Engineering and Quantifying Drug Response

In Small Cell Lung Cancer

By

Clayton Michael Wandishin

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Cellular Reprogramming and Systems Biology

May 13, 2022

Nashville, Tennessee

Approved:

Vito Quaranta, M.D.

Carlos Lopez, Ph.D.

Christine Lovly, M.D., Ph.D.

Rachelle Johnson, Ph.D.

Linda Sealy, Ph.D.

To my friends, family, and health. Cheers.

Acknowledgments

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Amp ……………………………………………………………………….Ampicillin
c-Myc……………………………………………………………………Cellular MYC
CAG……………………………Cytomegalovirus-beta Actin-beta Globin synthetic promoter
Cas9………………………………………………………..CRISPR associated protein 9
CMV………………………………………………………Cytomegalovirus promoter
CRISPR……………………….Clustered regularly interspersed short palindromic repeats
CRISPRa…………Clustered regularly interspersed short palindromic repeats activation
CRISPRi……………Clustered regularly interspersed short palindromic repeats inhibition
cyTOF……………………………………………………….Cytometry by time of flight
dCas9……………………………………………………Dead CRISPR associated protein 9
DIP…………………………………………………Drug-induced proliferation (rate)
DMEM/F12……………...Dulbecco's modified eagle's medium supplemented with Ham's
………………………………………………………………………...F12 nutrient mix
DNA…………………………………………………………………Deoxyribonucleic acid
dsDNA………………………………………………Double-stranded deoxyribonucleic acid
EC50 …………………………………………………Excitatory concentration fifty percent
Ef1a………………………………………………….Translation elongation factor 1A promoter
E_{max}……………………………………………………………Maximum effect
E_{min}……………………………………………………………Minimum effect
EMT……………………………………………….....Epithelial to mesenchymal transition
FACS………………………………………………..Fluorescence assisted cell sorting
FBS……………………………………………………Fetal bovine serum
H2A………………………………………………………………Histone 2A
hUbc…………………………………………………Human Ubiquitin c promoter
I/O………………………………………………………Immuno-oncology
IC50………………………………………………Inhibitory concentration fifty percent
Kan……………………………………………………………Kanamycin
mAzamiGreen……………………………………………Monomeric azami green protein
mKate2……………………………………………………Monomeric katushka two protein
ML………………………………………………………………...……Mesenchymal like
mRuby2……………………………………………...Monomeric ruby two protein
MT……………3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide derivative
mTagBFP……………………………………………Monomeric tag blue fluorescent protein
MTK……………………………………………………………Mammalian tool kit
MTS……3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-
………………………………………………………………………tetrazolium
MTT…………………………3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide
NE………………………………………………………………………Neuroendocrine
NEv1…………………………………………………Neuroendocrine variant one
NEv2…………………………………………………Neuroendocrine variant two
NIH……………………………………………………National institutes of health
NLS……………………………………………………Nuclear localization signal
nNE……………………………………………………...…Non-neuroendocrine
Opti-MEM……………………………Reduced serum improved minimal essential medium

Chapter 1: Introduction

1.1 History of cancer

Cancer was first described in form as early as 3000 BC in Egypt on what is now known as the "Edwin Smith Papyrus", a copy of a contemporary textbook of that time on trauma surgery. It describes eight types of tumors and ulcers, ending with the foreboding mention that, "There is no treatment". A few millennia later, Cancer was first described as "Cancer" by the Greek physician Hippocrates, using the terms "*carcinos*" and *"carcinoma"* to describe the "finger-like" projections emanating from the tumors he was dissecting. In Greek, both terms refer to a crab, and when they were subsequently translated into Latin by the roman physician Celsus they became "Cancer", the Latin word for crab. Similarly, another Greek physician studying these tumors, Galen, described them as *"oncos"* meaning "swelling" which is how we gleaned the modern term for cancer specialists, oncologists ("Understanding What Cancer Is: Ancient Times to Present," n.d., p. Cancer.org).

In modern times we have greatly improved our understanding of this incredibly complex and unfortunately common disease. In fact, over the last 75-100 years the mortality rate for some cancers has decreased by over 50% ("History of Cancer Screening and Early Detection: 20th Century to Present," n.d., p. Cancer.org). This is primarily due to both the advent of chemotherapeutic treatments, as well as the popularization of a variety of early screening and detection methods. However, not all cancers have enjoyed this renaissance of both knowledge and positive patient outlooks.

To rectify the lack of progress being made in a subset of cancer types, in 2014 the United States Congress and the National Institutes of Health successfully passed a bill naming a list of cancers "recalcitrant", and subsequently provided substantial funding for research into both characterizing their methodology as well as improving their clinical outcomes. Specifically, Small Cell Lung Cancer (SCLC) was named as part of this list, leading to a resurgence in research interest ("Lung Cancer - Small Cell: Statistics | Cancer.Net," n.d., p. Cancer.net).

1.2 Small cell lung cancer

SCLC was first described as a lung cancer in 1926, and then as its own distinct form in 1959 (Figure 1.2.1) (Gazdar, Bunn, & Minna, 2017; Semenova, Nagel, & Berns, 2015). Globally, it kills about 250,000 people per year and comprises roughly 15-20% of all lung cancer incidences, the rest being Non-Small Cell Lung Cancer (Gazdar et al., 2017; Sabari, Lok, Laird, Poirier, & Rudin, 2017). Only 7% of all patients diagnosed with SCLC successfully make it to the 5-year mark, perhaps due in part to its aggressive metastatic preferences as well as its unusually high propensity to settle in the brain ("Lung Cancer Survival Rates | 5-Year Survival Rates for Lung Cancer," n.d.). In fact, SCLC metastases are so aggressive that ~70% of patients will already have at least one metastatic site at the time of diagnosis, and brain metastases are so widespread that prophylactic cranial irradiation has become a standard of care (Gazdar et al., 2017; Ko, Winslow, & Sage, 2021). Additionally, SCLC is known to exhibit a characteristic "tumor recurrence" in regards to its clinical treatment where there is an initial highly effective tumor response to combination chemotherapeutics (Cisplatin/Etoposide), followed by an aggressive rebounding of the tumor population that is no longer
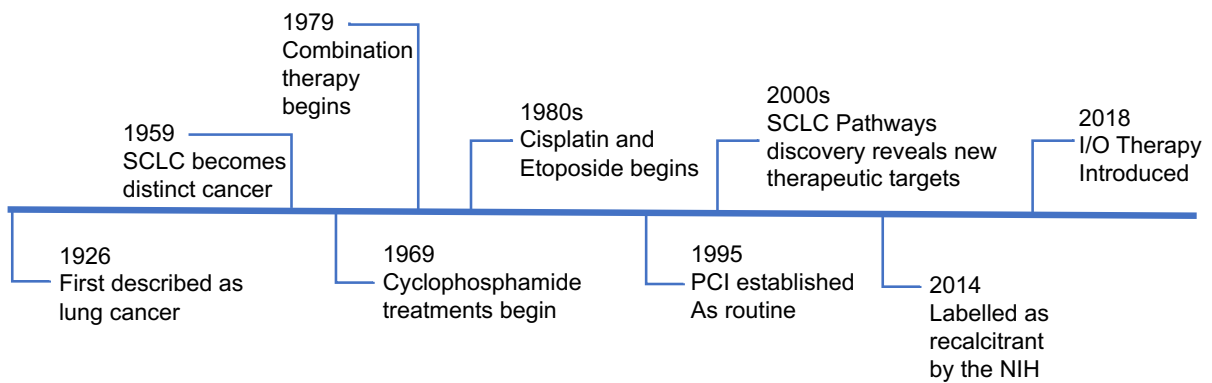
Figure 1.2.1 Timeline of small cell lung cancer milestones. A timeline of major milestones in small cell lung cancer shows that despite being first described in 1926, and initial treatments beginning in 1969, from the 1980s through 2019 the standard treatment of cisplatin/etoposide remained largely unchanged, even in the face of genetic research advancements in the early 2000s. 2018 marked the first change in small cell lung cancer treatment in over two decades with the administration of immuno-oncological therapy involving anti PD-L1 antibodies, in addition to cisplatin/etoposide, for extensive-stage small cell lung cancer patients (Horn et al., 2018).

responsive to treatment, culminating in very poor patient outcomes (Asai, Ohkuni, Kaneko, Yamaguchi, & Kubo, 2014). Perhaps the most interesting statistic though is the fact that 97% of all SCLC patients are characterized as "heavy smokers", suggesting there is a strong environmental and/or epigenetic component to this disease (Gazdar et al., 2017; Haddadin & Perry, 2011; Semenova et al., 2015).

1.3 Genomic subtyping of SCLC

One of the most promising trends in current SCLC research is efficient and specific characterization of cell lines and tumors by genomic analyses (McFadden et al., 2014; Polley et al., 2016; Zheng et al., 2017). This has been pursued by gene microarray, RNAseq, and genome sequencing, revealing distinct gene expression profiles between the two classical morphological subtypes, Neuroendocrine like (NE) and non-Neuroendocrine like (nNE/"Mesenchymal Like" ML). Additionally, recent data from the Quaranta lab using a systems biological approach to cluster gene correlates in to phenotypes has identified two further phenotypes within the classical designations, Neuroendocrine variant 1 (NEv1), and Neuroendocrine variant 2 (NEv2) (Wooten et al., 2019). These two additional phenotypes allow for a much more exact classification of individual tumors and cell lines, which has yielded preliminary computational results suggesting that the phenotypic gene expression profile for each can be used as a predictor of drug response and effectiveness. Moreover, assuming this phenotypic landscape is dynamic, I hypothesized that by transitioning from one phenotype's gene expression profile to another, the drug sensitivity could be similarly transitioned (Figure 1.3.1).
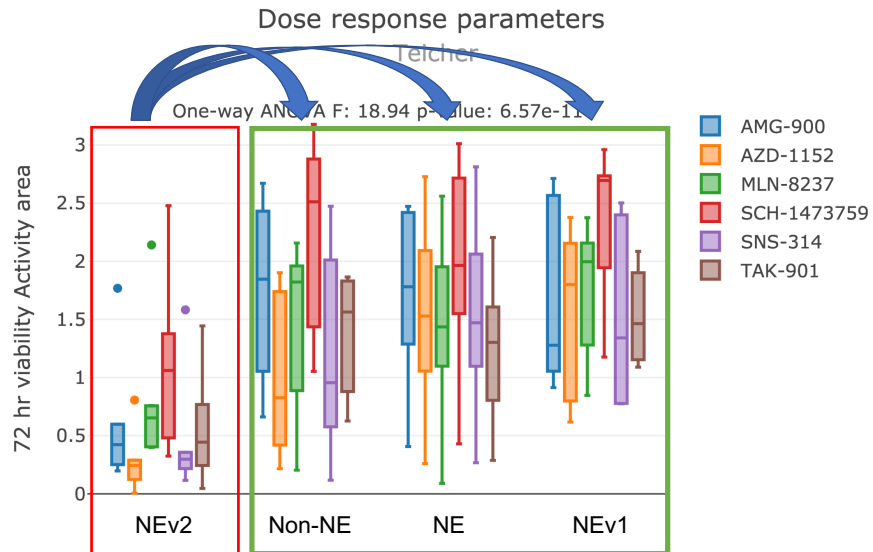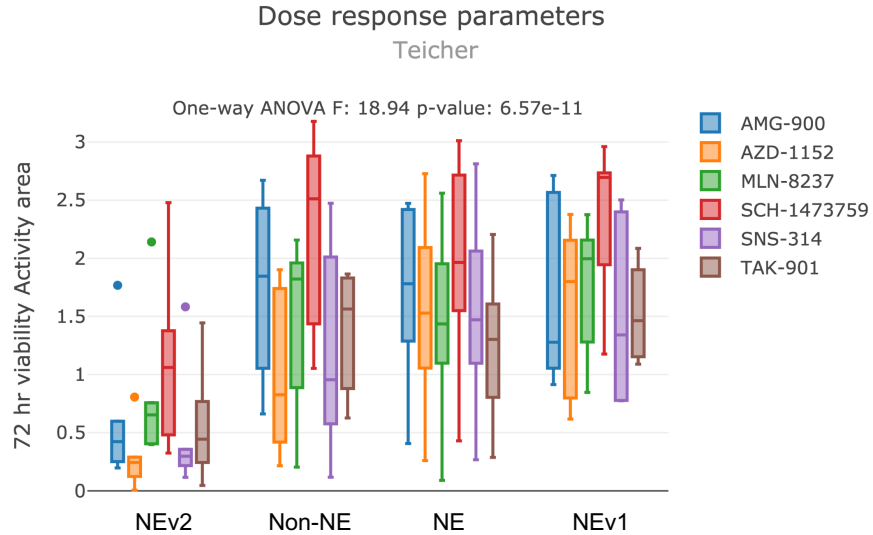
Figure 1.3.1 Diagram illustrating proposed SCLC phenotypic transition. 72 hour viability (activity area) drug-response measures of a selection of Aurora-Kinase inhibitors in SCLC cell lines grouped together by subtype. Observed differential drug response among SCLC subtypes (Top) and Proposed subtype transitioning (Bottom). The goal of the project is to shift the transcription factor network profile of a resistant subtype (red) to that of a more sensitive subtype (green) in the hope that the drug response profile will similarly transition.

1.4 Phenotypic transitions in SCLC

Within SCLC, phenotypic transitions have been thoroughly investigated and are thought to be a major contributor to the disease's aggressiveness (Krohn et al., 2014; Nieto, Huang, Jackson, & Thiery, 2016). The ability of SCLC to transition from a "neuroendocrine like" progenitor cell, for circulation in bodily fluid, to a more adherent "mesenchymal like" morphology, for colony establishment in tissue types outside the lung, is what makes it such a metastatic threat (Ca et al., n.d.; Nieto et al., 2016). While the exact mechanism of this transition has yet to be determined, the prevailing findings are that it is both highly complex and nuanced, relying on transcription factors (TF) as "super-enhancers" or "master-regulators" for a variety of oncogenic and anti-apoptotic pathways. However, it is the strength of the SCLC TF network that may also be its greatest weakness.

1.5 Epigenetic modulation in SCLC using CRISPR

Since its inception, CRISPR-Cas9 and its related technologies have changed the way modern genetic editing is done (Haurwitz et al., 2010). One of the newest CRISPR-Cas9 tools is the dCas9 genomic targeting method. Through the use of two amino acid mutations in the Cas9 protein, its endonuclease activity is abolished while still retaining the ability to bind gRNAs and subsequently traffic to specific genomic loci. By fusing this dCas9 protein to transcriptional activators (CRISPRa) or repressors (CRISPRi), highly efficient gene expression modulation can be accomplished (Konermann et al., 2015; Maeder et al., 2013). Furthermore, by using this system to activate or repress TFs instead of specific genes, global gene expression profile changes can be made with minimal amounts of perturbation, very similar in concept to

biological cascades. By redirecting SCLCs already highly influential TF network towards a profile more correlated with drug sensitive phenotypes, I planned to engineer drug sensitivity in previously insensitive cell lines.

In order to effectively probe the genetics and epigenetics of the TF network in regards to drug response, I developed a cell line library composed of single knockout lines, as well as tunable CRISPRa/i incorporated ones, all broadly representing the four previously identified phenotypic clusters from RNAseq analysis (NE, NEv1, NEv2, nNE). This library then served as an indispensable tool for iterative modulation of the TF network, as subsequent experiments characterized the individual contributions of each TF to both the stability of the phenotype being probed. Next, I developed two parallel approaches to quantitatively analyze the drug responses from this cell line library.

1.6 Real-time luminescence in SCLC

First, I used the components of a commercially available cellular viability assay to design a custom, continuous protocol for quantifying the metabolic activity of a cell in response to drug (Promega Real-Time Glo MT Cellular Viability Assay). Then, I designed a custom mathematical analysis algorithm via Python to assess the raw continuous data using analogous principles to those outlined previously by colleagues in the Quaranta lab (Harris et al., 2016b). These results were then benchmarked against direct cell counting measurements taken from within the exact same wells which lead to my second parallel approach, linking real-time luminescence measurements to cell counts by way of mathematically modelling the dynamics of the biochemical system.

1.7 Computational modelling of real-time luminescence in SCLC

In order to model the relationship between real-time luminescence and cell count, PySB (Python for Systems Biology) was used as a framework to create the initial components of the system as a whole (Lopez, Muhlich, Bachman, & Sorger, 2013). From there, a mix of experimental results and computational parameter space optimization techniques were used to define values for each of the individual components and rates. Once the initial parameters were defined, the model as a whole was used to compare simulated luminescence data based on cell counts to experimental luminescence data, as well as simulated cell counts based on raw luminescence data to experimental cell counts in order to assess congruency and iteratively refine to best fit the data.

1.8 Outline of dissertation

These completed tasks can be broadly grouped into the three specific aims of my dissertation project: 1) Engineer a SCLC cell line library to enable experimental testing of SCLC phenotype shift and its relationship to drug response 2) Develop an assay and analysis pipeline capable of continuous quantitation of drug response in SCLC across broad morphologies 3) Computationally model the relationship between real-time luminescence and cell count in order to integrate luminescence derived data into existing direct cell counting analyses. The current progress and results of these aims are discussed in the following three chapters.

Chapter 2: Engineering a Small Cell Lung Cancer CRISPR Cell Line Library

2.1 Introduction

In 2020 the Nobel Prize in Chemistry was awarded jointly to Emmanuelle Charpentier and Jennifer Doudna for their discovery of CRISPR/Cas technology in bacteria. Since then, this biological process that originally evolved for protection from, and excision of, viral DNA has become the most efficient and powerful genetic editing tool the world over. CRISPR (CRISPR/Cas), which stands for "Clustered Regularly Interspaced Short Palindromic Repeats", works by creating double-stranded breaks in DNA by way of a nuclease bound to a "guide" RNA molecule that is complementary to a specific genomic site (Jinek et al., 2012). Once this ribonucleoprotein Cas9/gRNA complex binds to its complementary target region of the genomic DNA, the Cas9 nuclease creates a double stranded DNA break, leading to DNA repair cascades within the cell that often result in base insertions or deletions, "indels". These indels in turn can cause a genetic frame-shift leading to abrogation of meaningful genetic sequences following the double-stranded break. Depending on where in a gene these indels occur, this can result in truncated proteins or complete knockouts. Because of this, CRISPR technologies have become an incredibly popular tool within molecular biology for specifically and efficiently creating cells that contain one or more gene knockouts. Within my first thesis aim of engineering a SCLC cell line library, I used CRISPR extensively (Jinek et al., 2012).

In initial brainstorming of what I wanted the SCLC cell line library to contain, I settled on four components to guide my constructions. First, I wanted to pick two cell lines from each of the subtypes previously identified by David Wooten and Sarah Groves, relying on their RNAseq k-means clustering analysis to give guidance on "exemplar" candidates that were firmly sorted into their respective subtype with little ambiguity (Wooten et al., 2019). Had I not taken this into consideration, cell lines which lie somewhat towards the edge of a cluster, or contain "mixed" phenotypic traits could have been included in the library, unnecessarily complicating and potentially obfuscating downstream analyses of response difference between the phenotypes. The decision on which cell lines were labelled as exemplar was entirely based on the work of, and guidance from, Sarah Groves and could not have been accomplished without. Second, I wanted to make sure that the library accurately depicted the common morphologies found across SCLC lines (e.g. Adherent, Mixed, Suspension, Aggregate) which was made serendipitously facile by nature of the exemplar lines chosen; all culture morphologies were well represented. Third, I wanted each line to be fluorescently nuclear labelled to give the option of direct counting as a quantification measure downstream. And fourth, I wanted each individual line to have multiple variants of both the fluorescent label, as well as CRISPR machinery, in order to make increasing, decreasing, or completely abolishing individual genes possible.

2.2 Results

2.2.1 Cloning of CRISPR vectors

Initial attempts at cloning the necessary vectors for these lines immediately proved to be troublesome. Because of factors such as secondary and quaternary

structure of the genetic materials, size of the intended insertion sequences, and relative complexity of the vector as a whole, the vast majority of modern cloning techniques proved to be ineffective or overly inefficient for my cloning purposes. Techniques tried include but are not limited to: Gibson Assembly, Restriction Enzyme Cloning, Overlap-Extension PCR, and Ligation Independent Cloning. However, what did eventually work was Golden Gate Assembly.

Golden Gate Assembly is a relatively new cloning method that relies on both the outside recognition site cutting of Type IIS restriction enzymes, as well as the use of a modular insert library that contains these Type IIS sequences either outside (insert) or inside (destination) of the sequences of interest. The general principle of this cloning method is illustrated in Figure 2.2.1.1.

In order for Golden Gate Assembly to be effective, it is best used as part of a modular cloning library. For this project, I was fortunate enough to start with a "base library" generously shared by the El-Samad lab at UCSF, along with their "Mammalian Tool Kit" (MTK) Protocol (Fonseca et al., 2019). From there, I only had to domesticate about 30 additional parts in order to fully assemble the plasmids for the library. This domestication and assembly process is illustrated across Figures 2.2.1.1-2.2.1.4. In brief, genetic material is ordered as oligos and annealed into dsDNA. From there, Gibson Assembly is used to create and "domesticate" individual parts of an MTK plasmid to incorporate a specific four base pair overhang. Each part is grouped according to its role (e.g. linker sequence, promoter, coding sequence, label, etc.). Once domesticated, multiple parts are combined into a one-pot Golden Gate reaction in order to create a

Golden Gate Assembly
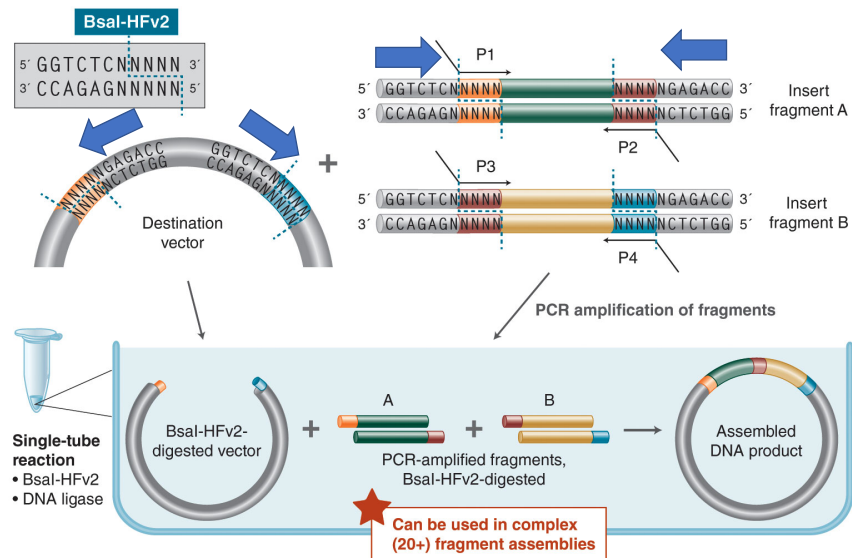
Figure 2.2.1.1 Illustration of golden gate assembly mechanism. Golden gate assembly works by using Type IIS restriction enzymes that cut DNA outside of their recognition sequence to create four base-pair overhangs. Because the nuclease activity occurs outside of the recognition site, by engineering destination sequences that have cutting occur from the inside out, and insert sequences where cutting occurs from the outside in, ordered and scarless ligation products can be efficiently and easily assembled from a one-pot thermal cycling reaction.

MTK Part Structures

Figure 2.2.1.2 MTK part structure assembly diagram. An assembled plasmid is illustrated here, composed of MTK parts 1-8. Each individual MTK part type is specifically designed to include the correct four base-pair overhang so that the assembly is ordered. Parts 1-5 complete the transcriptional unit of the plasmid, while parts 6-8 consist of components needed for bacterial maintenance (amplification).

Complete Transcriptional Units
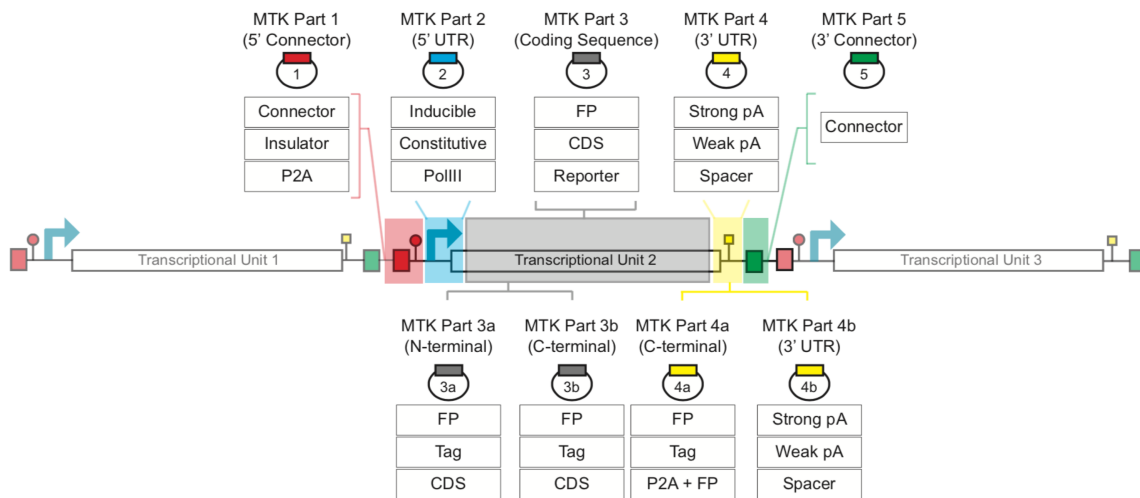
Figure 2.2.1.3 Diagram of a complete transcriptional unit and part designation. A complete MTK transcriptional unit is composed of MTK parts 1-5. Additionally, MTK parts 3 and 4 can be split into sub-parts 3a,3b,4a, and 4b. Parts 1-5 are described in the illustration above but can generally be classified as: 1) 5' Connector 2) 5' UTR 3) Coding Sequence 4) 3' UTR and 5) 3' Connector.

MTK Workflow
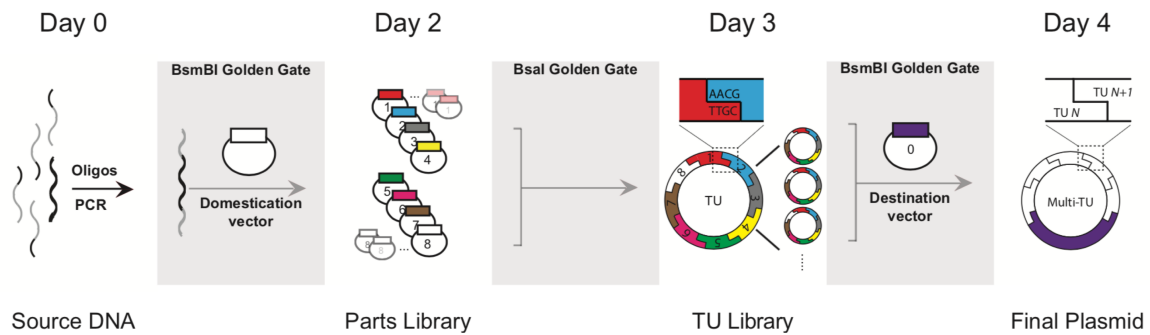
From Lee et. Al.

Figure 2.2.1.4 Diagram of overall MTK workflow. The MTK workflow illustrated above generally involves the following steps and timelines: Day 0-1: Anneal DNA oligos and Gibson assemble into defined domestication vectors. Day 2: Golden gate assemble a complete TU from domesticated parts. Day 3: Golden gate assemble TU(s) into a destination vector for expression. Day 4: Amplify, purify, store, and use the fully assembled final plasmid.

complete "Transcriptional Unit" (TU) plasmid. Afterwards, TU plasmids can either be combined with additional TUs to create increasingly complex vectors, or directly inserted into a destination plasmid by way of a further Golden Gate Assembly reaction.

The extensive modularity of the MTK system allowed me to quickly and easily create multiple versions of similar vectors by exchanging out individual parts in the assembly reaction. For example, when looking to create a vector with either a blue, red, or green fluorescent label, all I had to do was swap out the appropriate part (4a) while keeping the other remaining seven the same. These could then be simultaneously assembled within a singular PCR protocol. In total I was able to assemble three fluorescent variants of three CRISPR technology types (a/i/ko) of three different destination vectors for a grand total of 27 unique plasmids. An example TU for my CRISPRa plasmid is shown in Figure 2.2.1.5.

2.2.2 Stable integration of CRISPR vectors

Once the cloning of vectors was complete, I moved on to stable integration via lentiviral transduction. This did not work well. Likely because of the size of the inserts being used, efficiency of both viral packaging as well as transduction was consistently low. As an alternative, and because I already had the vectors on hand from my modular cloning, I tried PiggyBac integration by dual transfection of a hyperactive PiggyBac recombinase alongside of a PiggyBac destination vector containing my gene of interest. The general principles and procedures of this are outlined in Figures 2.2.2.1 and 2.2.2.2. PiggyBac based integration proved to be significantly more efficient than viral transduction in every single case, and from then on was used as the standard.

CRISPRa TU
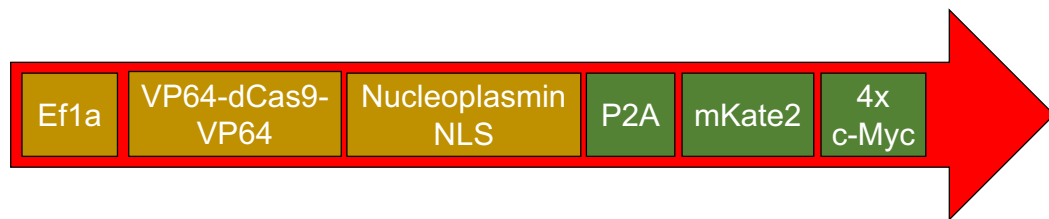
Figure 2.2.1.5 Example transcriptional unit (CRISPRa plasmid). This diagram of an example TU from a CRISPRa vector is composed of the following parts: A strong constitutive mammalian promoter (Ef1a), a transcriptional activator domain fused dCas9 (VP64-dCas9-VP64), a strong bipartite NLS (Nucleoplasmin NLS), a 2A peptide for multicistronic expression under a single promoter (P2A), a fluorescent reporter (mKate2), and a strong monomeric NLS (4x c-Myc). Together, this encodes for a targetable transcriptional activation system that also allows for efficient nuclear labelling for downstream cell counting applications.

PiggyBac System
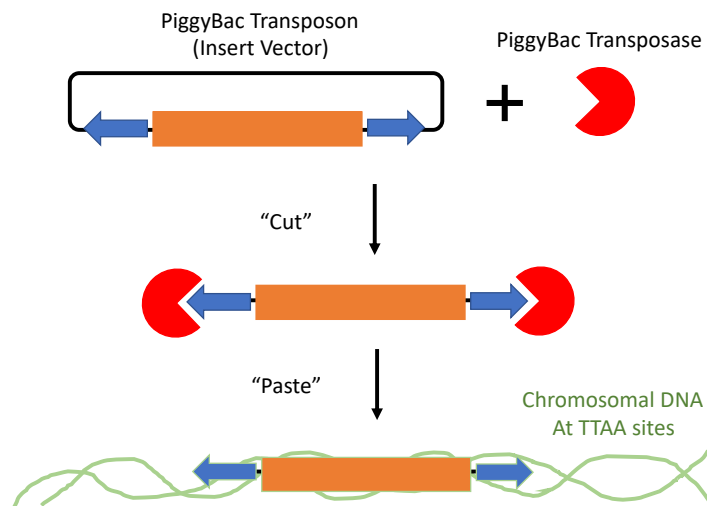
Figure 2.2.2.1 Diagram of PiggyBac integration system. The PiggyBac recombinase mediated integration system begins with dual transfection of an insert vector (flanked by PiggyBac transposable elements sequences) and either a plasmid containing a PiggyBac transposase expression system, or purified PiggyBac transposase. Once transfected, the PiggyBac transposase excises or "cuts" the intended insertion sequence from the insert vector and inserts or "pastes" it between genomic TTAA sites.

Figure 2.2.2.2 PiggyBac integration protocol

PiggyBac Integration Protocol (6-well)
By Clayton Wandishin

Materials:

pCMV_hyPBase plasmid (Amp Resistant)
XXXXXX_PB plasmid (Kan Resistant) (from Golden Gate Assembly)
Lipofectamine 3000 reagent
P3000 reagent
Opti-MEM medium (COLD)
6-well culture plate

Method: (all amounts are based on a single well)
1. Seed cells at 70-90% confluency at transfection (90% is better) with 1-3 mL of appropriate medium (2 mL is usually fine) in a 6-well culture plate (roughly 1E6 cells per well if adherent)
2. Aliquot 2 x 125uL Opti-MEM medium (COLD) in to 1.5 mL Eppendorf tubes
3. Add 6 uL Lipofectamine 3000 to one tube of 125 uL Opti-MEM and vortex 2-3 seconds (Tube A)
4. Add 2500 ng of TOTAL DNA in appropriate ratio (1:2.5 or 1:5 [Transposase{pCMV…}:Transposon{XXXX_PB}]) to 125 ul of Opti-MEM (Tube B)
   a. **1:5** is **416 ng** pCMV_hyPBase and **2,084 ng** XXXX_PB
   b. **1:2.5** is **714 ng** pCMV_hyPBase and **1,786 ng** XXXX_PB
5. Add 5uL P3000 reagent to Tube B (Mix Well and DO NOT VORTEX)
6. Add Tube A to Tube B and gently swirl (DO NOT VORTEX and solution may start to occlude as lipid vesicles form [this is good!])
7. Incubate mixture for 30 minutes at room temperature
8. Add lipid mixture to cells drop-wise and incubate at 37 degrees for 2-4 days before analysis

Notes:

All PB vectors are hygromycin resistant since it is built in to the PiggyBac integration cassette. Most vectors are additionally Blasticidin or Puromycin resistant, as well as fluorescently marked by either mTagBFP, mAzamiGreen, or mKate2. To fully select by an antibiotic requires a minimum of 10 days of selection pressure until a stably expressing population exists, which can be further assessed by imaging (making sure all cells glow). Relative integration copy number can further be quantified downstream by qPCR (not always necessary).

However, although integration was now efficient and successful, promoter and nuclear localization signal issues began to become apparent.

2.2.3 Promoter and nuclear localization signal issues in CRISPR vectors

The promoter issue with the SCLC lines was observed when compared to the control line (HEK293FT) in regards to both brightness of the fluorescent signal, as well as the ability to undergo antibiotic selection. In each case, SCLC lines were severely underperforming when a CMV promoter was used. In order to solve this, alternative promoters such as Ef1a, Ef1a core, CAG, PGK, and hUbc were tested using a cytosolic mAzamiGreen (GFP) expression plasmid. It was found that Ef1a and CAG performed the best, followed by PGK, Ef1a core, and hUbc. However, all of them greatly outperformed CMV. Based on these results, all of the cloning plasmids were adjusted to use an Ef1a promoter for CRISPR machinery, and a PGK promoter for antibiotic selection. Once the promoter issues were solved, it became noticeable that certain fluorescent reporters were leaky in regards to nuclear localization. To address this, each of the vectors was edited to use a Nucleoplasmin (bipartite) NLS for the Cas9 machinery, and a custom 4x c-Myc NLS combined with an H2A fusion protein for the fluorescent nuclear labels. This resolved all observed issues with leaky expression outside of the nucleus (Cardarelli, Bizzarri, Serresi, Albertazzi, & Beltram, 2009; Hodel et al., 2006; Hu, Zhao, Zhang, Li, & Zu, 2018; Lu et al., 2021; Ray, Tang, Jiang, & Rotello, 2015).

2.2.4 Fluorescent assisted cell sorting (FACS) of SCLC

Once all cloning, expression, selection, and localization issues were resolved, the next step was to enrich the stable expressing population by FACS. This went poorly. For

unknown reasons, SCLC lines have a strong propensity to die before, during, and after the flow-sorting process. To increase viability from FACS, I designed a custom protocol wherein each cell line was pelleted, digested in TrypLE, re-pelleted, resuspended in culture media (-Phenol Red) containing 30% FBS, and strained through a filter before being placed in a flow cytometry tube. This entire process was done over the course of 10 minutes at room temperature, where immediately following preparation the sample was run for 60 minutes (collected in a tube NOT a plate) and promptly returned to warmed media supplemented with 30% FBS. This custom protocol worked much better and led to consistent recovery of cells post sorting.

2.2.5 Issues with SCLC culture conditions

After addressing hurdles during cloning, integration, and sorting there was still one more challenge in the SCLC cell line library, culture conditions. For reasons unknown to myself, SCLC cell lines are particularly sensitive to culture conditions. This includes factors such as cell density, aggregate size, time in culture, FBS percentage, and whether or not the media was fully exchanged during splitting. Each of these factors had at least a partial contribution to the overall health of the cells. In general, guidelines for optimal culture conditions were sparse, and the ones that were available often seemed incorrect or a "bare minimum" of nutrient mix for cells to survive temporarily. This was further exacerbated by the fact that, because many of the lines were originally established at different institutions, there was a variety of media indicated for culturing. To address this, and before optimizing any other factors, I set about unifying all of the cellular media to RPMI-1640 w/v 10% FBS and 1% PEN-STREP by way of stepwise replacement (e.g. 80/20, 60/40, 40/60, 20/80, 0/100) over

the course of five splits. Following this, I spent months with individual lines slowly

changing conditions to figure out what worked best. What I learned for each line is

briefly described below:

2.2.6 Optimized culture conditions of SCLC lines

2.2.6.1 CORL279

Suspension line, prefers to be monodisperse, kept at 70% for 2-3 days then split,

sensitive to pH changes (dies), if cultured long enough will start to adhere to the plate

(likely EMT transition).

2.2.6.2 DMS53

Adherent line, very slow growing (weeks) if not cultured in greater than 70% confluency,

when cultured at >70% confluent they can be split every 3 or 4 days, if cultured too low

DMS53 forms small "islands" of 8-10 cells then enters into some sort of idling state

where total confluency of the culture vessel will not be reached unless the cells are

trypsinized and redistributed. Culture conditions are improved if no more than 50% of

the original culture media is removed during a split, similar to primary cell culture.

Highly resistant to trypsinization and if split below 25% will die off.

2.2.6.3 DMS114

Adherent line, average rate of growth, prefers to not be split to <35% but will slowly

recover if done so, >50% confluency for optimal growth, can be split very 2-3 days, do

not split lower than 25%.

2.2.6.4 DMS454

Adherent line, very slow growing (weeks) if not cultured in greater than 70% confluency,

when cultured at >70% confluent they can be split every 4 or 5 days, if cultured too low

DMS454 forms small "islands" of 10-15 cells then enters into some sort of idling state where total confluency of the culture vessel will not be reached unless the cells are trypsinized and redistributed. Culture conditions are improved if no more than 50% of the original culture media is removed during a split, similar to primary cell culture. Highly resistant to trypsinization and if split below 30% will die off.

2.2.6.5 NCIH69

Suspension line with a commercially available adherent variant (H69v), average rate of growth, can be split every 2-3 days, prefers 60-80% confluency, prefers monodisperse culture and if clumps greater than 5-7 cells begin to form they should be gently broken up mechanically by pipetting. When clumps do form they tend to have the appearance of "torn fragments of fabric".

2.2.6.6 NCIH82

Suspension line that clumps into spheroids during culture, average rate of growth, can be split every 2-3 days, prefers 50-70% confluency, prefers "mildly clumped" culture of 15-20 cells per clump, larger clumps should be gently broken up not to the point of monodispersion. If cultured long enough, a portion of the cell population will adhere (likely EMT transition).

2.2.6.7 NCIH146

Suspension line that prefers monodisperse culture with occasional clumps of 3-5 cells, average rate of growth, can be split every 2-3 days, prefers >70% confluency, best to split by dilution of existing culture instead of pelleting and resuspension.

2.2.6.8 NCIH196

Highly adherent line, grows in an islet shape and is extremely clear, difficult to tell confluency, prefers the 60-70% confluency range, average rate of growth, can be split every 3-4 days, sensitive to pH (dies). Culture conditions are improved if no more than 50% of the original culture media is removed during a split, similar to primary cell culture.

2.2.6.9 NCIH446

Mixed suspension and adherent line that will repopulate the dynamic mixture even if only the suspension or adherent portion is moved to subsequent culture. Difficult to work with and judge confluency but prefers the 60-80% confluency range. Average rate of growth, can be split every 3-4 days.

2.2.6.10 NCIH524

Suspension line that prefers spheroid/clumped culture of 20-25 cells, average rate of growth, can be split every 3-4 days, prefers >70% confluency, best to split by dilution of existing culture instead of pelleting and resuspension. Clumps will need to be broken up if they become too large, there is a hypothesis within the field that these clumps contain a "necrotic core" and only the outer surface are still dividing. Because of this, viability is traditionally low.

2.2.6.11 NCIH526

Suspension line that prefers spheroid/clumped culture of 25-30 cells, average rate of growth, can be split every 3-4 days, prefers >70% confluency, best to split by dilution of existing culture instead of pelleting and resuspension. Clumps will need to be broken up if they become too large, there is a hypothesis within the field that these clumps

contain a "necrotic core" and only the outer surface are still dividing. Because of this,

viability is traditionally low.

2.2.6.12 NCIH841

Adherent line that prefers ~60% confluency, average rate of growth, can be split every

2-3 days, very easy to work with, can be split as low as 10%.

2.2.6.13 NCIH1048

Adherent line that prefers ~60% confluency, average rate of growth, can be split every

2-3 days, very easy to work with, can be split as low as 10%.

2.2.6.14 NCIH1930

Suspension line that prefers monodisperse culture, slow rate of growth, can be split

every 5-6 days, prefers >80% confluency, best to split by dilution of existing culture

instead of pelleting and resuspension, sensitive to pH changes.

2.2.6.15 SW1271

Adherent line, extremely "Mesenchymal Like", commercially available in "L-15 media"

which requires a 0% CO_2 environment but can be transitioned to RPMI. Very sensitive

to confluency and highly prefers growth in the 65-90% confluency range. Average

growth rate, can be split every 2-3 days, should not be split <50%. Culture conditions

are improved if no more than 50% of the original culture media is removed during a

split, similar to primary cell culture.

2.3 Methods

2.3.1 Cloning Methods

2.3.1.1 Gibson Assembly

Gibson assembly was accomplished by use of the NEB Gibson Assembly Master Mix (NEB #E5510S). In brief, oligos were designed using the NEBuilder online assembly tool to ensure correct overlap of homology arms. These oligos were then used to PCR out an insert and destination vector from various sources, before being ligated and assembled together in a 1 hour long 60°C thermal cycling reaction as prescribed by the kit instructions. For the specific purposes of this cell line library this was primarily utilized as a way to domesticate new genetic sequences into the MTK part system.

2.3.1.2 Oligonucleotide annealing

Individual oligos were often ordered as a cost-saving measure for acquiring novel genetic sequences for the MTK library. In brief, a gene fragment was broken up into 60-80bp oligonucleotide fragments and ordered piece-meal from Integrated DNA Technologies in both the forward and reverse directions, taking care to make sure their annealing profiles were staggered to create four base-pair sticky-end overhangs. Pairs of oligonucleotides (equimolar mixture) were annealed by dissolving them in nuclease free water and heating them in an Eppendorf tube to a temperature of 95°C for 5 minutes, before unplugging the machine and leaving it to slowly cool to room temperature over 2-3 hours.

2.3.1.3 Golden gate assembly

Golden gate assembly was accomplished primarily as part of the MTK workflow, by combining the appropriate parts in a 20 uL PCR tube and placing in a thermal cycler at (37°C for 5 min then 16°C for 5 min) cycled 34 times finishing with a final 5 min

ligation step at 60°C. Appropriate parts for assembly of a TU include 3uL T4 ligase, 1uL

BsaI-HFv2, 2uL T4 ligase buffer (10x), 1uL each of MTK parts plasmids 1-8, fill to 20uL

with nuclease free water. Appropriate parts for assembly of final vector include 3uL T4

ligase, 3uL Esp3I, 2 uL T4 ligase buffer (10X), 1uL of TU plasmid, 1 uL of vector

plasmid, 10uL nuclease free water.

2.3.2 PiggyBac integration

The process of PiggyBac integration is best described in figures 2.2.2.1 and

2.2.2.2. In brief, PiggyBac integration is accomplished through co-transfection of an

insert plasmid with TTAA recombinase sites flanking the gene of interest to be

integrated, alongside an expression plasmid coding for a hyperactive version of

PiggyBac recombinase. Once transfected, the recombinase is expressed and excises

the gene of interest from the insert plasmid, and pastes it randomly into the genome at

available TTAA site locations. Downstream screening either by FACS or antibiotic

selection is then used to select for cells where the integration was stable and

successful.

2.3.3 FACS preparation

In preparation for flow sorting, each cell line was pelleted, digested in TrypLE, re-

pelleted, resuspended in culture media (-Phenol Red) containing 30% FBS, and

strained through a filter before being placed in a flow cytometry tube. This entire

process was done over the course of 10 minutes at room temperature, where

immediately following preparation the sample was run for 60 minutes (collected in a

tube NOT a plate) and promptly returned to warmed media supplemented with 30%

FBS.

2.3.4 Cell culture

See section 2.2.6 for cell line specific culture details.

2.4 Discussion

Once the growth mechanics of each individual lines was taken in to consideration, all of the necessary integrations were performed and stable expressing lines were sorted and stored. In total, eleven parental lines were generated, each containing 1-3 fluorescent variants (mAzamiGreen, mRuby2, mKate2) of three separate CRISPR technologies (CRISPRa/i/koi). This library and vectors is available upon request from the Quaranta lab.

Two main factors contributed to the difficulty of the library construction. First, the plastic nature of SCLC in regards to its ability to evade the immune system as well as standard chemotherapies is likely the main driver in its resistance to genetic perturbation. It is interesting to note as well, that while in the literature PiggyBac integration is usually 10-15% efficient, the efficiency in SCLC was often orders of magnitude higher. This could perhaps be due to the molecular mechanism of the recombinase integration that is preferential to regions of the genome that are being the most highly transcribed. One could imagine that the strong plasticity characteristics of SCLC require near constant transcription of certain genes in order to be maintained, and by integrating in to these regions, stable integration is grossly preferred. The other main driver of difficulty was the complexity of the constructs themselves. Each were large and highly complex, and if this library were to be recapitulated in another cell type, it could be highly beneficial to pursue one-time knockouts instead of stably generated lines, for ease of implementation and cloning.

Chapter 3: Using real-time luminescence to quantify drug-response in small cell lung cancer

3.1 Introduction

Assessing cellular drug response across multiple cell lines and types is an integral component of modern cancer research. This is primarily done by taking a single cellular viability measurement before and after the addition of a drug across a range of concentrations in what is known as a "fixed-endpoint assay". These measurements are then used to produce a dose-response curve to assess efficacy and potency. However, fixed-endpoint assays contain a multitude of inherent biases such as the time delay effect (slow-acting drug bias), seeding density variability (T_0), exponential growth vs. percent viability (ratio changes based on how far out the endpoint is taken), cellular growth rate dependence, and the lack of ability to produce negative values (minimum efficacy of zero) that can result in inaccurate determinations of both efficacy and potency in a variety of scenarios, potentially mischaracterizing both effective and ineffective treatments (Harris et al., 2016a). A more robust alternative is to assess viability via a continuous metric. Continuous viability assays have gained substantial interest in the scientific community as they overcome the biases associated with a fixed endpoint and provide a more detailed representation of cellular drug response over time. Continuous viability assays are conducted by taking intermediate measurements across a given time interval, with short measurement intervals and extended time courses giving the most detailed information. While fixed-endpoint data yields a single number that can easily be used in dose-response curve generation, continuous assays generate multiple values, and thus require derivation to distill

responses across a time period down to a single value. Assays such as EZ-MTT address this most simply by taking the slope of the dataset for dose-response curve generation, while alternative approaches such as the GR metrics and DIP rate address it by expressing each individual data series as a ratio of the basal response (Fallahi-Sichani, Honarnejad, Heiser, Gray, & Sorger, 2013; Hafner, Niepel, Chung, & Sorger, 2016; Harris et al., 2016a). Continuous assays also have their own experimental hurdles that have prevented widespread adoption of the platform, such as requiring a live cell fluorescent label (direct cell counting), inefficient cell segmentation algorithms, and an inability to work well with suspension cell lines (limited by imaging ability).

Recently, a new continuous luminescence-based viability assay has been developed that indirectly measures the cellular reductive capacity through metabolic conversion of a pro-substrate to substrate (Figure 3.1.1). The novel low-toxicity and membrane permeable NanoLuc luciferase pro-substrate rapidly diffuses into cells and is converted to active substrate (Furimazine) primarily by NAD(P)H oxidoreductase, a ubiquitous and established enzyme in the cellular metabolic process (Altman, 1976; Bernas & Dobrucki, 2002; Berridge, Herst, & Tan, 2005; Berridge & Tan, 1993; Cory, Owen, Barltrop, & Cory, 1991; Duellman et al., 2015; England, Ehlerding, & Cai, 2016; Hall et al., 2012; Riss et al., 2004). Once the substrate is generated, binding to the luciferase and subsequent enzymatic cleavage produces luminescence. These luminescence values correlate well with cell counts in static measurements (Figure 3.1.2) suggesting that this system could also be used for continuous luminescence measurements as an alternative to obtaining proliferation rates by direct cell counting. This is especially promising for suspension cell cultures, where direct cell counting is
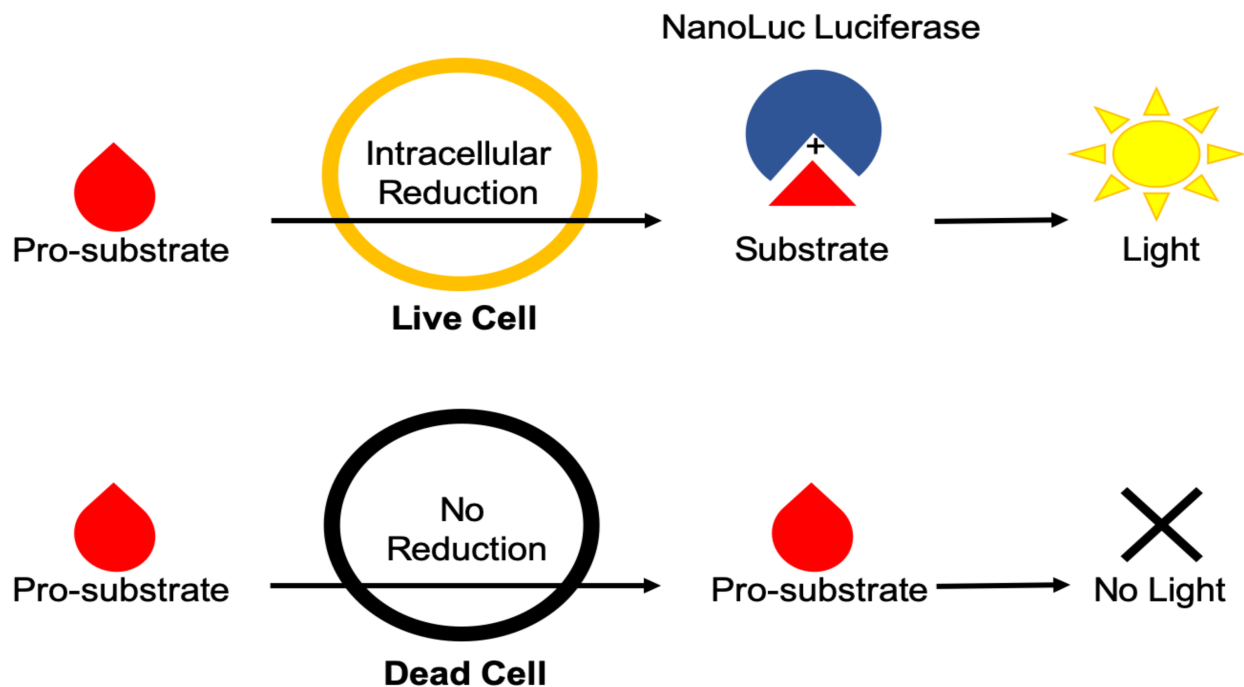
Figure 3.1.1. Diagram of real-time luminescence dynamics. Pro-substrate added to the culture media is rapidly metabolized by live cells via intracellular reduction into active substrate. The active substrate then reacts with NanoLuc luciferase to produce light. Dead cells are not able to metabolize the pro-substrate and therefore do not contribute to the amount of active substrate produced and subsequent light generation within the assay.

Adherent Cell Lines

Suspension Cell Lines
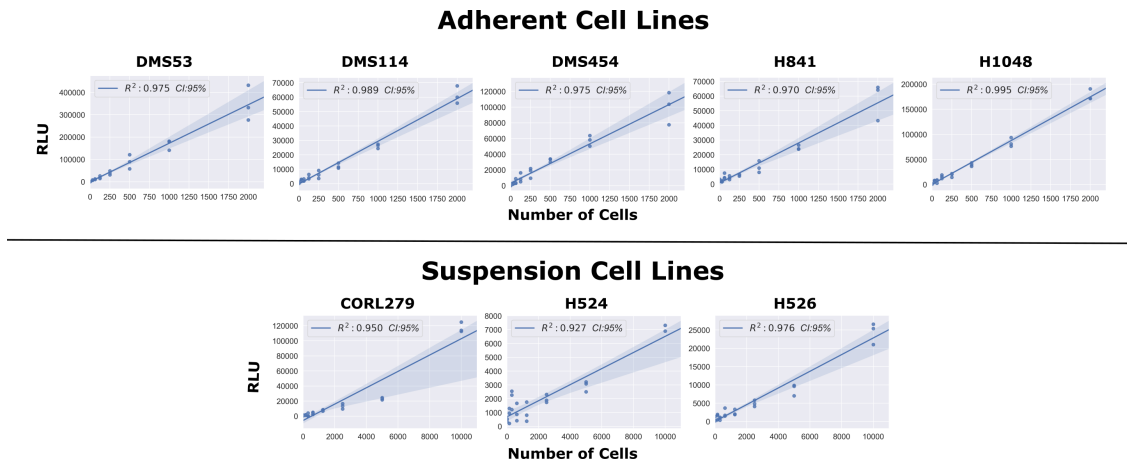
Figure 3.1.2 Static luminescent signal is correlated well to cell count. A range of cell lines were serially diluted by a factor of 2 from either 10,000 cells (suspension lines) or 2,000 cells (adherent lines). Assay reagents were then added to the wells and the plate was allowed to equilibrate for 1 hour. The luminescence measurements were then obtained, with the above graph showing the regression values among the static measurements of luminescence compared to varying cell seeding densities.

often not a feasible option. Here, we show that by modifying and optimizing the commercial assay protocol for single reagent-addition, the continuous luminescence data can be used as an alternative for direct cell counting measurements. Briefly, by focusing on the rate of luminescence change in drugged cell conditions and normalizing to the basal rate of change in an undrugged population, the continuous luminescence data can be reduced to a single value, reflecting the expansion and contraction of the cell population in response to drug.

This streamlines the quantification of the response to the level of a fixed-endpoint assay, while remaining continuous in origin (Harris et al., 2016a; Hsieh et al., 2017; Isherwood et al., 2011; Riss et al., 2004). Furthermore, we addressed challenges in the data interpretation by developing a freely available open-source analytical process (coding algorithm). Overall, using continuous luminescence to measure cellular drug response allows quantification regardless of cells being in suspension or adherent culture.

3.2 Results

3.2.1 Optimizing the commercial assay for single reagent-addition continuous experiments

In order to utilize the commercial NanoLuc luciferase assay for continuous experiments, we adjusted the supplied protocol. After testing of a variety of conditions addressing NanoLuc enzyme concentration, MT substrate concentration, Solubilization temperature and duration, cell seeding density, and confluency of culture prior to experimentation (data not shown), the following tenets were obtained. First and foremost, the optimal reagent preparation was found to be 20 uL of both the NanoLuc enzyme (1000X supplied)

and the MT substrate (1000X supplied) dissolved in to 25mL of culture medium supplemented with 10% FBS. We found the solubility of the MT substrate specifically, to be highly dependent on temperature.

During optimization, it was observed that the assay was more sensitive to temperature fluctuations during reads than previously anticipated. In order to address this, travel time between the plate incubator and reader was reduced to a minimum, and an additional incubation delay within a pre-warmed reader was added. The resulting optimized protocol based on these findings is available at https://github.com/QuLab-VU/RT-Glow/tree/master/RT-Glo%20Paper and **Figure 3.2.1.1**.

3.2.2 Comparing luminescence to direct cell counts in proliferating cell populations

We first confirmed the relationship between luminescence signal and cell number by comparing luminescence readings and direct cell counts in cultured wells with predefined numbers of cells (Figure 3.1.2, and section 3.4.3). To this end, we took luminescence reads across serially diluted cell concentrations after addition of assay reagents followed by one hour of equilibration. These static, single time-point measurements revealed a strong linear correlation between luminescence signal intensity and cell number in five adherent and three suspension cell lines (Figure 3.1.2). These results suggested that it is possible to monitor cell proliferation via luminescence in continuous culture over time, as a substitute for the more laborious direct cell count sampling.

To test the feasibility of continuous luminescence as an alternative for direct cell counting, we cultured multiple adherent cell lines (see 3.4 Methods) and took both luminescence and direct cell counts every 4 hours for 100 hours (Figure 3.2.2.1).

Figure 3.2.1.1 Optimized real-time glo protocol

Reagents required:

RealTime-Glo MT Cell Viability Assay (Promega G9711)

NOTE: If obtaining the commercially available assay, and not a custom formulation (20uL aliquots), 2 complete RealTime-Glo kits will be used during this protocol.

Cells in log-phase culture

Culture Medium

Sytox or CellTox Green (ThermoFisher S7020 or Promega G8741) *optional*

Reagent Preparation:

1. Prepare MT Viability Substrate and NanoLuc Luciferase from kit by moving from -80˚C or -20˚C to room temperature under cover (light sensitive) for at least 70 minutes before moving to 37˚C to equilibrate for 10-30 min
2. Simultaneuously, prepare culture medium by warming to 37˚C
   a. IT IS EXTREMELY IMPORTANT THAT THE MT SUBSTRATE, NANOLUC LUCIFERASE, AND CULTURE MEDIUM ARE ALL AT 37˚C DURING SOLUBILIZATION TO AVOID CRYSTALLIZATION OF MT SUBSTRATE, ADDITIONALLY DURING THE 37˚C INCUBATION STEP, CAREFUL ATTENTION SHOULD BE PLACED NOT TO LET THE SOLVENT OF THE MT SUBSTRATE EVAPORATE DUE TO EXTENDED INCUBATION. IF THE MT SUBSTRATE APPEARS HIGHLY VISCOUS EXCESS SOLVENT HAS LIKELY EVAPORATED AND THE SOLUBILIrATION PROCESS WILL SUFFER GREATLY. DEPENDING ON WHETHER OR NOT A WATER OR DRY BATH IS BEING USED FOR INCUBATION THE 10-30 MIN INCUBATION TIME MAY NEED TO BE SHORTENED IN ORDER TO PREVENT THIS.

NOTE: The following volumes are to create a 1X RT-Glo medium. Up to a 5X concentration can be prepared if desired and the prepared medium is stable for up to two weeks from preparation at 4˚C (including assay length) but will require re-warming and vortexing to completely re-dissolve components.

3. Pipette 25mL culture medium in to a 50 mL Falcon tube
   a. Optional: Supplement with 20uL 1000X Sytox Green or CellTox Green
4. Pipette 950uL culture medium from the 50 mL falcon tube and slowly use it to resuspend the MT Substrate (10uL) [1000x] by gentle pipetting up and down within the MT substrate tube before adding back in to the 50 mL Falcon Tube
5. Invert Falcon Tube 3-5 Times and repeat step 4 (20uL total of MT substrate will be added to 25mL of culture media)
6. Vortex the 50mL Falcon tube for AT LEAST 45 seconds
7. Place in Dry Bath at 37˚C until ready to use [4 hours max] or store at 4˚C for up to two weeks (including assay length)

Figure 3.2.1.1 optimized real-time glo protocol (continued)

Protocol:

1. Add cells at desired culture density appropriate for the assay length so that confluency is not reached to pre-warmed RT-Glo medium (supplemented with Sytox or CellTox green if desired)

   a. Alternatively, if cells have already been plated, aspirate medium and add 1X RT-Glo medium [Or aspirate a fractional portion of medium and add 1.5-5X RT-Glo medium]

2. Plate cells
3. Let plate incubate for 30 minutes at 37˚C before beginning to measure luminescence (and fluorescence if supplemented with Sytox or CellTox green)
4. Measure plate at desired intervals across the length of the assay, making sure that the measurements are done in a plate reader with the ability to incubate at 37˚C, this is Imperative for clean and accurate luminescence data

   a. Recommended: Add a minimum of 5 minutes delay within a pre-incubated plate reader so the plate has a chance to equilibrate before being read

   b. Assays requiring media change: If an assay length is such that it requires changing of culture medium, it is recommended that only a fractional medium changed is performed to reduce noise across the duration of the experiment.

   c. APPLICATION NOTE: For workflows utilizing "lid off" reading it is imperative that attention is paid to total evaporative loss for the duration of the assay. The addition of RT-Glo reagents can greatly speed up the evaporative process, leading to excess media loss, especially during extended periods of incubation with the plate lid off as can be common in the case of multi-read format experiments that include imaging as an aspect. e.g. If the plate workflow requires an imaging step, fluorescent read, and luminescence read, and the total time of all three is >5 min, replacing the lid for some or all of the reads can greatly decrease the total evaporative loss for the duration of the experiment.

      i. If medium must be exchanged do not exchange at an interval >72 hours if at all possible

5. Results may vary but if this is not possible, further optimization may need to be performed to ensure that MT substrate in culture is not depleted between medium changes (you might have to add more MT substrate initially)
6. When the assay is complete, determine the basal luminescence rate of the undrugged controls by taking the slope of the luminescence from the assay onset until either the end of the assay, or the maximum luminescent signal; whichever is first.
7. To determine the luminescent rates of drugged conditions, first constrain all drugged luminescent timepoints to the final timepoint used in determining the basal rate. Next, if the final timepoint is the maximum luminescent measurement for that condition, determine the rate by taking the slope. If the final timepoint is not the maximum, constrain the slope determination to the region between the maximum luminescence measurement and the endpoint of the basal luminescence rate. This can be done algorithmically by sampling every possible

Figure 3.2.1.1 Optimized real-time glo protocol (continued)

slice size of length N >= 4 from the determined final timepoint, and picking the slice with the greatest R^2 value. This ensures that the slope being used in the rate determination is indeed a stable drug-response for that condition.

    a. If Sytox or CellTox green were added, additional information on the cytotoxicity over time can be also analyzed. Namely, a cytostatic vs. cytotoxic drug-response can be inferred by a decreasing luminescence signal paired with a stable or increasing fluorescent signal respectively.

    NOTE: More nuanced cytotoxicity information can be obtained using the Sytox or CellTox as part of an imaging nuclear label akin to the process described in Harris et al.

Proliferation rates were then generated by taking the slope of both the raw luminescence and log transformed direct cell counting values and compared (Figure 3.2.2.1). The coefficient of determination (R^2) between the two proliferation rates was found to be greater than 0.92 in each.

Next, we took continuous luminescence measurements on suspension cell lines, where direct cell counting is not available, to assess if their luminescence remained linear for the duration of the experiment. Since linearity of luminescence signal is a requirement for straightforward analysis of continuous luminescence measurements (taking the slope) it was necessary to confirm this prior to using it as a metric for cell proliferation (See 3.4.4 Determining linear assay range). All three of the suspension lines tested (CORL279, H526, H1930) satisfied this requirement (Figure 3.2.2.2). Taken together, these results from both adherent and suspension cell cultures indicate that continuous luminescent measurements are a viable alternative to direct cell counting to assess cell proliferation over time.

3.2.3 Quantifying drug-response using continuous luminescence measurements

To explore the usefulness of the assay for continuous measurements of cell proliferation in response to drugs, we treated eight cell lines with several known anticancer agents and cultured them with the assay reagents for five days while taking luminescence measurements. Luminescence offers several advantages over conventional cell count assays (see Introduction and Discussion for more details), including speed and ease of execution and analysis for both adherent and suspension cell lines. By combining luminescence with drug-response data, continuous dose-response curves can be rapidly and efficiently generated by quantifying the rate of change
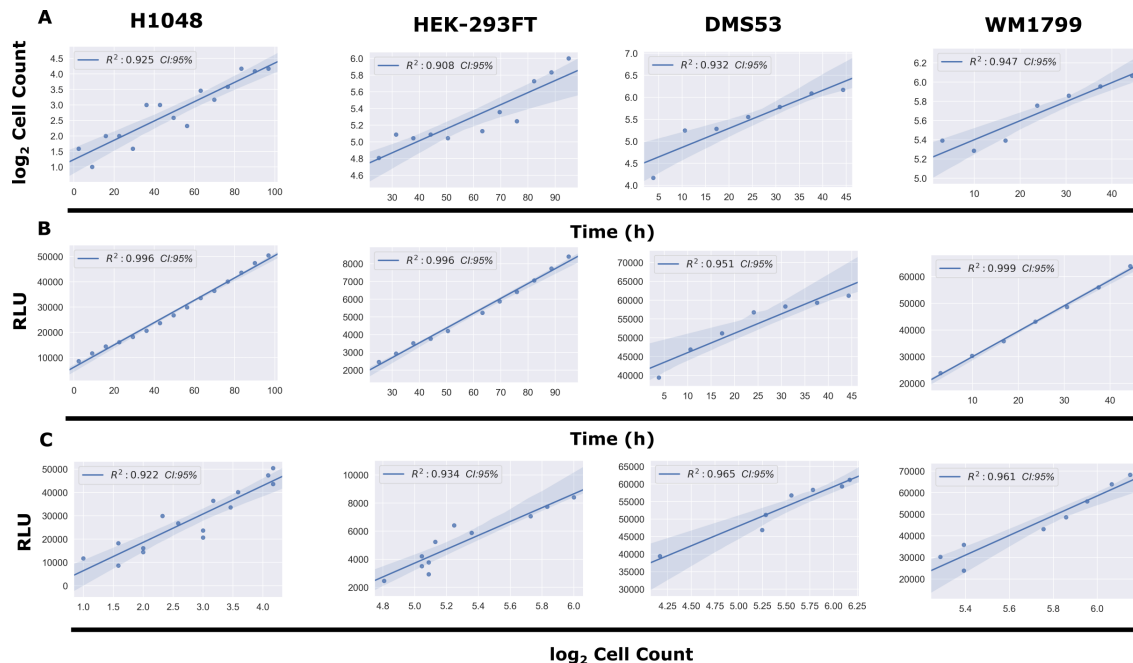
Figure 3.2.2.1 Continuous luminescence is correlated well to cell counts over time. (A) Comparison of the $\log_2$ transformed cell counts over time in four adherent cell lines. Cell counts were $\log_2$ transformed in order to linearize the data for subsequent comparisons. (B) Comparison of the continuous luminescent signal over time for the same four adherent lines from panel A. (C) Comparison of the correlation between continuous luminescent signal and $\log_2$ transformed cell count over time using a best-fit linear regression model. All conditions show R^2 correlation coefficients >0.92 .

CORL279 | **H526** | **H1930**

$R^2 : 0.941$ *CI:95%* | $R^2 : 0.912$ *CI:95%* | $R^2 : 0.957$ *CI:95%*

RLU

Time (h)
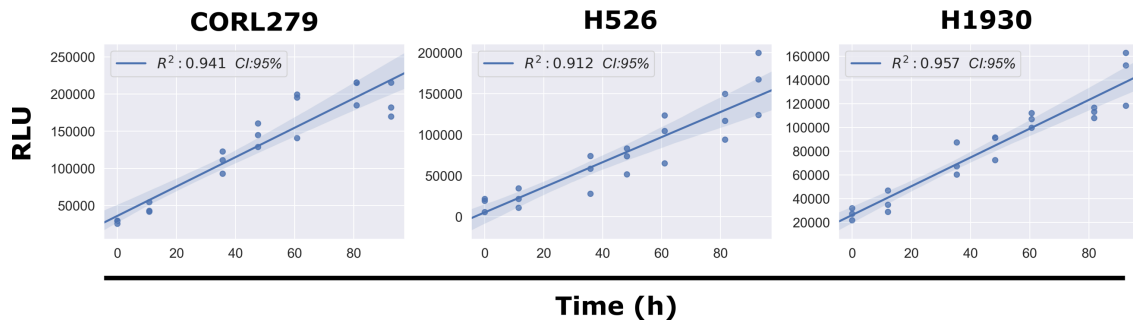
Figure 3.2.2.2 Continuous luminescence of suspension cell lines remains linear over time. A best-fit linear regression model of continuous luminescence in all suspension cell lines tested shows that minimum luminescent linearity requirements (R^2 >0.90) are met. Real-time luminescent signal maintains a sufficient linearity for the duration of the assay.

in luminescence (slope). Moreover, because luminescence measurements are an indirect quantification of every single cell within a well, the data gleaned from them is anticancer agents and cultured them with the assay reagents for five days while taking luminescence measurements. Luminescence offers several advantages over conventional cell count assays (see Introduction and Discussion for more details), including speed and ease of execution and analysis for both adherent and suspension cell lines. By combining luminescence with drug-response data, continuous dose-response curves can be rapidly and efficiently generated by quantifying the rate of change in luminescence (slope). Moreover, because luminescence measurements are an indirect quantification of every single cell within a well, the data gleaned from them is much more sensitive and less variable than taking direct cell imaging counts. This is most exemplified when comparing luminescence measurements to direct cell counts produced from imaging only a fraction of a given well (standard practice).

To generate rates from the continuous luminescence data, we took the slopes of the best fit linear regression lines of the raw luminescence data. An algorithm was developed to compare increasing slices of data points from the end of the assay (defined as peak luminescence in the control condition) by calculating an R^2 value for each slice, and using the highest R^2 value's linear regression slope as the basal rate for which subsequent drug dilution luminescence rates were normalized to. For drugged conditions, a similar process was used, but constrained to the region between the peak luminescence of the drugged condition, and the final timepoint of the assay determined by the peak luminescence of the control (Figure 3.2.3.1). Once the slopes of the
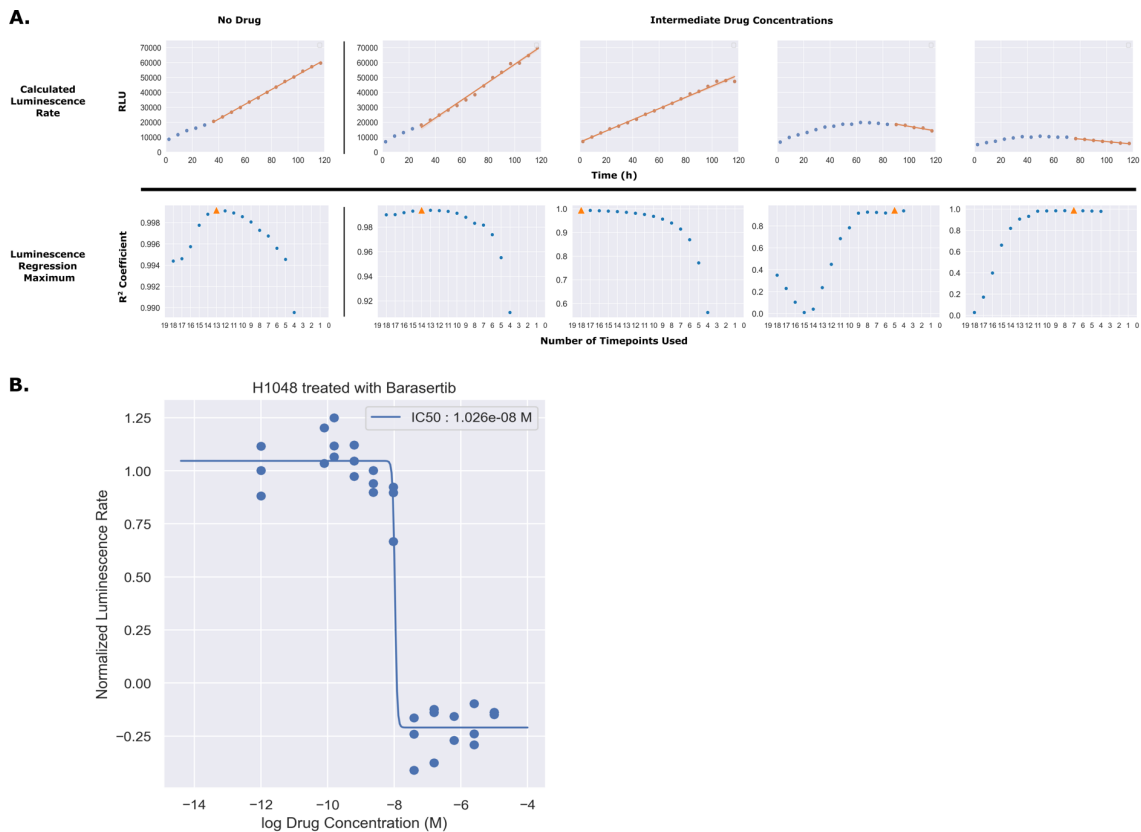
Figure 3.2.3.1 Slicing of luminescence data enables quantification of a steady maximum luminescence rate for dose-response curve generation. (A) Luminescence rates for each individual drug concentration were calculated by fitting the raw luminescence data to a linear regression model. For each concentration, the number of timepoints used in the regression (slice) was determined by calculating the R^2 for every possible slicing vector containing more than four points, originating from the end of the assay. The slice producing the maximum R^2 value is denoted in orange as a triangle. (B) To generate dose-response curves, each of the calculated luminescence rates was normalized to the luminescence rate in the absence of drug and plotted as a normalized rate in respect to the log of the drug concentration. These data were then fitted to a four parameter log logistic function.

Table 3.2.3.1. Extracted parameters from best-fit dose-response models

Cell Line	Drug	Data Type	Hill Coef	Max Resp	EC50	residuals
DMS114	barasertib	Lum	8.754	-0.079	7.86E-09	9.812
DMS114	AMG-900	Lum	0.544	-0.234	1.69E-09	9.254
DMS114	TAK-901	Lum	4.181	-0.074	1.6E-08	4.990
DMS114	YM-155	Lum	9.527	-0.034	6.91E-08	13.579
DMS114	SCH-1473759	Lum	0.485	-0.131	5.74E-09	7.134
DMS114	etoposide	Lum	14.618	-0.049	4.83E-07	16.446
DMS114	SNS-314	Lum	0.143	-0.564	2.88E-09	11.205
DMS454	barasertib	Lum	0.192	-5.000	733	17.492
DMS454	AMG-900	Lum	2.816	-1.841	1.17	23.981
DMS454	TAK-901	Lum	0.187	-4.834	6.25E-04	61.902
DMS454	YM-155	Lum	14.273	-0.480	7.6E-08	19.917
DMS454	SCH-1473759	Lum	0.490	-5.000	3.81E-05	30.889
DMS454	etoposide	Lum	10.439	-1.163	2.99E-06	305.999
DMS454	SNS-314	Lum	0.535	-0.511	1.44E-08	70.104
CORL279	barasertib	Lum	0.482	0.143	5.13E-09	2.275
CORL279	AMG-900	Lum	0.047	-0.414	5.44E-09	4.109
CORL279	TAK-901	Lum	1.027	-0.114	4.42E-08	1.755
CORL279	YM-155	Lum	0.572	-0.172	2.83E-09	8.129
CORL279	SCH-1473759	Lum	0.313	-0.391	1.55E-09	4.560
CORL279	etoposide	Lum	0.088	-0.967	5.5E-09	19.581
CORL279	SNS-314	Lum	0.000	-1.608	65.5	5.726
H1930	barasertib	Lum	0.396	0.134	2.57E-08	4.891
H1930	AMG-900	Lum	0.950	0.500	7E-10	6.743
H1930	TAK-901	Lum	1.350	0.080	9.01E-08	6.072
H1930	YM-155	Lum	1.528	-0.253	2E-09	48.869
H1930	SCH-1473759	Lum	0.318	-0.550	5.59E-07	3.666
H1930	etoposide	Lum	0.126	-1.089	3.37E-09	7.514
H526	barasertib	Lum	2.818	-1.859	1.15	93.562
H526	AMG-900	Lum	2.821	-1.887	1.11	144.650
H526	TAK-901	Lum	14.720	0.402	2.39E-06	23.102
H526	YM-155	Lum	0.299	0.084	5.12E-08	9.928
H526	SCH-1473759	Lum	13.750	0.101	2.47E-06	24.866
H526	SNS-314	Lum	0.000	-0.103	2.18E-09	101.411
H1048	barasertib	Lum	3.431	-0.209	1.44E-08	10.332
H1048	hygromycin_b	Lum	0.577	0.635	1.1E-04	13.097
H1048	trametinib	Lum	1.054	0.238	1.37E-07	12.033
H1048	SCH-1473759	Lum	1.333	-0.219	2.86E-09	17.029
H1048	YM-155	Lum	0.995	-0.186	2.74E-09	4.925
H1048	TAK-901	Lum	2.508	-0.209	1.45E-08	17.176
H1048	SNS-314	Lum	2.507	-0.069	4.95E-09	14.051
H841	barasertib	Lum	0.520	-0.427	3.96E-09	44.403

H841	hygromycin_b	Lum	2.817	-1.851	1.16	17.366
H841	trametinib	Lum	0.799	-0.281	2.1E-08	17.401
H841	SCH-1473759	Lum	3.777	-0.363	4.17E-09	59.188
H841	YM-155	Lum	0.999	-0.794	9.79E-09	86.686
H841	TAK-901	Lum	2.342	-0.380	1.45E-08	13.628
H841	SNS-314	Lum	2.465	-0.377	8.42E-09	14.323
DMS53	barasertib	Lum	2.817	-1.855	1.15	1395.696
DMS53	hygromycin_b	Lum	2.717	0.857	5.28E-08	117.582
DMS53	trametinib	Lum	1.032	-0.659	1.38E-08	85.699
DMS53	SCH-1473759	Lum	0.201	-4.430	2.17E-03	47.346
DMS53	YM-155	Lum	0.483	-0.696	2.89E-09	46.906
DMS53	TAK-901	Lum	0.203	-5.000	4.42E-03	45.003
DMS53	SNS-314	Lum	0.781	-5.000	1.54E-04	138.641
DMS53	vemurafenib	Lum	0.230	-5.000	5.49E-03	43.273
H1048	SNS-314	Direct	0.339	-1.673	1.31E-08	3.812
H1048	trametinib	Direct	1.009	0.137	2.29E-08	9.877
H1048	SCH-1473759	Direct	0.601	-1.588	3.32E-09	2.783
H1048	YM-155	Direct	0.730	-1.703	3.01E-10	22.686
H1048	TAK-901	Direct	1.692	-1.294	2.11E-08	1.317
H1048	barasertib	Direct	0.322	-2.742	6.29E-07	6.424
H841	SNS-314	Direct	0.550	-0.351	1.12E-09	1.208
H841	trametinib	Direct	2.111	0.588	5.46E-07	0.124
H841	SCH-1473759	Direct	1.312	-0.072	5.33E-09	3.166
H841	YM-155	Direct	1.216	-0.818	2.61E-10	4.012
H841	TAK-901	Direct	2.328	-0.633	6.33E-08	0.634
H841	barasertib	Direct	0.442	-5.000	1.55E-05	0.544

continuous luminescent signals were obtained, they were normalized and plotted against the drug concentration series to obtain dose-response curves (Table 3.2.3.1). In comparing dose-response curves generated from luminescent or direct cell counting data; overall fitting, data variation, and EC50 values were broadly found to be in agreement (Figure 3.2.3.2 and Table 3.2.3.2). Across both suspension and adherent cell lines, dose-response curves from luminescence-based rates were generated successfully. The code and associated data are freely accessible in this github repository, https://github.com/QuLab-VU/RT-Glow/tree/master/RT-Glo%20Paper and Appendix 1.1.

3.2.4 Mathematical comparison of direct counted and luminescence based dose- response curves

Mathematical comparison of dose-response curve fits is a common task within the biological community, however the most correct way to do so is often discussed (Carpenter, 1986; Meddings, Scott, & Fick, 1989). Because dose-response curves are traditionally generated using either a three or four parameter log-logistical model, the variation among each individual parameter as well as the errors for some parameters being normally distributed in linear space, while others are normally distributed in logarithmic space, often contributes to nearly qualitatively identical curves being found to be statistically significantly different. Moreover, when comparing dose-response metrics garnered from entirely different methods, differences in scaling of the raw data can further exacerbate this issue. Because of this, often the most useful comparison between curves is through their respective E_{max}, E_{min}, and EC/IC$_{50}$ values.

When comparing dose-response curves derived from luminescence measurements and those derived from cell counting measurements, an E_{max} comparison

is nonsensical. This is because all dose-response curves in our analysis are fitted using a three parameter log-logistical function that defines the E_{max} value as 1, so it is identical in every case. For E_{min} comparison, which is the comparison of maximal drug effect as the raw data for drugged conditions is represented as a normalized percentage of the undrugged metric, it is not appropriate to analyze the differences statistically as the scales and magnitudes of the raw data are vastly different. While initially it would seem as though both cell count and luminescence are measured in continuous integers, this is not the case. The individual cell counts are derived from counting the total number of fluorescently labelled nuclei (red) within an image, and subtracting the number of dead cells labelled by a fluorescent nuclear label in another channel (green) from that total. This then gives the total number of live cells, which is plotted over time to obtain a rate for the dose-response curve. The biological method of staining these dead cells uses a fluorescent dye which only permeates the nucleus when membrane integrity is compromised. This membrane integrity can be thought of as a boolean trait where each cell is either alive with an intact membrane (+1 to cell count) or dead without membrane integrity (+0 to cell count). Luminescence data differs from this greatly, as it measures overall production of light across the entire population within a well related to the metabolic activity of the cells. In theory, this helps to capture the partial luminescent production of cells as they are actively dying, but have not yet lost membrane integrity to be considered dead by cell counting. As an analogy, cell death measured by counting assumes mechanics like that of an on/off light switch, while cell death measured by luminescence allows for both boolean and continuous death mechanics, more akin to a dimmer switch. Since the two methods are not of equal scaling, neither are their

variances, making direct statistical comparison ill-advised. Moreover, because the magnitudes of each method are so different between single time steps, likelihood of a direct statistically significant difference is also increased. Because of this, the decision was made to base the comparisons of the dose-response curves primarily on their EC50 value, followed by whether or not the E_{min} values were in agreement on drug effect (Antiproliferative, Cytostatic, Cytotoxic, or Null).

For the statistical analysis of EC50 values, the decision was made to perform the most conservative paired sample analysis with the least assumptions, the Wilcoxon signed-rank test. The results of this showed that among 12 paired EC50 samples (six drugs across two cell lines), a Wilcoxon value of 31 was obtained with a p-value of 0.5693359 (Figure 3.2.3.2, Panel A). For the EC50 sample size (N=12) a Wilcoxon value of 31 exceeds the two-tailed critical value threshold of 13 ($\alpha = .05$) and we fail to reject the null hypothesis that the EC50 value pairs are not significantly different. Therefore, whether the EC50 was obtained through direct counting measurements or luminescence did not make a significant difference in its value across the cases tested here.

To further assess the congruency of dose-response curve generation between luminescence derived datasets and those from direct counting, the E_{min} values for each cell line and drug pairing were compared based on whether or not the E_{min} values were positive (anti-proliferative/cytostatic), negative (cytotoxic), or equal to 1 (no drug effect). This is an important observation to make if this type of assay were to be utilized as a first pass drug screening application in order to correctly discern cytotoxic lead compounds from those that are anti-proliferative/cytostatic, or have no effect at all. From 14 paired samples, 13 pairs were found to be in agreement on drug effect mechanism while only
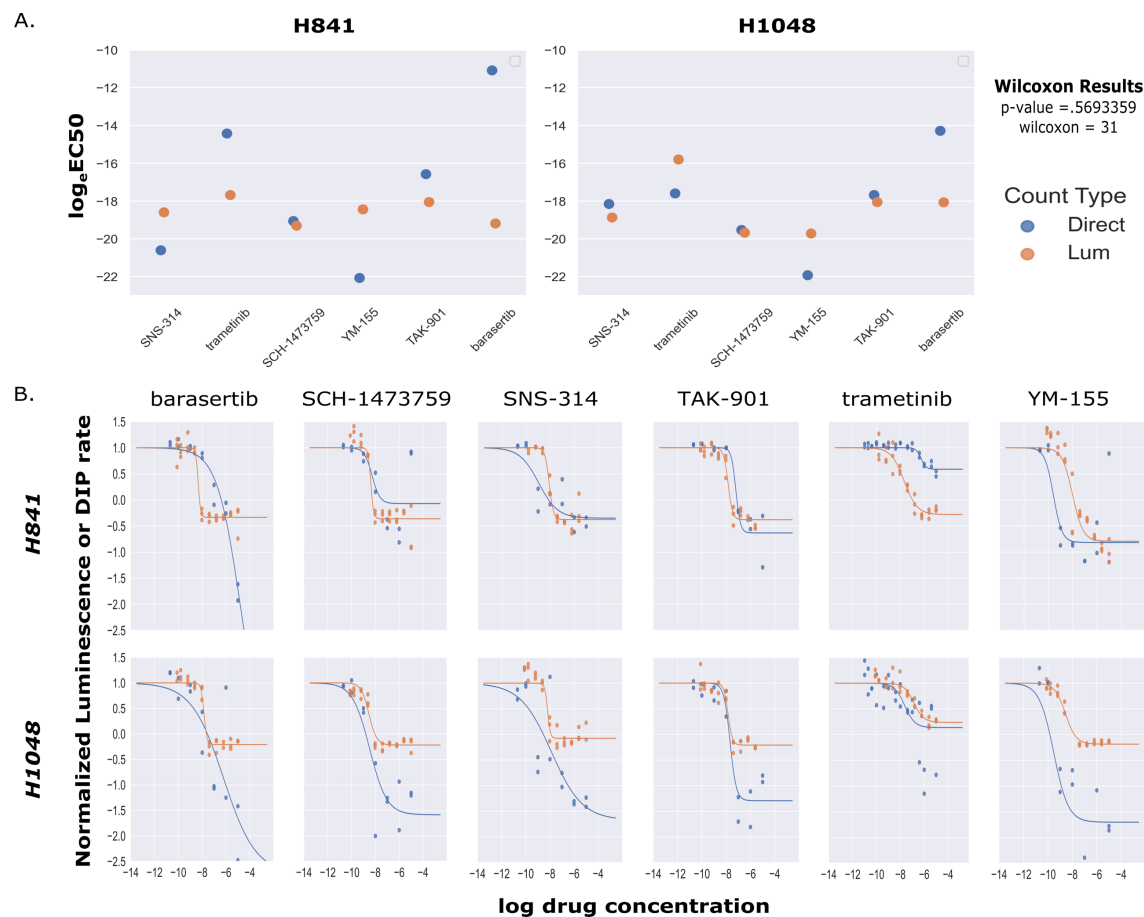
Figure 3.2.3.2 Comparison of EC50 values and dose-response curve fits between luminescent and direct cell counting measurements shows congruency among the two methods. (A) Scatter plot comparison of calculated $\log_e$EC50 values for both luminescence based and direct cell counting measurements. Across all paired values tested, there was no significant difference between luminescence based $\log_e$EC50 values and those obtained from direct cell counting (Wilcoxon Signed Rank Test, p-value=0.569, W=31, N=12). (B) Comparisons of dose-response curves generated by either luminescence (orange) or direct cell counting (blue) for two cell lines across a panel of six drugs.

Table 3.2.3.2 E_min drug effect mechanism shows congruency between luminescence and direct counting dose-response curves

Cell Line	Drug	Lum Emin	Direct Emin
H841	Barasertib	-0.335	-0.171
H841	SCH-1473759	-0.364	-0.069
H841	SNS-314	-0.373	-0.182
H841	TAK-901	-0.374	-0.634
H841	Trametinib	-0.271	0.586
H841	Vemurafenib	null	null
H841	YM-155	-0.804	-0.727
H1048	Barasertib	-0.21	-0.772
H1048	SCH-1473759	-0.209	-1.299
H1048	SNS-314	-0.079	-1.668
H1048	TAK-901	-0.191	-1.26
H1048	Trametinib	0.221	0.136
H1048	Vemurafenib	null	null
H1048	YM-155	-0.167	-1.47

An analysis of 14 pairs of E_{min} parameters across two cell lines and seven drugs shows congruency of drug effect mechanism in 13 out of 14 cases.

one pair (H841/trametinib) was found to lack congruency (Figure 3.2.3.2 Panel B, and Table 3.2.3.1).

3.3 Discussion

Here we have outlined the development and application of a non-lytic luminescence-based assay to extract rate-based metrics of drug response. Implementation of our analysis and workflow has the potential to greatly expedite and modernize large-scale screening and characterization of drug response in a variety of disease models and culture methods. This work has traditionally been accomplished using fixed-endpoint viability metrics, which contain a significant degree of inherent biases, ultimately leading to a large potential for mischaracterization of drug effect in a variety of indices, both positive and negative. We and others have shown the value in taking continuous measurements across the duration of an experiment at multiple timepoints (Fallahi-Sichani et al., 2013; Frick, Paudel, Tyson, & Quaranta, 2015; Hafner et al., 2016; Harris et al., 2016a; Riss et al., 2004; Uzunoglu et al., 2010). However, despite the clear advantages in data quality, adoption of continuous viability assays has been relatively slow, likely due primarily to the difficulties in integrating a continuous assay into an existing setup designed for fixed-endpoint measurements. Previously, we have described the DIP rate as an unbiased metric for drug proliferation when using direct cell counting. Our analysis of continuous luminescence utilizes the same mathematical ideology, while going one step further, with a protocol that is easily adaptable to existing fixed-endpoint workflows. What this means is, by changing only the reagent preparation method and data analysis pipeline, laboratories currently setup for drug screening using a fixed-endpoint protocol could rapidly pivot to a much more

quantitatively robust method with little to no adjustment of established automation. Our hope is that this additional analytical rigor at the basic science level could lead to fewer cases of therapeutic candidates failing to translate to higher order biological models.

Like any assay, NanoLuc luciferase based continuous luminescence does have its limitations, and suffers many of the same issues surrounding MTT/MTS based measurements such as potential overestimation of viability from active mitochondrion, and inability of use for drugs targeting redox pathways (Riss et al., 2004; Wang, Henning, & Heber, 2010). These features are hardly unique to this assay, and have been generally accepted in the field for quite some time (Butcher, 2005; Cory et al., 1991; Garnett et al., 2012; Mosmann, 1983; Riss et al., 2004). By structuring experiments to avoid these known factors, complex drug-response analysis can easily be simultaneously achieved across cell lines, independent of their morphology (Wang et al., 2010).

For cell lines that are able to maintain a linear trend in luminescence for the duration of an experiment (without drug), continuous luminescence measurements offer a simple and scalable option for generating dose-response curves. This is of particular interest for cell lines that are cultured in suspension, as direct counting of suspension line cultures is not currently feasible in most situations. Based on the results of our experimentation, we intend to further explore the utility of NanoLuc luciferase based luminescence by computationally modelling the dynamics of the system, potentially using luminescence rates to predict DIP rates, as well as testing its usefulness in quantifying drug-response in three dimensional cultures (organoids). Lastly, our most

51

immediate goal for this work is to showcase its utility with the successful integration into a high-throughput in vitro drug screening platform.

3.4 Methods

3.4.1 Cell culture

All cell lines were cultured for a minimum of two weeks prior to experimentation in T75 (Corning 430641U) flasks containing appropriate media (see below) at 37 °C and 5% CO_2. Additionally, prior to any experimentation, absence of mycoplasma was confirmed using a MycoAlert Mycoplasma Detection Kit (Lonza LT07-118).

3.4.2 Appropriate media

CORL-279, DMS53, DMS114, DMS454, H524, H526, H1048, H1930

RPMI 1640 medium (Corning 10-040-CV) supplemented with 10% FBS (Gibco 26140079) and 1% Pen-Strep (Gibco 15140122)

WM1799

DMEM/F12 medium (Gibco 11320033) supplemented with 10% FBS (Gibco 26140079), and 15mM HEPES (Gibco 15630080)

HEK293FT

DMEM medium containing 4.5 g/L glucose (Gibco 11965092) supplemented with 10% FBS (Gibco 26140079), and 1% Pen-Strep (Gibco 15140122)

3.4.3 Static luminescence measurements

Cells were cultured for two weeks, spun down, and resuspended at a density of 2.86E4 cells/mL in appropriate media, NanoLuc Enzyme (Promega E499A), and MT pro-substrate (Promega G971A). Each cell line was plated on to a 384 well GreinerOne Imaging plate (Greiner 781096) at a density of 2000 cells per well serially diluted across

10 wells (2000-4) with a total well volume of 70 uL in each. Additionally, each cell line was plated in triplicate. The plate was then incubated in a BioTek Synergy H1 at 37 °C and 5% CO_2 for 5 minutes before luminescence measurements were taken (lid on).

3.4.4 Determining linear assay range

Initial cell concentrations for the linearity range of the assay were determined by following the guidelines in the "Promega RealTime-Glo MT Cell Viability Assay Protocol Handbook" under subsection four, "Determining Assay Linearity for the Endpoint or Continuous-Read Format". Briefly, cells were serially diluted and plated with RT-Glo reagents, incubated for the proposed length of experiment (120 hours), while luminescence measurements were taken every four hours. Upon completion, the luminescence trend lines were analyzed by linear regression to find a suitable cell concentration that would maintain a linear regression coefficient of >.90 for the duration of the assay (Data Not Shown).

3.4.5 Continuous luminescence measurements

Cells were cultured for two weeks, spun down, and resuspended at a density of 4.39E3 cells/mL in appropriate media, 10 nM Sytox Green (Invitrogen S7020), NanoLuc Enzyme (Promega E499A), and MT pro-substrate (Promega G971A). Each cell line was plated on to a 384 well GreinerOne imaging plate (Greiner 781096) at a density of 300 cells per well across 10 wells with a total well volume of 70 uL in each. Additionally, each cell line was plated in triplicate. The plate was then incubated in a BioTek Synergy H1 at 37° C and 5% CO_2 for 5 minutes before initial luminescence and fluorescence measurements were taken (lid on). The plate was then stored at 37° Celsius and 5% CO_2 in an incubator. Every 12 hours, the plate was removed, left to equilibrate for 5 minutes

in the BioTek Synergy H1, and luminescence measurements were recorded. This continued for a total of 100 hours, at which time the plates were discarded.

3.4.6 Direct cell counting

To facilitate automated image processing, cells were engineered to express the monomeric red fluorescent protein mRuby2, integrated by dual transfection of a modified PiggyBac recombinase expressing plasmid and a custom mRuby2 containing transposon plasmid (Li et al., 2013; Yusa, Zhou, Li, Bradley, & Craig, 2011). Cells were seeded at 300 cells per well in 384 well GreinerOne imaging plates (Greiner 781096). DMSO (Sigma D8418) and phosphate-buffered saline (Corning 21-040-CV) were used as vehicle controls, as appropriate. Images were acquired through a 10x or 20x objective with a Cellavista HighEnd Bioimager (SynenTec Bio Services, Meunster, Germany) every 12 h as 3 x 3 or 5 x 5 montages for 120 hours. Image processing to obtain counts of cell nuclei at each time point was performed as previously described (Frick et al., 2015).

Chapter 4: Computationally modelling the relationship between real-time luminescence and cell count

4.1 Introduction

From the onset of using real-time luminescence to quantify drug-response in SCLC, the same critique was echoed time and time again by scientists firmly rooted in the exponential growth of cellular division, "If luminescence is related to cell count, and cells grow exponentially, why is the luminescence linear?" Personally, having done extensive work with luminescence-based assays, linear luminescence trends were something I expected. Additionally, cellular division in an exponential sense is more akin to a conceptual convenience (ala "particle in a box") than an infallible principle, as evidenced in experimental observations over the years (Anderson, Bell, Petersen, & Tobey, 1969; Cermak et al., 2016; Sinclair & Ross, 1969). I was eventually able to show that the strongest correlation between luminescence and cell count was between the linear luminescence and the log2 transformed counts, making it the most mathematically correct way to benchmark the relationship. Additionally, while a relationship between luminescence and cell count does exist, I did not believe it to be necessary to define that relationship in an exact sense in order to quantify drug-response (see Chapter 3). However, the conversations surrounding this phenomenon persisted long enough that the general consensus of the research group was that it should be explored and characterized, leading to the initial construction of the first systems biological model of real-time luminescence and cell count.

4.2 Results and Discussion

4.2.1 Initial modelling of real-time luminescence and cell count

The initial model for real-time luminescence and cell count was created using the PySB framework and coded in Python. It took into consideration the initial number of cells, those cells' rate of division, those cells' rate of death, an initial concentration of pro-substrate, an initial concentration of NanoLuc luciferase, the rate at which pro-substrate was converted to substrate by intracellular reduction, and the rate at which NanoLuc luciferase consumed substrate to generate luminescence. This is represented as an equation and with simulated data in Figures 4.2.1-4.2.2. Once a model was established, nC0 (minimum/initial luminescence) and m (slope of luminescence over time) were estimated from control wells, while Kdiv-Kdeath, Kdiv*-Kdeath*, and Koff/Kon were estimated using particle swarm optimization with a cost function based around the raw luminescence data. These parameter sets were then used to compare predicted cell counts to direct cell counting data, and iteratively refined. As a first pass, the models used were all able to produce results, however to varying degrees of success depending on both the cell line in question as well as the drug being used. This is illustrated across figures 4.2.3-4.2.14 when comparing between experimental and predicted values of luminescence and cell count. This cell line and drug specific modelling success and or failure is likely due to the different population dynamics of each of the lines, as well as varying degrees of labelling efficiency for the cell lines from which direct counting measurements were gleaned. A large variance in labelling efficiency leads to a large difference in image segmentation efficiency, subsequently resulting in significant variance in the cell count data over time. This variance in the cell

$$Cell \xrightarrow{k_{div}} Cell + Cell \qquad Cell + Drug \xrightarrow{k_{on}} Cell^* + Drug \qquad Cell^* \xrightarrow{k_{div^*}} Cell^* + Cell^*$$

$$Cell \xrightarrow{k_{death}} \emptyset \qquad Cell^* \xrightarrow{k_{off}} Cell \qquad Cell^* \xrightarrow{k_{death^*}} \emptyset$$
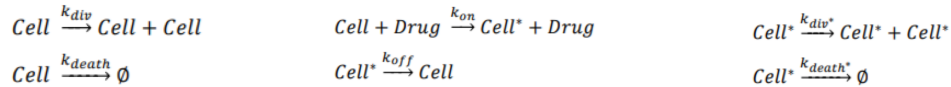
$$\ln \frac{Cell_T}{Cell_T(0)} = \frac{\frac{k_{off}}{k_{on}}(k_{div}-k_{death}) + Drug(k_{div^*}-k_{death^*})}{\frac{k_{off}}{k_{on}} + Drug} \cdot t.$$

Figure 4.2.1 Equations for cell population dynamics. These equations are all derived from the mathematical model for cellular division and drug response as proposed by Harris et al.. In brief (from left to right and top to bottom) One cell becomes two cells at a rate of k_{div}, A cell and drug becomes a drugged cell and drug at a rate of k_{on}, a drugged cell becomes two drugged cells at a rate of k_{div^*}, a cell becomes no cell (death) at a rate of k_{death}, a drugged cell becomes an undrugged cell at a rate of k_{off}, and a drugged cell becomes no cell (death) at a rate of k_{death^*}. This is all summarized as a single equation in linear space as a function of time in the final equation above.

57

$$l' = nC_0e^{kt}$$

*Assume Δ(t) is independent of drug concentration
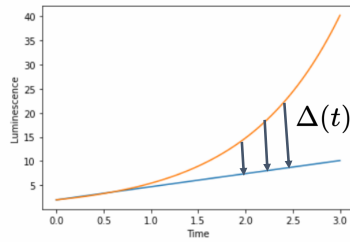
$$\Delta(t)$$

$$l = nC_0 + mt$$

$$l(d,t) = nC(d,t) - \left[nC_0(e^{kt} - 1) - mt\right]$$

Courtesy of Michael Quan

Figure 4.2.2 Simulated data comparing real-time luminescence and cell count dynamics suggests a correction factor can be applied to link the two. This diagram outlines predicted luminescence values if luminescence was directly correlated with cell count (orange) and compares that to what is observed experimentally (blue). The relationship between l' (idealized luminescence) and l (observed luminescence) can be modelled using a correction factor (Δ) in respect to time and is described mathematically above. (nC = # of cells, m = rate of change of luminescence over time, t = time)
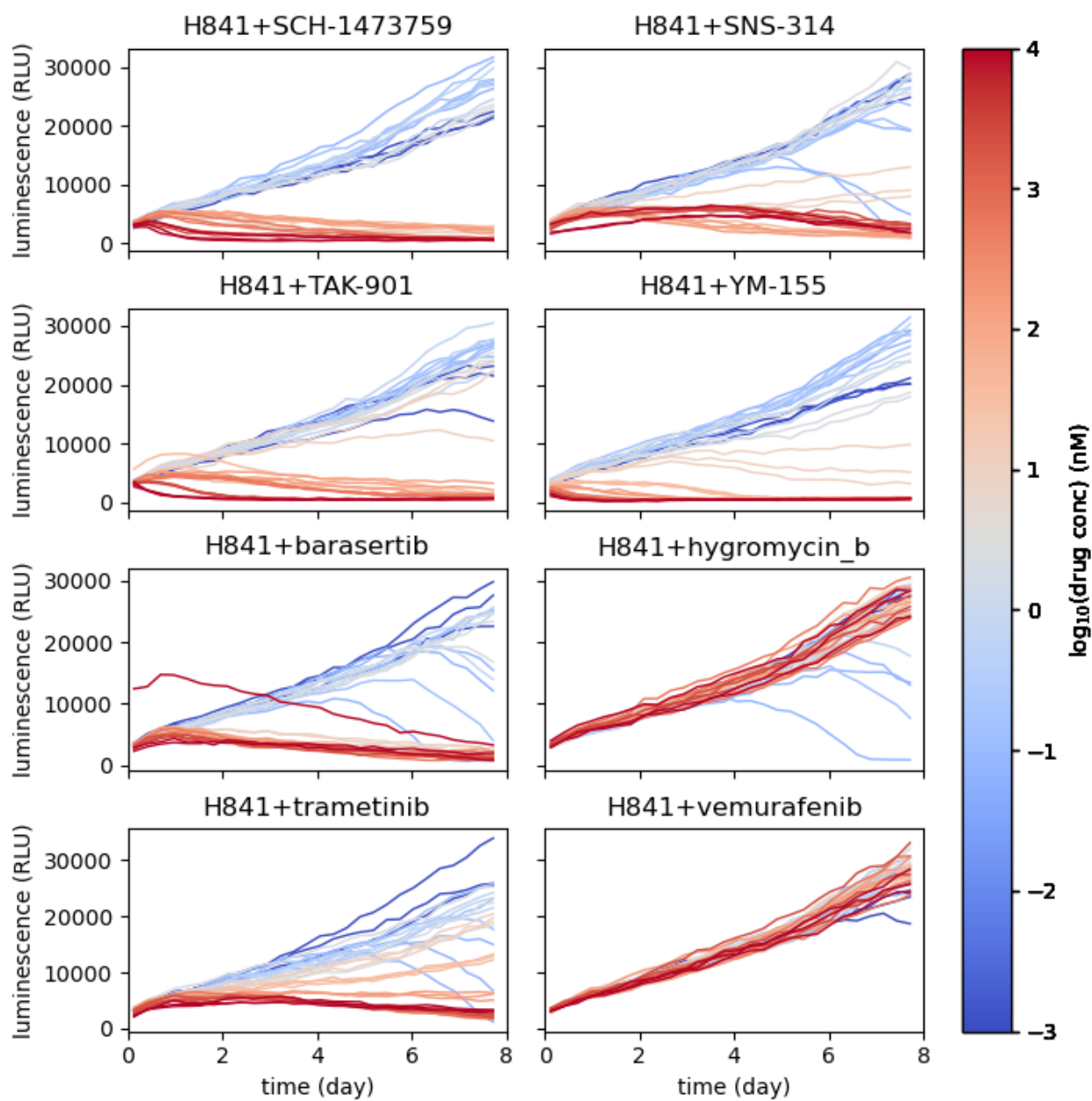
Figure 4.2.3 Real-time luminescence of H841 over time (EXPERIMENTAL).
Experimental luminescence data for eight drugs over the course of eight days were
generated for the H841 cell line (see 4.2.4 for modelling comparison).

Courtesy of L. Harris
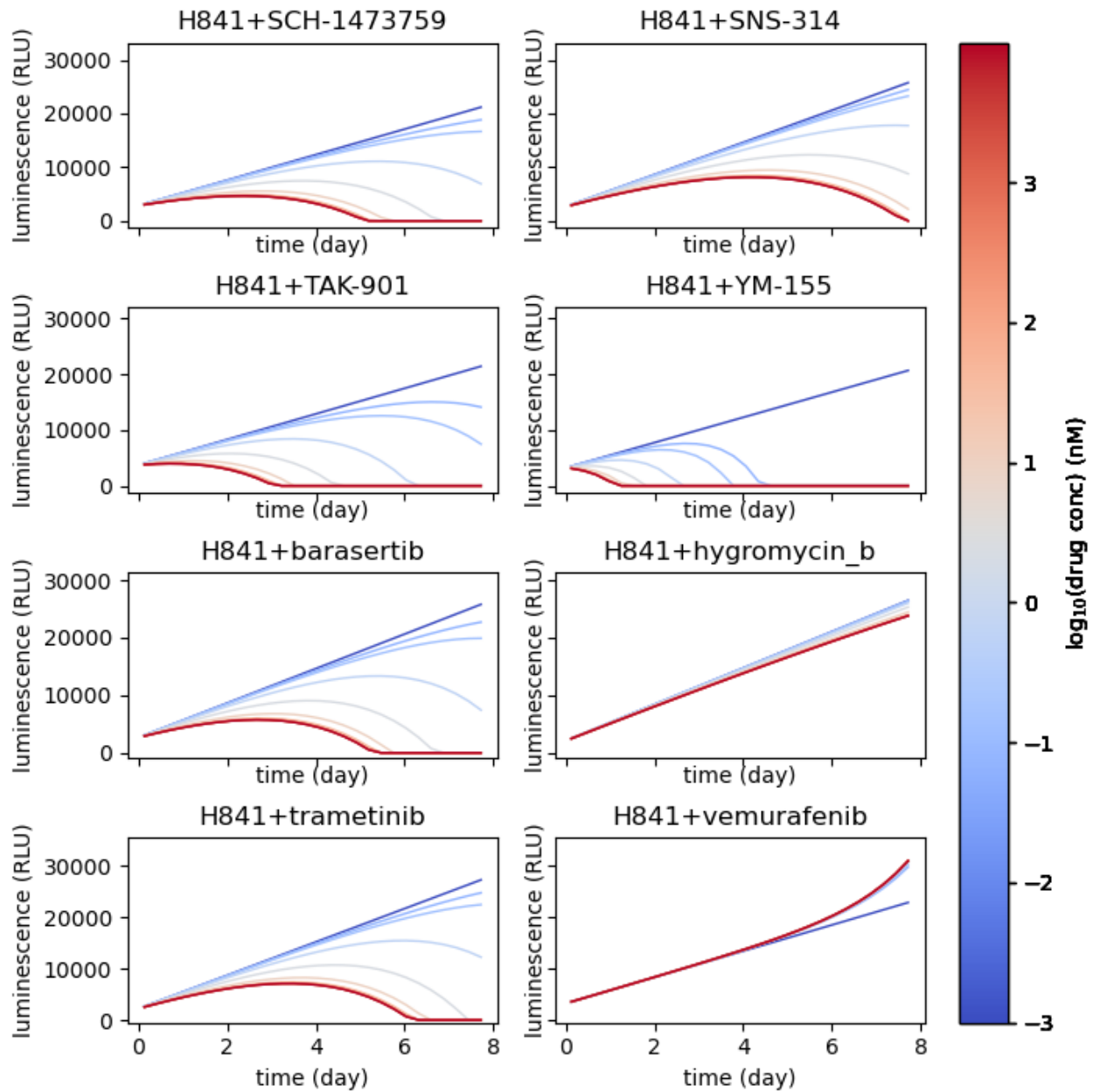
Figure 4.2.4 Modelled luminescence of H841 over time (SIMULATED). Simulated luminescence data for eight drugs over the course of eight days were modelled for the H841 cell line in PySB (see 4.2.3 for experimental comparison).

Courtesy of L. Harris
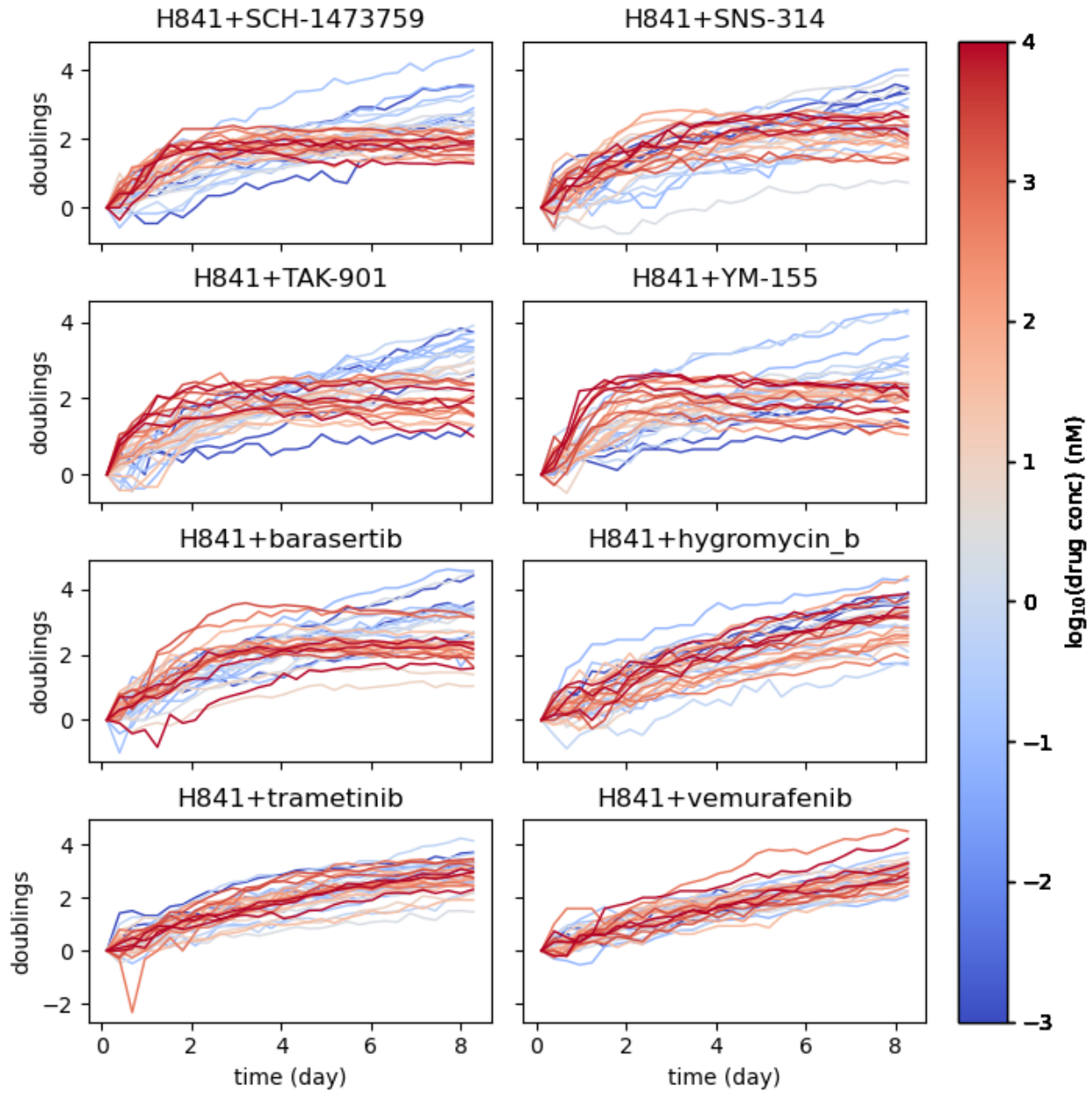
Figure 4.2.5 Direct cell counting of H841 over time (EXPERIMENTAL). Experimental cell count data for eight drugs over the course of eight days were generated for the H841 cell line (see 4.2.6 for modelling comparison).

Courtesy of L. Harris
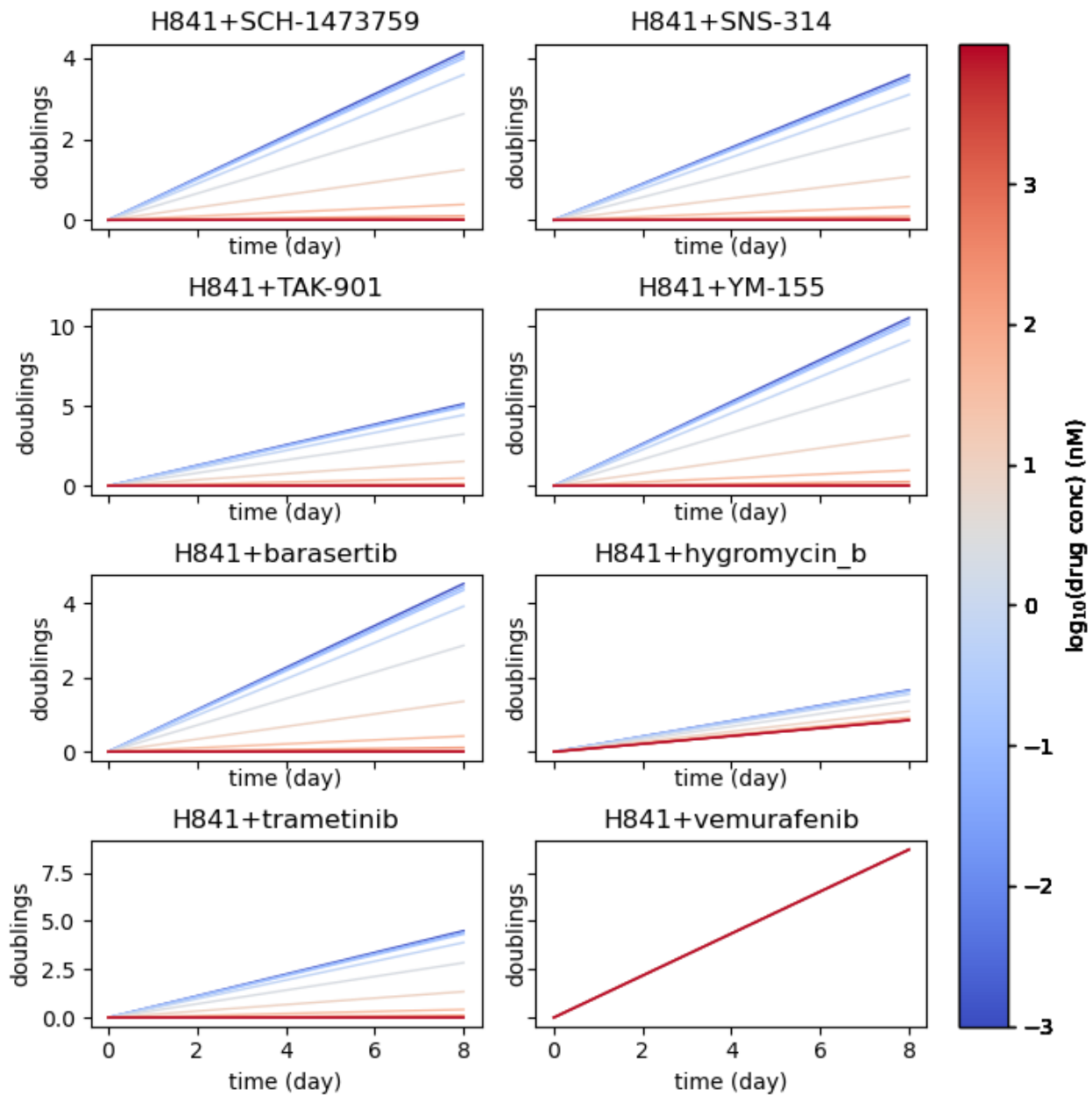
Figure 4.2.6 Predicted cell counts of H841 over time (SIMULATED). Simulated cell count data for eight drugs over the course of eight days were modelled for the H841 cell line in PySB (see 4.2.5 for experimental comparison).

Courtesy of L. Harris
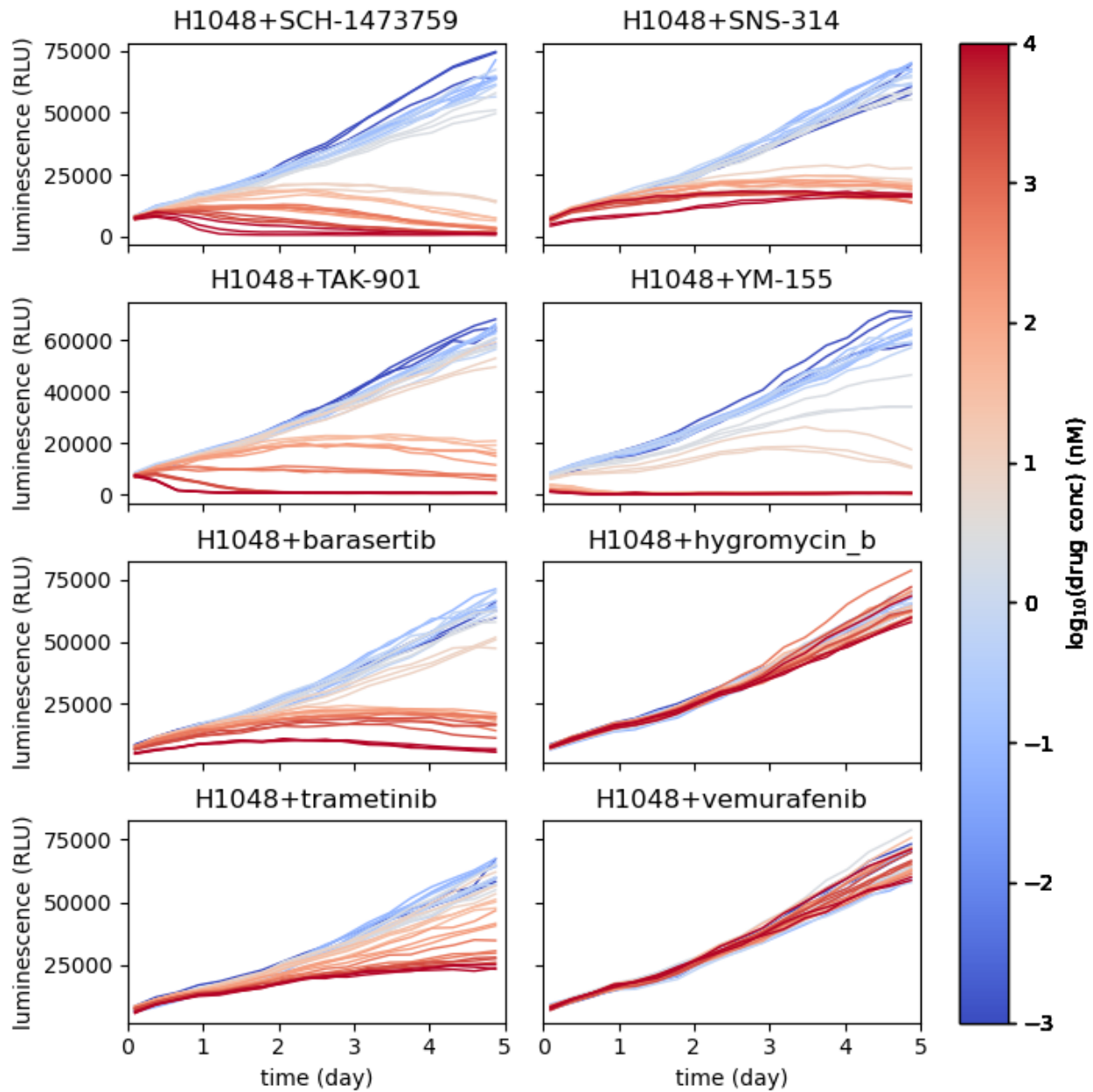
Figure 4.2.7 Real-time luminescence of H1048 over time (EXPERIMENTAL).
Experimental luminescence data for eight drugs over the course five days were
generated for the H1048 cell line (see 4.2.8 for modelling comparison).

Courtesy of L. Harris
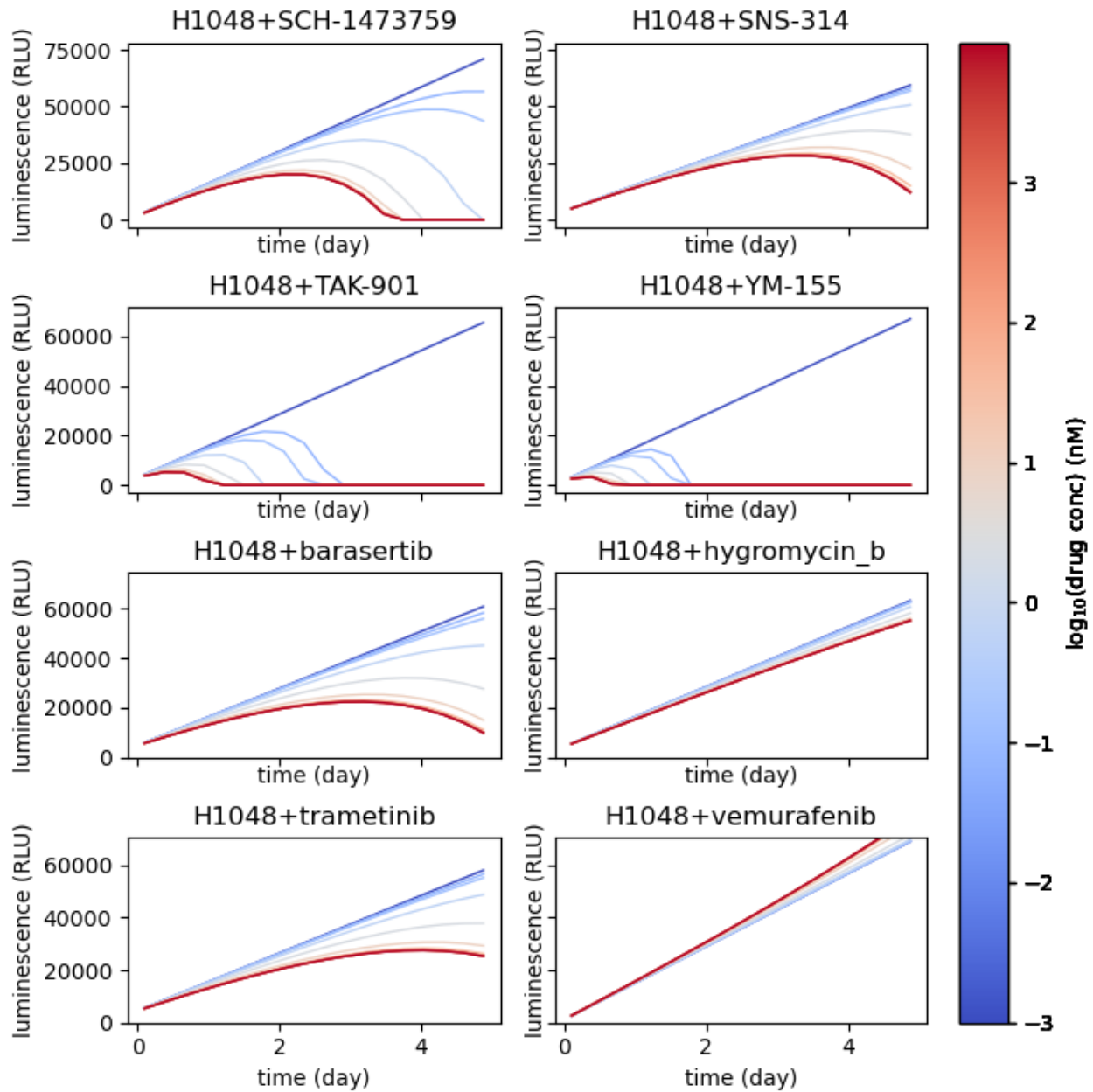
Figure 4.2.8 Modelled luminescence of H1048 over time (SIMULATED). Simulated luminescence data for eight drugs over the course of five days were modelled for the H1048 cell line in PySB (see 4.2.7 for experimental comparison).

Courtesy of L. Harris
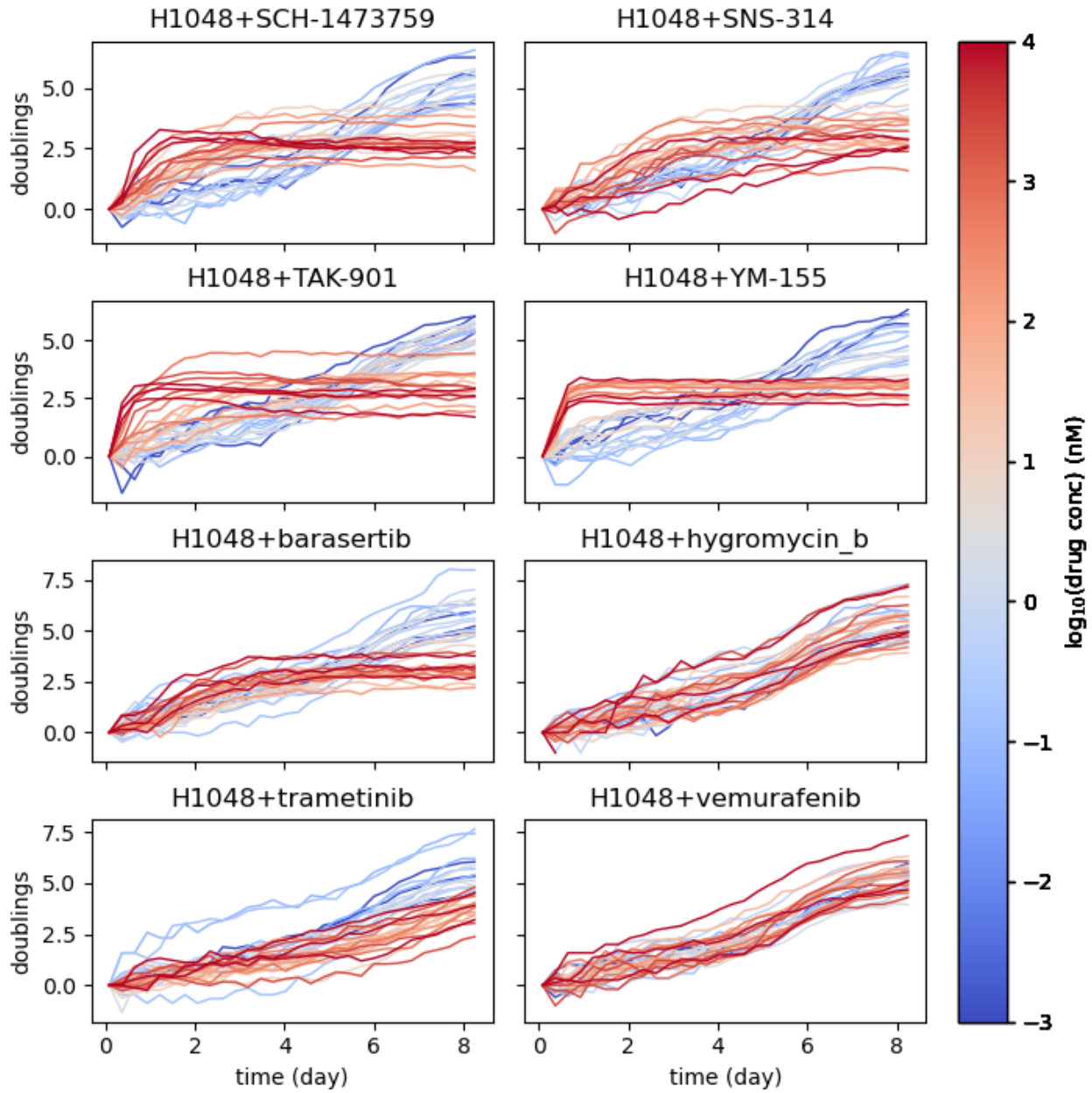
Figure 4.2.9 Direct cell counting of H1048 over time (EXPERIMENTAL).
Experimental cell count data for eight drugs over the course of eight days were
generated for the H1048 cell line (see 4.2.10 for modelling comparison).

Courtesy of L. Harris
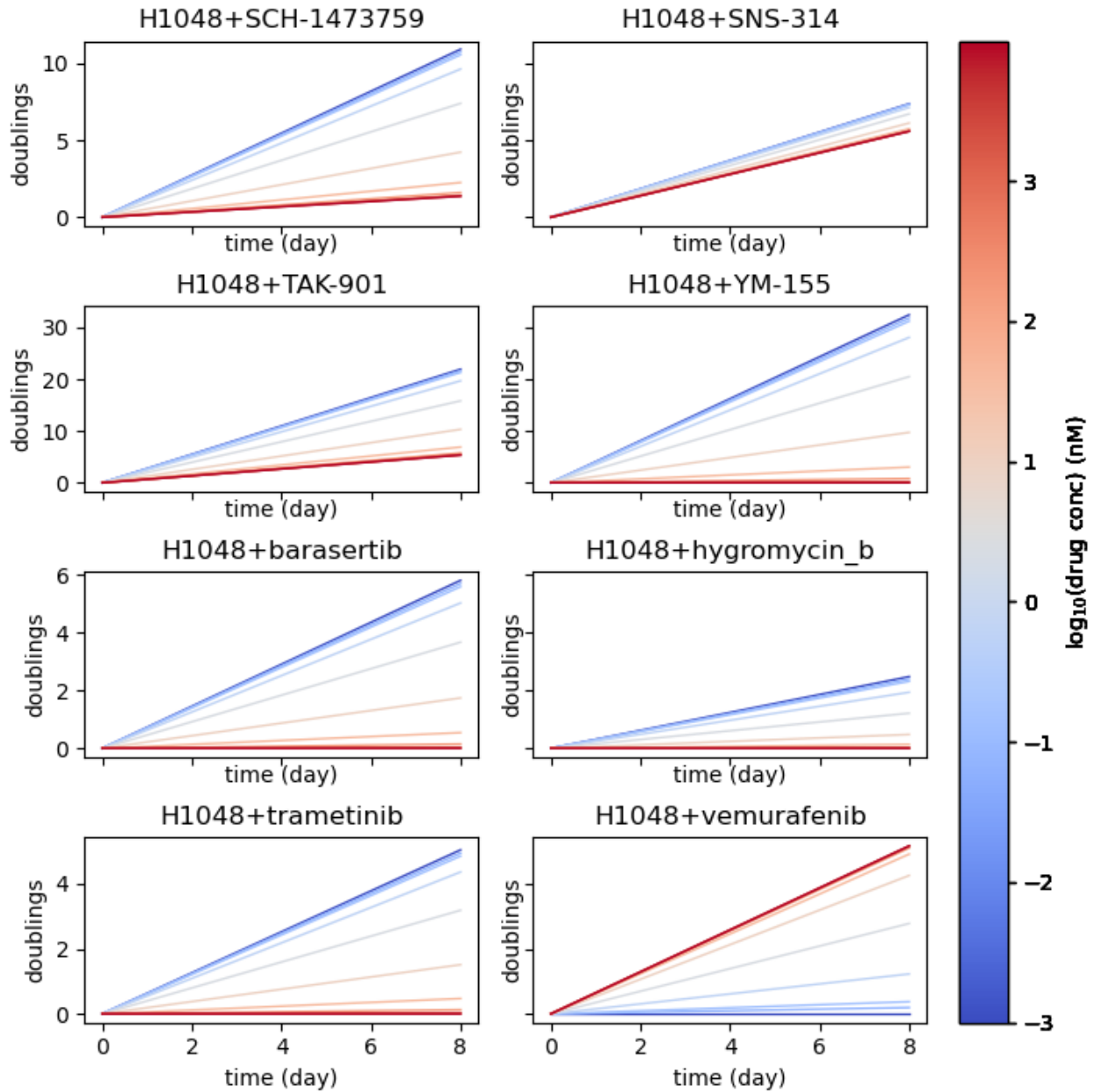
Figure 4.2.10 Predicted cell counts of H1048 over time (SIMULATED). Simulated cell count data for eight drugs over the course of eight days were modelled for the H1048 cell line in PySB (see 4.2.9 for experimental comparison).

Courtesy of L. Harris
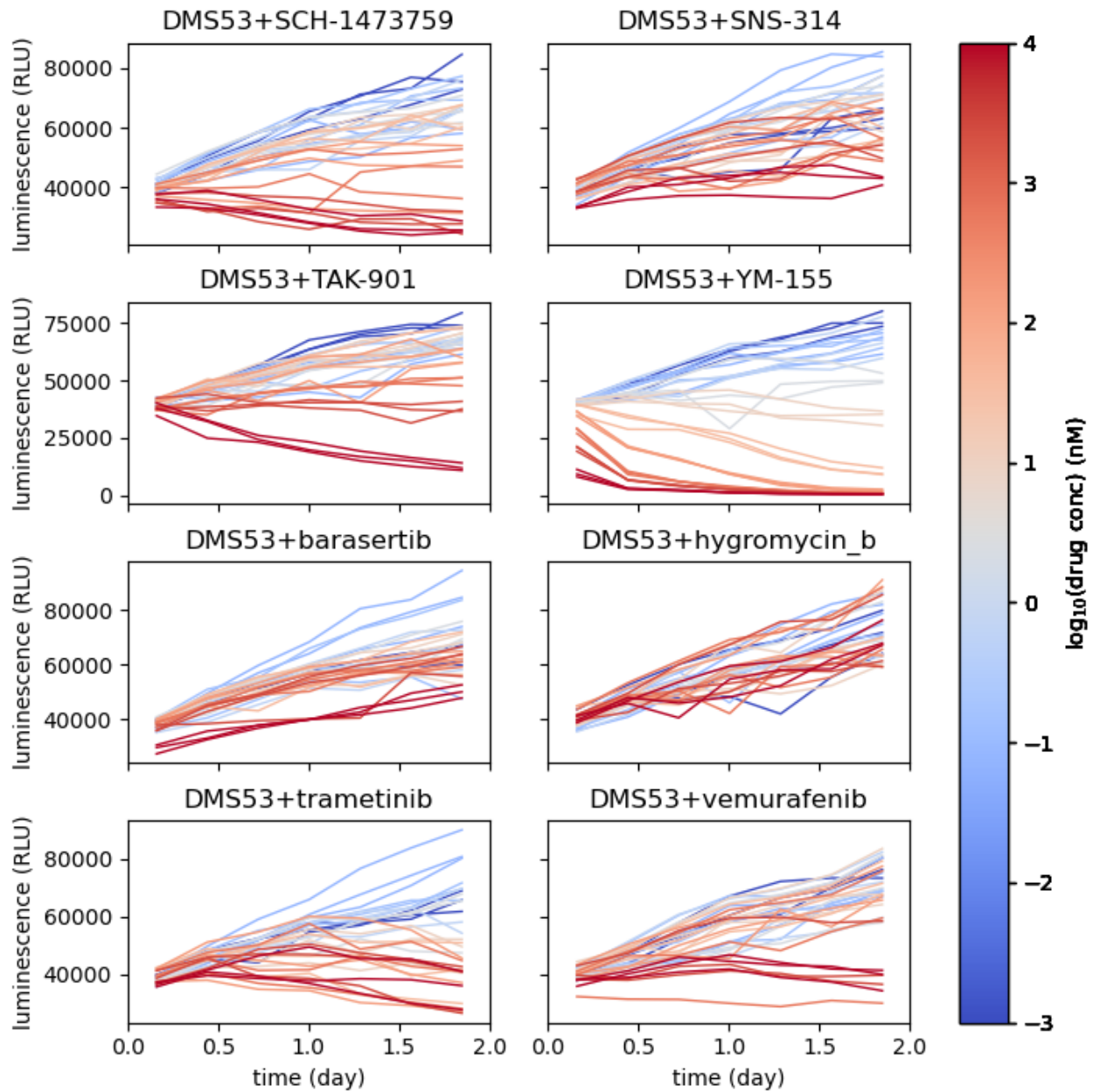
Figure 4.2.11 Real-time luminescence of DMS53 over time (EXPERIMENTAL).
Experimental luminescence data for eight drugs over the course of two days were
generated for the DMS53 cell line (see 4.2.12 for modelling comparison).

Courtesy of L. Harris
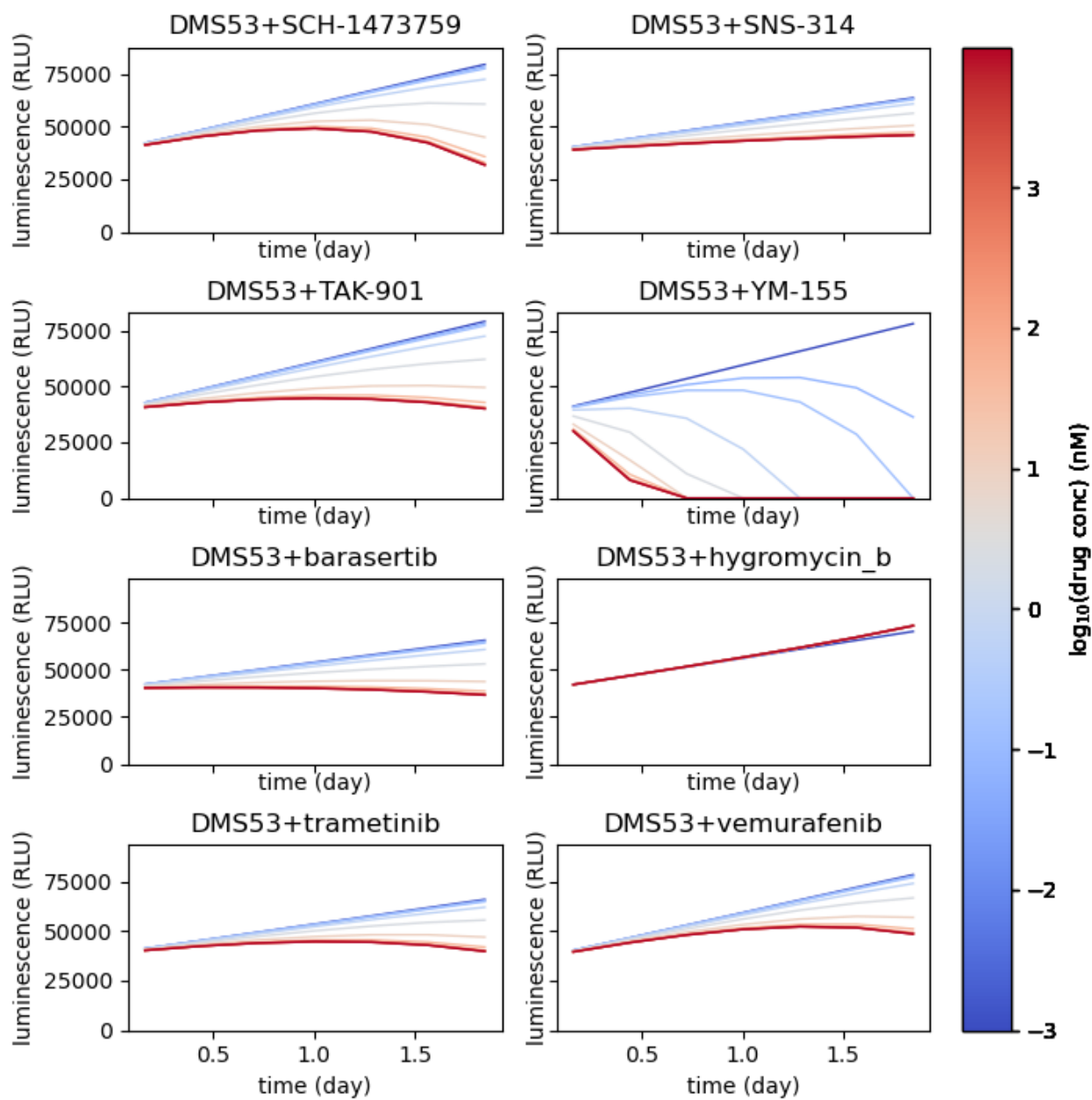
Figure 4.2.12 Modelled luminescence of DMS53 over time (SIMULATED).
Simulated luminescence data for eight drugs over the course of two days were modelled for the DMS53 cell line in PySB (see 4.2.11 for experimental comparison).

Courtesy of L. Harris
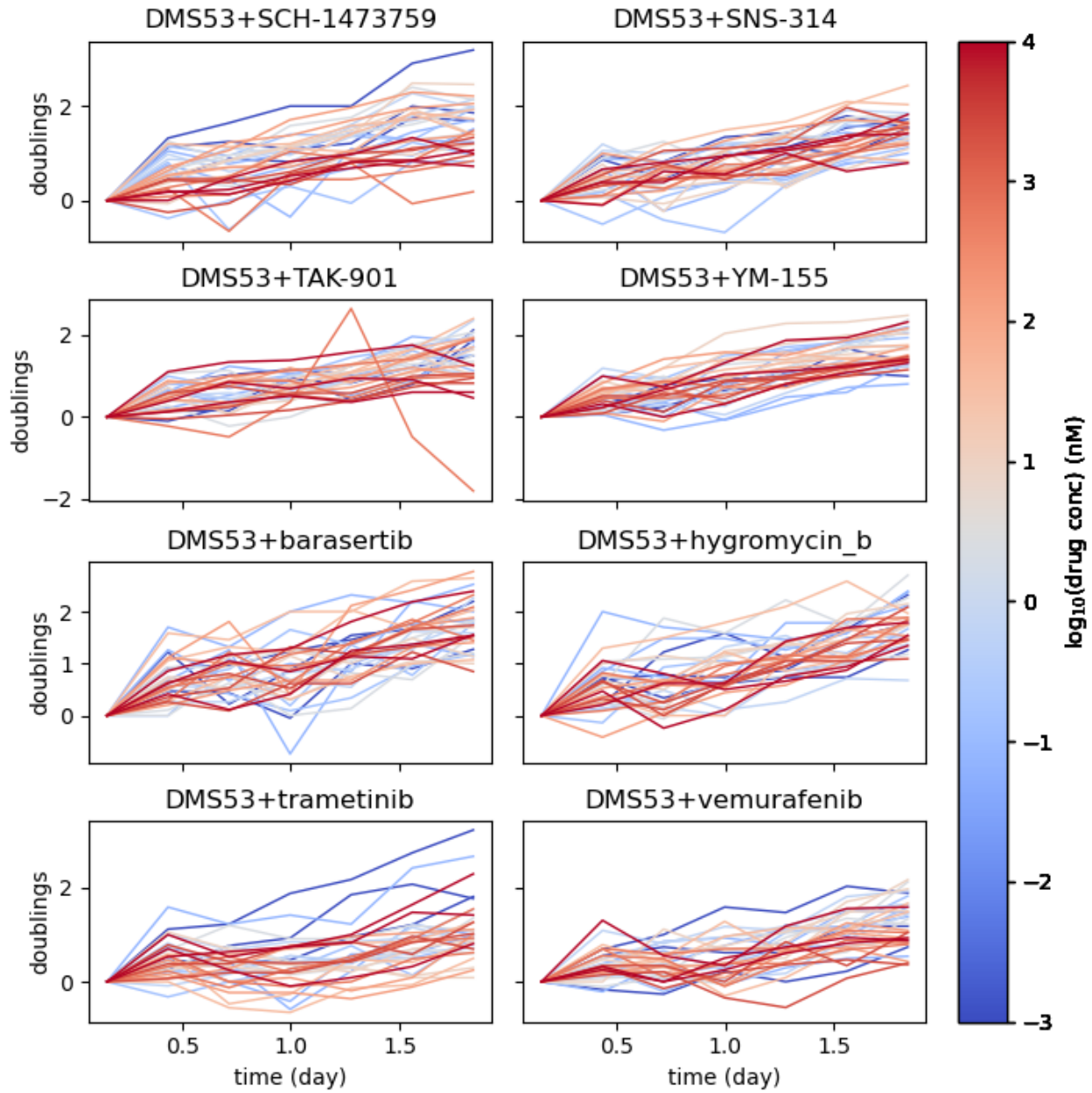
Figure 4.2.13 Direct cell counting of DMS53 over time (EXPERIMENTAL).
Experimental cell count data for eight drugs over the course of two days were generated for the DMS53 cell line (see 4.2.14 for modelling comparison).

Courtesy of L. Harris
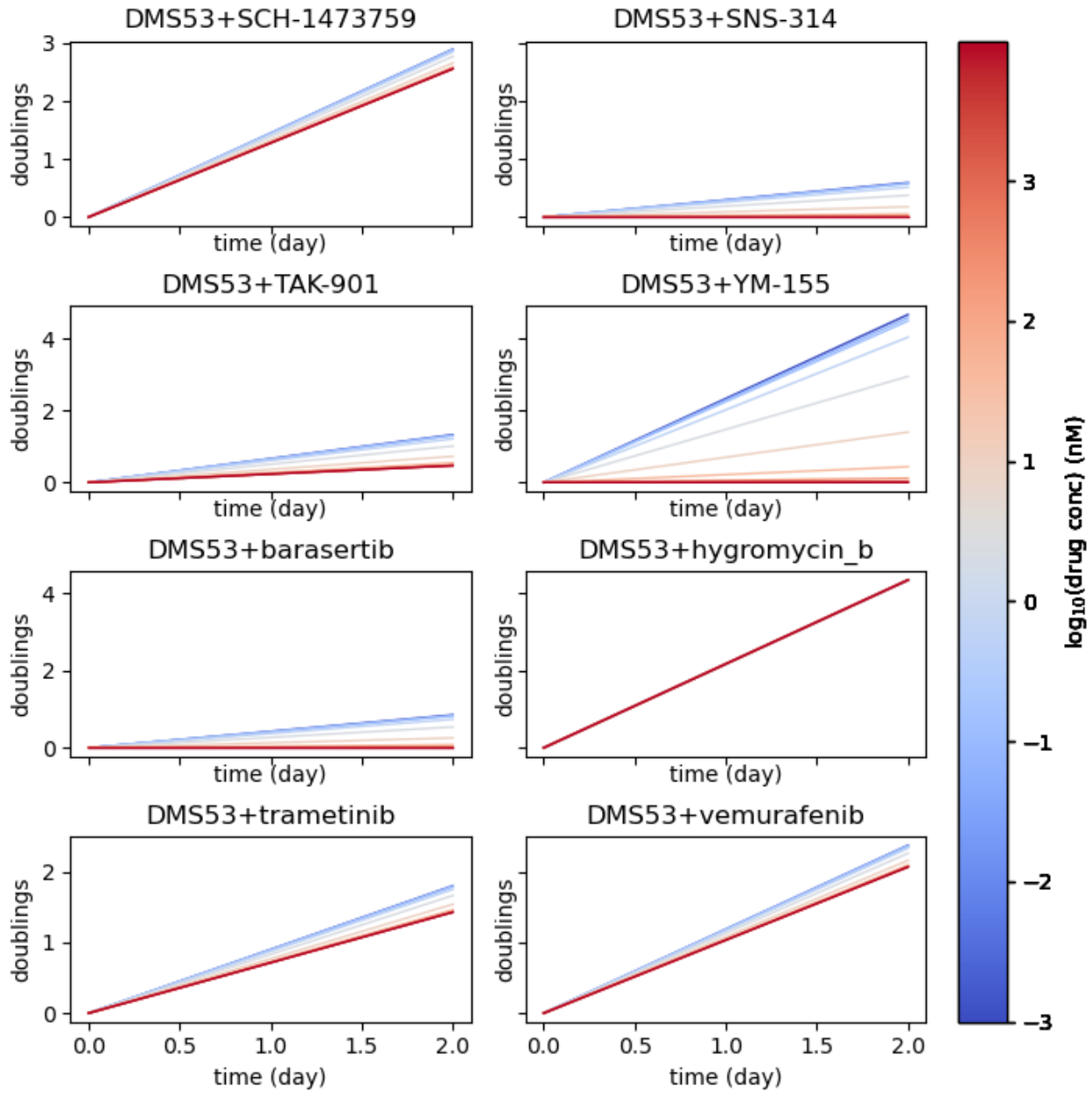
Figure 4.2.14 Predicted cell counts of DMS53 over time (SIMULATED).
Simulated cell count data for eight drugs over the course of two days were modelled for the DMS53 cell line in PySB (see 4.2.13 for experimental comparison).

Courtesy of L. Harris
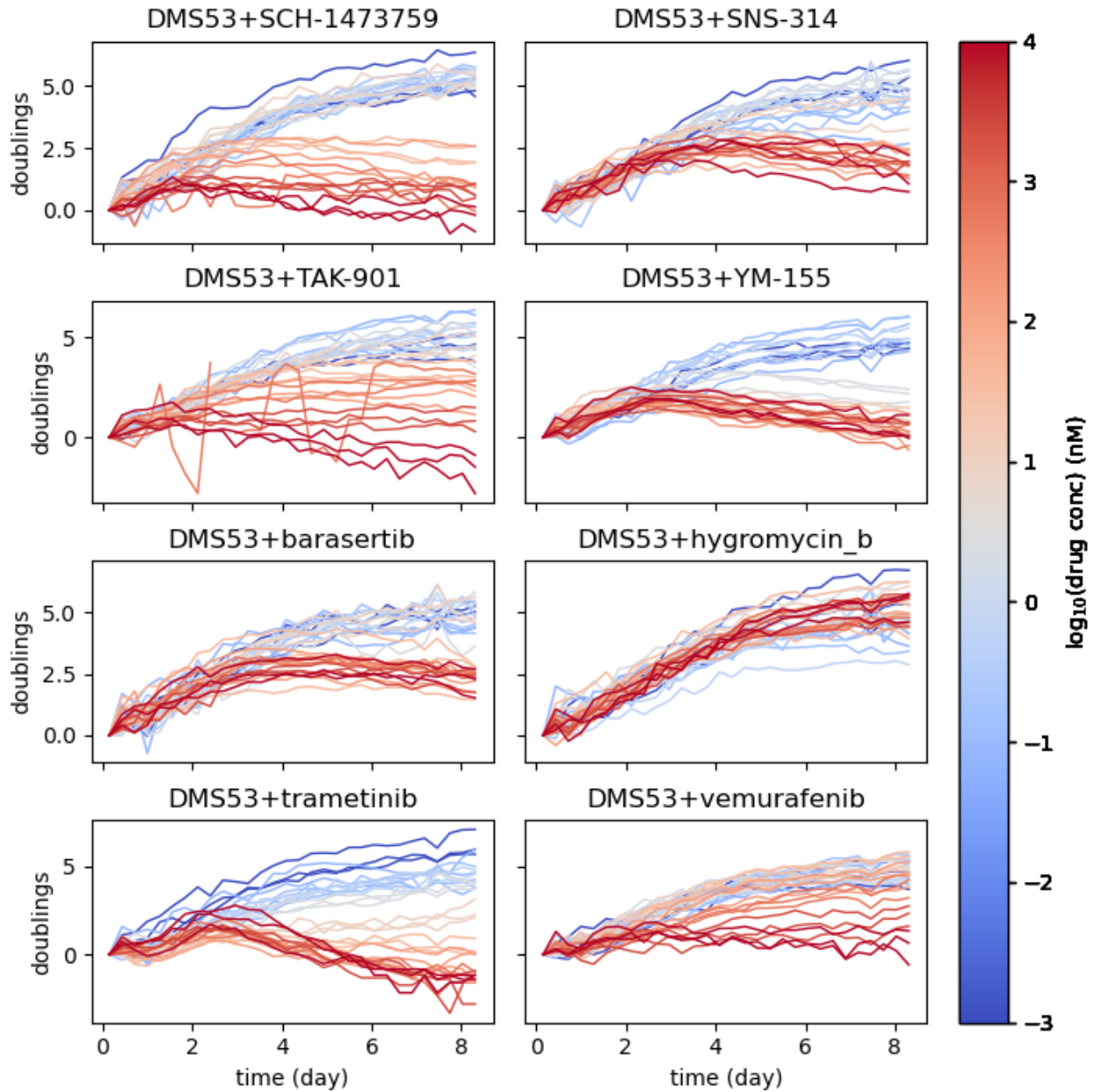
Figure 4.2.15 Direct cell counting of DMS53 over time (>50 hours EXPERIMENTAL). Experimental cell count data for eight drugs over the course of eight days were generated for the DMS53 cell line. Of key interest is the Hygromycin B response which continues to grow exponentially, but becomes decoupled from luminescence results at around 2 days (see Figure 4.2.16).

Courtesy of L. Harris

71
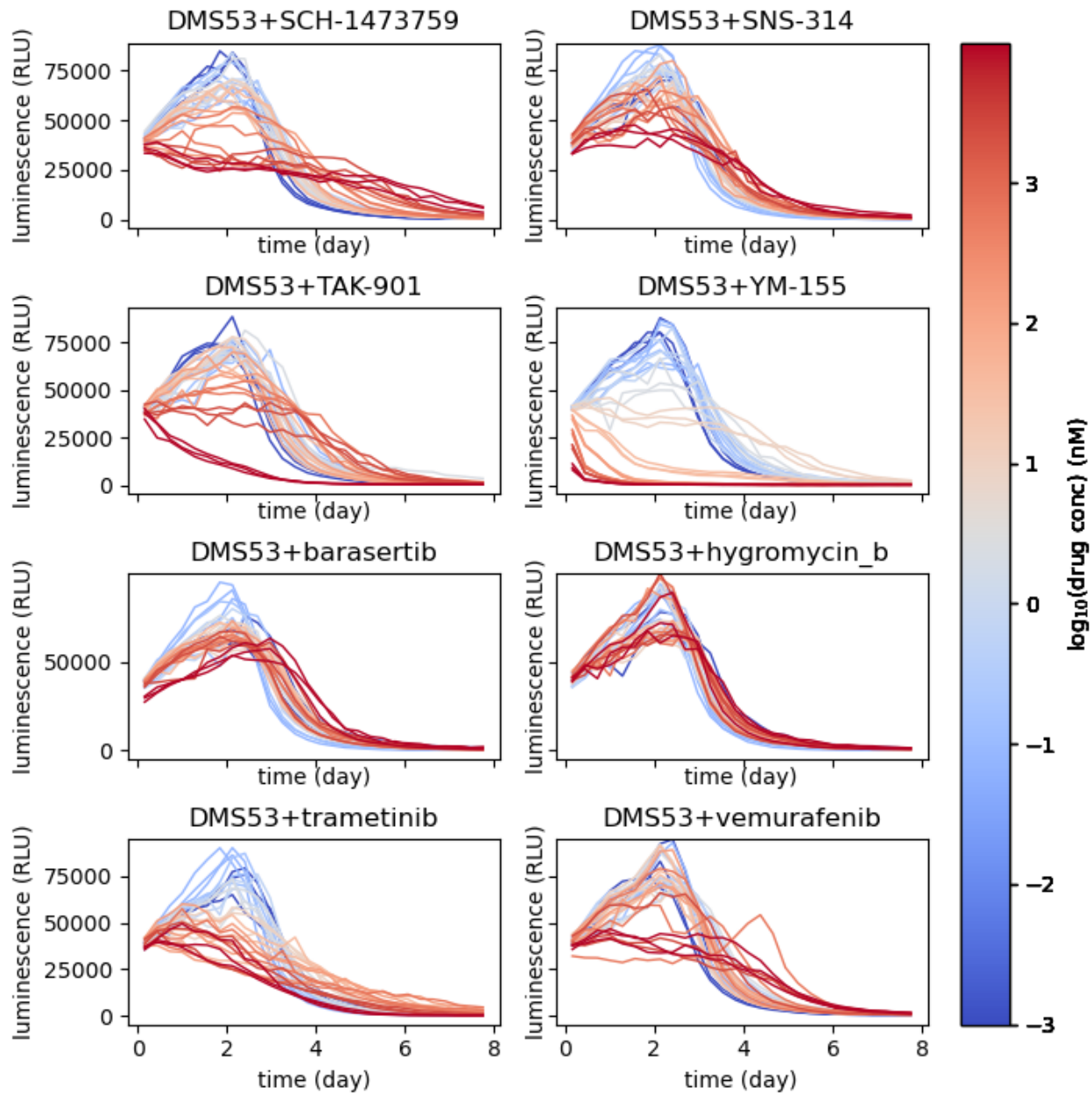
Figure 4.2.16 Real-time luminescence of DMS53 over time (>50 hours EXPERIMENTAL). Experimental luminescence data for eight drugs over the course of eight days were generated for the DMS53 cell line. Of key interest is the Hygromycin B response which becomes decoupled from cell counting results at around 2 days even though cell counting data suggests continued exponential growth (see Figure 4.2.15).

Courtesy of L. Harris

count then effects the model's overall fit as there is a wide range of parameters that can satisfy the observed data. Further experimentation needs to be done to lower the variance in the experimental data so that the current models can be more rigorously evaluated. Upon collection of higher quality data, other phenomenon can be explored such as the apparent uncoupling of luminescence and cell count for some lines (DMS53 post 50 hours, Figures 4.2.15-4.2.16), as well as the tendency of the model to overestimate drug effects in cases where a minimum luminescence level is reached (all cells die) [FIGURE 22 – H1048+TAK-901]. The latter, is likely a mathematical artifact of multiple "zero" values after all cells have died, which are incorrectly included as part of the data involved in the cost function, biasing the parameter space towards the "maximum effect" condition. This can be addressed rather simply by not including timepoint values once a "minimum luminescence threshold" has been met for a condition in the particle swarm optimization. Overall, as a first approximation, the real-time luminescence model performs well, but a significant amount of work is still needed to accurately characterize the relationship between luminescence and cell count.

4.3 Methods

In order to effectively model the real-time luminescence system, an iterative process between experiments and model attenuation was necessary. Our very first attempt at predicting cell counts from luminescence involved the most basic idea of a mathematical model, a standard curve. We tested if this model would work by taking luminescence measurements at different seeding densities and corresponding specific luminescent values to cell numbers. However, it was immediately apparent that while this works for the initial timepoint, using a simplified standard curve model for relating

continuous luminescence measurements to cell counts does not work at all, and the accuracy of the model quickly falls apart. This failure taught us that more information from the real-time luminescence system was needed to accurately simulate the dynamics of the luminescence and cell count relationship. Out of this, an initial PySB model comprised of Cells, NanoLuc luciferase, pro-substrate, and substrate was constructed to begin the modelling process. While this worked initially in roughly approximating cell growth over time in an undrugged sense, it no longer produced sensicle data when drug effects were introduced. This failure then brought us to the third iteration of the model which added a new component (cells under the effect of drugs) in addition to analogous rates for the drugged cells as the undrugged ones, as well as a rate of conversion for cells to go from their initial state to a drugged state. This addition allowed for the luminescence rates to decrease for the first time and marked the closest we have come currently to effective modelling of the system. Additionally, during this time, repeated experiments were conducted to get higher quality data after it was discovered during the first modelling attempt that the fitting statistics were extremely sensitive to the accuracy of the cell count, and whether or not that count came from the exact same well as the luminescence reads or an equivalent replicate. This back and forth iterative process of experimentation and model refinement is a hallmark of the computational modelling process.

Chapter 5: Conclusion and Further Research

5.1 Discussion

5.1.1 Initial intention of thesis

If I had to summarize the entirety of my dissertation work in a single sentence it would be, "Science is hard, and nothing ever works." However, I feel that the expectation is that I should be a tad more verbose. When I was initially brought on to this project, my assumption was that this would not be something that would take very long at all. In fact, the vast majority of the work I've described up to this point was first intended only as part one of my thesis, followed by a large screen comparing drug response in SCLC between genetically modified and unmodified cell lines. As it turns out, the genetic and epigenetic factors that make SCLC so resistant to treatment and interesting to study, also greatly increase the difficulty with which one can conduct and quantify experiments.

5.1.2 Summary of SCLC CRISPR cell line difficulties and findings

The creation of the SCLC cell line library was an incredibly difficult and iterative process straight from the start. However, I gained an absolutely invaluable wealth of knowledge on how to independently problem solve molecular biological and cellular culture issues. From the molecular biology perspective, it allowed me to gain experience in almost every modern cloning technique, as well as cultivate a deep understanding of the nuances of cell culture in difficult to culture lines. The most useful findings of the SCLC cell line library are unfortunately what "doesn't work". SCLC lines have a clear aversion to the use of certain promoters and other genetic elements, but our current understanding of them is not sufficient enough to do anything beyond

speculate as to why. Specifically, for the promoter and NLS signal optimization, the fact that Ef1a and a c-Myc NLS performed the best is not surprising as this is in line with previously published results comparing promoters and NLS' across cell lines (Ray et al., 2015). However, what was surprising was the absolute deft of commercially available plasmids containing promoters other than CMV, and NLS' other than SV40. This gives the general impression, that the majority of plasmids used in research are used and/or assembled purely for their convenience, and not the product of iterative rounds of optimization, as was the case in the SCLC cell line library construction. Moving forward, I think the most beneficial area of molecular biological research in regards to the SCLC cell line library would be to expand the current modular cloning library that was created during this process to include a wider range of preassembled TUs to make the iterative testing of ideal genetic components more facile and streamlined in the future.

5.1.3 Future directions in utilizing the SCLC CRISPR cell line library

Now that the SCLC CRISPR cell line library exists, the next step would be to utilize it as a tool in mechanistic studies. As a starting point, the top 10 most associated transcription factors from each subtype described in Wooten et al. could be knocked out in an effort to destabilize the subtype they are associated with, or they could be activated via CRISPRa in an effort to destabilize a subtype they are not associated with. In practice, for the CRISPRko experiments this would involve the design and ordering of a minimum of 160 sgRNAs (4 sgRNAs per gene) followed by 40 multiplexed sgRNA electroporations per cell line being investigated (ideally 2 per subtype). From there each individual knockout line would need to be verified by ICE analysis of sanger sequencing, to confirm the knockout of each individual genes. Once the knockout lines

were established, scRNAseq and cyTOF could be done on each line to compare the wild-type transcriptomic and proteomic profiles to those of the knockouts. The knockout data could then be re-clustered alongside the wild types to see if the perturbations led to any difference in subtype classification. Furthermore, differential drug response studies could be done to look for increased or decreased sensitivity to certain drug classes as the result of these genetic perturbations. For the CRISPRa experiments, a completely different set of 80 sgRNAs (2 per gene and 10 per subtype) would need to be designed targeting the promoter and early exon regions. This sgRNA library would then need to be cloned into PiggyBac destination vectors under an RNA Pol II promoter system, and randomly integrated into each respective cell lines genome. Once integrated, successful incorporation would need to be quantified by qPCR initially, followed by analogous characterization experiments to the CRISPRko lines such as scRNAseq, cyTOF, and differential drug response studies. Overall, the results of these studies could help indicate novel TF network pathways for use in combination therapy design. Perhaps combining a specific TF inhibitor with a chemotherapeutic drug to sensitize the cells to its mechanism of action. It is also of particular importance to note that the experiments described above already take into account the massive degree of uncertainty and influence among genes between the genomic, transcriptomic, and proteomic levels. When probing any gene network it is important to conceptualize potential pitfalls and hurdles in designing experiments. Particularly with TFs, where a single TF can interact with multiple genes, it is important to build an experiment from its most basic principles. In this case, when trying to perturb the TF network from a genetic level, it is first important to know that both the gene exists in the cell line being edited

(done through genomic sequencing), and that gene also persists at downstream levels such as transcriptomically (checked by scRNAseq abundance), or proteomically (checked by Western blot, or cyTOF mass spectrometry). Absence at downstream levels could indicate silencing at the genomic levels through methylation, or perhaps epigenetic inactivation by another transcription factor. If neither of those are found to be the cause (perhaps tested by ChIP or Infinium methylation assay), the loss could be at the transcriptomic level as part of RNA degradation, or similarly at the proteomic level as well. However, the experimental directions described above can largely avoid the likelihood of such troubleshooting events by rigorously characterizing the cellular abundance of the targeted transcription factors gleaned from clustering by an additional cyTOF screen to confirm their presence at the protein level, subsequently confirming their presence at the levels above as well (DNA and RNA) prior to CRISPR experimentation. This same level of rigor should then also be applied after CRISPR editing (a/i/ko) of lines to quantify abundance and presence at the genomic (qPCR), transcriptomic (scRNAseq), and proteomic (cyTOF mass spectrometry) again. However, with the extreme complexity and interconnectedness of gene networks, it is also probable that single gene perturbations lead to no measurable difference in phenotypic behavior. This is to say that as the network being perturbed becomes more complex, the likelihood of it being resistant to small changes increases exponentially. Contrastingly, perturbing a single gene within a relatively non-redundant and interdependent network could also quickly lead to lethality if the gene being perturbed's function cannot be substituted or ignored. Overall, this SCLC cell line library has enormous potential to interrogate a variety of gene networks related to SCLC disease

progression and drug response, but in order to do so, an extremely rigorous approach to experimental design is absolutely necessary.

5.1.4 Summary of SCLC culture condition optimization processes

In regards to the large amount of culture condition optimization that had to be done for each individual cell line, I would like to see a broader cultural shift within cell line based research. When biologists request cell lines from commercial sources like Sigma, or academic sources like the ATCC, they tend to treat the published cell culture conditions and notes as immutable, rather than think in a broader sense what those culture conditions represent. Each of those culture conditions is purely the result of what a researcher was able to get cells originally isolated from a whole organism to grow in. By no means are these the absolute only conditions in which those cells can thrive, nor should those conditions be presumed optimal. Culture conditions should be regarded as a starting point, and based on the specific needs of the research, be adjusted accordingly.

5.1.5 Summary of quantification of drug-response by real-time luminescence

Moving on to the real-time luminescence portion of my work, overall it was an amazing success. Despite the initial obstacles of optimizing the reagent protocol to suit a single-addition continuous workflow, the data generated and subsequent analysis pipeline now allow for a highly efficient way of characterizing drug response in cell lines without the need for a label. Moreover, because the protocol utilizes microplates and commonly available luminometer instrumentation, it is easily scalable and integrated in to larger screening endeavors than what were attempted in the proof of concept experiments.

5.1.6 Future directions in real-time luminescence

There are three main directions that I see the real-time luminescence method improving on from here. One, testing of the method on 3-Dimensional cultures and organoids to see if it remains suitable to drug-response quantitation. Two, expanding the real-time analysis past 100 hours by exchanging media components after a given time, potentially opening the door for experiments that last days, weeks, or months. And three, testing the real-time luminescence method on drugs and compounds that do not exhibit a single-phase linear response. For example, Triton-X 100 which, due to its mechanism of action (membrane destabilization), would be predicted to show an exponentially increasing drug-response, or long-term treatment with cisplatin/etoposide which would likely be a bi-phasic response wherein there is an initial single-phase linear response followed by subsequent resistance and loss of response altogether. In each case, it would be very interesting to adjust the algorithmic method wherein the continuous response is distilled down into a single value. One can imagine that for an exponentially increasing drug effect a derivative analysis could be used, and for a biphasic response, perhaps the data could be fit to an nth term polynomial function. Regardless, the initial experiments and results from this work lay the foundation for a variety of extended applications in the future, without the immediate need for implementation protocol adjustment.

5.1.7 Summary of modelling the relationship between real-time luminescence and cell count

For the dynamic modelling of the relationship between real-time luminescence and cell count, as it currently stands the model works roughly as intended. From raw

luminescence values, simulated cell counts can be generated and compared to actual direct counting. These current results are mixed in results of prediction success, but it represents a strong start to a much longer endeavor than the scope of my thesis. In general, the most promising data is for drugs that do not reach the "minimum luminescence threshold" during the experiment, as the fitting data is not flooded with zeroes. Additionally, because of the clear differences in cell imaging efficiency among the lines tested, it is difficult to comment on whether the differences between experimental and simulated results are due to poor parameter set optimization, or low-quality data being fed in to the model. One key observation is that the relationship between luminescence and cell count is clearly cell line dependent as the optimal parameter sets for each individual line were markedly different. This was the expected result however as each line exhibits completely different cell growth mechanics (as evidenced by doubling time and morphology of culture) as well as different reductive capacities for converting pro-substrate to substrate. Overall, the cell line with the most accurate modelling behavior compared to the experimental data was clearly DMS53. However, this can only be said for the first 50 hours used in the analysis as constrained by the luminescence rate algorithm. Most peculiarly, after the 50 hour mark, DMS53 seems to decouple the relationship between luminescence and cell count, even in cases where there was no drug effect. This is most explicitly evidenced in the Hygromycin B condition where the cells continue to grow exponentially by cell counting, but luminescence rapidly begins to decrease. This decoupling and luminescence decrease could be due to a variety of factors with the most likely two being either a "metabolic shift" in the cell population leading to a decrease in reductive capacity after a

50 hour latency period, or a total consumption of the available pro-substrate reagents (i.e. lack of available substrate is causing the decrease). Prior to testing the metabolic shift theory, the consumption theory would need to be tested to see if either adding a portion or multiple of pro-substrate at the beginning of the assay shifted the decoupling time earlier or later respectively. If no shift is seen in the decoupling observation, further interrogation of the metabolic pathways of the cells before and after the decoupling would need to be done to hone in on a proposed molecular mechanism.

5.1.8 Future directions of modelling the relationship between real-time luminescence and cell count

In order to further optimize the parameter values and applicability of the model, multiple experiments will need to be repeated utilizing higher quality imaging that gives a more accurate and less varied estimate of the cellular growth over time. Furthermore, as the analysis pipeline of the real-time luminescence is also extended into 3-Dimensional culture and multi-phase drug response, the computational model can similarly be expanded. Prior to any adjustment on the computational side, higher quality experimental data is an absolute necessity in order to accurately tune the cost function. However, as evidenced by the entirety of Chapter 3, linking cell count and luminescence is not a requirement for using continuous luminescence to gather drug response data. The utility of linking luminescence and cell count lies solely within interrogating the cellular relationship between cancer metabolism and cell growth. Lastly, the decoupling of luminescence and cell count could be re-confirmed through experiments containing varied levels of pro-substrate as mentioned in section 5.1.8, as well as in conjunction

with standard continuous metabolic assays such as Seahorse XF. If it is found that this metabolic shift is consistently occurring, it would pave the way for a multitude of experiments exploring the cancer utility of this mechanism, especially since it was only initially observed in cell lines that were part of the generally drug-resistant NEv2 subtype.

5.1.9 Closing

In closing, the work accomplished over the duration of my graduate studies has morphed and changed as the years have gone by. My initial expectations and goals at the outset were lofty, and quickly tempered by the realities of modern biomedical research. However, what I was able to accomplish in regards to constructing a labelled and genetically tunable SCLC cell line library, development of a label-free continuous method of drug-response analysis, and establishment of an initial computational model linking real-time luminescence to cell count, have laid a strong groundwork for future researchers to continue to improve upon and apply to a variety of biological questions.

References

Altman, F. P. (1976). Tetrazolium salts and formazans. *Progress in Histochemistry and Cytochemistry*, *9*(3), 3–6. https://doi.org/10.1016/S0079-6336(76)80015-0

Anderson, E. C., Bell, G. I., Petersen, D. F., & Tobey, R. A. (1969). Cell Growth and Division: IV. Determination of Volume Growth Rate and Division Probability. *Biophysical Journal*, *9*(2), 246. https://doi.org/10.1016/S0006-3495(69)86383-6

Asai, N., Ohkuni, Y., Kaneko, N., Yamaguchi, E., & Kubo, A. (2014). Relapsed small cell lung cancer: treatment options and latest developments. *Therapeutic Advances in Medical Oncology*, *6*(2), 69. https://doi.org/10.1177/1758834013517413

Bernas, T., & Dobrucki, J. (2002). Mitochondrial and nonmitochondrial reduction of MTT: Interaction of MTT with TMRE, JC-1, and NAO mitochondrial fluorescent probes. *Cytometry*, *47*(4), 236–242. https://doi.org/10.1002/CYTO.10080

Berridge, M. V., Herst, P. M., & Tan, A. S. (2005). Tetrazolium dyes as tools in cell biology: New insights into their cellular reduction. *Biotechnology Annual Review*, *11*(SUPPL.), 127–152. https://doi.org/10.1016/S1387-2656(05)11004-7

Berridge, M. V., & Tan, A. S. (1993). Characterization of the Cellular Reduction of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT): Subcellular Localization, Substrate Dependence, and Involvement of Mitochondrial Electron Transport in MTT Reduction. *Archives of Biochemistry and Biophysics*, *303*(2), 474–482. https://doi.org/10.1006/ABBI.1993.1311

Butcher, E. C. (2005). Can cell systems biology rescue drug discovery? *Nature Reviews. Drug Discovery*, *4*(6), 461–467. https://doi.org/10.1038/NRD1754

Ca, I., Rojo, F., Taus, A., Arp, O., Arum -Ur, M., Pijuan, L., … Arriola, E. (n.d.).
Targeting Epithelial-to-Mesenchymal Transition with Met Inhibitors Reverts
Chemoresistance in Small Cell Lung Cancer. *Clin Cancer Res*, 1–13.
https://doi.org/10.1158/1078-0432.CCR-13-1330

Cardarelli, F., Bizzarri, R., Serresi, M., Albertazzi, L., & Beltram, F. (2009). Probing
nuclear localization signal-importin alpha binding equilibria in living cells. *The
Journal of Biological Chemistry*, *284*(52), 36638–36646.
https://doi.org/10.1074/jbc.M109.036699

Carpenter, J. R. (1986). A method for presenting and comparing dose-response curves.
Journal of Pharmacological Methods, *15*(4), 283–303. https://doi.org/10.1016/0160-
5402(86)90009-4

Cermak, N., Olcum, S., Delgado, F. F., Wasserman, S. C., Payer, K. R., Murakami, M.
A., … Manalis, S. R. (2016). High-throughput measurement of single-cell growth
rates using serial microfluidic mass sensor arrays. *Nature Biotechnology*, *34*(10),
1052. https://doi.org/10.1038/NBT.3666

Cory, A. H., Owen, T. C., Barltrop, J. A., & Cory, J. G. (1991). Use of an Aqueous
Soluble Tetrazolium/Formazan Assay for Cell Growth Assays in Culture. *Cancer
Communications*, *3*(7), 207–212. https://doi.org/10.3727/095535491820873191

Duellman, S. J., Zhou, W., Meisenheimer, P., Vidugiris, G., Cali, J. J., Gautam, P., …
Vidugiriene, J. (2015). Bioluminescent, Nonlytic, Real-Time Cell Viability Assay and
Use in Inhibitor Screening. *Assay and Drug Development Technologies*, *13*(8),
456–465.
https://doi.org/10.1089/ADT.2015.669/ASSET/IMAGES/LARGE/FIGURE5.JPEG

England, C. G., Ehlerding, E. B., & Cai, W. (2016). NanoLuc: A Small Luciferase is

Brightening up the Field of Bioluminescence. *Bioconjugate Chemistry*, *27*(5), 1175.

https://doi.org/10.1021/ACS.BIOCONJCHEM.6B00112

Fallahi-Sichani, M., Honarnejad, S., Heiser, L. M., Gray, J. W., & Sorger, P. K. (2013).

Metrics other than potency reveal systematic variation in responses to cancer

drugs. *Nature Chemical Biology*, *9*(11), 708.

https://doi.org/10.1038/NCHEMBIO.1337

Fonseca, J. P., Bonny, A. R., Kumar, G. R., Ng, A. H., Town, J., Wu, Q. C., … El-

Samad, H. (2019). A Toolkit for Rapid Modular Construction of Biological Circuits in

Mammalian Cells. *ACS Synthetic Biology*, *8*(11), 2593–2606.

https://doi.org/10.1021/ACSSYNBIO.9B00322/SUPPL_FILE/SB9B00322_SI_001.P

DF

Frick, P. L., Paudel, B. B., Tyson, D. R., & Quaranta, V. (2015). Quantifying

heterogeneity and dynamics of clonal fitness in response to perturbation. *Journal of

Cellular Physiology*, *230*(7), 1403–1412. https://doi.org/10.1002/JCP.24888

Garnett, M. J., Edelman, E. J., Heidorn, S. J., Greenman, C. D., Dastur, A., Lau, K. W.,

… Benes, C. H. (2012). Systematic identification of genomic markers of drug

sensitivity in cancer cells. *Nature 2012 483:7391*, *483*(7391), 570–575.

https://doi.org/10.1038/nature11005

Gazdar, A. F., Bunn, P. A., & Minna, J. D. (2017). Small-cell lung cancer: what we

know, what we need to know and the path forward. *Nature Reviews Cancer*,

17(12), 725–737. https://doi.org/10.1038/nrc.2017.87

Haddadin, S., & Perry, M. C. (2011). History of Small-Cell Lung Cancer. *Clinical Lung*

Cancer, *12*(2), 87–93. https://doi.org/10.1016/J.CLLC.2011.03.002

Hafner, M., Niepel, M., Chung, M., & Sorger, P. K. (2016). Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs. *Nature Methods*, *13*(6), 521. https://doi.org/10.1038/NMETH.3853

Hall, M. P., Unch, J., Binkowski, B. F., Valley, M. P., Butler, B. L., Wood, M. G., … Wood, K. V. (2012). Engineered LuciferaseReporter from a Deep Sea ShrimpUtilizing a Novel Imidazopyrazinone Substrate. *ACS Chemical Biology*, *7*(11), 1848. https://doi.org/10.1021/CB3002478

Harris, L. A., Frick, P. L., Garbett, S. P., Hardeman, K. N., Paudel, B. B., Lopez, C. F., … Tyson, D. R. (2016a). An unbiased metric of antiproliferative drug effect in vitro. *Nature Methods*, *13*(6), 497–500. https://doi.org/10.1038/nmeth.3852

Harris, L. A., Frick, P. L., Garbett, S. P., Hardeman, K. N., Paudel, B. B., Lopez, C. F., … Tyson, D. R. (2016b). An unbiased metric of antiproliferative drug effect in vitro. *Nature Methods 2016 13:6*, *13*(6), 497–500. https://doi.org/10.1038/nmeth.3852

Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., Doudna, J. A., Snijders, A. P., … Oost, J. van der. (2010). Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. *Science*, *329*(5997), 1355–1358. https://doi.org/10.1126/science.1192272

History of Cancer Screening and Early Detection: 20th Century to Present. (n.d.). Retrieved March 11, 2022, from https://www.cancer.org/cancer/cancer-basics/history-of-cancer/screening-early-detection.html

Hodel, A. E., Harreman, M. T., Pulliam, K. F., Harben, M. E., Holmes, J. S., Hodel, M. R., … Corbett, A. H. (2006). Nuclear Localization Signal Receptor Affinity

Correlates with in Vivo Localization in Saccharomyces cerevisiae * Downloaded from. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, *281*(33), 23545–23556. https://doi.org/10.1074/jbc.M601718200

Horn, L., Mansfield, A. S., Szczęsna, A., Havel, L., Krzakowski, M., Hochmair, M. J., … Liu, S. V. (2018). First-Line Atezolizumab plus Chemotherapy in Extensive-Stage Small-Cell Lung Cancer. *New England Journal of Medicine*, *379*(23), 2220–2229. https://doi.org/10.1056/NEJMOA1809064/SUPPL_FILE/NEJMOA1809064_DATA-SHARING.PDF

Hsieh, J. H., Huang, R., Lin, J. A., Sedykh, A., Zhao, J., Tice, R. R., … Auerbach, S. S. (2017). Real-time cell toxicity profiling of Tox21 10K compounds reveals cytotoxicity dependent toxicity pathway linkage. *PloS One*, *12*(5). https://doi.org/10.1371/JOURNAL.PONE.0177902

Hu, P., Zhao, X., Zhang, Q., Li, W., & Zu, Y. (2018). Comparison of Various Nuclear Localization Signal-Fused Cas9 Proteins and Cas9 mRNA for Genome Editing in Zebrafish. *G3 (Bethesda, Md.)*, *8*(3), 823–831. https://doi.org/10.1534/G3.117.300359

Isherwood, B., Timpson, P., Mcghee, E. J., Anderson, K. I., Canel, M., Serrels, A., … Carragher, N. O. (2011). Live Cell in Vitro and in Vivo Imaging Applications: Accelerating Drug Discovery. *Pharmaceutics*, *3*(2), 141. https://doi.org/10.3390/PHARMACEUTICS3020141

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, *337*(6096), 816.

https://doi.org/10.1126/SCIENCE.1225829

Ko, J., Winslow, M. M., & Sage, J. (2021). Mechanisms of small cell lung cancer metastasis. *EMBO Molecular Medicine*, *13*(1), e13122. https://doi.org/10.15252/EMMM.202013122

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., … Zhang, F. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, *517*(7536), 583–588. https://doi.org/10.1038/nature14136

Krohn, A., Ahrens, T., Yalcin, A., Plönes, T., Wehrle, J., Taromi, S., … Burger, M. (2014). Tumor cell heterogeneity in Small Cell Lung Cancer (SCLC): phenotypical and functional differences associated with Epithelial-Mesenchymal Transition (EMT) and DNA methylation changes. *PloS One*, *9*(6), e100249. https://doi.org/10.1371/journal.pone.0100249

Li, X., Burnight, E. R., Cooney, A. L., Malani, N., Brady, T., Sander, J. D., … Craig, N. L. (2013). PiggyBac transposase tools for genome engineering. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(25). https://doi.org/10.1073/pnas.1305987110

Lopez, C. F., Muhlich, J. L., Bachman, J. A., & Sorger, P. K. (2013). Programming biological models in Python using PySB. *Molecular Systems Biology*, *9*(1), 646. https://doi.org/10.1038/MSB.2013.1

Lu, J., Wu, T., Zhang, B., Liu, S., Song, W., Qiao, J., & Ruan, H. (2021). Types of nuclear localization signals and mechanisms of protein import into the nucleus. *Cell Communication and Signaling*, *19*(1), 1–10. https://doi.org/10.1186/S12964-021-

00741-Y/FIGURES/1

Lung Cancer - Small Cell: Statistics | Cancer.Net. (n.d.). Retrieved March 11, 2022,

from https://www.cancer.net/cancer-types/lung-cancer-small-cell/statistics

Lung Cancer Survival Rates | 5-Year Survival Rates for Lung Cancer. (n.d.). Retrieved

March 26, 2022, from https://www.cancer.org/cancer/lung-cancer/detection-

diagnosis-staging/survival-rates.html

Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H., & Joung, J. K. (2013).

CRISPR RNA-guided activation of endogenous human genes. *Nature Methods*,

10(10), 977–979. https://doi.org/10.1038/nmeth.2598

McFadden, D. G., Papagiannakopoulos, T., Taylor-Weiner, A., Stewart, C., Carter, S.

L., Cibulskis, K., … Jacks, T. (2014). Genetic and Clonal Dissection of Murine

Small Cell Lung Carcinoma Progression by Genome Sequencing. *Cell*, *156*(6),

1298–1311. https://doi.org/10.1016/j.cell.2014.02.031

Meddings, J. B., Scott, R. B., & Fick, G. H. (1989). Analysis and comparison of

sigmoidal curves: application to dose-response data.

Https://Doi.Org/10.1152/Ajpgi.1989.257.6.G982, *257*(6).

https://doi.org/10.1152/AJPGI.1989.257.6.G982

Mosmann, T. (1983). Rapid colorimetric assay for cellular growth and survival:

Application to proliferation and cytotoxicity assays. *Journal of Immunological*

Methods, *65*(1–2), 55–63. https://doi.org/10.1016/0022-1759(83)90303-4

Nieto, M. A., Huang, R. Y.-J., Jackson, R. A., & Thiery, J. P. (2016). EMT: 2016. *Cell*,

166(1), 21–45. https://doi.org/10.1016/j.cell.2016.06.028

Polley, E., Kunkel, M., Evans, D., Silvers, T., Delosh, R., Laudeman, J., … Teicher, B.

A. (2016). Small Cell Lung Cancer Screen of Oncology Drugs, Investigational

    Agents, and Gene and microRNA Expression. *Journal of the National Cancer*

    Institute, *108*(10), djw122. https://doi.org/10.1093/jnci/djw122

Ray, M., Tang, R., Jiang, Z., & Rotello, V. M. (2015). Quantitative tracking of protein

    trafficking to the nucleus using cytosolic protein delivery by nanoparticle-stabilized

    nanocapsules. *Bioconjugate Chemistry*, *26*(6), 1004–1007.

    https://doi.org/10.1021/acs.bioconjchem.5b00141

Riss, T. L., Moravec, R. A., Niles, A. L., Duellman, S., Benink, H. A., Worzella, T. J., &

    Minor, L. (2004). *Cell Viability Assays*. *Assay Guidance Manual*. Eli Lilly &

    Company and the National Center for Advancing Translational Sciences. Retrieved

    from http://www.ncbi.nlm.nih.gov/pubmed/23805433

Sabari, J. K., Lok, B. H., Laird, J. H., Poirier, J. T., & Rudin, C. M. (2017). Unravelling

    the biology of SCLC: implications for therapy. *Nature Reviews. Clinical Oncology*,

    14(9), 549. https://doi.org/10.1038/NRCLINONC.2017.71

Semenova, E. A., Nagel, R., & Berns, A. (2015). Origins, genetic landscape, and

    emerging therapies of small cell lung cancer. *Genes & Development*, *29*(14), 1447–

    1462. https://doi.org/10.1101/gad.263145.115

Sinclair, W. K., & Ross, D. W. (1969). Modes of Growth in Mammalian Cells.

    Biophysical Journal, *9*(8), 1056. https://doi.org/10.1016/S0006-3495(69)86436-2

Understanding What Cancer Is: Ancient Times to Present. (n.d.). Retrieved March 11,

    2022, from https://www.cancer.org/cancer/cancer-basics/history-of-cancer/what-is-

    cancer.html

Uzunoglu, S., Karaca, B., Atmaca, H., Kisim, A., Sezgin, C., Karabulut, B., & Uslu, R.

(2010). Toxicology Mechanisms and Methods Comparison of XTT and Alamar blue

   assays in the assessment of the viability of various human cancer cell lines by AT-

   101 (−/− gossypol). https://doi.org/10.3109/15376516.2010.508080

Wang, P., Henning, S. M., & Heber, D. (2010). Limitations of MTT and MTS-Based

   Assays for Measurement of Antiproliferative Activity of Green Tea Polyphenols.

   PLOS ONE, *5*(4), e10202. https://doi.org/10.1371/JOURNAL.PONE.0010202

Wooten, D. J., Groves, S. M., Tyson, D. R., Liu, Q., Lim, J. S., Albert, R., … Quaranta,

   V. (2019). Systems-level network modeling of Small Cell Lung Cancer subtypes

   identifies master regulators and destabilizers. *PLOS Computational Biology*,

   15(10), e1007343. https://doi.org/10.1371/JOURNAL.PCBI.1007343

Yusa, K., Zhou, L., Li, M. A., Bradley, A., & Craig, N. L. (2011). A hyperactive piggyBac

   transposase for mammalian applications. *Proceedings of the National Academy of

   Sciences of the United States of America*, *108*(4), 1531–1536.

   https://doi.org/10.1073/PNAS.1008322108

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., …

   Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells.

   Nature Communications, *8*, 14049. https://doi.org/10.1038/ncomms14049

```python
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Sat Jul 31 14:06:51 2021

@author: claytonwandishin
"""
import os
import pandas as pd
from itertools import repeat
import datetime
import math
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn import preprocessing
from scipy import stats
import scipy.optimize as opt

#this script is for determing the slope of the luminescence or direct count
data when accounting for minimum luminescence being reached, since if you
don't it slopes up after the first total potency concentration

def ll4(x,b,c,d,e):
    '''This function is basically a copy of the LL.4 function from the R drc
package with
    - b: hill slope
    - c: min response
    - d: max response
    - e: EC50'''
    return(c+(d-c)/(1+np.exp(b*(np.log(x)-np.log(e)))))
def ll3u(x,b,c,e):
    '''This function is basically a copy of the LL.3u function from the R drc
package with
    - b: hill slope
    - c: min response
    - e: EC50'''
    #return(c+((1-c)/(1+(np.log(x/e)**b))))
    return(c+((1-c)/(1+np.exp(b*(np.log(x)-np.log(e))))))
def pDose(x):
    '''This is just a helper function, to compute easily log transformed
concentrations used in drug discovery'''
    return(np.log10(x))

def IC50(EC50,hill,maxv,minv):
    return(np.exp(np.log(EC50)+(1/hill)*np.log(maxv/(maxv-(2*minv)))))
debugdf = pd.read_csv('/Users/claytonwandishin/December 14 RT glow
run/20201216_Lum_CellCounts_TOTAL.csv')
data = pd.read_csv('/Users/claytonwandishin/December 14 RT glow
run/050819_RTGlowDF.csv')
#data =data.rename(columns={'cell.line':'Cell_Line','well':'Well',
'drug1':'Drug','drug1.conc':'Drug_Conc'})

#df.rename(columns={"A": "a", "B": "c"})
```

```
data = data.rename(columns={'TotHour':'TotHour_Lum', 'well':'Well',
'drug1':'Drug','cell.line':'Cell_Line', 'drug1.conc':'Drug_Conc'})
debugdf = debugdf.rename(columns={'TotHour':'TotHour_Lum', 'well':'Well',
'drug1':'Drug','cell.line':'Cell_Line', 'drug1.conc':'Drug_Conc'})

drugLL = data['Drug'].unique()


#####used during debug



#data = data[data.Well != 'J12']
#data = data[data.Well != 'K12']
#data = data[data.Well != 'K10']
#data = data[data.Well != 'K11']




##### THIS IS WHERE THE LUM RATES ARE STORED
dimrate =
pd.DataFrame(columns=['Well','DIMr','Cell_Line','Drug','R2_DIM_Slice','DIM_Sl
ice'])

IC50df = pd.DataFrame(columns=['Cell_Line','Drug','IC50_Lum'])
cline = ['H1930']
drug = ['SNS-314']
druglistref = list(data.Drug.unique())
#conc = [0, 9.96e-06,2.49e-06,6.23e-07]
#conc = [0,9.96e-06]
rrLIST = []
conc = list(data.Drug_Conc.unique())
conc.sort()
#cldata = data.loc[data['Cell_Line'] == 'CORL279']
#cldrugdata = cldata.loc[cldata['Drug'] == 'barasertib']
#for d in conc:
    #sns.lineplot(data=cldrugdata.loc[cldrugdata['Drug_Conc'] == d],
x='TotHour_Lum', y='RLU')
    #plt.show()
dcrr = []
dcrrslice =[]
dcrrlabel =[]
dcrrwell=[]
for CL in cline:
    for dd in drug:
        for cc in conc:
            cldata = data.loc[data['Cell_Line'] == CL]
            cldrugdata = cldata.loc[cldata['Drug'] == dd]
            cldrugdata11c = cldrugdata.loc[cldrugdata['Drug_Conc'] == cc]
```

```
            welly = list(cldrugdata11c.Well.unique())
            for w in welly:
                cldrugdata1c = cldrugdata11c.loc[cldrugdata11c['Well'] == w]
                if cc == 0.0:
                    maxlum0 = cldrugdata1c['RLU'].max()
                    #maxlum0 = 120
                    #the addition of iloc at the end specifies it as a single
value instead of what would be a series object but make sure this works
otherwise you need to manually put in the maxtime to generate slopes from
                    #maxtime0 = cldrugdata1c.query('RLU ==
'+str(maxlum0))['TotHour_Lum'].iloc[0]
                    #maxtime CORL279 AMG900
                    #maxtime0=100
                    #maxtime for H1930 SCH1473759
                    maxtime0 = 93
                    #maxtime0 for H841 AZD1152 (barasertib)
                    #maxtime0= 117.82805555555555
                    #maxtime0 for CORL279 SCH1473759
                    #maxtime0 = 47.5536
                    #H1048 adjustment for TAK901
                    #maxtime0 = 70
                    #H1048 Barasertib adjustment
                    #maxtime0=120
                    #maxtime0=85

                    cldrugdata1c =
cldrugdata1c.loc[cldrugdata1c['TotHour_Lum'] <= maxtime0]
                    cldrugdata1c=cldrugdata1c.loc[cldrugdata1c['Well'] == w]
                    '''
                    slope, intercept, r_value, pv, se =
stats.linregress(cldrugdata1c['TotHour_Lum'],cldrugdata1c['RLU'])
                    rr = r_value**2
                    rrLIST.append(rr)
                    dimrate = dimrate.append({'Well':w,'DIMr':slope},
ignore_index=True)
                    '''
                    sns.set(style='darkgrid')
                    sns.regplot(data=cldrugdata1c,
x=cldrugdata1c['TotHour_Lum'],    y=cldrugdata1c['RLU'], ci=95,
fit_reg=False)
 #############################

                    maxtime=maxtime0
                    cldrugdata3c =
cldrugdata1c.loc[cldrugdata1c['TotHour_Lum'] <= maxtime]
                         #here a loop needs to be written to take the unique
time points that are left and put them in a list, then from that list take
increasing slices as ilocs from the end stored as a new dataframe until the
R2 value reaches >.90, then store that time as an object mintime and slice
the original dataframe from that mintime >=
                    timepoints = cldrugdata3c['TotHour_Lum'].unique()
                    numtp = len(timepoints)-3
                    #something gets messed up here sometimes but what needs
to be accomlished is not allowing the algo to take a slope of a 3 point line
                    intloclisttp = [*range(0,numtp)]
                    rrlist2=[]
                    ltp = len(timepoints)
```

```
                        for t in intloclisttp:
                            slic=timepoints[t]
                            slicesize=ltp-t
                            mintime = slic
                            #print(str(mintime))
                            cldrugdata2c =
cldrugdata3c.loc[cldrugdata3c['TotHour_Lum'] >= mintime]
                            slope, intercept, r_value, pv, se =
stats.linregress(cldrugdata2c['TotHour_Lum'],cldrugdata2c['RLU'])
                            rr=r_value**2
                            dcrr.append(rr)
                            dcrrslice.append(slicesize)
                            dcrrlabel.append(cc)
                            dcrrwell.append(w)
                            rrlist2.append(rr)
                            #print(str(rr))
                        #CONVERT THIS TO A SERIES
                        rrlist3=pd.Series(rrlist2)
                        rrmax = rrlist3[rrlist3 == rrlist3.max()].index[0]
                        mintime=timepoints[rrmax]
                        cldrugdata1c =
cldrugdata1c.loc[cldrugdata1c['TotHour_Lum'] >= mintime]
                        SlicePoints = list(cldrugdata1c['TotHour_Lum'])
                        slope, intercept, r_value, pv, se =
stats.linregress(cldrugdata1c['TotHour_Lum'],cldrugdata1c['RLU'])
                        rr=r_value**2
                        dimrate =
dimrate.append({'Well':w,'DIMr':slope,'Cell_Line':CL,'Drug':dd,'R2_DIM_Slice'
:rr,'DIM_Slice':SlicePoints}, ignore_index=True)
                        print('Maximum R2 for '+CL+' treated with '+dd+' at
'+str(cc)+'in well '+w+'is '+str(rr))
                        sns.regplot(data=cldrugdata1c,
x=cldrugdata1c['TotHour_Lum'],   y=cldrugdata1c['RLU'], ci=95)
                        #sns.regplot(data=indivdrugdata,
x=indivdrugdata['TotHour'],    y=indivdrugdata['RLU'])
                        sns.set(style='darkgrid')
                        #plt.xlim(0, 100)
                        plt.ylabel('')
                        plt.xlabel('')
                        plt.xlim(0,maxtime)
                        plt.legend(title=None, loc='upper left',
labels=[r'$R^2:{0:.3f}$  CI:95%'.format(r_value**2)])
                        #plt.rcParams.update({'font.size': 25})
                        plt.title(CL+' treated with '+str(cc)+'M '+dd+' in well
'+w)
                        plt.savefig('/Users/claytonwandishin/December 14 RT glow
run/plots/PaperFigs/DRC_Lum/H1048/AlgoV2/'+CL+'_'+dd+'/IndivRegLines/'+w+'at'
+str(cc)+'Reg'+'.png', dpi=300, bbox_inches='tight')
                        plt.show()
```

```python
    ########
                        if rr <.90:
                            print('Regression model for '+CL+' treated with
'+dd+' at '+str(cc)+'M is <.95 MODEL DOES NOT FIT R value ='+str(r_value**2))
                            sns.regplot(data=cldrugdata1c,
x=cldrugdata1c['TotHour_Lum'],    y=cldrugdata1c['RLU'], ci=95)
                            #sns.regplot(data=indivdrugdata,
x=indivdrugdata['TotHour'],    y=indivdrugdata['RLU'])
                            sns.set(style='darkgrid')
    #plt.xlim(0, 100)
                            plt.ylabel('')
                            plt.xlabel('')
                            plt.xlim(0,maxtime0)
                            plt.legend(title=None, loc='upper left',
labels=[r'$R^2:{0:.3f}$  CI:95%'.format(r_value**2)])
                            #plt.rcParams.update({'font.size': 25})
                            plt.title(CL+' treated with '+str(cc)+'M '+dd+' in
'+w)
                            #plt.savefig('/Users/claytonwandishin/December 14 RT
glow
run/plots/PaperFigs/DRC_Lum/H1048/AlgoV2/H1048_Barasertib/'+w+'at'+str(cc)+'.
png', dpi=300, bbox_inches='tight')
                            plt.show()
                    elif rr>.90:
                        print('Regression model for '+CL+' treated with
'+dd+' at '+str(cc)+' is >.95 PASS')
                        sns.regplot(data=cldrugdata1c,
x=cldrugdata1c['TotHour_Lum'],    y=cldrugdata1c['RLU'], ci=95)
                        #sns.regplot(data=indivdrugdata,
x=indivdrugdata['TotHour'],    y=indivdrugdata['RLU'])
                        sns.set(style='darkgrid')
    #plt.xlim(0, 100)
                        plt.ylabel('')
                        plt.xlabel('')
                        plt.xlim(0,maxtime0)
                        plt.legend(title=None, loc='upper left',
labels=[r'$R^2:{0:.3f}$  CI:95%'.format(r_value**2)])
                        #plt.rcParams.update({'font.size': 25})
                        plt.title(CL+' treated with '+str(cc)+'M '+dd+' in
'+w)
                        #plt.savefig('/Users/claytonwandishin/December 14 RT
glow
run/plots/PaperFigs/DRC_Lum/H1048/AlgoV2/H1048_Barasertib/'+w+'at'+str(cc)+'.
png', dpi=300, bbox_inches='tight')
                        plt.show()

                elif cc != 0:
                    #this minimun value may be cell line and or drug depndent
and this code should be edited to reflect that
                    minlum=1000
                    cldrugdata1c =
cldrugdata1c.loc[cldrugdata1c['TotHour_Lum'] <= maxtime0]
                    cldrugdata1c=cldrugdata1c.loc[cldrugdata1c['Well'] == w]
                    mincheckRLUcol = cldrugdata1c['RLU'].min()
                    maxldrug = cldrugdata1c['RLU'].max()
```

```
                              #max time needs to first check if the minimum value is
reached for that drug concentration, then if not it needs to use the maxtime
value from the zero concentration and use that as the maxtime
                              #This plotting function shows ALL of the points without a
regression which is necessary for pretty figures, by adding to the regplot
below we get the orange regression line
                              sns.regplot(data=cldrugdata1c,
x=cldrugdata1c['TotHour_Lum'],    y=cldrugdata1c['RLU'], ci=95,
fit_reg=False)
                              #This should be adjusted to loop through like the else
function below
                              if mincheckRLUcol <= minlum:
                                  maxtime = cldrugdata1c.query('RLU <=
'+str(minlum))['TotHour_Lum'].iloc[0]
                                  cldrugdata1c =
cldrugdata1c.loc[cldrugdata1c['TotHour_Lum'] <= maxtime]
                                  slope, intercept, r_value, pv, se =
stats.linregress(cldrugdata1c['TotHour_Lum'],cldrugdata1c['RLU'])
                                  rr=r_value**2
                                  SlicePoints = list(cldrugdata1c['TotHour_Lum'])
                                  dimrate =
dimrate.append({'Well':w,'DIMr':slope,'Cell_Line':CL,'Drug':dd,'R2_DIM_Slice'
:rr,'DIM_Slice':SlicePoints}, ignore_index=True)
                                  sns.regplot(data=cldrugdata1c,
x=cldrugdata1c['TotHour_Lum'],    y=cldrugdata1c['RLU'], ci=95,
fit_reg=False)
                                  #sns.regplot(data=individrugdata,
x=individrugdata['TotHour'],    y=individrugdata['RLU'])
                                  sns.set(style='darkgrid')
                                  #plt.xlim(0, 100)
                                  plt.ylabel('')
                                  plt.xlabel('')
                                  plt.xlim(0,maxtime)
                                  plt.legend(title=None, loc='upper left',
labels=[r'$R^2:{0:.3f}$  CI:95%'.format(r_value**2)])
                                  #plt.rcParams.update({'font.size': 25})
                                  plt.title(CL+' treated with '+str(cc)+'M '+dd+' in
well '+w)
                                  plt.savefig('/Users/claytonwandishin/December 14 RT
glow
run/plots/PaperFigs/DRC_Lum/H1048/AlgoV2/'+CL+'_'+dd+'/IndivRegLines/'+w+'at'
+str(cc)+'.png', dpi=300, bbox_inches='tight')
                                  plt.show()
                              else:
                                  maxtime = maxtime0
                                  mintime = 0
                                  cldrugdata3c =
cldrugdata1c.loc[cldrugdata1c['TotHour_Lum'] <= maxtime]
                                  #here a loop needs to be written to take the unique
time points that are left and put them in a list, then from that list take
increasing slices as ilocs from the end stored as a new dataframe until the
R2 value reaches >.90, then store that time as an object mintime and slice
the original dataframe from that mintime >=
                                  timepoints = cldrugdata3c['TotHour_Lum'].unique()
                                  numtp = len(timepoints)-3
```

```python
                            #something gets messed up here sometimes but what
needs to be accomlished is not allowing the algo to take a slope of a 2 point
line
                            intloclisttp = [*range(0,numtp)]
                            rrlist2=[]
                            ltp = len(timepoints)

                            for t in intloclisttp:
                                slicesize=ltp-t
                                slic=timepoints[t]
                                mintime = slic
                                #print(str(mintime))
                                cldrugdata2c =
cldrugdata3c.loc[cldrugdata3c['TotHour_Lum'] >= mintime]
                                slope, intercept, r_value, pv, se =
stats.linregress(cldrugdata2c['TotHour_Lum'],cldrugdata2c['RLU'])
                                rr=r_value**2
                                dcrr.append(rr)
                                dcrrslice.append(slicesize)
                                dcrrlabel.append(cc)
                                dcrrwell.append(w)
                                rrlist2.append(rr)
                                #print(str(rr))
                            #CONVERT THIS TO A SERIES
                            rrlist3=pd.Series(rrlist2)
                            rrmax = rrlist3[rrlist3 == rrlist3.max()].index[0]
                            mintime=timepoints[rrmax]
                            cldrugdata1c =
cldrugdata1c.loc[cldrugdata1c['TotHour_Lum'] >= mintime]
                            SlicePoints = list(cldrugdata1c['TotHour_Lum'])
                            slope, intercept, r_value, pv, se =
stats.linregress(cldrugdata1c['TotHour_Lum'],cldrugdata1c['RLU'])
                            rr=r_value**2
                            dimrate =
dimrate.append({'Well':w,'DIMr':slope,'Cell_Line':CL,'Drug':dd,'R2_DIM_Slice'
:rr,'DIM_Slice':SlicePoints}, ignore_index=True)
                            print('Maximum R2 for '+CL+' treated with '+dd+' at
'+str(cc)+'in well '+w+'is '+str(rr))
                            sns.regplot(data=cldrugdata1c,
x=cldrugdata1c['TotHour_Lum'],    y=cldrugdata1c['RLU'], ci=95)
                            #sns.regplot(data=indivdrugdata,
x=indivdrugdata['TotHour'],    y=indivdrugdata['RLU'])
                            sns.set(style='darkgrid')
                            #plt.xlim(0, 100)
                            plt.ylabel('')
                            plt.xlabel('')
                            plt.xlim(0,maxtime)
                            plt.legend(title=None, loc='upper left',
labels=[r'$R^2:{0:.3f}$  CI:95%'.format(r_value**2)])
                            #plt.rcParams.update({'font.size': 32})
                            plt.title(CL+' treated with '+str(cc)+'M '+dd+' in
well '+w)
                            plt.savefig('/Users/claytonwandishin/December 14 RT
glow
run/plots/PaperFigs/DRC_Lum/H1048/AlgoV2/'+CL+'_'+dd+'/IndivRegLines/'+w+'at'
+str(cc)+'.png', dpi=300, bbox_inches='tight')
                            plt.show()
```

```python
#dimrate = dimrate.fillna()
full_df = cldrugdata.copy()
full_df = full_df.merge(dimrate)
dimratezero = full_df.loc[full_df['Drug_Conc'] == 0]
dimratezeromean = pd.Series(dimratezero['DIMr'].unique()).mean()
dzmean = dimratezeromean
full_df['DIMrNorm'] = full_df['DIMr']/dzmean
#double check to make sure that plotting a datapoint for zero is the best way
to do this if the data has already been normalized
#something is weird with the scaling
full_df['DClog']= np.log10(full_df['Drug_Conc']+.000000000001)
#full_df['DIMrNormal'] =
#sns.lmplot(x='DClog',y='DIMrNorm',data=full_df,fit_reg=False)
#removes erroneous points for the H1048 YM-155 data
#full_df = full_df[full_df.Drug_Conc != 9.340000000000001e-09]
#full_df = full_df[full_df.Drug_Conc != 3.89e-08]
#removes errod in H841 TAK901 dataset
#full_df = full_df[full_df.Drug_Conc != 9.96e-06]
#removes error in the H841trametinib dataset
#full_df = full_df[full_df.Drug_Conc != 7.779999999999999e-11]
#removes error in the H841 SNS314 dataset
#full_df = full_df[full_df.Drug_Conc != 7.779999999999999e-11]
#this well error could also be fixed by forcing the slope
#full_df = full_df[full_df.Well != 'D22']
#this well error could also be fixed by forcing slope H526 with barasertib
#full_df = full_df[full_df.Well != 'H11']
#this is well errors for DMS114 with barasertib
#full_df = full_df[full_df.Well != 'F11']
#full_df = full_df[full_df.Well != 'G11']
#full_df = full_df[full_df.Well != 'H11']
#full_df = full_df[full_df.Well != 'F07']
#full_df = full_df[full_df.Well != 'G07']
#full_df = full_df[full_df.Well != 'H07']
#DMS114 with AMG900
#full_df = full_df[full_df.Well != 'C12']
#fixes well error for DMS114TAK901
#full_df = full_df[full_df.Well != 'D20']
#fixes well error for H1048 with barasertib
#full_df = full_df[full_df.Well != 'K23']
#full_df = full_df[full_df.Well != 'K16']
#fixes H1048 YM-155 errors
#full_df = full_df[full_df.Well != 'E08']
#full_df = full_df[full_df.Well != 'I08']
#full_df = full_df[full_df.Well != 'M08']
#fixes h1048 trametinib
#full_df = full_df[full_df.Well != 'I22']
#fixes h1048 SNS314
#full_df = full_df[full_df.Well != 'L13']
#full_df = full_df[full_df.Well != 'D14']
#fixes H841 barasertib
#full_df = full_df[full_df.Well != 'K13']
#fixes H841 hygromycin
#full_df = full_df[full_df.Well != 'N23']
#fixes H841 SNS314
#full_df = full_df[full_df.Well != 'H22']
#before applying this, the entire dataframe needs to be truncated so that the
residuals make more sense
```

```python
#DMS1114 TAK901
#full_df = full_df[full_df.Well != 'E20']
#full_df = full_df[full_df.Well != 'E19']
#DMS114 YM155
#full_df = full_df[full_df.DIMrNorm < 1.24]
#full_df = full_df[full_df.Well != 'F18']
#full_df = full_df[full_df.Well != 'G18']
#full_df = full_df[full_df.Well != 'H18']
#CORL279 SCH1473759 fix
#full_df = full_df[full_df.Well != 'K10']
#full_df = full_df[full_df.Well != 'K11']
#full_df = full_df[full_df.Well != 'K12']
#H1930 SCH1473759
#full_df = full_df[full_df.DIMrNorm > -1.0]
#H1930 etoposide
'''
full_df = full_df[full_df.Well != 'J22']
full_df = full_df[full_df.Well != 'K22']
full_df = full_df[full_df.Well != 'J21']
full_df = full_df[full_df.Well != 'K21']
full_df = full_df[full_df.Well != 'J20']
full_df = full_df[full_df.Well != 'K20']
full_df = full_df[full_df.Well != 'I19']
full_df = full_df[full_df.Well != 'I18']
full_df = full_df[full_df.Well != 'K18']
full_df = full_df[full_df.Well != 'K14']
full_df = full_df[full_df.Well != 'K15']
full_df = full_df[full_df.Well != 'I15']
full_df = full_df[full_df.Well != 'K16']
full_df = full_df[full_df.Well != 'I16']
full_df = full_df[full_df.Well != 'J17']
'''
#H1930 SNS-314 data does not look good
'''
full_df = full_df[full_df.Well != 'L12']
full_df = full_df[full_df.Well != 'M12']
full_df = full_df[full_df.Well != 'N12']
full_df = full_df[full_df.Well != 'L11']
full_df = full_df[full_df.Well != 'M11']
full_df = full_df[full_df.Well != 'N11']
full_df = full_df[full_df.Well != 'M10']
full_df = full_df[full_df.Well != 'N10']
full_df = full_df[full_df.Well != 'N07']
full_df = full_df[full_df.Well != 'N03']
'''
compoundData = full_df.groupby(['Drug'])
fitData = []
for name,group in compoundData:
    fitCoefs, covMatrix = opt.curve_fit(ll4, group.Drug_Conc,
group.DIMrNorm)
    resids = group.DIMrNorm-group.Drug_Conc.apply(lambda x: ll4(x,*fitCoefs))
    curFit = dict(zip(['b','c','d','e'],fitCoefs))
    curFit['Drug']=name
    curFit['residuals']=sum(resids**2)
    fitData.append(curFit)
    fitCompound = [ item['Drug'] for item in fitData]
    fitTable = pd.DataFrame(fitData).set_index('Drug')
```

```python
    #Export this fit table into a physical file




    IC50valLum =
IC50(fitTable['e'],fitTable['b'],fitTable['d'],fitTable['c']).copy()
    IC50valLum=IC50valLum.reset_index()
    #This is the variable containing the IC50 Value for the cell line and
drug combo
    IC50valLum2=IC50valLum.iat[0,1]
    IC50df = IC50df.append({'Cell_Line':CL,'Drug':dd,'IC50_Lum':IC50valLum2},
ignore_index=True)
    IC50lum =r'IC50 : {:0.3e}'.format(IC50valLum2)+' M'
    refDose = -
np.linspace(min(full_df['DClog'])*1.2,max(full_df['DClog'])*0.8,256)
    refDose = (10**-refDose)
    sns.lmplot(x='DClog',y='DIMrNorm',data=full_df,fit_reg=False)
    for fit in fitData:
        plt.plot([pDose(i) for i in refDose],[ll4(i,*[fit[i] for i in
['b','c','d','e']]) for i in refDose])
    sns.set(style='darkgrid')
    plt.ylabel('Normalized Luminescence Rate')
    plt.xlabel('log Drug Concentration (M)')
    plt.title(CL+' treated with '+dd)
    plt.legend(title=None, loc='upper right', labels=[IC50lum])
    plt.savefig('/Users/claytonwandishin/December 14 RT glow
run/plots/PaperFigs/DRC_Lum/H1048/AlgoV2/DRC/'+CL+'_'+dd+'.png', dpi=300,
bbox_inches='tight')
    plt.show()
full_df = full_df.merge(IC50df)
#####THIS GENERATES THE R2 vs. Slice size dataframe
r2df = pd.DataFrame()
r2df['DrugCon'] = dcrrlabel
r2df['Well'] = dcrrwell
r2df['SliceSize'] = dcrrslice
r2df['R2_Value'] = dcrr

well_list_r2df = r2df['Well'].unique()

def set_custom_palette(series, max_color = 'tab:orange', other_color =
'tab:blue'):
    max_val = pd.Series(series).max()
    print(max_val)
    pal = []

    for item in series:
        if item == max_val:
            pal.append(max_color)
        else:
            pal.append(other_color)
        print(pal)
    return pal

def set_custom_markers(series, max_mark = '^', other_mark = 'o'):
    max_val = series.max()
```

```python
    mar = []

    for item in series:
        if item == max_val:
            mar.append(max_mark)
        else:
            mar.append(other_mark)
    return mar

def set_custom_marker_size(series, max_mar_size = 200, other_mar_size = 50):
    max_val = series.max()
    marsize = []

    for item in series:
        if item == max_val:
            marsize.append(max_mar_size)
        else:
            marsize.append(other_mar_size)
    return marsize

for uw in well_list_r2df:
    r2dfplot = r2df.loc[r2df['Well'] == uw]
    cuspal=set_custom_palette(r2dfplot['R2_Value'])
    cusmar=set_custom_markers(r2dfplot['R2_Value'])
    cussize=set_custom_marker_size(r2dfplot['R2_Value'])
    r2dfplot['cpallette']=cuspal
    sns.set(style='darkgrid')
    sns.scatterplot(data=r2dfplot, x=r2dfplot['SliceSize'],
y=r2dfplot['R2_Value'],
hue=r2dfplot['R2_Value'],hue_order=r2dfplot['R2_Value'],
palette=cuspal,style=r2dfplot['R2_Value'],markers=cusmar,
style_order=r2dfplot['R2_Value'], size = r2dfplot['R2_Value'],
sizes=cussize,size_order=r2dfplot['R2_Value'], legend=False)
    plt.ylabel('')
    plt.xlabel('')
    plt.xlim(19,3)
    #plt.title(CL+' well '+uw)
    #plt.savefig('/Users/claytonwandishin/December 14 RT glow
run/plots/PaperFigs/DRC_Lum/H1048/AlgoV2/'+CL+'_'+dd+'/'+CL+'_well_'+uw+'R2pl
ot.png', dpi=300, bbox_inches='tight')
    plt.show()
#pmdf = pd.read_csv('/Users/claytonwandishin/December 14 RT glow
run/20201216_Lum_CellCounts_TOTAL_DIPcomp.csv')
#full_df = full_df.merge(pmdf)


#full_df.to_csv('/Users/claytonwandishin/December 14 RT glow
run/20201216_Lum_CellCounts_TOTAL_DIM_33.csv')
'''
dclist_r2df = r2df['DrugCon'].unique()
for d in dclist_r2df:
    r2dfplot = r2df.loc[r2df['DrugCon'] == d]
    sns.set(style='darkgrid')
    sns.scatterplot(data=r2dfplot, x=r2dfplot['SliceSize'],
y=r2dfplot['R2_Value'])
    plt.xlim(19,3)
    plt.ylabel('$R^2 Value')
```

```
    plt.xlabel('Timepoint Slice Size (points)')
    plt.title(CL+' with '+str(d)+' Barasertib')
    plt.show()
'''


#####################################################
#####################################################
######################################################
######################################################

# This part creates rate-based DataFrames so that the DRC generated from Lum
and Direct Counting can be compared
'''
NoDrugLumandCounts = full_df.loc[full_df['Drug_Conc'] == 0]
NDLCC = NoDrugLumandCounts
NDLCC['CC+1']=NDLCC['Live_Dead']+1
NDLCC['log2_Live_Dead+1']=np.log2(NDLCC['CC+1'])

Lnorm = NDLCC.loc[:,'log2_Live_Dead+1']
Lnorm = pd.DataFrame(Lnorm)
Lnorm['TotHour_Image']= NDLCC.loc[:,'TotHour_Image']

DrugCon =
['0','0.00000996','0.00000249','0.000000623','0.000000156','0.0000000389','0.
00000000934','0.00000000233','0.000000000623','0.000000000156','0.00000000007
78']

DrugConfloat = [float(x) for x in DrugCon]

cell_line_list = ['H1048']
#drug_list_fig = ['AZD-1152','SNS-314','trametinib']
drug_list_fig = ['barasertib']
#drug_list_fig = completeDF['Drug'].unique()
#etoposide and DMSO fail
#remove 0 in Lum for SCH
#remove 0 for vemurafinib
#remove 7.78e-11 Lum and 9.96 e-06 for direct
#remove 2.49e-6 and 7.78e-11 for azd1152
DrugConfloat.sort()
DrugConfloatDIRECT = DrugConfloat.copy()
#DrugConfloatDIRECT.remove(0)
#DrugConfloatDIRECT.remove(2.33e-09)
#DrugConfloat.remove(2.49e-06)
#DrugConfloat.remove(0)
#DrugConfloatDIRECT.remove(1.56e-7)
#DrugConfloat.remove(9.96e-06)
#DrugConfloat.remove(7.78e-11)
DrugConfloatDIRECT.sort()
DipDF = pd.DataFrame()
DipDF['Drug_Con']=[]
DipDF['DIP_Rate']=[]
DipDF['Count_Type']=[]
DipDF['DIP_Rate_Norm']=[]
DipDF['Response_Ratio']=[]
DipDFw = []
count_type = ['Direct','Lum']
for CLL in cell_line_list:
```

```
    for DD in drug_list_fig:
        oneCLdf = full_df.loc[full_df['Cell_Line'] == CLL]
        OCLDF = oneCLdf
        onedrugdf = OCLDF.loc[OCLDF['Drug'] == DD]
        ODDF = onedrugdf
        ODDFt100 = ODDF.loc[ODDF['TotHour_Image']<maxtime0]
        ODDftLum = ODDF.loc[ODDF['TotHour_Lum']<maxtime0]


        #DrugConfloat = completeDF['Drug_Conc'].unique()

        for c in DrugConfloatDIRECT:
            individrugcondip = ODDFt100.loc[ODDFt100['Drug_Conc'] == c]
            slope, intercept, r_value, pv, se =
stats.linregress(individrugcondip['TotHour_Image'],individrugcondip['log2_Live_
Dead'])
            DipDF=DipDF.append([{'Drug_Con':c, 'DIP_Rate':slope,
'Count_Type':'Direct'}], ignore_index=True)
            print(str(c)+" "+str(slope)+"Direct")



        for c in DrugConfloat:
            individrugcondip = ODDftLum.loc[ODDftLum['Drug_Conc'] == c]
            slope, intercept, r_value, pv, se =
stats.linregress(individrugcondip['TotHour_Lum'],individrugcondip['RLU'])
            DipDF=DipDF.append([{'Drug_Con':c, 'DIP_Rate':slope,
'Count_Type':'Lum'}], ignore_index=True)
            print(str(c)+" "+str(slope)+"Lum")




#This needs to be fixed to reflect the different scales of lum and direct
counting
            LumDip = DipDF.loc[DipDF['Count_Type'] == 'Lum']


            DirectDip = DipDF.loc[DipDF['Count_Type'] == 'Direct']

            LumDipMax = LumDip['DIP_Rate'].max()
            DirectDipMax = DirectDip['DIP_Rate'].max()
            for i, row in DipDF.iterrows():
                if DipDF.at[i,'Count_Type'] == 'Lum':
                    DipNorm = DipDF.at[i,'DIP_Rate']/LumDipMax
                    DipDF.at[i,'DIP_Rate_Norm'] = DipNorm


            for i, row in DipDF.iterrows():
                if DipDF.at[i,'Count_Type'] == 'Direct':
                    DipNorm = DipDF.at[i, 'DIP_Rate']/DirectDipMax
                    DipDF.at[i,'DIP_Rate_Norm'] = DipNorm

            LumDipNormMin = LumDip['DIP_Rate_Norm'].min()
            LumDipNormMax = LumDip['DIP_Rate_Norm'].max()
            DirectDipNormMin = DirectDip['DIP_Rate_Norm'].min()
            DirectDipNormMax = DirectDip['DIP_Rate_Norm'].max()
```

```python
        for i, row in DipDF.iterrows():
            if DipDF.at[i,'Count_Type'] == 'Lum':
                DipRR = (DipDF.at[i,'DIP_Rate_Norm']-
LumDipNormMin)/(LumDipNormMax - LumDipNormMin)
                DipDF.at[i,'Response_Ratio'] = DipRR


        for i, row in DipDF.iterrows():
            if DipDF.at[i,'Count_Type'] == 'Direct':
                DipRR = (DipDF.at[i,'DIP_Rate_Norm']-
DirectDipNormMin)/(DirectDipNormMax - DirectDipNormMin)
                DipDF.at[i,'Response_Ratio'] = DipRR


    DipDF['DClog']=pDose(DipDF['Drug_Con']+.0000000001)

    #DipDF['Response_Ratio']=
DipDF.groupby(['Count_Type']).transform(lambda x: (DIP_Rate_Norm -
DIP_Rate_Norm.min())/ LumDipDIP_Rate_Norm.max() - DIP_Rate_Norm.min()))
    #DipDF['DClog']=np.log(DipDF['Drug_Con']+.0000000001)


#FITTING FUNCTION NEEDS TO BE SPLIT INTO TWO
        #DipDFw=DipDF.loc[DipDF['Count_Type'] == f]
    compoundData = DipDF.groupby(['Count_Type'])
    fitData = []
    for name,group in compoundData:
        fitCoefs, covMatrix = opt.curve_fit(ll4, group.Drug_Con,
group.Response_Ratio)
        resids = group.Response_Ratio-group.Drug_Con.apply(lambda x:
ll4(x,*fitCoefs))
        curFit = dict(zip(['b','c','d','e'],fitCoefs))
        curFit['Count_Type']=name
        curFit['residuals']=sum(resids**2)
        fitData.append(curFit)
        fitCompound = [ item['Count_Type'] for item in fitData]
        fitTable = pd.DataFrame(fitData).set_index('Count_Type')

        if name == 'Direct':
            IC50valDirect =
IC50(fitTable['e'],fitTable['b'],fitTable['d'],fitTable['c']).copy()
            IC50valDirect=IC50valDirect.reset_index()
            IC50valDirect2=IC50valDirect.at[0,0]
            IC50direct = r'IC50 : {:0.3e}'.format(IC50valDirect2)+' M'
        if name == 'Lum':
            IC50valLum =
IC50(fitTable['e'],fitTable['b'],fitTable['d'],fitTable['c']).copy()
            IC50valLum=IC50valLum.reset_index()
            IC50valLum2=IC50valLum.at[1,0]
            IC50lum =r'IC50 : {:0.3e}'.format(IC50valLum2)+' M'
        refDose = -
np.linspace(min(DipDF['DClog'])*0.6,max(DipDF['DClog'])*1.5,256)
        refDose = (10**-refDose)*1e6

sns.lmplot(x='DClog',y='Response_Ratio',data=DipDF,hue='Count_Type',fit_reg=F
alse)
        for fit in fitData:
```

```
                plt.plot([pDose(i) for i in refDose],[ll4(i,*[fit[i] for i in
['b','c','d','e']]) for i in refDose])




#IC50lum =r'IC50 : {:0.3e}'.format(IC50valLum2.iloc[0])+' M'
#IC50direct = r'IC50 : {:0.3e}'.format(IC50valDirect2.iloc[0])+' M'
sns.set(style='darkgrid')
plt.ylabel('Response Ratio')
plt.xlabel('log Drug Concentration (M)')
plt.title(CL+" treated with "+DD)
plt.legend(title=None, loc='upper right', labels=[IC50direct,IC50lum])
#plt.show()
#plt.savefig('/Users/claytonwandishin/December 14 RT glow
run/plots/PaperFigs/DRC_Lum/'+CL+'/'+CL+'DirectandLumCOMPARISON_DRC_RESPONSER
ATIO_RAWlum'+DD+'DECrun.png', dpi=300, bbox_inches='tight')
#ODDF.to_csv('/Users/claytonwandishin/December 14 RT glow
run/plots/PaperFigs/DRC_Lum/'+CL+'/'+CL+'YM155_OneDrugDF.csv')
'''
```