5-2019

# THE ROLE OF GENE EXPRESSION NOISE IN MAMMALIAN CELL SURVIVAL

Kevin Farquhar

## Recommended Citation

THE ROLE OF GENE EXPRESSION NOISE IN MAMMALIAN CELL SURVIVAL

by

*Kevin Scott Farquhar, B.S.*

APPROVED:

_____
Gábor Balázsi, Ph.D.
Advisory Professor


_____
William Mattox, Ph.D.


_____
Nicholas Navin, Ph.D.


_____
Eric Brouzes, Ph.D.


_____
Han Liang, Ph.D.


APPROVED:


_____
Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

THE ROLE OF GENE EXPRESSION NOISE IN MAMMALIAN CELL SURVIVAL


A


DISSERTATION


Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment


of the Requirements


for the Degree of


DOCTOR OF PHILOSOPHY


by


Kevin Scott Farquhar, B.S.

Houston, Texas


May 2019

**Dedication**

*To my family, Jean Farquhar, Lindsay Farquhar Pfisterer, and Blake Pfisterer, who offered their much support and love in any situation, no matter how daunting,*

*To my Uncle Craig and Aunt Sharon, who sprouted my early interest in science as a child,*

*To my late grandma and grandfather Alice and Gordon (Glenn) Farquhar, who made this journey possible with their love and generous support,*

*And to my late father, Randy, who we all miss dearly.*
*All of you made this journey possible.*

appreciated the chance to mentor. All members of the lab helped in their own way with the projects and with our culture in the lab.

My committee members gave great advice during my time at GSBS. I would like to thank them all, Dr. Nicholas Navin, Dr. William Mattox, Dr. Han Liang, Dr. Eric Brouzes, and the chair Dr. Gábor Balázsi for their participation. The Genes and Development (now Genetics and Epigenetics) Graduate Program has been a great program to associate with. I enjoyed our annual Retreats, classes, and events, both research-oriented and social. I enjoyed my conversations and interactions with Dr. Swathi Arur and Dr. Jessica K. Tyler. I appreciate my friendships developed in the program with Dr. Avinash Venkatanarayan, Dr. Esmeralda Ramirez-Peña, and Dr. Alejandro Vila.

Lastly, I would like to acknowledge my cherished friends Dr. Mohit Kumar Jolly and Jason George. You both have been my anchor during this journey. I always enjoyed our conversations and seeing each other progress as scientists.

# THE ROLE OF GENE EXPRESSION NOISE IN MAMMALIAN CELL SURVIVAL

**Kevin Scott Farquhar, B.S.**

**Advisory Professor: Gábor Balázsi, Ph.D.**

Drug resistance and metastasis remain obstacles to effective cancer treatment. A major challenge contributing to this problem is cellular heterogeneity. Even in the same environment, cells with identical genomes can display cell-to-cell differences in gene expression, also known as gene expression noise. Gene expression noise can vary in magnitude in a population or in fluctuation time scales, which is influenced by gene regulatory networks.

Currently, it is unclear how gene expression noise from gene regulatory networks contributes to drug survival outcomes in mammalian cells. An isogenic cell line with a noise-modulating genetic system tuned to the same mean is required. Additionally, how modulating endogenous mean gene expression and noise in living cells influences pro-survival metastatic state transitions remains unanswered.

To address these knowledge gaps, I implemented an exogenous synthetic biology approach to control noise for the drug resistance gene *PuroR* in drug survival while complementing with endogenous expression measurements of the pro-metastatic gene *BACH1* as a correlate for metastatic survival. For exogenous control, I developed synthetic gene circuits in Chinese Hamster Ovary (CHO) cells based on positive and negative feedback that tune noise for *PuroR* at identical mean expression. At a decoupled noise point, isogenic cells were treated with various Puromycin concentrations. Evolution experiments revealed that noise hurts drug resistance during low drug dosage while facilitating resistance at a high Puromycin concentration. Drug adaptation for the low-noise gene circuit relied on intra-circuit mutations while the high-noise circuit did not and became re-sensitized to drug after removing circuit induction.

To implement the endogenous approach, I tagged the endogenous BACH1 gene with the mCherry fluorescent protein in six HEK293 clones. Molecular perturbations such as serum starvation and long-term hemin treatment altered mean fluorescence in at least one clone.

Additionally, monitoring migration after cell wounding revealed increased non-uniform fluorescence at the wound edge. The increased mean fluorescence for the potentially bistable HEK293 clone 2C10 during hemin treatment may reflect altered BACH1 state transitions. Overall, noise enhanced the probability of cells to reach an expression level that confers survival during drug treatment while hemin perturbations may induce a pro-survival metastatic transition via BACH1 expression.

# Table of Contents

# List of Illustrations

Figure 1: Gene expression noise is a form of nongenetic heterogeneity.

Figure 2: Drug-tolerant persister cells can recapitulate the original drug-sensitive population after treatment removal.

Figure 3: CRISPR-Cas9 is gene editing technology re-engineered from a natural anti-bacteriophage immune system in bacteria.

Figure 4: The core epithelial-to-mesenchymal transition (EMT) transcriptional network cross-talks with the MAP kinase pathway.

Figure 5: Gene expression noise enhancing accessibility of pro-survival states.

Figure 6: Stress-dependent effect of network noise on drug resistance.

Figure 7: FACS-sorting of non-expressing and basal CHO-PuroR subpopulations revealed no switching from the non-expressing state.

Figure 8: Dose-response of the mPF-PuroR gene circuit.

Figure 9: The mPF-GFP gene circuit amplifies gene expression noise.

Figure 10: Dose-response of the mNF-PuroR gene circuit.

Figure 11: Assessment of linearity for the mNF-PuroR and mNF-GFP mean dose-responses by L1-norm and curve fitting.

Figure 12: The Flp-In mNF-GFP circuit uniformly expresses *EGFP*.

Figure 13: A decoupled noise regime provides a series of decoupled noise points prior to drug treatment.

Figure 14: Relaxation of high and low expression subpopulations after flow sorting for memory estimates.

Figure 15: Single copy integration for *PuroR* circuits confirmed by qPCR.

Figure 16: Puromycin treatment experimental design.

Figure 17: Replicate expression data of decoupled noise points prior to treatment.

# List of Tables

# Chapter 1: Introduction

**1.1 Background**

**1.1.1 Biological heterogeneity.**

Biological systems can vary phenotypically within and between populations at multiple scales, from single cells to the level of tissues[3], organs[4], and organisms[5]. Indeed, evolution requires phenotypic heterogeneity for selection and adaptation[6]. Phenotypic heterogeneity can be established through genetic variability involving single nucleotide polymorphisms, copy number variations[7], chromosomal translocations[8], and genetic mosaicism[9]. Phenotypic heterogeneity can also depend on the environment[10] and in fact become an adaptive response to the environment, which is called phenotypic plasticity[11]. Phenotypic plasticity has a role in development[12] and can be heritable by principles outside of genetics (epigenetic).

Epigenetics is a concept first described by Conrad Waddington for any heritable alterations that are not governed by genetic processes[13]. Now the field of epigenetics has grouped together disparate mechanisms including histone modifications, non-coding RNA, DNA methylation, and chromatin remodeling as underlying epigenetic processes that do not require DNA modifications[14,15]. Despite the uncertain time scales, cells may demonstrate epigenetic inheritance only through DNA methylation and certain histone modifications[16]. Epigenetic processes that are heritable include genomic imprinting[17], mammalian X-chromosome inactivation[18], paramutation[19], and position-effect variegation (PEV)[20]. Epigenetic heterogeneity can develop in response to the environment[21]. In disease, increased epigenetic heterogeneity, such as genome-wide DNA methylation profiles, can predict neoplastic transformation[22]. Both genetic and epigenetic mechanisms of heterogeneity involve heritable transfer of information in the form of phenotypes, though it is possible for cells displaying phenotypic heterogeneity to have identical genomes and epigenetic marker patterns.

Lately, there is confusion over the term 'epigenetic' in its original meaning compared to the definition heavily used in the epigenetics field in the early 21$^{st}$ century. In the case of the Waddington landscape, the original concept has more to do with *nongenetic* mechanisms and heterogeneity than 'epigenetic' mechanisms. The critical difference between these two types of

mechanisms is the role of stochastic or random processes in nongenetic heterogeneity, which can take the form of cells with the same genome and identical environment demonstrating cell-to-cell variation in phenotypes[23]. As major driver of nongenetic heterogeneity, protein and transcript production in a single cell can randomly fluctuate, which is defined as gene expression noise[24]. A measure of gene expression noise amplitude in a population of cells is the coefficient of variation, which is the standard deviation over the mean (**Figure 1**, right side). Moreover, the time cells spent fluctuating away from mean expression is called cellular memory or the relaxation time (**Figure 1**, left side)[25,26]. Additional molecular and cellular processes driving nongenetic heterogeneity include aneuploidy[27], stochastic partitioning of the cell during division[28], and molecular chaperones[29].

Heterogeneity from nongenetic mechanisms has functional uses in cells. For example, nongenetic heterogeneity in terms of gene expression noise for a stress-response gene can promote survival in a subpopulation of isogenic cells during fluctuating environments, which is known as bet hedging[30]. Evolutionary processes can also accelerate through nongenetic heterogeneity by priming for selection of mutations in viruses[31], unicellular organisms[32], and possibly cancer[33]. Understanding gene expression noise requires expression measurements at the single-cell level to avoid masking the impact of heterogeneity through population-level averaging[34]. Indeed, one cannot understand stochastic gene expression without looking at single cells, especially in higher eukaryotes such as mammals.

**1.1.2 Stochastic gene expression.**

Novick and Weiner first demonstrated in 1957 that induction of enzyme synthesis between individual cells was highly variable and that induction scaled with the proportion of cells maximally producing the enzyme[35]. This stochasticity was later theoretically predicted to occur in gene expression[36] using stochastic kinetic simulations developed by Gillespie in 1977[37]. A major experimental breakthrough in *E. coli* first introduced the concepts of intrinsic and extrinsic sources of gene expression noise[24]. In the experiments, two fluorescent proteins controlled by the same promoter in separate copies were introduced into *E. coli* and measured
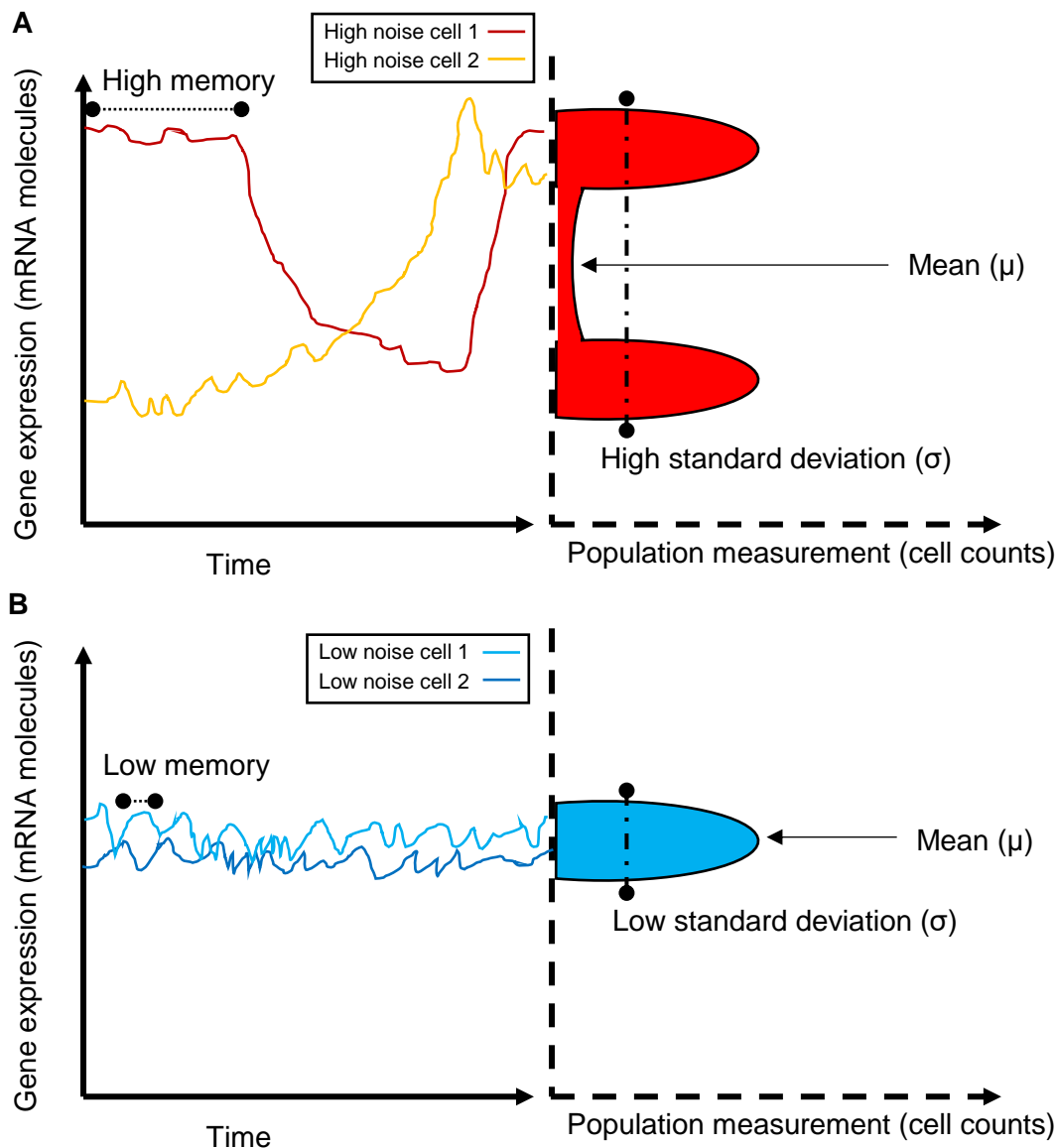
**Figure 1: Gene expression noise is a form of nongenetic heterogeneity.**

High **(A)** and low **(B)** gene expression noise in cells can comprise of noise amplitude and noise frequency. The noise frequency, also called cellular memory or the relaxation time, is the amount of time a cell spends in a consistent expression level (left) deviating from the mean. The gene expression amplitude is quantified by the coefficient of variation ($\sigma/\mu$) and is a population level statistic at one time point. Both the gene expression noise amplitude and frequency depend on the network structure regulating the measured gene.

with flow cytometry. Whenever the cyan and yellow fluorescence intensities directly correlated with each other in the same *E.coli* cell, the gene expression noise in a cell population arose

from extrinsic processes, such as the cell cycle and other global or pathway-specific processes. On the other hand, completely uncorrelated fluorescence levels indicated fluctuations from intrinsic sources, such as random biochemical processes in transcription and translation[38].

The mechanism of stochastic gene expression at the promoter level in prokaryotes follows a two-state model of promoter activity[39,40], which is not governed by a Poisson stochastic process[40]. In eukaryotes, a two-state model predicts that promoters randomly transition between an 'off'' state, which represents a completely silenced promoter, and an 'on' state that can transcribe mRNA[41]. The $k_{on}$ and $k_{off}$ rates in the two-state model determine how long promoters take to transition into the 'on' or 'off' state, respectively[42]. Additionally, the rate that mRNA molecules are produced when RNA Polymerase II is bound to the promoter is termed $k_m$[41,43]. Biologically, chromatin remodeling in eukaryotic cells can mediate the transcriptional promoter state transitions by altering promoter accessibility for binding of the transcriptional pre-initiation complex[38,41]. Thus, positional effects at specific genomic loci can influence fluctuations in gene expression whenever chromatin is accessible or non-accessible, leading to random silencing and activation[44].

When the promoter transition rates are slow, as in eukaryotes[45], transcriptional activity can produce 'bursts' of mRNA, which is quantified by the average number of mRNA molecules created per transcriptional burst (burst size) at a measurable rate (burst frequency)[46]. The two-state model predicts the transcriptional burst size equals the active mRNA production rate over the transition rate to an inactive promoter state ($k_m/k_{off}$) while the burst frequency equals the transition rate to the active state ($k_{on}$)[41]. In a more complex eukaryotic model, there are two 'on' states: active promoter bound to a transcriptional pre-initiation complex along with the TATA-box binding protein (TBP) either 1) without RNA Polymerase II or 2) with RNA Polymerase II[43,46]. An experimentally integrated three-state model of eukaryotic stochastic gene expression demonstrated that transcriptional bursting also depends on TATA box sequence, with mutations in the TATA box lowering fluctuations in gene expression, presumably by altering the *TBP* binding rate[43]. The two-state and three-state models are not mutually exclusive. In fact,

evidence is accumulating that chromatin-mediated promoter state transitions from the two-state model can depend on enhancer elements[47]. Additionally, the TATA-box binding protein central to the three-state model is associated with increased gene expression fluctuations depending on which cofactors in the pre-initiation complex (PIC) bind *TBP* and whether the turnover of *TBP*-bound pre-initiation complexes increases, mediated by PIC-disruptors such as *Mot1p*[48]. Thus, promoter element sequences such as the TATA-box can influence gene expression noise[43].

Protein production can also display noisy bursting dynamics. Translational bursting depends on the number of mRNA molecules available for protein synthesis, increasing with low levels of mRNA and higher translational efficiency, also called translational burst size (proteins synthesized per mRNA molecule)[49]. Noise strength, quantified by the variance divided by the mean (Fano factor), can determine the source of protein noise based on translational bursting[42]. The noise strength will increase over a span of mean protein levels whenever the translational burst size also increases with constant transcription rate. On the other hand, noise strength will not change with higher translational burst size whenever the transcription rate increases[49]. The coefficient of variation would not show this dependency, allowing noise strength to distinguish sources of protein noise.

Lowering cell volume enhances stochastic gene expression by increasing the relative proportion of changes in protein subcellular compartment concentrations during diffusion between compartments. For example, in the nucleus, the number of proteins relative to the cytoplasm is vastly lower though equal in concentration, which increases proportional changes between compartments during nucleus protein number fluctuation. This phenomenon is called the 'finite-number effect', which roughly scales noise (coefficient of variation, standard deviation over the mean) to $1/\sqrt{N}$, where N is the average protein number[42]. Interestingly, mammalian cell volume correlates with transcriptional burst size while progression in the cell cycle negatively correlates with transcriptional burst frequency[50]. Though it is unclear how extensively gene expression noise affects mammalian cell phenotypes, it would not be

6

surprising to uncover additional relationships for noise in mammalian cells given the adequate tools.

### 1.1.3 Synthetic biology.

While recombinant DNA technologies have revolutionized the field of biotechnology and genetic engineering, most techniques tinkered with natural living systems. More recently, the development of artificial genetic constructs in novel, unnatural applications have coalesced in the field of synthetic biology[51]. Fundamentally, synthetic biology takes natural and unnatural biological components to quantitatively engineer novel biological systems, such as synthetic gene circuits akin to electrical circuits[52], metabolic engineering of biosynthesis pathways[53], and even whole chromosomes[54] and genomes[55]. The standardization of biological parts and components form the core basis of engineering increasingly complex artificial biological systems[56], which has become a central effort in adopting a Synthetic Biology Open Language (SBOL) for interchangeable part design[57].

At their core, the composition of synthetic gene circuits includes a gene(s) of interest and their interactions in the circuit, whether through transcriptional[58], translational[59], RNA-based[60] or protein-protein-based[61] mechanisms. Thus, network motifs such as positive or negative feedback loops, feed-forward loops, and cascades shape the interaction architecture and dynamics of a synthetic gene circuit[62]. For instance, negative feedback loops (through repressors) tend to lower the amount of time expression levels deviate from the mean (quick relaxation time/lower cellular memory)[63] and overall reduce gene expression fluctuations[64]. In contrast, positive feedback loops (through activators) slow down the response time of the circuit (longer relaxation time/higher cellular memory)[65] while enhancing stochastic fluctuations in gene expression[66]. Indeed, feedback loops allow systematic engineering of inducible synthetic gene circuits to manipulate nongenetic heterogeneity both temporally (cellular memory)[67] and within a population (gene expression noise amplitude)[67,68].

In the early days of synthetic biology, simple synthetic gene circuits such as a toggle switch in *E. coli* demonstrated the transferability of an electrical engineering framework to

biology[58]. The toggle switch network pioneered the use of the tetracycline repressor (*TetR*) family of transcriptional repressors[69,70], which has expanded to various orthogonal repressors[71] and the related reverse tetracycline transcriptional activator (*rtTA*)[72]. Synthetic gene circuits with regulatory components can act as oscillators[73], edge detectors[74], and function in more complex systems such as multi-cellular biological computers[75]. Now the field is becoming increasingly applied, with clinical and commercial interests in synthetic gene circuits. As therapeutic devices, engineered cells with synthetic gene circuits can detect pathogenic biomarkers[76], monitor and adjust physiological pH in response to diabetic ketoacidosis[77], and correct insulin resistance[78].

Recent advances in mammalian synthetic biology have significant clinical potential[77,78]. Mammalian synthetic gene circuits can modulate expression in response to light through optogenetic regulatory factors, which can regulate blood glucose to promote physiological levels as a proof-of-concept therapy shown in mice[79]. Additionally, synthetic biology principles have found usefulness in human chimeric antigen receptor (CAR) therapy (CAR-T), where receptors implemented an AND logic gate for targeting cancer cells with two antigens[80]. The engineered T cells avoid targets with just one antigen, thereby enhancing CAR-T specificity. As in bacteria and yeast, inducible mammalian synthetic gene circuits can adjust cellular memory[81] and can reduce the magnitude of fluctuations in gene expression[82]. Mammalian synthetic biology has progressed far enough that synthetic signaling pathways, such as synNotch[83], are scaling towards interactions in complex systems with multiple cells and synthetic intercellular communication[84]. Still, the field of mammalian synthetic biology is nascent with many technical challenges. Quantitative prediction of engineered mammalian synthetic gene circuits has lagged compared to other organisms due to cell-type-specific endogenous gene regulation and locus insertion positional effects[44,85]. Future studies in mammalian synthetic biology should comprehensively address these challenges by using, for example, recombinase-mediated landing pads[86].

**1.1.4 Drug resistance in mammalian cells.**

Chemotherapy resistance is a major contributor to cancer-related deaths in the United States, with most of the 609,640 cancer deaths in 2018 ultimately caused by resistance to chemotherapy treatments[87]. The mechanisms underlying drug resistance in cancer and more generally in mammalian cells are complex and still under active investigation[88]. The distinction between tumor chemoresistance and drug resistance in mammalian cells is a matter of scale; effective understanding of chemotherapy and tumor drug resistance requires model organisms with organs affecting drug metabolism, drug distribution, and the tumor microenvironment[89]. On the other hand, the molecular and cellular biology of drug resistance in single mammalian cells can find just as much importance in mechanisms as diverse as altered cellular metabolism, rewiring of signaling pathways and molecular targets, induction of stem-like cell reprogramming, evasion of apoptosis, genomic instability, epigenetic changes, and phenotypic heterogeneity[90,91]. Indeed, the crucial mechanisms for mammalian and tumor drug resistance determine survival at the cellular level.

Mammalian cells can employ a multi-drug resistance (MDR) phenotype to cope with cytotoxic drugs[92]. Increased *P-Glycoprotein* (*ABCB1*, otherwise known as *MDR1*) expression can mediate the multidrug resistance phenotype in a multitude of cancers and mammalian species[93]. *MDR1* is an ATP-binding cassette (ABC) family transporter in the plasma membrane that pumps cytotoxic drugs out of the cell, thereby conferring drug resistance[94]. In addition to the MDR phenomenon, mutations in drug targets[95], resistance-associated amplification of genes such as *MDR1*[96], and genomic instability[97] contribute to mammalian drug resistance, leading to genetic heterogeneity, selection, and adaptation in response to treatment[98,99]. Mammalian cells can also inactivate drugs to confer resistance through chemical modifications, such as the case of thiol glutathione binding to platinum drugs[100].

Drug resistance in mammalian cells can also rely on drug-specific mechanisms. The aminonucleoside Puromycin, which shares chemical similarity to the 3'-end of an aminoacylated tRNA, inhibits translation and is cytotoxic in both prokaryotes and eukaryotes[101].

Puromycin disrupts protein synthesis by binding to the A site in the ribosome and covalently binding to the translating peptide chain through the peptidyl transferase reaction in the ribosomal P site, which leads to pre-mature release of the polypeptide chain[102]. The bacterium species *Streptomyces alboniger*, which synthesizes Puromycin, also produces the corresponding resistance gene *Puromycin N-acetyltransferase* (*pac* or *PuroR*), which acetylates Puromycin and its toxic precursor *O*-demethylpuromycin, thereby simultaneously inactivating Puromycin while functioning as an enzyme in the Puromycin biosynthesis pathway[103]. In mammalian cells selected for natural resistance, the *P-glycoprotein* (*MDR1*) was reported to confer resistance to Puromycin along with multiple drugs, linking the multi-drug resistance phenotype to Puromycin resistance[93].

In addition to genetic mechanisms, epigenetic and nongenetic mechanisms can promote drug resistance in mammalian cells. Epigenetic modifications such as chromatin inaccessibility from remodeling confers resistance in 'persister' cells, where a drug-induced subpopulation of non-growing cells can transiently tolerate a cytotoxic drug and subsequently repopulate the sensitized main subpopulation after drug removal (**Figure 2**)[104-106]. The chromatin transitions rely on histone demethylase *KDM5A*[104]. By opening chromatin accessibility, histone deacetylase complex inhibitors (HDACi) can disrupt the persister state to re-sensitize the cells to a cytotoxic drug (**Figure 2**)[104]. Mammalian persister cells display tolerance to molecularly targeted drugs, such as the *EGFR* tyrosine kinase inhibitors[104,105,107]. In addition to HDAC inhibitors, disrupting a dependency on the lipid hydroperoxidase *GPX4* can re-sensitize persister cells to drug and induce a lipid reactive oxygen species (ROS) form of cell death called ferroptosis[105]. In a population with high nongenetic heterogeneity, a small persister-like subpopulation of human melanoma cells displayed very high resistance marker (*EGFR*, *JUN*, *WNT5A*, *AXL*, *PDGFRB*) expression levels[108].

Related to the concept of persister cells, cancer stem cells (CSCs) are hypothesized to comprise of a set of tumor cells, not necessarily a small or quiescent subpopulation, that have aggressive tumor-initiating potential, self-renewal of the CSC state, and the ability to
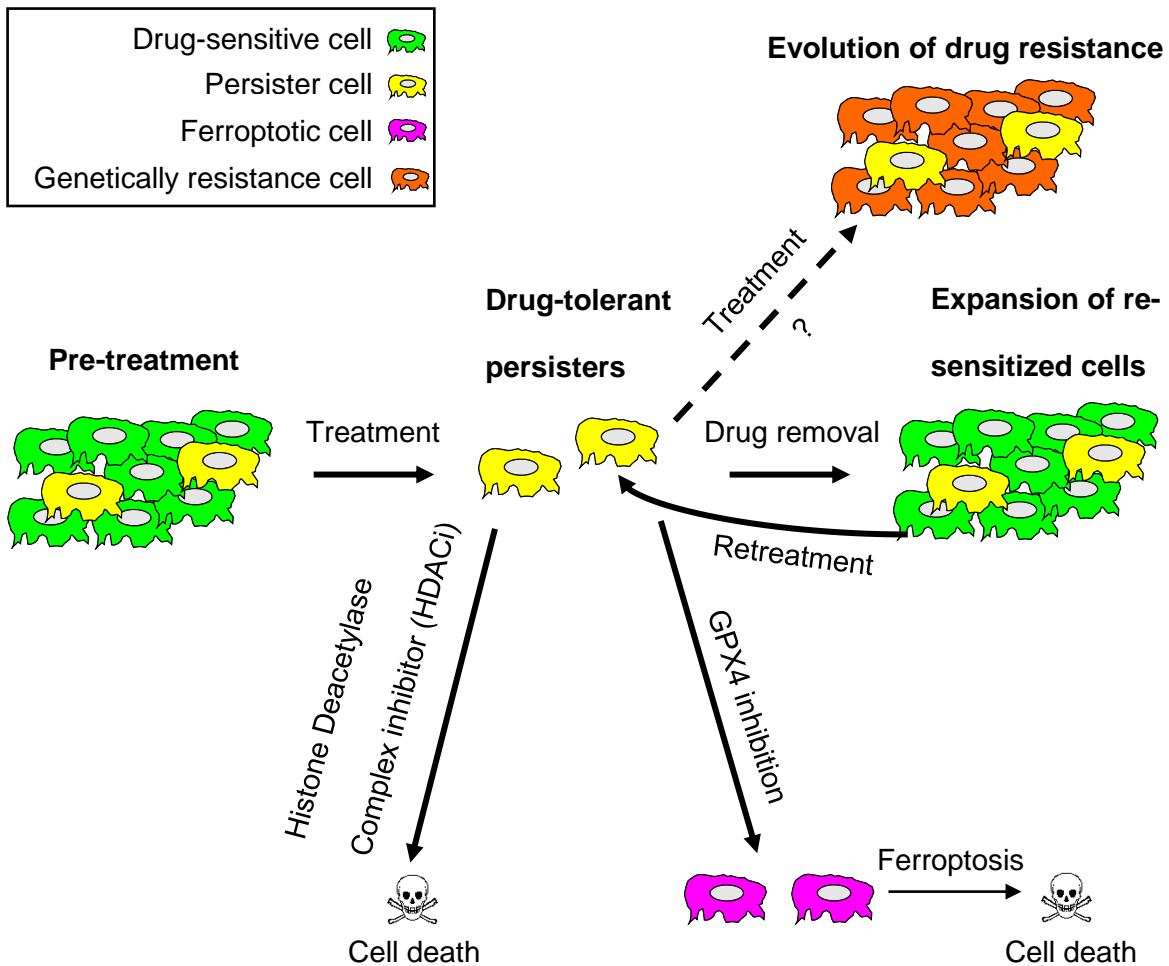
**Figure 2: Drug-tolerant persister cells can recapitulate the original drug-sensitive population after treatment removal.**

Bacteria and cancer cells can have a subpopulation of non-growing persister cells that tolerate antibiotics or chemotherapy, respectively. Remove of drug treatment allows the persisters to phenotypically switch to a proliferative state that recapitulates the original population, leaving a majority of cells drug-sensitive. At least two mechanisms can disrupt the persister state in mammalian cells: inhibition of histone deacetylase complexes or GPX4. It's unclear whether mammalian persister cells can evolve to become genetically drug resistant.

progressively transition into a hierarchy of tumor cell types with little ability to initiate or propagate tumors[109]. Cancer stem cells exhibit enhanced drug resistance properties that arise from quiescence, increased drug efflux activity, robust DNA repair pathways, and lower

predisposition toward apoptosis[110]. Overexpression of *MDR1* in cancer stem cells contributes to their high drug efflux rate[111]. While less understood, polyploid cells with multiple nuclei fused together can display cancer stem-like properties like aggressive tumorigenic capacity and enhanced drug resistance[112,113].

In addition to phenotypic heterogeneity from cancer stem cells and the persister state, gene expression heterogeneity may aid drug resistance in cancer[114,115]. Expression for the drug-pumping *MDR1* gene, for example, displays features of cellular memory, with high expressing cells returning the original distribution after flow sorting[115]. Instead of a single gene, fractional apoptosis in response to TRAIL ligand can rely on variability in the rate of initiator *CASP8/10* cleavage of the *BID* protein, which propagates expression noise for downstream pro- and anti-apoptotic proteins[116]. The studies on *MDR1* and *BID* did not consider the effect of mean expression on survival, which is inherent to gene expression noise. Future experiments must control gene expression noise while minimizing the effect of mean expression on differential survival to drug.

**1.1.5 Experimental evolution.**

Evolution as a process leads to diversification of living organisms over time. Four major mechanism underlie the ability of populations to change: natural selection, random mutation, genetic drift, and gene flow. Natural selection is a fundamental mechanism for cells and organisms to adapt to their environment, which enriches for genotypes that optimally survive and reproduce (i.e., maximize fitness)[117]. Adaptation requires heritable, differential fitness between multiple genotypes. Fundamentally, adaptative evolution occurs on the level of genes. However, phenotypic interactions with the environment can influence adaptative evolution, and in some cases accelerate the rate of mutations for further adaptation[118]. Importantly, differences or alterations in gene expression can act as an adaptive response to a changing or hostile environment, thereby promoting phenotypic heterogeneity[119]. Investigations of adaptative evolution can take place in a natural or experimental setting.

Experimental evolution systematically imposes controlled environmental conditions in the laboratory to explore how populations adapt through fundamental evolutionary processes[120]. Pioneering evolution experiments in bacteria demonstrated that fitness benefits decelerate in the long term[121]. Experimentally evolved populations can adapt through pre-existing genetic heterogeneity or by increasing overall heterogeneity in response to the environment[122]. For instance, experimental evolution of *P. fluorescens* by passaging colonies with novel morphology and periodic shaking of liquid cultures led to adaptation through stochastic morphology state switching[123]. Because selecting against colonies with state-switching led to the genotype's elimination, the adaptation developed as a form of bet-hedging in the fluctuating environment. Bet-hedging through nongenetic heterogeneity can also lead to drug treatment adaptation in bacteria[124]. Consequentially, cancer is an evolutionary process. Therefore, mammalian cells, especially in tumors, may exploit bet-hedging strategies for survival under chemotherapy[125]. However, evolution experiments with mammalian cells have not rigorously tackled how controlled phenotypic heterogeneity affects drug resistance independently from other traits.

More recently, experiments in mammalian cells have evolved novel proteins and antibodies for biotechnological purposes[126]. Indeed, therapeutic evolution experiments of proteins require the physiology of mammalian cells for proper folding and maturation[127]. As biotechnology toolkits advance, evolution experiments in mammalian cells have become more automated with diverse measurements of novel evolutionary adaptation[128]. Now, emerging gene editing tools can directly drive genetic diversification[129]. Lacking in current basic research studies, experimental evolution of mammalian cells with constrained conditions, such as constant mean expression, a lack of genetic diversity, and differential gene expression noise, may reveal how mammalian cells with varying levels of gene expression noise adapt to drug treatment or tumorigenic cells escape from the primary tumor.

**1.1.6 CRISPR-Cas9.**

The gene editing tool CRISPR-Cas9 originated from a bacterial anti-phage defense mechanism. Initiated by bacteriophage invasion, the natural type II CRISPR system works (**Figure 3A**) by first incorporating phage-specific protospacers, which are flanked by direct repeats for self-recognition, into a CRISPR locus expressing a noncoding RNA (crRNA). Direct repeats from the transcribed pre-crRNA bind to a trans-activating CRISPR RNA (tracrRNA), leading to cleavage and maturation of the crRNA. The CRISPR-associated enzyme Cas9 then cleaves the RNA duplex for eventual maturation of individual crRNAs each with a spacer and repeat. By targeting 20 nucleotides directly followed by a protospacer adjacent motif (PAM), crRNA guides Cas9 cleavage of invading viral DNA while avoiding targeting of endogenous spacers lacking PAM sites to avoid self-recognition[130].

Discovery of the CRISPR adaptive immunity system led to the realization that combining the crRNA duplex with tracrRNA while removing unessential nucleotides simplifies the system into two components: the Cas9 and a newly engineered single guide RNA (sgRNA) that is a fusion of the tracrRNA and crRNA[131]. A DNA sequence with 20 nucleotides complementary to a guide RNA followed by a 3' PAM site can be targeted by the engineered CRISPR-Cas9 gene editing system for Cas9-mediated cleavage, usually leading to inefficient DNA damage repair through the non-homologous end joining (NHEJ) pathway (**Figure 3B**)[132]. Since NHEJ repair causes indels, complete gene knock-out systems can elucidate gene function more effectively than RNA interference[133].

In addition to indel creation, the double-stranded breaks in DNA cleaved by Cas9 can promote homology directed repair (HDR) in the presence of DNA with homologous sequence, which leads to replacement of the region surrounding the double-stranded break with homologous donor sequence (**Figure 3B**)[132]. Therefore, any synthetic sequence in the donor DNA molecule flanked by sequence homologous to the region of Cas9-mediated cutting can insert itself into a genomic region with a PAM site, allowing CRISPR-Cas9-mediated knock-in gene editing. The homologous DNA can comprise of homology arms in a relatively large
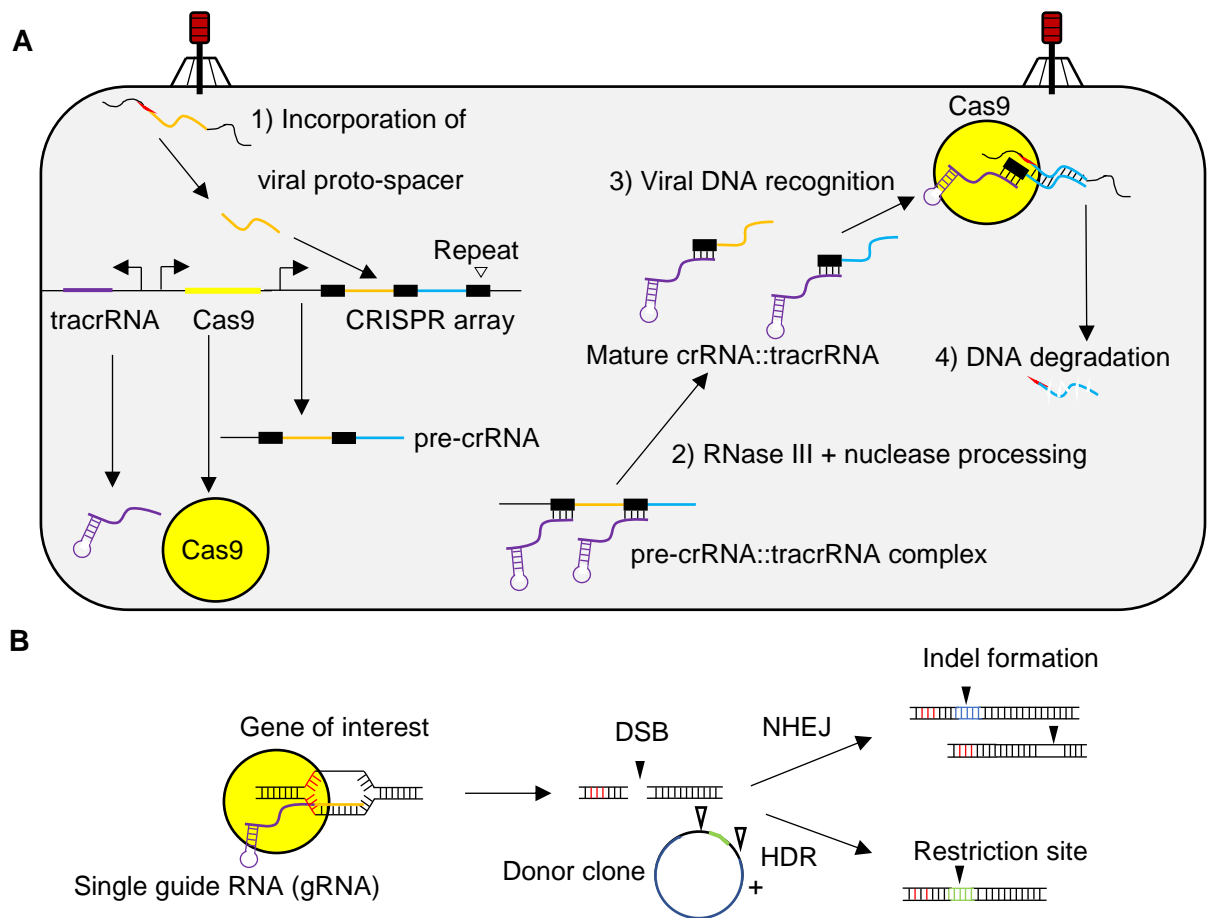
14

**Figure 3: CRISPR-Cas9 is gene editing technology re-engineered from a natural anti-bacteriophage immune system in bacteria.**

**(A)** The type II CRISPR/Cas bacterial viral immune system incorporates viral DNA as proto-spacers (orange and blue) in the noncoding CRISPR array flanked by repeating elements (thick black rectangles). The tracrRNA (purple) and pre-crRNA duplex are processed by nucleases into mature crRNA bound to tracrRNA. The Cas9 protein binds to the RNA complex, which then binds to and facilitates cleavage of viral DNA containing a proto-spacer adjacent motif (PAM, red, NGG). **(B)** The tracrRNA and crRNA are fused as a guide RNA (gRNA) in the CRISPR-Cas9 gene editing system. Repair of double strand breaks (DSB, black arrow) cut by Cas9 can disrupt gene function with the error-prone non-homologous end joining (NHEJ) pathway or introduce a novel sequence with the homology directed repair pathway given the presence of a homologous donor clone (plasmid, unfilled arrows for homology arms).

double-stranded plasmid vector or single-stranded oligodeoxynucleotides (ssODNs)[134]. Replacement of deleterious disease-associated mutations in gene therapy[135], introduction of fluorescence proteins for gene expression tagging[136], and insertion of restriction sites at genomic loci are common applications of CRISPR-Cas9-mediated knock-in. In fact, CRISPR-Cas9 knock-in of a fluorescence reporter tag at the native loci for a gene allows live-cell endogenous gene expression monitoring, which can reveal dynamics that have a basis in network structure or how specific expression levels affect a phenotype, like metastasis as an example[137].

A major drawback of CRISPR-Cas9 gene editing is off-target cleavage from partially complementary guide RNA sequence[138]. To improve the specificity of CRISPR-Cas9 gene editing, directed protein evolution experiments uncovered more specific Cas9 variants[139]. Engineering Cas9 variants with mutations affecting DNA-Cas9 binding to improve specificity can also greatly reduce off-target effects[140]. Bioinformatics tools help in the design of guide RNAs to avoid off-target cleavage[141]. Targeting two close PAM sites with a gRNA pair and a double-nickase Cas9 can dramatically decrease off-target cleavage[142]. In mammalian cells, lowering expression of proteins involved with the NHEJ pathway promotes the repair of Cas9-induced double-stranded breaks through the HDR pathway[143].

**1.1.7 Metastasis regulatory factors.**

Along with chemoresistance, metastasis is a leading cause of cancer-related deaths[87]. In breast carcinomas, a major cellular program driving dissemination from the primary tumor is the epithelial-to-mesenchymal transition (EMT). Although it is hotly disputed whether EMT is necessary for metastasis[144], EMT has a role in early dissemination by promoting invasion of the basement membrane[145]. Specifically, EMT inhibits α6β4 integrins that normally maintain contact with the basement membrane[146] while requiring α5β3 and α3β1 integrins for EMT progression through interactions with the TGF-β signaling pathway[147,148]. Indeed, the TGF-β ligand is a major inducer of the core EMT network (**Figure 4**)[149].
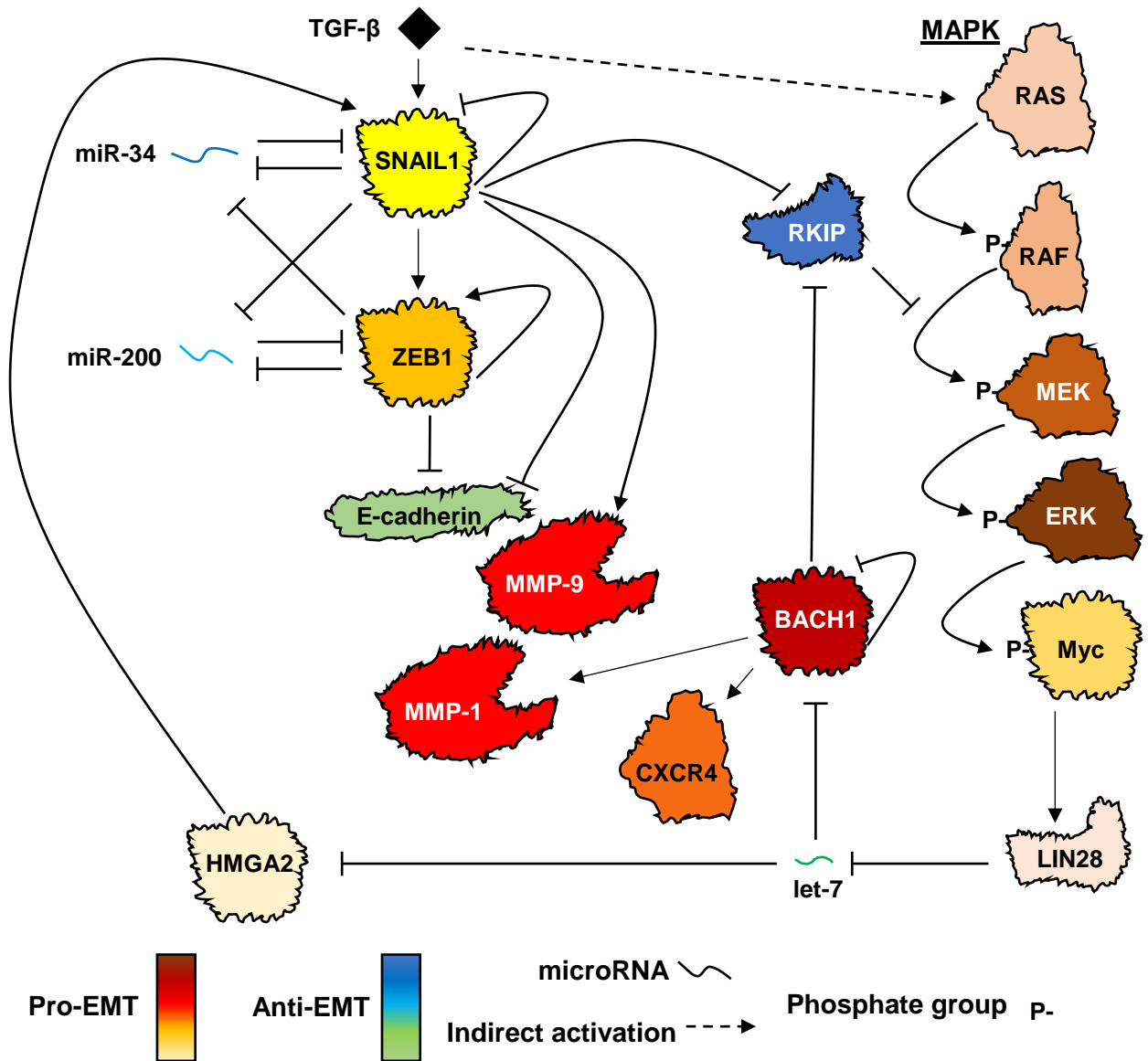
**Figure 4: The core epithelial-to-mesenchymal transition (EMT) transcriptional network cross-talks with the MAP kinase pathway.**

TGF-β ligand induces the expression of *SNAIL1*, a pro-EMT transcription factor, part of a transcriptional activation cascade with the pro-EMT *ZEB1* transcription factor. Both repress epithelial-associated genes, including *E-cadherin*. Each EMT-TF is repressed by an anti-EMT microRNA. Ras is indirectly activated by TGF-β, which indirectly activates *SNAIL1* expression. Inhibition of *RAF kinase* activity by *RKIP* is disrupted by *SNAIL1* transcriptional repression. The microRNA *let-7* can de-repress *RKIP* by reducing expression of *BACH1*. Matrix metalloproteases are activated by EMT for basement membrane degradation.

17

Increased expression of EMT core transcription factors can promote early metastasis[150]. In the core EMT transcriptional network, TGF-β ligand transduces extra-cellular signaling by increasing the expression of the pro-EMT transcription factors *SNAIL1/2* and *TWIST*[151] while indirectly activating the RAS-MAPK pathway[152]. The microRNA *miR-34* negatively regulates *SNAIL1* expression while itself being transcriptionally repressed by *SNAIL1* in a toggle switch network motif[153]. *SNAIL1* can also repress itself and activate the pro-EMT transcription factor *ZEB1*[154], which also inhibits *miR-34*[155]. *ZEB1* activates itself and is directly repressed by the microRNA *miR-200*, which is also inhibited by *ZEB1* and *SNAIL1*[156]. *SNAIL1* and *ZEB1* transcriptionally repress E-cadherin, which leads to the disruption of epithelial adherens junction between cells[149]. Important for basement membrane invasion, *SNAIL1* also promotes the expression of matrix metalloprotease 9 (*MMP-9*)[157].

In addition to the core EMT pathway, RAS-MAPK can contribute to signal pathway cross-talk (**Figure 4**), where TGF-β and RAS-MAPK indirectly promote EMT through de-repressing *HMGA2* expression that then promotes *SNAIL1* expression[149]. By lowering RAS-MAPK activity, the Raf kinase inhibitory protein (*RKIP*) can indirectly repress *HMGA2* through de-repression of the microRNA *let-7*[158]. Additional cross-talk is mediated by inhibition of *RKIP* by *SNAIL1* and *BACH1*. *BACH1* represses its own expression and is indirectly repressed by *RKIP* through *let-7*[1]. Important for metastatic phenotypes, *BACH1* directly activates *CXCR4* and *MMP-1* expression[159]. Although *MMP-1* is usually silenced in breast cancer cells, *MMP-1* from stromal cells in the tumor microenvironment can facilitate breast carcinoma invasion by degrading the *PAR1* receptor, thereby activating pro-migration pathways independently from MMP-mediated ECM degradation[160].

Heterogeneity shapes the development and progression of metastasis. Single breast cancer cells were found to vary in gene copy number within a multiclonal tumor, which could determine the extent of tumor subpopulation clonal expansion between a primary tumor and a secondary metastatic tumor[161]. The tumor microenvironment has a heterogenous population of stroma that can enhance metastatic progression through secreting cytokines, various types of

growth factors, and components of the extra-cellular matrix[162]. Matrix stiffness in the microenvironment can facilitate breast cancer EMT and metastasis[149,163]. Spatial organization and heterogeneity in the microenvironment can also influence genetic heterogeneity through clonal evolutionary dynamics during cell survival[164].

In addition to genetic and microenvironmental heterogeneity, phenotypically plastic cancer stem cells, for example, aggressively initiate metastasis[165]. Partial EMT states may have the highest proclivity for initiating metastasis[166]. Importantly, partial EMT states could reflect nongenetic heterogeneity from the EMT transcriptional network[167]. Given the self-activation of *ZEB1* in the core EMT network (**Figure 4**), mutual inhibition of EMT microRNA *miR-200* and the transcription factor *ZEB1* can lead to tristability, where cells can transition between three stable *miR-200*/*ZEB1* expression states[167]. Additionally, expression of the *BACH1* and *RKIP* toggle switch network was found to display bistable-like transitions correlating with pro- and anti-metastatic states, respectively[1]. Although the characteristics of the epithelial-to-mesenchymal transition in living cells is well-established[149,168], it is unclear how stochastic transitions between pro- and anti-metastatic states relate to endogenous *BACH1* expression noise *in vivo*.

**1.2 Core hypothesis and significance.**

Although studies on gene expression noise have mainly focused on unicellular prokaryotes and eukaryotes, understanding the role of gene expression noise in higher eukaryotes, particularly in living mammalian cells, is lagging. As a major challenge, mammalian cells have more complex gene regulatory networks and genetic constructs tend to randomly integrate into the genome. In addition, investigation of noise independently of the mean requires control of gene expression noise at similar mean expression levels, which is very difficult for endogenous genes. Such control feats are easier with exogenous genes that are integrated at a targeted genomic site. However, endogenous gene expression noise should not be ignored; concepts like incomplete penetrance have a deep connection to gene expression noise that required measuring endogenous noise to uncover[169].

As introduced earlier, gene expression noise can drive phenotypic state transitions, like the Waddington landscape conceptualized for cellular differentiation, which can depend on network structure. As an example, Bacteria evolved gene regulatory networks to exploit gene expression noise for survival in response to environmental stress[62,170]. The surviving bacteria then enter a cellular state that allows drug tolerance, and these cells can be driven from and back to this pro-survival state through gene expression fluctuations[171]. It is not clear that mammalian cells can survive drug treatment solely based on gene expression noise, even in a short time-scale.

Though studies in relation to nongenetic (Waddington) landscapes theoretically predict gene expression noise as driving pro-survival state transitions in mammalian cells, particularly in a drug resistance or metastatic context[172], experimental verification is mostly lacking and the studies that do find a relationship suffer from two major weaknesses. First, gene expression noise is inherently tied to mean expression. Without making gene expression noise independent of mean expression by keeping the mean constant between two mammalian cell populations, it is difficult to rigorously separate the contribution of mean expression from gene expression noise on an outcome. Second, the cell populations must be isogenic, keeping genetics and the environment the same but varying gene expression noise. These two weaknesses are difficult to overcome for studying noise in endogenous genes, but even studies with exogenous systems do not meet the mark[173]. Therefore, how gene expression noise influences access to pro-survival cellular states in mammalian cells, as seen in bacteria, remains an open question.

In this thesis, I hypothesize that gene expression noise in mammalian cells increases the probability to reach a cellular state leading to a survival outcome that is bounded by an expression threshold, whether it is in response to drug treatment (*scenario 1*) or metastatic progression beyond a primary tumor (*scenario 2*). In the first scenario, the level of drug will determine the expression threshold for the drug resistance gene *Puromycin N-acetyltransferase* (*PuroR*) required to survive Puromycin treatment (**Figure 5A**). When the
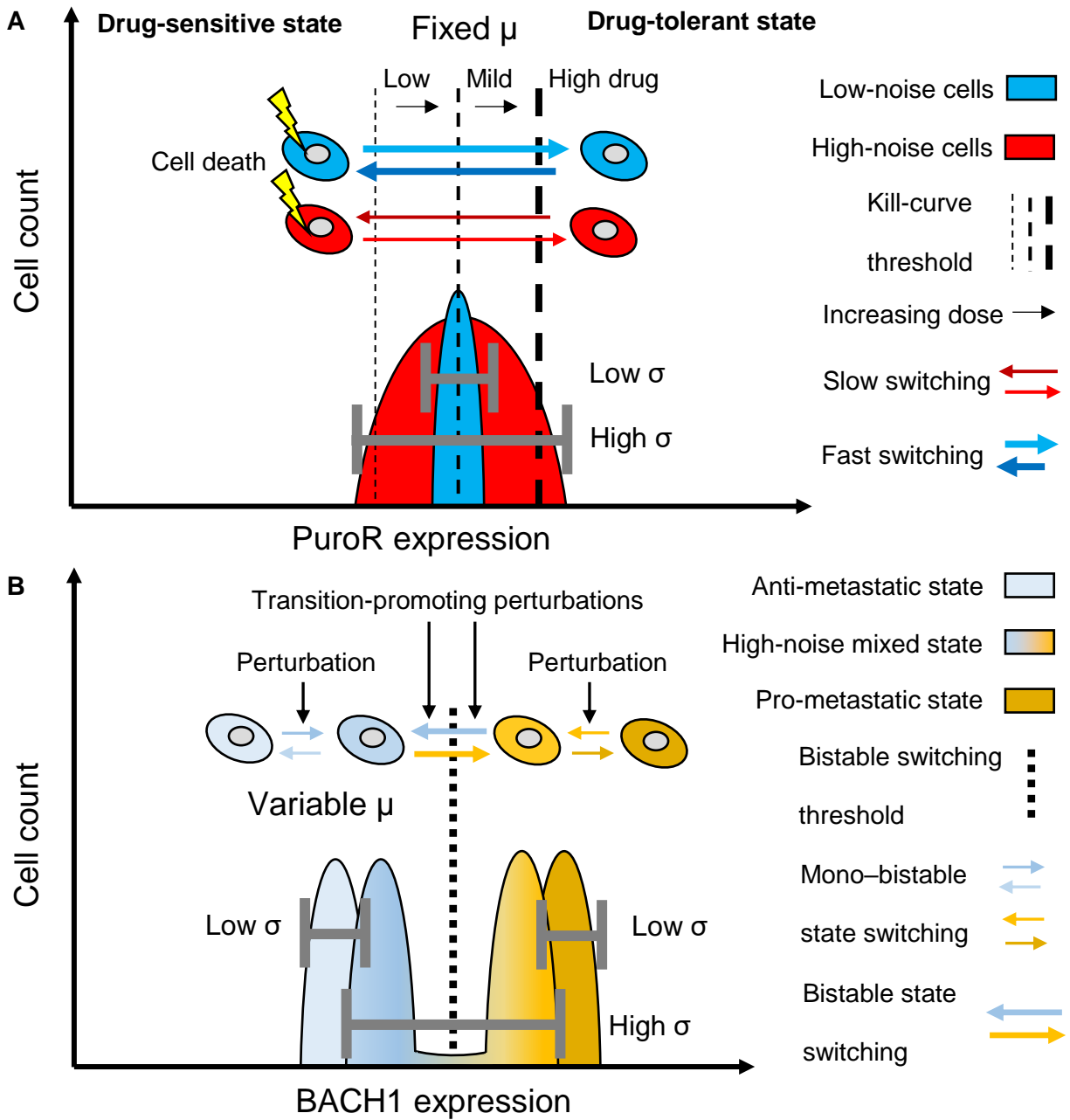
**Figure 5: Gene expression noise enhancing accessibility of pro-survival states.**

Cells express either **(A)** an exogenous drug resistance gene *PuroR* or **(B)** an endogenous reporter from the pro-metastatic gene *BACH1*. Single-cell *PuroR* expression above a variable kill-curve threshold confers survival in response to various drug concentrations. A fixed bistable *BACH1* transition threshold determines entrance into metastatic states. The mean ($\mu$) is fixed during drug treatment while molecular and physical perturbations may change both *BACH1* mean ($\mu$) and noise ($\sigma/\mu$) sufficiently large enough to initiate bistable transitions. Cell expression can fluctuate slowly or quickly. Legends are on the right.

mean *PuroR* expression is fixed, high *PuroR* noise will lead to more cells entering the drug

survival state than cells with low *PuroR* noise. Likewise, for the self-repressed pro-metastatic

regulator *BACH1*, which is engaged in a toggle switch network by reciprocal repression with the

anti-metastatic regulator *RKIP*, cells with *BACH1* gene expression above a specific expression

threshold will enter a pro-metastatic state (**Figure 5B**). The pro-metastatic threshold can

change depending on network perturbations, such as perturbations altering the strength of

BACH1 self-repression and anti-metastatic RKIP protein degradation[1]. The experimental

perturbations in this thesis should not disrupt these parameters, leaving the threshold fixed. If

the pro-metastatic *BACH1* mean expression is too low, the chance of cells entering the pro-

metastatic state may diminish to zero depending on the noise level. Nonetheless, the highest

expressing cells in a high noise population at each perturbation-driven mean expression level

have a higher probability of entering the pro-metastatic state compared to the highest

expressing cells in a low-noise population at the same mean because of the higher cellular

memory and typically higher noise amplitude in gene expression.

My investigation of noise affecting entrance into the drug survival state (*scenario 1*) is

through <u>exogenous</u> control of the drug resistance gene *PuroR* at a fixed mean (**Figure 5A**),

and the chance of being in the pro-survival state is predicted to depend on single-cell *PuroR*

expression levels. As for *BACH1* (*scenario 2*), I aim to estimate the occupancy in the pro-

metastatic state by measuring <u>endogenous</u> single-cell *BACH1* expression levels in response to

various perturbations and identifying bistable state transitions whenever I observe large shifts

in expression levels (**Figure 5B**). Incorporating an endogenous system for studying gene

expression noise will complement insights brought about by the more rigorous exogenous

approach.

The contrasting focus on drug resistance and metastatic regulators expands the scope

and significance of the thesis. It is also possible that *BACH1* has a relationship with the cancer

persister state because *NRF2* target genes shared with BACH1 are downregulated in cancer

persister cells[105] and *RKIP* regulates ferroptosis, which as a form of cell death mediates

persister cell drug resensitization upon loss of the lipid peroxidase *GPX4*[174]. By estimating state occupancy through single cell expression, I avoid the complexities of phenotypically quantifying the metastatic state with an *in vivo* model of metastasis, which could vastly increase the number of variables affecting survival outside the primary tumor. Unlike the drug resistance scenario, endogenous BACH1 mean and noise cannot be rigorously controlled by perturbations while keeping the mean the same. However, because the metastatic state threshold is fixed, the effect of mean expression on the cell state occupancy will be more consistent to infer the effect of noise compared to the case of a varying threshold. For example, high noise is predicted to help survival in high stress (high expression threshold) while hurting survival in low stress (low expression threshold), but this may simplify in the case of a constant threshold with varying mean. Thus, interpreting the effect of noise from an endogenous gene transition into cellular states with a constant threshold can still occur in future studies, though not perfectly rigorous. This thesis will focus on perturbing endogenous *BACH1* expression levels without specifically modulating gene expression noise or studying noise exclusively. Rather, the thesis will monitor potential metastatic state transitions.

The hypothesis of this thesis is significant because all cancer patient deaths are ultimately rooted in chemotherapy resistance from cells within metastatic lesions escaping the primary tumor. Studying mammalian drug resistance with an approach that captures the effect of gene expression noise will provide illuminating insights on how tumor cell heterogeneity affects efficacy of chemotherapy treatment and may influence therapeutic regimes to reduce the number of noisy cells. In addition, endogenous expression as a correlate for metastatic state occupancy has not been demonstrated in living cells. Without creating an endogenous gene expression readout system in living cells, we cannot take the next steps towards understanding how gene expression noise affects metastasis, especially since metastasis is influenced by the tumor microenvironment. Thus, cells allowing measurement of endogenous gene expression of a pro-metastatic gene provide the means for more physiological experiments in the future.

**Chapter 2: Role of network-mediated stochasticity in mammalian drug resistance.**

This chapter is based upon Farquhar, K. S., Charlebois, D., Szenk, M., Cohen, J., Nevozhay, D., & Balázsi, G. Role of network-mediated stochasticity in mammalian drug resistance. Submitted, in review.

**2.1 Introduction.**

**2.1.1 Background.**

More than a decade after the completion of the Human Genome Project, understanding how genes control mammalian cells and organisms remains a daunting task[175]. A major factor contributing to this challenge is the complexity of gene regulation at various scales, from underlying molecular mechanisms to large-scale regulatory networks[176]. Adding to the conundrum is that genetically identical cells can differ drastically due to microenvironmental and stochastic factors[23,38]. Numerous examples over the last two decades indicate that a population of isogenic cells in the same environment can exhibit single cell-level stochastic fluctuations in gene expression, also known as gene expression noise[23,44]. Two main characteristics of gene expression noise are its amplitude and its memory. The amplitude (often measured by the coefficient of variation or CV) defines how far cells deviate from the average. The memory describes the time for which cells remain deviant once they depart from the average[26,177]. These noise characteristics of a gene depend strongly on the regulatory network that embeds it. Positive regulatory feedback typically increases both the amplitude and memory of noise, while negative feedback tends to have the opposite effect[62], implying that network structure and noise characteristics are deeply intertwined and difficult to separate.

Traditional measurements have generated numerous insights by focusing on the gene expression mean and its cellular effects, but we still need to understand the phenotypic roles of gene expression noise in many circumstances[23,38,114,178]. Likewise, approaches that perturb cells in bulk by over-expression, down-regulation, or knockout try to control only the gene expression mean, without precisely adjusting cell-to-cell stochasticity or considering its phenotypic effects[34], which include aiding cell survival during drug treatment[26,43,115] and other forms of environmental stress[179,180]. These effects depend on the amplitude and memory of noise, both of which are network-dependent. The network conferring noise can evolve[181] while improving the adaptive impact of other beneficial mutations under stress[32]. Studies in human cells seemingly suggest that cellular heterogeneity and gene expression noise in general

promote chemotherapy resistance[108,115], evasion of apoptosis[116], and metastasis[1,173]. However, prior demonstration that noise can also be harmful in low stress[43,180] cautions against the generality of such conclusions. Moreover, prior work implies that examining the phenotypic effects of noise requires proper, mean-decoupled noise control[43,182,183], which has not been established for mammalian cells. Therefore, despite the growing interest in the role of mammalian gene expression noise, its precise role in mammalian cell survival and evolution remain open questions. Addressing these questions requires establishing mammalian cell lines that are as similar as possible, differing only in the networks controlling their gene expression noise. To achieve this, one might manipulate the expression of genes by selecting and mixing cells[173], controlling transcriptional regulators, or applying noise-altering chemicals[184]. However, the regulatory networks that control mammalian gene expression are large, complex[185], and incompletely known, making predictable and mean-decoupled noise control for specific individual genes in their native context difficult. Thus, unraveling how gene expression noise of specific networks affects mammalian cell evolution remains a serious challenge.

The field of synthetic biology builds bottom-up synthetic regulatory circuits, which often mimic natural network structures[51,186]. While gene expression noise is difficult to control endogenously, simple synthetic gene circuits have been specifically engineered to modulate noise independently of mean gene expression levels in yeast[43,187] and bacteria[182,183]. In such cases, two non-overlapping noise vs mean curves have decoupled noise regimes (**Figure 6A**), which consist of decoupled noise points (DNPs) where two different noise values correspond to the same mean. Low noise gene circuits for this purpose could include synthetic microRNA-based feedforward loops[188,189] or negative autoregulation[64,68,190] (**Figure 6A**). In contrast, synthetic gene circuits that incorporate positive auto-regulation (**Figure 6A**) or ultrasensitivity have high gene expression noise in yeast[66,67] and mammalian cells[191,192]. Enforcing similar means, but different noise levels in yeast indicated (**Figure 6B**) that noise aids survival in high stress whereas it hinders survival in low stress if the kill curve is sharp. For gradual kill curves, cells with high noise always have a survival advantage regardless of the stress level.
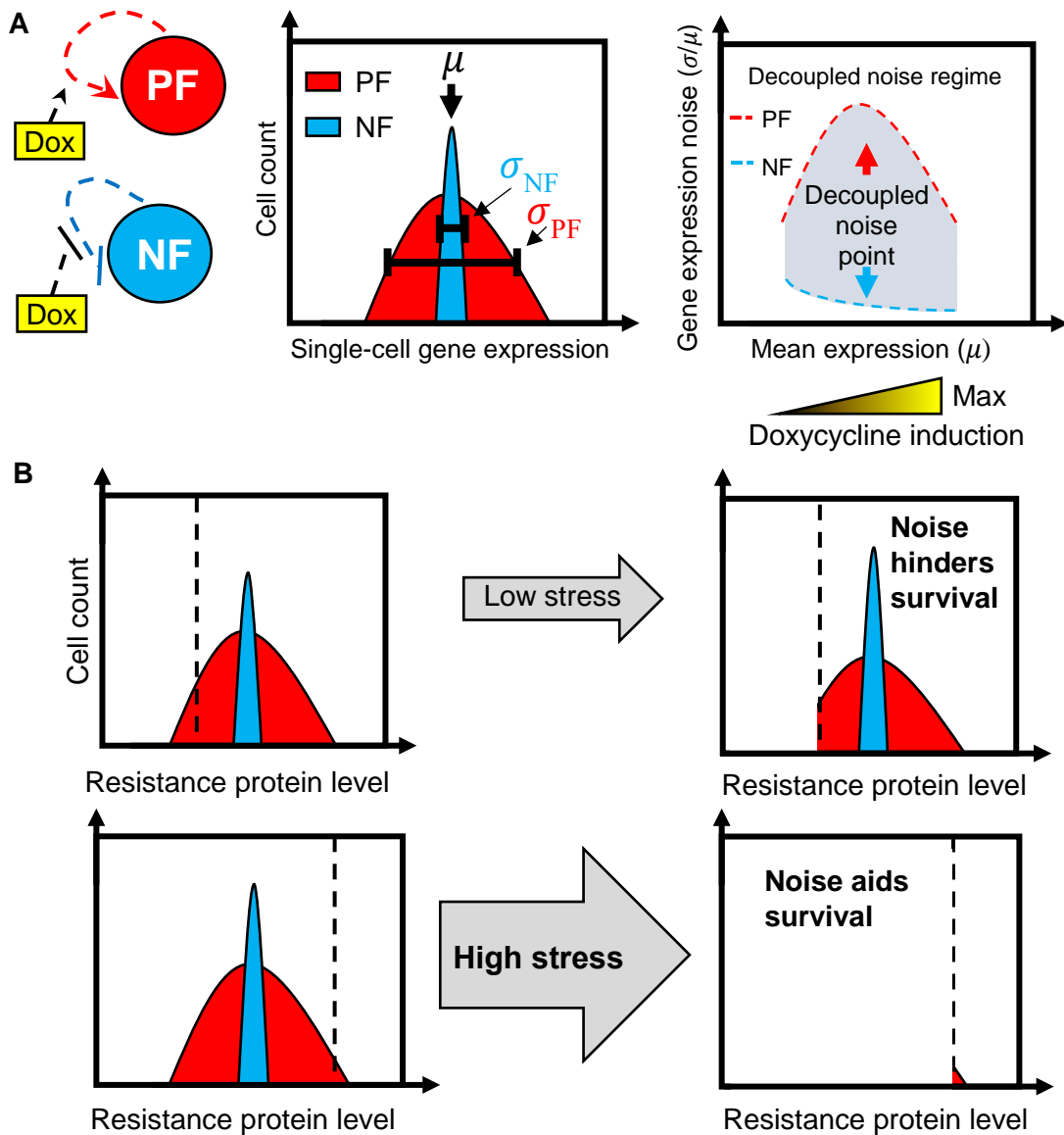
**Figure 6: Stress-dependent effect of network noise on drug resistance.**

**(A)** Mammalian positive (mPF) or negative (mNF) feedback synthetic gene circuits confer high and low gene expression noise while the mean expression is identical, thus decoupling gene expression noise amplitude (standard deviation divided by the mean; $\sigma/\mu$) from the mean within a decoupled noise regime composed of decoupled noise points (arrows). **(B)** Fractional viability under low or high levels of drug (stress) for cells with high (red distribution) or low (blue distribution) drug resistance gene expression noise. Relative survival upon treatment will depend on network noise relative to the fitness function (dashed line).

Testing the role of network structure and noise in mammalian cell evolution as in yeast requires a similar control feat. However, genes or gene circuits integrate randomly into mammalian genomes, which can impose locus-dependent effects on gene expression[44], compromising rigorous noise control in mammalian cells. Therefore, noise-decoupling gene circuits should be reliably integrated at the same transcriptionally active locus to minimize such locus-dependent effects.

Here, we integrate mammalian-optimized high-noise positive-feedback (mPF) and low-noise negative feedback (mNF) synthetic gene circuits (**Figure 6A**) into separate, but isogenic Chinese Hamster Ovary (CHO) cells at the same well-expressed genomic locus by utilizing the Flp-In™ system[193]. By comparing gene expression in CHO cell lines carrying each gene circuit, we establish decoupled noise points with different gene expression noise levels but with similar mean expression. By using these gene circuits to control the expression of the *Puromycin N-acetyl-transferase* (*PuroR*) gene that confers resistance to the antibiotic Puromycin, we investigate how mNF and mPF gene expression noise influences mammalian drug resistance evolution. We find that the mPF gene circuit with high PuroR expression noise can aid long-term evolutionary adaptation of mammalian cells at high stress (Puromycin) level, whereas it has the opposite effect at low stress. Moreover, by withdrawing and re-adding the drug we find that the gene circuit can mutate to adapt stably in mNF cells. On the contrary, cells with the mPF gene circuit do not adapt by intra-network mutations and their resistance is unstable without circuit induction. Overall, combining mammalian synthetic biology with experimental evolution indicates that the noisy mPF network aids adaptation of mammalian cells to high drug levels, while the opposite is true at low drug levels. These findings may have implications for cancer treatment with known regulatory mechanisms of resistance.

**2.1.2 Significance.**

The dichotomy of noise being helpful or hurtful in the short-term depending on the drug treatment dosage has relevance to initial chemotherapy resistance and refractory treatment. When the gene expression noise for a chemoresistance-driving gene is low, the maximum

tolerable concentration of chemotherapeutic agents should optimally kill the resistant tumor cells. By inhibiting gene expression noise for a resistance gene with small molecule noise-suppressors[184], maximal treatment can then effectively kill resistant cells with reduced noise. This chapter will present results supporting this framework and will test the core hypothesis for scenario 1.

**2.2 Methods and materials.**

**2.2.1 Cell culture and maintenance.**

Chinese hamster ovary (CHO) cells with the single stably integrated FRT site (Invitrogen, R75807) were grown in Ham's F-12 Nutrient Mix (Gibco, 11765) with 10% fetal bovine serum (Gibco, 10437) and 100 U/mL Penicillin and Streptomycin (Gibco, 15140). The cells were grown at 37ºC in an incubator filled with 5% $CO_2$. Hygromycin B (Invitrogen, 10687-010) at 700 µg/mL was used as a selection agent that killed untransfected CHO cells and CHO cells with randomly integrated constructs.

During subculture passaging, cells were washed with 1X Dulbecco's Phosphate-Buffered Saline (DPBS) without calcium or magnesium (Life Technologies, 14190250), which was aspirated and replaced with 0.25% Trypsin containing 2.21 mM EDTA (Corning, 25-053-CI). After 2-5 minutes of incubation at 37 ºC, the trypsin/cell mixture was diluted 1:10 with growth medium containing serum and then passaged into a new flask with fresh media in another 1:10 dilution.

**2.2.2 Preparation of reagent working stocks.**

Doxycycline (Fisher Scientific, BP26531) stock solution was stored at -20 ºC at 5 mg/mL and diluted in media at 4 ºC storage for no longer than 7 days after initial preparation. Prior to imaging experiments, CHO cell nuclei were stained with the live cell dye NucBlue (Invitrogen, R37605) at a concentration of 1 droplet per 90 mL of media.

**2.2.3 mPF and mNF plasmid construction.**

During the development of the humanized *TetR* (*hTetR*), an extra glycine amino acid was inserted directly after the start codon to complete a consensus Kozak sequence and thereby enhance translational efficiency[82]. Thus, the amino acid coordinates are shifted by +1 compared to the original *TetR* class B protein sequence (UniProtKB: P04483; PDB: 4AC0). Plasmids integrated into CHO Flp-In™ cells were constructed using restriction cloning on commercial and custom vectors and constructs. The addition of T2A::*PuroR* to both plasmids

resulted in mNF-PuroR and mPF-PuroR constructs. The molecular cloning extensively used overlap PCR extension to fuse DNA pieces together.

The oligonucleotides used in the construction of the mPF and mNF plasmids can be found in **Table 1**. The first intermediate plasmid pDN-D2irTNG5kwh was created by cutting pcDNA5/FRT and pDN-D2irTNG4kwh[82] with the SpeI-HF and SphI-HF restriction enzymes. The 4000 bp fragment from the first (Flp-In expression backbone) and 3400 bp fragment from the second (*hTetR*::NLS::*EGFP* construct with enhancements) plasmids were ligated together to make pDN-D2irTNG5kwh. The P2A sequence was introduced into the circuit by sequential extension PCR with primers CMV-PacI-f, SV40-AscI-BbvCI-r (flanking) and TN-P2A-r, P2A-TN-f, P2A-i2-f, P2A-i1-EGFP-f, EGFP-P2A-f (internal) and then inserting the product fragment into pDN-D2irTNG5kwh (cutting both with SpeI-HF, NotI-HF and ligating) resulting in pDN-D2irTN2AG5kwh (mNF-GFP).

Next, we built the mPF circuits. The pDN-MMa6h plasmid encoding the *rtTA* regulator was made by cutting pDN-D2irTN6kwh[82] with SpeI-HF and XhoI-HF restriction enzymes and using it as a backbone for a subsequent insertion of a PCR fragment amplified by CMV-SpeI-f and rtTA-XhoI-adv-r primers from the pTet-On Advanced plasmid (Clontech, 631069). Then, the pTRE-Tight promoter was PCR amplified from pTRE-Dual2 (Clontech, PT5038-5) with pTRE-Tight-SpeI-f and 2nd-OL-pTRE-hrtTA-r, while *rtTA* was amplified from pDN-MMa6h with 2nd-OL-pTRE-hrtTA-f and hrtTA-SalI-r, having overlap between the proximal ends. After overlap PCR extension of pTRE-*rtTA*, the intermediate plasmid pKF-P14MM5h was created by cutting pTRE::*rtTA* and the pDN-MMa6h vector with SpeI and SalI. We introduced the P2A and *EGFP* sequence with overlap PCR extension by adding P2A to pTRE-*rtTA* with pTRE-Tight-MluI-f and pTRE-rtTA-OL-2A-r and to hEGFP sequentially with 2A-OL-hEGFP-f, 2A-Bridge-OL-f and hEGFP-BamHI-r. Finally, the resulting pTRE-*rtTA*::P2A::*EGFP* construct and the pcDNA5/FRT vector was cut by MluI and BamHI, and ligated to produce the mPF-GFP circuit plasmid pKF-P14MM2AG5h.

| Primer name | Sequence (5' -> 3') |
|---|---|
| CMV-SpeI-f | GCGCACTAGTTATTAATAGTAATCAATTACG |
| rtTA-XhoI-adv-r | GCGCCTCGAGTTACCCGGGGAGCATGTCAAGGTC |
| CMV-PacI-f | GCGCTTAATTAATGACATTGATTATTGACTAGTTATTAATAG |
| SV40-AscI-BbvCI-r | CAGAAGGCTGAGGTAGCGGCGCGCCCCCATAGAGCCCACCGCATCCCCAGC |
| TN-P2A-r | GCTGAAGTTAGTAGCTCCGCTTCCCTTTCTCTTCTTTTTTGGCCCGCCGC |
| P2A-TN-f | GCGGCGGGCCAAAAAAGAAGAGAAAGGGAAGCGGAGCTACTAACTTCAGC |
| P2A-i2-f | GGAAGCGGAGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGA |
| P2A-i1-EGFP-f | CTGCTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCTATGGTGAG |
| EGFP-P2A-f | GGAGGAGAACCCTGGACCTATGGTGAGCAAGGGCGAGGAGCTG |
| pTRE-Tight-SpeI-f | GCGCACTAGTCGAGGCCCTTTCGTCTTCA |
| 2nd-OL-pTRE-hrtTA-r | CTTTGCTCTTGTCCAGTCTAGACATTCCAGGCGATCTGACGGTTCACTAA |
| 2nd-OL-pTRE-hrtTA-f | TTAGTGAACCGTCAGATCGCCTGGAATGTCTAGACTGGACAAGAGCAAAG |
| hrtTA-SalI-r | AAAAGTCGACTTACCCGGGGAGCATGTCAAG |
| pTRE-Tight-MluI-f | GCGCACGCGTCGAGGCCCTTTCGTCTTCA |
| pTRE-rtTA-OL-2A-r | CAGCAGGCTGAAGTTAGTAGCTCCGCTTCCCCCGGGGAGCATGTCAA |
| 2A-OL-hEGFP-f | GCTGGAGACGTGGAGGAGAACCCTGGACCTATGGTGAGCAAGGGCGAGGA |
| 2A-Bridge-OL-f | AGCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAACC |
| hEGFP-BamHI-r | GCGCGGATCCTTACTTGTACAGCTCGTCCATGC |
| T2A-f | GAGGGCAGAGGAAGTCTTCTAACATG |
| NotI-AgeI-PuroR-r | GCGGCCGCACCGGTTCAGGCACCGGGCTTGCG |
| Intron-SbfI-f | CCTACAGGTCCTGCAGGCGCCAC |
| T2A-hEGFP-OL-r | AGGGCCGGGATTCTCCTCCACGTCACCGCATGTTAGAAGACTTCCTCTGCCCTCCTTGTACAGCTCGTCCATGCCG |
| SacI-pCMV-f | GCAGAGCTCGTTTAGTGAACCGT |

**Table 1: List of primers used for mPF and mNF plasmid construction.**

Primers were diluted with molecular-grade water to a working stock of 10 μM from 100 μM

frozen stocks.

T2A::*PuroR* was amplified from DC-RFP-SH01 (GeneCopoeia) with the T2A-f and NotI-

AgeI-PuroR-r primers. To make mNF-PuroR, the PCR product was combined in overlap PCR

extension with the mNF-GFP construct amplified with Intron-SbfI-f and T2A-hEGFP-OL-r. Both

the product and pDN-D2irTN2AG5kwh vector were restriction digested by SbfI-HF and AgeI-HF and ligated to make the pKF-D2irTNP2AG-T2APuroR-5kwh plasmid (mNF-PuroR). The mPF-GFP construct was amplified with SacI-pCMV-f and T2A-hEGFP-OL-r, and combined with T2A::*PuroR* in overlap PCR extension. The resulting product and the pKF-P14MMP2AG5h vector were cut with SacI and NotI-HF and ligated to make the final pKF-P14MMP2AG-T2APuroR-5h plasmid (mPF-PuroR).

### 2.2.4 Flpase-mediated integration of CHO Flp-In cells.

For the genetic constructs containing the *PuroR* gene, CHO Flp-In cells were transfected with plasmid DNA (up to 5 µg) using the Lipofectamine 3000 reagent (Life Technologies, L3000008) according to the manufacturer protocol. Plasmid DNA for mNF-GFP and mPF-GFP was introduced into cells using the Nucleofector$^{TM}$ 2b device (Lonza, Walkersville, MD), per manufacturer protocol, using 5-10 × $10^6$ cells, plasmid DNA (1-5 µg), and relevant buffers (Solution T, and program V-23). Site-specific integration of synthetic gene circuits was achieved by co-transfecting the pOG44 plasmid (Invitrogen, V600520) expressing Flp-recombinase with the Flp-expression vectors that encode an FRT-tagged Hygromycin B resistance gene without a start codon. Upon selecting with Hygromycin B, the resistance gene acts as a positive-selection promoter trap, which provides the resistance gene with a start codon only upon successful integration at the genomic FRT site, thus leading to survival. The clonal CHO populations were derived from bulk-transfected cells by fluorescence-activated cell sorting (FACS) with the FACSAriaIII.

### 2.2.5 Cell memory estimates.

During the CHO mPF and mNF memory estimation experiments, FACSAriaIII sorted the cells with the lowest and highest 15% of *PuroR* expression into separate, terminal wells corresponding to each subsequent day of expression measurement. We estimated the rates of gene expression fluctuations ("switching rates") by flow-sorting of high-expressing and low-expressing mPF and mNF cells induced to the decoupled noise point (DNP) and then monitoring these sorted cells as they returned to their original expression distributions. To

threshold high- and low-expressing subpopulation fractions over time post-sorting, we split the expression values from each sorted sample by the median expression level of a corresponding unsorted control sample.

The mNF high- and low-sort fractions drifted over time to a value higher than the unsorted high fraction (0.5). Since the original distribution did not shift accordingly, the trend did not represent a biological phenomenon; it possibly arose from technical noise after fractioning very similar distributions. To detrend the mNF high- and low-sort fractional data, we individually fit the linear equation $mx + b$, with $x$ equaling the time point (days), to three local time points in each curve with a clearly visible linear trend. The resulting time-dependent linear equation was then subtracted from each mNF high fraction value at each corresponding time point from the curve.

We fit the high subpopulation fractions of both high- and low-sorted cells over time using the following exponential curve, a general solution for two-state phenotype switching models[67,194]:

$$H(t) = A * \left(1 - e^{-(r+f)t}\right) + C, \tag{1}$$

where $H(t)$ is the high subpopulation fraction over time, $r$ is the switching rate from low to high expression (rise rate), $f$ is the switching rate from high to low expression (fall rate), $A$ and $C$ are constants. Fitting low-sorted cells resulted in consistent switching rate values.

Since the individual median threshold at every time point partitions the monitored unsorted, visibly unimodal distributions into two cellular states with equal cell numbers, and cell growth rates do not differ noticeably, the calculated fits for $r$ and $f$ from the sorted cells reflect symmetrical cell switching rates across a consistent midway boundary. Therefore, at each time point, the rise and fall rates are equivalent:

$$r = f = \frac{r+f}{2}. \tag{2}$$

The corresponding rates are listed in **Table 5** while plots of high fractions over time are displayed in **Figure 14**.

### 2.2.6 Flow cytometry expression measurements.

The BD Accuri™ C6 bench-top flow cytometer measured single-cell *EGFP* fluorescence for the *PuroR* experiments. Prior to measurement, trypsinized cell solutions were neutralized with media, then centrifuged for 5 minutes at 300 x g, and the cell pellet was resuspended in 1X DPBS without calcium or magnesium before straining into a 5 mL Polystyene round-bottom tube (Corning, 352235). Between cell measurements, Decontamination solution (BD, 653154) cleaned the flow cytometer liquid stream and the SIP to prevent cells from sticking in the fluidic tubing. In CHO experiments, up to 20,000 events were gated for analysis. CHO cell samples treated with Puromycin typically had lower cell counts. For memory experiments, I measured CHO mPF and mNF expression daily with a BD FACSCalibur Cell Analyzer flow cytometer at the Stony Brook School of Medicine Research Flow Cytometry Core facility.

### 2.2.7 Time-lapse microscopy.

The Nikon TiE motorized inverted microscope acquired 10x phase contrast and fluorescent images at pre-defined x- and y-coordinates for each experiment. Depending on the experiment, a large image consisting of multiple image frames was constructed with 10-15% overlap between frames in the motorized x- and y-coordinates. Exposure times and post-acquisition look-up tables varied between separate experiments but remained consistent within each experiment. The Nikon Perfect Focus System (PFS) prevented focus drift between time points, with off-sets pre-programmed for each individual imaging field per experiment and sample.

### 2.2.8 Kill curves to determine optimal Puromycin dosage.

CHO Flp-in cells (parental) were passaged at 2.5 x $10^5$ cells per well in a 6-well plate to establish toxicity over increasing Puromycin concentrations. Stock Puromycin solution was diluted to 10 µg/mL in media and further diluted to conduct a kill curve with Puromycin concentrations of 1, 3, 5, 7, and 10 µg/mL added a day after cell seeding. Cells were monitored

every day and media was replenished every 3-4 days up to 2 weeks. We found that 10 μg/mL Puromycin was not toxic for both uninduced circuits.

## 2.2.9 Puromycin treatment phase.

Prior to experimental design, a kill curve for ancestral CHO cells evaluated the minimal Puromycin concentration affecting cells without the resistance gene. In the initial drug treatment experiment, cells were seeded at 5 x 10$^4$ cells in 6-well plates and incubated 24h prior to Doxycycline induction. After 48 hours of induction, 8,000 cells were split in replicates of six into 24-well plates and induced for another 24 hours before treatment with Puromycin. At 72 hours, expression was measured by flow cytometry to determine the existence of a decoupled noise point between the two circuits, with the criteria being means that differ by less than 10%. Once the decoupled noise point was established, the cells were treated with varying levels of Puromycin. Plates were replenished with media, Doxycycline, Puromycin, and NucBlue every 24 hours before imaging (**Figure 16A**). If a well became confluent during the first treatment phase, we ran the sample under flow cytometry, temporarily removed Puromycin, split the sample into two new wells with or without Doxycycline, and the remaining cells were cryogenically frozen (**Figure 26A**).

## 2.2.10 Post-treatment and re-treatment phases.

Samples that survived the Puromycin treatment phase were separated into two conditions: i) no Doxycycline and no Puromycin; and ii) with Doxycycline and without Puromycin. To maintain the same concentration of Doxycycline between passages, we neutralized 100 μL of trypsinized solution with 1 mL of either 0.055 or 6.6 ng/mL Doxycycline to dilute the solutions to 0.05 and 6 ng/mL, respectively. The media was replenished with the appropriate induction levels after adherence. Expression was monitored by flow cytometry during each passage. The number of passages for the samples required to reach the re-treatment phase varied between two to nine. If the expression completely reset, the mean expression for the uninduced sample reached lower levels than induced cells over a substantial amount of time (weeks), or the uninduced sample mean expression levels did not change over a month, the samples were retreated with the

previously used Puromycin concentration. All lineages were cryogenically preserved between each passage and after the re-treatment phase. Additionally, uninduced mNF-PuroR or mPF-PuroR cells were measured by flow cytometry at each passage as a positive control for a successful reset.

**2.2.11 CHO genomic DNA extraction and sequencing analysis.**

Ancestral and evolved CHO populations were centrifuged for 5 minutes at 300 x g, and the genomic DNA from the cell pellet was either extracted with the DNeasy Blood & Tissue kit (QIAGEN, 69504) or immediately frozen at –80 Celsius for future extraction. For all samples except mNF-PuroR replicates 1, 3, 4, 5, and 6, the entire sample were immediately centrifuged after thawing from cryopreservation. Otherwise, one-tenth of the sample was grown to confluence up to a 6-well plate, which was then lysed for DNA extraction. Once purified, the genomic DNA acted as a template for PCR amplification of the mNF-PuroR or mPF-PuroR circuits using primers listed in **Table 2**. Various sequencing primers were used for Sanger sequencing such that the chromatographs covered the circuits with at least 2 reads.

| Primer name | Sequence (5' -> 3') |
|---|---|
| D2ir-Enh-Seq-f | ACGGGCCAGATATACGCGTT |
| NF-Edge-f | GCGTTGACATTGATTATTGACTAGTT |
| FRT-Circuit-3-r | GCTGGTTCTTTCCGCCTCAG |
| MluI-pTRE-f | ATATACGCGTCGAGGCCCTTTC |
| hEGFP-130-qPCR-f | ACGACGGCAACTACAAGAC |
| bGH-pA-r | GCTGGTTGCTAAGAGGGAGG |
| FRT-Circuit-2-r | CGAACGTGGCGAGAAAG |

**Table 2: List of primers for circuit sequencing.**

Blue shaded rows contain primers ampifying the mNF-PuroR circuit, red rows indicate primers used for amplifying mPF-PuroR circuits, and primers in purple rows were used for both circuits.

The ab1 files were aligned and peaks visualized with the SnapGene software (from GSL Biotech; available at snapgene.com). To assess for genetic heterogeneity in a sequencing read with peak mixtures, CRISP-ID[195] analyzed individual chromatographs for variant subpopulation sequence. Background cutoff percentages (ranging from 10-25%) were scaled to the quality of

nucleotide calls. Each read was evaluated starting at 85 bp and up to 600 bp depending on the location of poor read quality. In doing so, we can distinguish genetic heterogeneity of potential subpopulations in the sequence reads from mixed peaks and minimize false positives derived from poor read quality.

## 2.2.12 PuroR circuit copy number analysis with quantitative PCR.

The qPCR reactions were measured by the Applied Biosystems™ QuantStudio™ 3 Real-Time PCR system (ThermoFisher). PowerUp™ SYBR® Green Master Mix (ThermoFisher, A25742) amplified the genomic DNA templates and emitted fluorescence upon binding to double-stranded DNA. The standard reaction condition for the master mix was set on the qPCR instrument. Each reaction contained a total volume of 10 µL with 500 nM primers. Reactions were prepared as a serial dilution of each genomic DNA sample in triplicate over 5-fold changes in concentration. *PuroR* and *Vinculin* (*Vcl*) were targeted as the unknown target and known copy number reference standard, respectively. CHO cells contain two copies of *Vinculin*[196]. For each genomic DNA sample, we used the $2^{-\Delta Ct}$ relative quantification method, where $\Delta Ct = Ct_{target} - Ct_{reference}$. To assess the upper bounds of error, we chose to analyze the serial dilution concentration with the largest set of deviation between Ct values. Results indicating a single copy of each circuit per cell should overlap 0.5 in terms of relative copy number of *PuroR* over *Vcl*, with two copies overlapping 1, three overlapping 1.5, etc. Since the box plot comparison intervals (95%) for each circuit (bounded by red notch markers) overlap 0.5 but not 1 or higher, we reject the hypothesis that there is more than one circuit at the 0.05 significance level. The qPCR primers are listed in **Table 3**.

## 2.2.13 Flow data analysis.

We developed custom MATLAB scripts to gate and analyze flow cytometry data. Cells were adaptively gated with a density-threshold fit of log-transformed SSC and FSC values per sample to exclude debris. Specifically, flow cytometry data was exported as individual FCS files, which were then analyzed by custom MATLAB scripts. Raw forward and side scatter values were log-transformed, and then plotted as a 2-dimensional histogram. The number of

| Primer name | Sequence (5' -> 3') |
|---|---|
| PuroR-150-qPCR-f | CGAGTACAAGCCCACGGT |
| PuroR-150-qPCR-r | AGTTCTTGCAGCTCGGTGA |
| Vinculin fwd | GCTGGTTGCTAAGAGGGAGG |
| Vinculin rev | ATCAGAGGCAGCTTTCACGG |

**Table 3: List of primers used for qPCR copy number analysis.**

Primers were diluted with molecular-grade water to a working stock concentration of 10 µM

from 100 µM frozen stocks. The *Vinculin* primer sequences were obtained from the

literature[2].

bins was usually 60 unless adjusted whenever debris was unintentionally gated. The 2-

dimensional histogram counts were then plotted as a contour, which further subdivided the plot

with a density gate. We chose the second from the widest contour level on average, which

increased in density at the SSC-FSC coordinates harboring cellular events. Cells must make

up most events in an FCS file for the algorithm to work.

Fluorescent events that were less than 2,000 arbitrary units were filtered prior to

normalization using the following formula:

$$EGFP^{i}_{\text{filtered}} = EGFP^{i}_{\text{raw}} \mid (EGFP^{i}_{\text{raw}} > EGFP_{cutoff}), \tag{3}$$

where $EGFP^{i}_{filtered}$ is the filtered fluorescence for individual cells (i) that have an unsilenced

circuit, $EGFP^{i}_{raw}$ is the raw fluorescence from any cell (i) with the circuit, and $EGFP_{cutoff}$ is a

constant fluorescence threshold (2,000 arbitrary units) above auto-fluorescence but below

uninduced, basal circuit expression. The $EGFP_{cutoff}$ was shown to threshold non-fluorescent

cells from cells with uninduced circuits based on the threshold value exceeding three standard

deviations higher than the auto-fluorescence mean. Additionally, the non-expressing peak and

basal expression peak flank a local minimum (**Figure 13C**) that aligns with the $EGFP_{cutoff}$. The

filter was justified by the lack of transitions from the small subpopulation of non-expressing cells

to the basal expression level for both circuits (**Figure 7**).

To estimate technical variation in fluorescence under flow cytometry between

experiments, we measured fluorescent values from an auto-fluorescence reference (CHO Flp-
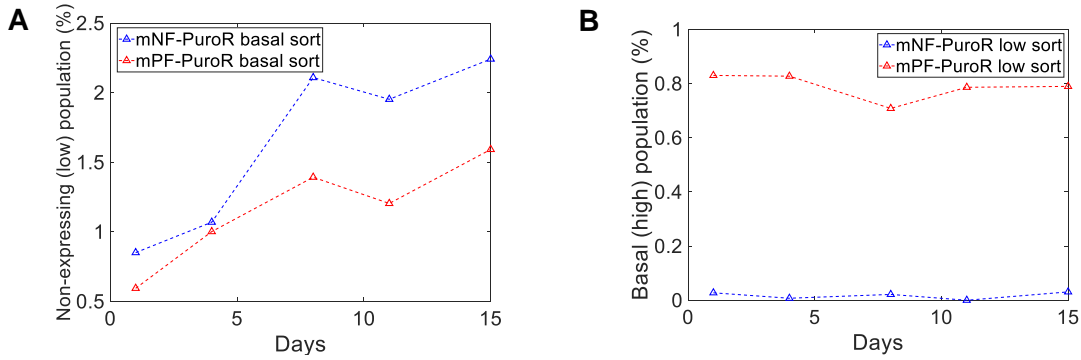
**Figure 7: FACS-sorting of non-expressing and basal CHO-PuroR subpopulations revealed no switching from the non-expressing state.**

**(A)** Percent fraction of the non-expressing subpopulation over time after sorting for cells with basal (uninduced) expression levels. **(B)** Percent fraction of the basal-expressing subpopulation over time after sorting for the non-expressing subpopulation.

In parental cell line) and 8-peak fluorescent calibration beads (BD, 653144). To normalize the gated, filtered fluorescence events, we applied the technical fluorescence variation control data to the following formula:

$$EGFP_{\text{norm}}^{\text{i}} = \frac{EGFP_{\text{filtered}}^{\text{i}} - EGFP_{\text{auto}}}{EGFP_{\text{max}}}, \tag{4}$$

where $EGFP_{\text{norm}}^{\text{i}}$ is the normalized fluorescence from individual cells with an unsilenced circuit, $EGFP_{\text{auto}}$ is the mean autofluorescence from the parental cell line without the circuit, and $EGFP_{\text{max}}$ is the mean fluorescence of the highest fluorescence intensity peak from the calibration beads control.

Using each individual normalized fluorescence reading, we calculated the normalized mean and CV directly from the $EGFP_{norm}^{i}$ values using standard formulas: $CV = \sigma(EGFP_{\text{norm}}^{\text{i}})/\mu(EGFP_{\text{norm}}^{\text{i}})$ where $\mu(EGFP_{\text{norm}}^{\text{i}}) = \frac{1}{\text{N}}\sum_1^{\text{N}} EGFP_{\text{norm}}^{\text{i}}$ and $\sigma^2(EGFP_{\text{norm}}^{\text{i}}) = \frac{1}{\text{N}}\sum_1^{\text{N}}(EGFP_{\text{norm}}^{\text{i}})^2$. We plotted the mean and standard error of the mean (SEM) over replicates for each condition.

**2.2.14 Linearity assessment with L1-norm and curve fitting.**

To assess the degree and range of linearity, we calculated the L1-norm statistic[68,82] for the mean expression levels from the mNF-PuroR and mNF-GFP dose-responses. The statistic can be thought of as the distance between a mean expression dose-response and an ideal linear function, with 0 representing perfect linearity and 0.5 being the least linear[68,82]. First, the doses and means were rescaled so that the first point in the new unit vector coordinate system was (0,0) and the last dose and mean was (1,1). From these discrete values, we finely interpolated the means between doses with the *interp1.m* MATLAB function. Then, the area between the ideal linear function and the rescaled interpolated dose-response was calculated with the *trapz.m* MATLAB function. This integrated area represents the L1-norm statistic up to each dose. We included 4 or more doses in each L1-norm statistic calculation. The plots in **Figure 11A,C** display L1-norms that are generally minimal over low to intermediate Doxycycline concentrations. The L1-norm abruptly increases once doses induce saturating means. To calculate the coefficients of determination or R-squared values of linearity for mean expression dose-responses, the mean expression data was also fit to a linear curve.

**2.2.15 Adaptation criteria.**

To calculate the adaptation time parameter, we fit the Baranyi environmental adjustment model[197] through minimization between the model and the data using Powell's method. We employed custom MATLAB scripts to estimate the local slope (growth rate) of the growth curves with a moving window of 3 timepoints (**Figures 19; 40C,D**). If the local growth rates up to saturation of the growth curve were equal to or greater than 0, then the adaptation time = 0. If any local growth rate of a curve was less than zero, then we extracted the half-saturation time and adaptation time from the Baranyi model. Replicate populations that completely die off have adaptation times = infinity.

**2.2.16 Statistical analysis.**

All statistical analyses employed custom MATLAB scripts. Given that the single-cell expression data consists of many cells per sample, the Central Limit Theorem states that the distribution of means for each sample is approximately normally distributed. Thus, parametric

tests were appropriate to infer differences in mean expression. Specifically, significant differences in mean expression were evaluated with the two-sample t-test with 0.05 set as the significance level. The coefficient of variation (standard error divided by the mean) of each distribution in a set were assumed to not follow a normal distribution. Thus, statistical inferences on CV and all other variables required the two-sample Mann-Whitney U test. The two-sample t-test and Mann-Whitney U test inferred significance at a p-value < 0.05.

**2.2.17 Image processing.**

In imaging, the number of cells per field was determined by spot detection of green-fluorescent cells and nuclei stained with the live cell dye NucBlue using NIS Elements AR software. Specifically, the motorized Nikon TiE microscope at a magnfication of 10x captured phase contrast and fluorescent images every 24 hours after initial Puromycin treatment. Each individual sample in the Puromycin treatment experiments have image fields tiled together in 2x3 fields with 15% overlap but not with active stitching (NIS Elements AR). Fluorescence background subtraction used an image that was a median of multiple background images of empty wells. The median background image was extracted with custom MATLAB scripts and then subtracted from the raw fluorescence images using the Nikon Elements AR software.

To determine the number of CHO cells in each field, distinct nuclear stained cells and green fluorescent cells were counted with a bright spot detection algorithm provided by the Nikon Elements AR software. A typical diameter between 9 to 12 µm and an average contrast of 22.7 was applied in each run of the algorithm. The resulting binary objects were merged with an OR gate and then the objects were morphologically separated with a structual element repeated throughout the image. Each object was then counted per image to construct the cell growth curves.

**2.2.18 Computational modeling.**

Used with permission from Daniel A. Charlebois, who developed the methodology, ran the simulations, and wrote the text in this section.

To relate the fraction of mPF-PuroR and mNF-PuroR cells surviving the death phase of Puromycin treatment to *EGFP* and Puromycin concentrations, we first log transformed the approximately lognormal experimental CHO *EGFP* histograms, which were then fit to the standard probability density function (**Figure 24A**):

$$f(EGFP_{mPF,mNF}) = \frac{1}{\sqrt{2\pi}\sigma_{mPF,mNF}} \exp\left(\frac{-(EGFP_{mPF,mNF} - \mu_{mPF,mNF})^2}{2\sigma^2_{mPF,mNF}}\right), \tag{5}$$

where *EGFP* is the level of the fluorescence reporter which corresponds to the expression level of the Puromycin resistance gene (*PuroR*), and $\mu$ and $\sigma$ are the mean and standard deviation of the population *EGFP* distribution for the mammalian positive or negative feedback circuit as indicated by the indices *mPF* and *mNF*, respectively.

The cumulative distribution function was directly obtained by integrating Equation (5) and is related to the fraction of cells that are initially killed by Puromycin ($A_D$):

$$A_{D_{mPF,mNF}}(EGFP) = \int_0^{A_{D_{thres}}} \frac{1}{\sqrt{2\pi}\sigma_{mPF,mNF}} \exp\left(\frac{-(EGFP'_{mPF,mNF} - \mu_{mPF,mNF})^2}{2\sigma^2_{mPF,mNF}}\right) dEGFP' . \tag{6}$$

The fraction of clonal CHO cells that initially survive Puromycin ($A_S$) is then simply 1-$A_D$ (**Figure 24B**). The *EGFP* expression threshold below which CHO cells were killed is related to Puromycin concentration ([Puro]) via a Michaelis-Menten type function (**Figure 24C**):

$$A_{D_{thres}}(Puro) = \beta\left(\frac{[Puro]}{K + [Puro]}\right). \tag{7}$$

The fraction of initial surviving cells can be further divided into a persister cell fraction ($A_P$) and a nongenetically resistant fraction ($A_N$) such that $A_D + A_P + A_N = 1$. The $A_P$ fraction was estimated using a lognormal distribution function (**Figure 24D**):

$$A_{P_{mPF,mNF}}([Puro]) = \frac{1}{\sqrt{2\pi}\sigma'_{mPF,mNF}} \exp\left(\frac{-(\ln([Puro]) - \mu'_{mPF,mNF})^2}{2\sigma'^2_{mPF,mNF}}\right). \tag{8}$$

The initial subpopulation fractions of $A_D$, $A_P$, and $A_N$ served as input to a stochastic population dynamics model [Equation (9)], which accounted for the phenotype switching

43

between *P* and *N* cells and the mutation of *N* or *P* cells to form a genetically drug resistant subpopulation fraction of *G* cells ($A_G$), which like *N* cells could grow and divide. The model predicted subpopulation fractions and adaptation (**Figure 23**):

$$\frac{dP}{dt} = r_{P,N}N - r_{N,P}P - r_{G,P}P$$

$$\frac{dN}{dt} = -r_{P,N}N + r_{N,P}P - r_{G,N}N + k_N N - g_N N \tag{9}$$

$$\frac{dG}{dt} = r_{G,P}P + r_{G,N}N + k_G G - g_G G$$

where $r_{i,j}$ is transition rate from genotype/phenotype *j* to *i*, $k_i$ is the growth rate of *i*, and $g_i$ is the death rate of *i*. The total population size $N_{tot}$ is given by:

$$\frac{dN_{tot}}{dt} = \frac{dP}{dt} + \frac{dN}{dt} + \frac{dG}{dt} = k_N N + k_G G - g_N N - g_G G . \tag{10}$$

All parameter values are given in **Table 4**.

| Function/Simulation Parameters | Strain | Symbol | Value |
|---|---|---|---|
| *f(EGFP)* and $A_{DmPF,mNF}$, Equations (5) & (6) in Methods section 2.2.18 | | | |
| | mPF | μ | 4.4094 |
| | | σ | 0.3835 |
| | mNF | μ | 4.5452 |
| | | σ | 0.1774 |
| $A_{Dthres}$, Equation (7) in Methods section 2.2.18 | | | |
| | | β | 5.1 |
| | | K | 4 |
| $A_{PmPF,mNF}$, Equation (8) in Methods section 2.2.18 | | | |
| | mPF | μ' | 3.4556 |
| | | σ' | 0.2599 |
| | mNF | μ' | 3.3176 |
| | | σ' | 0.2392 |
| Population Model | | | |
| | | μ | 0.0556 |
| | | $\eta_1$ | $[1, 5, 4, 10, 10]\times10^2$ |
| | | $\eta_2$ | $[1, 5, 2.5, 10, 5]\times10^2$ |
| | mPF | $N_0$ | $[0.1, 0.15, 0.2, 0.1, 0.1]\times10^4$ |
| | | $N_{max}$ | $[6, 3, 3.25, 2.5, 2]\times10^4$ |
| | | $r_{P,N}$ | $[1, 1, 1, 1, 1]\times10^{-4}$ |
| | | $r_{N,P}$ | $[1, 1, 1, 1, 1]\times10^{-2}$ |
| | | $r_{G,N}$ | $[0, 1, 1, 1, 1]\times10^{-7}$ |
| | | $k_N$ | $[1, 0.825, 0.6, 0.45, 0.4]\times\mu$ |
| | | $g_N$ | $[1, 2.5, 3, 3.5, 5]\times10^{-2}$ |
| | | $r_{G,P}$ | $[0, 5, 5, 5, 5]\times10^{-6}$ |
| | | $k_G$ | $[1, 0.825, 0.8, 0.45, 0.4]\times\mu$ |

| | | $g_G$ | $[1, 1, 1, 1, 1]\times10^{-2}$ |
|---|---|---|---|
| | mNF | $N_0$ | $[0.085, 0.2, 0.25, 0.1, 0.1]\times10^4$ |
| | | $N_{max}$ | $[6, 4, 3, 2.25, 2]\times10^4$ |
| | | $r_{P,N}$ | $[1, 1, 1, 1, 1]\times10^{-4}$ |
| | | $r_{N,P}$ | $[1, 1, 1, 1, 1]\times10^{-2}$ |
| | | $r_{G,N}$ | $[0, 1, 1, 1, 1]\times10^{-6}$ |
| | | $k_N$ | $[1, 0.825, 0.8, 0.45, 0]\times\mu$ |
| | | $g_N$ | $[1, 2.5, 3, 3.5, 5]\times10^{-2}$ |
| | | $r_{G,P}$ | $[0, 5, 5, 5, 5] \times10^{-6}$ |
| | | $k_G$ | $[1, 0.825, 0.8, 0.45, 0.4]\times\mu$ |
| | | $g_G$ | $[1, 1, 1, 1, 1]\times10^{-2}$ |

**Table 4: Parameters for stochastic population dynamics model.**

Used with permission from Daniel A. Charlebois, who constructed the table.

Parameters provided in an array correspond to increasing Puromycin concentrations ([0 10 22.5 35 50] μg/ml). Single parameter values indicate the value used for all Puromycin treatment conditions. The same parameters were used for mPF and mNF strains unless otherwise indicated.

**2.3 Results.**

**2.3.1 Developing mPF-PuroR, a high-noise mammalian puromycin resistance synthetic gene circuit.**

To obtain high gene expression noise amplitude and memory, we designed and assembled a Flp-In-compatible version of the positive feedback (PF) synthetic gene circuit[67]. We integrated this mammalian PF-PuroR (mPF-PuroR or mPF) gene circuit into the well-expressed genomic FRT site into clonal Chinese Hamster Ovary (CHO) Flp-In™ cells to avoid genomic locus-dependent variation in silencing. In mPF-PuroR, the reverse tetracycline Trans-Activator (*rtTA*) binds to Doxycycline (Dox) and activates the transcription of a tricistronic construct consisting of that same *rtTA* regulator, the fluorescent reporter *EGFP*, and the drug resistance gene *PuroR* (**Figure 8A**). Thus, with Doxycycline induction, the positive auto-regulatory network increases fluctuations in gene expression within a population of cells. We joined these coding sequences transcriptionally using the self-cleaving Porcine teschovirus-1 2A (P2A) and Thosea asigna virus 2A (T2A) peptides to prevent potential unwanted functional effects from protein fusion[198]. Once translated, the P2A and T2A peptide motifs cleave themselves, leading to the expression of three separated proteins from one transcript. This simple design, with a single common promoter, minimizes the number of genetic components in the mPF-PuroR gene circuit, facilitating genomic integration.

To characterize the expression of the mPF-PuroR gene circuit, we collected single cell-level *EGFP* fluorescence data at varying Doxycycline levels by flow cytometry. We normalized single cell-level EGFP fluorescence data to minimize technical variation from flow cytometry measurements by correcting for auto-fluorescence and then dividing the corrected fluorescence by the mean of one fluorescence peak from flow cytometry calibration beads [Equations (1)-(2), Methods section 2.2.13]. We characterized these normalized EGFP fluorescence distributions in terms of their gene expression mean and noise amplitude, quantified by the CV. The mean mPF-PuroR expression dose-response was sigmoidal with a steep response region (**Figure 8B**), similar to yeast[67]. Gene expression noise amplitude for uninduced mPF-PuroR cells was

46
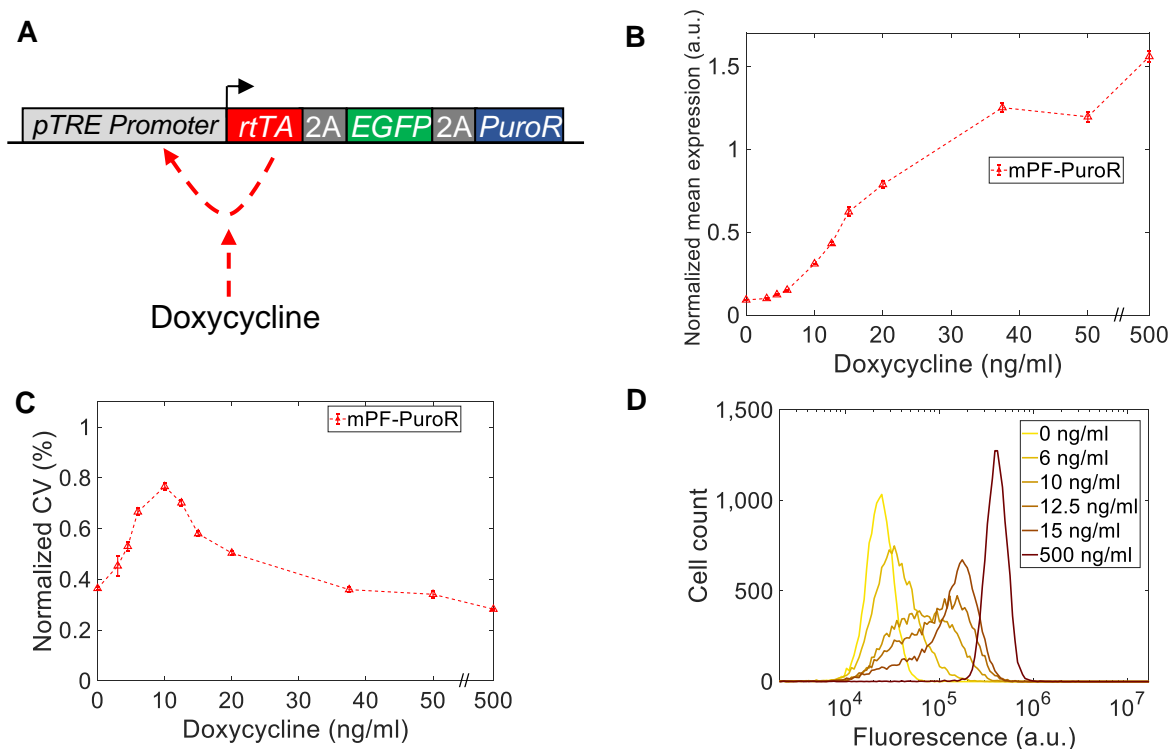
**Figure 8: Dose-response of the mPF-PuroR gene circuit.**

**(A)** Network schematic of the mPF-PuroR gene circuit induced by Doxycycline , which expresses the Puromycin resistance gene (*PuroR*) and *EGFP* separated by the self-cleaving 2A elements. **(B)** Mean expression under varying levels of Doxycycline induction. **(C)** Gene expression noise amplitude (coefficient of variation, CV) in response to Doxycycline induction. Error bars denote the standard error of the mean. There is an x-axis break (//) between 50 and 500 ng/mL Dox. All samples were measured in triplicate. **(D)** Single-cell gene expression distributions of mPF-PuroR cells with broad peaks at intermediate levels of Doxycycline (legend). Distributions are from representative replicates.

low, but then increased markedly upon Doxycycline induction (**Figure 8C**). The highest noise amplitude values corresponded to broad, yet visibly unimodal single-cell expression distributions (**Figure 8D**) in contrast to the bimodal distributions in yeast[67]. Importantly, the removal of *PuroR* did not impact the performance  (noise amplification) of the mPF circuit (**Figure 9**), indicating that it can impose high noise for various genes. To summarize, transferring the mPF-PuroR gene circuit into CHO Flp-In cells led to high noise amplitude with

broad, visibly unimodal distributions.
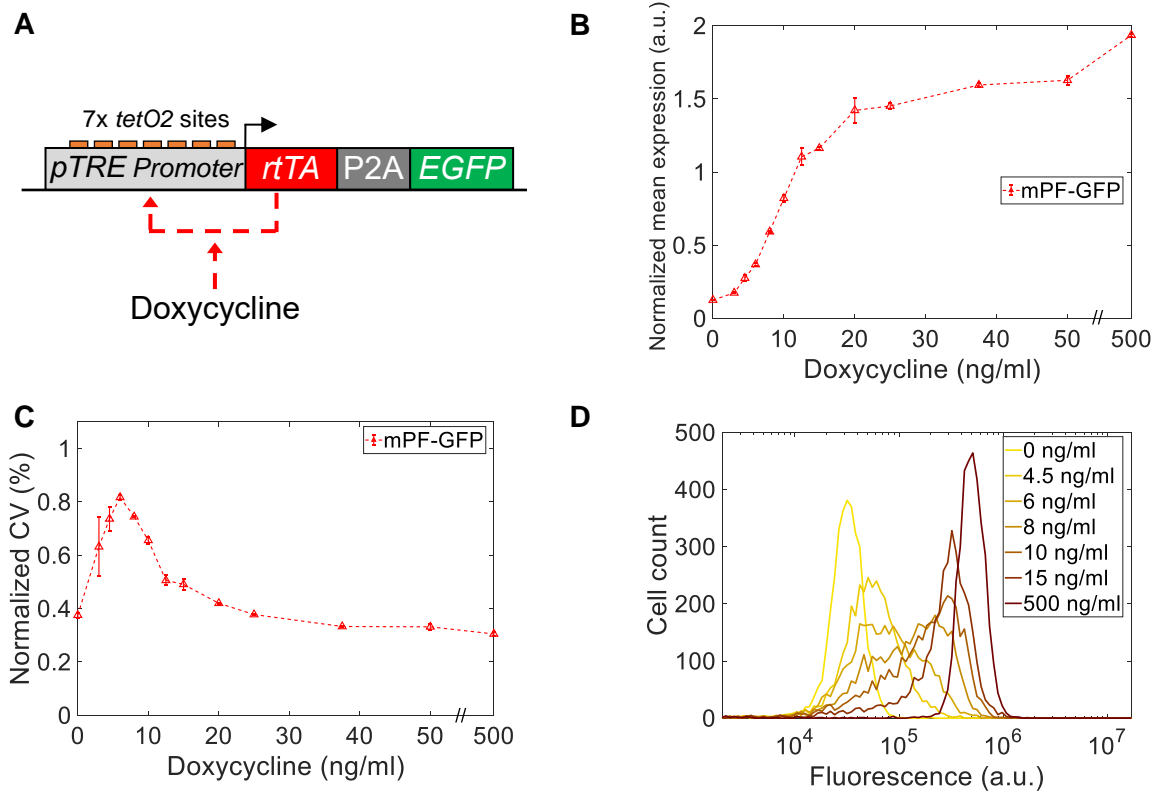


**Figure 9: The mPF-GFP gene circuit amplifies gene expression noise.**

**(A)** The Flp-In-compatible mammalian positive feedback circuit without *PuroR* (mPF-GFP) contains the Doxycycline-inducible transcriptional regulator *rtTA* which activates its own gene expression (red arrows) by binding to *tetO2* sites in the pTRE-Tight promoter. **(B)** Mean expression of the mPF-GFP circuit in response to increasing levels of Doxycycline. **(C)** Gene expression noise (coefficient of variation) over increasing concentrations of Doxycycline. All conditions were run in triplicate. Error bars represent the standard error of the mean. **(D)** Single-cell expression distributions in response to increasing Doxycycline concentrations (legend). The distributions are from a representative replicate.

**2.3.2 The low-noise mNF-PuroR gene circuit allows noise decoupling in conjunction with mPF-PuroR.**

To generate a low-noise gene circuit in the same genomic locus, we also integrated a Flp-In-compatible mammalian negative feedback (mNF-PuroR or otherwise called mNF) circuit

48

in the same ancestral CHO cell line (**Figure 10A**). With negative feedback, gene expression fluctuations are suppressed[190]. We preserved previous optimizations that enhanced gene expression, including mammalian-optimized codons, post-transcriptional expression elements, and a fold-activation-optimized promoter[82]. Again, we joined the humanized Tetracycline repressor  (*hTetR*), the *EGFP* reporter, and *PuroR* resistance genes with P2A and T2A peptide motifs, allowing co-translational separation of the three proteins.
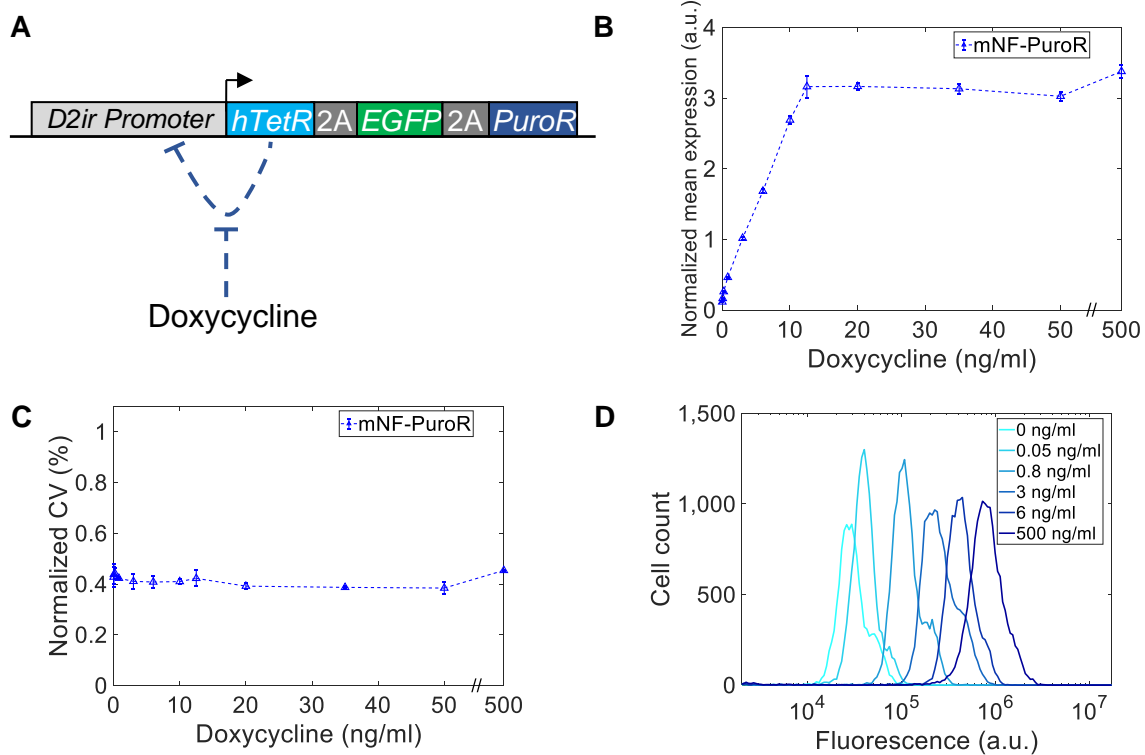


**Figure 10: Dose-response of the mNF-PuroR gene circuit.**

**(A)** The mNF-PuroR gene circuit controls expression of a Puromycin resistance gene and the *EGFP* reporter gene through negative feedback of a humanized tetracycline repressor (*hTetR*) gene separated by self-cleaving 2A peptides. **(B)** Normalized mean expression of mNF-PuroR cells under varying levels of Doxycycline (Dox). **(C)** Gene expression noise of mNF-PuroR cells in response to Dox. Error bars denote the standard error of the mean. There is an x-axis break (//) between 50 and 500 ng/mL Dox. Samples were measured in triplicate. **(D)** Single-cell gene expression distributions of the mNF-PuroR circuit. The legend indicates Dox concentrations. Distributions are from representative replicates.

To determine how the gene expression mean and noise amplitude of the mNF-PuroR circuit depend on Doxycycline, we obtained gene expression distributions by flow cytometry. As expected[82], the mNF-PuroR gene expression mean increased linearly with Doxycycline concentrations prior to saturation (**Figures 10B; 11A,B**). Linearity was lost when expression saturated (**Figure 11A,B**). We observed low gene expression noise amplitude in response to Doxycycline (**Figure 10C**), in agreement with narrow gene expression distributions (**Figure 10D**). Removing *PuroR* did not affect the performance of the mNF gene circuit (**Figures 11C,D; 12**).
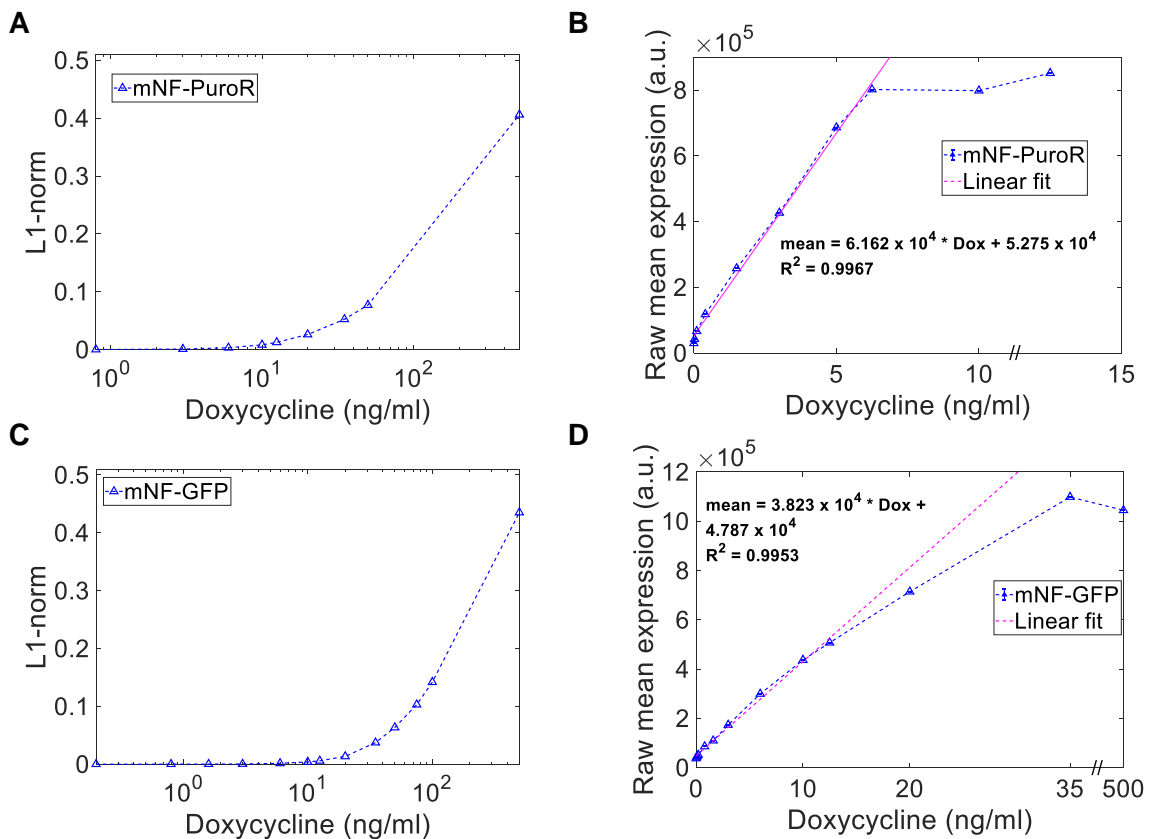


**Figure 11: Assessment of linearity for the mNF-PuroR and mNF-GFP mean dose-responses by L1-norm and curve fitting.**

**(A)** L1-norm statistics and **(B)** linear curve fits for the mNF-PuroR mean dose-response. **(C)** L1-norm and **(D)** linear curve fits for the mNF-GFP mean dose-response. The calculations incorporated average expression over replicates.
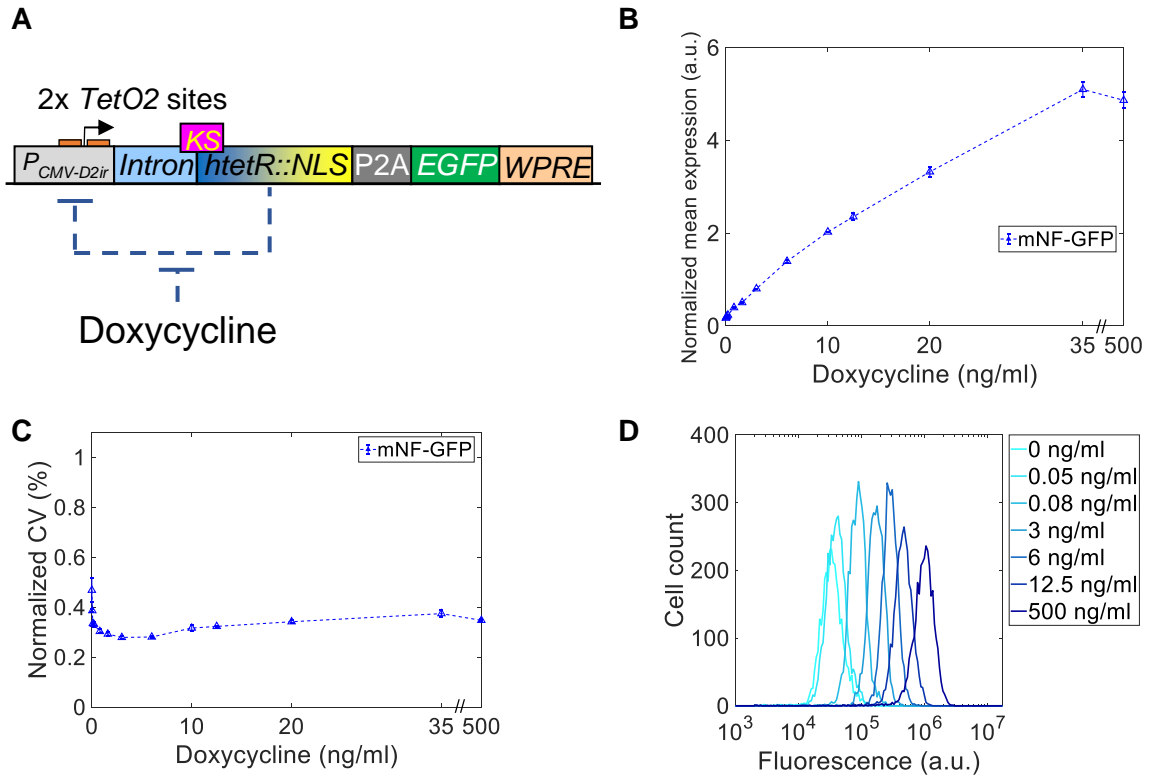
**Figure 12: The Flp-In mNF-GFP circuit uniformly expresses *EGFP*.**

**(A)** The Flp-In-compatible mammalian NF (mNF-GFP) circuit lacking *PuroR*, engineered to linearly tune mean expression of a gene with minimal gene expression noise, contains several genetic elements intended to enhance expression, which is also in mNF-PuroR. **(B)** Normalized mean expression of mNF-GFP cell populations in response to increasing levels of Doxycycline. **(C)** Gene expression noise amplitude for the mNF-GFP circuit as quantified by the coefficient of variation (CV). Samples were run in triplicate. Error bars represent the standard error of the mean. The data was normalized as described in the methods. **(D)** Single-cell expression distributions for the mNF-GFP circuit. Distributions are from representative replicates.

To test whether noise-mean decoupling is possible with mNF-PuroR and mPF-PuroR, we sought Doxycycline induction levels where the mean expression of the two gene circuits were similar while the differences in noise amplitude were high (**Figure 6A**). To identify such a decoupled *PuroR* noise regime from the dose-response, we analyzed noise amplitude as a function of mean expression (**Figure 13A**). We observed two decoupled noise regimes, one
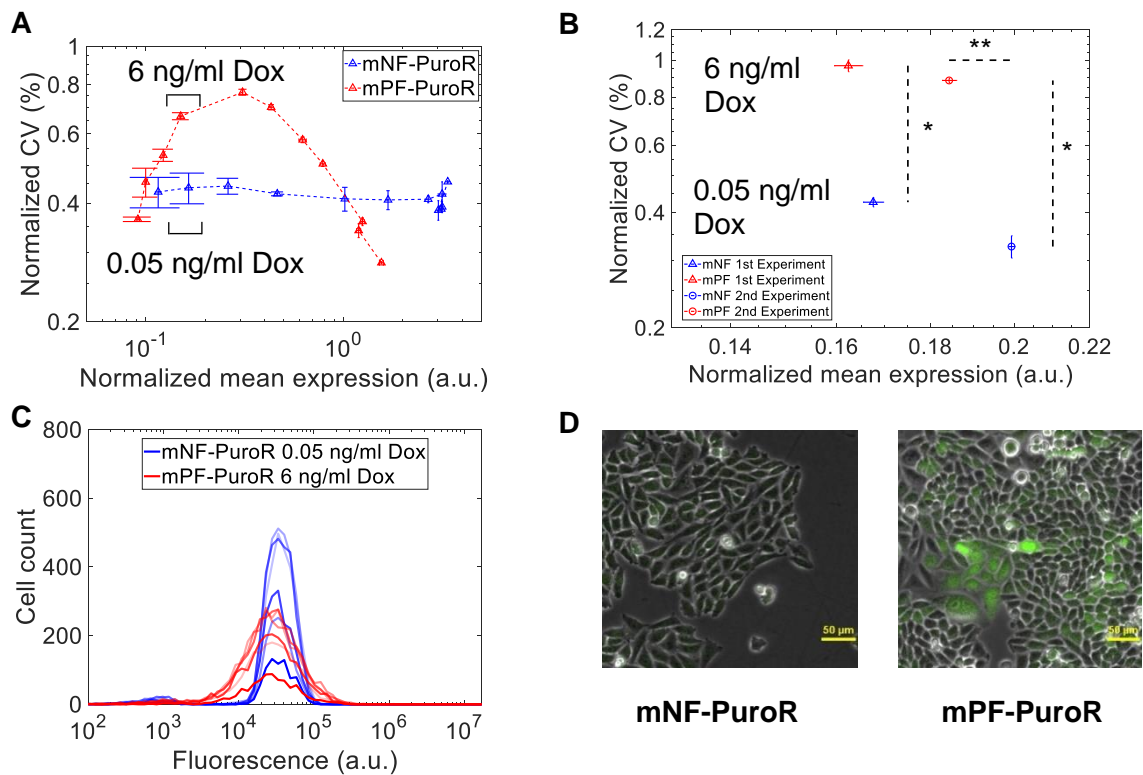
**Figure 13: A decoupled noise regime provides a series of decoupled noise points prior to drug treatment.**

**(A)** Plotting the noise (coefficient of variation, CV) as a function of mean gene expression for both gene circuits revealed two decoupled noise regimes. Black brackets indicate the expression range for the Doxycycline (Dox) concentrations used for noise-mean decoupling in **(B)**. Error bars represent the standard error of the mean. **(B)** Decoupled noise points (DNPs) at the beginning of the two drug treatment experiment sets. The noise was significantly different between gene circuits for both sets (*p-value = 0.0022; n = 6, Mann-Whitney U test). The mean expression was not significantly different for set 1 (~3% difference; p-value = 0.1770; n = 6, two-sample t-test) while it had significance for set 2 (~8% difference; **p-value = 7.9416 x 10$^{-6}$; n = 6, two-sample t-test). The dashed lines display the range of significant mean-noise differences between the circuits. **(C)** Gene expression distributions at the DNP from the first set. **(D)** Images of cells at the decoupled noise point from the first set. Statistical tests inferred significance at p-values < 0.05.

before and one after the mNF-PuroR and mPF-PuroR noise-mean curves intersect at high mean expression. Each regime has a set of decoupled noise points (DNPs) where circuits can have matching mean expression while ensuring distinct noise amplitudes. Thus, we decoupled gene expression noise amplitude of the *PuroR* gene from mean gene expression, providing a number of decoupled noise points to initiate the Puromycin treatment experiments.

**2.3.3 Estimating cellular memory by flow-sorting the mNF-PuroR and mPF-PuroR cells by high- and low-expressing subpopulations.**

Besides the noise amplitude, we also studied the memory, estimating the rate at which cells moved within the distributions. Flow-sorting high- or low-expressing subpopulations for both circuits at the drug treatment induction levels indicated that cells with the high-noise mPF-PuroR circuit have higher memory (~2 days) than cells with the low-noise mNF-PuroR circuit (~1/2 day) (**Figure 14; Table 5**). Overall, the mNF-PuroR and mPF-PuroR cell lines were
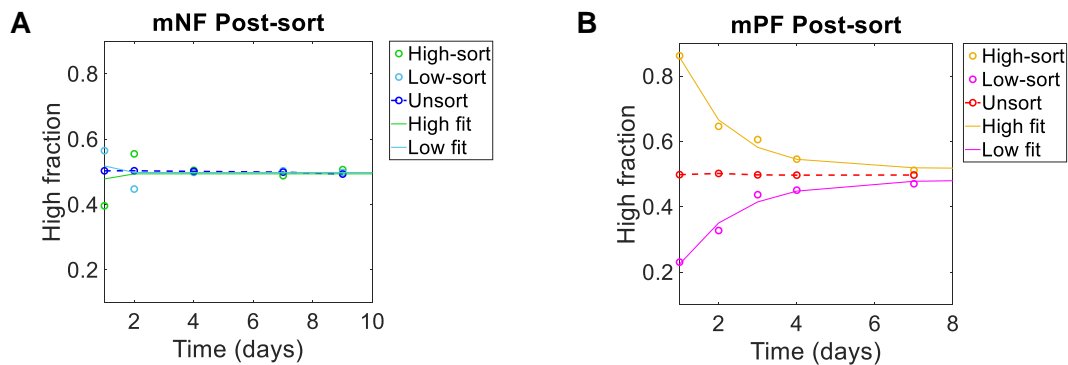


**Figure 14: Relaxation of high and low expression subpopulations after flow sorting for memory estimates.**

**(A)** The fraction of high-expressing cells for mNF-PuroR after flow sorting for high- and low-expressing subpopulations. **(B)** High fraction for mPF-PuroR after flow-sorting at the decoupled noise point. The exponential fits to the data are included in each figure. A control sample of cells was bulk-sorted without a gate as a reference for the original distribution. The mNF high- and low-sorted data were normalized as described in the methods.

equivalent except for the noise-controlling constructs, each integrated as a single copy (**Figure 15**) into the same genomic locus of a clonal cell line. Thus, the decoupled noise regimes

| Sorted sample | Parameter | Value (day$^{-1}$) |
|---|---|---|
| High-sorted mNF | f = r | 2.776 |
| Low-sorted mNF | f = r | 2.663 |
| High-sorted mPF | f = r | 0.4142 |
| Low-sorted mPF | f = r | 0.3343 |

**Table 5: Flow-sort subpopulation fitting of switching rates.**

Fall (f) and rise (r) switching rates were fitted with the H(t) solution for the high-expressing fraction in Equation 1.
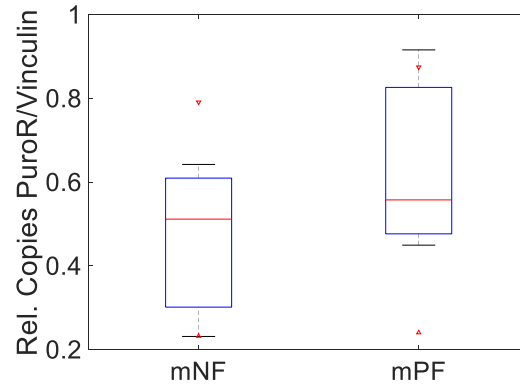


**Figure 15: Single copy integration for *PuroR* circuits confirmed by qPCR.**

Quantitative PCR results of relative copy number of the *PuroR* gene over the *Vinculin* (*Vcl*) gene, which served as a reference for known genomic copy number (*Vcl* copies = 2). The copy number was calculated using the $2^{-\Delta Ct}$ relative quantification method. Individual reactions were conducted in triplicate. The median is indicated by a red horizontal line inside each box plot. The top and bottom edge of the box plots represent the 75th and 25th percentiles, respectively. The whiskers extend to the most extreme point not considered outliers. The red notch markers indicate the comparison interval end points at the 5% significance level in the case of non-overlapping intervals.

provide DNPs to jointly control *PuroR* gene expression noise amplitude and frequency and test their role in drug resistance evolution.

**2.3.4 The noisy network helps or hinders drug resistance evolution in a stress-dependent manner.**

To uncover the role of mNF and mPF network noise in mammalian drug resistance evolution, we decoupled noise from mean *PuroR* expression in isogenic CHO cells. By following these cells in constant inducer concentrations through parallel flow cytometry and microscopy (**Figure 16A**), we identified a decoupled noise point for mNF-PuroR and mPF-
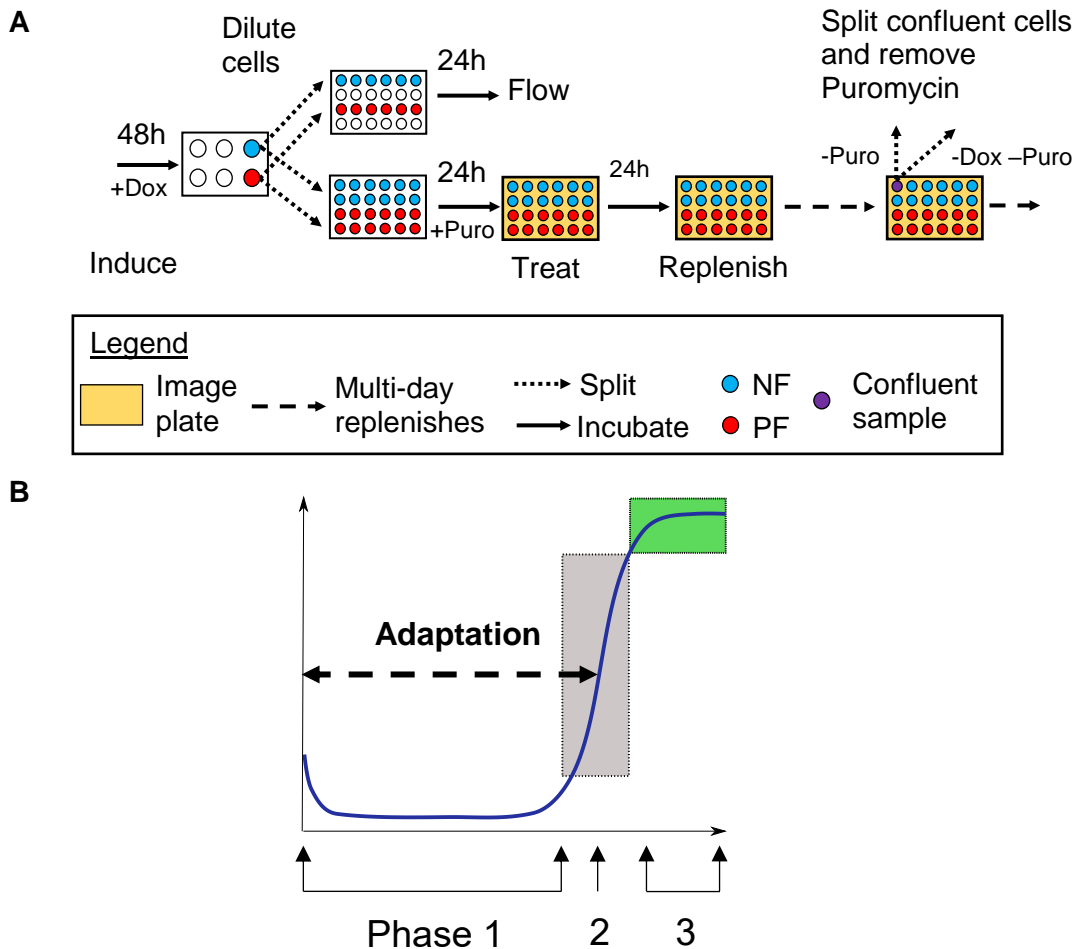


**Figure 16: Puromycin treatment experimental design.**

**(A)** Experimental workflow of the Puromycin treatment assay. Cells induced with Doxycycline (Dox) were treated with Puromycin in a parallel series of plates for imaging or for flow cytometry. Upon confluency, Puromycin was removed temporarily. **(B)** Illustration of a representative growth curve with three growth phases: 1) growth suppression, 2) regrowth (gray box), and 3) saturation (green box).

PuroR at 0.05 and 6 ng/mL Doxycycline, respectively, in two experimental sets. At the DNPs, the means differed by less than 10%. The gene expression noise amplitudes were significantly distinct prior to treatment in both sets (**Figures 13B; 17A**). Accordingly, the high-noise mPF-PuroR expression distribution consisted of a wide, visibly unimodal peak while the low-noise mNF-PuroR peak was narrow (**Figures 13C; 17B**), which is apparent in the imaging data (**Figure 13D**). Since the mNF-PuroR mean exceeded slightly the mPF-PuroR mean (non-significantly in the first experiment and significantly, but still within 10%, in the second experiment), observing better mPF-PuroR survival should then strengthen the evidence for noise-aided drug tolerance and subsequent resistance.
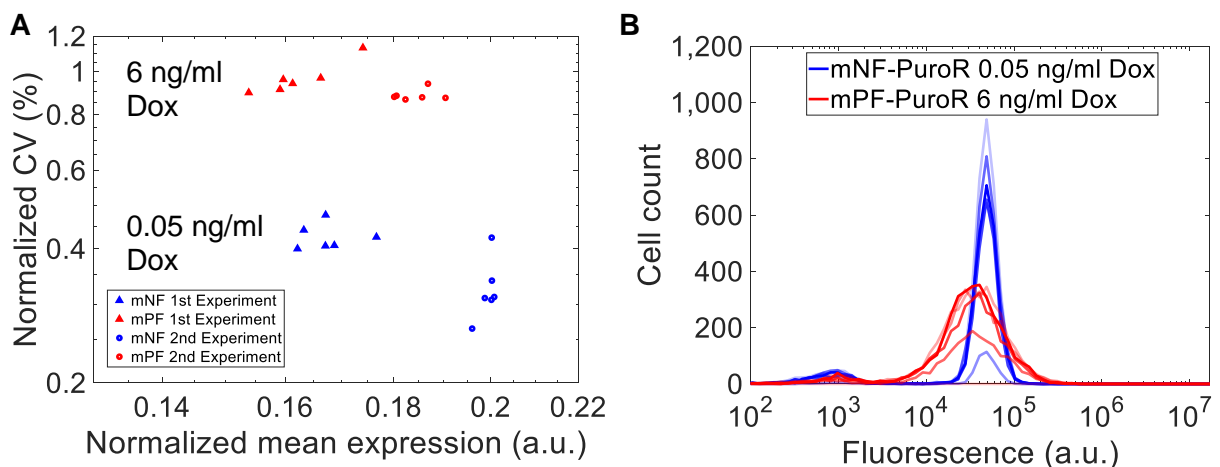


**Figure 17: Replicate expression data of decoupled noise points prior to treatment.**
**(A)** Replicates at the decoupled noise points. **(B)** Raw expression distributions from the second experimental set. There were six replicates per sample.

After preparing six mNF-PuroR and mPF-PuroR replicates at the DNP, we introduced various concentrations of Puromycin and performed sets of evolution experiments, each lasting until the adapting replicates have reached confluency. In the first experiment set, we maintained Puromycin concentrations of 0, 10, and 22.5 μg/mL while in the second set, we kept cells in Puromycin concentrations of 35 and 50 μg/mL. A total of five mPF-PuroR replicates (3 under 35 μg/mL Puromycin and 2 under 50 μg/mL Puromycin) were lost during sample maintenance.

To study the adaptation of CHO cells with low- and high-noise networks to Puromycin treatment, we constructed population-level adaptation curves at 0, 10, 22.5, 35, and 50 µg/mL Puromycin by detecting and counting single cells from daily microscope images. After examining these adaptation curves, we observed immediate growth without adaptation for low Puromycin doses, while for moderate to high Puromycin doses, the curves were initially flat, and fast growth resumed with a delay that increased with stress severity. We defined as "adaptation" only this latter behavior, which revealed three different phases of the adaptation process (**Figure 16B**): (1) growth suppression while the curve stayed mostly flat; (2) fast regrowth when the curve rose (grey region); and (3) saturation when the curve flattened again at confluency (green region). Based on these phases, we analyzed the adaptation time, which we defined as the time required for initially suppressed cells to reach half-saturation (indicated by a dashed arrow in **Figure 16B**). Interestingly, the length of the growth suppression phase became more variable between replicates of each circuit upon increasing Puromycin concentrations (**Figure 18A-E**).

We first investigated how decoupled *PuroR* expression noise and the corresponding networks affect the adaptation time at various levels of Puromycin. We calculated adaptation times only for replicates whose moving average growth rates (estimated for every 3 timepoints on the adaptation curves) fell below 0 at least once (**Figure 19**). Based on this definition, CHO cells with the mNF-PuroR and mPF-PuroR circuits grew without adaptation at 0 and 10 µg/mL Puromycin (**Figures 18A,B,F; 19A,B**). Under 22.5 µg/mL Puromycin, the high-noise mPF-PuroR cells adapted after a delay while the low-noise mNF-PuroR cells grew without adapting (**Figures 18C,F; 20A**). Likewise, the six low-noise mNF-PuroR replicates adapted faster with a shorter delay than the mPF-PuroR replicates (2 out of 3 surviving) under 35 µg/mL Puromycin (**Figures 18D,F; 20B**). Interestingly, cells with the high-noise network tended to exhibit morphological diversity, including signs of polyploidy (**Figures 21B-22B**). In contrast to the lower drug concentrations, at 50 µg/mL Puromycin all mNF-PuroR replicates perished whereas half (two) of the mPF-PuroR replicates eventually adapted and recovered (**Figures 18E-F;**
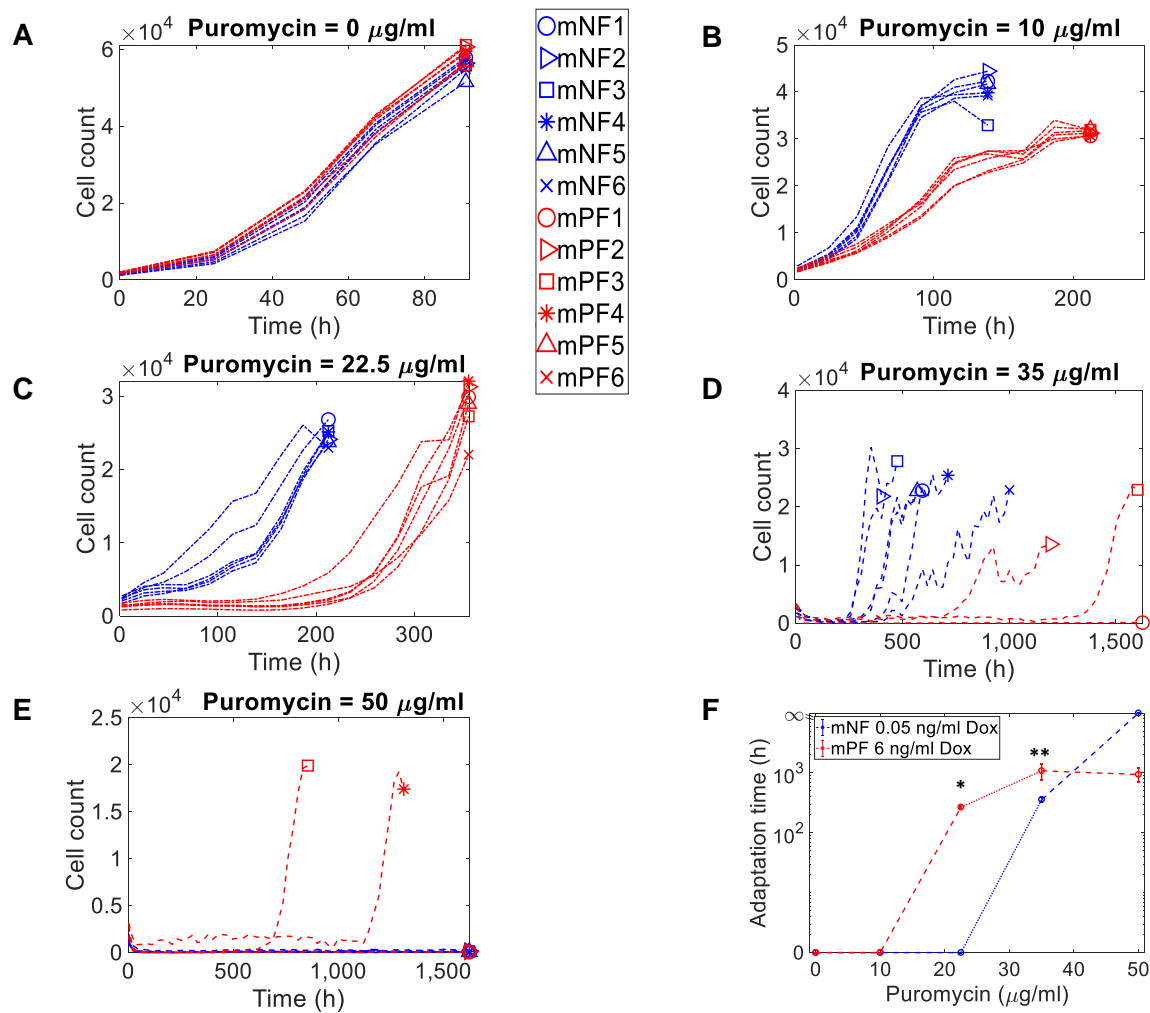
**Figure 18: High network noise hinders Puromycin resistance under low stress and facilitates adaptation under high stress.**

**(A-E)** Growth curves for cells initially tuned to the decoupled noise points (DNPs) under 0 **(A)**, 10 **(B)**, 22.5 **(C)**, 35 **(D),** and 50 **(E)** µg/mL Puromycin. Dash-dot growth curves indicate data from the first experimental set while dash-dash growth curves are from the second experiment set. **(F)** Mean adaptation times corresponding to **(A-E)** (*p-value = 0.0022, n = 6; **p-value = 0.0238, mPF-PuroR n = 3 and mNF-PuroR n = 6). The two-tailed Mann-Whitney U test inferred significant differences at p-values < 0.05, which included the non-growing mPF replicate 1 from 35 µg/mL Puromycin (infinite adaptation time), which is not shown in the mean in **(F)**. The error bars represent the standard error of the mean.

**20C; 22**). Importantly, visual inspection indicated that mNF-PuroR cells completely vanished
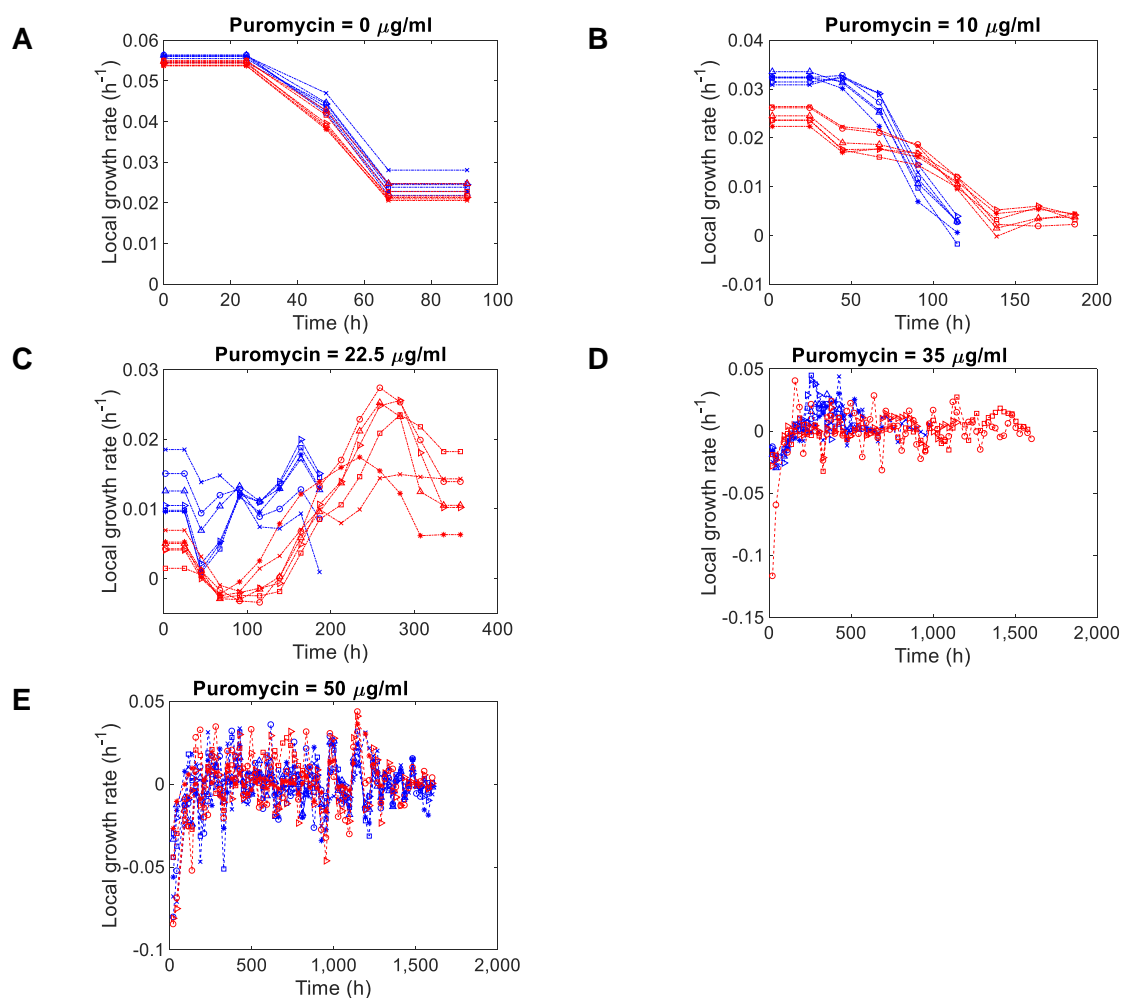
**Figure 19: Initial cell death criteria for identifying adaptation during treatment.**

**(a-e)** Local growth rates estimated from a moving slope of 3 timepoints for 0, 10, 22.5, 35, and 50 µg/mL Puromycin. If the local growth rates dipped below 0, adaptation can occur.
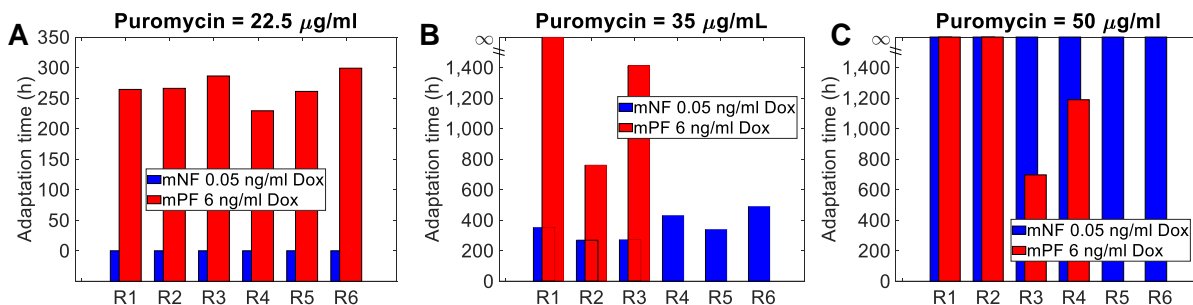


**Figure 20: Replicate adaptation times after initial treatment.**

Adaptation times for replicates with initial cell death after treatment under **(A)** 22.5, **(B)** 35, and **(C)** 50 µg/mL Puromycin. Cells quickly re-growing without initial death have adaptation times = 0 while treated replicates with no surviving cells have adaptation times = infinity.

59

**A** — Initial treatment, Growth suppression, Half-saturation, Growth saturation

**B**
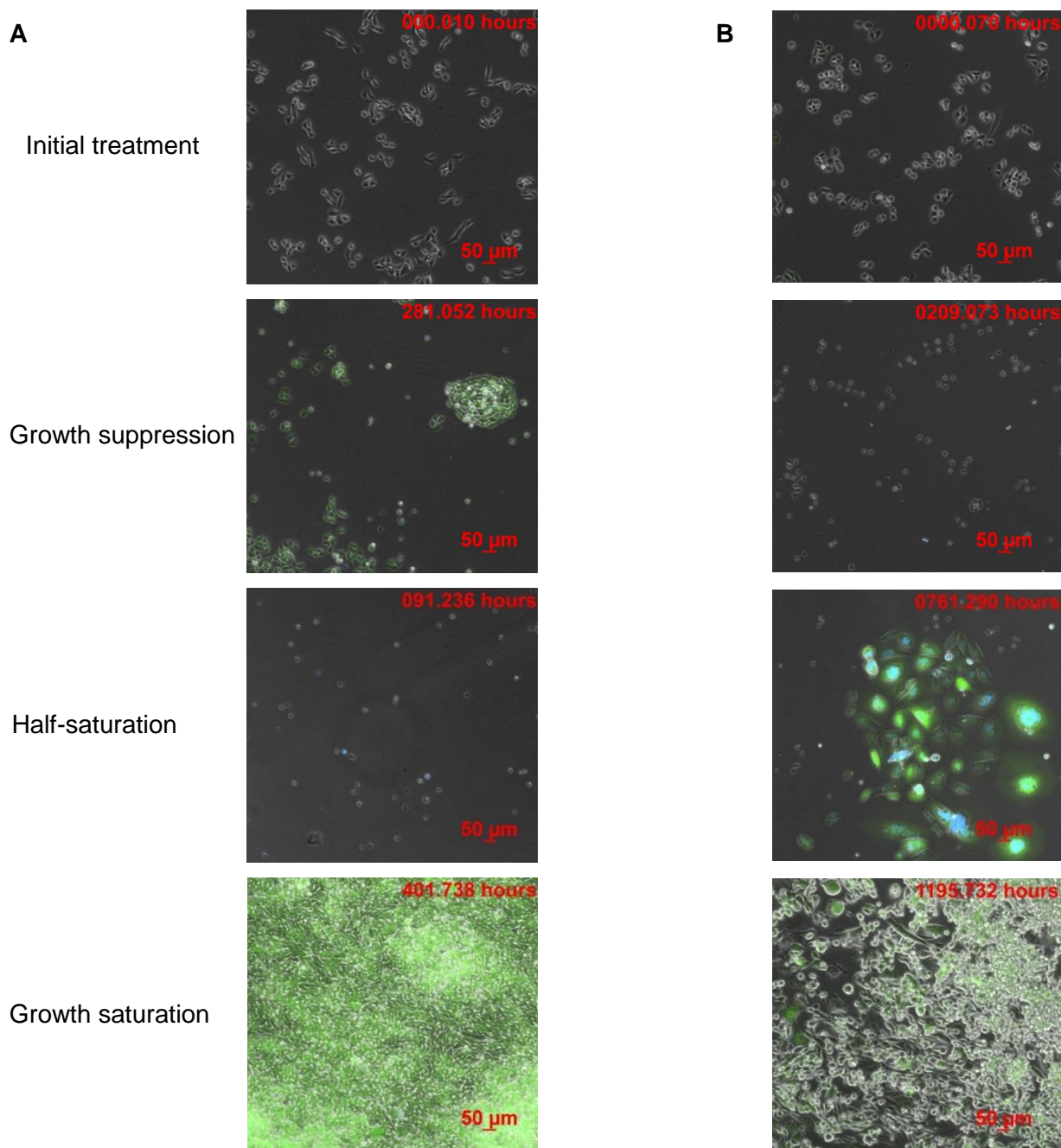
**Figure 21: Images of mNF-PuroR and mPF-PuroR replicate 2 growth curve phases at 35 µg/mL Puromycin.**

Images of (**A**) mNF-PuroR and (**B**) mPF-PuroR replicate 2 at initial treatment (0h) with 35 µg/mL Puromycin, then growth suppression, half-saturation (notice the large polynucleated cells for mPF-PuroR), and finally growth saturation phases. Time stamps are listed in the upper-right corner of each image, which come from **Figure 18D**. The images were cropped with consistent look-up table thresholds.

**A** — Initial treatment, Growth suppression, Half-saturation, Growth saturation

**B** — Initial treatment, Growth suppression, Half-saturation, Growth saturation
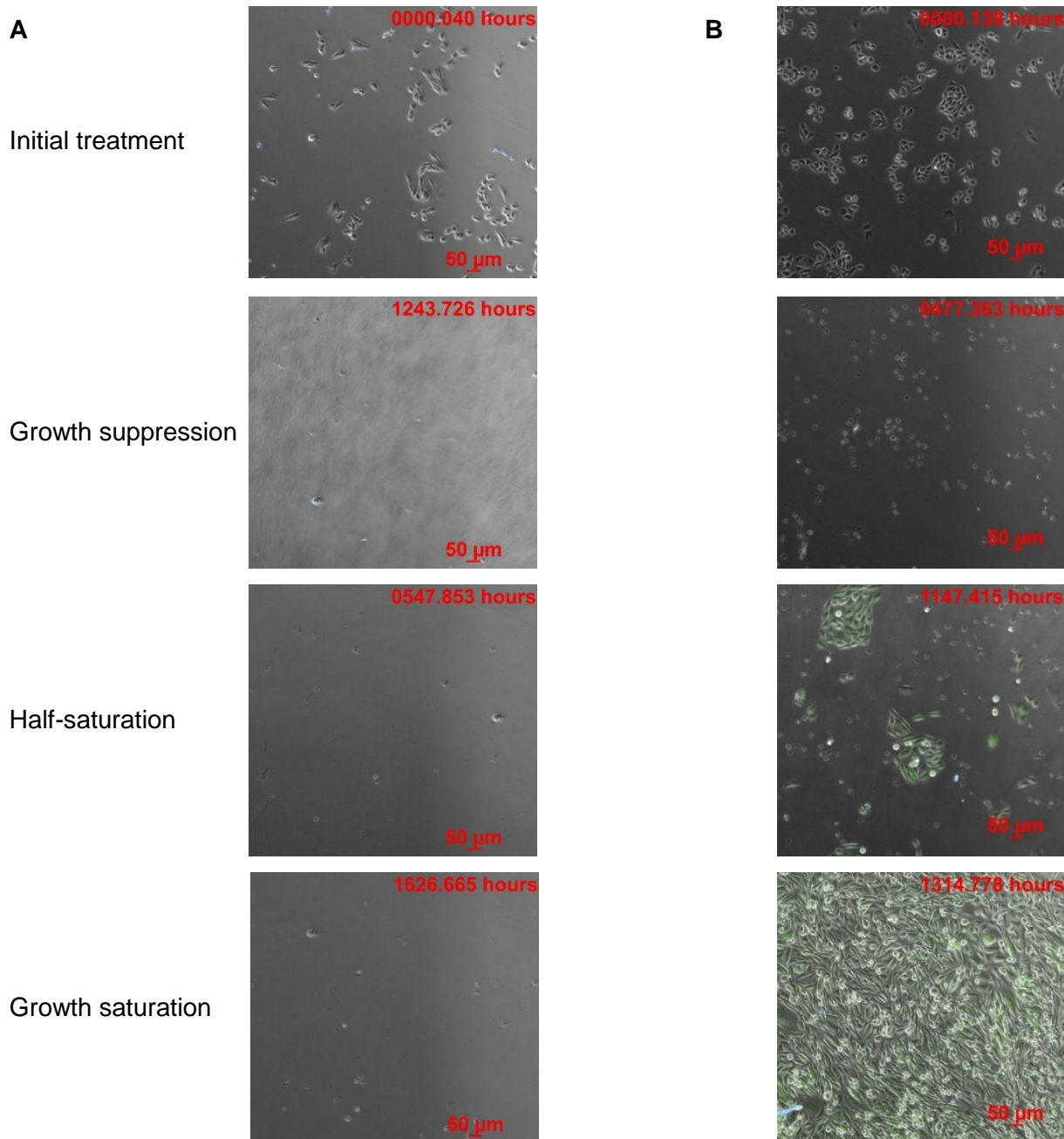
**Figure 22: Images of mNF-PuroR replicate 2 and mPF-PuroR replicate 4 growth curve phases at 50 µg/mL Puromycin.**

Images of **(A)** mNF-PuroR replicate 2 and **(B)** mPF-PuroR replicate 4 at initial treatment of 50 µg/mL Puromycin, growth suppression, half-saturation, and growth saturation. The times are listed in the upper-right corner, which come from **Figure 18E**. The images were cropped with consistent look-up table thresholds.

from the culture wells (**Figure 22A**) despite having slightly higher pre-treatment mean *PuroR*

expression, indicating their inability to adapt to the highest Puromycin concentration.

Overall, the adaptation times indicate that the noisy mPF network promotes evolutionary adaptation compared to mNF at high stress, while the reverse is true for low stress, which is consistent with the effects of noise on survival immediately after treatment for steep kill curves. Therefore, noisy mPF networks affect long-term mammalian drug resistance evolution similarly to noise-dependent short-term survival of other cell types[43]. The most pronounced evolutionary benefit from the noisy mPF network occurs at the highest stress level. Importantly, the temporal fluctuations in PuroR expression (memory) from the mPF network (~2 days) did not temporally overlap with the length of time before the emergence of drug resistance (weeks).

## 2.3.5 Computational model offers mechanistic insights into adaptation to drug treatment.

Used with permission from Daniel A. Charlebois, who contributed to the modeling in figures (**Figures 23 and 24**) and text in this section.

To predict the experimentally observed long-term evolutionary adaptation behaviors at high stress, we developed a stochastic population/evolutionary dynamics model[199]. This model assumed additional cellular states based on short-term experimental data before and immediately after drug treatment. Specifically, we assumed that cells die if their *PuroR* concentration is below a specific Puromycin-dependent threshold, which we estimated from the initial fraction of cells surviving Puromycin treatment [Equations (5)-(7), methods section 2.2.18]. We partitioned the remaining surviving cells into stress-induced "persister" cells that neither grow nor die and preexisting, nongenetically drug resistant cells that grow in the presence of Puromycin [Equation (8), methods section 2.2.18] (**Figure 23**). Thus, upon initial drug treatment, three different cell types exist within the cell population: dead (*D*), persister (*P*), and nongenetically drug-resistant (*N*) cells. We allowed phenotype switching between *P* and *N* cells. Additionally, we assumed that over time *P* cells and *N* cells can give rise to a fourth, stable (genetically or epigenetically) drug-resistant (*G*) cell type. Though growth rates of *N* and *G* cells were similar, the death rates of *N* cells were increasingly greater than *G* cells for higher

62
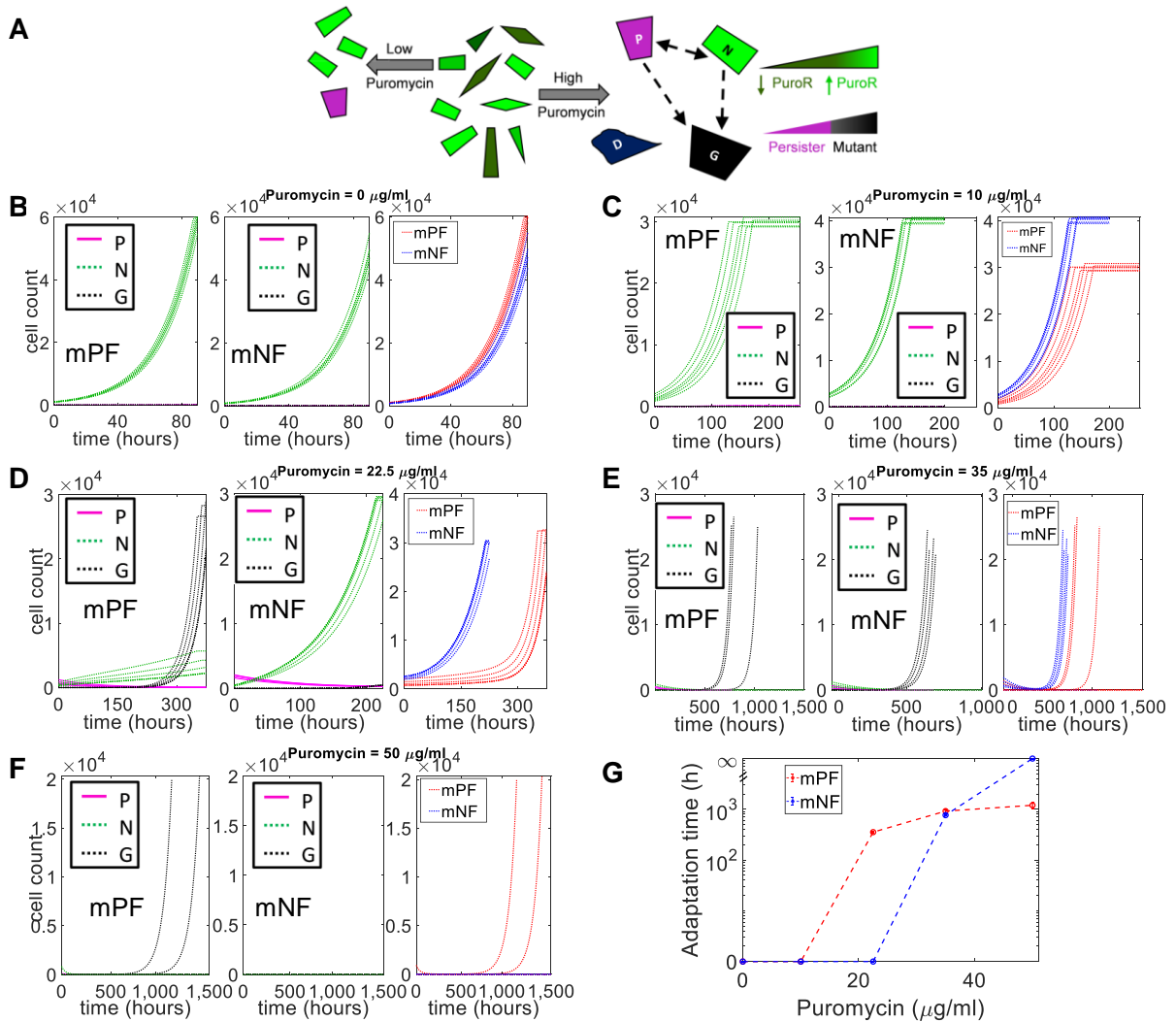
**Figure 23: Modeling the adaptation of mPF-PuroR and mNF-PuroR cells in various concentrations of Puromycin.**

Used with permission from Daniel A. Charlebois, who made the figures and wrote the legend text. **(A)** Schematic depicting the effects of Puromycin on CHO cell population composition and survival. Nongenetically Puromycin-resistant cells (green) and nongrowing persister cells (magenta) can switch phenotypes (bidirectional arrow), both of which can also become stably resistant cells (black). Upon increasing stress, cells die (dark blue). **(B-F)** Representative growth curves for simulated mPF and mNF CHO cell populations under **(B)** 0, **(C)** 10, **(D)** 22.5, **(E)** 35, and **(F)** 50 µg/mL Puromycin. Growth curves shown in panels correspond to: (left) mPF and (center) mNF subpopulations, and (right) mPF and mNF populations. **(G)** Adaptation times corresponding to the populations shown in panels **(B-F)**.

Puromycin doses (**Table 4**). Gene expression noise and the concentration of the drug imposing the selective pressure determine the $D$, $P$, and $N$ cell type percentages in the population shortly after treatment, and consequently, the ultimate evolutionary outcome in these simulations (**Figure 23B-F**). As Puromycin concentrations increase, the number of $D$ cells increases accordingly at the expense of surviving cells. Among surviving cells, the frequency of $P$ cells increases with stress levels until all surviving cells are $P$ cells at very high stress (**Figure 24**). $G$ cells emerge subsequently at rates reflecting the numbers of their $P$ and $N$ cell precursors. Only models that allowed $P$ to $G$ conversion captured the experimental adaptation time dynamics in all tested Puromycin conditions (**Figure 23G**). Therefore, the experiments and models jointly support that resistance to high Puromycin levels occurs by drug-induced
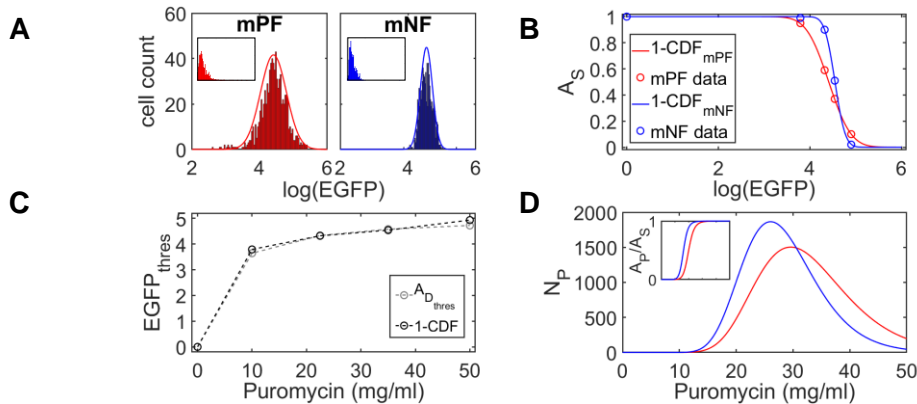


**Figure 24: Determination of initial dead, persister, and nongenetically drug resistant CHO subpopulation fractions.**

Used with permission from Daniel A. Charlebois, who made the figures and wrote the legend text. **(A)** Fits to log-transformed mPF and mNF expression data for Doxycycline = 6 and 0.05 ng/mL, respectively, using Equation (5). Insets show experimental non-transformed lognormal distributions. **(B)** Fraction of cells surviving initial Puromycin treatment ($A_S$) calculated from 1 - Equation (6) (solid lines) along with mPF and mNF data (open circles). **(C)** *EGFP* threshold below which cells are killed by Puromycin using Equation (7) (grey data points and dashed line) and the cumulative distribution function (CDF) (black data points and dashed line) from **(B)**. **(D)** The number of persister cells ($N_P$) as a function of Puromycin levels. Inset shows surviving cell fraction that are persisters.

formation of persister-like cells serving as reservoirs for fast-growing, heritably drug-resistant mutants.

## 2.3.6 Temporary removal of Puromycin and sequencing suggest mNF-PuroR adaptation by intra-circuit mutations.

As the computational model indicated, after initial cell death, adaptation to Puromycin stress could occur by multiple different mechanisms depending on the stress level. Specifically, at low stress, nongenetically resistant cells could continue growing, and eventually reestablish the population without any mutations or other heritable alterations (**Figure 25**)[26]. Alternatively,
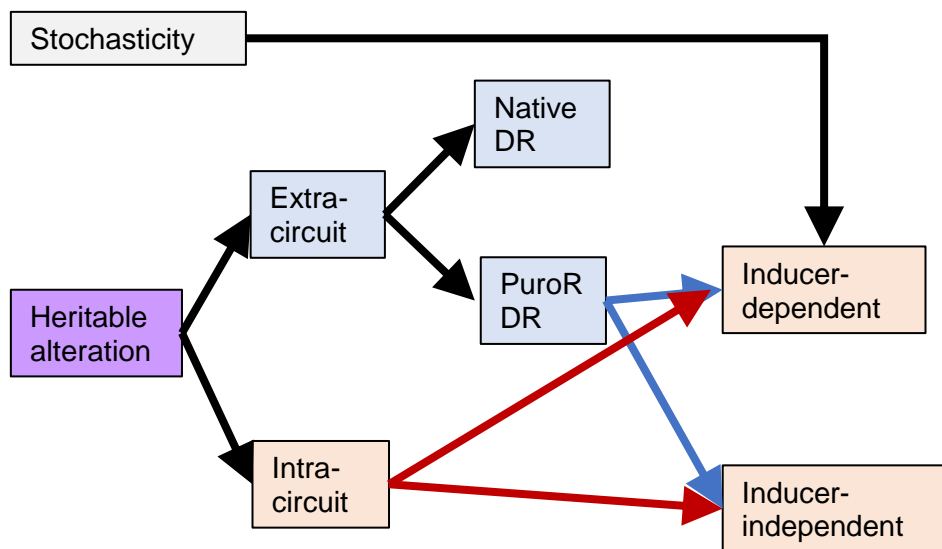


**Figure 25: Evolutionary scenarios for adaptation to Puromycin treatment.**

Two categories of mechanism can drive drug resistance evolution in the mNF and mPF cell populations: heritable alteration and stochastic gene expression. Heritable alterations can occur either inside or outside the gene circuit. Extra-circuit alterations could elevate *PuroR* expression (i.e., be *PuroR*-dependent) or could elicit *PuroR*-independent, native resistance mechanisms. *PuroR*-dependent adaptation could further be classified as inducer-dependent or independent. Gene expression stochasticity as a resistance mechanism should be inducer-dependent, based on the noise amplitude and noise memory at the given inducer level. Inducer-dependence should be indicated by expression reversion to pre-treatment levels in uninduced populations after drug removal.

at high stress, cells can acquire heritable (genetic or epigenetic) drug resistance alteration(s) after a significant delay, leading to stable resistance. Heritable mechanisms could be endogenous (based on native mechanisms independent of *PuroR* gene expression), or *PuroR*-dependent, elevating *PuroR* expression to a level enough for resistance. However, for all mNF and mPF replicates evolved at the highest 3 stress conditions, induced *PuroR* expression increased and stayed far above pre-treatment levels. Therefore, we concluded that adaptation always relied on elevated *PuroR* expression.

 *PuroR*-dependent mechanisms could occur inside or outside the synthetic gene circuit and may depend on network induction. To distinguish between such possibilities (**Figure 25**), and to formulate hypotheses about the nature of molecular events contributing to evolutionary adaptation, we removed Puromycin temporarily and then re-added it again for cells that have adapted under 22.5, 35, and 50 µg/mL Puromycin. Moreover, to test whether gene circuit induction was necessary for resistance, we split each evolved replicate into two separate wells, culturing them either without Doxycycline ("uninduced") or with Doxycycline ("induced") at the same concentration as before Puromycin removal (**Figure 26A**).

 Next, we studied the behavior of mNF-PuroR replicates after removal of 35 µg/mL Puromycin. All uninduced and induced mNF-PuroR replicates maintained constant *PuroR* expression levels well above corresponding induced but untreated ancestral cells for ~ a month (**Figures 26B; 27-28**), suggesting that inducer-independent, high *PuroR* expression has evolved. Accordingly, all uninduced and induced mNF-PuroR replicates grew without adaptation upon Puromycin re-addition, much quicker than their Puromycin-treated ancestors, further supporting stable, induction-independent drug resistance in each population (**Figure 26E**).

 To examine how inducer-independent, *PuroR*-dependent resistance arose in the mNF-PuroR circuit, we sequenced the gene circuit from the six induced replicates after drug removal at 35 µg/mL Puromycin. In replicate 2, we found an indel in *hTetR* that can reduce binding affinity to *tetO2* sites by 1000-fold[200] (**Figure 29**). Therefore, this mutation likely compromised
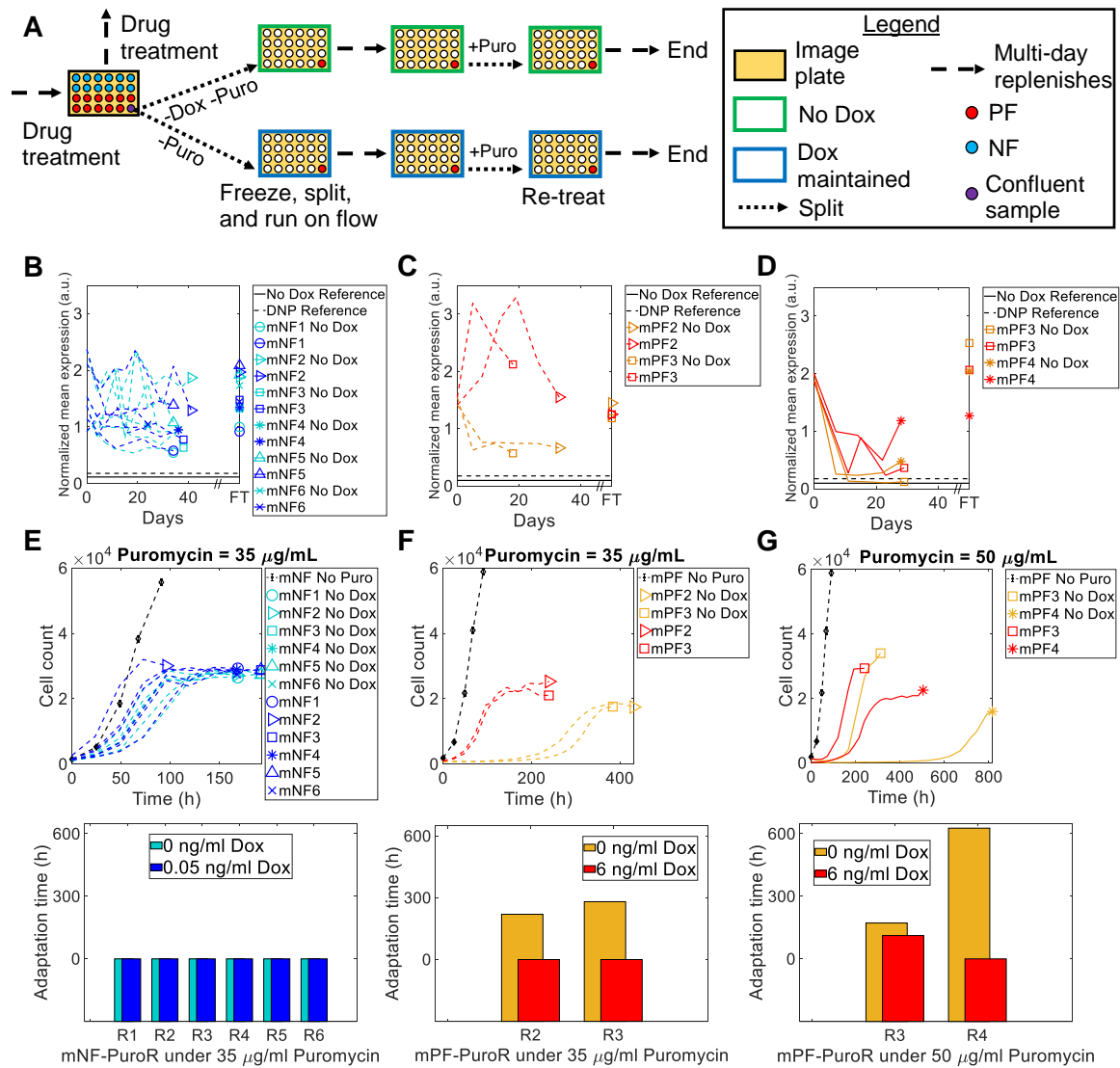
**Figure 26: Gene expression and cell counts during drug removal and re-addition suggest stable *PuroR*-dependent mechanisms of resistance at high drug levels.**
**(A)** Schematic for the drug removal and retreatment experiment. Doxycycline (Dox) was removed or maintained simultaneously with drug removal. **(B)** Mean expression for mNF during temporary removal of 35 µg/mL Puromycin and after final retreatment (FT). **(C-D)** Mean expression for mPF after removal of **(C)** 35 and **(d)** 50 µg/mL Puromycin and after final retreatment (FT). **(E-G)** Growth curves (top) and adaptation times (bottom) during re-treatment for **(E)** mNF cells under 35 µg/mL, and mPF cells under **(F)** 35 and **(G)** 50 µg/mL Puromycin. The black growth curves are averaged untreated cell counts from **Figure 18A**.

repressor functionality, leading to high *PuroR* expression and drug resistance. In replicate 3,
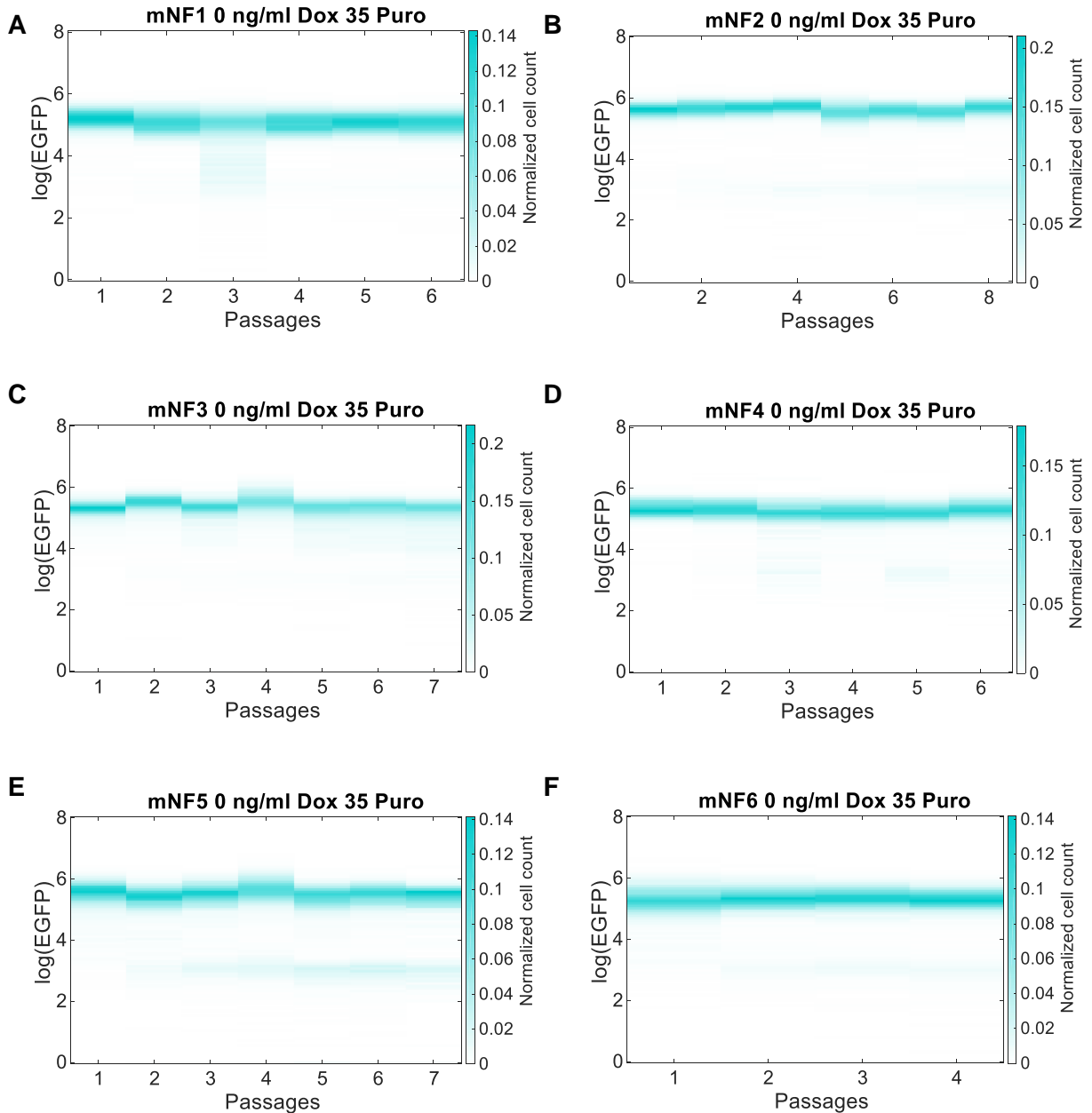
**Figure 27: Uninduced mNF expression distributions after removing 35 µg/mL Puromycin.**

**(A-F)** Log-transformed expression distributions for uninduced mNF-PuroR replicates 1-6 after removal of 35 µg/mL Puromycin. Each color bar represents normalized cell counts.

we found a single nucleotide polymorphism in the distal region of the promoter (**Figure 30**). Furthermore, the CRISP-ID[195] genotyping algorithm uncovered in replicate 1 two mutant variants in the same distal promoter region as in replicate 3 (**Figure 31**). Both variants contain the same mutation as replicate 3 (**Figures 30-31**), suggesting that both arose by selection for
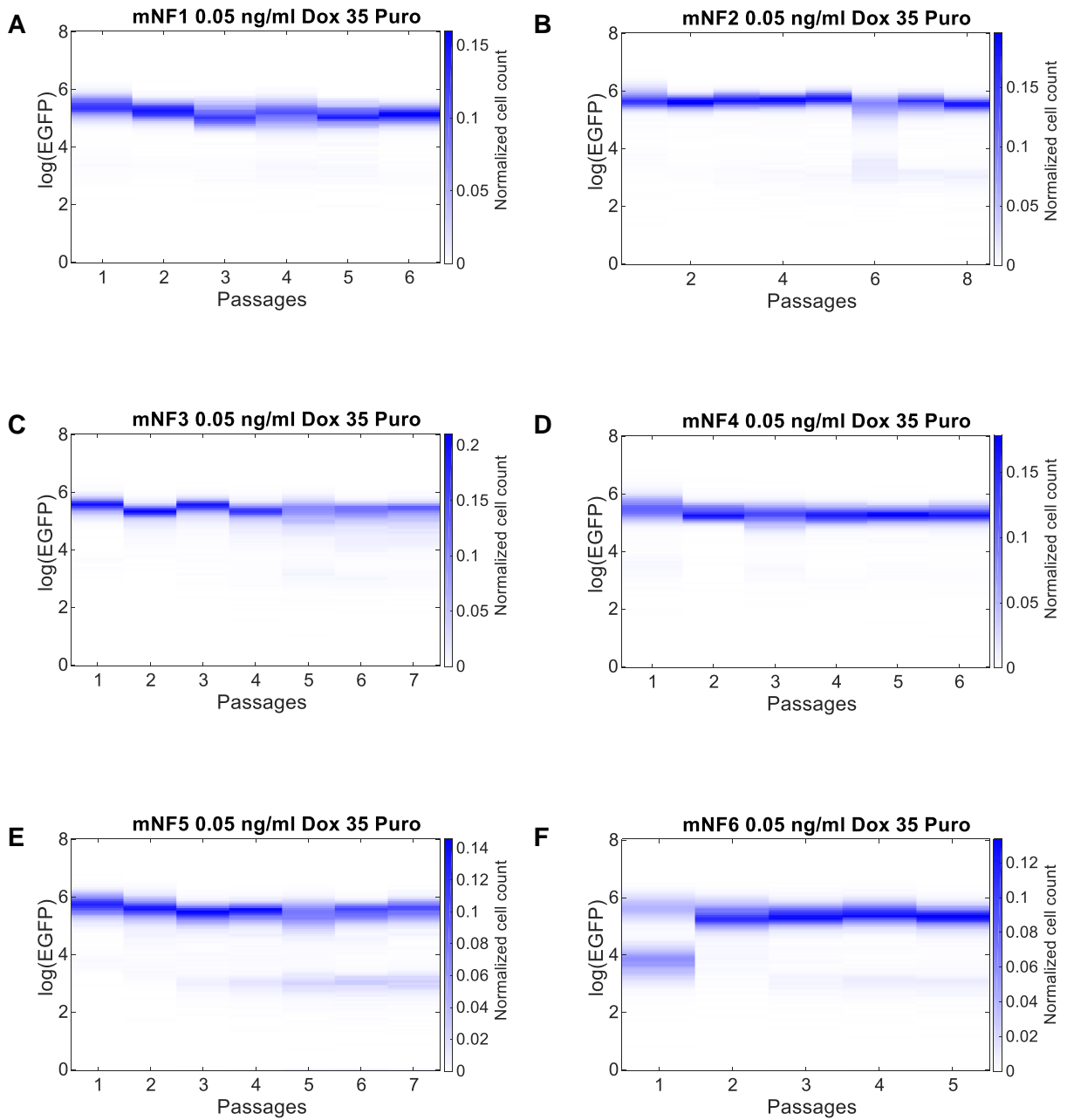
**Figure 28: Induced mNF expression distributions after removing 35 µg/mL Puromycin.**

**(A-F)** Log-transformed expression distributions for induced mNF-PuroR replicates 1-6 after removal of 35 µg/mL Puromycin. Each color bar represents normalized cell counts.

elevated *PuroR* expression. Therefore, mutations abrogating *hTetR* repression seem to occur repeatedly. This may occur because random mutations are typically deleterious rather than beneficial for biological function[181]. Despite identical phenotypes (stable inducer-independent
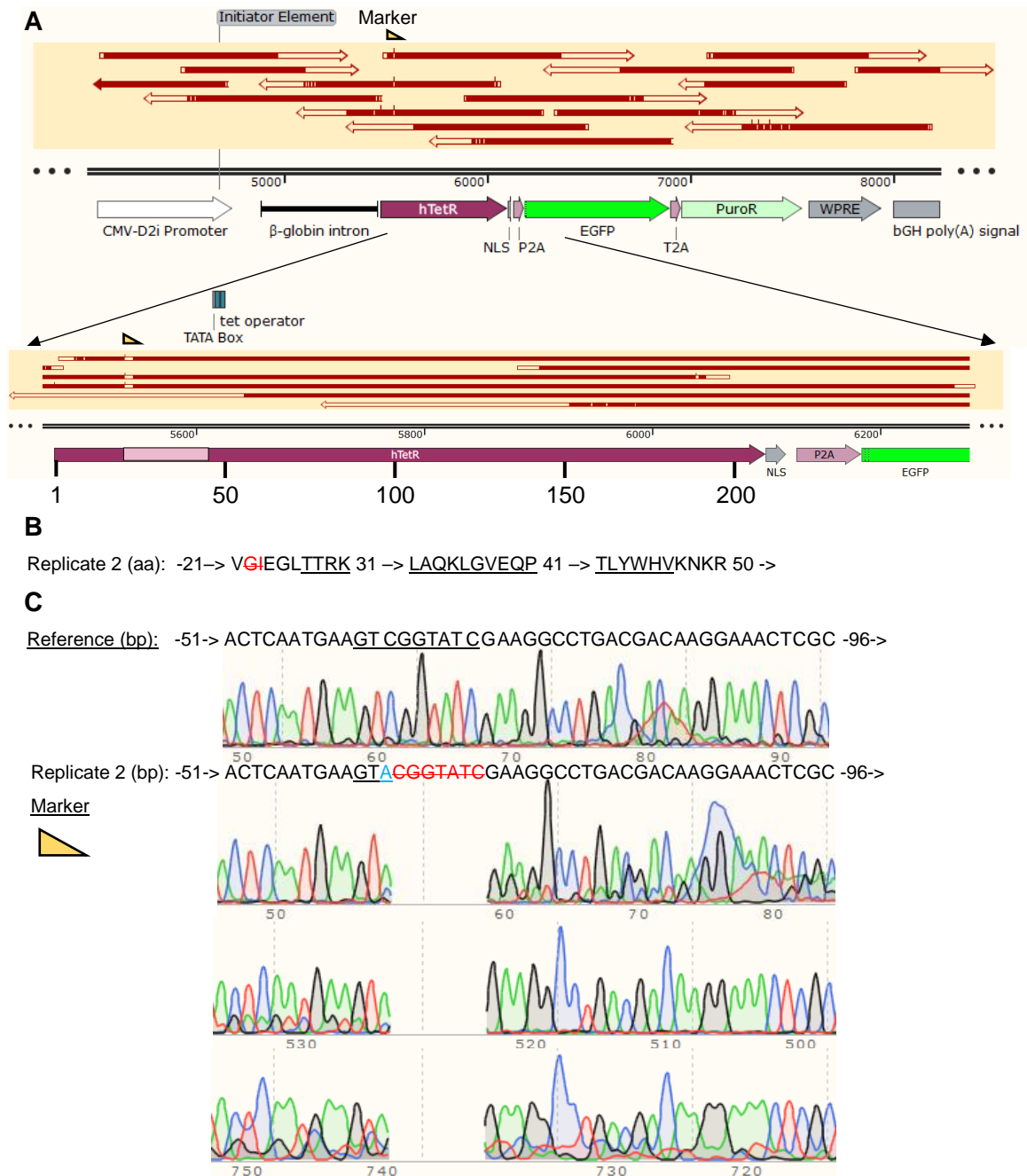
**Figure 29: Sequencing mNF-PuroR replicate 2 after 35 µg/mL Puromycin treatment revealed a DNA-binding-abolishing mutation in *hTetR*.**

**(A)** Sequencing map of the mNF-PuroR circuit from replicate 2 after 35 µg/mL Puromycin treatment with a zoom-in of *hTetR*. **(B)** Glycine 22, adjacent to the DNA-binding domain (underlined), was deleted along with Ile23. **(C)** Nucleotide sequences of affected amino acids (underlined) with chromatographs. Blue bases indicate insertions while crossed red nucleotides indicate deletions.
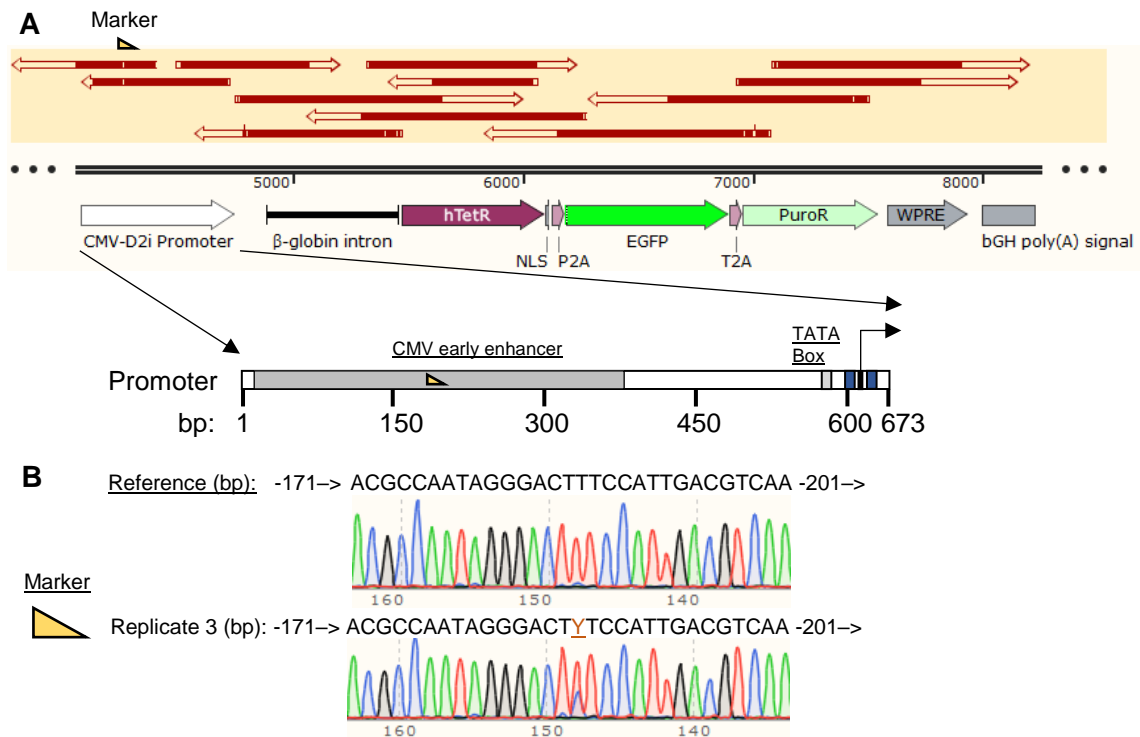
**Figure 30: Mutation in the early CMV enhancer suggests genetic heterogeneity in the mNF-PuroR circuit from replicate 3 after 35 µg/mL Puromycin treatment.**

**(A)** Sequencing map of the mNF-PuroR circuit from replicate 3. Zoom-in of the promoter has multiple elements including the CMV early enhancer (dark gray), TATA box (light gray), *tetO2* operator sites (blue), and Initiator element (black). A marker indicates the location of a mutation. **(B)** Nucleotide sequence for the reference and the replicate 3 promoter at the vicinity of the mutation in the CMV early enhancer at T186->Y (C/T mixture; underlined).

expression), we found no mutations in the mNF-PuroR circuit from replicates 4, 5, and 6 (**Figure 32**). In summary, the mNF gene circuit adapts through intra-circuit mutations or extra-circuit heritable alterations that corrupt *hTetR* repressor function to confer elevated, inducer-independent *PuroR* expression (**Figure 33**).

Finally, we followed the same procedure to gain insights for evolution in 22.5 µg/mL Puromycin, the lowest stress level, where mNF-PuroR cells grew instantaneously. At this stress level, after drug removal both uninduced and induced mNF-PuroR mean expression levels reverted quickly towards their pre-treatment levels (**Figure 34A**). This indicated non-genetic

71
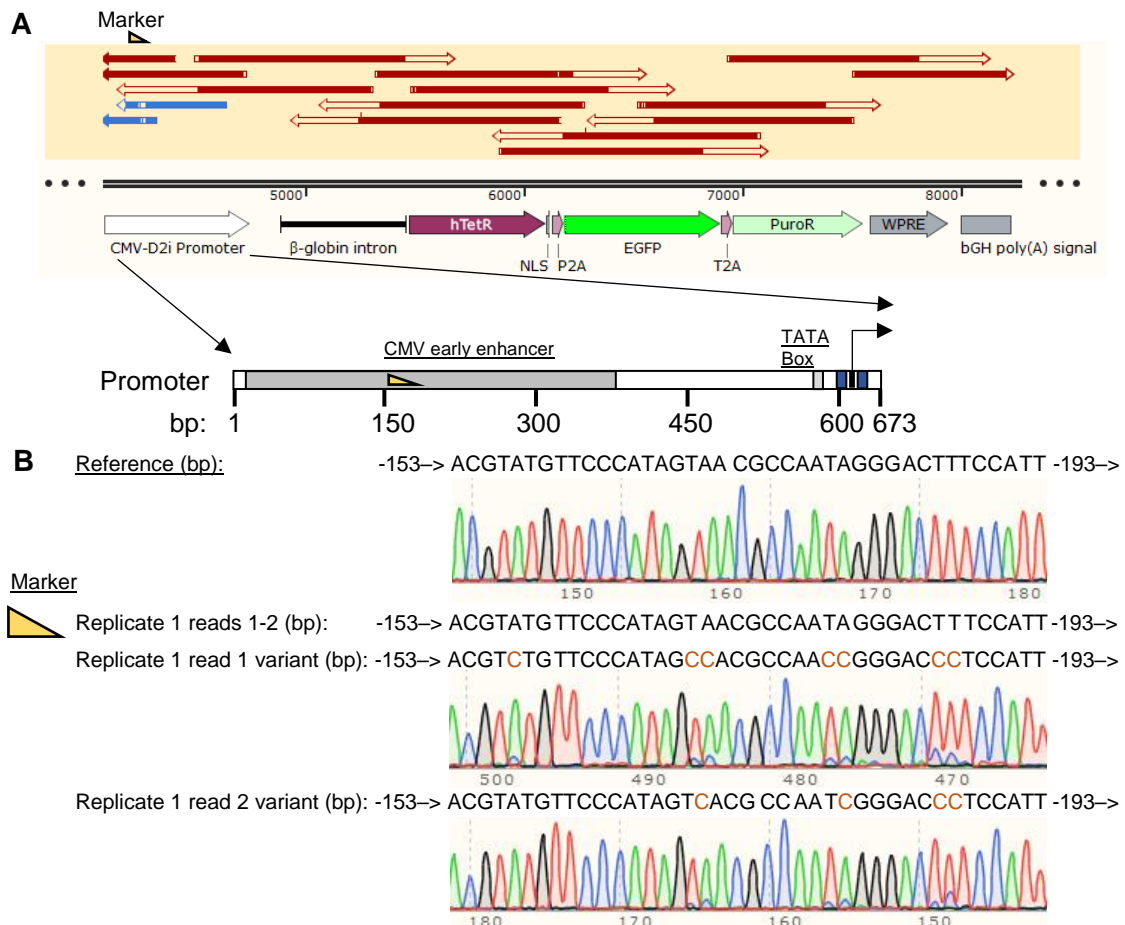
**Figure 31: Mutations in the early CMV enhancer suggest genetic heterogeneity in the mNF-PuroR circuit from replicate 1 after 35 μg/mL Puromycin treatment.**

**(A)** Sequencing map of the mNF-PuroR circuit from replicate 1 with a zoom-in of the CMV-D2ir promoter, which has two reads showing genetic heterogeneity (predicted variants shown as blue reads) from mixed sequencing traces. The marker indicates the location of the mutations. **(B)** Nucleotide sequence for the ancestral promoter, replicate 1 consensus read, and two variants at the vicinity of the mutations (orange) detected by CRISP-ID.

drug resistance purely from non-uniform *PuroR* expression. Sequencing revealed no intra-circuit mutations, further supporting these conclusions (**Figure 35A**). Overall, the lack of intra-circuit mutations and the quick reversion to pre-treatment expression levels suggest nongenetic drug resistance mechanisms relying on preexisting Doxycycline-dependent *PuroR* expression variability, as predicted computationally at sufficiently low stress levels (**Figure 23D**).
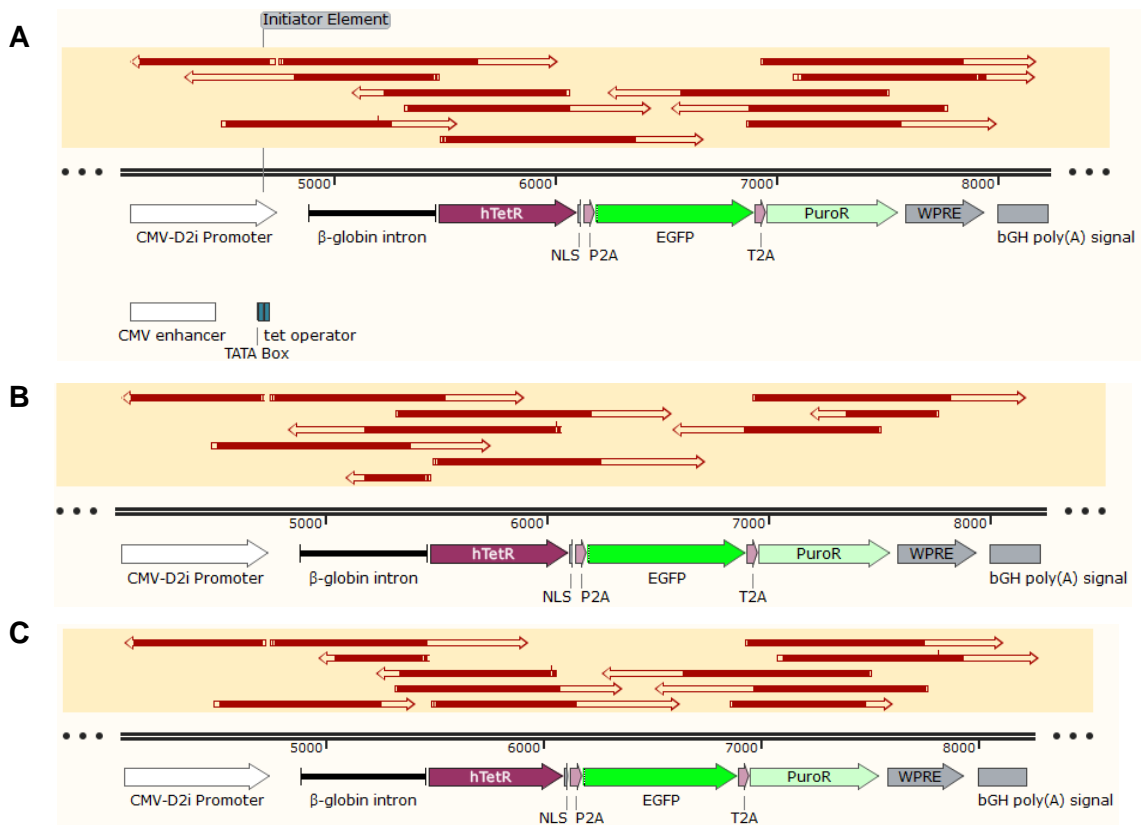
**Figure 32: Sequencing induced mNF-PuroR replicates 4 through 6 after removal of 35 µg/mL Puromycin treatment revealed no fixed mutations.**

Sequencing maps of the mNF-PuroR circuit from **(A)** replicate 4, **(B)** 5, and **(C)** 6 previously under 35 µg/mL Puromycin treatment.
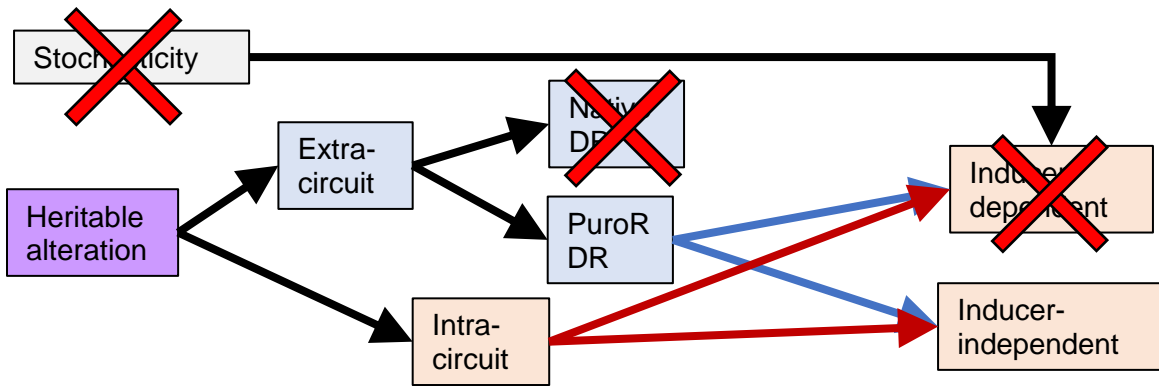
**Figure 33: mNF-PuroR at 35 µg/mL Puromycin adapts through inducer-independent genetic mechanisms.**

We ruled out stochasticity because the uninduced and induced cells did not return to pre-induction and pre-treatment expression levels, respectively. This implies *PuroR*-dependent resistance. Sequencing results for 3 mNF-PuroR replicates supports intra-circuit mutations leading to drug resistance. The remaining replicates do not have fixed circuit mutations, consistent with extra-circuit heritable drug resistance as an additional mechanism.
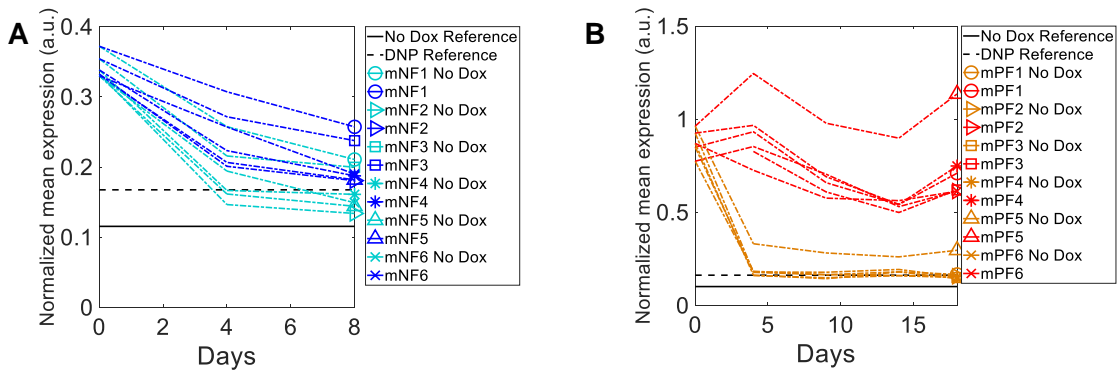


**Figure 34: Mean expression after temporary removal of 22.5 µg/mL Puromycin.**

Normalized mean expression of **(A)** mNF-PuroR and **(B)** mPF-PuroR after temporary removal of 22.5 µg/mL Puromycin. Uninduced samples that were not treated nor induced (black solid line) and pre-treatment expression at the decoupled noise point (black dash line) served as references for expression reversion.
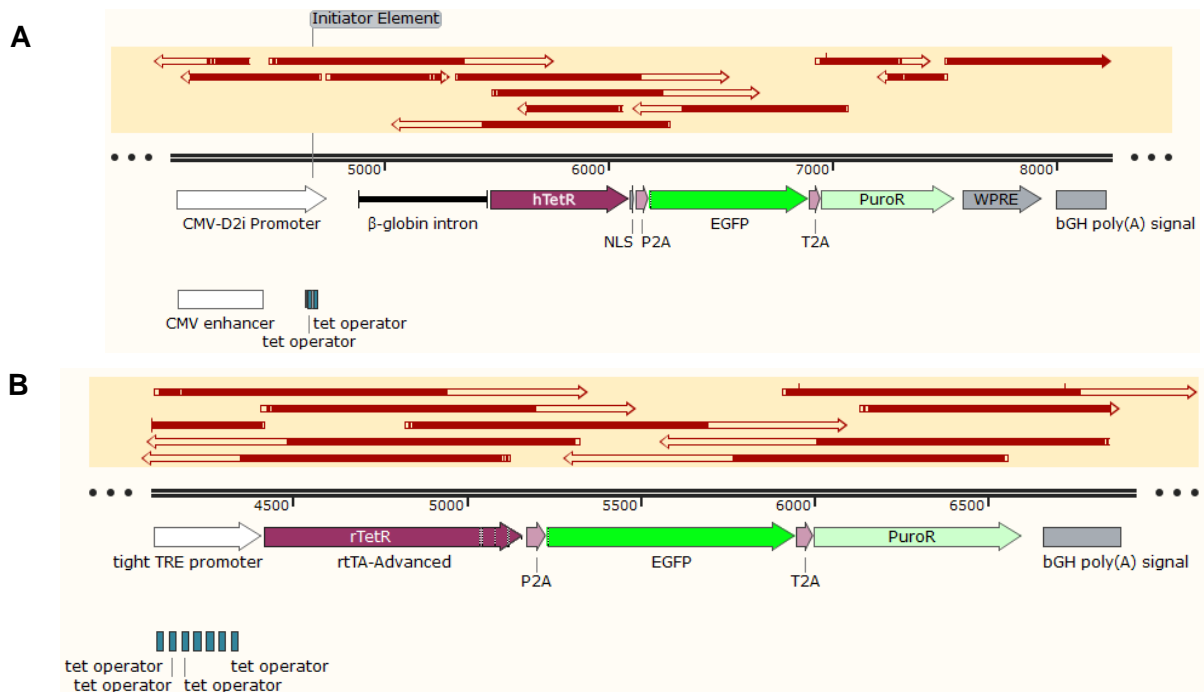
**Figure 35: Cells with the mNF-PuroR and mPF-PuroR circuit in replicate 4 after removal of 22.5 µg/mL Puromycin did not harbor intra-circuit mutations.**

**(A)** Sequencing results for the mNF-PuroR circuit from cells in replicate 4 after removal of 22.5 µg/mL Puromycin. **(B)** Sequencing reads for the mPF-PuroR circuit replicate 4 cells.

## 2.3.7 Temporary removal of Puromycin and sequencing suggest mPF-PuroR adaptation without intra-circuit mutations.

To investigate molecular adaptation mechanisms of mPF cells to 35 µg/mL Puromycin, as for the mNF gene circuit, we performed sequencing, which revealed no mutations in the high-noise mPF-PuroR gene circuit (**Figures 36-37**) for any replicate. Therefore, extra-circuit heritable alterations must exist that confer resistance by *rtTA* induction-dependent or independent mechanisms. To distinguish between these possibilities, as before, we compared "induced" versus "uninduced" cell count and gene expression time courses for mPF-PuroR replicates during drug removal and re-addition. In contrast to cells with the mNF-PuroR circuit, uninduced mPF-PuroR replicates showed signs of regulator (*rtTA*) induction-dependent adaptation, as their expression dropped closer, albeit not completely down to ancestral levels (**Figures 26C; 38A,B**), and they failed to grow initially after drug re-addition, adapting with a
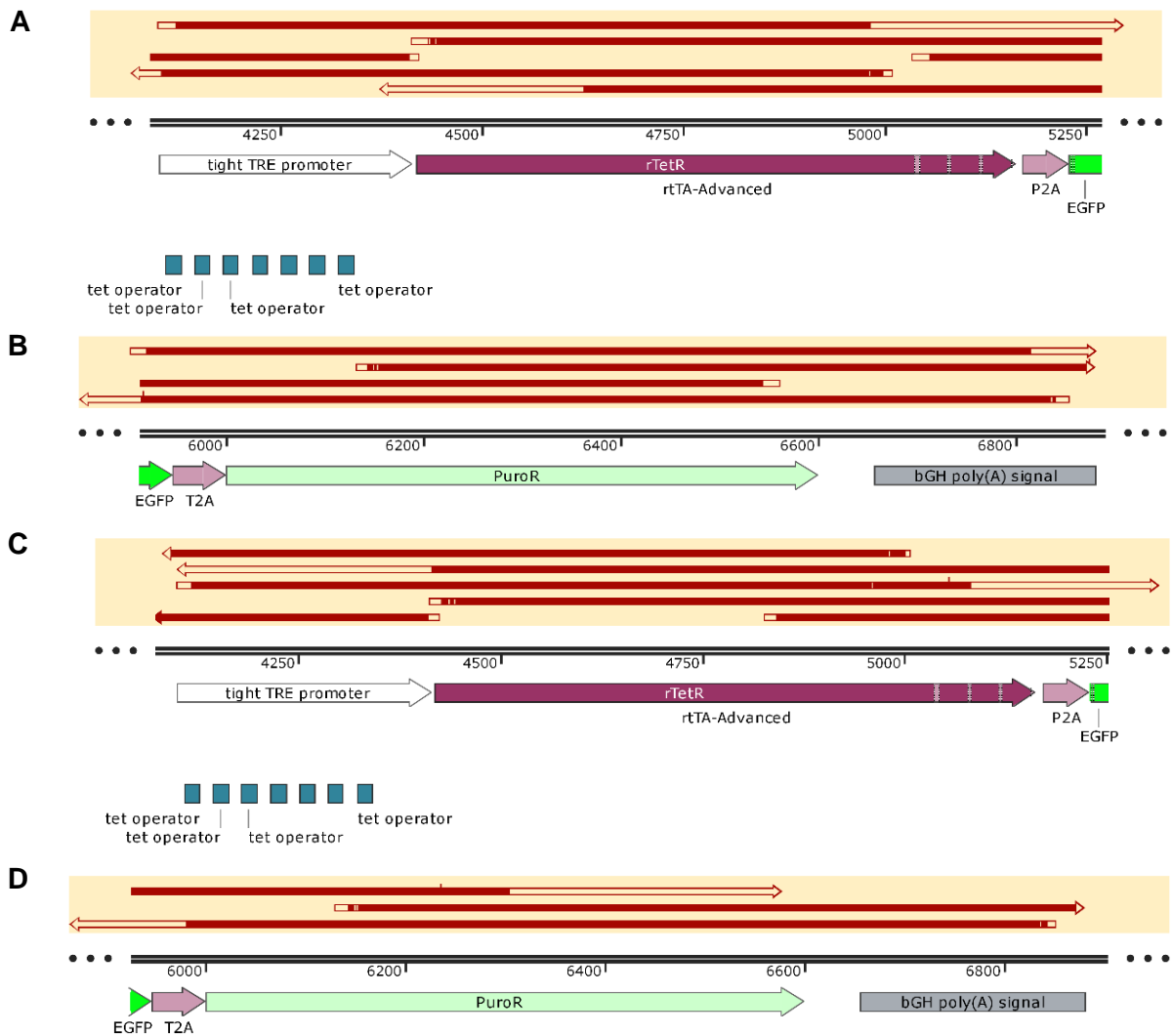
**Figure 36: Sequencing of induced mPF-PuroR replicates 2 and 3 circuits after removal of 35 µg/mL Puromycin indicated no mutations in the circuit.**

**(A-B)** Sequencing reads for the **(A)** pTRE promoter and *rtTA* regulator and **(B)** *PuroR* from induced mPF-PuroR replicate 2 after removing 35 µg/mL Puromycin. **(C-D)** Sequencing reads for the **(C)** pTRE promoter and *rtTA* regulator, and **(D)** *PuroR* from mPF-PuroR replicate 3 after removing 35 µg/mL Puromycin.

long delay (**Figure 26F**). Induced mPF-PuroR cells maintained their expression well above induced and uninduced ancestral cells **(Figures 26C; 38C,D)** and regrew quickly without adaptation upon retreatment (**Figure 26F**). Together with the lack of intra-circuit mutations and reacquisition of drug sensitivity after Doxycycline removal, the evidence supports *rtTA*

76
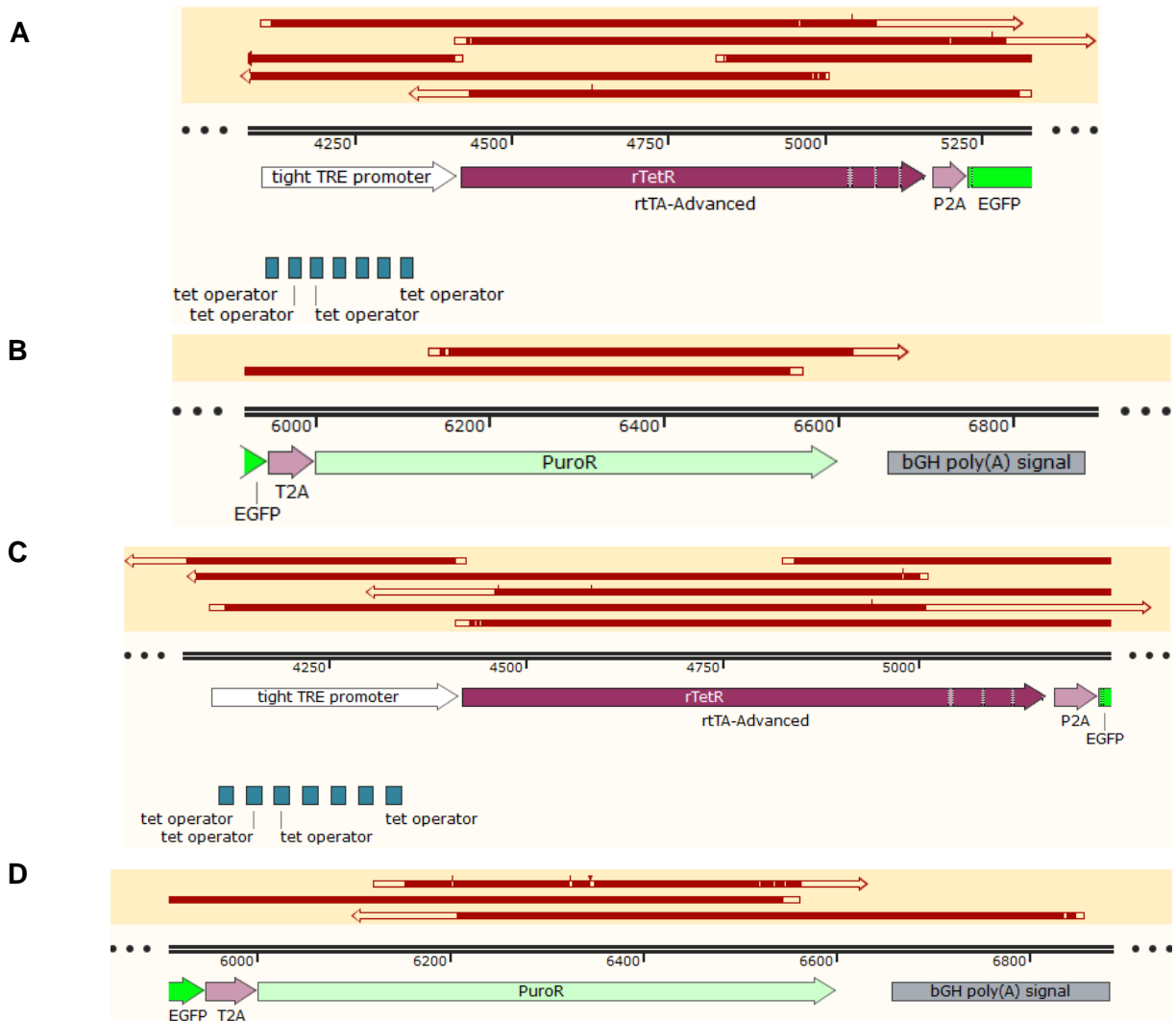
**Figure 37: Sequencing of induced mPF-PuroR replicates 3 and 4 circuits after removal of 50 µg/mL Puromycin did not reveal intra-circuit mutations.**

**(A-B)** Sequencing results for the **(A)** pTRE promoter and *rtTA* regulator and **(B)** *PuroR* from the induced mPF-PuroR replicate 3 circuit after removal of 50 µg/mL Puromycin. **(C-D)** Sequencing results for the **(C)** pTRE promoter and *rtTA* regulator, and **(D)** *PuroR* from the induced mPF-PuroR circuit replicate 4 after removal of 50 µg/mL Puromycin.

induction-dependent extra-circuit alterations that elevate *PuroR* expression to resist 35 µg/mL Puromycin (**Figure 39**).

At the highest level of 50 µg/mL Puromycin, two mPF-PuroR replicates recovered, demonstrating the evolutionary benefit of the noisy mPF-PuroR gene circuit over mNF-PuroR at very high stress levels. Once again, sequencing did not reveal any intra-circuit mutations
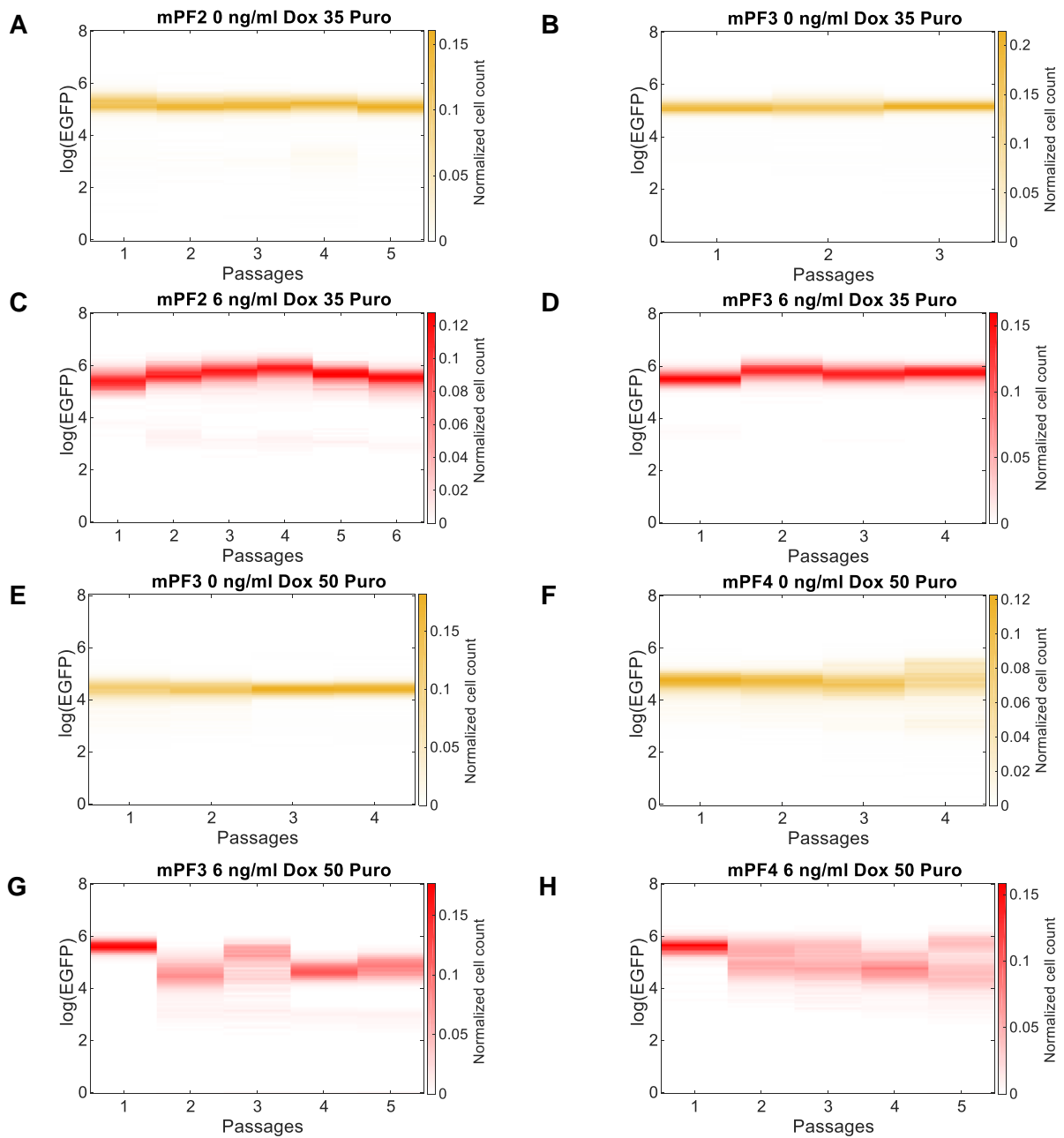
**Figure 38: Expression distributions of mPF-PuroR cells after removal of 35 and 50 μg/ml Puromycin.**

**(A-D)** Log-transformed expression distributions for **(A-B)** uninduced and **(C-D)** induced mPF-PuroR replicates 2 and 3 after removing 35 μg/mL Puromycin. **(E-H)** Expression distributions for **(E-F)** uninduced and **(G-H)** induced mPF-PuroR replicates 3 and 4 after removing 50 μg/mL Puromycin. Each color bar represents normalized cell counts.

(**Figure 37**). The expression of uninduced mPF-PuroR replicates dropped closer to the

baseline DNP mean over ~10 days (**Figures 26D; 38E,F**). Interestingly, for uninduced replicate

3 expression dropped farther down, and re-adaptation to Puromycin occurred even in the

induced condition (**Figure 26G**). Moreover, uninduced replicate 4 cells required more time to

adapt upon retreatment compared to replicate 3, despite slightly higher expression levels,

which suggests distinct heritable alterations contributed to resistance in each replicate. Overall,

we found evidence of distinct extra-circuit heritable inducer-dependent mechanisms

maintaining high *PuroR* expression at 50 µg/mL Puromycin (**Figure 39**).
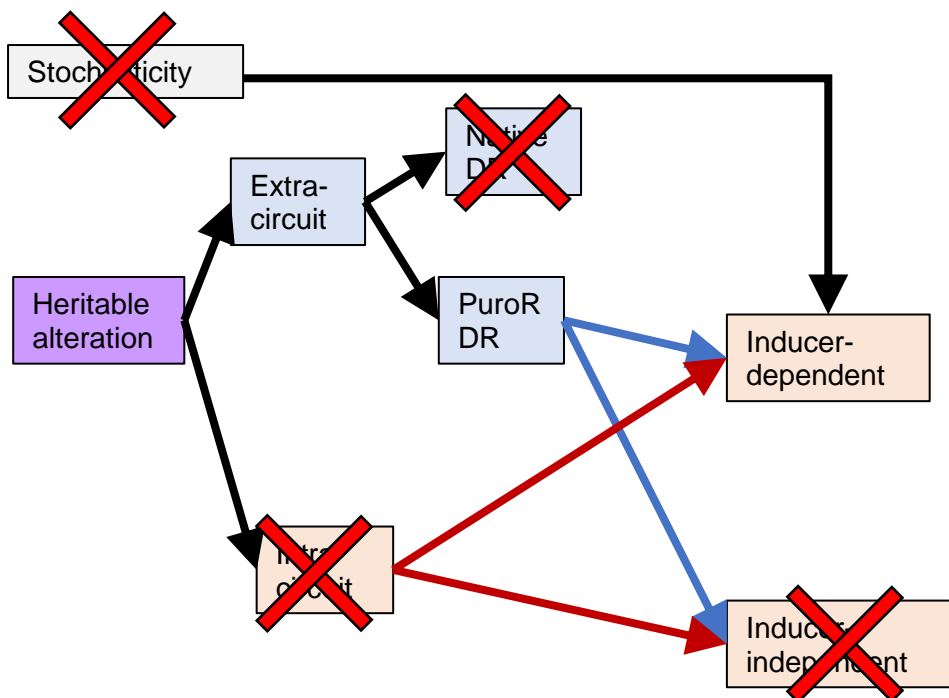


**Figure 39: Populations with mPF-PuroR at 35 and 50 µg/mL Puromycin adapt through extra-circuit inducer-dependent heritable mechanisms.**

We ruled out stochasticity out because the cells did not return to pre-treatment expression

levels. This also implies PuroR-dependent resistance. Sequencing results for mPF-PuroR

replicates 2 and 3 from 35 µg/mL Puromycin and replicates 3 and 4 from 50 µg/mL

Puromycin did not reveal intra-circuit mutations. Expression dropped after removal of

Doxycycline. Therefore, inducer-dependent extra-circuit heritable drug resistance is the

mechanism for mPF adaptation at these stress levels.

Finally, we applied similar criteria to gain insights for 22.5 µg/mL Puromycin, the lowest stress level where mPF-PuroR replicates adapted with a moderate delay. All induced mPF-PuroR replicates maintained their *PuroR* expression above the ancestral levels (**Figure 34B**), but uninduced replicates dropped close to baseline, indicating stable *PuroR* expression-dependent mechanisms of resistance requiring *rtTA*-induction. Accordingly, uninduced mPF-PuroR replicates failed to grow initially during retreatment, showing signs of adaptation (**Figure 40B,D**), as opposed to induced replicates, which grew instantaneously. Gene circuit sequencing revealed no mutations in any replicate (**Figure 35B**), suggesting heritable alterations outside the circuit contributing to inducer-dependent adaptation. Overall, the elevated expression in induced mPF-PuroR cells along with *rtTA*-dependent resistance
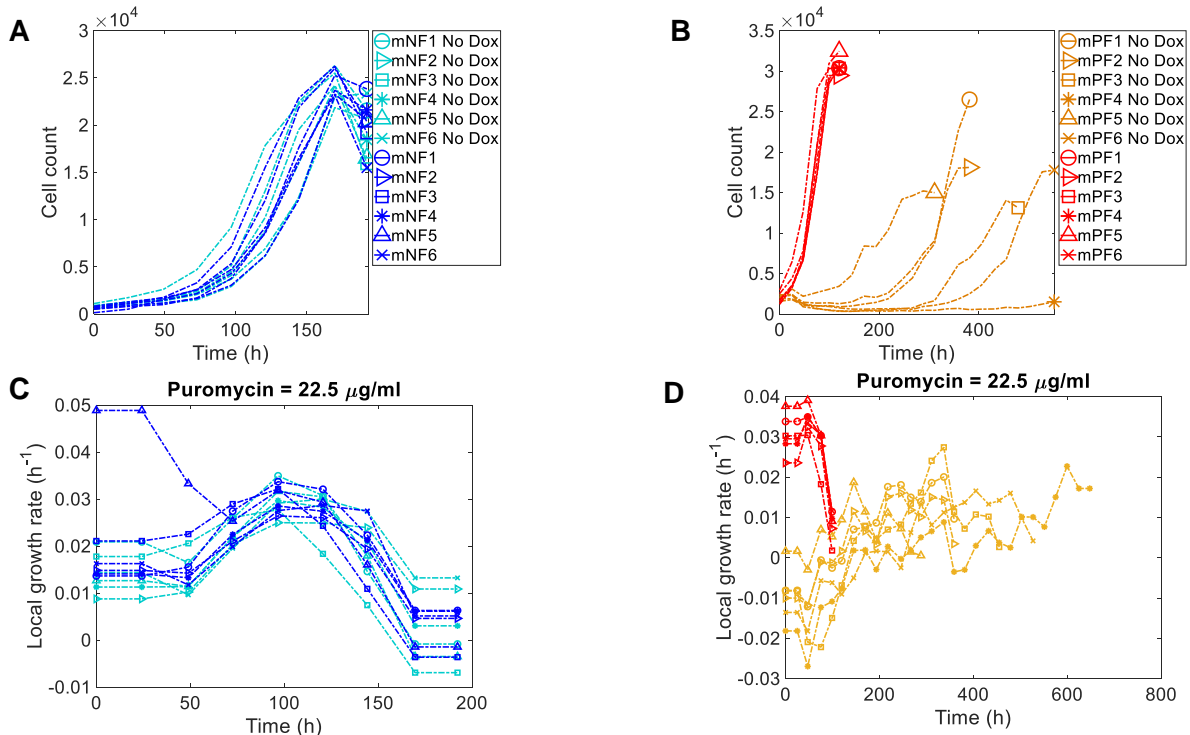


**Figure 40: Final treatment under 22.5 µg/mL Puromycin.**

**(A-B)** Growth curves for **(A)** mNF-PuroR and **(B)** mPF-PuroR replicates retreated with 22.5 µg/mL Puromycin. **(C-D)** Local growth rates for **(C)** mNF-PuroR and **(D)** mPF-PuroR retreated with 22.5 µg/mL Puromycin.

supports heritable extra-circuit genetic or epigenetic *PuroR*-dependent mechanisms of adaptation (**Figure 41**).
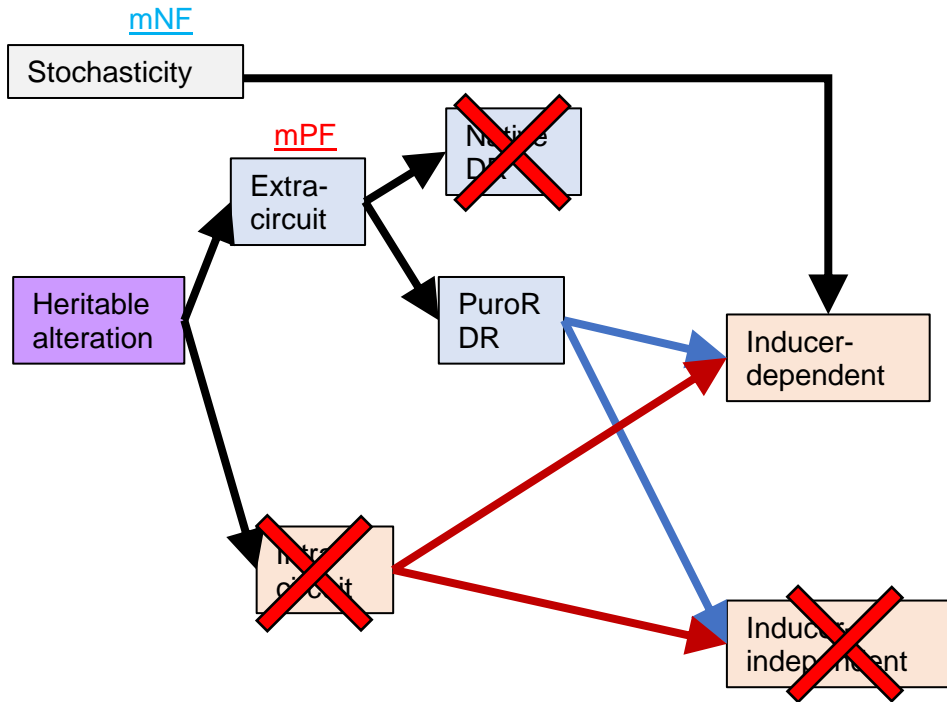


**Figure 41: Populations with mNF-PuroR at 22.5 µg/mL Puromycin adapt through stochastic mechanisms while mPF-PuroR cells adapt through extra-circuit heritable alterations in an inducer-dependent manner.**

Stochasticity cannot be ruled out for cells with the mNF-PuroR circuit because, after drug removal, expression of both induced and uninduced cells dropped to pre-treatment and pre-induction levels, respectively. On the other hand, the induced mPF-PuroR populations maintained their expression above pre-treatment levels, which is consistent with extra-circuit heritable mechanisms. Accordingly, sequencing results for mNF-PuroR and mPF-PuroR replicate 4 revealed no intra-circuit mutations.

## 2.4 Discussion.

Recent studies have raised awareness on cellular heterogeneity and gene expression noise, implying a general benefit for cell populations to overcome drug resistance or metastatic barriers[1,108,115,116,173]. However, earlier evidence for the harmful effects of noise in low stress[43,180] cautions against generalizing these recent observations. In fact, to rigorously study phenotypic effects of noise requires two cell populations with similar means, but different noise levels[43,182,183], which was lacking for mammalian cells. Without such control, we cannot exclude the possibility that the fitness benefit is from higher mean expression. Therefore, how gene expression noise affects mammalian cell survival and evolution were open questions, addressing which required isogenic mammalian cell lines with mean-decoupled noise control. We established such noise control with high- and low-noise gene circuits to study how network noise contributes to drug resistance evolution in CHO cells. While earlier work in yeast indicated that noise can aid or hinder short-term survival depending on the balance between drug and resistance protein levels[43,201], the evolutionary effects of noise are only recently being unraveled[32,181]. Here, by experimentally evolving synthetic gene circuit-harboring CHO cells in Puromycin, we show that noisy mPF networks hinder evolution at low stress, but aid evolution at high stress, mimicking the effects of noise on short-term survival[43].

We combined experimental evolution and synthetic gene circuits to drive evolutionary adaptation in mammalian cells. Since the pioneering studies of prokaryotic experimental evolution[202], the field has expanded into yeast[32,181,203] and fruit flies[204]. Mammalian cell evolution studies are timely and relevant to cancer[205], but they are still rare and have not involved synthetic gene circuits. Experimental evolution of artificial gene circuits in microbes[32,181,206] provided mechanistic insights across multiple scales of time and biological organization, by reducing the influence of complex and incompletely known native gene regulatory networks. Thus, synthetic gene circuits facilitate the development of predictive models that reveal unexpected, higher-level effects, which would be more difficult for natural gene networks.

The experimental system we developed is a feasible model[205] for the long-term evolutionary response of cancer cells to translational inhibitors. Puromycin compromises protein synthesis in a broad range of cell types, like emerging cancer therapeutics targeting mRNA translation[207,208]. Moreover, Puromycin itself has been proposed as a potent anticancer agent specifically released from a prodrug in cancer cells[209]. Considering that over 80% (19,711/24,383) of the predicted CHO protein-coding genes have homologs in human[196], studying drug resistance evolution in this cell line should be as relevant as mouse cell line models of drug resistance are to human cancers.

Overall, the data suggest that at the highest stress levels that cause prolonged growth suppression, cells with high-noise mPF networks recover through stable, but unknown extra-circuit genetic or epigenetic drug resistance mechanisms. At milder stress levels, cells with the low-noise mNF network adapt partly by mutating the circuit to abrogate repressor function. Surprisingly, adapting CHO cells always take advantage of the non-native *PuroR* gene. The mechanisms vary, and most likely include direct *PuroR* upregulation by intra- or extra-circuit alterations. The intra-circuit mutations or lack thereof reflect the fact that random evolutionary changes can more easily disrupt repression than facilitate activation. The exact extra-circuit heritable mechanisms behind the evolutionary adaptation remain to be studied as whole-genome and -transcriptome sequencing of CHO cells advance[210].

We used different (mNF and mPF) networks to control noise properties, keeping the role of networks and noise intertwined. We think noise properties (amplitude and memory and then switching to a persister state) should be more relevant for initial survival, when the protein level fluctuations make the difference between survival and death. On the other hand, network topology (repression versus activation of drug resistance) and how it can be beneficially altered seems to matter more at longer, evolutionary time scales. In the future, it will be interesting to try controlling noise while minimizing differences in network topology[187], to separate better the evolutionary effects of networks and gene expression noise.

Comparing the experimental evolution time courses with the evolutionary model and sequencing results suggested that persister cells convert to stably resistant proliferating cells at high stress levels. The mammalian drug-tolerant persister state could derive from a chromatin-mediated transition, which previously has shown sensitivity to HDAC inhibitors[104], or could depend on GPX4 expression[105]. However, in these experiments "persister" simply means cells that neither divide, nor die in stress – mediated by many possible mechanisms, such as the formation of polyploid cells[211]. Nonetheless, the successful elimination of low-noise populations without resistance at high stress levels provides hope for noise-controlling treatment strategies in cancer, like HIV[184].

The findings could also have relevance for metastasis, where gene expression noise may drive transitions into metastatic states[1]. Based on this chapter's findings, genes can confer drug survival in mammalian cells by expressing the gene above a critical expression threshold, which later leads to evolutionary adaptation. In a bistable network, an expression threshold can separate two stable expression states that switch. Both scenarios can lead to the same outcome: survival based on single-cell expression above a critical expression threshold. Survival for cells highly expressing metastasis-driving genes would be conferred in the environment outside the primary tumor. High gene expression noise then may lead to better survival in a metastatic setting, which has not been demonstrated with endogenous pro-metastatic regulatory genes. By complementing the exogenous system for studying gene expression noise in mammalian drug resistance, endogenous gene expression readout systems may reveal fundamental systems-level principles on noise in the context of an endogenous regulatory network.

**Chapter 3: Characterization of a live-cell endogenous expression reporter for the pro-metastatic BACH1 transcriptional repressor.**

**3.1 Introduction.**

**3.1.1 Background.**

Understanding of the dysregulated activity of transcription factors[212], signaling pathways[213], micro-environmental cues[214,215], non-coding RNAs[216], and metabolic processes[215,217] that facilitate cancer progression and survival beyond the primary tumor has become better understood at the population-average level. However, although next-generation sequencing methods have advanced knowledge of intra-tumor heterogeneity in terms of spatially distinct genomic alterations[161,218], fundamental insights regarding gene expression and cancer progression at the single-cell level are still lacking[34]. While protein-coding sequence mutations can drive cancer progression at the primary site[219,220], evidence for specific mutations driving metastasis is sparse[221] and only recently have novel mutations been identified[222]. This indicates that metastasis may rely on nongenetic mechanisms on top of the mutations already present in the primary tumor[33,114,173]. An important consequence of nongenetic mechanisms is the presence of nongenetic heterogeneity of tumor cell populations[223]. Nongenetic heterogeneity can take on the form of gene expression fluctuations (noise) in oncogenic or tumor-suppressing genes[1,173]. While single-cell genetic and transcriptomic sequencing has deepened our knowledge of tumor heterogeneity[224], the investigation of gene expression noise for metastasis regulatory factors in living cells has lagged. Part of the difficulty arises from the lack of single-cell endogenous expression reporters in cancer cell lines.

In breast cancer, the epithelial-to-mesenchymal transition (EMT) has generally been thought to mediate pro-metastatic phenotypes, though conflicting studies have created debate whether EMT is required at all[144,225]. Beyond the epithelial-mesenchymal transition regulatory program, breast cancer metastasis is associated with the pro-metastatic transcriptional factor *BACH1* and anti-metastatic Raf kinase inhibitory protein (*RKIP*), which both mutually repress each other[1]. This *RKIP-BACH1* toggle switch network (**Figure 42**) can predispose cells towards stochastic pro- and anti-metastatic state transitions mediated by the bistable overall positive feedback loop[1]. Additionally, *BACH1* inhibits its own expression by recruiting histone
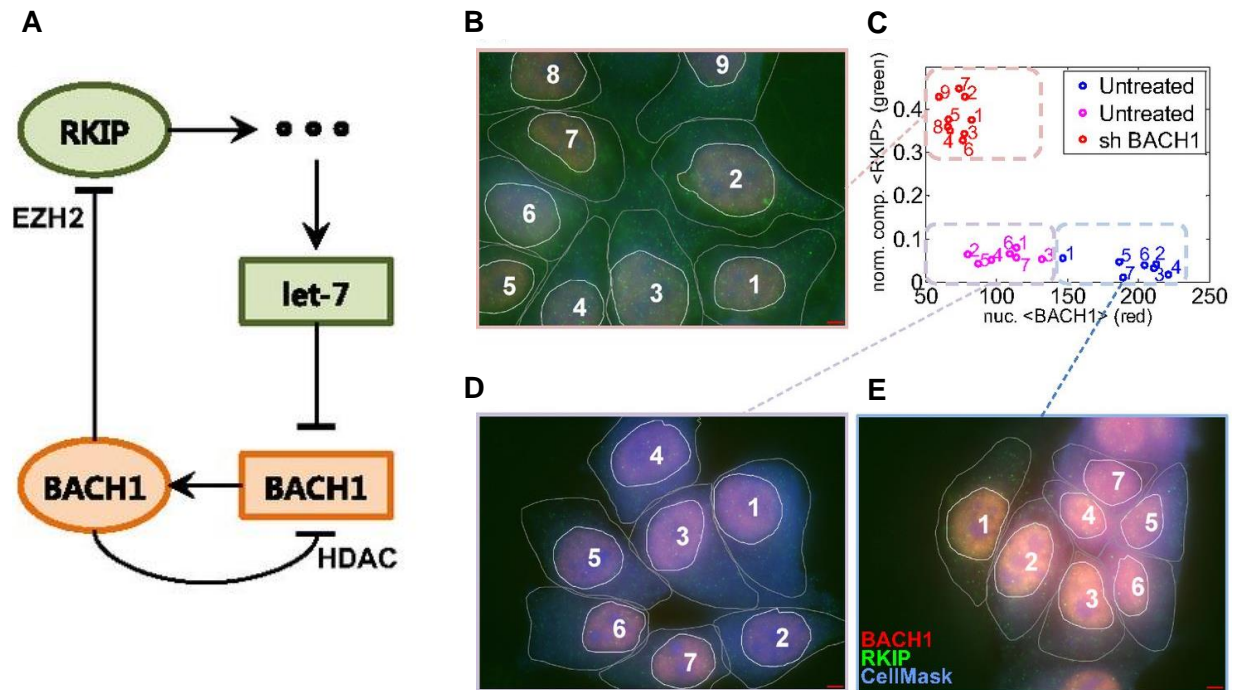
**Figure 42: The *RKIP-BACH1* regulatory axis controls metastatic state transitions.**

**(A)** The *BACH1* transcription factor participates in a double negative feedback look with the anti-metastatic *RKIP* gene, while *BACH1* negatively regulates its own expression in an HDAC-dependent manner. **(B-C)** Immunofluorescent staining of *RKIP* (green), *BACH1* (red), and the CellMaskBlue dye (blue), revealed lowered *BACH1* levels and increased *RKIP* in fixed MCF7 cells after **(B)** RNAi-mediated down-regulation of *BACH1*, leading to a shift from a **(C)** bimodal *BACH1* distribution with low *RKIP* to unimodal low *BACH1* and high *RKIP* distribution. **(D-E)** Immunofluorescent images of the *BACH1* **(D)** low- and **(E)** high-expressing subpopulations. The figures were derived from a co-authored publication[1].

deacetylase complexes (HDACs)[1]. Depending on the strength of the *BACH1* auto-regulatory feedback loop and *RKIP* levels, the regulatory network can transition between anti-metastatic and pro-metastatic states directly or could transverse a bistable regime where *BACH1* expression levels are noisy[1].

While immunofluorescent staining of fixed cells supports the network structure, it is unclear how *BACH1* expression in live cells behaves in response to various perturbations that

could transition cells into the pro- or anti-metastatic state. Doing so requires a readout of endogenous *BACH1* expression. An endogenous expression readout in living cells could involve tagging or replacing the endogenous gene of interest with a fluorescence marker, GFP binding to MS2 coat proteins encoded in mRNA, or exogenously introducing the BACH1 promoter driving a fluorescent reporter at a safe harbor locus. To address this knowledge gap, I designed a CRISPR-Cas9 knock-in procedure for fluorescently tagging *BACH1* by integrating the red fluorescent reporter mCherry downstream of a self-cleaving T2A element at the end of the genomic *BACH1* locus (**Figure 43A**). To this end, I report the successful construction of six HEK293 cell lines having the endogenous *BACH1* locus tagged with mCherry. Fluorescent tracking of *BACH1* expression levels in live cells could help uncover expression dynamics in response to various perturbations. HEK293 cells can also serve as a test bed for CRISPR-Cas9-mediated knock-in techniques based on their ease of transfection and robust growth characteristics. I also attempted to create a *BACH1* reporter system in the triple-negative MDA-MB-231 breast cancer cell line but could not find a successfully integrated clone. With the endogenous expression readout system, it is possible to uncover regulatory principles for the *BACH1-RKIP* network that provide a better understanding of gene expression noise in metastasis.

**3.1.2 Significance.**

Escape from and survival outside the primary tumor needs curtailing to reduce the number of cancer deaths. The ability of cancer cells to survive outside of the tumor may depend on gene expression of pro-metastatic genes, allowing intravasation into blood vessels in the case of *BACH1*[159]. Thus, we must study the expression dynamics associated with cell survival in a hostile environment, whether it is during drug treatment or metastatic progression. A fluorescent readout of endogenous gene expression will provide insights on single-cell gene expression dynamics associated with metastasis, complementing the exogenous approach for drug resistance in the previous chapter. In the most optimistic scenario, a better understanding of gene expression noise in a natural gene regulatory network may contribute to preventing

metastasis in the way that low noise prevented survival of cells at very high levels of drug stress.

**3.2 Method and materials.**

**3.2.1 Cell culture maintenance.**

HEK293 (ATCC® CRL-1573™) cells were grown in Dulbecco's Modified Eagle Medium (DMEM) containing 4500 mg/mL glucose (ThermoFisher Scientific, 11965092). The 1X DMEM media was supplemented with 10% fetal bovine serum and 100 units/mL and 100 µg/mL of Penicillin and Streptomycin, respectively. MDA-MB-231 (ATCC® HTB-26™) cells were grown in 1X RPMI 1640 media (Life Technologies, 11875093) containing a 5% fetal bovine serum supplement along with 1X Penicillin and Streptomycin. Each cell line was grown at 37ºC in an incubator filled with 5% $CO_2$. Sub-passaging was done as described in section 2.2.1.

**3.2.2 Preparation of reagent working stocks.**

DMEM media in the wound healing experiments contained 0.2% fetal bovine serum. Human Transforming Growth Factor-β1 (Sigma Aldrich, H8541-5UG) was aseptically reconstituted at a frozen stock concentration of 5 mg/mL in 4 mM Hydrochloric acid (Aqua Solutions, H2505-500ML), which contained 0.1% low-endotoxic bovine serum albumin supplement (Sigma Aldrich, A4919-1G). TGF-β1 was diluted to 5 ng/mL in DMEM media for use in perturbation experiments. Hemin (Sigma Aldrich, H9039-1G) was reconstituted at a stock concentration of 5 mM in 20 mM Sodium hydroxide solution (EMD Millipore, SX0607D-6) kept at 4ºC with minimal exposure to light. The final hemin concentration in DMEM media used in perturbation experiments was 20 µM. Cell culture plates with hemin were wrapped in aluminum foil to avoid exposure to light. During experiments with negative selection, MDA-MB-231 cells transfected with the donor clone modified to contain the *HSV-TK* gene were treated with 10 µg/mL Ganciclovir (InvivoGen, sud-gcv) after single-cell flow-sorting.

**3.2.3 CRISPR-Cas9 DNA donor clone design and construction.**

The BACH1::T2A::mCherry donor clone plasmid (GeneCopoeia, DC-HTN000623-PUC19-B) and the corresponding all-in-one CRISPR-Cas9 plasmid (GeneCopoeia, HCP000623-CG01-1-B) with a *BACH1*-targeting guide RNA induced a double-stranded break within 60 bp upstream of the endogenous stop codon of *Homo sapiens* BTB domain and CNC

homology 1 (*BACH1*) transcript variant 1 (NM_206866) from Chromosome 21. The system was designed to induce homology directed repair of the double-stranded break in the presence of a donor clone possessing two *BACH1* homology arms (655 bp long upstream and 586 bp long downstream of insertion site) flanking the end of the *BACH1* amino acid coding region prior to the stop codon. Between the homology arms, the donor clone contains a translational self-cleaving T2A element followed by the red fluorescent reporter mCherry.

The modified donor clone plasmid containing the *HSV-TK* negative selection marker outside of the *BACH1* homology arms was constructed using the HiFi DNA Assembly Master Mix (NEB, E2621S). In detail, the primers pKRAS-HSV.R and pKRAS-HSV.F amplified the vector backbone of a custom-built KRAS.A donor clone (GeneCopoeia, DC-HTN000555-D10-B), which includes the *HSV-TK* gene driven by a PGK promoter. Next, the BACH1 upstream and downstream homology arms flanking the T2A::mCherry genetic element were amplified with the BACH1-mCh-Ins.F and BACH1-mCh-Ins.R primers from the original BACH1::T2A::mCherry donor clone plasmid. Both fragments contain 20 bp of overlap sequence for the opposing fragment. About 0.088 pmol of the insert and vector backbone fragments were ligated in the HiFi assembly reaction. For the sequences of cloning primers, see **Table 6**.

| Primer name | Sequence (5'->3') |
|---|---|
| pKRAS-HSV.R | TGTTCCCTTTAGTGAGGG |
| pKRAS-HSV.F | CAATTCTACCGGGTAGGG |
| BACH1-mCh-Ins.F | gctcgaaattaaccctcactaaagggaacaCCAAAACCTGAAAAATGCTTTG |
| BACH1-mCh-Ins.R | gaaaagcgcctcccctacccggtagaattgGCTTAAAGCCTGATTTTAGG |

**Table 6: Primers for constructing the BACH1::T2A::mCherry donor clone with the *HSV-TK* negative selection marker outside the homology arms.**

The primer sequences are listed above. Lower-case nucleotides overlap with the complementary fragment during HiFi assembly.

### 3.2.4 CRISPR-Cas9 DNA transfection.

For optimization and hedging purposes, the Cas9/gRNA-expressing plasmid along with the original BACH1::T2A::mCherry donor clone plasmid were transfected in three wells of pre-

seeded HEK293 cells in concentrations of either 3.75 or 7.5 µL Lipofectamine 3000 reagent for the two samples with 1 µg total DNA (0.5 µg per plasmid) and 7.5 µL for the one sample with 2 µg total DNA (1 µg per plasmid). Based on the low amount of red fluorescent HEK293 cells seen after 2 days post-transfection (1%>), the three HEK293 transfection samples in the 6-well plate were pooled together for subsequent single cell flow-sorting into 96-well plates for clonal expansion.

MDA-MB-231 cells were nucleofected with 0.5 µg of the Cas9/gRNA plasmid and the improved BACH1::T2A::mCherry donor clone using the Nucleofector™ 2b device. Cells were seeded at $10^6$ per well in a 6-well plate the prior night. In the nucleofection experiment, trypsinized cells were centrifuged at 90 x g for 10 minutes. The cell pellet was then resuspended with 100 µL Cell Line Nucleofector™ Kit Solution V containing Supplement 1 (Lonza, VVCA-1003). After transferring the cell solution to a provided cuvette that is then placed inside the device, the Nucleofector™ program X-013 was selected for immediate nucleofection. The resulting nucleofected cells were diluted with 500 µL medium and transferred into a pre-prepared 6-well plate for subsequent fluorescent imaging and single-cell flow sorting.

**3.2.5 Fluorescence-activated cell sorting (FACS).**

The Stony Brook Medicine Research Flow Core facility FACSAriaIII instrument sorted the HEK293 and MDA-MB-231 cells. For the HEK293 BACH1 reporter clone expansion, single HEK293 cells were flow-sorted based on low or high red fluorescence, which is indicated in the first number (1 and 2, respectively) of each clone ID after expansion and genomic DNA isolation. The plate row follows as a single letter (A-H) after the 96-plate identifying number and the column represented by an additional number following the letter ID from 1-12 ($\#_{plate}$ Letter$_{row}$ $\#_{column}$). The clone ID system was applied for both potential HEK293 and MDA-MB-231 BACH1 clonal samples. The HEK293 clone 2D9 is an example, expanded from plate 2 (high fluorescence), row D, and column 9.

There were two separate flow-sorting clonal expansion experiments for MDA-MB-231 cells from the original nucleofected sample, with the second clonal expansion comprising remaining cells that were cryo-preserved. In the first MDA-MB-231 clonal flow-sorting experiment, there were three 96-well plates. For the first four rows in each plate, single cells were sorted in each well based on low fluorescence, whereas the last four rows had cells sorted for high fluorescence.

For the second MDA-MB-231 clonal screen, there were four 96-well plates that contained flow-sorted cells (Clone ID: $1_{plate}$ Letter$_{row}$ #$_{column}$ to $4_{plate}$ Letter$_{row}$ #$_{column}$). The first and second plate were seeded with single cells previously bulk-sorted prior to the first MDA-MB-231 screen by either low or high fluorescence, respectively. For these two plates, four additional gates of increasing fluorescence intensities sorted the cells into two rows for each gate. Additionally, the third plate consisted of MDA-MB-231 cells from the unsorted, original transfection having low levels of fluorescence. Finally, cells in the fourth plate from the originally unsorted transfection were then sorted based on mid-to-high fluorescence.

**3.2.6 Flow cytometry expression measurements.**

The BD Accuri™ C6 bench-top flow cytometer measured single-cell *mCherry* fluorescence for the *BACH1* experiments. The flow data was gated as described in section 2.2.13. Cells were prepared for flow as described in section 2.2.6. In HEK293 experiments, the number of gated cells per sample consistently analyzed in each experiment ranged from at least 1,000 in perturbation experiments to 110,000 in the initial clone characterization experiment. The mean and standard error of the mean from flow data were plotted.

**3.2.7 Time-lapse microscopy.**

The Nikon TiE motorized inverted microscope acquired 10x phase contrast and fluorescent images at pre-defined x- and y-coordinates for each experiment with the tiling of multiple frames into a large image as described in section 2.2.7. *BACH1* reporter experiments either consisted of one frame or 3x3 field frames. Fluorescence background subtraction was done as described in section 2.2.17. The look-up tables were consistent within experiments.

### 3.2.8 Junction PCR, genotyping PCR, and associated Sanger sequencing.

The junction PCR assay required a primer that uniquely binds to human genomic DNA on Chromosome 21 upstream (outside) of the first homology arm containing part of the last *BACH1* intron. This primer cannot target the homology arms or off-target loci in the genome. A reverse primer then targeted the synthetic portion of the knock-in. The assay only works on genomic DNA samples from clones that successfully recruited the homology directed repair pathway to close the double stranded break from Cas9 and replaced the break with the homology arms flanking the synthetic construct, leading to the introduction of the fluorescent tag. Thus, the junction PCR only amplifies the knock-in *BACH1* allele, not the wild-type *BACH1* allele or a donor clone that randomly integrated.

For the genotyping PCR assay, the same forward primer lying outside the upstream homology arm in the native BACH1 intron and a different reverse primer targeting the homology arm downstream of the synthetic fluorescent tag amplified the wild-type *BACH1* allele (1165 bp) from untransfected human genomic DNA. If the T2A::mCherry construct was integrated as a heterozygous allele, a second band (1927 bp) will amplify along with the wild-type *BACH1* allele. A third band could be the third HEK293 Chromosome 21[226] or a potential duplication or deletion event inherent in repair after CRISPR-Cas9 cutting[227]. The verified *BACH1* clones were screened for non-specific integration of the donor clone by amplifying the AmpR gene. Primers used for the junction PCR and genotyping assays are listed in **Table 7**.

| Primer name | Sequence (5'->3') | Assay |
|---|---|---|
| BACH1-Int-Seq-2-f | TCCCTTCAGTGGGTTTGATGTT | Junction and Genotyping PCR |
| mCh-Junct-2-r | CACGTCACCGCATGTTAGAAG | Junction PCR |
| BACH1-HA-R-GSeq-r | GACGCTGCCAAAACTTCAGG | Genotyping PCR |
| AmpR-split.rev | TTGATCGTTGGGAACCGGAG | Integration PCR |
| Origin-Seq-r | GCGTTGCTGGCGTTTTTCC | Integration PCR |

**Table 7: Primers for amplifying successful integrations with junction PCR and *BACH1* alleles with genotyping PCR.**

The primer sequences are listed above. BACH1-Int-Seq-2-f was used in both the junction PCR and genotyping PCR assays.

PCR products were submitted to the DNA Sequencing Facility at Stony Brook University and sequenced by the Sanger method with primers listed in **Table 7** and **Table 8**.

| Primer name | Sequence (5'->3') |
|---|---|
| mCh-Junct-r | CTCCATGTGCACCTTGAAGC |
| BACH1-C-Term-f | GTGGGATCTCAGATTTCTGTCAGCAG |
| BACH1-C-Term-Seq-r | CTGCTGACAGAAATCTGAGATCCCAC |
| BACH1-aa661-668-f | TGGTGAACTGGCGTTACCAT |

**Table 8: Sequencing primers for *BACH1* genotyping PCR.**

The primer sequences are listed above. All other primers used for sequencing the *BACH1* locus are listed in **Table 7**.

### 3.2.9 Wound healing experimental design.

The wound healing assay[228] required 24 hours of serum starvation (0.2% fetal bovine serum) prior to wounding to restrain cell proliferation, so that cells filling the wound gap did so through migration. Using a sterile micropipette tip, a straight vertical line of cells was scraped off from the cell monolayer in each well at the first time point (0h) in the wound healing assay.

There were two separate wound healing experiments at different time scales of imaging. In the four-day wound healing experiment, each well contained at least two wounds serving as replicates with or without 5 ng/mL TGF-β1 treatment added 48 hours before wounding. Images were acquired every 24 hours during the four-day experiment. The first experiment wounding event occurred in a biosafety cabinet before media replenishment, imaging, and standard incubation. During the second experiment, images were acquired every 5 minutes up to one hour. Cell monolayers in the second experiment were wounded directly on the microscope after acquisition of a pre-wound image and then immediately imaged for the remaining hour.

### 3.2.10 *Pri-let-7g* over-expression experiment.

The *pri-let-7g* microRNA was expressed from the pcDNA3-pri-let-7g vector, which was a gift from Narry Kim[229] (Addgene plasmid # 51381 ; http://n2t.net/addgene:51381 ; RRID:Addgene_51381). Cells were transfected with 1 μg of pcDNA3-pri-let-7g vector using the Lipofectamine 3000 kit. The transfected cells were incubated for four-days before flow cytometry measurement along with an untransfected set of replicates.

### 3.2.11 Optimized direct PCR methodology.

During optimization of the original direct PCR method[230], the necessity of flash-freezing or trypsin mixture was tested, the volume of trypsin to dilute with 1X DPBS was varied, and so was the number of cells extracted as template DNA. In the optimized protocol, the first step requires dilution of about 1/5th of trypsinized cells from a confluent 96-well plate with 5-fold volume 1X DPBS without calcium or magnesium. After spinning the diluted cells for 5 minutes at 300 x g and removing supernatant, the cell pellet, even if not visible, was resuspended with 10 µL 1X DPBS. Next, the trypsin-free cell solution was flash-frozen at -80⁰C. After 2 or more hours, the sample was heated at 65⁰C for two minutes. After this final step, about 1 µL of lysed solution served as a DNA template for PCR (see **Figure 57**).

### 3.2.12 Statistical analysis.

All statistical analyses employed custom MATLAB scripts. Statistical tests were conducted as described in section 2.2.16. To assess for bimodality (two peaks) in a single-cell expression distribution, a z-score statistic[67] tested whether two distribution extrema are significantly multi-modal. The z-score statistic was calculated after finding local minima/maxima extrema pairs from smoothened log-transformed single-cell expression distributions. Given that the smoothened distributions (with a moving average = 800) become Gaussian based on the Central Limit Theorem, two distribution extrema were significantly bimodal when the height difference between the local minimum and the right and left maxima peaks exceeded 4 standard deviations (see Equation 11 for score and 12 for standard deviation),

$$\left[\begin{matrix} |H_L - H_M| \\ |H_R - H_M| \end{matrix}\right] - 4 * \left[\begin{matrix} \sigma_L \\ \sigma_R \end{matrix}\right] + \left[\begin{matrix} 4 \\ 4 \end{matrix}\right] = \left[\begin{matrix} Z_L \\ Z_R \end{matrix}\right], \tag{11}$$

with the standard deviation equaling:

$$\left[\begin{matrix} \sigma_L \\ \sigma_R \end{matrix}\right] = \frac{\sqrt{\left[\begin{matrix} H_L \\ H_R \end{matrix}\right]}}{\sqrt{2*moving_{average}}}. \tag{12}$$

### 3.2.13 Image processing.

For the HEK293 BACH1 reporter 96-hour wound healing experiment, wound gap area was segmented using the Nikon Elements AR segmentation option called 'Wound area Detection in time'. Segmentation thresholds were first automatically determined and then manually adjusted to avoid including cells in the calculated gap region. After exporting the gap area values from Nikon Elements AR, the data was analyzed with custom MATLAB scripts. The calculated gap area for each image over time was normalized by the gap area from the first time point (0h) of each individual replicate (**Figure 50-51**).

## 3.3 Results.

### 3.3.1 Developing a red fluorescent reporter system tagging endogenous *BACH1* expression *in vivo.*

To measure endogenous BACH1 gene expression at its native locus as demonstrated with other genes[231], I designed a CRISPR-Cas9 knock-in plasmid that contains two separate homology arms for *BACH1* exon 5 and the *BACH1* 3' untranslated region flanking right before the *BACH1* stop codon a T2A self-cleaving motif and the red fluorescent mCherry reporter (**Figure 43A**). Along with a second plasmid expresing a *BACH1* guide RNA targeting the corresponding native *BACH1* genomic locus and Cas9, I transfected the *BACH1* donor clone plasmid into HEK293 cells and clonally expanded single cells after fluorescence-activating cell
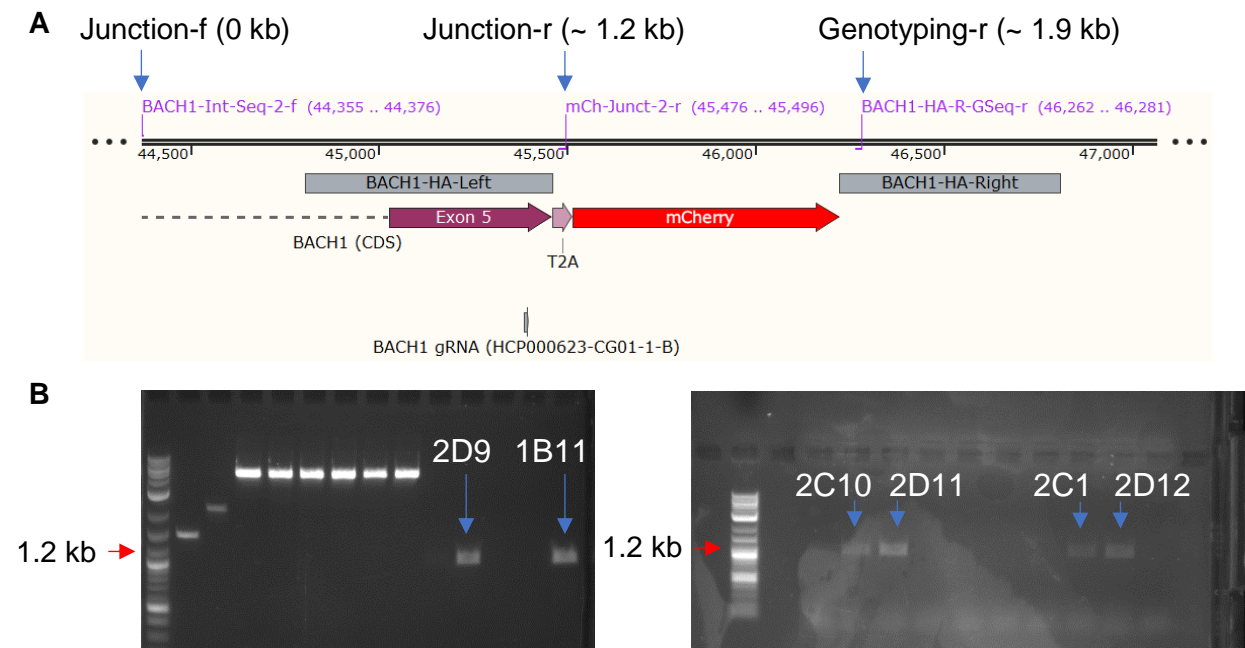
**Figure 43: Design and validation of HEK293 with mCherry-tagged BACH1 as an endogenous expression readout.**

**(A)** The T2A motif and mCherry reporter gene were inserted directly upstream of the BACH1 stop codon by providing two *BACH1* homology arms (BACH1-HA-Left and -HA-Right) during double-stranded DNA repair. PCR primers are indicated above. **(B)** Junction PCR results with integrant clone identifiers above the bands. 2-log ladder assessed size.

sorting. In summary, the CRISPR-Cas9 knock-in methodology will introduce mCherry at the *BACH1* genomic locus in HEK293 cells to allow live-cell monitoring of endogenous gene expression.

To validate mCherry integration at the native *BACH1* locus in flow-sorted HEK293 cells previously transfected with the donor clone and guide RNA/Cas9 plasmid, I amplified the synthetic T2A::mCherry target from genomic DNA samples with a primer outside of the upstream homology arm using the junction PCR technique. Subsequently, the results indicated successful integration of the reporter construct at the *BACH1* locus in 6 HEK293 clones (**Figure 43B**). In line with this finding, the HEK293 BACH1 2D9 clone displayed red fluorescence under microscopy compared to the ancestral HEK293 cell line (**Figure 44**). The remaining clones did not amplify during junction PCR, which indicates failure of mCherry integration at the endogenous BACH1 locus. Thus, the junction PCR indicated the knock-in experimental design worked as intended for six HEK293 clones, facilitating integration of the synthetic T2A::mCherry at the native *BACH1* locus.
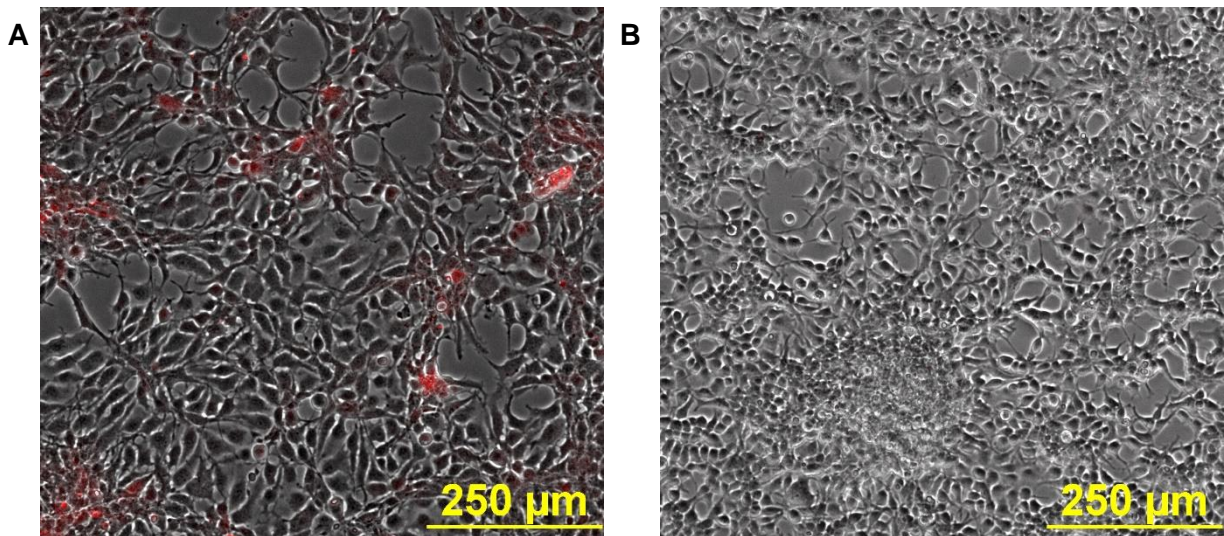


**Figure 44: Fluorescence from HEK293 BACH1 reporter clone 2D9.**

**(A-B)** Overlay bright field and fluorescent images of the **(A)** HEK293 clone 2D9 and **(B)** parental HEK293 cells.

To confirm whether the T2A:mCherry tag integrated into all *BACH1* alleles in each clone, I ran a genotyping PCR that amplified with a primer binding outside the upstream homology arm in the native *BACH1* intron and a primer inside the downstream homology arm, thereby allowing simultaneous amplification of both the wild-type, non-integrated *BACH1* allele and the knock-in allele. Genotyping PCR results suggested heterozygous integration in all six clones (**Figure 45A**). In fact, the genotyping assay revealed three bands after amplifying the genomic *BACH1* locus in all clones. Based on the fragment size, the ~1.2 kb band consisted of the wild-type allele with no knock-in, and the ~1.9 kb band matched the predicted size that includes the T2A::mCherry knock-in. Unexpectedly, a third band with a size of ~1.5 kb amplified that did not match *in silico* predictions arising from CRISPR-Cas9 modification. Since *BACH1* resides in Chromosome 21 with three chromosomal copies in the aneuploid HEK293 cell line[226], each band could have amplified from the three individual *BACH1* copies.

To confirm whether the *BACH1* genotyping bands contained either the wild-type allele or the T2A::mCherry knock-in, I sequenced the PCR products. For the 2D9 clone fragment with the putative knock-in, sequencing did not reveal evidence of substantial mutations (**Figure 45B**). Additionally, the wild-type allele (~1.2 kb) also lacked mutations (**Figure 45C**). Lastly, the anomalous allele aligned with the *BACH1* locus but displayed many mutations and indels (**Figure 45D**). A complex reconfiguration of the *BACH* locus in addition to multiple indel events inherent in CRISPR-Cas9 gene editing[227] may have led to the aberrant allele. Importantly, one knock-in allele sequencing read (blue in **Figure 45B**) from the 2D9 clone spanned from the synthetic T2A sequence across the upstream homology arm and ended with coverage in the native *BACH1* intron not contained in the donor clone homology arms, which strongly supports site-specific integration at the *BACH1* locus. There was no evidence of fixed mutations in the knock-in allele for the 2C1, 2C10, and 2D11 clones (**Figure 46**). Thus, sequencing of the three 2D9 clone alleles supported heterozygous integration of the T2A::mCherry tag in-frame with the C-terminus of *BACH1*, along with a non-integrated undamaged allele and a disrupted third allele.
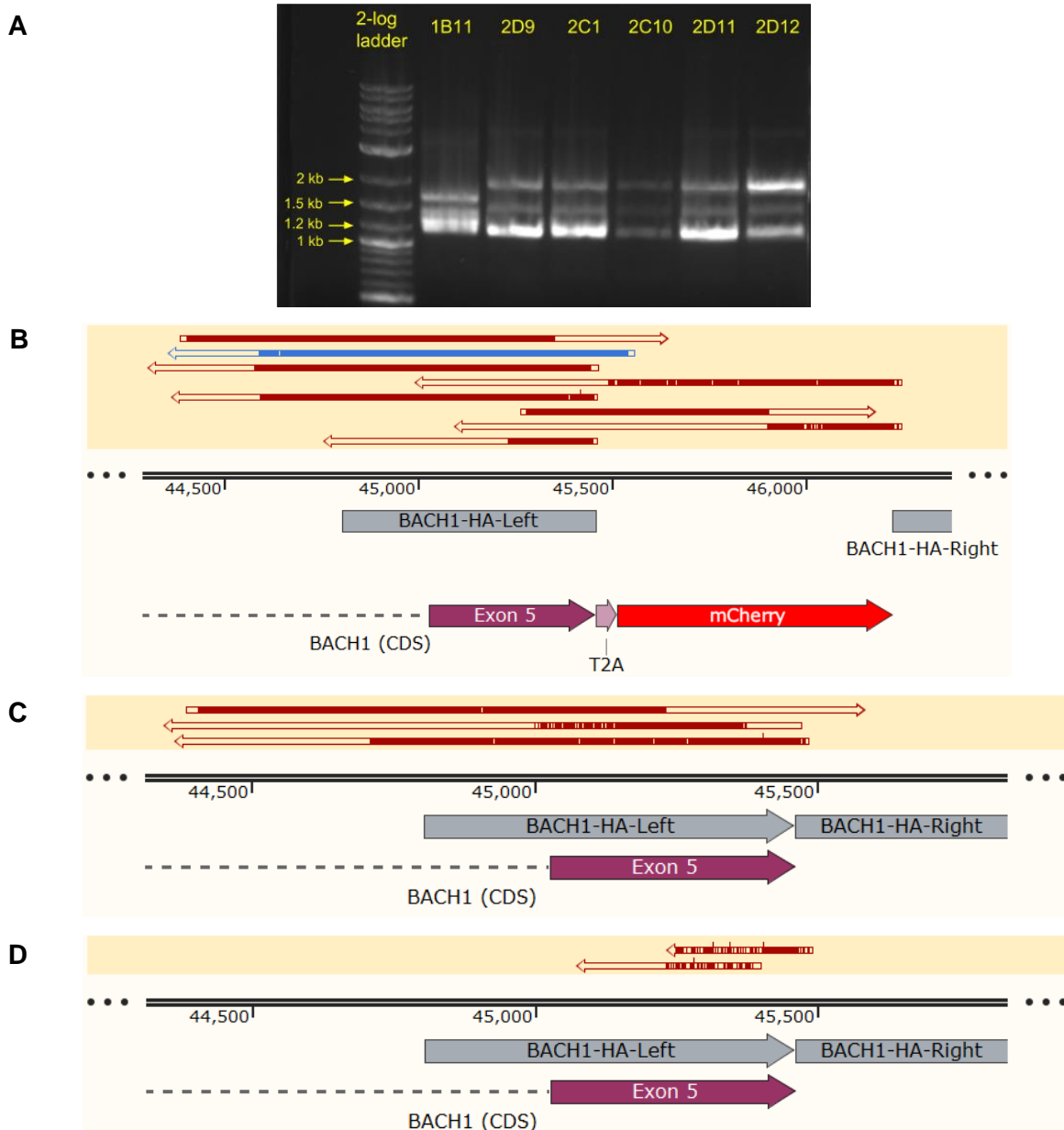
**Figure 45: Genotyping PCR for HEK293 BACH1 reporter clone 2D9 alleles.**

**(A)** The genotype of six successfully integrated BACH1 clones were analyzed by amplifying a region outside of the upstream homology arm, across the synthetic construct, and ending within the downstream homology arm. **(B-D)** Sequencing for the 2D9 clone alleles, including the **(B)** knock-in, **(C)** wild-type, and **(D)** a damaged third allele. The blue read in **(B)** includes a continuous read spanning from mCherry to outside the upstream homology arm.

To determine whether the verified integration event in the six HEK293 clones coincided with random integration of the entire donor clone, I amplified a region of the donor clone
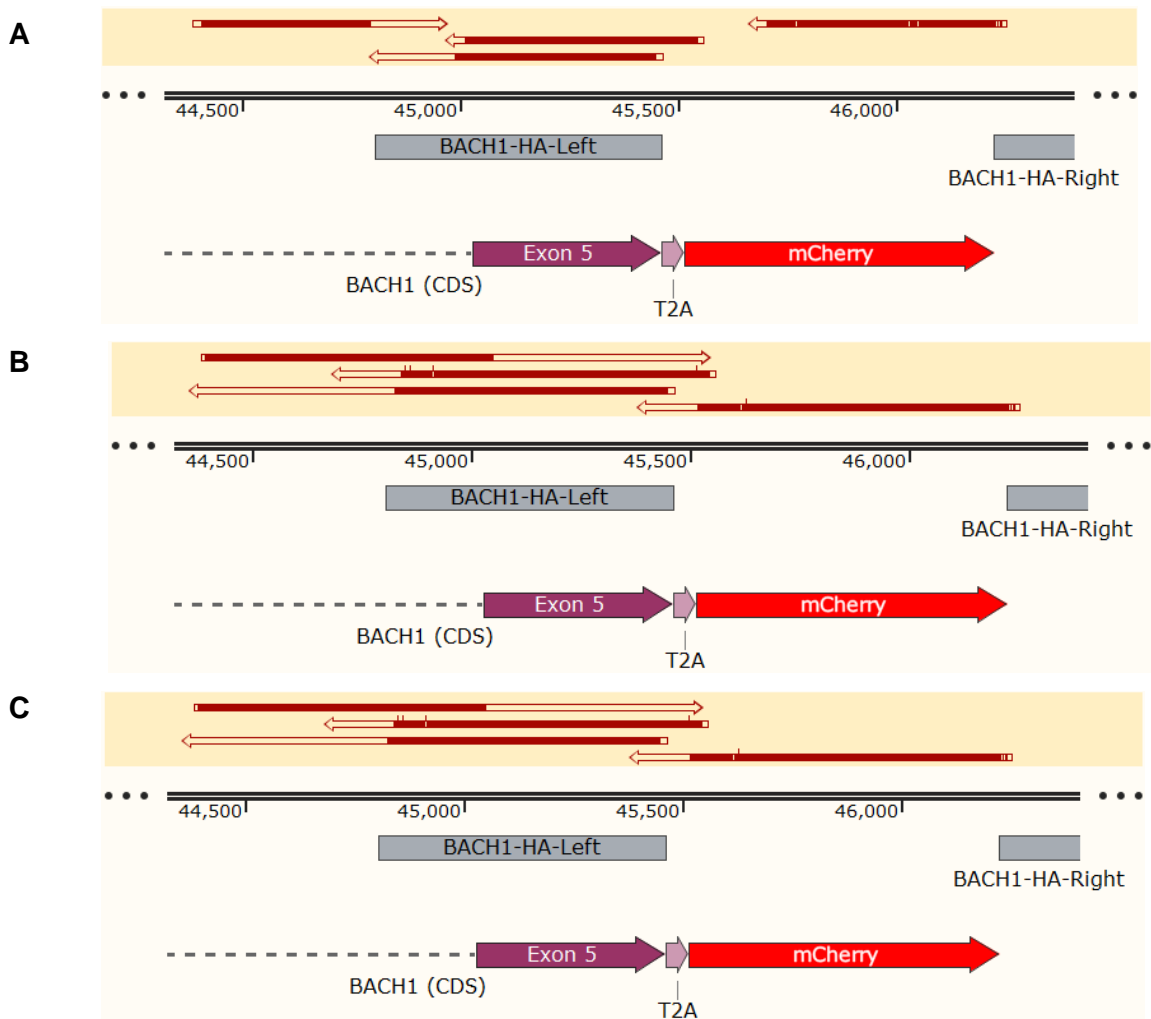
**Figure 46: Genotyping PCR for HEK293 BACH1 reporter clones 2C1, 2C10, and 2D11 knock-in alleles.**

Sequencing for the knock-in allele from HEK293 clones **(A)** 2C1, **(B)** 2C10, and **(C)** 2D11.

plasmid backbone outside of the homology arms from genomic DNA of the six verified HEK293 BACH1 reporter clones. Compared to a known randomly integrated clone lacking a tag at the *BACH1* locus serving as a positive control, the PCR experiment revealed non-specific integration of the donor clone in all six clones (**Figure 47**). Despite the unideal random integration, it may not lead to fluorescence from the randomly integrated locus. A lack of any expression from the native integrated *BACH1* locus along fluorescence from a random integration event would be too improbable to explain the fluorescence results. Otherwise, the low fluorescence arises from low *BACH1* expression in HEK293 cells, which has been
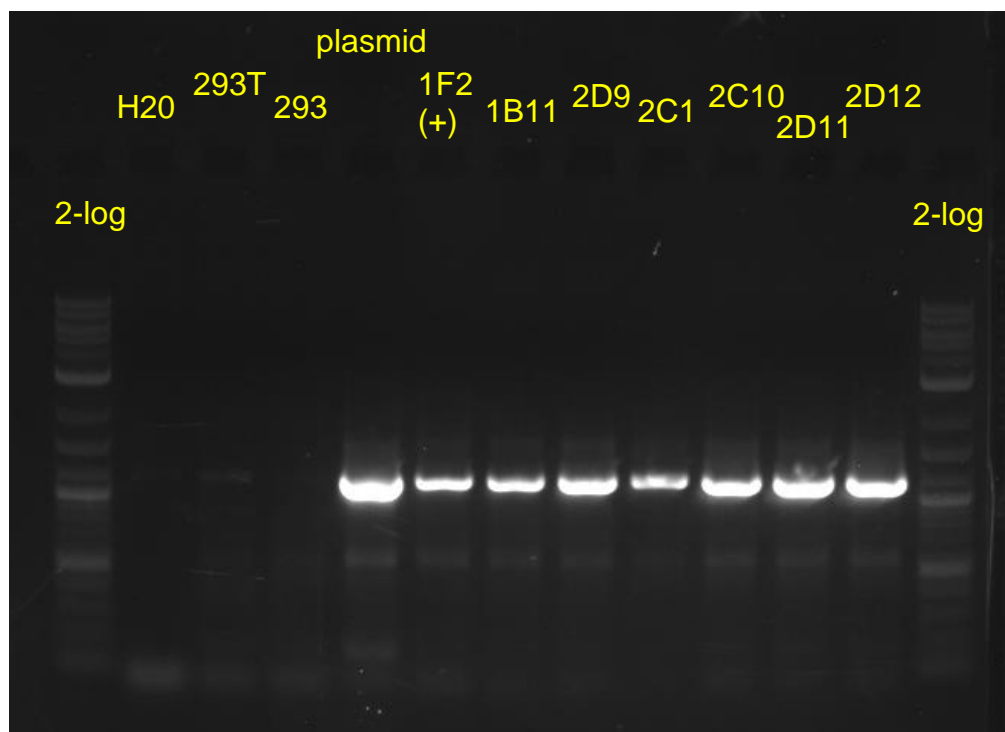
**Figure 47: Random integration of the donor clone in the six HEK293 BACH1 reporter clones.**

The presence of the AmpR gene was assessed from the genomic DNA of wild-type HEK293, a randomly integrated, but non-successful endogenous integration clone for BACH1 (1F2, positive control), the donor clone plasmid positive control, and the six HEK293 BACH1 clones. 2-log ladder served to assess product size.

previously demonstrated (see NCI60 cell line expression at BioGPS.com). Importantly, the BACH1::T2A::mCherry construct does not contain a promoter, leaving the mCherry reporter promoter-less at a random genomic locus. During a successful integration, mCherry will express if the endogenous *BACH1* promoter is active. Therefore, the observed fluorescence should come from the mCherry reporter tagging the *BACH1* locus, but this requires functional verification. In summary, I successfully detected six HEK293 clones with the endogenous *BACH1* locus tagged with a fluorescent protein, which despite the presence of at least one randomly integrated donor clone, can serve as a tool to measure single cell *BACH1* expression, including gene expression noise, in response to molecular and mechanical perturbations *in vivo.*

**3.3.2 Characterization of unperturbed endogenous HEK293 BACH1 reporter clone expression.**

To assess mCherry fluorescence from the six HEK293 BACH1 reporter clones, I characterized single-cell red fluorescence by flow cytometry. All six clones have higher mean fluorescence than a HEK293 parental cell auto-fluorescence control (**Figure 48A**). The 2D9 clone exhibited the highest mean mCherry fluorescence while 2C10 fluorescence displayed the lowest. Despite having the highest mean expression, 2D9 gene expression noise exceeded the five other clones (**Figure 48B**). Single-cell expression distributions for the HEK293 BACH1 2D9 clone suggest the noise arises from a relatively larger tail of high-expressing cells (**Figure 48C**; green distribution). Unlike the rest of the clones, the HEK293 clone 2C10 displayed a statistically significant bimodal distribution[67] (**Figure 48C**; cyan distribution in middle right; $Z_L$ = 4.6073 and $Z_R$ = 7.2478, significant bimodality at $Z_{L,R} > 4$). In all, the six BACH1 reporter clones exhibited various levels of fluorescence that are specific to each clone. Next, I will cover various chemical perturbations and one physical perturbation applied to live HEK293 cells to better understand the fluorescence response dynamics and functionally verify the reporter system.

**3.3.3 TGF-β1 induction of the HEK293 BACH1 reporter clones.**

To assess the functionality of the HEK293 BACH1 fluorescent reporter system with a chemical perturbation, we needed a test for a known *BACH1*-specific interaction that either increases or decreases *BACH1* expression. Transforming Growth Factor Beta I (TGF-β1) ligand may increase BACH1 expression[232], but single-cell verification has not confirmed this finding in living HEK293 cells. Therefore, I treated the HEK293 BACH1 2C1 and 2C10 clonal cells with 5 ng/mL TGF-β1 over 24 hours. The two clones represent distinct expression distribution modalities observed in the six HEK293 BACH1 clones: 2C1 having one peak and 2C10 having two peaks. Flow cytometry revealed no significant increase in mean fluorescence (**Figure 49A**). The gene expression noise also did not change significantly (**Figure 49B**). Since the study reported an increase in *BACH1* expression from a western blot in NMuMG cells,
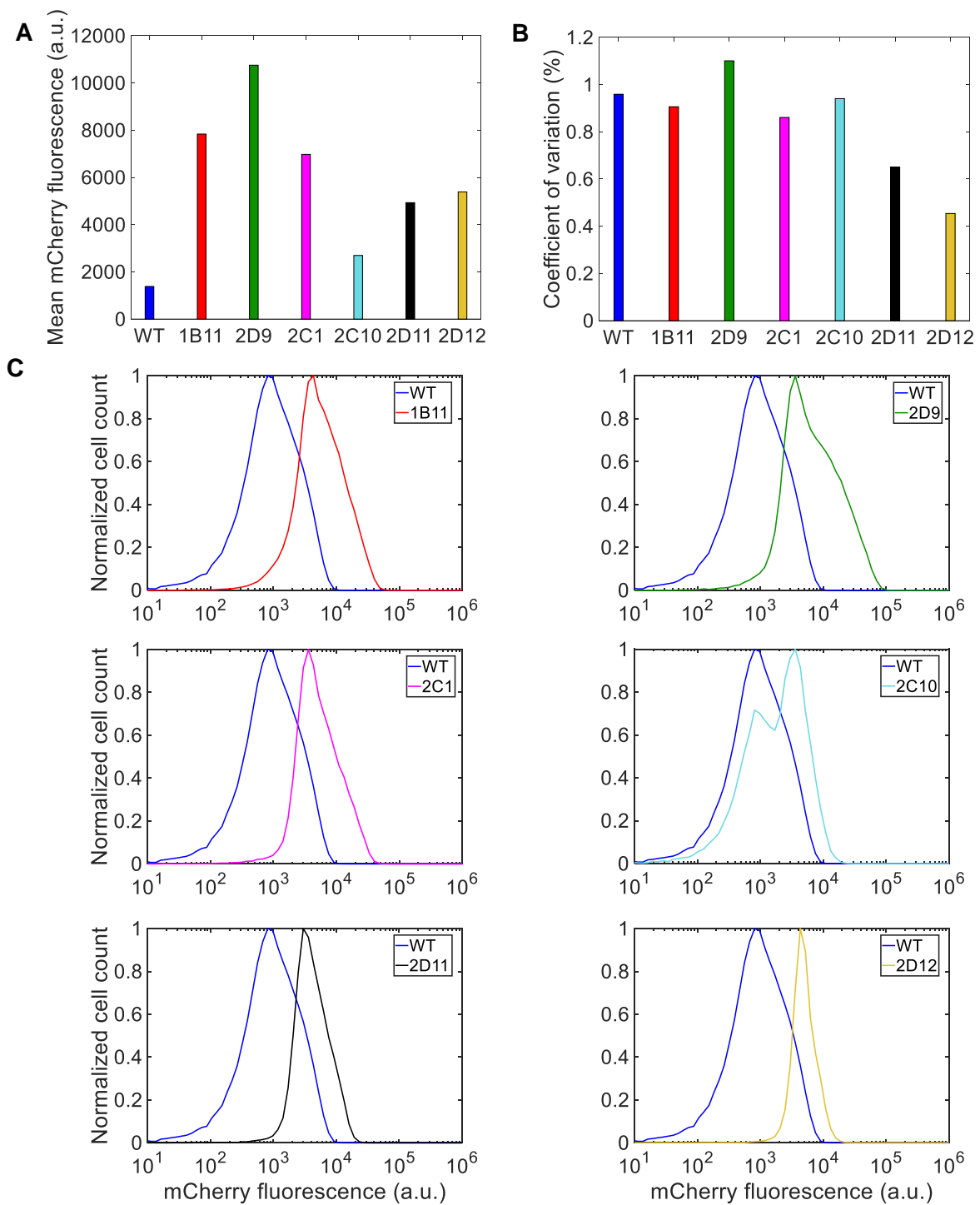
**Figure 48: Expression characterization of the HEK293 BACH1 reporter clones.**

**(A)** Mean expression for the HEK293 parental cells and six BACH1 reporter clones. **(B)** Gene expression noise (coefficient of variation). **(C)** Single-cell expression distributions for the HEK293 parental compared to each BACH1 reporter clone.

which are from *Mus musculus*, while HEK293 cells are from humans, the results may reflect
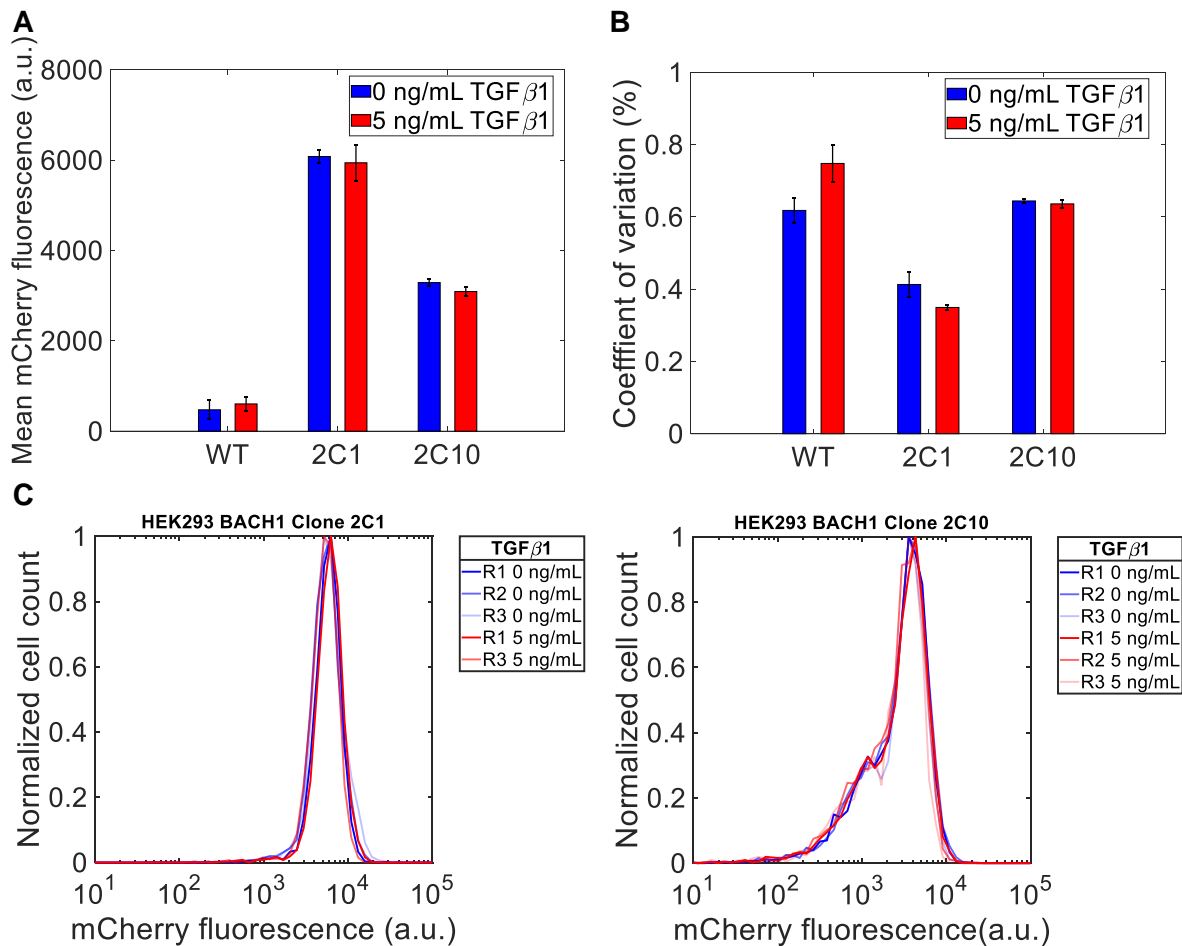
**Figure 49: TGF-β1 induction of HEK293 BACH1 reporter clones for 24 hours.**

**(A)** Mean expression of mCherry in two BACH1-tagged HEK293 reporter clones (2C1 and 2C10) after 24 hours of TGF-β1 stimulation. **(B)** Gene expression noise for two BACH1 reporter clones and wild-type cells. The experiments were conducted in triplicate. Error bars represent the standard error of the mean (SEM). **(C)** Single-cell expression distributions for HEK293 BACH1 clones 2C1 (left) and 2C10 (right). The cell counts were normalized by the bin with the highest number of cells.

species-specific differences in the regulatory network controlling *BACH1* expression or a non-active *BACH1* locus. It is also possible that 24 hours was not a sufficient amount of time to observe an effect on expression. To summarize, the cells treated with TGF-β1 did not respond with changes in mCherry fluorescence, indicating the need for a perturbation assay that shows an effect of TGF-β1 on human cells.

To phenotypically characterize the fluorescence response with a physical perturbation in conjunction with a chemical perturbation, I induced the HEK293 BACH1 2D9 clone with 5 ng/mL TGF-β1 vs untreated 2D9 cells prior to wounding a monolayer of cells in a scratch assay. We used the 2D9 clone over the unimodal 2C1 or bimodal 2C10 clones because the unimodal population of 2D9 cells displayed the highest gene expression noise (CV) during expression characterization, exceeding the bimodal 2C10 clone. By utilizing the clone with the highest noise, we can best address the core hypothesis in a physiologically-relevant phenotype (migration after wounding) for BACH1. Qualitatively, fluorescence tended to increase initially at the frontier edge of the wound for 2D9 cells (**Figure 50**), though the untreated sample monolayer of cells were flipped over after wounding, making interpretation between samples difficult. Over the course of four days, both untreated and treated cells that migrated to close the wound tend to have decreased fluorescence (**Figure 50**, 24h-96h). However, to check whether the fluorescence rapidly changed required a considerable amount of time to acquire images at time intervals less than 24 hours. As for the wound closure rate, the data was inconclusive regarding whether TGF-β1 induces migration (data not shown). In conclusion, the fluorescence at the wound edges increased after wounding, but the time scales of fluorescence dynamics required further investigation.

**3.3.4 Time scales of wound healing.**

To study the timescale of fluorescence dynamics during wound healing, I conducted a scratch assay for serum-starved low-noise HEK293 BACH1 reporter 2C1 clone without TGF-β1 induction and imaged every 5 minutes over the span of one hour. Qualitatively, the cells remained relatively uniform in red fluorescence before the scratch **(Figure 51A)**. Right after the scratch, fluorescence slightly increased for parts of the wound edge (**Figure 51B**). Moreover, fluorescence tended to non-uniformly increase even further at the edge during later time points (**Figure 51C**). Overall, the images visually demonstrated an increase in fluorescence at the wound edge within an hour of wounding.

**A**

0h

24h

48h

72h

**B**

0h

24h

48h

72h
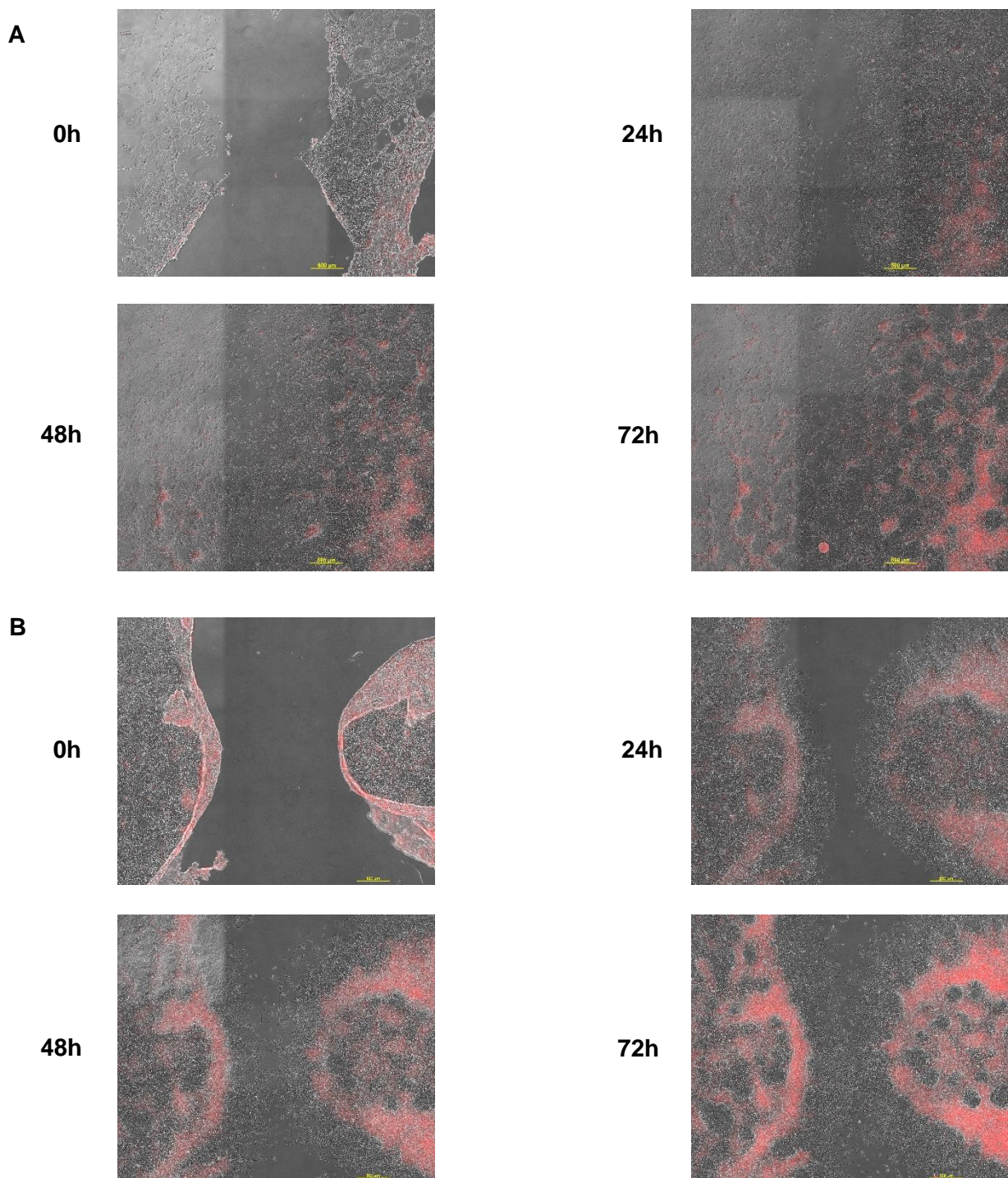
**Figure 50: Images of wound healing for BACH1 clone 2D9 after TGF-β1 induction.**

Images of HEK293 BACH1 reporter clone 2D9 cells **(A)** treated with 5 ng/mL TGF-β1 and

**(B)** untreated cells during a wound healing assay over time.

To quantify the fluorescence while retaining spatial information, the red fluorescence for

each x-coordinate was averaged over the y-axis pixels. As distance from the wound became

shorter, the overall fluorescence increased after wounding (**Figure 51D**). In conclusion, I

observed overall increased fluorescence at the edge of the wound over an hour after wounding

the serum-starved HEK293 BACH1 reporter clone 2C1 cells, with qualitative evidence of

increased gene expression noise at the wound edge.
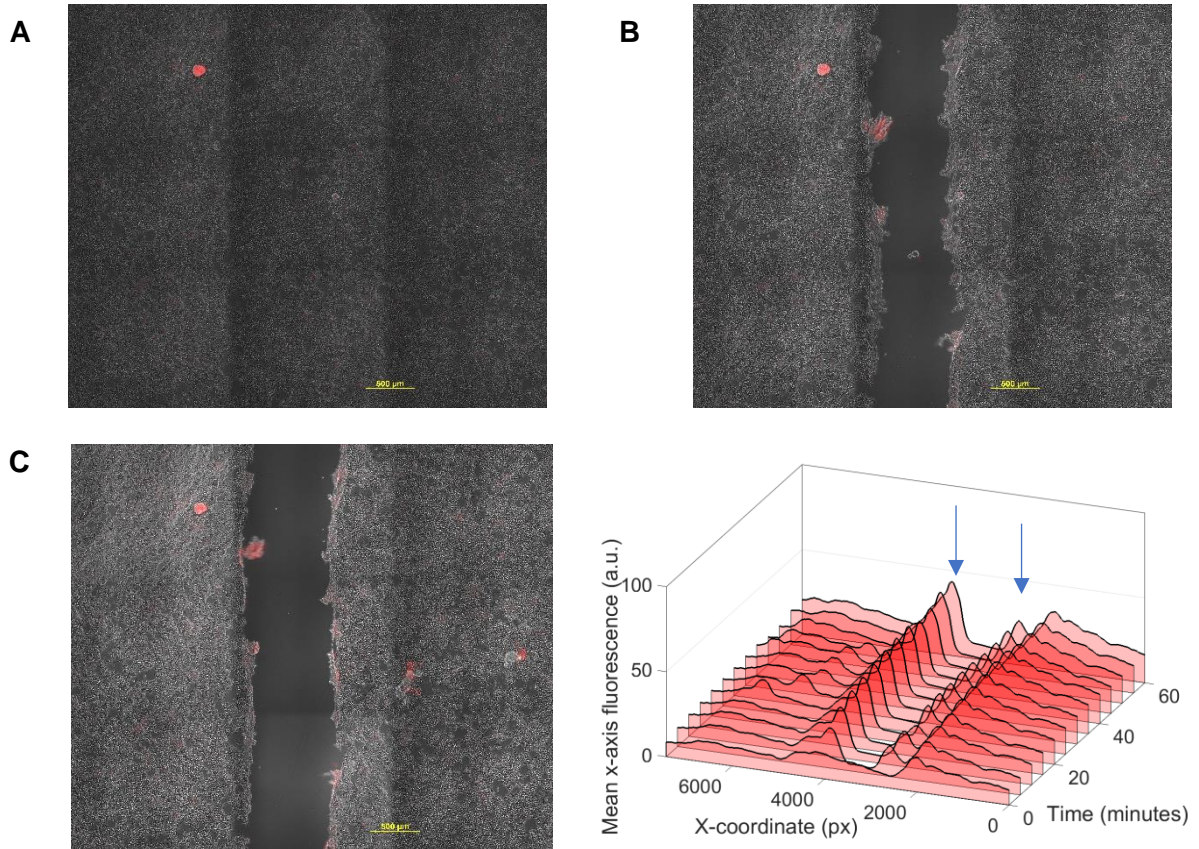


**Figure 51: Wound healing for BACH1 clone 2C1 over one hour.**

**(A)** Pre-scratch image of 2C1 cells starved of growth serum. **(B)** Post-scratch (0 minute)

image of serum-free 2C1 cells. **(C)** Image of 2C1 cells 40 minutes after the scratch. **(D)**

Mean intensity of 2C1 cells for each x-axis coordinate (average of y-axis pixel intensities)

with the wound edge indicated by blue arrows. X-coordinate 0 starts at the left image edge.

### 3.3.5 Effect of hemin treatment on HEK293 BACH1 reporter clone expression.

To test whether other molecular perturbations affect mCherry fluorescence in a manner

expected for *BACH1* expression, I treated the HEK293 BACH1 reporter clones 2C1 and 2C10

with hemin, which promotes cytoplasmic transport and subsequent degradation of BACH1

protein[233,234]. Since hemin-mediated degradation of BACH1 can take 4 hours[234], I treated the

HEK293 BACH1 clones 2C1 and 2C10 for four hours with 20 µM hemin. I found no significant

change in mean mCherry fluorescence for BACH1 clones 2C1 and 2C10 compared to

untreated samples (**Figure 52A**), though the treated samples tend to have slightly higher,

nonsignificant mean fluorescence. However, the effect of duration of the hemin treatment on

mCherry fluorescence was unclear, which could mask the response to the perturbation.

Additionally, a longer time series may reveal a stronger fluorescent response after the initiation

of BACH1 protein degradation.
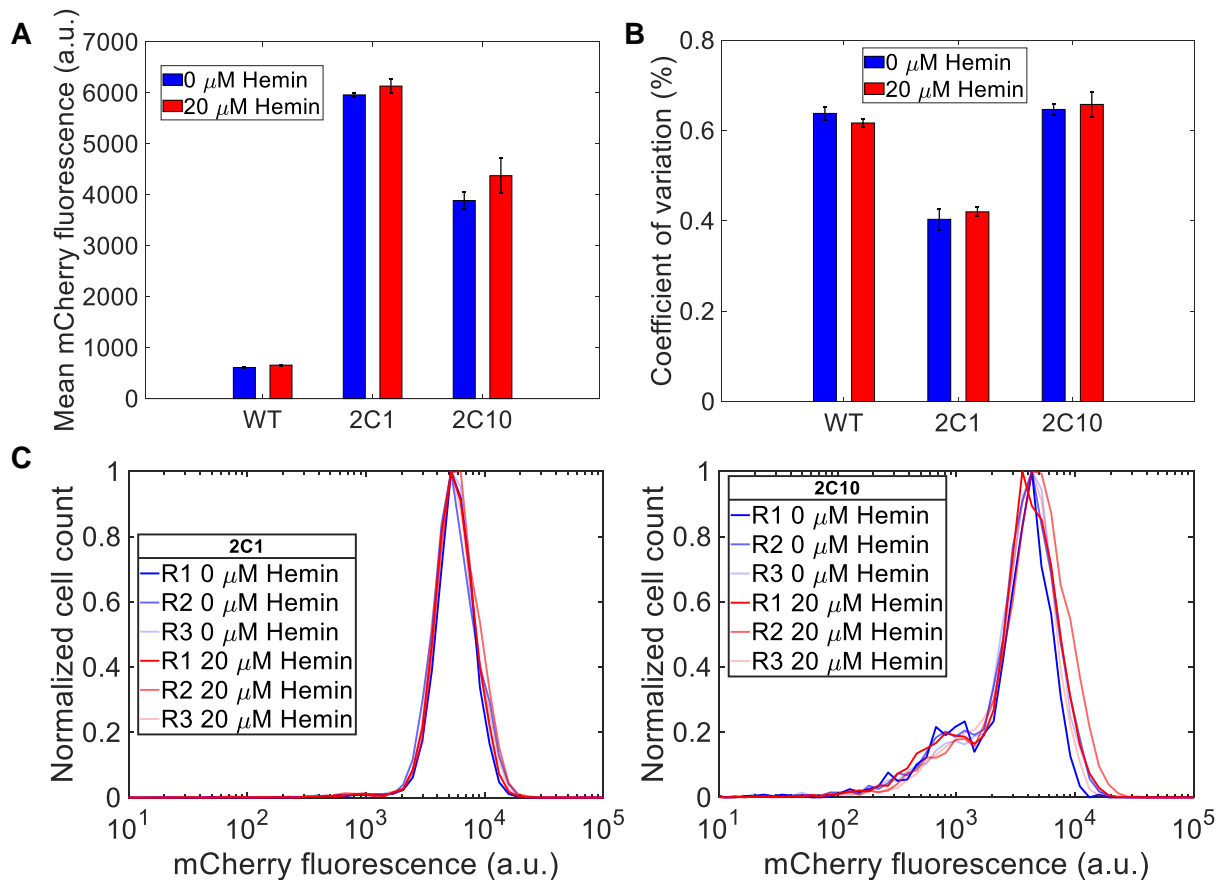


**Figure 52: Hemin treatment over 4 hours for HEK293 BACH1 clones 2C1 and 2C10.**

**(A)** Mean mCherry expression in BACH1 clones and wild-type cells after 4 hours with or

without hemin treatment. **(B)** Gene expression noise (CV). **(C)** Single-cell expression

distributions. The experiments were conducted in triplicate. Error bars represent the

standard error of the mean.

To uncover alterations in fluorescence from hemin at a larger time scale, I again treated the two HEK293 BACH1 reporter clones 2C1 (unimodal) and 2C10 (bimodal) with 20 μM hemin for 48 hours. Flow cytometry indicated the mean mCherry fluorescence for the 2C10 clone significantly increased after 48 hours of hemin treatment (**Figure 53A**; p-value = 0.0219, two-sample t-test). Based on this finding, it is possible that the build-up of mCherry protein may take more time than the hemin-mediated BACH1 protein degradation mechanism[233] and subsequent locus de-repression, leading to a stronger effect at longer time scales. Overall,
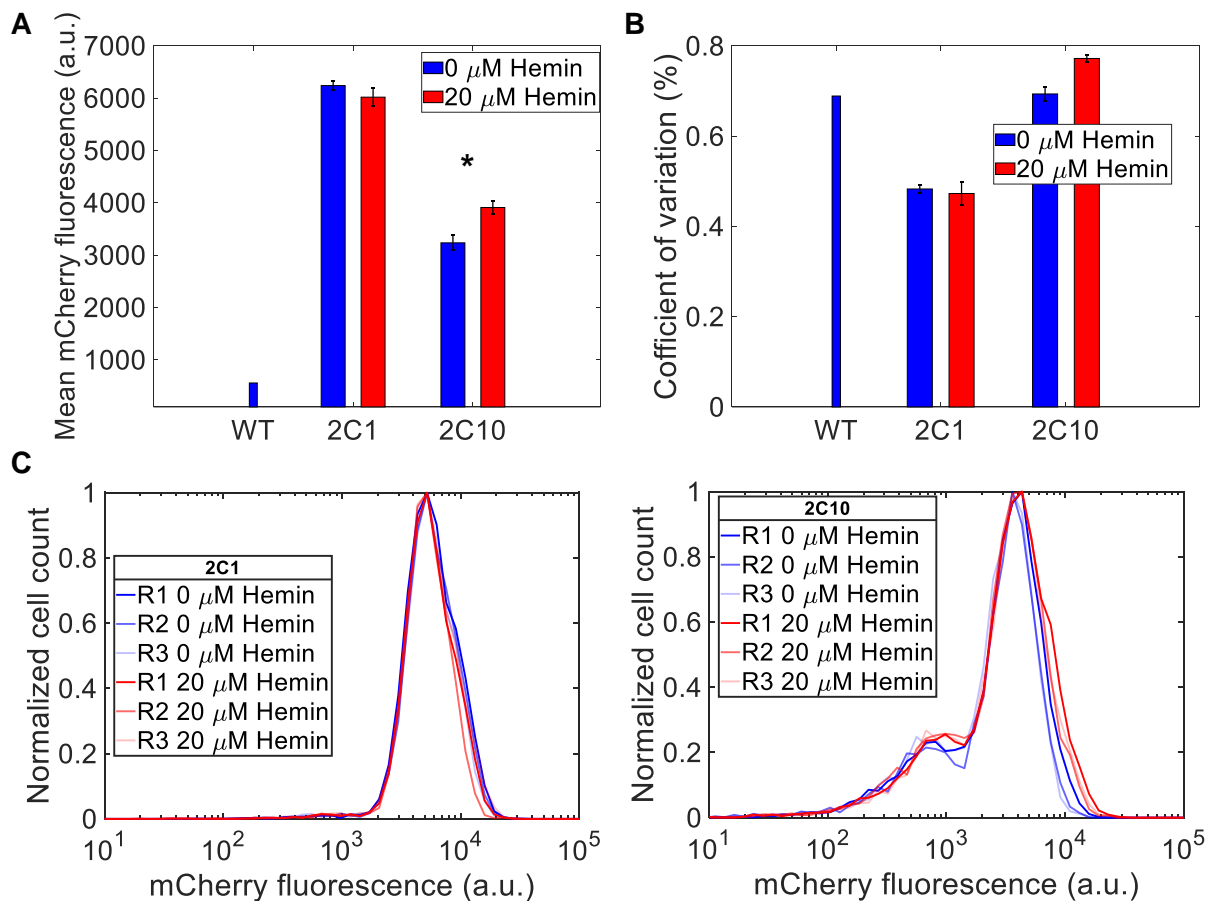


**Figure 53: Hemin treatment over 48 hours suggests increased fluorescence for the bimodal 2C10 HEK293 BACH1 clone.**

**(A)** Mean fluorescence for wild-type HEK293 and two BACH1 clones after 48 hours with or without hemin. **(B)** Gene expression noise (CV). **(C)** Single-cell expression distributions. The experiments were done in triplicate. Error bars indicate the standard error of the mean. *p-value < 0.05 from a two-sample t-test.

hemin increased mCherry fluorescence in the 2C10 clone, which remains consistent with the

expected expression dynamics arising from temporary inhibition of BACH1 self-repression.

### 3.3.6 Over-expression of *pri-let-7g* in the HEK293 BACH1 reporter clones.

To assess the effect of post-transcriptional perturbations on fluorescence from the

BACH1 reporter system, I transiently transfected a vector constitutively expressing the primary

microRNA *pri-let-7g* into the HEK293 BACH1 clones 2C1 and 2C10. The *pri-let-7g* matures into

*let-7g*, a known target for the *BACH1* mRNA 3' untranslated region[159]. I observed no significant

evidence of miRNA-induced decrease in mean expression, though the transfection non-

significantly increased mean 2C10 expression (**Figure 54A**). Gene expression noise also
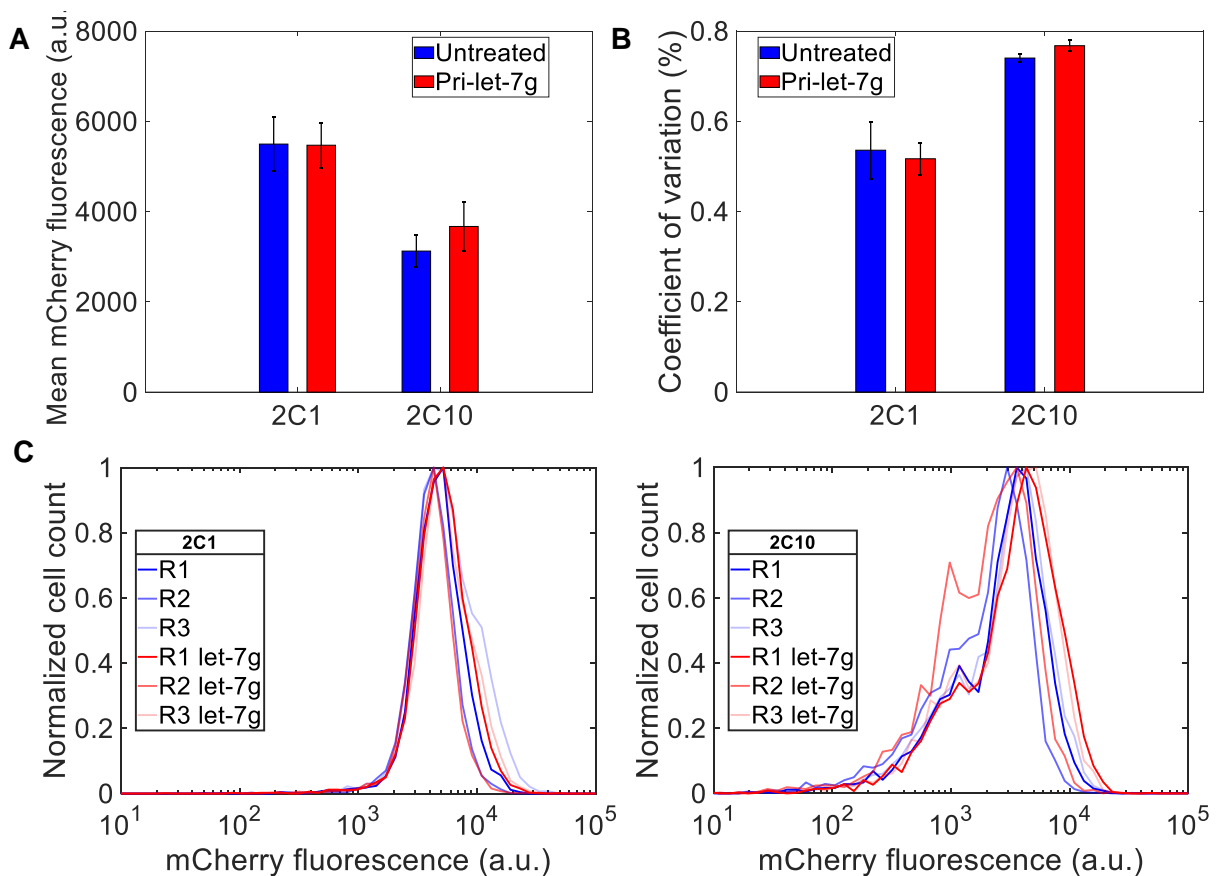


**Figure 54: Over-expression of *pri-let-7g* in HEK293 BACH1 2C1 and 2C10 clones.**

**(A)** Mean fluorescence for two BACH1 clones with or without *pri-let-7g* transfected over 93

hours. **(B)** Gene expression noise (CV). The experiments were conducted in triplicate. Error

bars represent the standard error of the mean. **(C)** Single-cell expression distributions.

remained significantly unchanged, though qualitatively 2C10 displayed slightly increased noise after transfection (**Figure 54B**). Since the *let-7g* seed sequence does not have a perfect complementary binding site in *BACH1* mRNA, the RISC complex should inhibit translation of the mRNA rather than degrading the BACH1 mRNA[235]. Since the BACH1 negative feedback loop depends on the level of BACH1 protein in the cell, a temporary decrease in BACH1 protein levels could de-repress the *BACH1* locus, leading to increased mCherry levels over time. Like hemin, this mechanism could explain the possible (non-significant) increase in 2C10 mean fluorescence. To summarize, *pri-let-7g* over-expression did not significantly alter mCherry fluorescence from the HEK293 BACH1 reporter system.

### 3.3.7 Negative selection as a mechanism to decrease random integration events.

To avoid expanding clones with non-specific integration of the BACH1::T2A::mCherry cassette vector, as seen in the six HEK293 BACH1 reporter clones with the verified tag at the endogenous BACH1 locus (**Figure 47**), I introduced the Herpes Simplex Virus-Thymidine Kinase *(HSV-TK)* negative selection marker outside of the *BACH1* homology arms flanking the reporter tag in a modified donor clone plasmid backbone (**Figure 55**). Doing so should
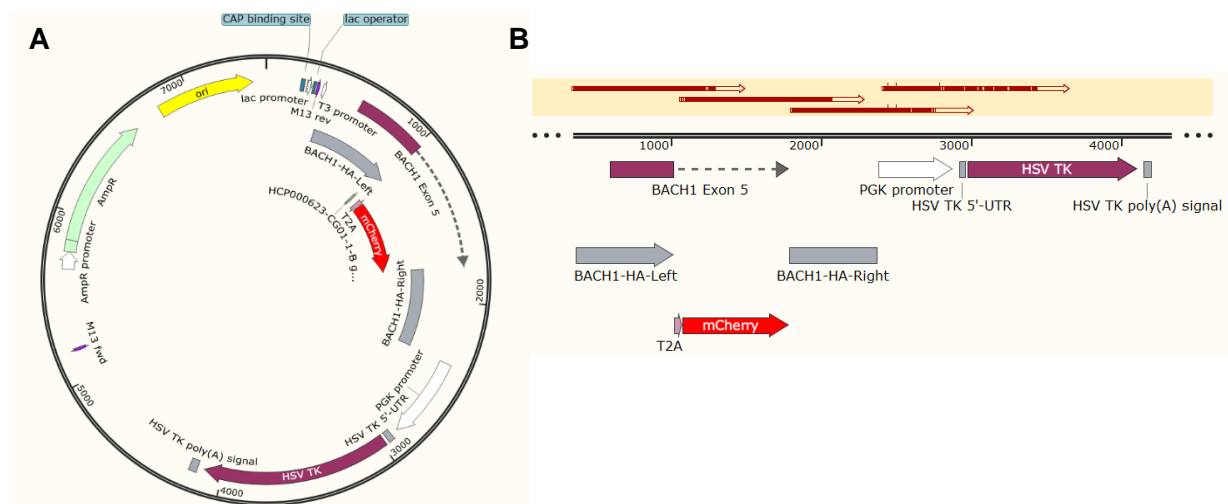


**Figure 55: Introduction of the Herpes Simplex Virus Thymidine Kinase (*HSV-TK*) negative selection marker for avoiding random integration cell enrichment.**

**(A)** Plasmid map and **(B)** sequencing results for the improved BACH1::T2A::mCherry donor clone containing the *HSV-TK* outside the homology arms.

sensitize cells possessing the HSV-TK gene to the drug Ganciclovir, which would occur when the entire backbone randomly integrates. After selection, the remaining cells will either not contain the construct at all or will have a successful integration event at the *BACH1* locus. In summary, I built a BACH1::T2A::mCherry donor clone with a negative selection marker guarding against random integration.

### 3.3.8 Designing the BACH1::T2A::mCherry reporter system in the MDA-MB-231 triple-negative breast cancer cell line.

After successfully constructing the improved donor clone with the *HSV-TK* negative selection marker and the characterization of the HEK293 BACH1 reporter cell lines, it was unclear how *BACH1* expression dynamics could reflect physiological characteristics in a cancer cell line, specifically the highly aggressive triple-negative MDA-MB-231 breast cancer cell line. Creating such a cell line could reveal how *BACH1* expression in living cells switches to pro-metastatic, anti-metastatic, and mixed bistable expression regimes previously uncovered in fixed cells[1]. Additionally, the *BACH1* expression levels in single breast cancer cells may respond to more perturbations than the non-tumorigenic HEK293 cells. Therefore, the development of such a BACH1 reporter clone in a breast cancer cell line will open lines of inquiry that are relevant to the pro-metastatic functions of *BACH1*.

To develop the MDA-MB-231 BACH1 reporter cell line, I transfected the improved donor clone vector with the BACH1-specific guide RNA into triple-negative MDA-MB-231 breast cancer cells. After flow-sorting by low or high mCherry fluorescence, I applied the negative selection agent Ganciclovir to the potential MDA-MB-231 BACH1 clones to select against random integration events. Since I lacked an *HSV-TK* positive control required for a proper Ganciclovir kill curve, I conducted a rudimentary kill curve experiment on wild-type MDA-MB-231 cells. Based on qualitatively slower growth of the untransfected MDA-MB-231 cells using Ganciclovir concentrations of at least 0.5 mM (data not shown), I treated the transfected cells once in exponential growth after flow sorting with a much lower dosage; about 0.03 mM (10 µg/mL) Ganciclovir. However, none of the clones growing in the 96-well plates died off after

Ganciclovir selection, suggesting either the ineffectiveness of that concentration or the lack of any clone harboring randomly integrated donor clone plasmids. Thus, one cannot make a conclusion on Ganciclovir's effectiveness without knowing the lethal dose concentration for MDA-MB-231 cells with the *HSV-TK* gene.

To ensure the archiving of all fluorescent clones, I cryogenically preserved the bulk, polyclonal population, and high- and low-fluorescent FACS-sorted cells. To uncover a MDA-MB-231 clone with a successful knock-in of T2A::mCherry at the *BACH1* locus, I conducted two separate screens with genotyping PCR targeting the region upstream the 5' homology arm and within the 3' downstream homology arm as in HEK293 cells (see section 3.3.1). I evaluated 40 clones in the first round of screening. In the second round, I genotyped 151 clones. Clones varied from having high to low or no mCherry expression (**Figure 56**). Therefore, the clones vary widely in mCherry expression, which was expected from the flow-sorting gating.
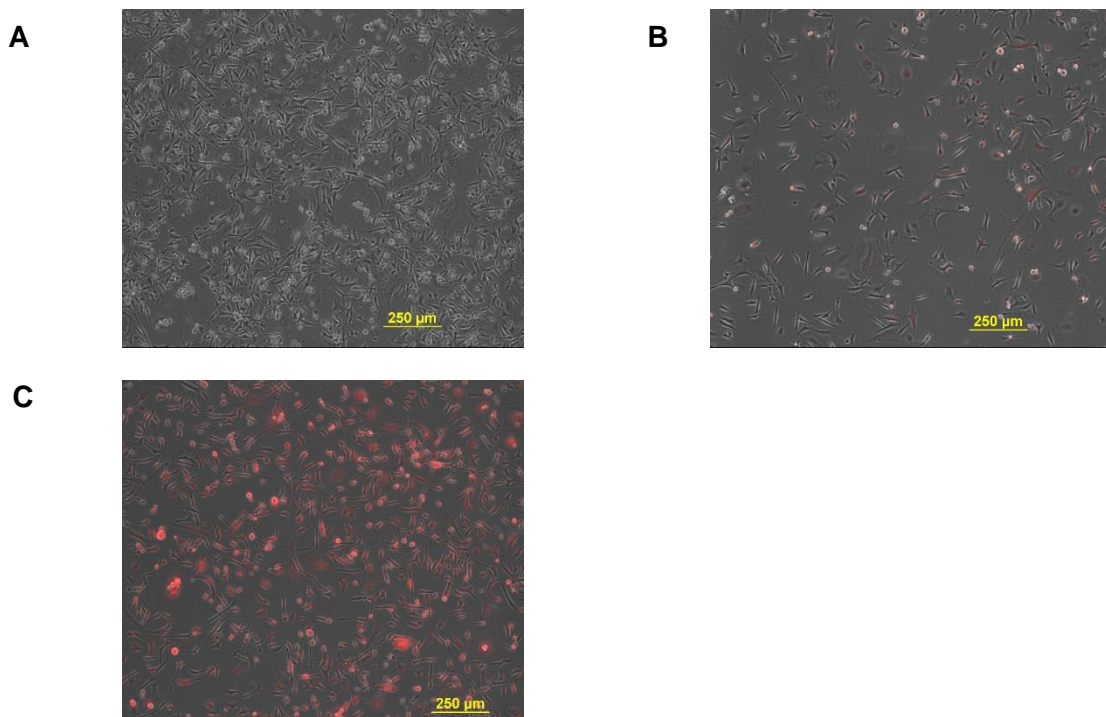


**Figure 56: Fluorescent images from putative MDA-MB-231 BACH1 clones.**

**(A)** Image of the non-fluorescent 2B12 clone cells. **(B)** Image of the mildly fluorescent 1G4 clone cells. **(C)** Image of the highly fluorescent 2E11 clone cells. The look-up tables were kept consistent between images.

### 3.3.9 Screening for MDA-MB-231 BACH1 clones with an optimized direct PCR method.

To reduce the time for detecting whether a particular MDA-MB-231 clone contained the knock-in, I modified a direct (colony) PCR detection method that directly lyses cells before amplification of DNA[230]. As a mammalian version of colony PCR, it can reduce the amount of time for genotype screening after flow-sorting from a month to just a week (**Figure 57**). Sixteen clones genotyped with direct PCR during the optimization stage consisted of 6 clones that did not amplify either the wild-type or knock-in allele. Of the six clones failing to amplify, one clone (1G4) eventually amplified the wild-type *BACH1* allele during a repeat direct PCR. Additionally, purified genomic DNA from a 17th clone (3B8) amplified just the wild-type *BACH1* allele (data not shown). Thus, I confirmed the amplification of the wild-type *BACH1* allele in 12 out of 17 MDA-MB-231 clones, but all clones lacked the correctly integrated mCherry tag.
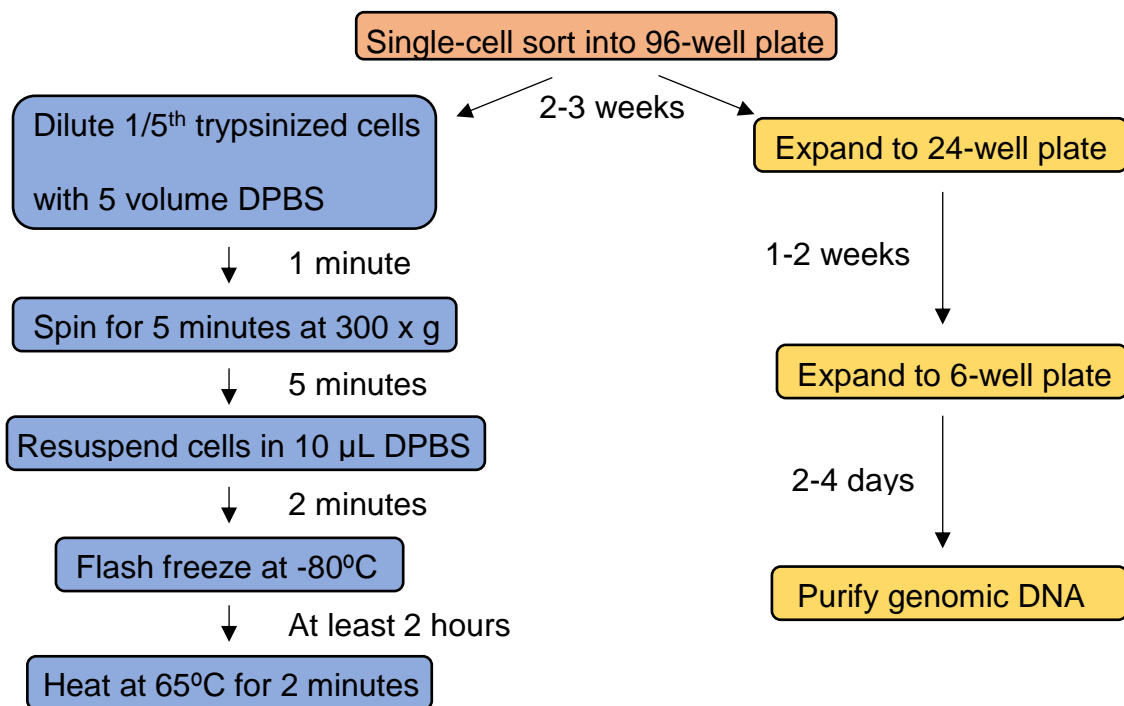


**Figure 57: Experimental workflow of the modified direct PCR method.**

On the right, single cells that are sorted into a 96-well plate are expanded into 24-well plates, then 6-well plates, further expanded in 25 cm$^2$ flasks, and then genomic DNA from each clone is purified. Right before expansion to 24-well plate, the direct PCR method quickly lyses the cell after trypsinization and flash freezing, saving weeks before screening.

**3.3.10 Two rounds of direct PCR screening for potential MDA-MB-231 BACH1 clones.**

To further test the direct PCR methodology in its optimized form, I genotyped the remaining 23 MDA-MB-231 putative BACH1 clones from the first screening round, which derived from clones directly expanded after the original flow-sorting experiment. In all cases, I did not detect a knock-in allele. Overall, the direct PCR methodology found the wild-type *BACH1* allele in 28 out of 40 clones. Moreover, traditional PCR for purified genomic DNA from 2 clones detected just the wild-type allele. It is possible that the integration rate via DNA repair mechanisms in MDA-MB-231 is too low to uncover one clone with the knock-in out of 40 samples (<1/40 odds). In the entire original screen, 30 clones that contained the wild-type *BACH1* allele lacked the fluorescent reporter knock-in (**Table 9**).

After exhausting the clones from the original flow sorting screen, I still had to check the cryo-preserved bulk-transfected MDA-MB-231 cells along with cryo-preserved high- and low-expressing bulk-sorted cells from the original CRISPR-Cas9 transfection experiment. After thawing both unsorted bulk-transfected and high- or low-fluorescing bulk-sorted cells, flow sorting separated cells into four 96-well plates based on mCherry fluorescence. In two plates with cells either previously sorted for high or low fluorescence, four gates further separated cells from low to high fluorescence intensities. This gating scheme may increase the chance of enriching for clones with successful mCherry integration by systematically scanning possible *BACH1* expression levels. The two other plates contained single cells from the remaining bulk-transfected sample, which I flow-sorted based on low or high fluorescence. Overall, the gating may enrich for the range of fluorescence that actual *BACH1* expression resides in.

To determine whether the mCherry tag integrated in the *BACH1* locus for the previously cryo-preserved samples, I applied the optimized direct PCR method for genotyping in a second screen. The second screen evaluated 151 clones. Out of the 151 clones, 4 clones failed to amplify anything. The remaining 147 clones, despite the presence of non-specific bands in some clone PCR products, all amplified the wild-type *BACH1* allele. However, not a single clone contained the knock-in allele. In summary, the second screen failed to verify MDA-MB-

117

231 BACH1 reporter clones. In all, I found the wild-type *BACH1* allele in 177 clones (**Table 9**).

| | Screen 1 | Screen 2 | Total |
|---|---|---|---|
| **MDA-MB-231 clones with wild-type BACH1** | 30 | 147 | 177 |
| **MDA-MB-231 clones with knock-in** | 0 | 0 | 0 |
| **Total MDA-MB-231 clones** | 40 | 151 | 191 |

**Table 9: Summary statistics of MDA-MB-231 BACH1 clone genotyping screens.**

Table includes the number of clones with just the wild-type *BACH1* allele, the number of

clones with the knock-in allele, and the total number of genotyped MDA-MB-231 clones,

including clones without amplification of the *BACH1* locus.

## 3.4 Discussion.

In this chapter, I described the construction and characterization of an endogenous

BACH1 locus reporter system that should co-translate the red fluorescent mCherry protein after

transcription in HEK293 cells. While none of the MDA-MB-231 clones successfully integrated

the T2A::mCherry fluorescent tag at the *BACH1* locus, the direct PCR technique facilitated

rapid screening, useful in future experiments. For the HEK293 BACH1 reporter clones,

inducing a wound increased fluorescence at the wound edge. Additionally, long-term hemin

treatment increased mCherry fluorescent reporter intensities for the bimodal 2C10 population.

**Chapter 4: Discussion and future directions.**

**4.1 Discussion.**

**4.1.1 Gene expression noise can enhance cellular access to a pro-survival state.**

Gene expression noise is a ubiquitous phenomenon in biology ranging from bacteria to humans. The functional aspects of gene expression noise have been investigated in unicellular organisms while studies on higher eukaryotes such as mammalian cells are limited by biological complexities. Mammalian cells do not easily integrate foreign DNA in a site-specific manner. Additionally, the endogenous gene regulatory networks in mammalian cells involve multiple feedback loops with dense, redundant layers of interactions facilitating robustness, especially in drug resistance. Thus, mammalian cells with noise-tunable genetic constructs must share the same genetic background while maintaining constant mean expression. In this manner, I modulated noise with exogenous synthetic gene circuits controlling the drug resistance gene *Puromycin N-acetyltransferase* (*PuroR*) to rigorously study gene expression noise in mammalian drug resistance .

The core hypothesis of my thesis argues that gene expression noise in mammalian cells enhances the probability of reaching a favorable survival fate in drug resistance and metastasis. To this end, I investigated the role of *PuroR* noise independently of mean expression using an exogenous approach in Chinese Hamster Ovary (CHO) cell drug resistance. Moreover, I developed an endogenous *BACH1* expression fluorescent reporter system in live HEK293 cells to observe metastatic state transitions in response to molecular and physical perturbations. In Chapter 2, the exogenous approach allowed tuning the mean expression at similar levels in two circuits with high or low *PuroR* gene expression noise, thereby fixing the mean with distinct noise levels in a decoupled noise point (DNP). Using this tunable system to control the non-native *PuroR* gene should isolate the impact of *PuroR* gene expression from all other cellular characteristics during short time-scales of drug treatment.

Improved drug survival from gene expression noise in mammalian cells under high drug concentrations clearly supports the core hypothesis, even with lower mean expression from the noisy circuit. In low stress levels, cells with high *PuroR* noise tend to have a subpopulation of

cells with very high expression of the gene, past the highest expressing cell in the low-noise population (**Figure 13C**). Consequently, cells with drug resistance in low stress (**Figure 18C,D,F**) that survived drug treatment have higher mean expression than at pre-treatment (**Figure 26B-C**). Furthermore, increasing Puromycin concentration from 35 to 50 µg/mL led to higher mean expression levels for the high-noise cells before drug removal (**Figure 26C-D**). Lastly, mean expression levels tended to jump even higher during the retreatment phase (**Figure 26B-D**), though induced mPF at 35 µg/mL Puromycin proved to be an exemption (**Figure 26C**). Summing it all up, gene expression noise was shown to enhance the opportunity for mammalian cells to escape drug-induced cell death by expressing enough of the drug resistance gene for survival, which is apparent in cellular memory experiments (**Figure 14B**).

In addition to exogenous control of the drug resistance *PuroR* gene, I investigated endogenous expression of the pro-metastatic *BACH1* transcriptional regulator that participates in a double negative feedback loop with the anti-metastatic *RKIP* gene (**Figures 4; 42**). If gene expression noise in *BACH1* can contribute to the survival of cancer cells in a hostile microenvironment after facilitating escape from the primary tumor, noise in the anti-metastatic regulator *RKIP* could also affect survival through bistable state transitions mediated by the network. Under this scenario, high *RKIP* gene expression noise and mean expression may stochastically drive *RKIP* expression in single cancer cells below the ultrasensitive bistable expression threshold, leading to stabilization of the pro-metastatic state with low *RKIP* and high *BACH1* levels in a cellular subpopulation[1]. Thus, the mean expression of *BACH1* and *RKIP* strongly influences whether cells can exit or enter the pro-metastatic state, with high noise facilitating survival and subsequent metastasis in a subpopulation of cells deviating from low *BACH1* and high *RKIP* mean expression despite the mean expression levels favoring the anti-metastatic state. On the other hand, low noise with low *BACH1* and high *RKIP* mean expression levels should stabilize the anti-metastatic state.

A potential tumor suppressor relevant to the hypothesis outside of metastasis is *TP53* (p53), which depends on the gene regulatory network with embedded feedback loops[236,237] as

predicted computationally[236-238]. Phosphorylation by the ATR and ATM kinase pathways activates the p53 protein after single and double stranded DNA breaks, respectively[239]. In both pathways, the MDM2 ubiquitin ligase and the WIP1 phosphatase proteins negatively regulate p53 at the protein level while p53 also transcriptionally activates *MDM2* and *WIP1* expression[240,241]. During ATM kinase transduction, WIP1 protein also dephosphorylates the ATM kinase, which adds an additional layer of negative feedback controlling the *TP53* gene. With both layers of feedback, *TP53* expression can pulse[237]. Thus, double stranded DNA breaks from gamma radiation may increase gene expression noise for the *TP53* tumor suppressor[238]. In this case, high noise in wild-type *TP53* expression may aid survival when single cancer cells express *TP53* below the mean, highlighting another oncogenic phenomenon where the core hypothesis regarding cancer cell survival is relevant to tumor suppressors. It is unclear how the hypothesis relates to mutant *TP53*, which will disrupt the underlying gene regulatory network architecture.

## 4.1.2 Evolution of network noise-mediated mammalian drug resistance depends on endogenous mechanisms outside the circuit.

The synthetic gene circuit, drawing on non-native genetic elements, should exhibit genetic isolation from the rest of the CHO genome. In line with this, two of six low-noise replicates at 35 µg/mL Puromycin developed mutations in the inducible promoter while a third replicate acquired an amino acid deletion in the *hTetR* gene that disrupted repression of the promoter, thereby increasing *PuroR* expression (**Figures 29-31**). Unexpectedly, the three remaining replicates did not contain mutations within the circuit (**Figure 32**) even though all replicates had elevated expression levels without induction (**Figure 26B**). For those three replicates, the induction-independent elevation of *PuroR* expression without circuit mutations suggests CHO cells either acquired mutations outside the circuit or epigenetically modulated *PuroR* expression (**Figures 26B; 32-33**). Since removing Doxycycline inducer did not lower *PuroR* expression in the three circuit-mutation-free replicates, logically the *hTetR* repressor cannot effectively repress the promoter. If the mechanism did not require disrupting *hTetR*

repression, the molecular mechanism of adaptation would depend on induction, leading to lower *PuroR* levels in the absence of Doxycycline.

Hypothetically, at least three processes outside the circuit could affect *hTetR* repression: 1) *hTetR*-specific protein degradation, 2) inhibition of *hTetR* binding to *tetO2* operators, and/or 3) general up-regulation of the FRT locus. The first case requires an endogenous mechanism that degrades the *hTetR* repressor. A stress response program could disrupt proper folding of *hTetR* while somehow leaving the *PuroR* gene functionally intact[242], possibly by not affecting the N-acetylation reaction. In the second case, altered protein-protein, protein-DNA, or protein-Doxycycline interactions may prevent binding of *hTetR* to the promoter. In yeast, synthetic control of efflux pumps increased inducer sensitivity of a *TetR*-based negative feedback circuit[243]. Although the exact mechanism is not clear, it depends on reduced *TetR* regulator activity. In the third case, the transcriptional machinery may strongly activate expression at the FRT circuit to overcome *hTetR*-mediated repression, either through chromatin-mediated or steric mechanisms. Since the Initiator element in the low-noise circuit is flanked by one *tetO2* site on each side, the RNA Polymerase II pre-initiation complex could possibly outcompete steric inhibition from *hTetR-tetO2* binding, leading to higher *PuroR* expression. Independent of extra-circuit processes, insufficient 2A cleavage from *EGFP* and *PuroR* may lower the degradation rate of the fused protein or completely inefficient cleavage, creating a *hTetR*::P2A::*EGFP*::T2A::*PuroR* fusion protein, may sterically prevent the regulator from binding to DNA.

For evolved cells with the high-noise circuit, induction with Doxycycline enhanced *PuroR* expression above uninduced cells, which themselves displayed elevated expression above the pre-treatment mean expression level. Chromatin-mediated activation of the genomic FRT site may explain the increased expression, where induced *rtTA* and epigenetic markers synergistically enhance transcriptional activation above the pre-treatment mean. For evolved mPF cells, it is not clear how enhanced activity of efflux pumps that should reduce intra-cellular

Doxycycline and Puromycin concentrations would lead to higher activity of the positive feedback circuit in a Doxycycline-dependent manner.

## 4.1.3 Impact of gene expression noise for various mechanisms of chemotherapy treatment.

We computationally predicted drug-tolerant persister cells must develop genetic resistance to explain the experimental results (**Figures 23-24**). The persister state may depend on histone demethylases or *GPX4*, which can be disrupted by histone deacetylase complex inhibitors (HDACi) and molecular inhibitors of *GPX4*[104,105]. If feasible, the findings in Chapter 2 support the design of a therapeutic regime that lowers gene expression noise of genes relevant to cell survival in early chemotherapy treatment resistance prior to increasing the dose regime to the highest tolerable level, thereby wiping out the low-noise, otherwise drug resistance population. Consistent with the core hypothesis, decreasing the opportunity for mammalian cells entering a survival state could depend on reduced gene expression noise. This is the case with the *PuroR* gene in Puromycin resistance.

Expanding the scope of the findings to various chemotherapeutic agents will depend on their mechanism of action. *PuroR*-based resistance, since the mechanism initially depends on a single gene, could mimic resistance from molecularly targeted chemotherapy agents, such as *EGFR* mutation-mediated resistance to Gefitinib in non-small cell lung cancer (NSCLC)[244] or *BCR-ABL* kinase domain mutations conferring resistance to Dasatinib in acute lymphoblastic leukemia (ALL)[245]. Gefitinib resistance can also rely on enhanced DNA methylation[246] and altered signaling pathway activity from genes such as *HER2*[247], which is consistent with extra-circuit mechanisms predicted during the evolution of Puromycin resistance. Additionally, anti-tumor activity from EGFR inhibition can be overcome through stochastic transitions into persister cell states[104,107]. Thus, targeted therapy with *EGFR* is one example where our findings have potential relevance.

Genotoxic chemotherapy agents such as Cisplatin may accelerate genetic mutations that subsequently influence resistance. However, the drugs damage DNA in general without

targeting a specific gene[248], making gene expression noise for a single gene uninfluential. Theoretically, cell-cycle disruption from genotoxic agents may increase global gene expression noise, potentially affecting and mediated by cell cycle regulators[249]. Chemotherapeutic agents that disrupt microtubules may become ineffective for killing cancer cells when expression of *βIII-tubulin* is high[250], which is a possibility with high *βIII-tubulin* expression noise. For inhibitors of DNA topoisomerase[251], a subpopulation of cancer cells with high *TOP2A* mean expression and noise may maintain drug resistance by facilitating low *TOP2A* levels in single cells deviating from the mean, but it is unclear which transcription factors, mainly tumor suppressors[251,252], would down-regulate the gene without affecting tumor progression. Overall, the postulated effect of gene expression noise in drug resistance for different classes of chemotherapy agents will depend on the mechanism of action and could be confounded by interactions with tumor suppressors.

## 4.1.4 HEK293 cell BACH1 expression reporter fluorescence responds to one chemical and physical perturbation.

In Chapter 3, I report the development of endogenous *BACH1* gene expression reporters in six clonal populations of the human cell line HEK293. The *BACH1* locus in six clones demonstrated heterozygotic CRISPR-Cas9 homology directed repair-mediated knock-in of an T2A::mCherry reporter prior to the *BACH1* stop codon. However, the donor clone with the promoter-less *BACH1* exon 5::T2A::mCherry element also randomly integrated into the HEK293 genome. The triple-negative breast cancer cell line MDA-MB-231 failed to successfully integrate the T2A::mCherry tag at the *BACH1* locus in almost 200 genotyped clones. The methodology for direct PCR genotyping readily increased the through-put of screening, making the isolation of very rare integration events easier as the number of screened clones scales upward. Importantly, the HEK293 reporter system clones proceeded to perturbation experiments to address the core hypothesis.

Mean mCherry expression levels only varied during long-term hemin treatment and possibly wounding, though wound healing fluorescence statistics were not statistically

quantified. During hemin treatment, an increase in mCherry fluorescence was not unexpected because hemin-mediated BACH1 protein degradation should temporally de-repress the *BACH1* promoter based on negative feedback (**Figure 42**), thereby increasing mCherry protein levels. Thus, the decoupling of degradation rates for mCherry and BACH1 protein along with the de-repression of the *BACH1* negative feedback loop may lead to lower BACH1 and higher mCherry protein levels. In addition to chemical perturbations, I characterized the fluorescence of the HEK293 BACH1 clones after physically inducing wound-mediated migration. Wounding tended to non-uniformly increase fluorescence at the wound edge (**Figures 50-51**). After migration, mCherry fluorescence for cells enclosing the gap qualitatively decreased (**Figure 50**, 24h-96h). At shorter-time scales (1 hour), the cells at the wound edge displayed increased fluorescence (**Figure 51**), which suggests alterations in protein translation rates may drive the wound-induced expression dynamics since protein synthesis is in general faster than transcription in mammalian cells[1,253].

### 4.1.5 mCherry fluorescence may not reflect *BACH1* expression.

It is possible that mCherry fluorescence arose from a randomly integrated donor clone. Although promoter-less, the construct could in principle integrate at a site that promotes the transcription of the *BACH1* homology arms and the mCherry element. Future qRT-PCR experiments will confirm correlation between *BACH1* mRNA levels and red fluorescence. Nevertheless, hemin treatment revealed the most promising evidence for fluorescence negatively correlating with *BACH1* expression given that the BACH1 protein degradation rate increased, potentially leading to anti-correlated BACH1 and mCherry protein levels.

### 4.1.6 Inconclusive evidence that *BACH1* expression transitions between metastatic states in HEK293 cells.

It is unclear that the fluorescence, if correlated with *BACH1* expression, reflects transitions into or out of a pro-metastatic state in the six HEK293 clones. The fluorescence changes during the two perturbation experiments and the failure of the remaining chemical perturbation experiments may instead reflect an inability of cells to push past a steep

Waddington valley (expression threshold) for a transition to occur. However, the hemin treatment experiments demonstrated a significant increase in mean fluorescence for the bimodal 2C10 clone, but not the unimodal 2C1 clone. Does this reflect a shift in the make-up of the bimodal subpopulation composition in 2C10? Repeat experiments may reveal a similar shift in the bimodal distribution (**Figure 53C**), which hints of cells in the low-expressing anti-metastatic subpopulation potentially transitioning into a high-expressing pro-metastatic state. If so, the five HEK293 clones with unimodal distributions may reside in the pro-metastatic state.

Except for long-term hemin treatment, mean fluorescence did not significantly change in the other chemical perturbation experiments. In these experiments, the very low red fluorescence (**Figure 48**) may indeed change, but the variation in fluorescence during a perturbation could be too small to notice. To address this issue, one could amplify the low expression levels with a novel synthetic gene circuit or measure fluorescence with a more sensitive technique.

Although gene expression noise did not change significantly in any chemical perturbation experiment, the standard error of the mean for the coefficient of variation (gene expression noise) for clone 2C10 during long-term hemin treatment did not overlap with the untreated control (**Figure 53B**), which requires further investigation. Additionally, single HEK293 cells at the frontier wound edge during the induction of wound-mediated migration qualitatively varied in gene expression based on fluorescence, which could represent spatially localized gene expression noise. In line with these findings, previous research[1] support a role for *BACH1* gene expression noise in pro-migratory phenotypes. Overall, the core hypothesis could not be addressed with the endogenous BACH1 reporter HEK293 cell lines. To continue this line of inquiry, a future approach could decouple endogenous *BACH1* gene expression noise from the mean during chemical and physical perturbations with noise-modulating molecules[184].

**4.2 Future studies.**

**4.2.1 Single-cell RNA-seq on evolved CHO mPF and mNF populations.**

How the cell recruits other endogenous genes to implement adaptation to Puromycin remains unaddressed. To gather as much information as possible, single-cell RNA-seq for evolved and ancestral cells could reveal the signaling and gene regulatory pathways differentially expressed in the evolved cells, possibly resulting from mutations within that pathway. Sequencing single cells will also preserve heterogeneity persisting after adaptation. In fact, enhanced gene expression noise in endogenous pathways associated with resistance could elegantly complement the exogenous approach for controlling *PuroR* mean and gene expression noise. Expected up-regulated pathways detected by RNA-seq could include WNT[115], MDR1[93], NF-κB[254], RAS-MAPK[255], NOTCH[256], PI3K[257], and NRF2[258]. Importantly, activation of the RAS-MAPK pathway could occur through reduced *RKIP* expression[259], which has direct relevance to the *BACH1* and *RKIP* pathway studied in Chapter 3.

Single cell RNA-seq has many technical drawbacks that makes subsequent analysis difficult[260,261]. A lack of reads for a gene may actually indicate RNA dropout arising from a lack of cells or partial lysis during a cell isolation event, which is hard to distinguish from true absence of expression[260]. RNA dropout can also lead to under-estimation of RNA levels for genes after insufficient RNA extraction. It is also challenging to distinguish gene expression noise from technical variation, which RNA spike-ins can help estimate for removal during analysis[262]. Coverage of the 3' end can introduce read bias when quantifying mean expression, which full-length RNA sequencing technologies such as STRT-Seq can better limit[263]. However, the cost of both 3' end and especially full-length single cell sequencing technologies can prohibit its use depending on how many cells to sequence, the resolution of gene coverage, and required statistical power[264]. Overall, one should cautiously interpret expression data from single-cell RNA-sequencing. In case the prohibitions are too high, bulk RNA-seq could act as an alternative approach that focuses attention on population-level differentially expressed genes.

**4.2.2 Molecular disruption of the drug-tolerant persister state.**

The persister cells predicted from the model could become re-sensitized to drug upon inhibition of histone deacetylase complexes (HDACs)[104]. Additionally, inhibition of the lipid hydroperoxidase *GPX4* may promote a drug-sensitive state in persister cells[105]. Future experiments on the evolved CHO populations to disrupt the persister state include treatment with trichostatin A (TSA), which inhibits class I and II HDACs. *GPX4* inhibitors associated with sensitizing persister cells will require screening to uncover their effect in CHO cells without prior knowledge or screening. Nevertheless, we can attempt persister state disruption by *GPX4* inhibition through RSL3 and ML210[105].

**4.2.3 Efflux dye assays for evolved drug resistance populations.**

If multi-drug resistance contributes to the evolution of drug resistance in the CHO populations surviving Puromycin treatment, one should expect increased efflux of intra-cellular drug, conferring resistance[90]. To test whether *MDR1* or another efflux pump contributes to resistance, an efflux dye assay can determine differential pumping of drug between cell populations. The Calcein Assay® fluorescent dye is activated within a cell but does not fluoresce after efflux pumping. Thus, the evolved cells with the dye may display lower fluorescence than ancestral cells, implicating drug pumping as an extra-circuit mechanism of adaptation, which may involve the *MDR1* gene. No differences in cellular fluorescence would rule out drug pump-based mechanisms of resistance.

**4.2.4 Sequencing the HEK293 BACH1 reporter clone 1B11 knock-in allele.**

Regarding the small size of the 1B11 knock-in allele (**Figure 45A**), future work should determine whether the C-terminus of mCherry carries a deletion. A deletion at the end of mCherry could explain why the size of the 1B11 junction PCR product was consistent with integration because the reverse junction PCR primer binds upstream of the mCherry C-terminus (**Figure 43**). Sequencing must address the issue before interpreting data from this clone.

**4.2.5 Flow-sorting each peak in the bimodal HEK293 BACH1 2C10 clone to assess cellular memory.**

The characterization of endogenous expression revealed a bimodal distribution for the HEK293 2C10 clone, which requires explanation. There are two possibilities for the bimodal distribution. The population may consist of two genetically distinct subpopulations, arising from two cells flow-sorted into the same well. On the other hand, isogenic cells could exhibit bimodality in a mixed bistable regime[1]. To rule out one scenario over the other, a flow sorting experiment enriching for each subpopulation could lead to either no change in fluorescence over time if the population is polyclonal, or the reestablishment of the original distribution if the bimodality arises from bistability. If the clone is in a bistable regime, then the hemin treatment data from 2C10 could provide evidence of state transitions, addressing the core hypothesis.

## 4.2.6 Repeating the long-term hemin treatment experiment with more samples.

The trend of higher gene expression noise for the 2C10 clone after 2 days of hemin treatment hints at a difference, but triplicates lack the statistical power to test this for either non-parametric tests or coefficient of variation equivalence testing[265]. Does gene expression noise also change during treatment with hemin? Repeat hemin treatment experiments with more replicates of the HEK293 clones can resolve this question, strengthening a functional verification interpretation from hemin treatment data.

## 4.2.7 Parallel qRT-PCR experiments to verify whether BACH1 mRNA levels correlate with mCherry fluorescence.

Lastly, future fluorescence measurement experiments absolutely require corresponding qRT-PCR experiments to verify whether mCherry fluorescence changes in conjunction with *BACH1* mRNA levels. While experiments using CRISPRa on the *BACH1* locus showed ~1.6-fold increase in qPCR *BACH1* mRNA levels (data not shown), we could not properly apply image processing tools to rigorously quantify changes in fluorescence from microscopy. In conclusion, future experiments and analyses will address these outstanding questions, which will serve to independently confirm the findings in Chapter 3.

In an ending note, renowned physicist Max Plank once said,

*"The whole strenuous intellectual work of an industrious research worker would appear, after all, in vain and hopeless, if he were not occasionally through some striking facts to find that he had, at the end of all his criss-cross journeys, at last accomplished at least one step which was conclusively nearer the truth."*

Every journey has an ending, as in my educational progression to a Doctoral degree, but every journey's end is another journey's beginning. I look forward to my future as a scientist and the many successes ahead.

## Bibliography

1   Lee, J., Lee, J., Farquhar, K. S., Yun, J., Frankenberger, C. a., Bevilacqua, E., Yeung, K., Kim, E.-J., Balázsi, G. & Rosner, M. R. Network of mutually repressive metastasis regulators can promote cell heterogeneity and metastatic transitions. *P Natl Acad Sci USA* **111**, E364-E373, doi:10.1073/pnas.1304840111 (2014).

2   Lee, J. S., Kallehauge, T. B., Pedersen, L. E. & Kildegaard, H. F. Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci Rep-Uk* **5**, 8572, doi:10.1038/srep08572 (2015).

3   Yano, K., Gale, D., Massberg, S., Cheruvu, P. K., Monahan-Earley, R., Morgan, E. S., Haig, D., von Andrian, U. H., Dvorak, A. M. & Aird, W. C. Phenotypic heterogeneity is an evolutionarily conserved feature of the endothelium. *Blood* **109**, 613-615, doi:10.1182/blood-2006-05-026401 (2007).

4   Pabst, R. Plasticity and heterogeneity of lymphoid organs what are the criteria to call a lymphoid organ primary, secondary or tertiary? *Immunol Lett* **112**, 1-8, doi:10.1016/j.imlet.2007.06.009 (2007).

5   Lawton, J. H. Biological Diversity - the Coexistence of Species on Changing Landscapes - Huston,Ma. *Nature* **373**, 32-32, doi:DOI 10.1038/373032a0 (1995).

6   Piersma, T. & Drent, J. Phenotypic flexibility and the evolution of organismal design. *Trends Ecol Evol* **18**, 228-233, doi:10.1016/S0169-5347(03)00036-3 (2003).

7   Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., Redon, R., Bird, C. P., de Grassi, A., Lee, C., Tyler-Smith, C., Carter, N., Scherer, S. W., Tavare, S., Deloukas, P., Hurles, M. E. & Dermitzakis, E. T. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848-853, doi:10.1126/science.1136678 (2007).

8   Laughney, A. M., Elizalde, S., Genovese, G. & Bakhoum, S. F. Dynamics of Tumor Heterogeneity Derived from Clonal Karyotypic Evolution. *Cell Rep* **12**, 809-820, doi:10.1016/j.celrep.2015.06.065 (2015).

9    Gill, D. E., Chao, L., Perkins, S. L. & Wolf, J. B. Genetic Mosaicism in Plants and Clonal Animals. *Annu Rev Ecol Syst* **26**, 423-444, doi:DOI 10.1146/annurev.es.26.110195.002231 (1995).

10   Scheiner, S. M. Genetics and Evolution of Phenotypic Plasticity. *Annu Rev Ecol Syst* **24**, 35-68, doi:DOI 10.1146/annurev.es.24.110193.000343 (1993).

11   Draghi, J. A. & Whitlock, M. C. Phenotypic Plasticity Facilitates Mutational Variance, Genetic Variance, and Evolvability Along the Major Axis of Environmental Variation. *Evolution* **66**, 2891-2902, doi:10.1111/j.1558-5646.2012.01649.x (2012).

12   Bateson, P. & Gluckman, P. Plasticity and robustness in development and evolution. *Int J Epidemiol* **41**, 219-223, doi:10.1093/ije/dyr240 (2012).

13   Waddington, C. H. *The strategy of the genes.* (Routledge, 2014).

14   Felsenfeld, G. A Brief History of Epigenetics. *Csh Perspect Biol* **6**, doi:ARTN a018200 10.1101/cshperspect.a018200 (2014).

15   Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: A landscape takes shape. *Cell* **128**, 635-638, doi:10.1016/j.cell.2007.02.006 (2007).

16   Bird, A. Perceptions of epigenetics. *Nature* **447**, 396-398, doi:10.1038/nature05913 (2007).

17   Surani, M., Barton, S. C. & Norris, M. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**, 548 (1984).

18   Heard, E. Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. *Curr Opin Genet Dev* **15**, 482-489, doi:10.1016/j.gde.2005.08.009 (2005).

19   Stam, M. Paramutation: A Heritable Change in Gene Expression by Allelic Interactions In Trans. *Mol Plant* **2**, 578-588, doi:10.1093/mp/ssp020 (2009).

20   Elgin, S. C. & Reuter, G. Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. *Cold Spring Harb Perspect Biol* **5**, a017780, doi:10.1101/cshperspect.a017780 (2013).

21    Richards, E. J. Inherited epigenetic variation--revisiting soft inheritance. *Nat Rev Genet* **7**, 395-401, doi:10.1038/nrg1834 (2006).

22    Teschendorff, A. E., Jones, A., Fiegl, H., Sargent, A., Zhuang, J. J., Kitchener, H. C. & Widschwendter, M. Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation. *Genome Med* **4**, doi:ARTN 24 10.1186/gm323 (2012).

23    Balázsi, G., van Oudenaarden, A. & Collins, J. J. Cellular decision making and biological noise: from microbes to mammals. *Cell* **144**, 910–925, doi:10.1016/j.cell.2011.01.030 (2011).

24    Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183-1186, doi:10.1126/science.1070919 (2002).

25    Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N. & Alon, U. Variability and memory of protein levels in human cells. *Nature* **444**, 643-646, doi:10.1038/nature05316 (2006).

26    Charlebois, D. A., Abdennur, N. & Kaern, M. Gene expression noise facilitates adaptation and drug resistance independently of mutation. *Physical Review Letters* **107**, doi:10.1103/PhysRevLett.107.218101 (2011).

27    Beach, R. R., Ricci-Tam, C., Brennan, C. M., Moomau, C. A., Hsu, P. H., Hua, B., Silberman, R. E., Springer, M. & Amon, A. Aneuploidy Causes Non-genetic Individuality. *Cell* **169**, 229-242, doi:10.1016/j.cell.2017.03.021 (2017).

28    Huh, D. & Paulsson, J. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nat Genet* **43**, 95-U32, doi:10.1038/ng.729 (2011).

29    Hsieh, Y. Y., Hung, P. H. & Leu, J. Y. Hsp90 Regulates Nongenetic Variation in Response to Environmental Stress. *Mol Cell* **50**, 82-92, doi:10.1016/j.molcel.2013.01.026 (2013).

30    Levy, S. F., Ziv, N. & Siegal, M. L. Bet Hedging in Yeast by Heterogeneous, Age-Correlated Expression of a Stress Protectant. *Plos Biol* **10**, doi:ARTN e1001325 10.1371/journal.pbio.1001325 (2012).

31    Petrie, K. L., Palmer, N. D., Johnson, D. T., Medina, S. J., Yan, S. J., Li, V., Burmeister, A. R. & Meyer, J. R. Destabilizing mutations encode nongenetic variation that drives evolutionary innovation. *Science* **359**, 1542-1545, doi:10.1126/science.aar1954 (2018).

32    Bódi, Z., Farkas, Z., Nevozhay, D., Kalapis, D., Lázár, V., Csörgő, B., Nyerges, Á., Szamecz, B., Fekete, G., Papp, B., Araújo, H., Oliveira, J. L., Moura, G., Santos, M. A. S., Jr, T. S., Balázsi, G. & Pál, C. Phenotypic heterogeneity promotes adaptive evolution. *Plos Biol* **15**, e2000644, doi:10.1371/journal.pbio.2000644 (2017).

33    Frank, S. a. & Rosner, M. R. Nonheritable cellular variability accelerates the evolutionary processes of cancer. *Plos Biol* **10**, e1001296, doi:10.1371/journal.pbio.1001296 (2012).

34    Wills, Q. F., Livak, K. J., Tipping, A. J., Enver, T., Goldson, A. J., Sexton, D. W. & Holmes, C. Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. *Nat Biotechnol* **31**, 748-+, doi:10.1038/nbt.2642 (2013).

35    Novick, A. & Weiner, M. Enzyme Induction as an All-or-None Phenomenon. *Proc Natl Acad Sci U S A* **43**, 553-566 (1957).

36    McAdams, H. H. & Arkin, A. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A* **94**, 814-819 (1997).

37    Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical-Reactions. *J Phys Chem-Us* **81**, 2340-2361, doi:DOI 10.1021/j100540a008 (1977).

38    Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226, doi:10.1016/j.cell.2008.09.050 (2008).

39    Peccoud, J. & Ycart, B. Markovian Modeling of Gene-Product Synthesis. *Theor Popul Biol* **48**, 222-234, doi:DOI 10.1006/tpbi.1995.1027 (1995).

40    Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025-1036, doi:10.1016/j.cell.2005.09.031 (2005).

41    Singh, A., Razooky, B., Cox, C. D., Simpson, M. L. & Weinberger, L. S. Transcriptional Bursting from the HIV-1 Promoter Is a Significant Source of Stochastic Noise in HIV-1 Gene Expression. *Biophys J* **98**, L32-L34, doi:10.1016/j.bpj.2010.03.001 (2010).

42    Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nature reviews. Genetics* **6**, 451–464, doi:10.1038/nrg1615 (2005).

43    Blake, W. J., Balázsi, G., Kohanski, M. a., Isaacs, F. J., Murphy, K. F., Kuang, Y., Cantor, C. R., Walt, D. R. & Collins, J. J. Phenotypic consequences of promoter-mediated transcriptional noise. *Mol Cell* **24**, 853–865, doi:10.1016/j.molcel.2006.11.003 (2006).

44    Neildez-Nguyen, T. M., Parisot, A., Vignal, C., Rameau, P., Stockholm, D., Picot, J., Allo, V., Le Bec, C., Laplace, C. & Paldi, A. Epigenetic gene expression noise and phenotypic diversification of clonal cell populations. *Differentiation* **76**, 33-40, doi:10.1111/j.1432-0436.2007.00219.x (2008).

45    Paldi, A. Stochastic gene expression during cell differentiation: order from disorder? *Cell Mol Life Sci* **60**, 1775-1778, doi:10.1007/s00018-003-23147-z (2003).

46    Blake, W. J., M, K. A., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633-637, doi:10.1038/nature01546 (2003).

47    Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166**, 358-368, doi:10.1016/j.cell.2016.05.025 (2016).

48    Ravarani, C. N. J., Chalancon, G., Breker, M., de Groot, N. S. & Babu, M. M. Affinity and competition for TBP are molecular determinants of gene expression noise. *Nat Commun* **7**, doi:ARTN 10417 10.1038/ncomms10417 (2016).

49    Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *P Natl Acad Sci USA* **98**, 8614-8619, doi:DOI 10.1073/pnas.151588598 (2001).

50    Padovan-Merhar, O., Nair, G. P., Biaesch, A. G., Mayer, A., Scarfone, S., Foley, S. W., Wu, A. R., Churchman, L. S., Singh, A. & Raj, A. Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number through Independent Global

Transcriptional Mechanisms. *Mol Cell* **58**, 339-352, doi:10.1016/j.molcel.2015.03.005 (2015).

51    Benner, S. A. & Sismour, A. M. Synthetic biology. *Nature reviews. Genetics* **6**, 533-543, doi:10.1038/nrg1637 (2005).

52    Slusarczyk, A. L., Lin, A. & Weiss, R. Foundations for the design and implementation of synthetic genetic circuits. *Nature reviews. Genetics* **13**, 406-420, doi:10.1038/nrg3227 (2012).

53    Nielsen, J. & Keasling, J. D. Engineering Cellular Metabolism. *Cell* **164**, 1185-1197, doi:10.1016/j.cell.2016.02.004 (2016).

54    Dymond, J. S., Richardson, S. M., Coombes, C. E., Babatz, T., Muller, H., Annaluru, N., Blake, W. J., Schwerzmann, J. W., Dai, J., Lindstrom, D. L., Boeke, A. C., Gottschling, D. E., Chandrasegaran, S., Bader, J. S. & Boeke, J. D. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**, 471-476, doi:10.1038/nature10403 (2011).

55    Hutchison, C. A., 3rd, Chuang, R. Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., Gill, J., Kannan, K., Karas, B. J., Ma, L., Pelletier, J. F., Qi, Z. Q., Richter, R. A., Strychalski, E. A., Sun, L., Suzuki, Y., Tsvetanova, B., Wise, K. S., Smith, H. O., Glass, J. I., Merryman, C., Gibson, D. G. & Venter, J. C. Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253, doi:10.1126/science.aad6253 (2016).

56    Arkin, A. P. A wise consistency: engineering biology for conformity, reliability, predictability. *Curr Opin Chem Biol* **17**, 893-901, doi:10.1016/j.cbpa.2013.09.012 (2013).

57    Galdzicki, M., Clancy, K. P., Oberortner, E., Pocock, M., Quinn, J. Y., Rodriguez, C. A., Roehner, N., Wilson, M. L., Adam, L., Anderson, J. C., Bartley, B. A., Beal, J., Chandran, D., Chen, J., Densmore, D., Endy, D., Grunberg, R., Hallinan, J., Hillson, N. J., Johnson, J. D., Kuchinsky, A., Lux, M., Misirli, G., Peccoud, J., Plahar, H. A., Sirin, E., Stan, G. B., Villalobos, A., Wipat, A., Gennari, J. H., Myers, C. J. & Sauro, H. M. The Synthetic

Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nat Biotechnol* **32**, 545-550, doi:10.1038/nbt.2891 (2014).

58    Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339-342, doi:10.1038/35002131 (2000).

59    Isaacs, F. J., Dwyer, D. J., Ding, C., Pervouchine, D. D., Cantor, C. R. & Collins, J. J. Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol* **22**, 841-847, doi:10.1038/nbt986 (2004).

60    Win, M. N. & Smolke, C. D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci U S A* **104**, 14283-14288, doi:10.1073/pnas.0703961104 (2007).

61    Lim, W. A. Designing customized cell signalling circuits. *Nat Rev Mol Cell Biol* **11**, 393-403, doi:10.1038/nrm2904 (2010).

62    Alon, U. Network motifs: theory and experimental approaches. *Nature reviews. Genetics* **8**, 450-461, doi:10.1038/nrg2102 (2007).

63    Rosenfeld, N., Elowitz, M. B. & Alon, U. Negative autoregulation speeds the response times of transcription networks. *J Mol Biol* **323**, 785-793 (2002).

64    Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590-593, doi:10.1038/35014651 (2000).

65    Maeda, Y. T. & Sano, M. Regulatory dynamics of synthetic gene networks with positive feedback. *J Mol Biol* **359**, 1107-1124, doi:10.1016/j.jmb.2006.03.064 (2006).

66    Becskei, A., Séraphin, B. & Serrano, L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *The EMBO journal* **20**, 2528–2535, doi:10.1093/emboj/20.10.2528 (2001).

67    Nevozhay, D., Adams, R. M., Van Itallie, E., Bennett, M. R. & Balázsi, G. Mapping the environmental fitness landscape of a synthetic gene circuit. *PLOS Computational Biology* **8**, e1002480, doi:10.1371/journal.pcbi.1002480 (2012).

68    Nevozhay, D., Adams, R. M., Murphy, K. F., Josic, K. & Balázsi, G. Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *P Natl Acad Sci USA* **106**, 5123–, doi:10.1073/pnas.0809901106 (2009).

69    Ramos, J. L., Martinez-Bueno, M., Molina-Henares, A. J., Teran, W., Watanabe, K., Zhang, X., Gallegos, M. T., Brennan, R. & Tobes, R. The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* **69**, 326-356, doi:10.1128/MMBR.69.2.326-356.2005 (2005).

70    Hillen, W., Gatz, C., Altschmied, L., Schollmeier, K. & Meier, I. Control of expression of the Tn10-encoded tetracycline resistance genes. Equilibrium and kinetic investigation of the regulatory reactions. *J Mol Biol* **169**, 707-721 (1983).

71    Stanton, B. C., Nielsen, A. A., Tamsir, A., Clancy, K., Peterson, T. & Voigt, C. A. Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat Chem Biol* **10**, 99-105, doi:10.1038/nchembio.1411 (2014).

72    Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W. & Bujard, H. Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766-1769 (1995).

73    Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S. & Hasty, J. A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516-519, doi:10.1038/nature07389 (2008).

74    Tabor, J. J., Salis, H. M., Simpson, Z. B., Chevalier, A. A., Levskaya, A., Marcotte, E. M., Voigt, C. A. & Ellington, A. D. A synthetic genetic edge detection program. *Cell* **137**, 1272-1281, doi:10.1016/j.cell.2009.04.048 (2009).

75    Regot, S., Macia, J., Conde, N., Furukawa, K., Kjellen, J., Peeters, T., Hohmann, S., de Nadal, E., Posas, F. & Sole, R. Distributed biological computation with multicellular engineered networks. *Nature* **469**, 207-211, doi:10.1038/nature09679 (2011).

76    Courbet, A., Endy, D., Renard, E., Molina, F. & Bonnet, J. Detection of pathological biomarkers in human clinical samples via amplifying genetic switches and logic gates. *Sci Transl Med* **7**, 289ra283, doi:10.1126/scitranslmed.aaa3601 (2015).

77    Auslander, D., Auslander, S., Charpin-El Hamri, G., Sedlmayer, F., Muller, M., Frey, O., Hierlemann, A., Stelling, J. & Fussenegger, M. A synthetic multifunctional mammalian pH sensor and CO2 transgene-control device. *Mol Cell* **55**, 397-408, doi:10.1016/j.molcel.2014.06.007 (2014).

78    Ye, H., Xie, M., Xue, S., Charpin-El Hamri, G., Yin, J., Zulewski, H. & Fussenegger, M. Self-adjusting synthetic gene circuit for correcting insulin resistance. *Nat Biomed Eng* **1**, 0005, doi:10.1038/s41551-016-0005 (2017).

79    Ye, H., Daoud-El Baba, M., Peng, R. W. & Fussenegger, M. A synthetic optogenetic transcription device enhances blood-glucose homeostasis in mice. *Science* **332**, 1565-1568, doi:10.1126/science.1203535 (2011).

80    Kloss, C. C., Condomines, M., Cartellieri, M., Bachmann, M. & Sadelain, M. Combinatorial antigen recognition with balanced signaling promotes selective tumor eradication by engineered T cells. *Nat Biotechnol* **31**, 71-75, doi:10.1038/nbt.2459 (2013).

81    Burrill, D. R., Inniss, M. C., Boyle, P. M. & Silver, P. A. Synthetic memory circuits for tracking human cell fate. *Genes Dev* **26**, 1486-1497, doi:10.1101/gad.189035.112 (2012).

82    Nevozhay, D., Zal, T. & Balázsi, G. Transferring a synthetic gene circuit from yeast to mammalian cells. *Nat Commun* **4**, 1451, doi:10.1038/ncomms2471 (2013).

83    Morsut, L., Roybal, K. T., Xiong, X., Gordley, R. M., Coyle, S. M., Thomson, M. & Lim, W. A. Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell* **164**, 780-791, doi:10.1016/j.cell.2016.01.012 (2016).

84    Toda, S., Blauch, L. R., Tang, S. K. Y., Morsut, L. & Lim, W. A. Programming self-organizing multicellular structures with synthetic cell-cell signaling. *Science* **361**, 156-162, doi:10.1126/science.aat0271 (2018).

85    Lienert, F., Lohmueller, J. J., Garg, A. & Silver, P. A. Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nat Rev Mol Cell Biol* **15**, 95-107, doi:10.1038/nrm3738 (2014).

86 Duportet, X., Wroblewska, L., Guye, P., Li, Y. Q., Eyquem, J., Rieders, J., Rimchala, T., Batt, G. & Weiss, R. A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res* **42**, 13440-13451, doi:10.1093/nar/gku1082 (2014).

87 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J Clin* **68**, 7-30, doi:10.3322/caac.21442 (2018).

88 Nikolaou, M., Pavlopoulou, A., Georgakilas, A. G. & Kyrodimos, E. The challenge of drug resistance in cancer treatment: a current overview. *Clin Exp Metastasis* **35**, 309-318, doi:10.1007/s10585-018-9903-0 (2018).

89 Housman, G., Byler, S., Heerboth, S., Lapinska, K., Longacre, M., Snyder, N. & Sarkar, S. Drug Resistance in Cancer: An Overview. *Cancers* **6**, 1769-1792, doi:10.3390/cancers6031769 (2014).

90 Holohan, C., Van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer* **13**, 714-726, doi:10.1038/nrc3599 (2013).

91 Kessler, D. A., Austin, R. H. & Levine, H. Resistance to Chemotherapy: Patient Variability and Cellular Heterogeneity. *Cancer Res* **74**, 4663-4670, doi:10.1158/0008-5472.Can-14-0118 (2014).

92 Gottesman, M. M., Fojo, T. & Bates, S. E. Multidrug resistance in cancer: role of ATP-dependent transporters. *Nat Rev Cancer* **2**, 48-58, doi:10.1038/nrc706 (2002).

93 Kartner, N., Riordan, J. R. & Ling, V. Cell-Surface P-Glycoprotein Associated with Multidrug Resistance in Mammalian-Cell Lines. *Science* **221**, 1285-1288, doi:DOI 10.1126/science.6137059 (1983).

94 Ambudkar, S. V., Dey, S., Hrycyna, C. A., Ramachandra, M., Pastan, I. & Gottesman, M. M. Biochemical, cellular, and pharmacological aspects of the multidrug transporter. *Annu Rev Pharmacol* **39**, 361-398, doi:DOI 10.1146/annurev.pharmtox.39.1.361 (1999).

95 Kobayashi, S., Boggon, T. J., Dayaram, T., Janne, P. A., Kocher, O., Meyerson, M., Johnson, B. E., Eck, M. J., Tenen, D. G. & Halmos, B. EGFR mutation and resistance of

non-small-cell lung cancer to gefitinib. *New Engl J Med* **352**, 786-792, doi:DOI 10.1056/NEJMoa044238 (2005).

96    Riordan, J. R., Deuchars, K., Kartner, N., Alon, N., Trent, J. & Ling, V. Amplification of P-Glycoprotein Genes in Multidrug-Resistant Mammalian-Cell Lines. *Nature* **316**, 817-819, doi:DOI 10.1038/316817a0 (1985).

97    Lee, A. J. X., Endesfelder, D., Rowan, A. J., Walther, A., Birkbak, N. J., Futreal, P. A., Downward, J., Szallasi, Z., Tomlinson, I. P. M., Howell, M., Kschischo, M. & Swanton, C. Chromosomal Instability Confers Intrinsic Multidrug Resistance. *Cancer Res* **71**, 1858-1870, doi:10.1158/0008-5472.Can-10-3604 (2011).

98    Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338-345, doi:10.1038/nature12625 (2013).

99    Snijders, A. M., Fridlyand, J., Mans, D. A., Segraves, R., Jain, A. N., Pinkel, D. & Albertson, D. G. Shaping of tumor and drug-resistant genomes by instability and selection. *Oncogene* **22**, 4370-4379, doi:10.1038/sj.onc.1206482 (2003).

100   Meijer, C., Mulder, N. H., Timmer-Bosscha, H., Sluiter, W. J., Meersma, G. J. & de Vries, E. G. Relationship of cellular glutathione to the cytotoxicity and resistance of seven platinum compounds. *Cancer Res* **52**, 6885-6889 (1992).

101   Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilipenko, E. V., Shatsky, I. N., Agol, V. I. & Hellen, C. U. Molecular mechanisms of translation initiation in eukaryotes. *Proc Natl Acad Sci U S A* **98**, 7029-7036, doi:10.1073/pnas.111145798 (2001).

102   Nathans, D. Puromycin Inhibition of Protein Synthesis - Incorporation of Puromycin into Peptide Chains. *P Natl Acad Sci USA* **51**, 585-&, doi:DOI 10.1073/pnas.51.4.585 (1964).

103   Vara, J., Perezgonzalez, J. A. & Jimenez, A. Biosynthesis of Puromycin by Streptomyces-Alboniger - Characterization of Puromycin N-Acetyltransferase. *Biochemistry-Us* **24**, 8074-8081, doi:DOI 10.1021/bi00348a036 (1985).

104     Sharma, S. V., Lee, D. Y., Li, B., Quinlan, M. P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M. a., Wong, K.-K., Brandstetter, K., Wittner, B., Ramaswamy, S., Classon, M. & Settleman, J. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* **141**, 69–80, doi:10.1016/j.cell.2010.02.027 (2010).

105     Hangauer, M. J., Viswanathan, V. S., Ryan, M. J., Bole, D., Eaton, J. K., Matov, A., Galeas, J., Dhruv, H. D., Berens, M. E., Schreiber, S. L., McCormick, F. & McManus, M. T. Drug-tolerant persister cancer cells are vulnerable to GPX4 inhibition. *Nature* **551**, 247-250, doi:10.1038/nature24297 (2017).

106     Knoechel, B., Roderick, J. E., Williamson, K. E., Zhu, J., Lohr, J. G., Cotton, M. J., Gillespie, S. M., Fernandez, D., Ku, M., Wang, H., Piccioni, F., Silver, S. J., Jain, M., Pearson, D., Kluk, M. J., Ott, C. J., Shultz, L. D., Brehm, M. A., Greiner, D. L., Gutierrez, A., Stegmaier, K., Kung, A. L., Root, D. E., Bradner, J. E., Aster, J. C., Kelliher, M. A. & Bernstein, B. E. An epigenetic mechanism of resistance to targeted therapy in T cell acute lymphoblastic leukemia. *Nat Genet* **46**, 364-370, doi:10.1038/ng.2913 (2014).

107     Ramirez, M., Rajaram, S., Steininger, R. J., Osipchuk, D., Roth, M. A., Morinishi, L. S., Evans, L., Ji, W. Y., Hsu, C. H., Thurley, K., Wei, S. G., Zhou, A. W., Koduru, P. R., Posner, B. A., Wu, L. F. & Altschuler, S. J. Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. *Nat Commun* **7**, doi:ARTN 10690 10.1038/ncomms10690 (2016).

108     Shaffer, S. M., Dunagin, M. C., Torborg, S. R., Torre, E. A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P. A., Xiao, M., Ggan, E. E., Anastopoulos, I. N., Vargas-Garcia, C. A., Singh, A., Nathanson, K. L., Herlyn, M. & Raj, A. Rare cell variability and drug- induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431-+, doi:10.1038/nature22794 (2017).

109     Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat Med* **23**, 1124-1134, doi:10.1038/nm.4409 (2017).

110 Borst, P. Cancer drug pan-resistance: pumps, cancer stem cells, quiescence, epithelial to mesenchymal transition, blocked cell death pathways, persisters or what? *Open Biol* **2**, doi:ARTN 120066 10.1098/rsob.120066 (2012).

111 Begicevic, R. R. & Falasca, M. ABC Transporters in Cancer Stem Cells: Beyond Chemoresistance. *Int J Mol Sci* **18**, doi:ARTN 2362 10.3390/ijms18112362 (2017).

112 Zhang, S., Mercado-Uribe, I., Xing, Z., Sun, B., Kuang, J. & Liu, J. Generation of cancer stem-like cells through the formation of polyploid giant cancer cells (vol 33, pg 116, 2013). *Oncogene* **33**, 134-134, doi:10.1038/onc.2013.318 (2014).

113 Puig, P. E., Guilly, M. N., Bouchot, A., Droin, N., Cathelin, D., Bouyer, F., Favier, L., Ghiringhelli, F., Kroemer, G., Solary, E., Martin, F. & Chauffert, B. Tumor cells can escape DNA-damaging cisplatin through DNA endoreduplication and reversible polyploidy. *Cell Biol Int* **32**, 1031-1043, doi:10.1016/j.cellbi.2008.04.021 (2008).

114 Brock, A., Chang, H. & Huang, S. Non-genetic heterogeneity–a mutation-independent driving force for the somatic evolution of tumours. *Nature reviews. Genetics* **10**, 336–342, doi:10.1038/nrg2556 (2009).

115 Pisco, A. O., Brock, A., Zhou, J., Moor, A., Mojtahedi, M., Jackson, D. & Huang, S. Non-Darwinian dynamics in therapy-induced cancer drug resistance. *Nat Commun* **4**, 2467, doi:10.1038/ncomms3467 (2013).

116 Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M. & Sorger, P. K. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* **459**, 428–432, doi:10.1038/nature08012 (2009).

117 Orr, H. A. Fitness and its role in evolutionary genetics. *Nat Rev Genet* **10**, 531-539, doi:10.1038/nrg2603 (2009).

118 Moxon, E. R., Rainey, P. B., Nowak, M. A. & Lenski, R. E. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol* **4**, 24-33 (1994).

119  Lopez-Maury, L., Marguerat, S. & Bahler, J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet* **9**, 583-593, doi:10.1038/nrg2398 (2008).

120  Kawecki, T. J., Lenski, R. E., Ebert, D., Hollis, B., Olivieri, I. & Whitlock, M. C. Experimental evolution. *Trends Ecol Evol* **27**, 547-560, doi:10.1016/j.tree.2012.06.001 (2012).

121  de Visser, J. A. & Lenski, R. E. Long-term experimental evolution in Escherichia coli. XI. Rejection of non-transitive interactions as cause of declining rate of adaptation. *BMC Evol Biol* **2**, 19 (2002).

122  Barrick, J. E. & Lenski, R. E. Genome dynamics during experimental evolution. *Nature Reviews Genetics* **14**, 827-839, doi:10.1038/nrg3564 (2013).

123  Beaumont, H. J., Gallie, J., Kost, C., Ferguson, G. C. & Rainey, P. B. Experimental evolution of bet hedging. *Nature* **462**, 90-93, doi:10.1038/nature08504 (2009).

124  Sanchez-Romero, M. A. & Casadesus, J. Contribution of phenotypic heterogeneity to adaptive antibiotic resistance. *Proc Natl Acad Sci U S A* **111**, 355-360, doi:10.1073/pnas.1316084111 (2014).

125  Jolly, M. K., Kulkarni, P., Weninger, K., Orban, J. & Levine, H. Phenotypic Plasticity, Bet-Hedging, and Androgen Independence in Prostate Cancer: Role of Non-Genetic Heterogeneity. *Front Oncol* **8**, 50, doi:10.3389/fonc.2018.00050 (2018).

126  Majors, B. S., Chiang, G. G. & Betenbaugh, M. J. Protein and genome evolution in Mammalian cells for biotechnology applications. *Mol Biotechnol* **42**, 216-223, doi:10.1007/s12033-009-9156-x (2009).

127  Wurm, F. M. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol* **22**, 1393-1398, doi:10.1038/nbt1026 (2004).

128  Piatkevich, K. D., Jung, E. E., Straub, C., Linghu, C., Park, D., Suk, H. J., Hochbaum, D. R., Goodwin, D., Pnevmatikakis, E., Pak, N., Kawashima, T., Yang, C. T., Rhoades, J. L., Shemesh, O., Asano, S., Yoon, Y. G., Freifeld, L., Saulnier, J. L., Riegler, C., Engert, F.,

Hughes, T., Drobizhev, M., Szabo, B., Ahrens, M. B., Flavell, S. W., Sabatini, B. L. & Boyden, E. S. A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nat Chem Biol* **14**, 352-360, doi:10.1038/s41589-018-0004-9 (2018).

129   Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y. & Chang, X. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat Methods* **13**, 1029-1035, doi:10.1038/nmeth.4027 (2016).

130   Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

131   Rodgers, K. & McVey, M. Error-Prone Repair of DNA Double-Strand Breaks. *J Cell Physiol* **231**, 15-24, doi:10.1002/jcp.25053 (2016).

132   Cong, L., Ran, F. A., Cox, D., Lin, S. L., Barretto, R., Habib, N., Hsu, P. D., Wu, X. B., Jiang, W. Y., Marraffini, L. A. & Zhang, F. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).

133   Evers, B., Jastrzebski, K., Heijmans, J. P., Grernrum, W., Beijersbergen, R. L. & Bernards, R. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat Biotechnol* **34**, 631-633, doi:10.1038/nbt.3536 (2016).

134   Yoshimi, K., Kunihiro, Y., Kaneko, T., Nagahora, H., Voigt, B. & Mashimo, T. ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. *Nat Commun* **7**, doi:ARTN 10431 10.1038/ncomms10431 (2016).

135   Barrangou, R. & Doudna, J. A. Applications of CRISPR technologies in research and beyond. *Nat Biotechnol* **34**, 933-941, doi:10.1038/nbt.3659 (2016).

136   Koch, B., Nijmeijer, B., Kueblbeck, M., Cai, Y., Walther, N. & Ellenberg, J. Generation and validation of homozygous fluorescent knock-in cells using CRISPR-Cas9 genome editing. *Nat Protoc* **13**, 1465-1487, doi:10.1038/nprot.2018.042 (2018).

137  Ye, X., Tam, W. L., Shibue, T., Kaygusuz, Y., Reinhardt, F., Ng Eaton, E. & Weinberg, R. A. Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature* **525**, 256-260, doi:10.1038/nature14897 (2015).

138  Zhang, X. H., Tee, L. Y., Wang, X. G., Huang, Q. S. & Yang, S. H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther-Nucl Acids* **4**, doi:ARTN e264 10.1038/mtna.2015.37 (2015).

139  Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W. X., Chen, L. W., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z. & Liu, D. R. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-+, doi:10.1038/nature26155 (2018).

140  Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z. & Joung, J. K. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495, doi:10.1038/nature16526 (2016).

141  Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* **42**, W401-407, doi:10.1093/nar/gku410 (2014).

142  Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y. & Zhang, F. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380-1389, doi:10.1016/j.cell.2013.08.021 (2013).

143  Chu, V. T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K. & Kuhn, R. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* **33**, 543-U160, doi:10.1038/nbt.3198 (2015).

144  Fischer, K. R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S. T., Choi, H., El Rayes, T., Ryu, S., Troeger, J., Schwabe, R. F., Vahdat, L. T., Altorki, N. K., Mittal, V. & Gao, D.

Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472-476, doi:10.1038/nature15748 (2015).

145     Nieto, M. A., Huang, R. Y., Jackson, R. A. & Thiery, J. P. Emt: 2016. *Cell* **166**, 21-45, doi:10.1016/j.cell.2016.06.028 (2016).

146     Yang, X., Pursell, B., Lu, S., Chang, T. K. & Mercurio, A. M. Regulation of beta 4-integrin expression by epigenetic modifications in the mammary gland and during the epithelial-to-mesenchymal transition. *J Cell Sci* **122**, 2473-2480, doi:10.1242/jcs.049148 (2009).

147     Kim, Y., Kugler, M. C., Wei, Y., Kim, K. K., Li, X., Brumwell, A. N. & Chapman, H. A. Integrin alpha3beta1-dependent beta-catenin phosphorylation links epithelial Smad signaling to cell contacts. *J Cell Biol* **184**, 309-322, doi:10.1083/jcb.200806067 (2009).

148     Mori, S., Kodaira, M., Ito, A., Okazaki, M., Kawaguchi, N., Hamada, Y., Takada, Y. & Matsuura, N. Enhanced Expression of Integrin alpha v beta 3 Induced by TGF-beta Is Required for the Enhancing Effect of Fibroblast Growth Factor 1 (FGF1) in TGF-beta-Induced Epithelia-lMesenchymal Transition (EMT) in Mammary Epithelial Cells. *Plos One* **10**, doi:ARTN e0137486 10.1371/journal.pone.0137486 (2015).

149     Gomez Tejeda Zanudo, J., Guinn, M. T., Farquhar, K., Szenk, M., Steinway, S. N., Balazsi, G. & Albert, R. Towards control of cellular decision-making networks in the epithelial-to-mesenchymal transition. *Phys Biol* **16**, 031002, doi:10.1088/1478-3975/aaffa1 (2019).

150     Puisieux, A., Brabletz, T. & Caramel, J. Oncogenic roles of EMT-inducing transcription factors. *Nat Cell Biol* **16**, 488-494, doi:10.1038/ncb2976 (2014).

151     Thuault, S., Valcourt, U., Petersen, M., Manfioletti, G., Heldin, C. H. & Moustakas, A. Transforming growth factor-beta employs HMGA2 to elicit epithelial-mesenchymal transition. *J Cell Biol* **174**, 175-183, doi:10.1083/jcb.200512110 (2006).

152     Lee, M. K., Pardoux, C., Hall, M. C., Lee, P. S., Warburton, D., Qing, J., Smith, S. M. & Derynck, R. TGF-beta activates Erk MAP kinase signalling through direct phosphorylation of ShcA. *Embo J* **26**, 3957-3967, doi:10.1038/sj.emboj.7601818 (2007).

153  Siemens, H., Jackstadt, R., Hunten, S., Kaller, M., Menssen, A., Gotz, U. & Hermeking, H. miR-34 and SNAIL form a double-negative feedback loop to regulate epithelial-mesenchymal transitions. *Cell Cycle* **10**, 4256-4271, doi:10.4161/cc.10.24.18552 (2011).

154  Dave, N., Guaita-Esteruelas, S., Gutarra, S., Frias, A., Beltran, M., Peiro, S. & de Herreros, A. G. Functional Cooperation between Snail1 and Twist in the Regulation of ZEB1 Expression during Epithelial to Mesenchymal Transition. *J Biol Chem* **286**, doi:10.1074/jbc.M110.168625 (2011).

155  Ahn, Y. H., Gibbons, D. L., Chakravarti, D., Creighton, C. J., Rizvi, Z. H., Adams, H. P., Pertsemlidis, A., Gregory, P. A., Wright, J. A., Goodall, G. J., Flores, E. R. & Kurie, J. M. ZEB1 drives prometastatic actin cytoskeletal remodeling by downregulating miR-34a expression. *J Clin Invest* **122**, 3170-3183, doi:10.1172/JCI63608 (2012).

156  Burk, U., Schubert, J., Wellner, U., Schmalhofer, O., Vincan, E., Spaderna, S. & Brabletz, T. A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *Embo Rep* **9**, 582-589, doi:10.1038/embor.2008.74 (2008).

157  Wu, W. S., You, R. I., Cheng, C. C., Lee, M. C., Lin, T. Y. & Hu, C. T. Snail collaborates with EGR-1 and SP-1 to directly activate transcription of MMP 9 and ZEB1. *Sci Rep-Uk* **7**, doi:ARTN 17753 10.1038/s41598-017-18101-7 (2017).

158  Dangi-Garimella, S., Yun, J., Eves, E. M., Newman, M., Erkeland, S. J., Hammond, S. M., Minn, A. J. & Rosner, M. R. Raf kinase inhibitory protein suppresses a metastasis signalling cascade involving LIN28 and let-7. *Embo J* **28**, 347-358, doi:10.1038/emboj.2008.294 (2009).

159  Yun, J., Frankenberger, C. a., Kuo, W.-L., Boelens, M. C., Eves, E. M., Cheng, N., Liang, H., Li, W.-H., Ishwaran, H., Minn, A. J. & Rosner, M. R. Signalling pathway for RKIP and Let-7 regulates and predicts metastatic breast cancer. *The EMBO journal* **30**, 4500–4514, doi:10.1038/emboj.2011.312 (2011).

160    Boire, A., Covic, L., Agarwal, A., Jacques, S., Sherifl, S. & Kuliopulos, A. PAR1 is a matrix metalloprotease-1 receptor that promotes invasion and tumorigenesis of breast cancer cells. *Cell* **120**, 303-313, doi:DOI 10.1016/j.cell.2004.12.018 (2005).

161    Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D., Muthuswamy, L., Krasnitz, A., McCombie, W. R., Hicks, J. & Wigler, M. Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90-94, doi:10.1038/nature09807 (2011).

162    Finger, E. C. & Giaccia, A. J. Hypoxia, inflammation, and the tumor microenvironment in metastatic disease. *Cancer Metastasis Rev* **29**, 285-293, doi:10.1007/s10555-010-9224-5 (2010).

163    Wei, S. C., Fattet, L., Tsai, J. H., Guo, Y. R., Pai, V. H., Majeski, H. E., Chen, A. C., Sah, R. L., Taylor, S. S., Engler, A. J. & Yang, J. Matrix stiffness drives epithelial mesenchymal transition and tumour metastasis through a TWIST1-G3BP2 mechanotransduction pathway. *Nature Cell Biology* **17**, 678-U306, doi:10.1038/ncb3157 (2015).

164    Gonzalez-Garcia, I., Sole, R. V. & Costa, J. Metapopulation dynamics and spatial heterogeneity in cancer. *P Natl Acad Sci USA* **99**, 13085-13089, doi:10.1073/pnas.202139299 (2002).

165    Celia-Terrassa, T. & Kang, Y. Distinctive properties of metastasis-initiating cells. *Genes Dev* **30**, 892-908, doi:10.1101/gad.277681.116 (2016).

166    Aiello, N. M., Maddipati, R., Norgard, R. J., Balli, D., Li, J., Yuan, S., Yamazoe, T., Black, T., Sahmoud, A., Furth, E. E., Bar-Sagi, D. & Stanger, B. Z. EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. *Dev Cell* **45**, 681-695 e684, doi:10.1016/j.devcel.2018.05.027 (2018).

167    Lu, M., Jolly, M. K., Levine, H., Onuchic, J. N. & Ben-Jacob, E. MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. *Proc Natl Acad Sci U S A* **110**, 18144-18149, doi:10.1073/pnas.1318192110 (2013).

168    Beerling, E., Seinstra, D., de Wit, E., Kester, L., van der Velden, D., Maynard, C., Schafer, R., van Diest, P., Voest, E., van Oudenaarden, A., Vrisekoop, N. & van Rheenen, J. Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. *Cell Rep* **14**, 2281-2288, doi:10.1016/j.celrep.2016.02.034 (2016).

169    Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913-918, doi:10.1038/nature08781 (2010).

170    Kittisopikul, M. & Suel, G. M. Biological role of noise encoded in a genetic network motif. *P Natl Acad Sci USA* **107**, 13300-13305, doi:10.1073/pnas.1003975107 (2010).

171    Veening, J. W., Smits, W. K. & Kuipers, O. P. Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* **62**, 193-210, doi:10.1146/annurev.micro.62.081307.163002 (2008).

172    Pisco, A. O. & Huang, S. Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse: 'What does not kill me strengthens me'. *Br J Cancer* **112**, 1725-1732, doi:10.1038/bjc.2015.146 (2015).

173    Nguyen, A., Yoshida, M., Goodarzi, H. & Tavazoie, S. F. Highly variable cancer subpopulations that exhibit enhanced transcriptome variability and metastatic fitness. *Nat Commun* **7**, 11246, doi:10.1038/ncomms11246 (2016).

174    Wenzel, S. E., Tyurina, Y. Y., Zhao, J., St Croix, C. M., Dar, H. H., Mao, G., Tyurin, V. A., Anthonymuthu, T. S., Kapralov, A. A., Amoscato, A. A., Mikulska-Ruminska, K., Shrivastava, I. H., Kenny, E. M., Yang, Q., Rosenbaum, J. C., Sparvero, L. J., Emlet, D. R., Wen, X., Minami, Y., Qu, F., Watkins, S. C., Holman, T. R., VanDemark, A. P., Kellum, J. A., Bahar, I., Bayir, H. & Kagan, V. E. PEBP1 Wardens Ferroptosis by Enabling Lipoxygenase Generation of Lipid Death Signals. *Cell* **171**, 628-641 e626, doi:10.1016/j.cell.2017.09.044 (2017).

175    Lehner, B. Genotype to phenotype: lessons from model organisms for human genetics. *Nature reviews. Genetics* **14**, 168–178, doi:10.1038/nrg3404 (2013).

176  Albert, R. Scale-free networks in cell biology. *Journal of Cell Science* **118**, 4947-4957, doi:10.1242/jcs.02714 (2005).

177  Acar, M., Becskei, A. & van Oudenaarden, A. Enhancement of cellular memory by reducing stochastic transitions. *Nature* **435**, 228-232, doi:10.1038/nature03524 (2005).

178  Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467**, 167-173, doi:10.1038/nature09326 (2010).

179  Bishop, A. L., Rab, F. A., Sumner, E. R. & Avery, S. V. Phenotypic heterogeneity can enhance rare-cell survival in 'stress-sensitive' yeast populations. *Molecular Microbiology* **63**, 507-520, doi:10.1111/j.1365-2958.2006.05504.x (2007).

180  Fraser, D. & Kærn, M. A chance at survival: gene expression noise and phenotypic diversification strategies. *Molecular Microbiology* **71**, 1333-1340, doi:10.1111/j.1365-2958.2009.06605.x (2009).

181  González, C., Ray, J. C. J., Manhart, M., Adams, R. M., Nevozhay, D., Morozov, A. V. & Balázsi, G. Stress-response balance drives the evolution of a network module and its host genome. *Molecular Systems Biology* **11**, 827, doi:10.15252/msb.20156185 (2015).

182  Çağatay, T., Turcotte, M., Elowitz, M. B., Garcia-Ojalvo, J. & Süel, G. M. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell* **139**, 512-522, doi:10.1016/j.cell.2009.07.046 (2009).

183  Maamar, H., Raj, A. & Dubnau, D. Noise in gene expression determines cell fate in Bacillus subtilis. *Science* **317**, 526–529, doi:10.1126/science.1140818 (2007).

184  Dar, R. D., Hosmane, N. N., Arkin, M. R., Siliciano, R. F. & Weinberger, L. S. Screening for noise in gene expression identifies drug synergies. *Science* **1392**, doi:10.1126/science.1250220 (2014).

185  Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., Ding, H., Koh, J. L. Y., Toufighi, K., Mostafavi, S., Prinz, J., Onge, R. P. S., VanderSluis, B., Makhnevych, T., Vizeacoumar, F. J., Alizadeh, S., Bahr, S., Brost, R. L., Chen, Y., Cokol, M., Deshpande, R., Li, Z., Lin, Z.-Y., Liang, W., Marback, M., Paw, J., Luis, B.-J. S.,

Shuteriqi, E., Tong, A. H. Y., Dyk, N. v., Wallace, I. M., Whitney, J. A., Weirauch, M. T., Zhong, G., Zhu, H., Houry, W. A., Brudno, M., Ragibizadeh, S., Papp, B., Pál, C., Roth, F. P., Giaever, G., Nislow, C., Troyanskaya, O. G., Bussey, H., Bader, G. D., Gingras, A.-C., Morris, Q. D., Kim, P. M., Kaiser, C. A., Myers, C. L., Andrews, B. J. & Boone, C. The genetic landscape of a cell. *Science* **327**, 425-431, doi:10.1126/science.1180823 (2010).

186   Andrianantoandro, E., Basu, S., Karig, D. K. & Weiss, R. Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology* **2**, 2006.0028, doi:10.1038/msb4100073 (2006).

187   Aranda-Díaz, A., Mace, K., Zuleta, I., Harrigan, P. & El-Samad, H. Robust synthetic circuits for two-dimensional control of gene expression in yeast. *ACS Synthetic Biology* **6**, 545-554, doi:10.1021/acssynbio.6b00251 (2017).

188   Osella, M., Bosia, C., Corá, D. & Caselle, M. The role of incoherent microRNA-mediated feedforward loops in noise buffering. *PLOS Computational Biology* **7**, e1001101, doi:10.1371/journal.pcbi.1001101 (2011).

189   Quarton, T., Ehrhardt, K., Lee, J., Kannan, S., Li, Y., Ma, L. & Bleris, L. Mapping the operational landscape of microRNAs in synthetic gene circuits. *npj Systems Biology and Applications* **4**, 6, doi:10.1038/s41540-017-0043-y (2018).

190   Shimoga, V., White, J. T., Li, Y., Sontag, E. & Bleris, L. Synthetic mammalian transgene negative autoregulation. *Molecular Systems Biology* **9**, 670, doi:10.1038/msb.2013.27 (2013).

191   Longo, D. M., Hoffmann, A., Tsimring, L. S. & Hasty, J. Coherent activation of a synthetic mammalian gene network. *Systems and synthetic biology* **4**, 15–23, doi:10.1007/s11693-009-9044-5 (2010).

192   Siciliano, V., Menolascina, F., Marucci, L., Fracassi, C., Garzilli, I., Moretti, M. N. & di Bernardo, D. Construction and modelling of an inducible positive feedback loop stably integrated in a mammalian cell-line. *PLOS Computational Biology* **7**, e1002074, doi:10.1371/journal.pcbi.1002074 (2011).

193  O'Gorman, S., Fox, D. T. & Wahl, G. M. Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251**, 1351-1355 (1991).

194  Belete, M. K. & Balazsi, G. Optimality and adaptation of phenotypically switching cells in fluctuating environments. *Physical Review E* **92**, doi:10.1103/PhysRevE.92.062716 (2015).

195  Dehairs, J., Talebi, A., Cherifi, Y. & Swinnen, J. V. CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. *Sci Rep-Uk* **6**, 28973, doi:10.1038/srep28973 (2016).

196  Xu, X., Nagarajan, H., Lewis, N. E., Pan, S. K., Cai, Z. M., Liu, X., Chen, W. B., Xie, M., Wang, W. L., Hammond, S., Andersen, M. R., Neff, N., Passarelli, B., Koh, W., Fan, H. C., Wang, J. B., Gui, Y. T., Lee, K. H., Betenbaugh, M. J., Quake, S. R., Famili, I., Palsson, B. O. & Wang, J. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat Biotechnol* **29**, 735-U131, doi:10.1038/nbt.1932 (2011).

197  Baranyi, J., Roberts, T. & McClure, P. A non-autonomous differential equation to model bacterial growth. *Food Microbiology* **10**, 43-59, doi:10.1006/fmic.1993.1005 (1993).

198  Miyawaki, A., Sawano, A. & Kogure, T. Lighting up cells: labelling proteins with fluorophores. *Nature Cell Biology* **Suppl**, S1-7 (2003).

199  Charlebois, D. A., Intosalmi, J., Fraser, D. & Kaern, M. An algorithm for the stochastic simulation of gene expression and heterogeneous population dynamics. *Commun Comput Phys* **9**, 89-112, doi:10.4208/cicp.280110.070510a (2011).

200  Isackson, P. J. & Bertrand, K. P. Dominant negative mutations in the Tn10 Tet repressor - evidence for use of the conserved helix-turn-helix motif in DNA-binding. *P Natl Acad Sci USA* **82**, 6226-6230, doi:DOI 10.1073/pnas.82.18.6226 (1985).

201  Zhuravel, D., Fraser, D., St-Pierre, S., Tepliakova, L., Pang, W. L., Hasty, J. & Kærn, M. Phenotypic impact of regulatory noise in cellular stress-response pathways. *Systems and synthetic biology* **4**, 105-116, doi:10.1007/s11693-010-9055-2 (2010).

202 Travisano, M. Long-term experimental evolution in Escherichia coli. VI. Environmental constraints on adaptation and divergence. *Genetics* **146**, 471-479 (1997).

203 Dunham, M. J., Badrane, H., Ferea, T., Adams, J., Brown, P. O., Rosenzweig, F. & Botstein, D. Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. *P Natl Acad Sci USA* **99**, 16144-16149, doi:10.1073/pnas.242624799 (2002).

204 Stearns, S. C., Ackermann, M., Doebeli, M. & Kaiser, M. Experimental evolution of aging, growth, and reproduction in fruitflies. *P Natl Acad Sci USA* **97**, 3309-3313, doi:DOI 10.1073/pnas.060289597 (2000).

205 Taylor, T. B., Johnson, L. J., Jackson, R. W., Brockhurst, M. A. & Dash, P. R. First steps in experimental cancer evolution. *Evol Appl* **6**, 535-548, doi:10.1111/eva.12041 (2013).

206 Ellefson, J. W., Meyer, A. J., Hughes, R. A., Cannon, J. R., Brodbelt, J. S. & Ellington, A. D. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat Biotechnol* **32**, 97-101, doi:10.1038/nbt.2714 (2014).

207 Malina, A., Mills, J. R. & Pelletier, J. Emerging therapeutics targeting mRNA translation. *Cold Spring Harb Perspect Biol* **4**, a012377, doi:10.1101/cshperspect.a012377 (2012).

208 Bhat, M., Robichaud, N., Hulea, L., Sonenberg, N., Pelletier, J. & Topisirovic, I. Targeting the translation machinery in cancer. *Nature reviews. Drug Discovery* **14**, 261-278, doi:10.1038/nrd4505 (2015).

209 Ueki, N., Wang, W., Swenson, C., McNaughton, C., Sampson, N. S. & Hayman, M. J. Synthesis and preclinical evaluation of a highly improved anticancer prodrug activated by histone deacetylases and cathepsin L. *Theranostics* **6**, 808-816, doi:10.7150/thno.13826 (2016).

210 Monger, C., Motheramgari, K., McSharry, J., Barron, N. & Clarke, C. in *Heterologous Protein Production in Cho Cells: Methods and Protocols* Vol. 1603  (ed Meleady P.)  169-186 (2017).

211    Coward, J. & Harding, A. Size does matter: Why polyploid tumor cells are critical drug targets in the War on Cancer. *Frontiers in Oncology* **4**, 123, doi:10.3389/fonc.2014.00123 (2014).

212    Bradner, J. E., Hnisz, D. & Young, R. A. Transcriptional Addiction in Cancer. *Cell* **168**, 629-643, doi:10.1016/j.cell.2016.12.013 (2017).

213    Giancotti, F. G. Deregulation of cell signaling in cancer. *FEBS Lett* **588**, 2558-2570, doi:10.1016/j.febslet.2014.02.005 (2014).

214    Bissell, M. J. & Hines, W. C. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. *Nature Medicine* **17**, 320-329, doi:10.1038/nm.2328 (2011).

215    Reina-Campos, M., Moscat, J. & Diaz-Meco, M. Metabolism shapes the tumor microenvironment. *Current Opinion in Cell Biology* **48**, 47-53, doi:10.1016/j.ceb.2017.05.006 (2017).

216    Adams, B. D., Kasinski, A. L. & Slack, F. J. Aberrant Regulation and Function of MicroRNAs in Cancer. *Current Biology* **24**, R762-R776, doi:10.1016/j.cub.2014.06.043 (2014).

217    Cairns, R. A., Harris, I. S. & Mak, T. W. Regulation of cancer cell metabolism. *Nat Rev Cancer* **11**, 85-95, doi:10.1038/nrc2981 (2011).

218    Sottoriva, A., Spiteri, I., Piccirillo, S. G., Touloumis, A., Collins, V. P., Marioni, J. C., Curtis, C., Watts, C. & Tavare, S. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A* **110**, 4009-4014, doi:10.1073/pnas.1219747110 (2013).

219    Wood, L. D., Parsons, D. W., Jones, S., Lin, J., Sjoblom, T., Leary, R. J., Shen, D., Boca, S. M., Barber, T., Ptak, J., Silliman, N., Szabo, S., Dezso, Z., Ustyanksky, V., Nikolskaya, T., Nikolsky, Y., Karchin, R., Wilson, P. A., Kaminker, J. S., Zhang, Z., Croshaw, R., Willis, J., Dawson, D., Shipitsin, M., Willson, J. K., Sukumar, S., Polyak, K., Park, B. H., Pethiyagoda, C. L., Pant, P. V., Ballinger, D. G., Sparks, A. B., Hartigan, J., Smith, D. R.,

Suh, E., Papadopoulos, N., Buckhaults, P., Markowitz, S. D., Parmigiani, G., Kinzler, K. W., Velculescu, V. E. & Vogelstein, B. The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108-1113, doi:10.1126/science.1145720 (2007).

220    Bailey, M. H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M. C., Kim, J., Reardon, B., Ng, P. K.-S., Jeong, K. J., Cao, S., Wang, Z., Gao, J., Gao, Q., Wang, F., Liu, E. M., Mularoni, L., Rubio-Perez, C., Nagarajan, N., Cortés-Ciriano, I., Zhou, D. C., Liang, W.-W., Hess, J. M., Yellapantula, V. D., Tamborero, D., Gonzalez-Perez, A., Suphavilai, C., Ko, J. Y., Khurana, E., Park, P. J., Van Allen, E. M., Liang, H., Lawrence, M. S., Godzik, A., Lopez-Bigas, N., Stuart, J., Wheeler, D., Getz, G., Chen, K., Lazar, A. J., Mills, G. B., Karchin, R. & Ding, L. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371-385.e318, doi:10.1016/j.cell.2018.02.060 (2018).

221    Vanharanta, S. & Massague, J. Origins of metastatic traits. *Cancer Cell* **24**, 410-421, doi:10.1016/j.ccr.2013.09.007 (2013).

222    Kroigard, A. B., Larsen, M. J., Laenkholm, A. V., Knoop, A. S., Jensen, J. D., Bak, M., Mollenhauer, J., Thomassen, M. & Kruse, T. A. Identification of metastasis driver genes by massive parallel sequencing of successive steps of breast cancer progression. *Plos One* **13**, e0189887, doi:10.1371/journal.pone.0189887 (2018).

223    Marusyk, A. & Polyak, K. Tumor heterogeneity: causes and consequences. *Biochimica et biophysica acta* **1805**, 105–117, doi:10.1016/j.bbcan.2009.11.002 (2010).

224    Alizadeh, A. A., Aranda, V., Bardelli, A., Blanpain, C., Bock, C., Borowski, C., Caldas, C., Califano, A., Doherty, M., Elsner, M., Esteller, M., Fitzgerald, R., Korbel, J. O., Lichter, P., Mason, C. E., Navin, N., Pe'er, D., Polyak, K., Roberts, C. W., Siu, L., Snyder, A., Stower, H., Swanton, C., Verhaak, R. G., Zenklusen, J. C., Zuber, J. & Zucman-Rossi, J. Toward understanding and exploiting tumor heterogeneity. *Nat Med* **21**, 846-853, doi:10.1038/nm.3915 (2015).

225  Ye, X., Brabletz, T., Kang, Y., Longmore, G. D., Nieto, M. A., Stanger, B. Z., Yang, J. & Weinberg, R. A. Upholding a role for EMT in breast cancer metastasis. *Nature* **547**, E1-E3, doi:10.1038/nature22816 (2017).

226  Lin, Y.-C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., Chen, J., Speleman, F., Lambrechts, D., Van de Peer, Y., Tavernier, J. & Callewaert, N. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat Commun* **5**, 4767, doi:10.1038/ncomms5767 (2014).

227  Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol*, doi:10.1038/nbt.4192 (2018).

228  Kramer, N., Walzl, A., Unger, C., Rosner, M., Krupitza, G., Hengstschläger, M. & Dolznig, H. In vitro cell migration and invasion assays. *Mutation research* **752**, 10–24, doi:10.1016/j.mrrev.2012.08.001 (2013).

229  Heo, I., Joo, C., Cho, J., Ha, M., Han, J. & Kim, V. N. Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell* **32**, 276-284, doi:10.1016/j.molcel.2008.09.014 (2008).

230  Peixoto, A., Monteiro, M., Rocha, B. & Veiga-Fernandes, H. Quantification of Multiple Gene Expression in Individual Cells. *Genome Research* **14**, 1938-1947, doi:10.1101/gr.2890204 (2004).

231  Lo, C. A., Kays, I., Emran, F., Lin, T. J., Cvetkovska, V. & Chen, B. E. Quantification of Protein Levels in Single Living Cells. *Cell Rep* **13**, 2634-2644, doi:10.1016/j.celrep.2015.11.048 (2015).

232  Okita, Y., Kamoshida, A., Suzuki, H., Itoh, K., Motohashi, H., Igarashi, K., Yamamoto, M., Ogami, T., Koinuma, D. & Kato, M. Transforming Growth Factor-β Induces Transcription Factors MafK and Bach1 to Suppress Expression of the Heme Oxygenase-1 Gene. *J Biol Chem* **288**, 20658-20667, doi:10.1074/jbc.M113.450478 (2013).

233   Zenke-Kawasaki, Y., Dohi, Y., Katoh, Y., Ikura, T., Ikura, M., Asahara, T., Tokunaga, F., Iwai, K. & Igarashi, K. Heme induces ubiquitination and degradation of the transcription factor Bach1. *Molecular and cellular biology* **27**, 6962–6971, doi:10.1128/MCB.02415-06 (2007).

234   Suzuki, H., Tashiro, S., Hira, S., Sun, J., Yamazaki, C., Zenke, Y., Ikeda-Saito, M., Yoshida, M. & Igarashi, K. Heme regulates gene expression by triggering Crm1-dependent nuclear export of Bach1. *Embo J* **23**, 2544-2553, doi:10.1038/sj.emboj.7600248 (2004).

235   Iwakawa, H. O. & Tomari, Y. The Functions of MicroRNAs: mRNA Decay and Translational Repression. *Trends Cell Biol* **25**, 651-665, doi:10.1016/j.tcb.2015.07.011 (2015).

236   Kolch, W., Halasz, M., Granovskaya, M. & Kholodenko, B. N. The dynamic control of signal transduction networks in cancer cells. *Nat Rev Cancer* **15**, 515-527, doi:10.1038/nrc3983 (2015).

237   Batchelor, E., Loewer, A., Mock, C. & Lahav, G. Stimulus-dependent dynamics of p53 in single cells. *Mol Syst Biol* **7**, 488, doi:10.1038/msb.2011.20 (2011).

238   Geva-Zatorsky, N., Rosenfeld, N., Itzkovitz, S., Milo, R., Sigal, A., Dekel, E., Yarnitzky, T., Liron, Y., Polak, P., Lahav, G. & Alon, U. Oscillations and variability in the p53 system. *Mol Syst Biol* **2**, 2006 0033, doi:10.1038/msb4100068 (2006).

239   Meek, D. W. & Anderson, C. W. Posttranslational modification of p53: cooperative integrators of function. *Cold Spring Harb Perspect Biol* **1**, a000950, doi:10.1101/cshperspect.a000950 (2009).

240   Wu, X., Bayle, J. H., Olson, D. & Levine, A. J. The p53-mdm-2 autoregulatory feedback loop. *Genes Dev* **7**, 1126-1132 (1993).

241   Takekawa, M., Adachi, M., Nakahata, A., Nakayama, I., Itoh, F., Tsukuda, H., Taya, Y. & Imai, K. p53-inducible wip1 phosphatase mediates a negative feedback regulation of p38

MAPK-p53 signaling in response to UV radiation. *Embo J* **19**, 6517-6526, doi:10.1093/emboj/19.23.6517 (2000).

242   Ma, Y. & Hendershot, L. M. The role of the unfolded protein response in tumour development: friend or foe? *Nat Rev Cancer* **4**, 966-977, doi:10.1038/nrc1505 (2004).

243   Diao, J., Charlebois, D. A., Nevozhay, D., Bodi, Z., Pal, C. & Balazsi, G. Efflux Pump Control Alters Synthetic Gene Circuit Function. *ACS Synth Biol* **5**, 619-631, doi:10.1021/acssynbio.5b00154 (2016).

244   Morgillo, F., Della Corte, C. M., Fasano, M. & Ciardiello, F. Mechanisms of resistance to EGFR-targeted drugs: lung cancer. *ESMO Open* **1**, e000060, doi:10.1136/esmoopen-2016-000060 (2016).

245   Fei, F., Stoddart, S., Muschen, M., Kim, Y. M., Groffen, J. & Heisterkamp, N. Development of resistance to dasatinib in Bcr/Abl-positive acute lymphoblastic leukemia. *Leukemia* **24**, 813-820, doi:10.1038/leu.2009.302 (2010).

246   Li, X. Y., Wu, J. Z., Cao, H. X., Ma, R., Wu, J. Q., Zhong, Y. J. & Feng, J. F. Blockade of DNA methylation enhances the therapeutic effect of gefitinib in non-small cell lung cancer cells. *Oncol Rep* **29**, 1975-1982, doi:10.3892/or.2013.2298 (2013).

247   Van Schaeybroeck, S., Karaiskou-McCaul, A., Kelly, D., Longley, D., Galligan, L., Van Cutsem, E. & Johnston, P. Epidermal growth factor receptor activity determines response of colorectal cancer cells to gefitinib alone and in combination with chemotherapy. *Clin Cancer Res* **11**, 7480-7489, doi:10.1158/1078-0432.CCR-05-0328 (2005).

248   Dasari, S. & Tchounwou, P. B. Cisplatin in cancer therapy: molecular mechanisms of action. *Eur J Pharmacol* **740**, 364-378, doi:10.1016/j.ejphar.2014.07.025 (2014).

249   Gut, G., Tadmor, M. D., Pe'er, D., Pelkmans, L. & Liberali, P. Trajectories of cell-cycle progression from fixed cell populations. *Nat Methods* **12**, 951-954, doi:10.1038/nmeth.3545 (2015).

250   Kavallaris, M. Microtubules and resistance to tubulin-binding agents. *Nat Rev Cancer* **10**, 194-204, doi:10.1038/nrc2803 (2010).

251    Nitiss, J. L. Targeting DNA topoisomerase II in cancer chemotherapy. *Nat Rev Cancer* **9**, 338-350, doi:10.1038/nrc2607 (2009).

252    Wang, Q. J., Zambetti, G. P. & Suttle, D. P. Inhibition of DNA topoisomerase II alpha gene expression by the p53 tumor suppressor. *Molecular and Cellular Biology* **17**, 389-397, doi:Doi 10.1128/Mcb.17.1.389 (1997).

253    Yan, X., Hoek, T. A., Vale, R. D. & Tanenbaum, M. E. Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell* **165**, 976-989, doi:10.1016/j.cell.2016.04.034 (2016).

254    Bentires-Alj, M., Barbu, V., Fillet, M., Chariot, A., Relic, B., Jacobs, N., Gielen, J., Merville, M. P. & Bours, V. NF-kappaB transcription factor induces drug resistance through MDR1 expression in cancer cells. *Oncogene* **22**, 90-97, doi:10.1038/sj.onc.1206056 (2003).

255    Martz, C. A., Ottina, K. A., Singleton, K. R., Jasper, J. S., Wardell, S. E., Peraza-Penton, A., Anderson, G. R., Winter, P. S., Wang, T., Alley, H. M., Kwong, L. N., Cooper, Z. A., Tetzlaff, M., Chen, P. L., Rathmell, J. C., Flaherty, K. T., Wargo, J. A., McDonnell, D. P., Sabatini, D. M. & Wood, K. C. Systematic identification of signaling pathways with potential to confer anticancer drug resistance. *Sci Signal* **7**, ra121, doi:10.1126/scisignal.aaa1877 (2014).

256    Wang, Z., Li, Y., Ahmad, A., Azmi, A. S., Banerjee, S., Kong, D. & Sarkar, F. H. Targeting Notch signaling pathway to overcome drug resistance for cancer therapy. *Biochim Biophys Acta* **1806**, 258-267, doi:10.1016/j.bbcan.2010.06.001 (2010).

257    Guerrero-Zotano, A., Mayer, I. A. & Arteaga, C. L. PI3K/AKT/mTOR: role in breast cancer progression, drug resistance, and treatment. *Cancer Metastasis Rev* **35**, 515-524, doi:10.1007/s10555-016-9637-x (2016).

258    Wang, X. J., Sun, Z., Villeneuve, N. F., Zhang, S., Zhao, F., Li, Y., Chen, W., Yi, X., Zheng, W., Wondrak, G. T., Wong, P. K. & Zhang, D. D. Nrf2 enhances resistance of cancer cells to chemotherapeutic drugs, the dark side of Nrf2. *Carcinogenesis* **29**, 1235-1243, doi:10.1093/carcin/bgn095 (2008).

259 Chatterjee, D., Bai, Y., Wang, Z., Beach, S., Mott, S., Roy, R., Braastad, C., Sun, Y., Mukhopadhyay, A., Aggarwal, B. B., Darnowski, J., Pantazis, P., Wyche, J., Fu, Z., Kitagwa, Y., Keller, E. T., Sedivy, J. M. & Yeung, K. C. RKIP sensitizes prostate and breast cancer cells to drug-induced apoptosis. *J Biol Chem* **279**, 17515-17523, doi:10.1074/jbc.M313816200 (2004).

260 Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol Cell* **58**, 610-620, doi:10.1016/j.molcel.2015.04.005 (2015).

261 Svensson, V., Vento-Tormo, R. & Teichmann, S. A. Exponential scaling of single-cell RNA-seq in the past decade. *Nat Protoc* **13**, 599-604, doi:10.1038/nprot.2017.149 (2018).

262 Jiang, L., Schlesinger, F., Davis, C. A., Zhang, Y., Li, R., Salit, M., Gingeras, T. R. & Oliver, B. Synthetic spike-in standards for RNA-seq experiments. *Genome Res* **21**, 1543-1551, doi:10.1101/gr.121095.111 (2011).

263 Islam, S., Kjallquist, U., Moliner, A., Zajac, P., Fan, J. B., Lonnerberg, P. & Linnarsson, S. Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nat Protoc* **7**, 813-828, doi:10.1038/nprot.2012.022 (2012).

264 Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I. & Enard, W. Comparative Analysis of Single-Cell RNA Sequencing Methods. *Mol Cell* **65**, 631-643 e634, doi:10.1016/j.molcel.2017.01.023 (2017).

265 Feltz, C. J. & Miller, G. E. An asymptotic test for the equality of coefficients of variation from k populations. *Stat Med* **15**, 646-658 (1996).

# Vita

Kevin Scott Farquhar was born in Conroe, Texas on September 6, 1988, the son of Jean Ann Farquhar and Randall James Farquhar. After completing his work at Klein High School, Klein, Texas in 2007, he entered Baylor University in Waco, Texas. He received the degree of Bachelor of Science with a major in biology from Baylor in May, 2011. In August of 2011 he entered The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences.

Permanent address:

6506 Spring Leaf Dr.

Spring, Texas 77379