplification (fluorescent+brightfield microscopy including the result of the image analysis pipeline in CellProfiler). **C.** The quantification of dead cells confirms a significant difference in cell death in cells treated with antimiRs compared to cells treated with the negative control RNA for HT29 and SW-620 cells (n=3, the differences becomes significant after 36 hours of incubation for HT29 cells and after 24 hours for the SW-620 cells with a $p<0.01$).

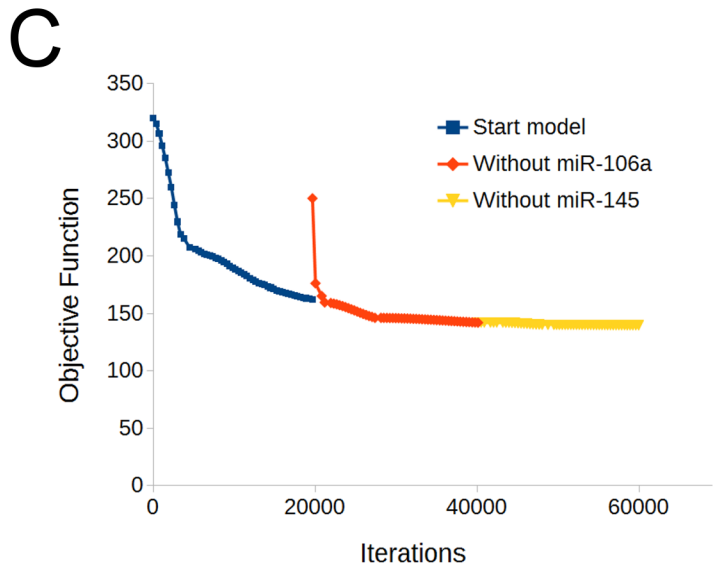
Figure 6. The gene dosage compensation depends on the kinetic parameters of the interactions of basic regulatory network motifs. A. Rationale of a biocomputational pipeline for the identification of motifs mediating gene dosage compensation. **B.** Evaluation of gene dosage compensation for the networks containing feedback and feedforward loops for pgr, *STAT5A* and *STAT5B*. **C.** The postulated network topologies of the minimal models of gene dosage compensation include 1 feedforward loop for *STAT5A* and 3 feedforward loops for *STAT5B*. **D.** These minimal models recapitulate the gene dosage compensation for *STAT5A* and *STAT5B*. **E.** A three dimensional landscape of gene dosage compensation dependent on the values of ks (synthesis rate of the miRNA), ka (activation parameter of the TF) and kr (repression parameter of the miRNA) demonstrates that the gene dosage compensation mediated by some basic regulatory network motifs depends on the values of these kinetic parameters. Individual cases of regulatory motifs are located within this regulatory landscape.

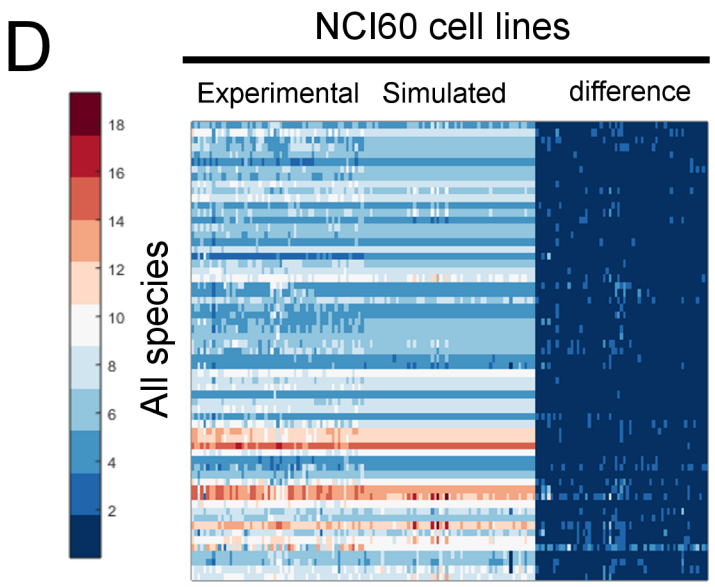
Figure 7. Breast Cancer patients show heterogeneity in gene dosage compensation leading to differences in patient survival. A. Assessment of gene dosage compensation by parameter scans of the gene copy numbers of *MYC*, *STAT3*, *STAT5A* and *STAT5B* for each individual patient-tailored model. **B.** Patient case separation by the amplifications and compensation behavior for *MYC*, *STAT5A* and *STAT5B*. **C.** Survival probabilities for the patient cases divided according to the amplification and compensation behavior of *MYC*. **D.** Compensation landscape of model parameters for the models of the TCGA breast cancer patients. **E.** The interpolation of each individual patient model within the compensation landscape indicates that miR-19a mediates *MYC* dosage compensation in most breast cancer patients.

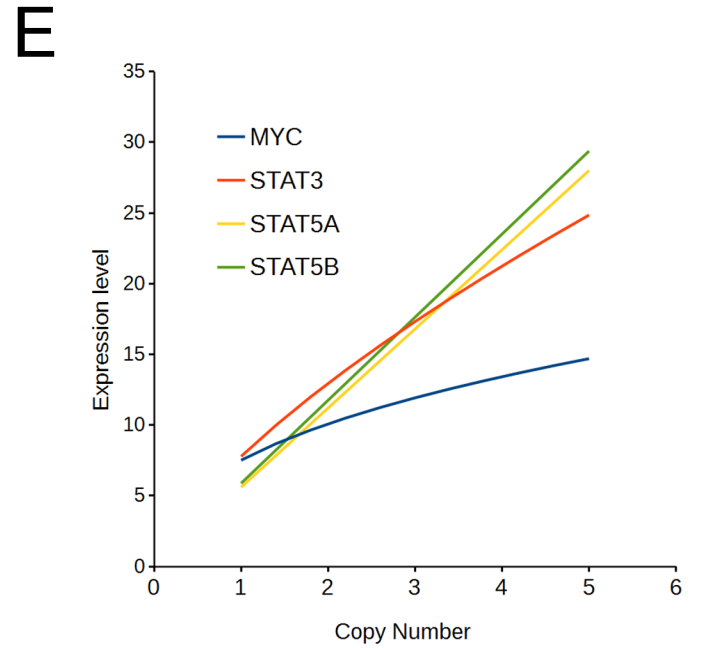

A
Absolute values
Normalized to diploid
Protein
RNA
DNA
NCI60 panel (cell line)
B
SD Protein
SD RNA
SD DNA
C
Probability density
SD RNA
SD DNA
D
SD Protein
SD RNA
SD DNA
SD Protein
SD DNA
SD RNA
SD DNA
E
Absolute values
Normalized to diploid
Protein
RNA
DNA
NCI60 panel (cell line)
F
MYC
RAB5C
SAV1
Protein
RNA
DNA
NCI60 panel (cell line)
G
Gene Copy Number
Selected Genes
NCI60 panel (cell line)

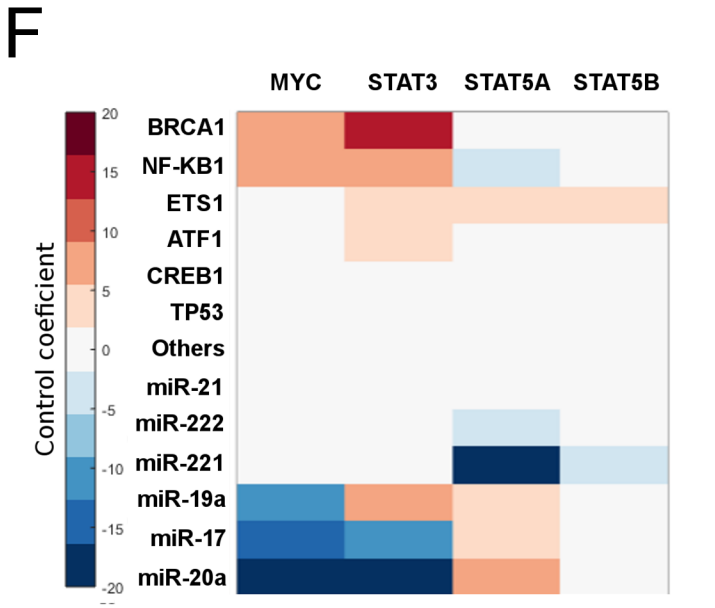

A
miRNAs
Target genes
Other Transcription factors

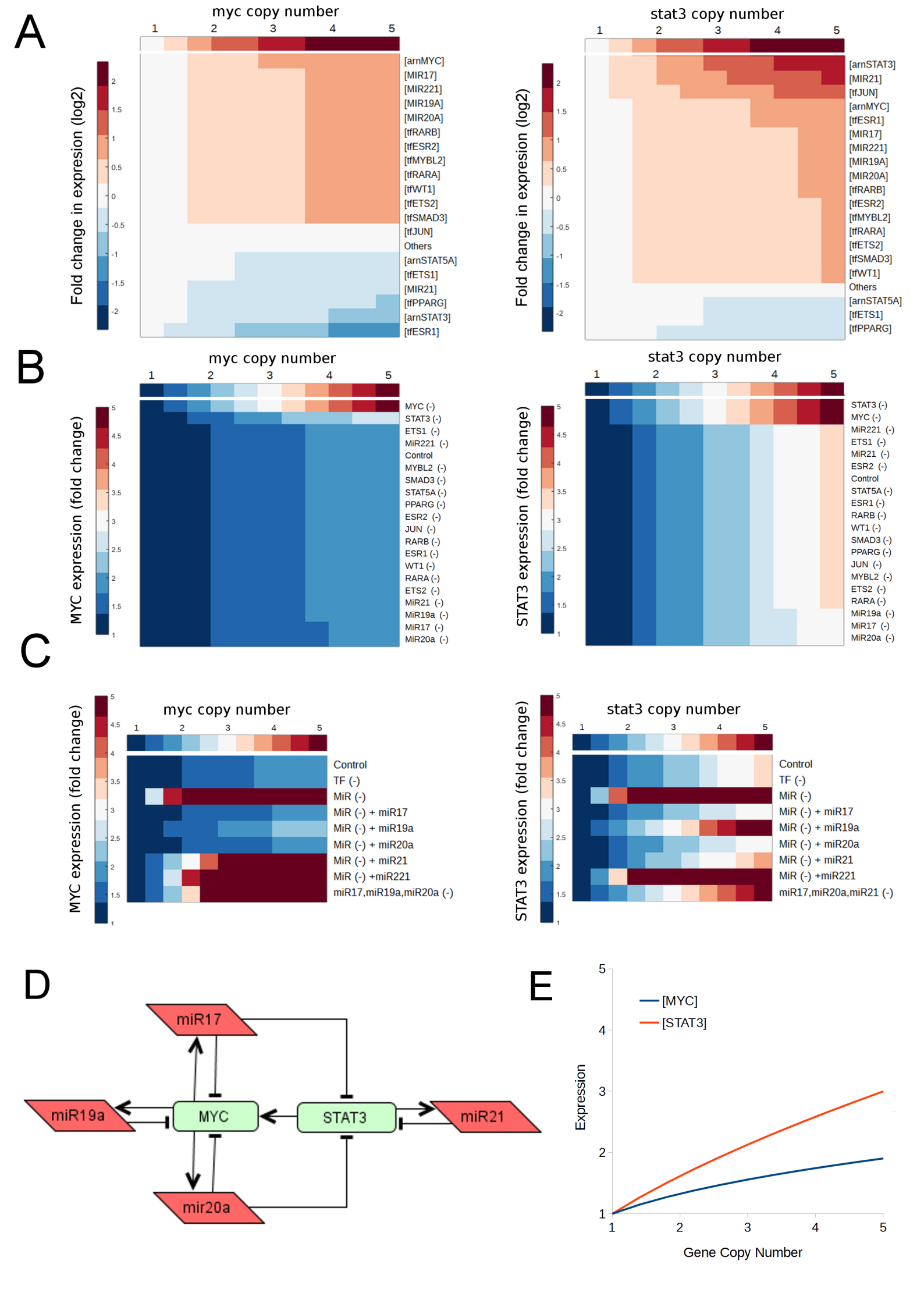

B
Target gene
TF1
microRNA
gene
RNA
TF2
microARN
TF1
gene
microRNA
TF2
Transcription factor
microRNA
gene
TF

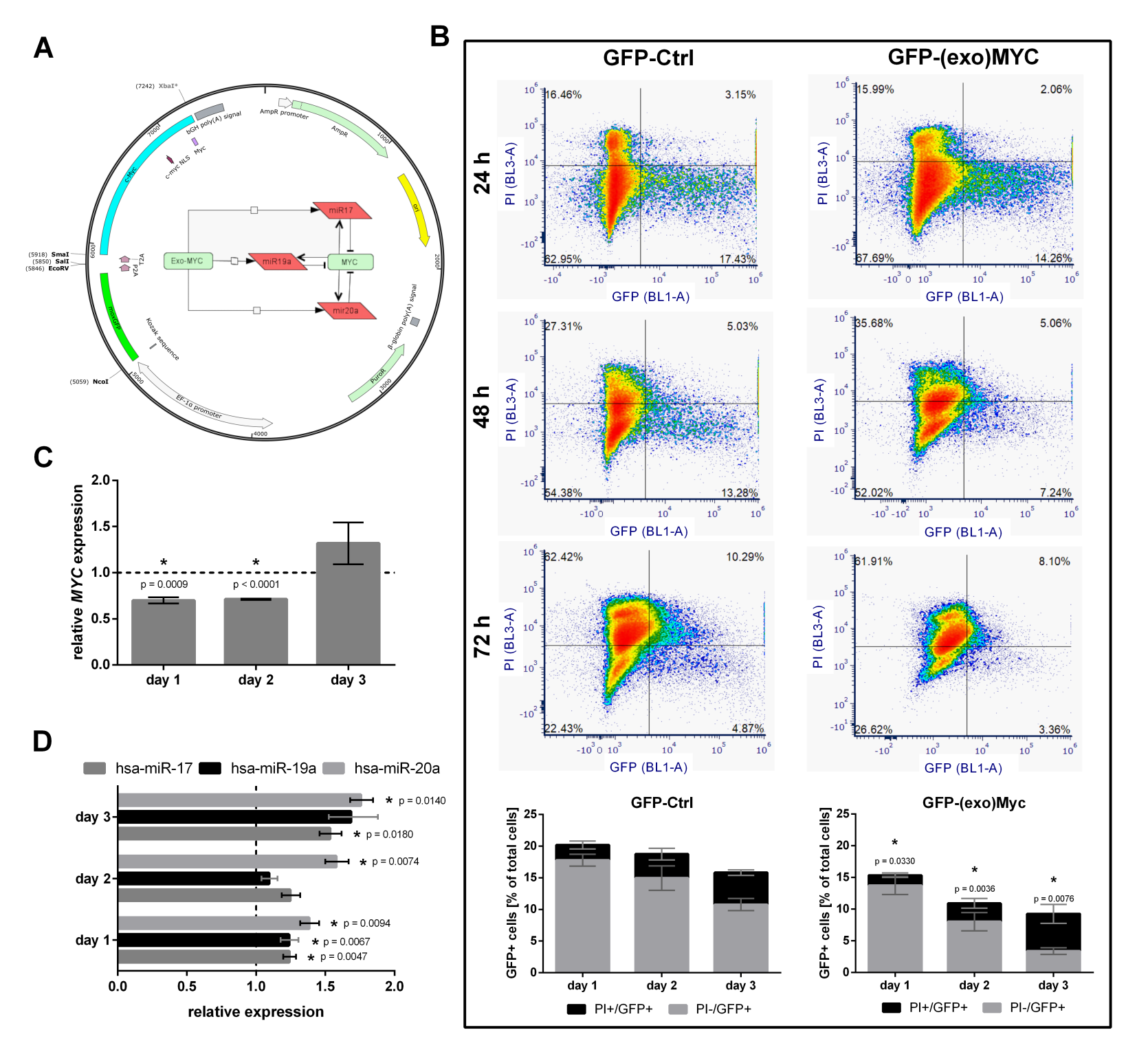

C
Start model
Without miR-106a
Without miR-145
Objective Function
Iterations
0
50
100
150
200
250
300
350
0
20000
40000
60000


D
NCI60 cell lines
Experimental
Simulated
difference
All species
18
16
14
12
10
8
6
4
2


E
MYC
STAT3
STAT5A
STAT5B
Expression level
Copy Number
0
5
10
15
20
25
30
35
0
1
2
3
4
5
6


F
MYC
STAT3
STAT5A
STAT5B
BRCA1
NF-KB1
ETS1
ATF1
CREB1
TP53
Others
miR-21
miR-222
miR-221
miR-19a
miR-17
miR-20a
Control coeficient
20
15
10
5
0
-5
-10
-15
-20


A
myc copy number
1 2 3 4 5
Fold change in expresion (log2)
2 1.5 1 0.5 0 -0.5 -1 -1.5 -2
[arnMYC]
[MIR17]
[MIR221]
[MIR19A]
[MIR20A]
[tfRARB]
[tfESR2]
[tfMYBL2]
[tfRARA]
[tfWT1]
[tfETS2]
[tfSMAD3]
[tfJUN]
Others
[arnSTAT5A]
[tfETS1]
[MIR21]
[tfPPARG]
[arnSTAT3]
[tfESR1]
stat3 copy number
1 2 3 4 5
Fold change in expresion (log2)
2 1.5 1 0.5 0 -0.5 -1 -1.5 -2
[arnSTAT3]
[MIR21]
[tfJUN]
[arnMYC]
[tfESR1]
[MIR17]
[MIR221]
[MIR19A]
[MIR20A]
[tfRARB]
[tfESR2]
[tfMYBL2]
[tfRARA]
[tfETS2]
[tfSMAD3]
[tfWT1]
Others
[arnSTAT5A]
[tfETS1]
[tfPPARG]
B
myc copy number
1 2 3 4 5
MYC expression (fold change)
5 4.5 4 3.5 3 2.5 2 1.5 1
MYC (-)
STAT3 (-)
ETS1 (-)
MiR221 (-)
Control
MYBL2 (-)
SMAD3 (-)
STAT5A (-)
PPARG (-)
ESR2 (-)
JUN (-)
RARB (-)
ESR1 (-)
WT1 (-)
RARA (-)
ETS2 (-)
MiR21 (-)
MiR19a (-)
MiR17 (-)
MiR20a (-)
stat3 copy number
1 2 3 4 5
STAT3 expression (fold change)
5 4.5 4 3.5 3 2.5 2 1.5 1
STAT3 (-)
MYC (-)
MiR221 (-)
ETS1 (-)
MiR21 (-)
ESR2 (-)
Control
STAT5A (-)
ESR1 (-)
RARB (-)
WT1 (-)
SMAD3 (-)
PPARG (-)
JUN (-)
MYBL2 (-)
ETS2 (-)
RARA (-)
MiR19a (-)
MiR17 (-)
MiR20a (-)
C
myc copy number
1 2 3 4 5
MYC expression (fold change)
5 4.5 4 3.5 3 2.5 2 1.5 1
Control
TF (-)
MiR (-)
MiR (-) + miR17
MiR (-) + miR19a
MiR (-) + miR20a
MiR (-) + miR21
MiR (-) +miR221
miR17,miR19a,miR20a (-)
stat3 copy number
1 2 3 4 5
STAT3 expression (fold change)
5 4.5 4 3.5 3 2.5 2 1.5 1
Control
TF (-)
MiR (-)
MiR (-) + miR17
MiR (-) + miR19a
MiR (-) + miR20a
MiR (-) + miR21
MiR (-) +miR221
miR17,miR20a,miR21 (-)
D
miR17
miR19a
MYC
STAT3
miR21
mir20a
E
[MYC]
[STAT3]
Expression
1 2 3 4 5
Gene Copy Number
1 2 3 4 5

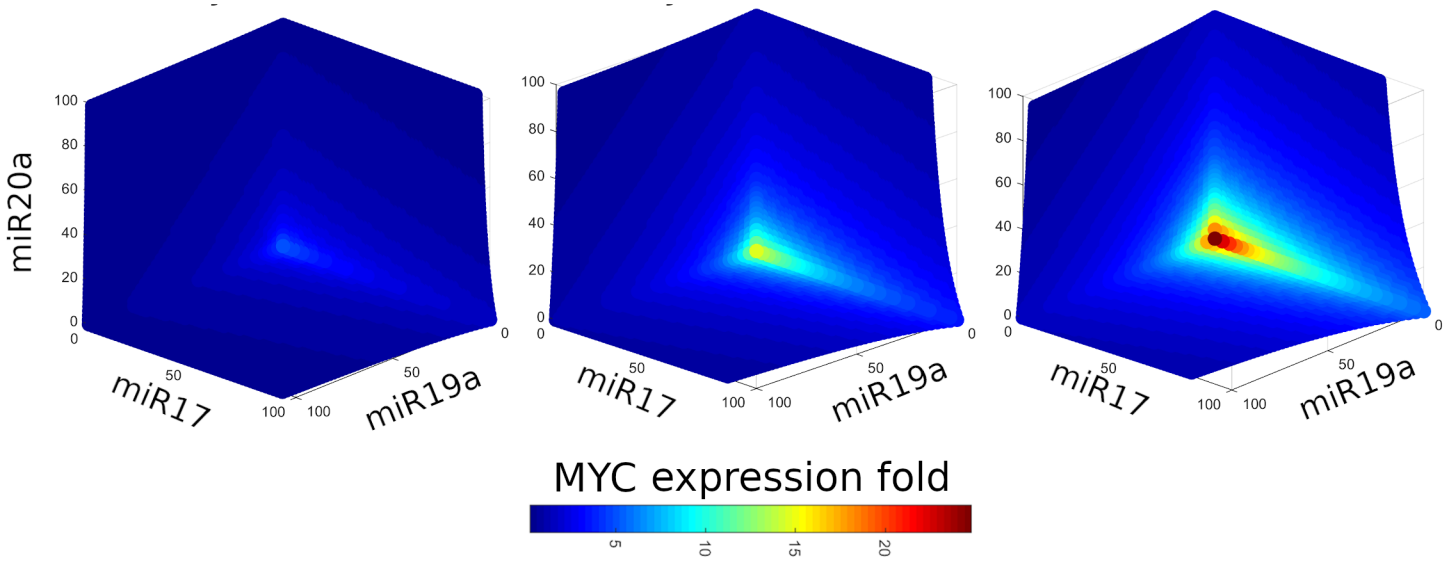

A
(7242) XbaI*
AmpR promoter
AmpR
bGH poly(A) signal
c-myc NLS
Myc
c-Myc
ori
miR17
Exo-MYC
miR19a
MYC
mir20a
(5918) SmaI
(5850) SalI
(5846) EcoRV
T2A
P2A
kozak sequence
β-globin poly(A) signal
PuroR
(5059) NcoI
EF-1α promoter
B
GFP-Ctrl
GFP-(exo)MYC
24 h
48 h
72 h
PI (BL3-A)
GFP (BL1-A)
16.46%
3.15%
62.95%
17.43%
15.99%
2.06%
67.69%
14.26%
27.31%
5.03%
54.38%
13.28%
35.68%
5.06%
52.02%
7.24%
62.42%
10.29%
22.43%
4.87%
61.91%
8.10%
26.62%
3.36%
GFP-Ctrl
GFP-(exo)Myc
GFP+ cells [% of total cells]
day 1
day 2
day 3
PI+/GFP+
PI-/GFP+
p = 0.0330
p = 0.0036
p = 0.0076
C
relative MYC expression
p = 0.0009
p < 0.0001
D
hsa-miR-17
hsa-miR-19a
hsa-miR-20a
p = 0.0140
p = 0.0180
p = 0.0074
p = 0.0094
p = 0.0067
p = 0.0047
relative expression

A

HCT-15 (myc copy number = 2)

HT-29 (myc copy number = 4)

SW-620 (myc copy number = 7)

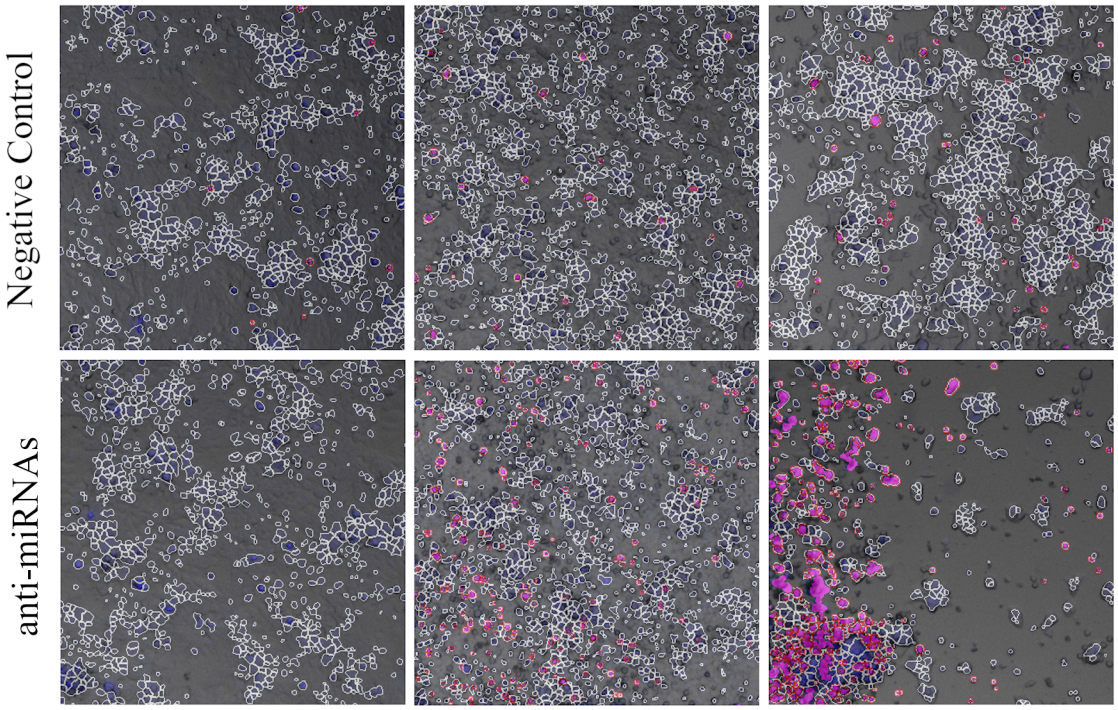


B

HCT-15 (myc copy number = 2)

HT-29 (myc copy number = 4)

SW-620 (myc copy number = 7)

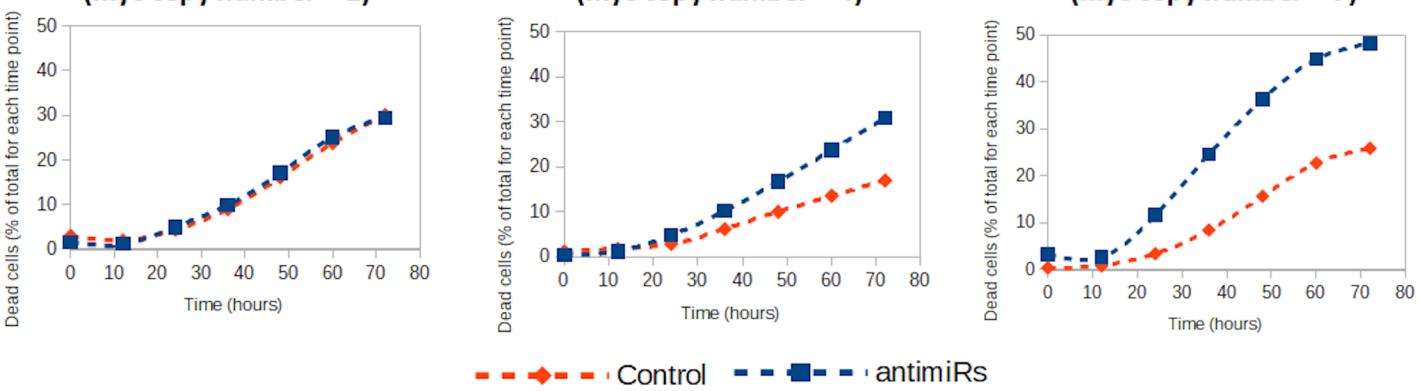


C

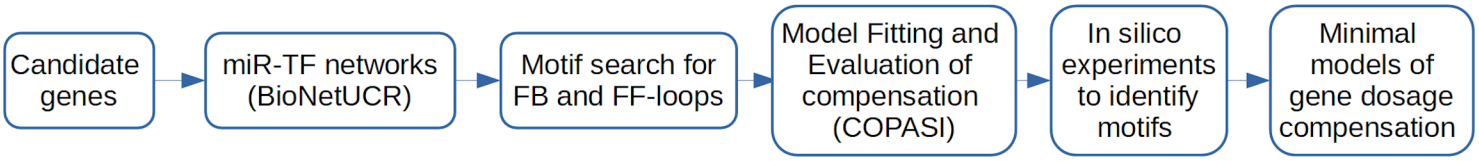
HCT-15 (myc copy number = 2)

HT-29 (myc copy number = 4)

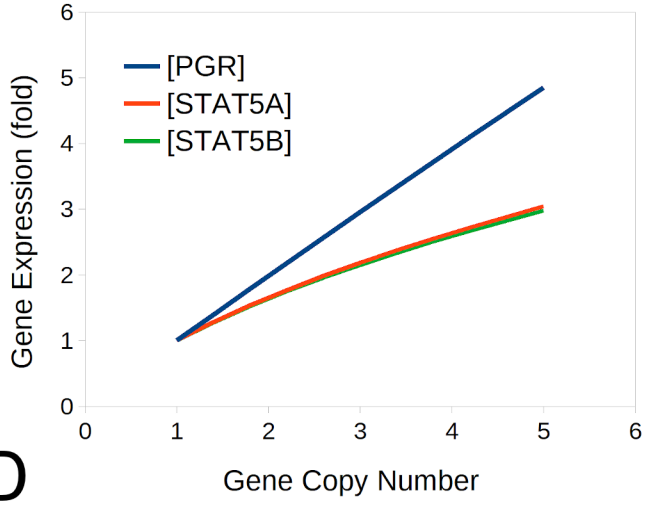
SW-620 (myc copy number = 7)



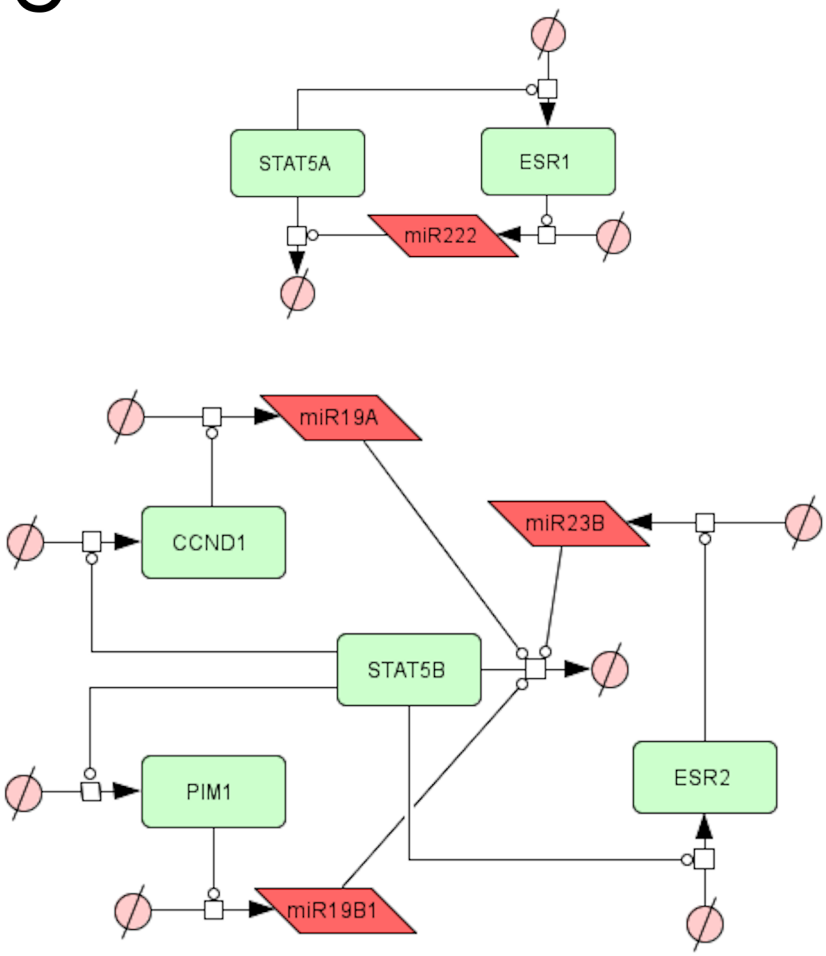
A


Candidate genes
miR-TF networks (BioNetUCR)
Motif search for FB and FF-loops
Model Fitting and Evaluation of compensation (COPASI)
In silico experiments to identify motifs
Minimal models of gene dosage compensation

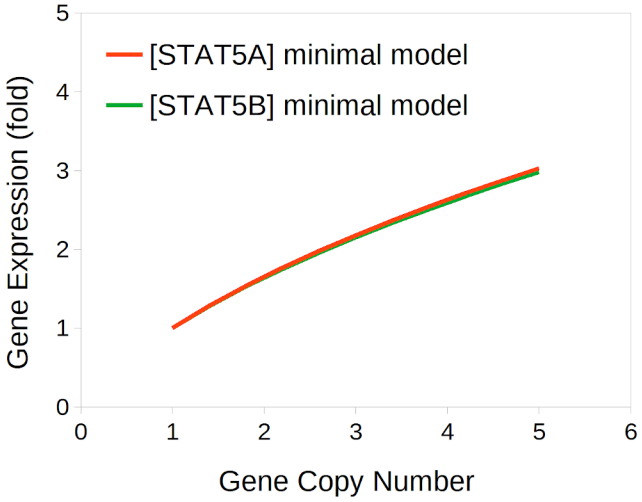
B


[PGR]
[STAT5A]
[STAT5B]
Gene Expression (fold)
Gene Copy Number
0
1
2
3
4
5
6

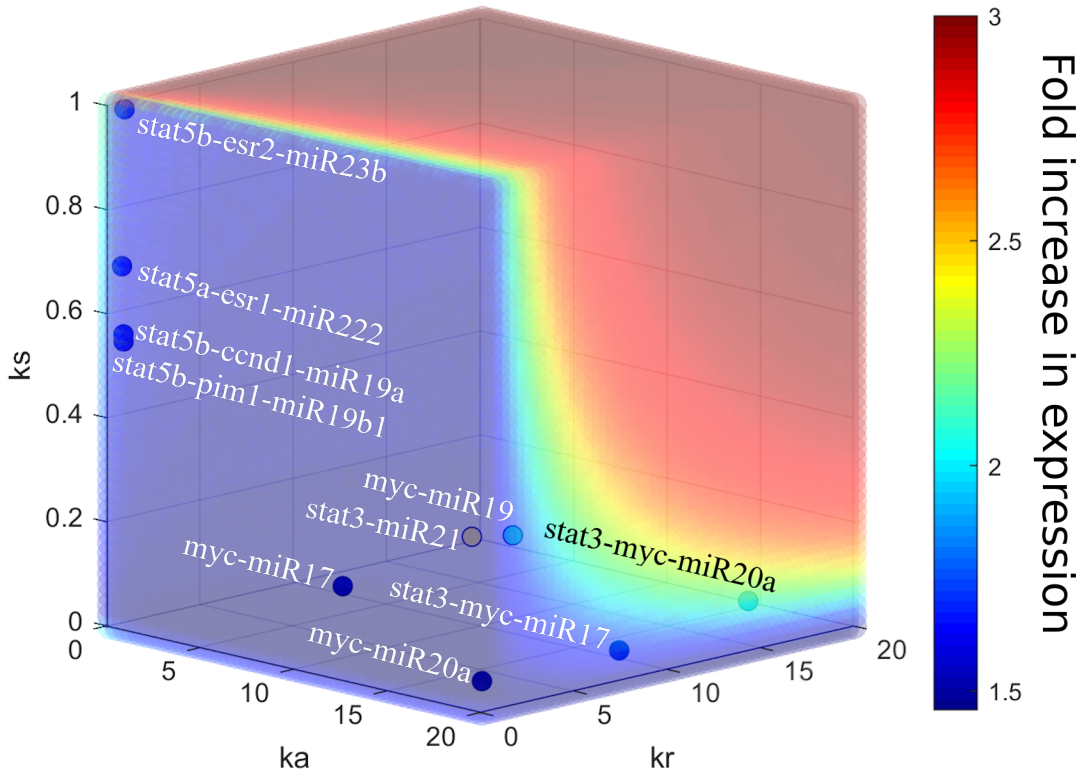
C

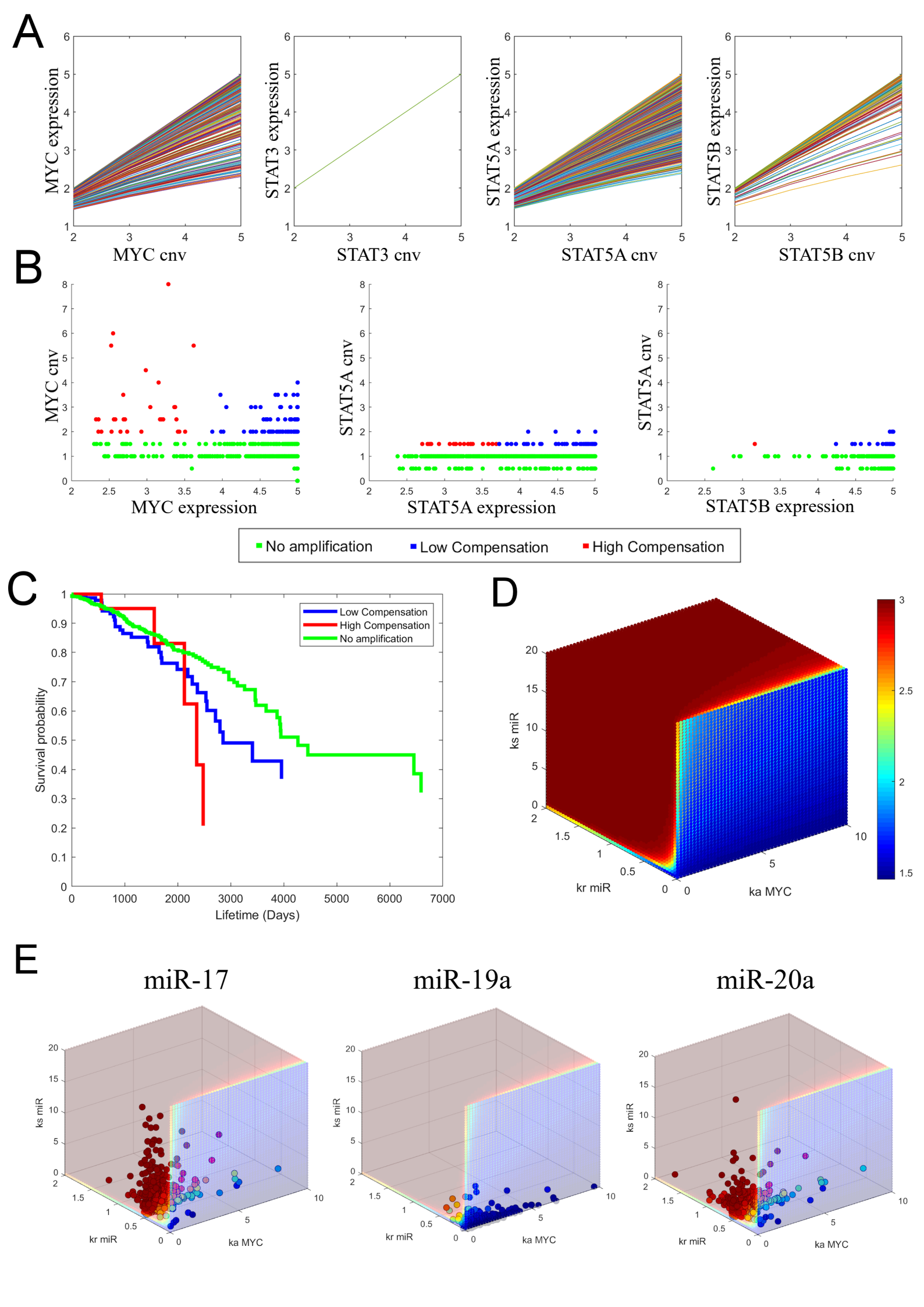

STAT5A
ESR1
miR222
miR19A
CCND1
miR23B
STAT5B
PIM1
ESR2
miR19B1

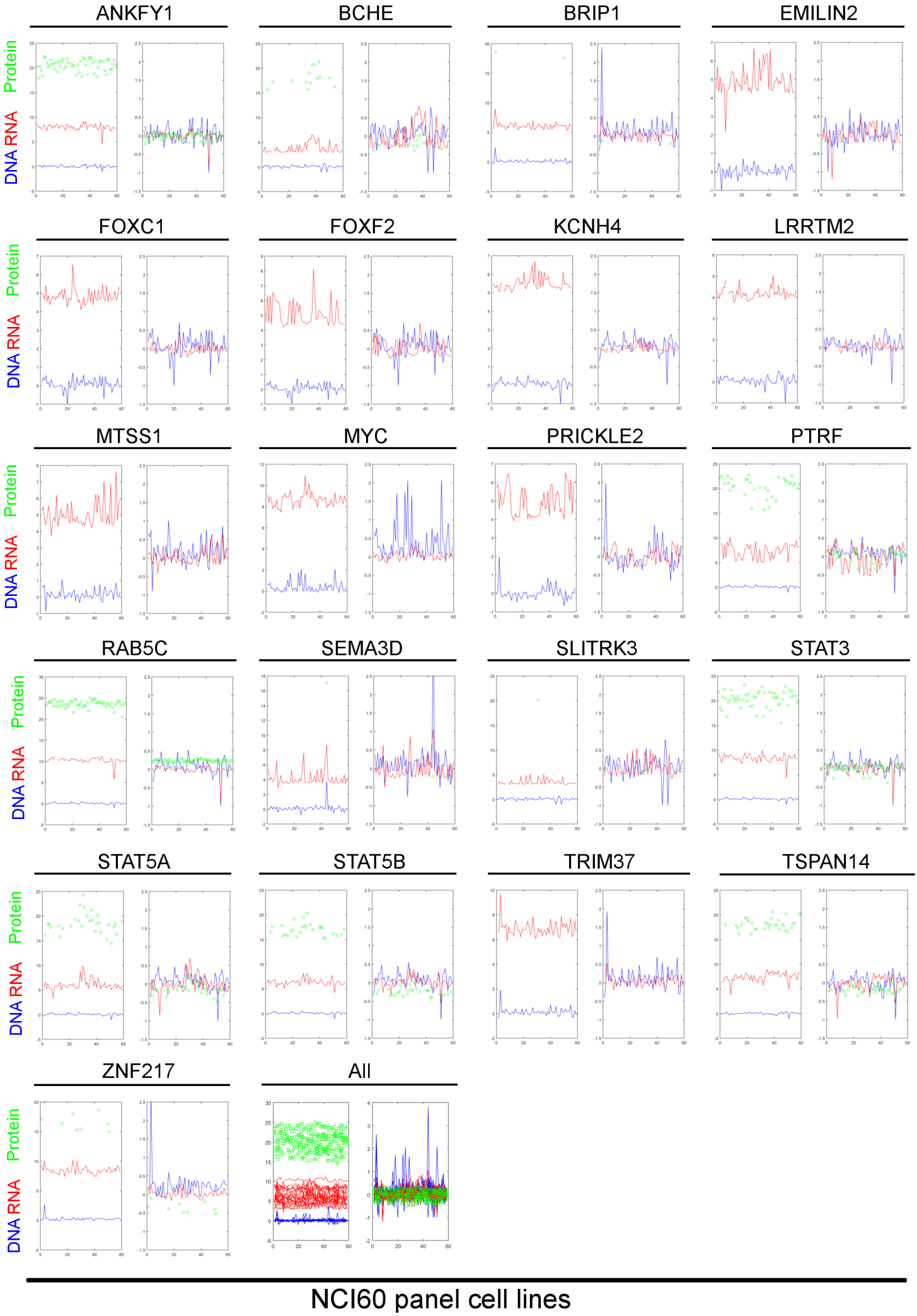
D

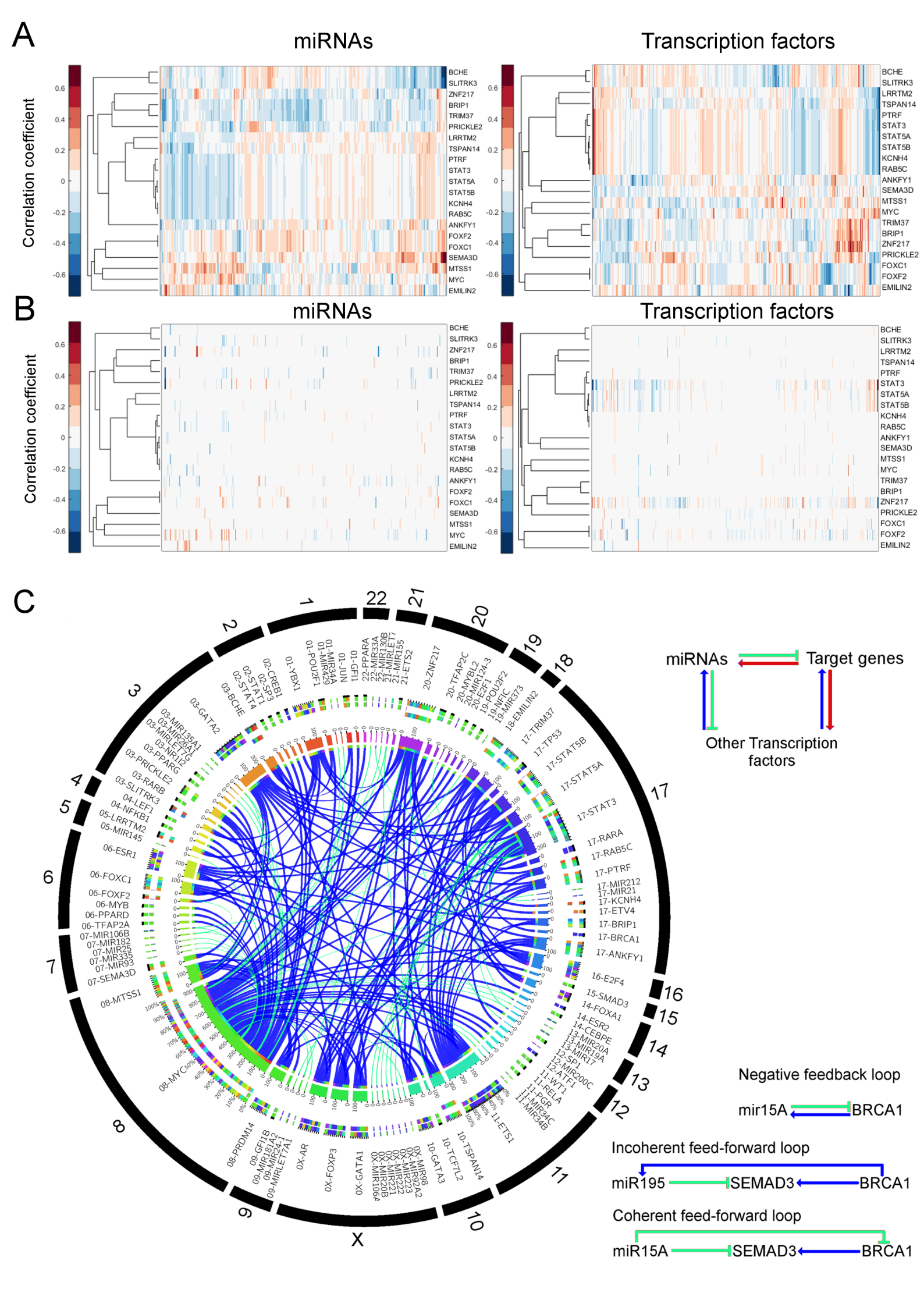

[STAT5A] minimal model
[STAT5B] minimal model
Gene Expression (fold)
Gene Copy Number
0
1
2
3
4
5
6

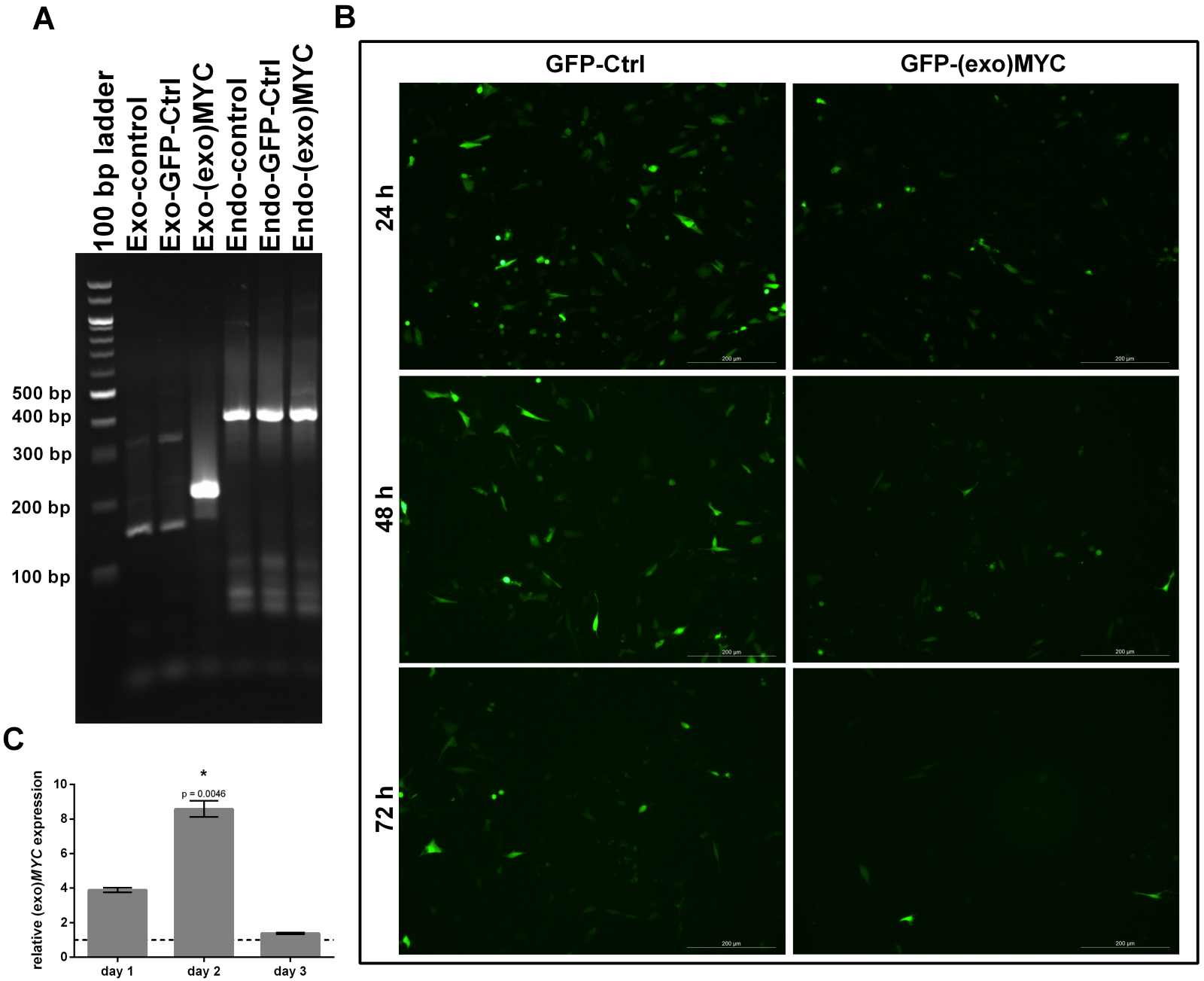
E


stat5b-esr2-miR23b
stat5a-esr1-miR222
stat5b-ccnd1-miR19a
stat5b-pim1-miR19b1
myc-miR19
stat3-miR21
stat3-myc-miR20a
myc-miR17
stat3-myc-miR17
myc-miR20a
ks
ka
kr
Fold increase in expression
1.5
2
2.5
3


A
MYC expression
MYC cnv
STAT3 expression
STAT3 cnv
STAT5A expression
STAT5A cnv
STAT5B expression
STAT5B cnv
B
MYC cnv
MYC expression
STAT5A cnv
STAT5A expression
STAT5A cnv
STAT5B expression
No amplification
Low Compensation
High Compensation
C
Low Compensation
High Compensation
No amplification
Survival probability
Lifetime (Days)
D
ks miR
kr miR
ka MYC
E
miR-17
miR-19a
miR-20a
ks miR
kr miR
ka MYC


ANKFY1
BCHE
BRIP1
EMILIN2
FOXC1
FOXF2
KCNH4
LRRTM2
MTSS1
MYC
PRICKLE2
PTRF
RAB5C
SEMA3D
SLITRK3
STAT3
STAT5A
STAT5B
TRIM37
TSPAN14
ZNF217
All
Protein
RNA
DNA
NCI60 panel cell lines


A
miRNAs
Transcription factors
Correlation coefficient
BCHE
SLITRK3
ZNF217
BRIP1
TRIM37
PRICKLE2
LRRTM2
TSPAN14
PTRF
STAT3
STAT5A
STAT5B
KCNH4
RAB5C
ANKFY1
FOXF2
FOXC1
SEMA3D
MTSS1
MYC
EMILIN2
B
miRNAs
Transcription factors
Correlation coefficient
C
miRNAs
Target genes
Other Transcription factors
Negative feedback loop
mir15A
BRCA1
Incoherent feed-forward loop
miR195
SEMAD3
BRCA1
Coherent feed-forward loop
miR15A
SEMAD3
BRCA1


A
100 bp ladder
Exo-control
Exo-GFP-Ctrl
Exo-(exo)MYC
Endo-control
Endo-GFP-Ctrl
Endo-(exo)MYC
500 bp
400 bp
300 bp
200 bp
100 bp
B
GFP-Ctrl
GFP-(exo)MYC
24 h
48 h
72 h
200 µm
C
relative (exo)MYC expression
10
8
6
4
2
0
*
p = 0.0046
day 1
day 2
day 3