High-Throughput siRNA Screening Reveals Epigenetic Proteins that Differentially Regulate Intratumoral Heterogeneity in Triple-Negative Breast Cancer

by

Garrett Johnson

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Biology)
in the University of Michigan
2021

Doctoral Committee:

Professor Duxin Sun, Chair
Associate Professor Jolanta Grembecka
Associate Professor Zaneta Nikolovska-Coleska
Associate Professor Raymond Trievel

Garrett William Johnson

johngarr@umich.edu

ORCID iD: 000-0001-8764-0454

Dedication

This dissertation is dedicated my mom and dad who have supported me in my educational pursuits, even if that meant they couldn't understand anything I learned, and to my family that supported me physically, emotionally, and spiritually when I couldn't take care of myself due to a certain cranial passenger. Thank you to you mom who cooked for me when I couldn't feed myself, but please stop overcooking your vegetables, don't make me say it again. Lastly, I do not dedicate this dissertation to cancer since you have made my life more annoying than it needed to be.

Acknowledgements

First, I would like to thank Dr. Duxin Sun for accepting me into his lab and giving me the opportunity to explore novel areas in triple negative breast cancer research. He broke bad habits and encouraged better ones; he scolded me, kindly, when I needed to refocus; he improved my outlook of the data compared to my rain-cloud opinions. I wouldn't be the person I am without him. Aside from science, he has shown personal interest in ensuring I am well, especially in this past year following my cancer diagnosis (Anaplastic Astrocytoma, Grade III). He placed no pressure on me to return but encouraged me to rest and recover. In fact, he told me to take the whole year off, which unfortunately was not in line with my budgetary constraints. I would like to thank Dr. Zaneta Nikolovska-Coleska, who was my first rotation professor as a naïve and joyful graduate student. She both provided guidance when I needed it but freedom to choose the best path forward. I can't say I ever used my knowledge of FMOC peptide synthesis ever again, but it did earn me a co-authored paper. I would like to thank Dr. Jolanta Grembecka, who terrified me during my graduate school interview by taking notes—who takes notes during an interview? After that, she always provided sound and fair critiques of my work, and the project was better for her input. I would also like to thank Dr. Raymond Trievel, whom in a calm manner always provided key critiques that others had not noticed. During our meetings, he never devolved to sesquipedalian loquaciousness, but he instead listened, observed, and said what would be critical to improve my research.

I would also like to thank my supervisor, Dr. Joseph Burnett, without whom I would have had a far more confusing time without his guidance and direction. Thank you to all the doctors, surgeons,

nurses, nurse practitioners, oncologists, orderlies, and all the others that ensured my cranial passenger remained as small as could be managed by medicine, scalpels, and poisons. Specifically, Dr. Al-Holou, my neurosurgeon and his PA-C Stephanie Catherine Roath, Karen Kluin, my speech pathologist, Dr. Michelle Kim, my radioncologist, and Dr. Yoshie Umemura, my ongoing neuroncologist who gives me all the toxins a man could want.

From the Program in Chemical Biology, I'd like to thank Traci Swan and Laura Howe for managing way too much throughout the year, from recruitment weekends, to program get-togethers, to Holiday parties, and duties that are beyond my ken. None of us PhD Candidates could do what we do without your tireless effort.

Thank you Mel for saving my life and forcing me into the ER; thank you (other) Mel for keeping me (mostly) sane, to Sarah with whom I've had a lot of fun writing (and not writing) together, and all my other friends that entertained and distracted me from the weekly graduate school grind. Thank you to my Dungeons and Dragons groups (Andrew, Caleb, Carolyn, Erin, Mel, Tjiske, and alumni Ashley and Henry) that allowed me to explore and imagine other worlds. Also thank you to Pastor Jeremy and his wife Heather for being a blessing to my family and I during this difficult time—I will never forget your kindness and compassion. Lastly, I would like to thank all current Sun Lab members, especially, Joe Burnett, Nathan Truchan, Madi Traore, and Alek Matvekas for either experiment help or simply conversation to keep me sane.

Table of Contents

List of Tables

List of Figures

Abstract

Triple negative breast cancer (TNBC) is a particularly deadly subtype of breast cancer, owing to the intratumoral heterogeneity present in late-stage metastatic breast cancer. This heterogeneity is characterized by different gene expression programs between cellular subpopulations. This heterogeneity is also observed in a number of TNBC cell lines and efforts have been made to understand both types of heterogeneity. These efforts have focused on genetic mutations, mRNA expression, cellular responses to chemotherapy, or population dynamics. While this research is valuable, there have been fewer investigations into the epigenetic mechanisms of TNBC heterogeneity. This knowledge is critical since aggressive and undifferentiated cells expand in response to chemotherapy and in fewer than 72 hours—a clear epigenetic response. However, there has been an inability previously to address this unknown due to the difficulty of population modeling using the standard methods of CD44/CD24 antibodies and aldehyde dehydrogenase activity. These methods are time consuming and require expertise. Instead, we used CRISPR/Cas-9 gene knock-ins of stem cell marker ALDH1A3_mCherry and epithelial differentiation marker CD24_EGFP. These four interconvertible cell states include stem-like states (ALDH+/CD24-, red), progenitor-like states (ALDH+/CD24+, yellow), epithelial-like states (CD24+, green), and mesenchymal like states (CD24-, no color). Those two fluorescent proteins allowed for imaging the internal hierarchy in the chosen cell line. In this thesis, we intend to answer fundamental questions in TNBC heterogeneity, namely: (1) what are the epigenetic proteins that regulate subpopulations in TNBC? (2) What are the complexes and protein-protein interactions that may play a role? (3) Can this heterogeneity be targeted by protein disruption? (4) Do any of these

protein targets highlight any future translational opportunities? To address these questions, we used this SUM149 reporter cell line to conduct a high-throughput siRNA screen focused on 505 epigenetic gene targets in TNBC. Each population was quantified by flow cytometry. Of the four effects studied, namely changes to epithelial progenitor-like, differentiated mesenchymal-like, differentiated epithelial-like cells, and total viability, 120 top targets were selected for single siRNA high-throughput screening. Of those that had the same consistent effect across the campaign, forty were confirmed by this secondary screening: Five in the epithelial progenitor-like, twenty-three in mesenchymal-like, and twelve in differentiated epithelial-like cell populations. While most of these targets were confirmed by traditional CD44 and CD24 antibody staining and aldehyde dehydrogenase activity, we focused on two proteins: TRIM24 and EPC2 for further validation. TRIM24 had a consistent effect reducing undifferented cells across four TNBC cell lines, while EPC2 successfully increased epithelial-like cells in two. TRIM24's effect is critical since undifferentiated cells are responsible for the aggressive nature of metastatic TNBC. Accordingly, we investigated if a TRIM24 degrader, dTRIM24, could decrease the chemotherapy-induced expansion in undifferented cells, which it successfully attenuated. In summary, this high-throughput campaign successfully identified two proteins that affect undifferented cells in opposite directions. While EPC2 will be useful to study why this population expands, TRIM24 will serve as the main focus as its effect has clear and necessary clinical impact. This point is furthered by dTRIM24, which significantly blocks the expansion of undifferentiated cells. Further work will push towards translational application that can address an unmet need in triple-negative breast cancer therapy. Overall, this work highlights that protein populations can be investigated for their role in establishing or altering intratumoral heterogeneity.

Chapter 1

Background

Abstract

Breast cancer causes the second most cancer deaths in US women. However, for tumors that remain localized, prognosis is good for most types of breast cancer, nearing 100%. Metastatic breast cancer, however, has a far worse five-year relative survival rate. Even worse is triple-negative breast cancer (TNBC). Treatment of TNBC is difficult owing to there being no established targets as well as the high levels of intratumoral heterogeneity. This heterogeneity, as described by the cancer stem cell model, can be separated according to CD44, CD24, and aldehyde dehydrogenase presence and activity. Previous research used different methods, such as RNAseq, scRNAseq, gene sequencing, and lineage tracing to decipher the origins of this intratumoral heterogeneity; however, we and others believe that the answer will be found in the epigenetics of TNBCs. Epigenetics is the study of the processes that genetic material can be stored yet remain useable for cells. This storage is undertaken by enzymes and complexes that alter chromatin structure, histone insertion, removal, or modification, and DNA methylation. However, no other group has yet investigated on a broad scale this field, primarily due to the difficulty of conducting high-throughput screening using traditional antibody and Aldefluor™ staining. Importantly, we

have developed a reporter cell system from SUM149 that allows direct visualization and quantification of different cell subpopulations based on knowledge generated by CD44, CD24, and Aldefluor™ staining. In conjunction with the Center of Chemical Genomics at The University of Michigan, we conducted a live-cell high-throughput screen using siRNA in the reporter cell system. After conducting high-throughput primary and secondary FACS screening, we discovered two proteins: EPC2 and TRIM24. EPC2 is a member of the NuA4 histone acetylase complex. When knocked down by siRNA interference, the percentage of undifferentiated cells increased in the reporter cell line ($p < 0.01$). TRIM24 is multifunctional as a ubiquitin ligase, a transcriptional activator by its associations with BRD4, and a regulator of nuclear receptors and retinoic acid signaling. When TRIM24 is knocked down by siRNA interference, undifferentiated cells ($ALDH^{HI}$) significantly decreased ($p < 0.05$) in multiple cell lines (SUM149, BT-20, and MDA-MB-468). Furthermore, dTRIM24, a TRIM24 degrader, attenuates docetaxel-induced expansion of undifferentiated cells. These undifferentiated cells are a major problem with TNBC therapeutics. This expansion is likely a major contributing factor in metastasis of TNBC. These experiments highlight the translation importance of future work concerning TRIM24, including its protein and gene targets.

Introduction

Targeted therapies are very successful for early-stage breast cancers and allow for the high median relative five-year survival of 98.7% (1, 2). Specifically, these therapies disable certain survival and growth pathways specific for different breast cancer subtypes. For instance, anti-estrogen therapies have been very effective at treating both Luminal A and B cancers (3). In HER2 enriched cancers, anti-HER2 antibodies have produced high levels of remission in those patients (4). While

these therapies are incredibly effective in early stage cancers, metastatic patients suffer a far worse prognosis, and an overall five-year relative survival rate drops to 23% (2). Current therapies fail to treat effectively these late-stage patients where numerous metastases have set in (5). Even in patients that show initial response to chemotherapy inevitably relapse (6). For these late-stage, metastatic patients, single target therapies fail to produce treatment or prolong the survival time or time until relapse occurs. Particularly worse off are patients diagnosed with triple-negative breast cancer.

Triple negative breast cancer (TNBC) is unique among breast cancer subtypes in that the tumors lack expression of estrogen, progesterone, and HER2 receptors, rendering any other targeted therapies irrelevant (7). While early stage TNBCs can be treated effectively by either adjuvant or neo-adjuvant therapies, metastatic triple-negative breast cancer remains largely untreatable (8). Median relative five-year survival for these patients is far lower at 11.5% (2). This poor survival rate can be explained by cancer recurrence, developed drug resistance, or further metastasis by the cancer (9). Current therapies are ineffective at treating the whole cancer, both primary and the numerous metastatic sites. We hypothesize that intratumoral heterogeneity present in triple-negative breast cancers lead to these clinical outcomes.

Triple negative breast cancers are highly heterogeneous diseases with significantly worse progression and survival compared to their non-TNBC counterparts (10-12). Literature has also shown that in TNBC different cancer cell subpopulations are present in any given tumor, each with its own unique chemo-response and resistance (1). Furthermore, the extent of intratumoral heterogeneity is likely to increase over disease progression as cancer cells spread to distant and diverse microenvironments (13). Targeted therapies that inhibit several subpopulations of cancer cells without eliminating all sub-populations are bound to relapse after therapy although they may

show early clinical remission (6). It is therefore necessary to treat each TNBC with rationally designed and selected combination therapies in order to effectively treat patients; however, although there has been increased effort in recent years to understand the origin and cause of this heterogeneity in clinical samples, the true causes remain unclear.

In the last few years, more effort has been undertaken to understand this heterogeneity in TNBC and provide future directions of how it may be treated. A recent work sought to clarify the poor prognosis of TNBC by grouping them by gene expression clusters (14). Another work sought to find ways to target the drug-persister population by targeting untransformed cells with the BET inhibitor JQ1; this treatment produced cell-death (15). Another group sought to understand why TNBC is so aggressive by single cell sequencing, whereby they concluded genomic expression, particularly of the glycosphingolipid metabolism pathway, led to an aggressive phenotype in some patient populations (16). Each of these efforts point towards the likely true cause of TNBC prognosis: lineage heterogeneity. However, these works do not examine the mechanisms by which intratumoral heterogeneity arises.

The Cancer Stem Cell Model Best Describes TNBC's Lineage Heterogeneity

Intratumoral heterogeneity can be described as the variation of cellular identity, and thus variation in gene expression, within a given tumor. This variation can be different even among primary and secondary tumor sites (13). While genetic mutation will be the ultimate answer in the long term, many of these metastatic sites arise before extensive genetic mutations can occur, and in fact, are populated with pre-existing tumor identities (13). Furthermore, a heterogeneous tumor can adapt to treatment and adopt survival mechanisms to escape the effect of taxanes (17). In this case, clear

gene expression patterns can be observed before and after treatment (17). These facts may explain why existing treatments fail to treat the full suite of tumors within a patient: they are not all the same tumor, and even then, the tumor may shift into a chemo-resistant state to avoid cell arrest, apoptosis, or necrosis. Recently, this shift in tumor identity was discovered to occur after treatment had elapsed, and gene signatures indicating an increase of epithelial-mesenchymal transition, extracellular matrix breakdown, and de-differentiation (17).

Any drug alters a given tumor's microenvironment, and certain subtypes within a cancer can survive this "microenvironment + drug" change better than other cell types (18-21). Furthermore, intratumoral heterogeneity implies reasons why targeted therapies will likely fail despite their own merits. For any given target, there will be variation in its expression level within the tumor. Some cells will highly expressive, while others may show no expression whatsoever, such as with HER2+ breast cancer (22). While some other works hint at therapeutic pathways in resistant cells, more work must be done to understand how this intratumoral heterogeneity arises.

There are two competing explanations for intratumoral heterogeneity for any cancer: the stochastic model and the cancer stem cell model. The stochastic model predicts heterogeneity arises at the genetic level by random mutations that produce distinct yet related cell populations. This process can produce cells that may or may not be susceptible to current standards of care. Genomic and screening such as RNAseq attempt to discover to what the tumor or cancer may be susceptible (23-25). While these techniques may not change the first treatment option, they may highlight other future treatment options should first line therapies fail to cause full remission. However, these approaches are more relevant in cancers known to be limited in heterogeneity as the samples are taken from the bulk tumor; thus, these approaches are less relevant in more heterogeneous cancers such as TNBC. In these cancers, particular data is lost in favor of bulk data for the whole tumor.

Other techniques, such as single-cell RNAseq can map sub-populations far better than RNAseq, but scRNAseq and other single-cell technologies are still primarily used in academic settings (26). The other main and competing theory is the cancer stem cell theory. Briefly, this hypothesis proposes that a small population of stem-like cells that are responsible for tumor initiation, for renewing the tumor following treatment, and that these cells are capable of infinite self-renewal and asymmetrical division into more differentiated cells (27). Within breast cancer research, the Wicha lab and others have conducted studies to characterize breast cancer stem cells (BCSCs) using cell surface markers such as CD44 and CD24 as well as the Aldefluor™ enzymatic assay (ALDH+) (28-31). These BCSCs are $CD44^{+}CD24^{-}$ and $ALDH^{+}$ with the Aldefluor™ assay (31, 32). Furthermore, these cells are capable of recapitulating the whole population of cancer cells, including progenitor and differentiated cells of both epithelial and mesenchymal lineages (11). Furthermore, not only are these cells resistant to most forms of chemotherapy, chemotherapy increases the percentage of ALDH+ cells (17). Aside from contributing to recurrence, these BCSCs are also likely responsible for the metastasis typical of TNBC (29). Importantly, these cells are claimed to underlie the heterogeneity of TNBC and the capability of the tumor to metastasize, resist treatment, and recur due to their ability to divide asymmetrically (11, 33-35).

However, targeting the cancer stem cell population has not produced any first line treatments in solid tumors and especially not in TNBC (11, 35). In the earlier days of BCSC research, it was believed that targeting stem-associated pathways would be enough to reduce their ability to divide and recapitulate the tumor (36, 37). However, since these cells are such a small percentage of the overall population, a drug targeting WNT, NOTCH, Hedgehog, or any such stem cell associated pathway would be ineffective against differentiated cells. Thus, the thinking developed that a combination therapy involving an anti-stem drug combined with an anti-differentiated cell

treatment could decrease overall tumor mass and potentially establish a complete response (38). In this situation, rather than try again and find new compounds or combination, more basic biology ought to be performed to discover which proteins control a given lineage cell identity.

When lineage changes occur, that is most likely due to alterations in the transcriptional profile of a given population of cells. Long term, this is most likely due to additional genetic mutations that permanently alters these cells. However, in the short term these "lineage switches" are most likely due to changes in epigenetic regulation within the cells. One such example is the transcription factor ZEB1 that induces EMT in breast cancer cells (39, 40). Importantly, these switches involve recruitment of histone deacetylase enzymes (HDACs) to specific locations to suppresses E-Cadherin, which is an epigenetic process (40). Additionally, SOX4 directs EZH2, the catalytic member of the H3K27 trimethylase PRC2, is a master regulator for that same process (41). If epigenetic processes can produce cancer cells that are more invasive and resistant to chemotherapy, that field bears further investigation in lineage heterogeneity.

Epigenetic Regulation Encompasses the Whole Structure of the Genome

Epigenetics is the study of genetic arrangements in space and time. These arrangements are controlled by proteins that determine the shape of cell's individual genome through processes such as chromatin organization, histone insertion, histone tail modifications, and DNA methylation (42). They control raveling or unraveling the chromosomes, X-chromosomes specific deactivation in females, the insertion or removal of histones, the third-dimensional organization of DNA, the silencing of vast untranslated regions within the chromosome, and they control the extent of expression of specific gene based on how accessible those regions are to DNA transcriptional

machinery (42-45). Each level of regulation, chromatin organization, nucleosome, and DNA itself, has its own member proteins that effect or remove alterations.

Chromatin organizers

Chromatin is the highest level of organization for genetic material, and it is composed of both DNA and protein (46). The DNA itself is wrapped around proteins called histones, and this DNA-protein complex is called a nucleosome. Where these nucleosomes are placed is critical for gene expression or repression as the DNA wrapped around the nucleosomes will be unable to be transcribed without additional actions taken (47, 48). This is one method of gene repression. In order to allow for gene expression, the nucleosome must either have its histones ejected, altered to have pro-transcription histones, or the nucleosome must be shifted to allow DNA access to transcription factors and other regulatory agents (47, 49-51).

For each of these functions, there are three main families of chromatin remodelers: Chromodomain-helicase DNA binding (CHD), SWItch/Sucrose Non-Fermentable (SWI/SNF), INOsitol requiring (INO) 80 family. Each of these families of remodelers may have different functions based on the individual members of the groups. In general, the CHD family spaces nucleosomes, allows access to promoters, and allows for editing of which histone variants are included in the nucleosome; the SWI/SNF family slide or eject nucleosomes and alter chromatid access, whether for repression or expression; the INO family primarily edits histone composition of nucleosomes (52). Each of these families play a role in the regulation of gene expression in all cells. Important in that process is how each of these chromatin modifiers associate with locations of interest.

Histone Modifiers

However, there are finer methods for gene control than histone insertion, sliding, or removal.

The H3 and H4 histones proteins possess “tails” that may be post-translationally modified, such as at lysine, arginine, and serine locations, though the locations vary between proteins. These locations can be marked with methyl-, acetyl-, or phospho- groups. For instance, COMPASS-like chromatin modifiers can simultaneously recognize H3K27me3, demethylating that generally repressive mark, while simultaneously trimethylating H3K4, which is a well-studied activating mark (53). The member protein Menin is responsible for this type of association (54). Additionally, H2A and H2B histones can be ubiquitinated, which can be activating or repressive, respectively (55). An example of a ubiquinating complex is PRC1, which then recruits the H3K27me3 trimethylase complex PRC2 through cofactors AEBP2 and JARID2 for further epigenetic repression (56).

Different marks in different locations affects the gene structure around these regions. The combination of all marks on each histone octamer forms an “epigenetic code” that augment chromatin association by nuclear proteins (57). This code is critical for recognition by readers, a class of epigenetic proteins, that may draw in erasers or writers for a given location (58-60). Additionally, chromatin modifiers are also drawn in by specific histone marks (61). Importantly, epigenetic changes are the main process by which stem cells produce additional differentiated cells in distinct lineages, likely by a primed H3k27me3/H3k4 potentiation at critical gene locations (41, 62, 63). A similar process may occur within a tumor—cells may use epigenetic mechanisms to escape cancer treatments, either directly altering pathways to escape cell death or by selective pressure that ultimately leads to changes in cell morphology, such as widespread hypomethylation and CpG sites or increased histone modification at given regions (64, 65). These epigenetic alterations may be a critical mechanism by which this heterogeneity is maintained and through which the cancer escapes treatment and become a separate, drug resistant population as has been

recently demonstrated in breast cancer (17). In the case BRCA1, which itself leads to hypermethylation, may have its promoter hypermethylated, ensuring the epigenetic silencing of that critical tumor suppressor (66, 67).

Many different histones can compose nucleosomes, predominantly H2A, H2B, H3, and H4 (68). These histones are inserted in doubles, and are inserted such that the resulting nucleosomes are octamers, composed of eight total histones of differing identities (69). Octamers containing H2A and H2B variants can be especially repressive to transcription, while in regions of transcription H3 and H4 histones can be modified to contain specific sequences that signal different operations (70-72). Each of these histones have N-terminal peptide tails that may be altered by different modifications (73, 74). However, many different modifications will be present on a given tail, and these effects may be varied based on the combinations present (75-80). The most well known modifications are H3k4me (meaning "on histone three, lysine 4, there is a monomethylation) indicate an actively transcribed region (81, 82). Several different enzymes catalyze this reaction, such as the complex SET1 or the protein MLL2 among other trithorax proteins (83-85). Conversely, H3K27me3 (histone 3, lysine 27, there is trimethylation) is considered a repressive mark and is catalyzed by the PRC2 complex (86, 87). The overall suite of modifications determines which proteins can associate with these histones due to the specific binding mode of the proteins. For instance, the histone chaperone RbBBP4 is unable to bind H3 when H3k4me3 is present (88).

DNA Modifiers

Epigenetic modifications can even occur at the level of DNA. DNA modifying enzymes such as DNA methyl transferases utilize S-adenosyl-L-methionine (SAM) to modify Deoxycytodine to 5-methyl-cytidine (89). The addition of a methyl group at the 5' position increases the hydrophobicity of that individual nucleotide, and if enough methylations occur in a given period,

that DNA region loses its hydrophilicity and will “coil” together to avoid a largely aqueous environment (90).

Localizing these epigenetic complexes, regardless of their given function, is critical for establishing an epigenetic identity at one given point in time. While all complexes possess readers that have recognition motifs, whether DNA itself or specific histone codes, other complexes must rely upon transcription factors (TFs) to direct them to regions of interest. Such transcription factors vary by recognition sequences, such as DNA itself, dimers with other transcription factors, TF-cofactor interactions, binding to DNA modifications, the shape of the DNA, and the overall genomic context of the binding sites (91). With all of these varied binding motifs, transcription factors can have a profound impact on gene expression or repression based on epigenetic modes.

Thus, a complex feedback loop develops between 1) extent of epigenetic protein expression 2) expression of guide proteins 3) overall 3D structure of the genome 4) availability of substrates for epigenetic modification, erasure, or silencers 5) rate of decay of specific proteins 6) presence of native inhibiting compounds within the cell 7) extent of ubiquitination of epigenetic proteins and 8) cell-cycle involved proteins that activate and deactivate certain epigenetic functions to enable proper mitosis.

Only a proper understanding of epigenetics allows for a better view of the Central Dogma, which is typically stated as DNA, when transcribed, produces a RNA code, that when translated, begets specific proteins. Epigenetics determines if and to what extent the DNA is transcribed.

Known Epigenetic Dysregulation in Cancer and Disease

Epigenetic regulation, and dysregulation, has long been examined for its ability to lead to or exacerbate human diseases and cancer (92-95). One such famous dysregulation is that of BRCA1/2 in triple negative breast cancer (67, 96, 97). Both isoforms are considered tumor suppressors by

hypermethylating oncogenes; however, if there is a mutation of either gene, that leads to higher expectancy of developing breast cancer (98). Additionally, the promoters to these genes may be hypermethylated themselves, leading to lower expression of those genes (99, 100). Genetic mutation and hypermethylation of both alleles can also occur (101). Fortunately, hypermethylation of all copies of BRCA1 may indicate susceptibility to PARP inhibitors (102).

However, there have been no systematic epigenetic studies performed to discover complexes that control differentiation or dedifferentiation patterns within the context of triple negative breast cancer, and specifically, which epigenetic proteins control or influence a given sub-population identity. This is largely due to the technical difficulty of visualizing distinct subpopulations of cells in real time. In TNBC, different subpopulations can be measured by several different methods: antibody staining and ALDEFLUOR assay activity, mammosphere formation, and cell tracing.

CD44, CD24, and the ALDEFLUOR Assay are Used to Distinguish Sub-populations

As a first line measure, antibody staining for the surface markers CD44 and CD24 are used to differentiate between epithelial and mesenchymal cell lines (103). A widely used cell line is SUM149, which is used due to the percentage of cells matching those observed in the clinic (28). In SUM149, $CD44^+$ and $CD24^+$ cells (80-90%) denote epithelial cells, while $CD44^+$ and $CD24^-$ cells denote mesenchymal cells (10-20%) (28). However, this antibody staining only detects differentiated cell populations; with the addition of the ALDEFLUOR™ assay, less differentiated cells may be detected. In SUM149, the $ALDH^+$ cells are measured at around 6-8% of the overall cell population (104, 105). When this assay is combined with antibody staining, the cell proportions are: $CD44^+$ $CD24^+$ $ALDH^+$ 6%, $CD44^+$ $CD24^+$ $ALDH^-$ 75%, $CD44^+$ $CD24^-$ $ALDH^-$ 5-15%, and $CD44^+$ $CD24^-$ $ALDH^+$ 0.3% of the population (106). The $CD24^+$ and $ALDH^+$ denote

epithelial progenitor cells and the $CD24^-$ $ALDH^+$ cells denote the apex stem cells. The progenitor cells only produced cells in that specific lineage, the differentiated cells only produced more differentiated cells, yet the stem cells produced both sets of lineages, both epithelial and mesenchymal. While very useful for *in vitro* and some *in vivo* applications, its applications are limited due to the technical skill and expense required for successful execution of staining methods. Furthermore, this method can only capture cell-states at a specific time, albeit both with accuracy and precision in skilled hands.

The mammosphere production method often used in conjunction with CD44 and CD24 antibody staining with the ALDEFLUOR™ assay (107). Mammospheres, called tumorspheres in other cancers, are spheroid structures that develop in an attachment free environment. In this environment, differentiated cells are unable to survive for long, and it is only less differentiated cells, the progenitor and stem cells, that can survive in this environment (108, 109). The number of spheroids that survive after a given treatment or procedure directly relate to the number of less differentiated cells in the overall population. This spheroid population is produced by primary mammosphere production. However, a second experiment, secondary mammosphere production, can confirm that the mammospheres are true mammospheres rather than clumps of differentiated cells. If a given treatment can truly reduce the less differentiated cell population, then that change will be reflected in the secondary mammosphere production assay. These sets of procedures are considered more accurate than antibody and the ALDEFLUOR assays, but are considerably more time consuming and technically challenging, both in assay execution, mammosphere counting, and establishing the secondary mammosphere assay.

A third method that can be used to trace cell lineages, plot cellular dynamics, and measure cell populations is through lineage tracing, whether by intercalating dyes or cellular barcoding. A

simplistic approach would be to use a DNA intercalating dye, such as 5-bromo-2′-deoxy-uridine, that can remain in stem cell populations for weeks (110). The rate of division and cellular location can be derived from the dilution of this dye, whether *in vitro* or *in vivo*, though there are better tools for *in vivo* research. While retroviral introduction of coded sequences can be used as well, more modern methods involve cellular barcoding. This can be accomplished by different methods, each with their advantages and disadvantages, such as Brainbow, Flpbow, *Polylox*, Gestalt, Scartrace, Linnaeus, mScribe, and MEMOIR (111). While each of these tools and methods are incredibly useful for lineage tracing, cell fate analysis, determining heterogenous cell division rates, mapping lineage trees, outputs of certain treatments, such as drug or gene knock-out, and other application, their execution and time inputs necessary are both at a high-skill level and take up a considerable amount of time to execute, valuable as the results may be. While theoretically very valuable avenues of investigation, use of cellular tracing is limited to a very select applications and select labs capable of performing such experiments, both in terms of price and time inputs.

It is clear that more high-throughput methods are needed to discover how lineage heterogeneity is established in triple-negative breast cancer. Each of these methods is too time consuming, too technically challenging, or too expensive to screen numerous candidates for their effect on overall populations. This necessary tool must allow for accurate determinations of 1) the proportions of cell subpopulations within a bulk cell line 2) those populations at a given time 3) the ability to trace changes of cell subpopulations over time 4) those details in a quick and convenient method. Thus, a fluorescent reporter cell line using already established markers was modified from the TNBC cell line SUM149. CRISPR was used to insert two genes encoding for fluorescent proteins, mCherry and EGFP, after the two main markers for antibody and $ALDH^+$ studies: ALDH1A3 and CD24 respectively. This was performed in such a way that when ALDH1A3 was expressed,

mCherry would also be expressed; likewise, when CD24 was expressed, it would also express EGFP. When induced at the same time, it allowed for exact measuring of each population with far more ease and far quicker results than combined antibody and ALDEFLUOR™ assays. In addition, these fluorescent proteins correlate very well with the other assay from which they were derived, meaning no accuracy or precision was lost with this increased ease and speed. Additionally, these cells may be used to model cellular dynamics in response to drugs, compounds, siRNA or other insults by using live-cell microscopy. The proportion of cell changes can be counted using other programs, and a more in-depth look at subpopulation changes can be postulated based on available evidence. Additionally, and finally, this two-color fluorescent protein reporter cell line can be used to measure changes to cell subpopulations in a high-throughput manner, whether with drugs, compounds, compound fragments, siRNA, or shRNA.

Using this cell line, I conducted a high-throughput siRNA screen targeting 505 epigenetic protein genes to determine which epigenetic proteins direct transitions between different sub-populations within a seventy-two hour time period. This cell line is a powerful tool that allows for fundamental questions to be asked and and answered in a swift manner. This was a first of its kind approach to discerning which proteins lead to the intractability of TNBC's critically poor prognosis in numerous patients.

Thus, in connection with the screening core at The University of Michigan's Center for Chemical Genomics, undertook a screen of 505 gene targets known to encode for epigenetic proteins, whether they be support proteins, localizing proteins, transcription factors, or the enzymes themselves. In conducting this screen, the main outputs would be decreases in set fluorescent signal relative to a negative siRNA control but also viability measured by DAPI signal present within the cells. This multifaceted and novel approach allowed the Sun lab to collect data by FACS

analysis. However, unlike drug high-throughput screening that only requires DMSO be kept below a certain level, siRNA screening is far more complex and requires far more optimization prior to campaign initiation. There are more critical components to a successful siRNA screen compared to compound screening, namely 1) transfection reagent chosen 2) transfection dose 3) selection of positive and negative siRNA controls 4) cell seeding number as well as 5) procedure for maximizing numbers of events in FACS analysis with 6) viability determination. However, it should be noted that viability is recorded not as a critical output but to add context to any alterations in cell populations. For instance, if a given target subpopulation is successfully reduced but overall cell numbers increase, which means that the gene target would be of little to no interest to develop and investigate further.

While epigenetic targets are the most likely to produce the short-term changes necessary for cancer survival and recurrence, there are currently no tools to screen for alterations in population heterogeneity, especially in a high-throughput fashion. This fluorescent reporter cell line overcame that obstacle. By using CRISPR-CAS9 in SUM149, ALDH1A3 is co-expressed mCherry and CD24 is co-expressed with EGFP, illuminating the distinct sub-populations within the overall population in real time. This cell line, combined with the 505 epigenetic-protein-targeting siRNA available from the CCG, it became possible to screen these targets in a high-throughput fashion using FACS analysis. We hypothesize that this two-color reporter cell model will allow us to screen for critical epigenetic regulators of specific subpopulations, and that we can confirm these results in genetically similar TNBC cell lines, both by knockdown and overexpression.

We will test this hypothesis in these two specific aims:

Specific Aims

Aim 1: Optimize and conduct primary and secondary high-throughput siRNA screening

Objectives: A) use fluorescent TNBC reporter cell line to plan and optimize for HTS screening B) use cell line to conduct siRNA screening against 505 epigenetic gene targets B) confirm selected hits with single-siRNA secondary screening C) confirm protein knockdown and effects by western blot, MTS assay, and antibody-based methods

Aim 2: Confirmation of hit gene targets proteins and preliminary translational investigation

Objectives: A) use reporter cell line to repeat selected hits from siRNA screen B) Investigate those hits using new siRNA screening analysis C) Investigate the translation application of dTRIM24 to reduce the negative effects of Docetaxel

References

1. Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, et al. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61-70.

2. National Cancer Institute D, Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov). SEER*Stat Database: Populations - Total U.S. (1969-2018) Breast Cancer Positive - Total U.S., 1969-2018 Counties. Surveillance Research Program: National Cancer Institute; December 2019.

3. Tamoxifen for early breast cancer: an overview of the randomised trials. The Lancet. 1998;351(9114):1451-67.

4. Slamon D, Eiermann W, Robert N, Pienkowski T, Martin M, Press M, et al. Adjuvant Trastuzumab in HER2-Positive Breast Cancer. 2011;365(14):1273-83.

5. Harbeck N, Gnant M. Breast cancer. The Lancet. 2017;389(10074):1134-50.

6. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res. 2007;13(8):2329-34.

7. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001;98(19):10869-74.

8. Masoud V, Pagès G. Targeted therapies in breast cancer: New challenges to fight against resistance. World J Clin Oncol. 2017;8(2):120-34.

9. Barnard ME, Boeke CE, Tamimi RM. Established breast cancer risk factors and risk of intrinsic tumor subtypes. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer. 2015;1856(1):73-85.

10. Granit RZ, Gabai Y, Hadar T, Karamansha Y, Liberman L, Waldhorn I, et al. EZH2 promotes a bi-lineage identity in basal-like breast cancer cells. Oncogene. 2013;32(33):3886-95.

11. Creighton CJ, Li X, Landis M, Dixon JM, Neumeister VM, Sjolund A, et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proc Natl Acad Sci U S A. 2009;106(33):13820-5.

12. Keller PJ, Lin AF, Arendt LM, Klebba I, Jones AD, Rudnick JA, et al. Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. Breast Cancer Res. 2010;12(5):R87.

13. Echeverria GV, Powell E, Seth S, Ge Z, Carugo A, Bristow C, et al. High-resolution clonal mapping of multi-organ metastasis in triple negative breast cancer. Nature Communications. 2018;9(1):5079.

14. Wang D-Y, Jiang Z, Ben-David Y, Woodgett JR, Zacksenhaus E. Molecular stratification within triple-negative breast cancer subtypes. Sci Rep. 2019;9(1):19107.

15. Risom T, Langer EM, Chapman MP, Rantala J, Fields AJ, Boniface C, et al. Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer. Nature Communications. 2018;9(1):3815.

16. Karaayvaz M, Cristea S, Gillespie SM, Patel AP, Mylvaganam R, Luo CC, et al. Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. Nature communications. 2018;9(1):3588-.

17. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, et al. Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell. 2018;173(4):879-93.e13.

18. Soysal SD, Tzankov A, Muenst SE. Role of the Tumor Microenvironment in Breast Cancer. Pathobiology : journal of immunopathology, molecular and cellular biology. 2015;82(3-4):142-52.

19. Käkönen SM, Mundy GR. Mechanisms of osteolytic bone metastases in breast carcinoma. Cancer. 2003;97(3 Suppl):834-9.

20. Dvorak HF. Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. The New England journal of medicine. 1986;315(26):1650-9.

21. Ma XJ, Dahiya S, Richardson E, Erlander M, Sgroi DC. Gene expression profiling of the tumor microenvironment during breast cancer progression. Breast Cancer Res. 2009;11(1):R7.

22. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. The New England journal of medicine. 2001;344(11):783-92.

23. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1-13.

24. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, et al. Genome-Wide Location and Function of DNA Binding Proteins. 2000;290(5500):2306-9.

25. Gatza ML, Lucas JE, Barry WT, Kim JW, Wang Q, D. Crawford M, et al. A pathway-based classification of human breast cancer. 2010;107(15):6994-9.

26. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. Science. 2014;343(6172):776.

27. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature. 2001;414(6859):105-11.

28. Honeth G, Bendahl P-O, Ringnér M, Saal LH, Gruvberger-Saal SK, Lövgren K, et al. The CD44+/CD24- phenotype is enriched in basal-like breast tumors. Breast Cancer Res. 2008;10(3):R53-R.

29. Mani SA, Guo W, Liao M-J, Eaton EN, Ayyanan A, Zhou AY, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell. 2008;133(4):704-15.

30. Charafe-Jauffret E, Ginestier C, Iovino F, Wicinski J, Cervera N, Finetti P, et al. Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. Cancer Res. 2009;69(4):1302-13.

31. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 2003;100(7):3983-8.

32. Charafe-Jauffret E, Ginestier C, Iovino F, Tarpin C, Diebel M, Esterni B, et al. Aldehyde dehydrogenase 1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer. Clin Cancer Res. 2010;16(1):45-55.

33. Liu S, Cong Y, Wang D, Sun Y, Deng L, Liu Y, et al. Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. Stem Cell Reports. 2013;2(1):78-91.

34. Hollestelle A, Nagel JHA, Smid M, Lam S, Elstrodt F, Wasielewski M, et al. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. 2010;121(1):53-64.

35. Li X, Lewis MT, Huang J, Gutierrez C, Osborne CK, Wu MF, et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. Journal of the National Cancer Institute. 2008;100(9):672-9.

36. Wang T, Narayanaswamy R, Ren H, Torchilin VP. Combination therapy targeting both cancer stem-like cells and bulk tumor cells for improved efficacy of breast cancer treatment. Cancer Biol Ther. 2016;17(6):698-707.

37. Naujokat C, Steinhart R. Salinomycin as a drug for targeting human cancer stem cells. Journal of biomedicine & biotechnology. 2012;2012:950658.

38. Yang L, Shi P, Zhao G, Xu J, Peng W, Zhang J, et al. Targeting cancer stem cell pathways for cancer therapy. Signal Transduction and Targeted Therapy. 2020;5(1):8.

39. Chaffer CL, Marjanovic ND, Lee T, Bell G, Kleer CG, Reinhardt F, et al. Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. Cell. 2013;154(1):61-74.

40. Cho HJ, Oh N, Park JH, Kim KS, Kim HK, Lee E, et al. ZEB1 Collaborates with ELK3 to Repress E-Cadherin Expression in Triple-Negative Breast Cancer Cells. Molecular cancer research : MCR. 2019;17(11):2257-66.

41. Tiwari N, Tiwari Vijay K, Waldmeier L, Balwierz Piotr J, Arnold P, Pachkov M, et al. Sox4 Is a Master Regulator of Epithelial-Mesenchymal Transition by Controlling Ezh2 Expression and Epigenetic Reprogramming. Cancer Cell. 2013;23(6):768-83.

42. Gayon J. From Mendel to epigenetics: History of genetics. Comptes rendus biologies. 2016;339(7-8):225-30.

43. Choi JD, Lee J-S. Interplay between Epigenetics and Genetics in Cancer. Genomics Inform. 2013;11(4):164-73.

44. Kiefer JC. Epigenetics in development. Developmental dynamics : an official publication of the American Association of Anatomists. 2007;236(4):1144-56.

45. Kanwal R, Gupta K, Gupta S. Cancer epigenetics: an introduction. Methods in molecular biology (Clifton, NJ). 2015;1238:3-25.

46. Kornberg RD. Chromatin Structure: A Repeating Unit of Histones and DNA. Science. 1974;184(4139):868.

47. Boeger H, Griesenbeck J, Strattan JS, Kornberg RD. Nucleosomes Unfold Completely at a Transcriptionally Active Promoter. Molecular Cell. 2003;11(6):1587-98.

48. Lorch Y, LaPointe JW, Kornberg RD. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. Cell. 1987;49(2):203-10.

49. Venkatesh S, Workman JL. Histone exchange, chromatin structure and the regulation of transcription. Nature reviews Molecular cell biology. 2015;16(3):178-89.

50. Kireeva ML, Walter W, Tchernajenko V, Bondarenko V, Kashlev M, Studitsky VM. Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. Mol Cell. 2002;9(3):541-52.

51. Lai WKM, Pugh BF. Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nature Reviews Molecular Cell Biology. 2017;18(9):548-62.

52. Clapier CR, Iwasa J, Cairns BR, Peterson CL. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. Nature Reviews Molecular Cell Biology. 2017;18(7):407-22.

53. Schuettengruber B, Martinez AM, Iovino N, Cavalli G. Trithorax group proteins: switching genes on and keeping them active. Nature reviews Molecular cell biology. 2011;12(12):799-814.

54. Yokoyama A, Cleary ML. Menin critically links MLL proteins with LEDGF on cancer-associated target genes. Cancer Cell. 2008;14(1):36-46.

55. Chandrasekharan MB, Huang F, Sun ZW. Histone H2B ubiquitination and beyond: Regulation of nucleosome stability, chromatin dynamics and the trans-histone H3 methylation. Epigenetics. 2010;5(6):460-8.

56. Kasinath V, Beck C, Sauer P, Poepsel S, Kosmatka J, Faini M, et al. JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications. Science. 2021;371(6527).

57. Kelly AD, Issa JJ. The promise of epigenetic therapy: reprogramming the cancer epigenome. Current opinion in genetics & development. 2017;42:68-77.

58. Torres IO, Fujimori DG. Functional coupling between writers, erasers and readers of histone and DNA methylation. Current opinion in structural biology. 2015;35:68-75.

59. Biswas S, Rao CM. Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. European journal of pharmacology. 2018;837:8-24.

60. Nakanishi S, Lee JS, Gardner KE, Gardner JM, Takahashi YH, Chandrasekharan MB, et al. Histone H2BK123 monoubiquitination is the critical determinant for H3K4 and H3K79 trimethylation by COMPASS and Dot1. The Journal of cell biology. 2009;186(3):371-7.

61. Morgan MAJ, Shilatifard A. Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. Nat Genet. 2020;52(12):1271-81.

62. Wang C, Lee J-E, Lai B, Macfarlan TS, Xu S, Zhuang L, et al. Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. Proc Natl Acad Sci U S A. 2016;113(42):11871-6.

63. Jørgensen HF, Giadrossi S, Casanova M, Endoh M, Koseki H, Brockdorff N, et al. Stem cells primed for action: polycomb repressive complexes restrain the expression of lineage-specific regulators in embryonic stem cells. Cell cycle (Georgetown, Tex). 2006;5(13):1411-4.

64. Shen L, Kondo Y, Guo Y, Zhang J, Zhang L, Ahmed S, et al. Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS Genet. 2007;3(10):2023-36.

65. Karlić R, Chung H-R, Lasserre J, Vlahovicek K, Vingron M. Histone modification levels are predictive for gene expression. Proc Natl Acad Sci U S A. 2010;107(7):2926-31.

66. Zhu Q, Pao GM, Huynh AM, Suh H, Tonnu N, Nederlof PM, et al. BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. Nature. 2011;477(7363):179-84.

67. Ford D, Easton DF, Bishop DT, Narod SA, Goldgar DE. Risks of cancer in BRCA1-mutation carriers. Breast Cancer Linkage Consortium. Lancet. 1994;343(8899):692-5.

68. Koyama M, Kurumizaka H. Structural diversity of the nucleosome. The Journal of Biochemistry. 2017;163(2):85-95.

69. Kouzarides T. Chromatin Modifications and Their Function. Cell. 2007;128(4):693-705.

70. Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, et al. Role of histone H2A ubiquitination in Polycomb silencing. Nature. 2004;431(7010):873-8.

71. Cole H, Ocampo J, Iben J, Chereji R, Clark D. Heavy transcription of yeast genes correlates with differential loss of histone H2B relative to H4 and queued RNA polymerases. Nucleic Acids Res. 2014;42.

72. Weber Christopher M, Ramachandran S, Henikoff S. Nucleosomes Are Context-Specific, H2A.Z-Modulated Barriers to RNA Polymerase. Molecular Cell. 2014;53(5):819-30.

73. Hyun K, Jeon J, Park K, Kim J. Writing, erasing and reading histone lysine methylations. Experimental & molecular medicine. 2017;49(4):e324.

74. Corujo D, Buschbeck M. Post-Translational Modifications of H2A Histone Variants and Their Role in Cancer. Cancers. 2018;10(3).

75. Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. Science. 1974;184(4139):868-71.

76. Kornberg RD, Thomas JO. Chromatin structure; oligomers of the histones. Science. 1974;184(4139):865-8.

77. Strahl BD, Allis CD. The language of covalent histone modifications. Nature. 2000;403(6765):41-5.

78. Zhou W, Zhu P, Wang J, Pascual G, Ohgi KA, Lozach J, et al. Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Molecular cell. 2008;29(1):69-80.

79. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. Cell. 2007;130(1):77-88.

80. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007;129(4):823-37.

81. Nagy PL, Griesenbeck J, Kornberg RD, Cleary ML. A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. Proc Natl Acad Sci U S A. 2002;99(1):90-4.

82. Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NC, et al. Active genes are tri-methylated at K4 of histone H3. Nature. 2002;419(6905):407-11.

83. Froimchuk E, Jang Y, Ge K. Histone H3 lysine 4 methyltransferase KMT2D. Gene. 2017;627:337-42.

84. Jeon J, McGinty RK, Muir TW, Kim JA, Kim J. Crosstalk among Set1 complex subunits involved in H2B ubiquitylation-dependent H3K4 methylation. Nucleic Acids Res. 2018;46(21):11129-43.

85. Dehé PM, Dichtl B, Schaft D, Roguev A, Pamblanco M, Lebrun R, et al. Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. J Biol Chem. 2006;281(46):35404-12.

86. Yu JR, Lee CH, Oksuz O, Stafford JM, Reinberg D. PRC2 is high maintenance. Genes Dev. 2019;33(15-16):903-35.

87. Margueron R, Justin N, Ohno K, Sharpe ML, Son J, Drury WJ, 3rd, et al. Role of the polycomb protein EED in the propagation of repressive histone marks. Nature. 2009;461(7265):762-7.

88. Moody RR, Lo MC, Meagher JL, Lin CC, Stevers NO, Tinsley SL, et al. Probing the interaction between the histone methyltransferase/deacetylase subunit RBBP4/7 and the transcription factor BCL11A in epigenetic complexes. J Biol Chem. 2018;293(6):2125-36.

89. Morgan HD, Dean W, Coker HA, Reik W, Petersen-Mahrt SK. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. J Biol Chem. 2004;279(50):52353-60.

90. Jones PA, Taylor SM. Cellular differentiation, cytidine analogs and DNA methylation. Cell. 1980;20(1):85-93.

91. Inukai S, Kock KH, Bulyk ML. Transcription factor-DNA binding: beyond binding site motifs. Current opinion in genetics & development. 2017;43:110-9.

92. Gargiulo G, Minucci S. Epigenomic profiling of cancer cells. Int J Biochem Cell Biol. 2009;41(1):127-35.

93. Iacobuzio-Donahue CA. Epigenetic changes in cancer. Annu Rev Pathol. 2009;4:229-49
.
94. Zhou W, Wang X, Rosenfeld MG. Histone H2A ubiquitination in transcriptional regulation and DNA damage repair. Int J Biochem Cell Biol. 2009;41(1):12-5.

95. Savage P, Stebbing J, Bower M, Crook T. Why does cytotoxic chemotherapy cure only some cancers? Nat Clin Pract Oncol. 2009;6(1):43-52.

96. Zhang H, Tombline G, Weber BL. BRCA1, BRCA2, and DNA damage response: collision or collusion? Cell. 1998;92(4):433-6.

97. Dobrovic A, Simpfendorfer D. Methylation of the BRCA1 gene in sporadic breast cancer. Cancer Res. 1997;57(16):3347-50.

98. Bertwistle D, Ashworth A. Functions of the BRCA1 and BRCA2 genes. Current opinion in genetics & development. 1998;8(1):14-20.

99. Mancini DN, Rodenhiser DI, Ainsworth PJ, O'Malley FP, Singh SM, Xing W, et al. CpG methylation within the 5' regulatory region of the BRCA1 gene is tumor specific and includes a putative CREB binding site. Oncogene. 1998;16(9):1161-9.

100. Rice JC, Massey-Brown KS, Futscher BW. Aberrant methylation of the BRCA1 CpG island promoter is associated with decreased BRCA1 mRNA in sporadic breast cancer cells. Oncogene. 1998;17(14):1807-12.

101. Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, et al. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. Journal of the National Cancer Institute. 2000;92(7):564-9.

102. Kondrashova O, Topp M, Nesic K, Lieschke E, Ho GY, Harrell MI, et al. Methylation of all BRCA1 copies predicts response to the PARP inhibitor rucaparib in ovarian carcinoma. Nat Commun. 2018;9(1):3970.

103. Fillmore CM, Kuperwasser C. Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. Breast Cancer Res. 2008;10(2):R25.

104. Tanei T, Morimoto K, Shimazu K, Kim SJ, Tanji Y, Taguchi T, et al. Association of Breast Cancer Stem Cells Identified by Aldehyde Dehydrogenase 1 Expression with Resistance to Sequential Paclitaxel and Epirubicin-Based Chemotherapy for Breast Cancers. Clinical Cancer Research. 2009;15(12):4234.

105. Vasiliou V, Pappa A, Petersen DR. Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism. Chemico-Biological Interactions. 2000;129(1):1-19.

106. Liu S, Cong Y, Wang D, Sun Y, Deng L, Liu Y, et al. Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. Stem Cell Reports. 2014;2(1):78-91.

107. Lombardo Y, de Giorgio A, Coombes CR, Stebbing J, Castellano L. Mammosphere formation assay from human breast cancer tissues and cell lines. J Vis Exp. 2015(97):52671.

108. Grimshaw MJ, Cooper L, Papazisis K, Coleman JA, Bohnenkamp HR, Chiapero-Stanke L, et al. Mammosphere culture of metastatic breast cancer cells enriches for tumorigenic breast cancer cells. Breast Cancer Res. 2008;10(3):R52.

109. Pham PV, Phan NL, Nguyen NT, Truong NH, Duong TT, Le DV, et al. Differentiation of breast cancer stem cells by knockdown of CD44: promising differentiation therapy. Journal of translational medicine. 2011;9:209.

110. Matatall KA, Kadmon CS, King KY. Detecting Hematopoietic Stem Cell Proliferation Using BrdU Incorporation. Methods in molecular biology (Clifton, NJ). 2018;1686:91-103.

111. Kebschull JM, Zador AM. Cellular barcoding: lineage tracing, screening and beyond. Nature Methods. 2018;15(11):871-9.

Chapter 2

Optimization and Execution of High-Throughput siRNA Screen Identifies 582 Gene Targets That Control Sub-Population Identity

Abstract

Despite the fact that a majority of triple-negative breast cancer patients have a high five-year relative survival rate (98%), patients with late-stage metastatic tumors face a grimmer prognosis: ten percent of patients survive five years relative to healthy populations. While this is due in part to a more advanced disease and metastasis at different sites, advanced disease alone cannot explain this dismal outlook. Instead, intratumoral heterogeneity should be the focus of future work in triple-negative breast cancer. Intratumoral heterogeneity is defined as the cell variation present within a tumor or within a metastatic site. Cell surface markers CD44 and CD24 assist in identifying cells, as does the commercially available ALDEFLUOR™ assay kit that measures the enzymatic activity of aldehyde dehydrogenases, which is indicative of stem-like cells. Within the past decade, more studies have been conducted to understand this issue, whether they be at the DNA, RNA, or single-cell RNAseq level. These efforts have focused on either direct patient samples or patient-derived xenographs—cancers that have already taken their course and killed their hosts. However, none of these studies have examined the origins of intratumoral heterogeneity. While there are many protein classes that could be investigated, I chose to investigate epigenetic proteins. This is due to the effect of many chemotherapies used in the clinic: not all of the cells die with treatment, but instead they shift to other cell states so that they might survive therapy. While further genetic mutation is a certainty, in the near time these are not due to genetic alterations but rather short-term epigenetic shifts that prolong a cell's life in this new

environment. There must be a population of epigenetic proteins that control this survival shift. Therefore, I designed, optimized, and executed a high-throughput siRNA screen focused on 505 epigenetic gene targets.

Introduction

Intratumoral heterogeneity remains the scourge of Triple-Negative Breast Cancer (1-4). Each individual cancer has its own set of populations, both mesenchymal and epithelial, differentiated and undifferentiated (3, 5-8). Each of these cell types respond differently to the primary treatment options, such as radiation and/or taxanes (2, 9). With docetaxel, the differentiated cells mostly die but the undifferentiated cells, both cancer stem cells and progenitors, are largely unaffected by this treatment, and according to our unpublished data, they de-differentiate to a more stem-like state (9, 10). These stem-like cells among the most deadly of cell types and can lead to cancer survival, recurrence, and metastasis (10). The intractability of these populations lead to an overall 5-year survival rate of 11% in TNBC, as there is currently no way clinically to separate out the distinct populations within a given cancer (11). For heterogenous cancers such as TNBC, the poor prognosis will never be addressed until this intratumoral heterogeneity is understood.

While previously there have been efforts to understand how TNBC cells can switch lineages, epithelial to mesenchymal and vice versa, these efforts focused only on switching lineages and an effort to understand how cancer stem-like cells might arise (12-14). Various stem-related pathways have been proposed as potential therapeutic pathways, such as NOTCH, WNT, Hh, but no therapeutic breakthroughs have occurred (15-17). This reality may indicate that two broad classifications, stem and differentiated cells, are insufficient in describing the populations present in any given cancer. Furthermore, such imprecise descriptions have not improved the grim prognosis in late-stage triple-negative breast cancer patients. A newer system, one that can recapitulate a broader heterogeneity, is necessary to reflect relevant breast cancer populations.

However, there were many directions we could take when dissecting this intratumoral heterogeneity, specifically which proteins should be investigated. While a whole exome siRNA screen would examine everything within the cell, such a broad search would be unnecessary. Instead, we focused on epigenetic proteins. When working with TNBC cell lines, cell state changes

are observed within 72 hours. Undoubtedly, this is not due to further genetic mutations but instead transcription changes effected by epigenetic disturbance. Therefore, I screened epigenetic proteins to increase the number of productive hits that the assay screen would identify. However, a means to detect heterogeneity needed to be developed.

Thus, we developed a two-color reporter cell using already established population markers: CD24 and ALDH1A3. CD24 is coexpressed with EGFP, and ALDH1A3 is co-expressed with mCherry. This tool allows researchers to avoid lengthy, expensive, and technically challenging antibody additions and Aldefluor assay kits. Furthermore, this tool allows researchers to determine effects of drugs, siRNA, or any such perturbation has on cellular heterogeneity with far less post-experiment preparation for fluorescence activated cell sorting (FACS) analysis, enhancing lab efficiency without sacrificing quality of data. This is the very tool necessary to push TNBC research to the next level.

Importantly, this two-color reporter cell system allows for siRNA and drug screening in a high throughput fashion. The time required to examine the individual cell populations using traditional approaches would render such an experiment extremely burdensome to researchers and would vastly limit the size and scope of the campaign (18). Additionally, this system permits researchers not only to measure the extent of cell death or arrest but also see if one or more populations is primarily affected. With further experiments, we can even see how a given population increases or decreases in response to an insult. However, this data-rich approach also demands great care in data analysis as there are multiple outputs that must be measured and compared to control wells; additionally, analysis is necessarily more complex as the three populations measured are more complex than typical assays focusing on extent of cell death, which is more “yes-no” rather than measuring an extent of change. While there is increased pressure put on data analysis, that pressure may be relieved in part by careful high-throughput screen design.

High-throughput screening (HTS) is a general method by which assays are miniaturized such that a given output can be measured on a smaller scale, such as 96, 384, 3456, or even 6144 well plates (19). HTS has been used successfully over the past thirty years to screen compounds and repurposed drugs against targets (19). Despite initial screens using natural products, this method has largely replaced hunting and hoping for natural compounds due to the expensive nature of the process and the time involved in discovering and replicating natural biological processes that develop natural product’s often complex structure (20). Although enzyme-based synthesis of these

natural products may have a positive impact in the future, HTS will remain the predominant method of finding hits against a well-validated drug target (20).

While most HTS campaigns focus on discovering and validating confirmed compound interaction with the protein of interest, TNBC research is not yet at that level of specificity. Instead, the targets must be discovered for future research and validation. To that end, I optimized and conducted a high-throughput siRNA screen against 505 epigenetic gene targets using a SUM149-derived two-color reporter cell line.

Results

The project goal for this high-throughput screening campaign is to develop a list of gene targets that control cell identity across multiple triple-negative breast cancer cell lines. While this present project is focused on basic biology, the future goal anticipates translational development of compounds, chemical probes, and ultimate therapeutics focused on a well-validated protein target. To that end, additional optimization is required for a high-throughput, live cell, and siRNA screen. These include 1) transfection agent and dose 2) transfection mode 3) selected class of gene targets 4) positive and negative control selection 5) cell-handling techniques 6) integrating live cell methods with FACS analysis 7) high-throughput methods of data analysis 8) statistical threshold for hit selection.

High-Throughput Optimization

For SUM149, RNAimax is the most efficient reagent for siRNA knockdown; however, the dose chosen contributed to unwanted cell death irrespective of the siRNA dosed in a given experiment. As decreased viability is unwanted in a high throughput screen, the right dose chosen must ensure 1) as little decrease in viability as possible as well as 2) effective transcriptional disruption. To determine effect on decreased cell viability, an RNAimax titration experiment was performed, and it was discovered that 0.3 uL/well decreased cell viability the least. (Figure 2.1). However, additional experiments would be required to determine if this level of transfection reagent is sufficient for siRNA delivery.

Every high-throughput screen must possess a positive control that reflects the demands of assay structure and output (21). In the case of siRNA screening, the positive control should demonstrate that transfection was successful in that given well rather than giving some maximum or minimum assay output as in other assay designs. Often times, non-targeted siRNA will be tethered to a fluorescent probe, and the level of fluorescence can be used to determine transfection efficiency for that well. However, since both mCherry, EGFP, and DAPI will be used, a fluorescent probe would largely be impractical. Instead, siRNA targeting ALDH1A3 and CD24 were used to decrease the signal of the co-expressed fluorescent proteins. As demonstrated in Figure 2.2, 25 nM of siRNA was sufficient to reduce mCherry and EGFP signal, respectively.

For negative controls, the focus was finding siRNA strand or strands that altered the cell populations the least at the chosen dose. After several experiments, Cell Signaling's non-targeting siRNA strands produced the least off-target induced cell death (Figure 2.3). Thus, ALDH1A3, CD24, and non-targeting siRNA would be suitable for future siRNA screening campaigns.

Indeed, the ultimate goal is to have as many events as possible while altering sub-populations. In terms of cell plating, the most practical approach to maximizing events would be to ensure close to 100% coverage after 72 hours of treatment (22). However, this approach requires that cells do not change morphology, gene expression, or epigenetic regulation as plate coverage increases. Unfortunately, SUM149 and its derivative cell lines undergo subpopulation changes once coverage exceeds 80%. These cells become more mesenchymal and less epithelial, which would introduce artifacts into the final data. Thus, the right cell seeding density must be selected to both maximize events without changing sub-population percentages as an artifact of the experiment. 3300 cells/well is that cell number for a 96 hour experiment (Figure 2.4).

Additionally, proper cell culture seeding and cultivation of the cells was necessary to develop as well. Seeding 500k of reporter cells in a T-75 flask would produce a suitable number of cells after four days (Figure 2.5, Panel A). It was also discovered that harvesting cells too soon after seeding would lead to an inaccurate measuring of EGFP signal, likely due to the cells insufficient expression of CD24 and thus insufficient amounts of co-expressed EGFP (Figure 2.5, Panel B). Both of these considerations were necessary for future experiments with reporter cells.

There was also the concern of siRNA transfection design, and whether that should be forward or reverse siRNA transfection. Forward transfection entails that both the cells and the siRNA are delivered into the well on the same day; however, reverse transfection requires the cells to have

24 hours to settle before media is changed to antibiotic-antimycotic free media and siRNA is delivered. However, due to procedural reasons for siRNA doses being delivered to the 96 well plates, forward transfection was selected for screening purposes. Mainly, it would be too difficult to seed the wells, deliver the seeded wells to the CCG prior to siRNA dosing, and return those seeded wells to Sun lab facilities. Instead, the siRNA could be plated and collected with far less timing coordination between the CCG and Sun lab members. Additionally, forward infection has the added benefit of increasing transfection efficiency, albeit at the cost of additional, unknown levels of cell death, but this is inescapable due to assay demands (23).

For any high throughput assay, a Z' score must be determined for the assay readout, namely to determine the ability to distinguish between positive and negative hits for the assay (24). This analysis ensures both positive and negative controls are suitable for their roles from a statistical perspective by analyzing the difference in averages from the positive and negative controls (19). The Z' score for the EGFP channel was 0.4 and the score for the mCherry channel was -0.1 (Figure 2.6, Panels A and B). For any high throughput assay, a suitable Z' score is between 0.5 and 1, but this goal is difficult to achieve in both live-cell assays as well as siRNA screening campaigns and are generally not expected for assay approval. However, these Z' scores demonstrate that the assay can determine positive hits for both $EGFP^+$ and EGFP- shifts, but discovering hits for $mCherry^+$ $EGFP^+$ shifts will be far more difficult and will require more additional testing.

Execution of high-throughput siRNA screen

Thus, after these assay optimization efforts, the general process for setting up and analyzing these plates are as followed: 96 well plates provided by the Sun lab would be dosed with sufficient pooled Dharmacon siGENOME siRNA to ensure 20 nM of siRNA per well. To each well was added 3300 cells. They were shaken then placed in the cell incubator at 35 C and 5% CO_2 for 96 hours. After that time had elapsed, cells analyzed by FACS. This process allowed for efficient measurements of numerous plates in a single day without a decline in cell numbers or cell viability. In executing this high-throughput screen, thousands of FCS files were produced, far more than could be analyzed by traditional well-by-well analysis. The data volume necessitated an automated method for developing data sets for hit determination. Thus, I used FCS Express to automate plate analysis. To minimize day-to-day and plate-to-plate variations, mCherry was used to set the gate

for that channel (set to 3% of mCherry$^+$ cells) while the gate for EGFP was set so that the double-positive population (mCherry$^+$ EGFP$^+$) was 6-8% of the overall number of live cell events. While the mCherry gating remained the same plate-to-plate, the double positive population was adjusted each data analysis to limit decreases in precision and ensure better quality data. FCS Express (Version 7) was used to generate Excel files that contained the well designation, the number of live cell events, and the percentage population for each subpopulation (mCherry$^+$ EGFP$^-$, mCherry$^+$ EGFP$^+$, mCherry$^-$ EGFP$^-$, mCherry$^-$ EGFP$^+$, or more colloquially, "red, yellow, grey, and green" due to fluorophore combination.

Analysis of primary screen data

Data analysis was conducted using an in-house R script. The script required files that identified the plate, the siRNA that the wells contained, and the data itself for each fluorescent channel.

The script imported each plate's replicate data, averaged each replicate, designated the control wells, and assigned the siRNA designation to each well. From this point, the data could be listed or graphed depending on additional inputs. Box-plots were generated to check the spread of each replicates and to ensure they remained concordant. For simple data analysis, bar charts were generated that contained average percentages by sub-population. These could further be modified by incorporating the viability data to display if the siRNA increased overall cell growth despite decreasing a certain sub-population. While viability was used to guide some siRNA selections in subsequent screening, this valuable contextualizing was integrated into all future data analysis.

Hit selection for each group ("yellow", "green", "grey", and viability) were chosen if the data deviated from the negative siRNA control by a p-value < 0.05 as generated by a two-tailed student's t-test. In this primary screen, any deviation was initially considered a hit. Of the 505 gene targets screened, 582 were considered hits by one of four outputs. Of those 582, 101 were viability hits (Figure 2.7, Panels A-F), 218 were mCherry$^+$ EGFP$^+$ hits (Figure 2.8, Panels A-R), 238 were mCherry$^-$ EGFP$^-$ hits (Figure 2.9, Panels A-V), and 25 were mCherry$^-$ EGFP$^+$ hits (Figure 2.10, Panel A and B),

However, this approach selected too many hits, and many of these hits proved redundant due to double-counting. For example, CHD2, ostensibly a mCherry$^-$ EGFP$^+$ hit by virtue of it decreasing that sub-population, would also be considered a hit due to it increasing the opposite population (mCherry$^-$ EGFP$^-$) (Figure 2.9, Panels E-F; Figure 2.10, Panels A-B). Furthermore, siRNA could

decrease a sub-population but increase the overall cell numbers, which lead to that sub-population technically increasing. This exclusion considered the total cell count for a treated well, which represents the first panel of Figures 2.7 through 2.10. This additional analysis would inform selections made for conformational secondary siRNA screening.

Of the remaining siRNA, the top 120 siRNA were chosen to be tested in secondary screening. Thirty from each FACS analysis output were selected, each from viability, "yellow", "grey", and "green".

Sub-population hit analysis

Each sub-population hits had distinct members based on prior research. While not all proteins aligned with these classifications, certain sub-populations were enriched for some protein classes. Protein knockdown that affected viability seemed to target proteins that are necessary for mitosis, cell-growth, and cell cycle progression. ACTL6B (actin-like 6B), similar in structure to actin, has roles in nuclear migration and chromatin remodeling due to its role as a sub-unit of the BAF complex (25). ATF2 (activating transcription factor 2) has been implicated in anti-apoptotic signaling, cell growth, DNA damage response, but also in context-dependent oncogenesis (26). Depending on cell-stress level, ATF2 may translocate to the cytosol and encourage apoptosis (27). However, its role in breast cancer is unclear, but these results indicate ATF2 encourages cell survival. This is in contrast to previous reports that ATF2 may sensitize cells to tamoxifen treatment (28). CDK1 and CDK2 are well-known kinases involved in cell-cycle progression, the former controlling G_1/S and G_2/M phase transitions and the latter presumed critical for G_1/S transitions (29, 30). Disrupting either of these kinases likely leads to disrupted cell cycles and apoptosis.

Double positive cell hits ($EGFP^+$ $mCherry^+$, or "yellow" cells) included a large proportion of E3 ubiquitin conjugating enzymes (HUWE1, G2E3, JADE2, RNF8) as well as demethylases (KDM3A, KDM3B, KDM5C, KDM8) (31-38). Interestingly, RNF8 is involved in mammary development by controlling NOTCH1 signaling (35). KDM8 is critical for regulating human embryonic stem cells (38). KDM3A is critical for early embryogenesis and maintenance of early-stem cells (39, 40). These previously performed studies may indicate the validity of these first-round hits. This information combined with the increased presence of chromatin modifying

proteins (HDGFL1, KANSL3, BRWD3), ability to maintain a stem-like state seems to be a matter of balancing protein composition and chromatin stability (41-43).

Different complexes affected the mCherry$^-$ EGFP$^+$ cells in both positive and negative fashion. Many ubiquitin ligases were found among the hits (RNF20, TRIM24, TRIM28, UBE2U, UBR2) (44-49). RNF20 also has roles in double-strand break repair, lineage restriction, and chromatin stabilization (50-53). TRIM24 possesses diverse roles in eukaryotes, including transcriptional co-regulation, heterochromatin organization, metabolism, and even playing a role as an oncoprotein by negatively regulating p53 (54-58). Importantly, TRIM24 overexpression indicates poor prognosis in breast cancer (57). UBE2U is an E2 ubiqutin conjugating enzyme that remains largely uncharacterized (47). UBR2, among other UBR-class ubiquitinases, are primarily involved in degrading mRNA transcripts (59).

TRIM28 is an E3 ubiquitinase and well-researched in many cancers and implicated in numerous pathways, namely p53 degradation, transcriptional co-repression, DNA damage response, EMT, and pluripotency maintenance (45). TRIM28 is implicated in being a key regulator of EMT in fibroblasts (60). Further, it also stabilizes TWIST1, a key regulator in EMT, in breast cancer (61). Since it has already been shown that EMT produces cells with cancer-initiating ability, it is unsurprising that TRIM28 also interacts with EZH2 and the chromatin remodeling complex SWI/SNF to promote mammosphere formation (62, 63). Loss of TRIM28 disrupts cancer stem cell maintenance and reduces the tumor population's ability to grow (64). To effect these changes, TRIM28 works with SETDB1, a H3K9 methyltransferase, to suppress differentiation-related genes (65). With these facts in mind, it is little surprise that TRIM28 had an overall increase of the CD24$^+$ population.

In constrast to the other subpopulation, double-negative cells (mCherry$^-$EGFP$^-$) showed little concordance among most hit identities. The most populous class of proteins was the polycomb group (L3MBTL1, PHC3, EZH2, RbBP7, EPC2).

L3MTBL1 is both a polycomb protein as well as a member of the malignant brain tumor (MBT) family of proteins (66, 67). It is involved in protein quality control by directing p53 activity (68). It also can lead to chromatin compaction by binding to H3K20 mono- and dimethylation (69, 70). In this way it can bind two nucleosomes simultaneously (71). Beyond that, there seems to be little investigation into the impacts this heterochromatin has on disease or cancer beyond the knowledge that it is targeted towards E2F regulated genes, implying there may be some role in regulating cell

cycle (72, 73). One study by Hoya-Arias and colleagues discovered differentiation dysfunction and restriction in embryonic stem cells that lacked L3MTBL1 (74).

PHC3 (PCGF) is a polycomb protein that helps define specific members of the polycomb complexes (75, 76). It can form repressive monomeric polymers that are open-ended, and can interact with E2F6, the loss of which can disrupt a cancerous cell from remaining in G_0 (77, 78). While this protein is largely uncharacterized in its function, its main role is to define PRC1 activities (79).

EZH2 is a well-researched histone methyltransferase that has roles in gene expression, cell survival, EMT, metastasis, and drug resistance (80). It is the catalytic component of the PRC2 complex; alongside, EED and SUZ12, it is a critical component for its function (81). The PRC2's role is also critical in cell fate transitions, especially with bivalent H3K4me3 and H3K27me3 sites for lineage decisions (82, 83). In addition to its canonical role in the PRC2 complex, it can also function as a transcriptional activator (84). Overall, due to its numerous roles in critical cancer gene programs, it is of high interest for drug discovery (85).

RbBP7 (RbAp46), along with its homologue RbBP4 (RbAp48), is a ubiquitous WD-40 repeat histone chaperones (86). Its main role is in associating with histones and transcription factors to localize the PRC2 complex to set locations, though it can be found in many chromatin-associated proteins, such as NuRD, PRC1, and SIN3A (87-89). It possesses an H3-binding site towards the top and an electrostatic H4 binding site along the side (90). Thus, knockdown of RbBP7 would be expected to disrupt a variety of processes, though that would be limited by its homologue's replacement in many of the complexes (91, 92).

Little is known about EPC2-specific function beyond being a member of the complex NuA4/TIP60 complex, which functions as an acetyltransferase that promote cell-cycle progression (93, 94). EPC2 interacts with MYC in tumor-promoting gene programs (95). Additionally, EPC2 and its paralogue EPC1 have been found to sustain the oncogenicity of AML with MLL-AF9 fusion proteins (96).

These findings will inform the siRNA to be ordered for secondary screening.

Discussion

Due to a lack of targeted therapies, triple-negative breast cancer (TNBC) remains an intractable disease (2, 6, 9, 10, 97). While recent advances have led to therapies for BRCA1/2 mutants, such as PARP therapies, there is no single target available for TNBC (98-100). There have been various methods undertaken to find such a target, such as genetic, RNA, and scRNA sequencing, but no target has been discovered, most certainly because there is no such target (101-111). This relates to how TNBC is diagnosed.

After a breast lump has been discovered, a biopsy sample is sent to histology where it is tested for the presence of estrogen, progesterone, androgen, and HER2 receptors (1). Once it has been tested, a treatment regimen is assigned based on the size, location, and spread of the cancer. However, if the tumor is negative for all those markers, it is assigned the designation of "triple negative breast cancer". While this is not an intrinsic subtype, triple-negative aligns nearly 80% with the basal subtype as determined by gene expression sequencing (97). Instead of considering it one monolithic disease, one should consider it a cluster of diseases, each with certain patterns of expression and behavior.

This fact is reflected in the past work performed in the past decade as many researchers attempted to ameliorate the ignorance that surrounds late-stage TNBC's natural intratumoral heterogeneity and potentially, finding a common protein with which to target. Most of these studies used siRNA or shRNA interference to identify new gene targets deserving of future investigation. Granit and colleagues conducted a study that involved knocking down 177 candidate genes with shRNA, followed by staining with CD14 and CD18, which are markers for luminal progenitors (18). They discovered that FOXA1 decreased symmetrical division decreased these progenitor population from symmetrically dividing, while NOTCH performed the opposite function (18). Another group performed siRNA screening in multiple TNBC cell lines and found that polo-like kinase 1 (PLK1) decreased the CD44+/CD24- population in SUM149 in contrast to first line therapies like doxorubicin (112). Petrocca and colleagues discovered through siRNA screening that the tested TNBC cell lines were dependent on proteasome activation, and that this addiction was directed by NOXA and MCL-1 (113). Singel and colleagues found that KIF14 and TLN1 sensitized multiple TNBC cell lines to docetaxel treatment (114).

Altogether, these efforts point to two main conclusions: 1) each identified weakness is likely true in some cases, but 2) these weaknesses are not uniform, owing to the well-established inter-tumoral heterogeneity present among the whole population of TNBC patients (115, 116). In all likelihood,

each of these studies will likely fit into TNBC research and treatment like a mosaic, each having their own role in different groups of triple-negative breast cancers. In that sense, there will never be one single treatment for all TNBCs, but instead a suite of approaches for a given patient's individual disease determined by RNAseq, and in the future scRNAseq.

While most early stage and localized cancers can be treated effectively with some combination of surgery, radiation, and chemotherapy, late stage TNBC is a death sentence with an 11% median relative survival rate (11). This metastatic stage of the disease can only be slowed and cannot be stopped (2, 111). This reality is likely due to the intratumoral heterogeneity presence within this triple-negative breast cancer (116, 117). This fact has led to recent efforts to discern the true nature of this intratumoral heterogeneity, but no effort has attempted to discern the origin of this heterogeneity.

To discover this origin, I conducted a high-throughput screen. Traditional HTS targets are most often proteins of some nature, whether GPCRs, kinases, phosphatases, or other enzymes (118-124). Most assay outputs are activity based with some manner of radioactive, fluorescent, or bioluminescent, among others (20). Fluorescence polarization, BRET, and FRET are examples (125). The goal of any assay is to disrupt (or increase) some signal based on binding to a specific location on the target of interest. For instance, Fluorescence polarization (FP) assays involve several components: protein or protein fragment of interest, a probe or peptide that binds to a specific region, that same probe with a fluorescent linker, and a library of compounds or other such entities. (125)

While there are other assay designs that are advantageous in different ways and for different goals, this cell-based fluorescent reporter system will be similar to other fluorescent-based assay designs, except there are four fluorescent outputs that must be measured and compared to the effect of a negative control siRNA. Additionally, this means that extensive optimization must be conducted, more so than other live-cell HTS campaigns.

In most HTS campaigns, there are several factors that must be optimized to ensure successful miniaturization: 1) homogenized and high quality materials 2) Sufficient signal 3) selection of positive and negative controls 4) buffer system (19).

For any HTS campaign, variation plate to plate must be minimized (126). That means for whatever component, protein, probe, buffer, etc., that component should be from the same batch and be sufficient in quantity for primary, secondary, and follow-up tests. For instance, if protein is

expressed and collected from an in-house bacterial, insect, or mammalian system, that protein should be expressed, extracted, and purified from the same batch. Otherwise, unexplained and non-systematic variations will be introduced into the testing system and reducing statistical power and likelihood of correctly identifying a potential hit.

In addition to concerns of the test components, the assay must be optimized such that sufficient signal can be produced on a per-well basis. That means that sufficient amount of protein, signal output, and tested compounds must be provided into each well without requiring too much material. Otherwise, the cost can increase exponentially, especially for larger campaigns (126). This step requires extensive optimization and is a critical factor in determining overall cost and data quality. It can be thought of as an early make-or-break point for an HTS campaign. Either the assay design must be altered or the campaign canceled.

Additionally, the positive and negative controls must be thoughtfully selected (21). Both controls define the assay window and are critical in determining hits from non-hits. The positive control defines a maximal effect, which can either be met or exceeded depending on assay type and design. The negative control often defines a non-effect and serves as a base level of non-activity for inactive hits. These tests ultimately determine once more if an assay is suitable for an HTS campaign.

Buffer selection and any additives that are necessary to ensure optimal assay function (21). Such buffers could be PBS, TRIS, or HEPES, and an alteration of pH or salinity may be necessary to increase assay signal. These buffers, besides optimizing protein structure, must also be able to tolerate relatively hydrophobic compounds that will be tested. Additionally, since many HTS assay designs rely on some form of light or emission, precipitation can be an ongoing concern.

Ultimately, each of these optimization steps must be performed in parallel and different results may necessitate changes in another area. The final test for any HTS assay is whether it can successfully produce a Z' score between 0.5 and 0.9. The Z' calculation compares the positive controls with the negative control results, determining if there is a sufficiently large assay window (24). Any lower than 0.5 indicates an assay without a sufficiently large assay window, and any larger than 0.9 to 1 indicates some other issue may be producing erroneous data.

Once each of these features have been optimized sufficiently, the campaign is ready to begin. Ongoing analysis can be routinely developed by measuring the signal-to-noise on a plate by plate

timing. Z' scores as well can be calculated for each plate, ensuring that the quality of data does not decrease relative to previous optimization studies.

However, these specific features are mostly meaningful for compound screening versus a given protein target; HTS campaigns using live cells and using siRNA, shRNA, and CRISPR technologies require significantly more optimization than compound screenings yet produce less powerful statistical significance, not to mention the additional expenses and inputs required to conduct live-cell campaigns. This present campaign uses a SUM149-derived cell line required far more optimization than a regular campaign.

While models of high-throughput screening have been proposed for this heterogeneity in the past, to my knowledge this was the first effort to understand how this heterogeneity arises and which proteins may be critical to maintaining those populations (127). Specifically, I chose to investigate epigenetic proteins. While further genetic alterations will ultimately occur, in the short term it is speculated that epigenetic adaptations allow for the cancer to endure chemotherapy, to alter gene expression to survive, and to ultimately metastasize to other regions in the body (116, 128-130). While the specific and initial genetic mutations may differ, there may be common epigenetic features that can be discovered and exploited in a future therapeutic.

To that end, a high-throughput screen using live cells and siRNA was successfully optimized to screen 505 gene targets. Due to the added complexity of these types of experiments, great care needed to be taken with each experimental variable. I optimized the type of transfection agent, the concentration of that agent, the mode of transfection, the number of cells seeded, the process by which they can be prepared for FACS analysis, and the final analysis (Figures 2.1-2.6). The final analysis required additional effort, as four outputs were considered: viability, "yellow", "grey", and "green" (Figures 2.7-2.10). While such optimization efforts may seem trivial compared to the final dataset, they are necessary for any successful screening to occur.

Various trends emerged after hit determination concluded. In the final data set, any hit must significantly decrease a selected sub-population at the $p < 0.05$ level. The majority of hits were classified as mesenchymal ("grey") hits, where there was some alteration in the grey population (Figure 2.9). In most cases, that was due to the epithelial ("green") cells decreasing in overall proportion. This makes sense as the epithelial cells comprise greater than 70% of the overall population, and thus the overall assay window was large enough to accommodate a great number of hits. Conversely, the number of "green" hits was not great in number due to the low levels of

mesenchymal cells in well-grown SUM149-derived cells (Figure 2.10). Even if there is a reduction in cell number, the effect is small, perhaps decreasing from 20% to 12%, which may not register at the $p > 0.05$ level. This same effect however did not occur in the epithelial progenitor population ("yellow"), which only comprise approximately 6% of the overall population (Figure 2.8). This population nearly had as many hits as the "grey" subpopulation. Proteins such as TRIM24, managed to reliably reduce this population, but a greater number of proteins increased this population such as KMT2D and EPC2. It remains to be seen whether these proteins can reliably decrease or increase these populations in other TNBC cell lines.

EGFP and mCherry assay window size was initially hypothesized to be the controlling factor for hit occurence, but that did not bear out in the primary screen's final dataset despite the Z' score of both ALDH1A3 and CD24 positive control siRNA (Figure 2.6). Another possibility is the difficulty in reducing an already small sub-population even further. There is some combination genes expressed that lead to the sub-population proportions that emerge, epithelial cells being the most predominant. However, that same geneset ensures that a smaller proportion of mesenchymal and stem-like cells are produced as well, and shifting these cells into non-existence is difficult for a single epigenetic protein. There may be some upper limit of siRNA effect on a given population in this cell line. Perhaps only a well-characterized and validated chemical probe or drug can surpass that limit.

Compared to $mCherry^{-}$ $EGFP^{-}$ hits, there were relatively fewer viability hits. This is likely due to epigenetic targets having limited impact on the fundamental functions of living cells. While knockdown of a certain target will have an impact on a certain function, be that DNA methylation or histone acetylation at a given lysine position, the cell itself must be dependent on that mark or series of marks for population-wide death to occur. This perhaps could be why few epigenetic drugs have been used in solid cancers. However, non-solid cancers like leukemia have had better success with research (131). This was the first cancer type that the cancer stem cell hierarchy was demonstrated, and it perhaps may be no surprise that epigenetic therapies are more likely to succeed than solid tumors (132). In AML leukemia, most patients experience some initial response to either chemotherapy or hemopoetic cell transplantation, relapse is still frequent, and second and third line therapies are in demand (131). While genetic mutations in AML are relatively rare, mutation rates are much higher in epigenetic processes that regulate DNA methylation. This fact drew industry and academic attention, and as a consequence the FDA has approved therapies for

two processes: DNA methylation (Azacitidine and Decitabine) and IDH1/2 mutant proteins, which regulate DNA methylation (Ivosidenib and Enasidenib) (131). Many other drug programs against other targets are in development, such as against TET, MLL, and SETD2, all of which have been demonstrated to play some role in AML development (131).

If knockdown of epigenetic proteins had any effect on viability, it is highly likely that this effect is due to disrupted non-epigenetic processes. For instance, EZH2 is the active catalytic member of the H3K27 trimethylase complex PRC2. When EZH2 is directly inhibited by GSK-353, viability is greatly reduced; however, if one inactivates the PRC2 complex allosterically by targeting EED's EZH2-activation (A-395), H3K27me3 is ablated but no significant change in viability occurs. This implies that PRC2 activity is unrelated to the EZH2-related drop in viability. This effect is possibly due to EZH2's role in transcriptional co-activation in conjunction with the JAK-STAT pathway rather than its canonical enzymatic activity (133). Ultimately, this work's present concern is not with the elimination of all cancer sub-populations but the regulations of those groups.

Overall, this primary screen demonstrated the viability of using fluorescent markers co-expressed with previously discovered identifiers for different sub-populations. By taking the necessary time to optimize each condition and the data analysis pipeline, significant changes could be easily distinguished from non-significant changes. This approach could be applied to any cancer type that is marked as containing multitudes of cell types, regardless of the means by which those cell types arise.

At this stage in the campaign, it is tempting to imagine the effects of certain hits, especially those detailed here, but such fantasies are premature, especially considering only the primary screen has been conducted. Secondary screening, featuring single siRNA strands per well, followed by testing in other TNBC cell lines, will be needed to further validate each of these gene target hits.

Materials and Methods

Cell Line Maintenance

SUM149 and its derivative reporter cell lines were cultured in a 37 C, 95% humidity, and 5% CO_2 environment. They were cultured in Ham's F-12 medium, (Invitrogen, Carlsbad, CA) supplemented with 5% fetal bovine serum (Fisher Scientific, Pittsburgh, PA), 1% antibiotic-

antimycotic (Invitrogen, Carlsbad, CA), 5 μg/ml insulin (Sigma-Aldrich, St Louis, MO), 1 μg/ml hydrocortisone (Sigma-Aldrich, St Louis, MO), and 4 μg/ml gentamicin (Invitrogen, Carlsbad, CA). Antibiotic/antimycotic free-media was produced by excluding both the standard antibiotic-antimycotic and gentamicin. FACS-specific media for high-throughput screening used the same but lowered supplement of fetal bovine serum to 2%.

RNAimax Titration Experiment

SUM149 cells were seeded in a 96 well plate at 2000 cells/well. After 24 hours, the media was replaced with antibiotic free Ham F-12 media (Invitrogen, Carlsbad, CA) containing a range of Lipofectamine RNAimax Transfection Reagent (ThermoFischer Scientific, Carlsbad, CA) of 1.0 to 0.1 uL in 0.1 uL incriments. After 72 hours, the MTS assay (Promega, Madison, WI) was conducted according to manufacturer's instructions. Fluorescent data was collected with a Cytation 5 plate reader (BioTek, Winooski, VT). Data represents the mean +/- standard deviation of the three samples following data normalization.

Positive Control Selection

500k SUM149 cells were seeded in triplicate 10 cm dishes and allowed to settle. 24 hours later, the media was changed to antibiotic free Ham F-12 media (Invitrogen, Carlsbad, CA). Each dish received Vehicle, non-targeting siRNA (Dharmacon, Lafayette, CO), ALDH1A3 siRNA (Dharmacon, Lafayette, CO), or CD24 siRNA (Dharmacon, Lafayette, CO) diluted in Opti-Mem™ (Fisher Scientific, Waltham, MA) and Lipofectamine RNAimax Transfection Reagent (ThermoFischer Scientific, Carlsbad, CA). Final concentration for each siRNA was 25 nM. After 72 hours elapsed, the cells were analyzed on a Zeti 5 flow cytometer (Bio-Rad, Hercules, CA). Data represent mean signal for either EGFP or mCherry +/- standard deviation.

Negative Control Selection

500k SUM149 cells were seeded in triplicate 10 cm dishes and allowed to settle. 24 hours later, the media was changed to antibiotic free Ham F-12 media (Invitrogen, Carlsbad, CA). Each dish received Vehicle, ON-TARGETplus Non-targeting Control Pool (Dharmacon, Lafayette, CO), or

SignalSilence® Control siRNA Unconjugated (Cell Signalling, Danvers, MA) diluted in Opti-Mem™ (Fisher Scientific, Waltham, MA) and Lipofectamine RNAimax Transfection Reagent (ThermoFischer Scientific, Carlsbad, CA). Final concentration for each siRNA was 80 nM. After 72 hours elapsed, the cells were analyzed on a Zeti 5 flow cytometer (Bio-Rad, Hercules, CA). Data represent mean signal for number of DAPI-negative cells +/- standard deviation.

Cell Density Titration and Cell Handling Optimization

In a clear-bottomed 96-well plate (Corning, Corning, NY), a range of SUM149 cells were added, ranging from 4500 cells to 1975 in increments of 225 cells/well. Each triplicate well was dosed with 0.6 uL of Lipofectamine RNAimax Transfection Reagent (ThermoFischer Scientific, Carlsbad, CA) in Opti-Mem™ (Fisher Scientific, Waltham, MA) and antibiotic-free Ham F12 media (Invitrogen, Carlsbad, CA). Cells were incubated for 96 total hours. After that elapsed time, the media was aspirated and 50 uL of Trypsin-EDTA (0.25%) (ThermoFisher Scientific, Carlsbad, CA). Cells were returned to incubator for 10 minutes, after which each well was mixed by a multichannel pippetteman fifty total times to further displace the cells. The cells were allowed to incubate for several more minutes. Trypsin was neutralized with FluoroBrite DMEM Media (ThermoFisher Scientific, Carlsbad, CA) containing 2% fetal bovine serum (Fisher Scientific, Pittsburgh, PA), 5 μg/ml insulin (Sigma-Aldrich, St. Louis, MO), 1 μg/ml hydrocortisone (Sigma-Aldrich, St Louis, MO), and 1 ug/mL DAPI (Sigma-Aldrich, St. Louis, MO). Cells were pipetted again to ensure a single cell suspension. Plate was analyzed on a Zeti 5 flow cytometer (Bio-Rad, Hercules, CA). Data are the mean DAPI-negative cells gathered in 50 uL of solution +/- standard deviation.

Sub-Population Analysis Over a 72 hour Time Course

500k SUM149 reporter cells were seeded in duplicate in 10 cm2 plates (Corning, Corning, NY) over a three day period, where 48, 72, and 96 hours would pass before FACS analysis. Cells seeded came from the same passage of reporter cells. After a total of 96 hours had passed, cells were detached using Trypsin-EDTA (0.25%) (ThermoFisher Scientific, Carlsbad, CA) and neutralized with complete Ham's F12 media. Cells were counted and resuspended in HBSS (Invitrogen, Carlsbad, CA) with 2% fetal bovine serum (Fisher Scientific, Pittsburgh, PA) with 1 ug/mL DAPI

(Sigma-Aldrich, St. Louis, MO). Cells were analyzed on a Zeti 5 flow cytometer (Bio-Rad, Hercules, CA). Data are means of each of the four sub-populations, arranged by percentage of total cell population.

Z-prime Score Determination

Five different passages of reporter cells were seeded at 3300 cells/well in a clear-bottomed 96-well plate (Corning, Corning, NY), each with four wells per group for a total of twenty sets of four. Six wells were seeded as negative controls. Cells were dosed with ALDH1A3 siRNA (Dharmacon, Lafayette, CO), CD24 siRNA (Dharmacon, Lafayette, CO) or SignalSilence® Control siRNA Unconjugated (Cell Signalling, Danvers, MA) diluted in Opti-Mem™ (Fisher Scientific, Waltham, MA) and Lipofectamine RNAimax Transfection Reagent (ThermoFischer Scientific, Carlsbad, CA). Final dose of all siRNA was 20 nM.

After 96 hours, the media was aspirated and 50 uL of Trypsin-EDTA (0.25%) (ThermoFisher Scientific, Carlsbad, CA). Cells were returned to incubator for 10 minutes, after which each well was mixed by a multichannel pippetteman fifty total times to further displace the cells. The cells were allowed to incubate for several more minutes. Trypsin was neutralized with FluoroBrite DMEM Media (ThermoFisher Scientific, Carlsbad, CA) containing 2% fetal bovine serum (Fisher Scientific, Pittsburgh, PA), 5 μg/ml insulin (Sigma-Aldrich, St. Louis, MO), 1 μg/ml hydrocortisone (Sigma-Aldrich, St Louis, MO), and 1 ug/mL DAPI (Sigma-Aldrich, St. Louis, MO). Cells were pipetted again to ensure a single cell suspension. Plate was analyzed on a Zeti 5 flow cytometer (Bio-Rad, Hercules, CA). EGFP and mCherry signal were analyzed using the Z' equation: $1 - (3SD^{+} + 3SD^{-})/(|Ave^{+} - Ave^{-}|)$

High-Throughput siRNA Screening

3300 cells/well were seeded in clear-bottomed 96 well plates (Corning, Corning, NY) prepared by the Center for Chemical Genomics with sufficient siRNA (siGENOME by Dharmacon, Lafayette, CO) for 20 nM final. Contained within the matrix was antibiotic free Ham's F-12 medium, (Invitrogen, Carlsbad, CA) supplemented with 5% fetal bovine serum (Fisher Scientific, Pittsburgh, PA), 5 μg/ml insulin (Sigma-Aldrich, St Louis, MO), and 1 μg/ml hydrocortisone (Sigma-Aldrich, St Louis, MO), 0.6 uL of Lipofectamine RNAimax Transfection Reagent

(ThermoFischer Scientific, Carlsbad, CA), and Opti-Mem™ (Fisher Scientific, Waltham, MA). Cells were incubated for 96 hours. After that elapsed time, the media was aspirated and 50 uL of Trypsin-EDTA (0.25%) (ThermoFisher Scientific, Carlsbad, CA). Cells were returned to incubator for 10 minutes, after which each well was mixed by a multichannel pippetteman fifty total times to further displace the cells. The cells were incubated for several more minutes. Trypsin was neutralized with FluoroBrite DMEM Media (ThermoFisher Scientific, Carlsbad, CA) containing 2% fetal bovine serum (Fisher Scientific, Pittsburgh, PA), 5 μg/ml insulin (Sigma-Aldrich, St. Louis, MO), 1 μg/ml hydrocortisone (Sigma-Aldrich, St Louis, MO), and 1 ug/mL DAPI (Sigma-Aldrich, St. Louis, MO). Cells were pipetted again to ensure a single cell suspension. Plate was analyzed on a Zeti 5 flow cytometer (Bio-Rad, Hercules, CA). Data are the mean of each gene target +/- standard deviation.

References

1. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001;98(19):10869-74.

2. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res. 2007;13(8):2329-34.

3. Wang D-Y, Jiang Z, Ben-David Y, Woodgett JR, Zacksenhaus E. Molecular stratification within triple-negative breast cancer subtypes. Sci Rep. 2019;9(1):19107.

4. Bareche Y, Venet D, Ignatiadis M, Aftimos P, Piccart M, Rothe F, et al. Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. Ann Oncol. 2018;29(4):895-902.

5. Beça FFd, Caetano P, Gerhard R, Alvarenga CA, Gomes M, Paredes J, et al. Cancer stem cells markers CD44, CD24 and ALDH1 in breast cancer special histological types. Journal of Clinical Pathology. 2013;66(3):187.

6. Honeth G, Bendahl P-O, Ringnér M, Saal LH, Gruvberger-Saal SK, Lövgren K, et al. The CD44+/CD24- phenotype is enriched in basal-like breast tumors. Breast Cancer Res. 2008;10(3):R53-R.

7. Grimshaw MJ, Cooper L, Papazisis K, Coleman JA, Bohnenkamp HR, Chiapero-Stanke L, et al. Mammosphere culture of metastatic breast cancer cells enriches for tumorigenic breast cancer cells. Breast Cancer Res. 2008;10(3):R52.

8. Fillmore CM, Kuperwasser C. Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. Breast Cancer Res. 2008;10(2):R25.

9. Li X, Lewis MT, Huang J, Gutierrez C, Osborne CK, Wu MF, et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. Journal of the National Cancer Institute. 2008;100(9):672-9.

10. Creighton CJ, Li X, Landis M, Dixon JM, Neumeister VM, Sjolund A, et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proc Natl Acad Sci U S A. 2009;106(33):13820-5.

11. National Cancer Institute D, Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov). SEER*Stat Database: Populations - Total U.S. (1969-2018) Breast Cancer Positive - Total U.S., 1969-2018 Counties. Surveillance Research Program: National Cancer Institute; December 2019.

12. Mani SA, Guo W, Liao M-J, Eaton EN, Ayyanan A, Zhou AY, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell. 2008;133(4):704-15.

13. Chaffer CL, Marjanovic ND, Lee T, Bell G, Kleer CG, Reinhardt F, et al. Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. Cell. 2013;154(1):61-74.

14. Liu S, Cong Y, Wang D, Sun Y, Deng L, Liu Y, et al. Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. Stem Cell Reports. 2013;2(1):78-91.

15. Harrison H, Farnie G, Howell SJ, Rock RE, Stylianou S, Brennan KR, et al. Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor. Cancer Res. 2010;70(2):709-18.

16. Liu S, Dontu G, Mantle ID, Patel S, Ahn NS, Jackson KW, et al. Hedgehog signaling and Bmi-1 regulate self-renewal of normal and malignant human mammary stem cells. Cancer Res. 2006;66(12):6063-71.

17. King TD, Suto MJ, Li Y. The Wnt/β-catenin signaling pathway: a potential therapeutic target in the treatment of triple negative breast cancer. Journal of cellular biochemistry. 2012;113(1):13-8.

18. Granit RZ, Masury H, Condiotti R, Fixler Y, Gabai Y, Glikman T, et al. Regulation of Cellular Heterogeneity and Rates of Symmetric and Asymmetric Divisions in Triple-Negative Breast Cancer. Cell Rep. 2018;24(12):3237-50.

19. Pereira DA, Williams JA. Origin and evolution of high throughput screening. Br J Pharmacol. 2007;152(1):53-61.

20. Archer JR. History, evolution, and trends in compound management for high throughput screening. Assay and drug development technologies. 2004;2(6):675-81.

21. Echeverri CJ, Perrimon N. High-throughput RNAi screening in cultured cells: a user's guide. Nature reviews Genetics. 2006;7(5):373-84.

22. Echeverri CJ, Beachy PA, Baum B, Boutros M, Buchholz F, Chanda SK, et al. Minimizing the risk of reporting false positives in large-scale RNAi screens. Nat Methods. 2006;3(10):777-9.

23. Parsons BD, Schindler A, Evans DH, Foley E. A direct phenotypic comparison of siRNA pools and multiple individual duplexes in a functional assay. PLoS One. 2009;4(12):e8471-e.

24. Zhang XD. Illustration of SSMD, z score, SSMD*, z* score, and t statistic for hit selection in RNAi high-throughput screens. Journal of biomolecular screening. 2011;16(7):775-85.

25. Olave I, Wang W, Xue Y, Kuo A, Crabtree GR. Identification of a polymorphic, neuron-specific chromatin remodeling complex. Genes Dev. 2002;16(19):2509-17.

26. Bhoumik A, Ronai Z. ATF2: a transcription factor that elicits oncogenic or tumor suppressor activities. Cell cycle (Georgetown, Tex). 2008;7(15):2341-5.

27. Lau E, Ronai ZA. ATF2 - at the crossroad of nuclear and cytosolic functions. J Cell Sci. 2012;125(Pt 12):2815-24.

28. Rudraraju B, Droog M, Abdel-Fatah TM, Zwart W, Giannoudis A, Malki MI, et al. Phosphorylation of activating transcription factor-2 (ATF-2) within the activation domain is a key determinant of sensitivity to tamoxifen in breast cancer. Breast Cancer Res Treat. 2014;147(2):295-309.

29. Sakurikar N, Eastman A. Critical reanalysis of the methods that discriminate the activity of CDK2 from CDK1. Cell cycle (Georgetown, Tex). 2016;15(9):1184-8.

30. Santamaría D, Barrière C, Cerqueira A, Hunt S, Tardy C, Newton K, et al. Cdk1 is sufficient to drive the mammalian cell cycle. Nature. 2007;448(7155):811-5.

31. Kao S-H, Wu H-T, Wu K-J. Ubiquitination by HUWE1 in tumorigenesis and beyond. J Biomed Sci. 2018;25(1):67-.

32. Brooks WS, Helton ES, Banerjee S, Venable M, Johnson L, Schoeb TR, et al. G2E3 is a dual function ubiquitin ligase required for early embryonic development. J Biol Chem. 2008;283(32):22304-15.

33. Liu YX, Zhang SF, Ji YH, Guo SJ, Wang GF, Zhang GW. Whole-exome sequencing identifies mutated PCK2 and HUWE1 associated with carcinoma cell proliferation in a hepatocellular carcinoma patient. Oncol Lett. 2012;4(4):847-51.

34. Huen MSY, Grant R, Manke I, Minn K, Yu X, Yaffe MB, et al. RNF8 transduces the DNA-damage signal via histone ubiquitylation and checkpoint protein assembly. Cell. 2007;131(5):901-14.

35. Li L, Guturi KKN, Gautreau B, Patel PS, Saad A, Morii M, et al. Ubiquitin ligase RNF8 suppresses Notch signaling to regulate mammary development and tumorigenesis. The Journal of clinical investigation. 2018;128(10):4525-42.

36. Sui Y, Gu R, Janknecht R. Crucial Functions of the JMJD1/KDM3 Epigenetic Regulators in Cancer. Molecular cancer research : MCR. 2020.

37. Brauchle M, Yao Z, Arora R, Thigale S, Clay I, Inverardi B, et al. Protein complex interactor analysis and differential activity of KDM3 subfamily members towards H3K9 methylation. PLoS One. 2013;8(4):e60549.

38. Zhu H, Hu S, Baker J. JMJD5 regulates cell cycle and pluripotency in human embryonic stem cells. Stem Cells. 2014;32(8):2098-110.

39. Kuroki S, Nakai Y, Maeda R, Okashita N, Akiyoshi M, Yamaguchi Y, et al. Combined Loss of JMJD1A and JMJD1B Reveals Critical Roles for H3K9 Demethylation in the Maintenance of Embryonic Stem Cells and Early Embryogenesis. Stem Cell Reports. 2018;10(4):1340-54.

40. Ramadoss S, Guo G, Wang CY. Lysine demethylase KDM3A regulates breast cancer cell invasion and apoptosis by targeting histone and the non-histone protein p53. Oncogene. 2017;36(1):47-59.

41. Meunier S, Shvedunova M, Van Nguyen N, Avila L, Vernos I, Akhtar A. An epigenetic regulator emerges as microtubule minus-end binding and stabilizing factor in mitosis. Nat Commun. 2015;6:7889.

42. Müller P, Kuttenkeuler D, Gesellchen V, Zeidler MP, Boutros M. Identification of JAK/STAT signalling components by genome-wide RNA interference. Nature. 2005;436(7052):871-5.

43. Li J, Lange LA, Duan Q, Lu Y, Singleton AB, Zonderman AB, et al. Genome-wide admixture and association study of serum iron, ferritin, transferrin saturation and total iron binding capacity in African Americans. Hum Mol Genet. 2015;24(2):572-81.

44. Fuchs G, Oren M. Writing and reading H2B monoubiquitylation. Biochim Biophys Acta. 2014;1839(8):694-701.

45. Czerwińska P, Mazurek S, Wiznerowicz M. The complexity of TRIM28 contribution to cancer. J Biomed Sci. 2017;24(1):63-.

46. Appikonda S, Thakkar KN, Barton MC. Regulation of gene expression in human cancers by TRIM24. Drug Discov Today Technol. 2016;19:57-63.

47. Sheng Y, Hong JH, Doherty R, Srikumar T, Shloush J, Avvakumov GV, et al. A human ubiquitin conjugating enzyme (E2)-HECT E3 ligase structure-function screen. Mol Cell Proteomics. 2012;11(8):329-41.

48. Tasaki T, Zakrzewska A, Dudgeon DD, Jiang Y, Lazo JS, Kwon YT. The substrate recognition domains of the N-end rule pathway. J Biol Chem. 2009;284(3):1884-95.

49. Kwon YT, Reiss Y, Fried VA, Hershko A, Yoon JK, Gonda DK, et al. The mouse and human genes encoding the recognition component of the N-end rule pathway. Proc Natl Acad Sci U S A. 1998;95(14):7898-903.

50. Tarcic O, Granit RZ, Pateras IS, Masury H, Maly B, Zwang Y, et al. RNF20 and histone H2B ubiquitylation exert opposing effects in Basal-Like versus luminal breast cancer. Cell Death Differ. 2017;24(4):694-704.

51. Foglizzo M, Middleton AJ, Day CL. Structure and Function of the RING Domains of RNF20 and RNF40, Dimeric E3 Ligases that Monoubiquitylate Histone H2B. J Mol Biol. 2016;428(20):4073-86.

52. Fuchs G, Shema E, Vesterman R, Kotler E, Wolchinsky Z, Wilder S, et al. RNF20 and USP44 regulate stem cell differentiation by modulating H2B monoubiquitylation. Mol Cell. 2012;46(5):662-73.

53. Moyal L, Lerenthal Y, Gana-Weisz M, Mass G, So S, Wang SY, et al. Requirement of ATM-dependent monoubiquitylation of histone H2B for timely repair of DNA double-strand breaks. Mol Cell. 2011;41(5):529-42.

54. Herquel B, Ouararhni K, Davidson I. The TIF1α-related TRIM cofactors couple chromatin modifications to transcriptional regulation, signaling and tumor suppression. Transcription. 2011;2(5):231-6.

55. Le Douarin B, Nielsen AL, Garnier JM, Ichinose H, Jeanmougin F, Losson R, et al. A possible involvement of TIF1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors. EMBO J. 1996;15(23):6701-15.

56. Maison C, Almouzni G. HP1 and the dynamics of heterochromatin maintenance. Nature Reviews Molecular Cell Biology. 2004;5(4):296-305.

57. Chambon M, Orsetti B, Berthe M-L, Bascoul-Mollevi C, Rodriguez C, Duong V, et al. Prognostic significance of TRIM24/TIF-1α gene expression in breast cancer. Am J Pathol. 2011;178(4):1461-9.

58. Pathiraja TN, Thakkar KN, Jiang S, Stratton S, Liu Z, Gagea M, et al. TRIM24 links glucose metabolism with transformation of human mammary epithelial cells. Oncogene. 2015;34(22):2836-45.

59. An JY, Kim E, Zakrzewska A, Yoo YD, Jang JM, Han DH, et al. UBR2 of the N-end rule pathway is required for chromosome stability via histone ubiquitylation in spermatocytes and somatic cells. PLoS One. 2012;7(5):e37414.

60. Venkov CD, Link AJ, Jennings JL, Plieth D, Inoue T, Nagai K, et al. A proximal activator of transcription in epithelial-mesenchymal transition. J Clin Invest. 2007;117(2):482-91.

61. Wei C, Cheng J, Zhou B, Zhu L, Khan MA, He T, et al. Tripartite motif containing 28 (TRIM28) promotes breast cancer metastasis by stabilizing TWIST1 protein. Sci Rep. 2016;6:29822.

62. Li J, Xi Y, Li W, McCarthy RL, Stratton SA, Zou W, et al. TRIM28 interacts with EZH2 and SWI/SNF to activate genes that promote mammosphere formation. Oncogene. 2017;36(21):2991-3001.

63. Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell. 2008;133(4):704-15.

64. Czerwińska P, Shah PK, Tomczak K, Klimczak M, Mazurek S, Sozańska B, et al. TRIM28 multi-domain protein regulates cancer stem cell population in breast tumor development. Oncotarget. 2017;8(1):863-82.

65. Miles DC, de Vries NA, Gisler S, Lieftink C, Akhtar W, Gogola E, et al. TRIM28 is an Epigenetic Barrier to Induced Pluripotent Stem Cell Reprogramming. Stem Cells. 2017;35(1):147-57.

66. Koga H, Matsui S, Hirota T, Takebayashi S, Okumura K, Saya H. A human homolog of Drosophila lethal(3)malignant brain tumor (l(3)mbt) protein associates with condensed mitotic chromosomes. Oncogene. 1999;18(26):3799-809.

67. Boccuni P, MacGrogan D, Scandura JM, Nimer SD. The human L(3)MBT polycomb group protein is a transcriptional repressor and interacts physically and functionally with TEL (ETV6). J Biol Chem. 2003;278(17):15412-20.

68. West LE, Roy S, Lachmi-Weiner K, Hayashi R, Shi X, Appella E, et al. The MBT repeats of L3MBTL1 link SET8-mediated p53 methylation at lysine 382 to target gene repression. J Biol Chem. 2010;285(48):37725-32.

69. Min J, Allali-Hassani A, Nady N, Qi C, Ouyang H, Liu Y, et al. L3MBTL1 recognition of mono- and dimethylated histones. Nat Struct Mol Biol. 2007;14(12):1229-30.

70. Kalakonda N, Fischle W, Boccuni P, Gurvich N, Hoya-Arias R, Zhao X, et al. Histone H4 lysine 20 monomethylation promotes transcriptional repression by L3MBTL1. Oncogene. 2008;27(31):4293-304.

71. Trojer P, Li G, Sims RJ, 3rd, Vaquero A, Kalakonda N, Boccuni P, et al. L3MBTL1, a histone-methylation-dependent chromatin lock. Cell. 2007;129(5):915-28.

72. Trojer P, Reinberg D. Beyond histone methyl-lysine binding: how malignant brain tumor (MBT) protein L3MBTL1 impacts chromatin structure. Cell cycle (Georgetown, Tex). 2008;7(5):578-85.

73. Lewis PW, Beall EL, Fleischer TC, Georlette D, Link AJ, Botchan MR. Identification of a Drosophila Myb-E2F2/RBF transcriptional repressor complex. Genes Dev. 2004;18(23):2929-40.

74. Hoya-Arias R, Tomishima M, Perna F, Voza F, Nimer SD. L3MBTL1 deficiency directs the differentiation of human embryonic stem cells toward trophectoderm. Stem Cells Dev. 2011;20(11):1889-900.

75. Hauri S, Comoglio F, Seimiya M, Gerstung M, Glatter T, Hansen K, et al. A High-Density Map for Navigating the Human Polycomb Complexome. Cell Rep. 2016;17(2):583-95.

76. Cao Q, Wang X, Zhao M, Yang R, Malik R, Qiao Y, et al. The central role of EED in the orchestration of polycomb group complexes. Nat Commun. 2014;5:3127.

77. Deshpande AM, Akunowicz JD, Reveles XT, Patel BB, Saria EA, Gorlick RG, et al. PHC3, a component of the hPRC-H complex, associates with E2F6 during G0 and is lost in osteosarcoma tumors. Oncogene. 2007;26(12):1714-22.

78. Robinson AK, Leal BZ, Nanyes DR, Kaur Y, Ilangovan U, Schirf V, et al. Human polyhomeotic homolog 3 (PHC3) sterile alpha motif (SAM) linker allows open-ended polymerization of PHC3 SAM. Biochemistry. 2012;51(27):5379-86.

79. Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, et al. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol Cell. 2012;45(3):344-56.

80. Gan L, Yang Y, Li Q, Feng Y, Liu T, Guo W. Epigenetic regulation of cancer progression by EZH2: from biological insights to therapeutic potential. Biomark Res. 2018;6:10.

81. Kirmizis A, Bartley SM, Kuzmichev A, Margueron R, Reinberg D, Green R, et al. Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. Genes Dev. 2004;18(13):1592-605.

82. Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. 2006;20(9):1123-36.

83. Cheedipudi S, Puri D, Saleh A, Gala HP, Rumman M, Pillai MS, et al. A fine balance: epigenetic control of cellular quiescence by the tumor suppressor PRDM2/RIZ at a bivalent domain in the cyclin a gene. Nucleic Acids Res. 2015;43(13):6236-56.

84. Kim J, Lee Y, Lu X, Song B, Fong KW, Cao Q, et al. Polycomb- and Methylation-Independent Roles of EZH2 as a Transcription Activator. Cell Rep. 2018;25(10):2808-20.e4.

85. Kim KH, Roberts CW. Targeting EZH2 in cancer. Nat Med. 2016;22(2):128-34.

86. van der Voorn L, Ploegh HL. The WD-40 repeat. FEBS Lett. 1992;307(2):131-4.

87. Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. 1999;13(15):1924-35.

88. Zhang Y, Iratni R, Erdjument-Bromage H, Tempst P, Reinberg D. Histone deacetylases and SAP18, a novel polypeptide, are components of a human Sin3 complex. Cell. 1997;89(3):357-64.

89. Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev. 2002;16(22):2893-905.

90. Murzina NV, Pei XY, Zhang W, Sparkes M, Vicente-Garcia J, Pratap JV, et al. Structural basis for the recognition of histone H4 by the histone-chaperone RbAp46. Structure. 2008;16(7):1077-85.

91. Qian YW, Lee EY. Dual retinoblastoma-binding proteins with properties related to a negative regulator of ras in yeast. J Biol Chem. 1995;270(43):25507-13.

92. Loyola A, Almouzni G. Histone chaperones, a supporting role in the limelight. Biochim Biophys Acta. 2004;1677(1-3):3-11.

93. Searle NE, Pillus L. Critical genomic regulation mediated by Enhancer of Polycomb. Curr Genet. 2018;64(1):147-54.

94. Doyon Y, Selleck W, Lane WS, Tan S, Côté J. Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. Mol Cell Biol. 2004;24(5):1884-96.

95. Kalkat M, Resetca D, Lourenco C, Chan PK, Wei Y, Shiah YJ, et al. MYC Protein Interactome Profiling Reveals Functionally Distinct Regions that Cooperate to Drive Tumorigenesis. Mol Cell. 2018;72(5):836-48.e7.

96. Huang X, Spencer GJ, Lynch JT, Ciceri F, Somerville TD, Somervaille TC. Enhancers of Polycomb EPC1 and EPC2 sustain the oncogenic potential of MLL leukemia stem cells. Leukemia. 2014;28(5):1081-91.

97. Cheang MCU, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK, et al. Basal-Like Breast Cancer Defined by Five Biomarkers Has Superior Prognostic Value than Triple-Negative Phenotype. 2008;14(5):1368-76.

98. Zaremba T, Curtin NJ. PARP inhibitor development for systemic cancer targeting. Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents). 2007;7(5):515-23.

99. Plummer R, Lorigan P, Steven N, Scott L, Middleton MR, Wilson RH, et al. A phase II study of the potent PARP inhibitor, Rucaparib (PF-01367338, AG014699), with temozolomide in patients with metastatic melanoma demonstrating evidence of chemopotentiation. Cancer chemotherapy and pharmacology. 2013;71(5):1191-9.

100. Lord CJ, Ashworth A. PARP inhibitors: Synthetic lethality in the clinic. Science. 2017;355(6330):1152-8.

101. Masoud V, Pagès G. Targeted therapies in breast cancer: New challenges to fight against resistance. World J Clin Oncol. 2017;8(2):120-34.

102. Chung W, Eum HH, Lee H-O, Lee K-M, Lee H-B, Kim K-T, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nature communications. 2017;8:15081-.

103. Gatza ML, Lucas JE, Barry WT, Kim JW, Wang Q, D. Crawford M, et al. A pathway-based classification of human breast cancer. 2010;107(15):6994-9.

104. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols. 2009;4(1):44-57.

105. Bouchal P, Schubert OT, Faktor J, Capkova L, Imrichova H, Zoufalova K, et al. Breast Cancer Classification Based on Proteotypes Obtained by SWATH Mass Spectrometry. Cell Rep. 2019;28(3):832-43.e7.

106. Ding YC, Steele L, Warden C, Wilczynski S, Mortimer J, Yuan Y, et al. Molecular subtypes of triple-negative breast cancer in women of different race and ethnicity. Oncotarget. 2019;10(2):198-208.

107. Kim IS, Gao Y, Welte T, Wang H, Liu J, Janghorban M, et al. Immuno-subtyping of breast cancer reveals distinct myeloid cell profiles and immunotherapy resistance mechanisms. Nature Cell Biology. 2019;21(9):1113-26.

108. Risom T, Langer EM, Chapman MP, Rantala J, Fields AJ, Boniface C, et al. Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer. Nature Communications. 2018;9(1):3815.

109. Karaayvaz M, Cristea S, Gillespie SM, Patel AP, Mylvaganam R, Luo CC, et al. Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. Nature Communications. 2018;9(1):3588.

110. Echeverria GV, Powell E, Seth S, Ge Z, Carugo A, Bristow C, et al. High-resolution clonal mapping of multi-organ metastasis in triple negative breast cancer. Nature Communications. 2018;9(1):5079.

111. Sledge GW, Jr. Curing Metastatic Breast Cancer. Journal of oncology practice. 2016;12(1):6-10.

112. Hu K, Law JH, Fotovati A, Dunn SE. Small interfering RNA library screen identified polo-like kinase-1 (PLK1) as a potential therapeutic target for breast cancer that uniquely eliminates tumor-initiating cells. Breast Cancer Res. 2012;14(1):R22.

113. Petrocca F, Altschuler G, Tan SM, Mendillo ML, Yan H, Jerry DJ, et al. A genome-wide siRNA screen identifies proteasome addiction as a vulnerability of basal-like triple-negative breast cancer cells. Cancer Cell. 2013;24(2):182-96.

114. Singel SM, Cornelius C, Batten K, Fasciani G, Wright WE, Lum L, et al. A targeted RNAi screen of the breast cancer genome identifies KIF14 and TLN1 as genes that modulate docetaxel chemosensitivity in triple-negative breast cancer. Clin Cancer Res. 2013;19(8):2061-70.

115. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature. 2012;486(7403):395-9.

116. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, et al. Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell. 2018;173(4):879-93.e13.

117. Januškevičienė I, Petrikaitė V. Heterogeneity of breast cancer: The importance of interaction between different tumor cell populations. Life sciences. 2019;239:117009.

118. Sittampalam GS, Kahl SD, Janzen WP. High-throughput screening: advances in assay technologies. Current Opinion in Chemical Biology. 1997;1(3):384-91.

119. Lu Y, Qin S, Zhang B, Dai A, Cai X, Ma M, et al. Accelerating the Throughput of Affinity Mass Spectrometry-Based Ligand Screening toward a G Protein-Coupled Receptor. Anal Chem. 2019;91(13):8162-9.

120. Gale M, Yan Q. High-throughput screening to identify inhibitors of lysine demethylases. Epigenomics. 2015;7(1):57-65.

121. Henderson MC, Azorsa DO. High-throughput RNAi screening for the identification of novel targets. Methods in molecular biology (Clifton, NJ). 2013;986:89-95.

122. Bibic L, Herzig V, King GF, Stokes L. Development of High-Throughput Fluorescent-Based Screens to Accelerate Discovery of P2X Inhibitors from Animal Venoms. J Nat Prod. 2019;82(9):2559-67.

123. Prével C, Pellerano M, Van TN, Morris MC. Fluorescent biosensors for high throughput screening of protein kinase inhibitors. Biotechnology journal. 2014;9(2):253-65.

124. Duong-Thi MD, Bergström M, Fex T, Isaksson R, Ohlson S. High-throughput fragment screening by affinity LC-MS. Journal of biomolecular screening. 2013;18(2):160-71.

125. Fang X, Zheng Y, Duan Y, Liu Y, Zhong W. Recent Advances in Design of Fluorescence-Based Assays for High-Throughput Screening. Anal Chem. 2019;91(1):482-504.

126. Aherne W, Garrett M, McDonald T, Workman P. CHAPTER 14 - MECHANISM-BASED HIGH-THROUGHPUT SCREENING FOR NOVEL ANTICANCER DRUG DISCOVERY. In: Baguley BC, Kerr DJ, editors. Anticancer Drug Development. San Diego: Academic Press; 2002. p. 249-67.

127. McDermott SP, Wicha MS. Targeting breast cancer stem cells. Molecular oncology. 2010;4(5):404-19.

128. Baylin SB, Ohm JE. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? Nature reviews Cancer. 2006;6(2):107-16.

129. Eden A, Gaudet F, Waghmare A, Jaenisch R. Chromosomal instability and tumors promoted by DNA hypomethylation. Science. 2003;300(5618):455.

130. Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet. 2005;37(4):391-400.

131. Gambacorta V, Gnani D, Vago L, Di Micco R. Epigenetic Therapies for Acute Myeloid Leukemia and Their Immune-Related Effects. Frontiers in cell and developmental biology. 2019;7:207.

132. Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature. 1994;367(6464):645-8.

133. Yan J, Li B, Lin B, Lee PT, Chung T-H, Tan J, et al. EZH2 phosphorylation by JAK3 mediates a switch to noncanonical function in natural killer/T-cell lymphoma. Blood. 2016;128(7):948-58.

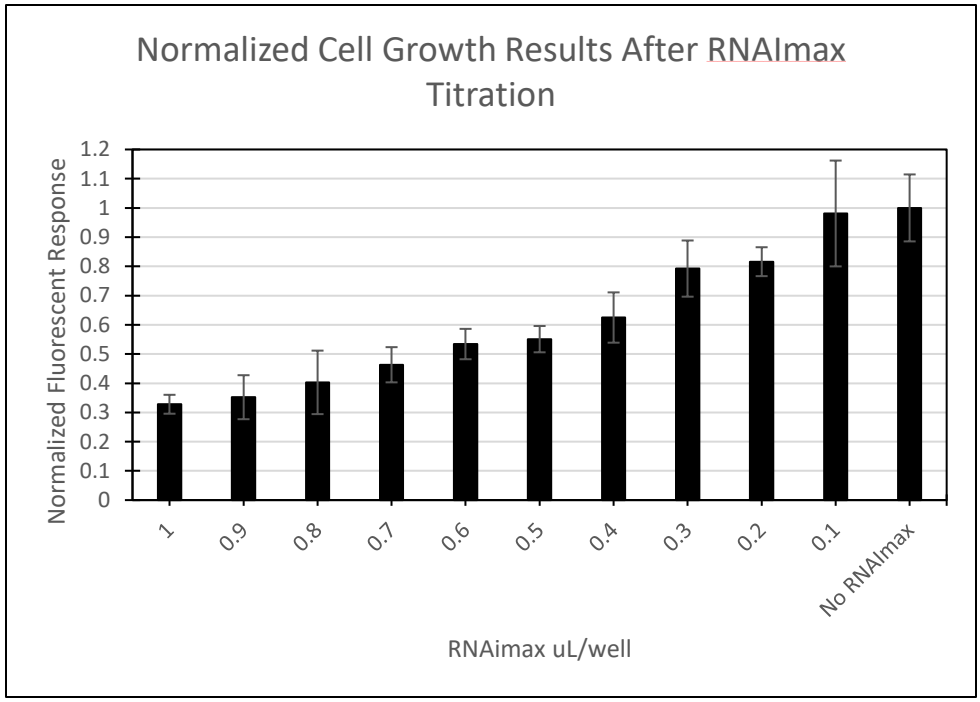

Normalized Cell Growth Results After RNAImax Titration
Normalized Fluorescent Response
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
No RNAImax
RNAimax uL/well

Figure 2.1 Optimization of RNAimax dose in SUM149 by cell proliferation (MTS).

Cells (n=3) were seeded in a 96 well plate, allowed to settle for 24 hours, and were dosed with RNAImax in levels ranging from 1 uL/well to 0.1 uL per well in 0.1 uL incriments. Cells were allowed to grow for 72 additional hours, and the cells were counted by cell proliferation according to Promega's instructions. Cells counts were divided by "No RNAImax" wells. Final values are mean +/- standard deviation.

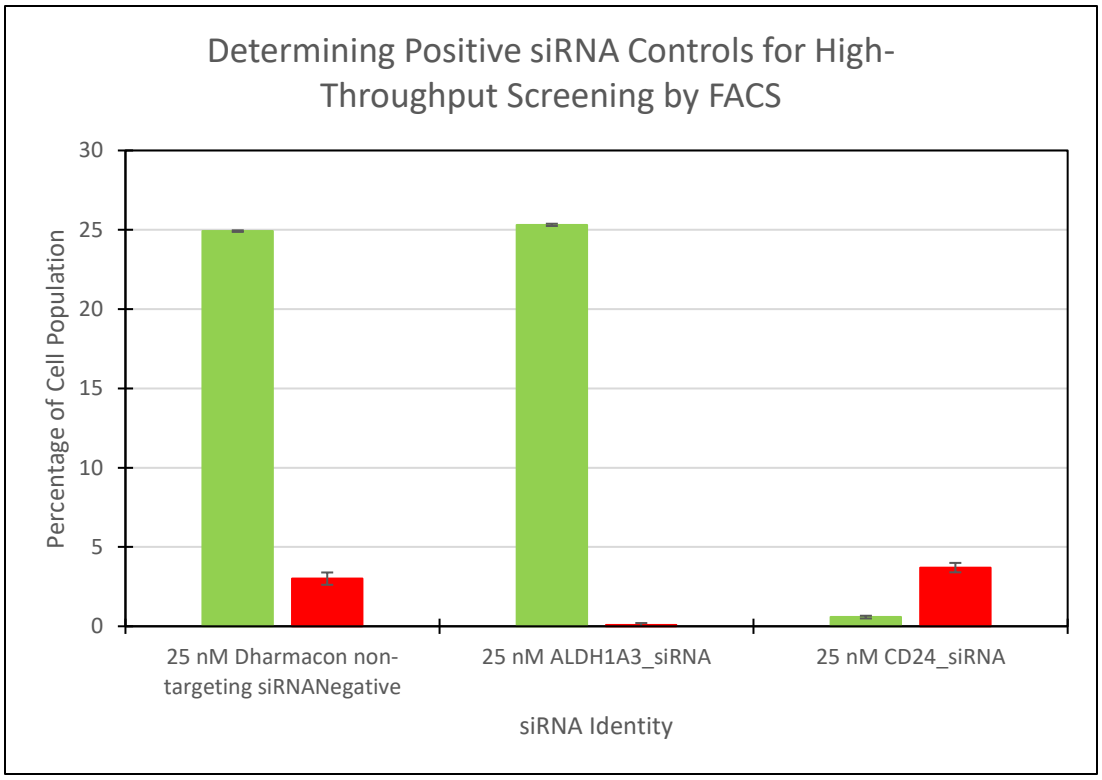

Determining Positive siRNA Controls for High-Throughput Screening by FACS
Percentage of Cell Population
30
25
20
15
10
5
0
25 nM Dharmacon non-targeting siRNANegative
25 nM ALDH1A3_siRNA
25 nM CD24_siRNA
siRNA Identity

Figure 2.2 Positive and negative siRNA control selection by FACS analysis.

Cells were seeded in 6 well plates at 50k cells per well (n=3). After settling for 24 hours, the cells were dosed with either 25 nM Dharmacon Non-Targeting siRNA, 25 nM ALDH1A3 siRNA, or 25 nM CD24 siRNA with RNAimax. Wells were allowed to grow for an additional 72 hours, where they were analyzed by FACS analysis. Cells in green represent the EGFP+ population, and cells in red represent the mCherry+ population.

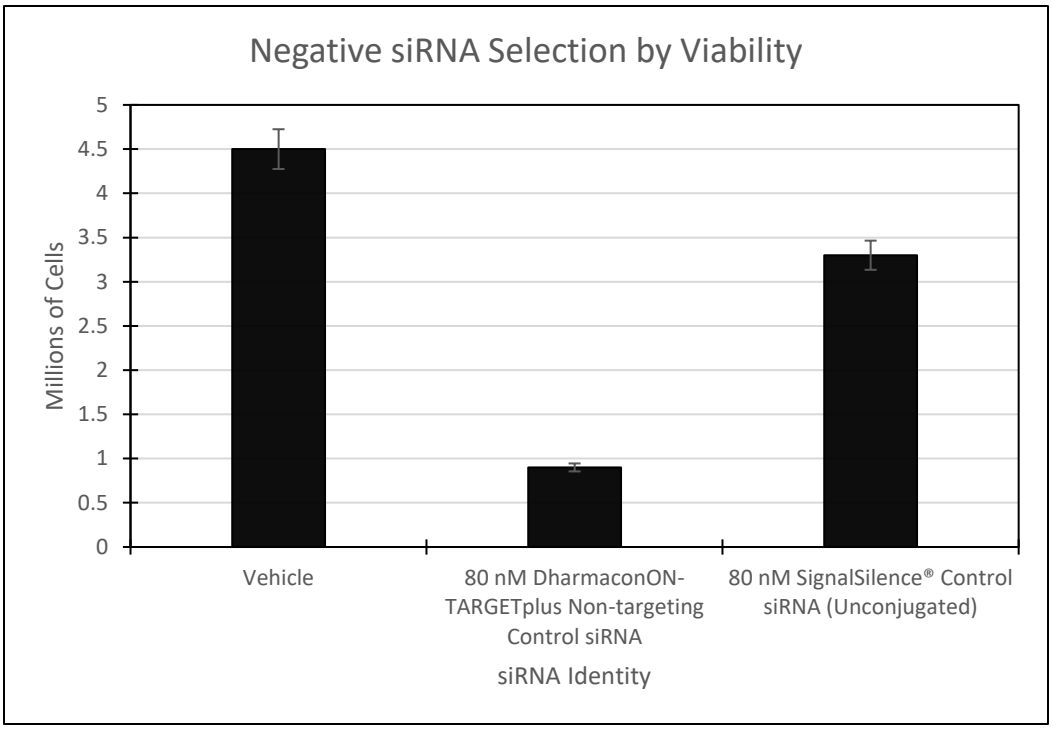

Negative siRNA Selection by Viability
Millions of Cells
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
Vehicle
80 nM DharmaconON-TARGETplus Non-targeting Control siRNA
80 nM SignalSilence® Control siRNA (Unconjugated)
siRNA Identity

Figure 2.3 Selection of Negative siRNA Control.

Cells were seeded in 6 well plates at 50k cells per well, three wells per group. After settling for 24 hours, the cells were dosed with either RNAimax alone, 25 nM Dharmacon Non-Targeting siRNA, or 25 nM Cell Signalling's Non-Targeting Control siRNA with RNAimax. Wells were allowed to grow for an additional 72 hours, where they were analyzed by FACS analysis. DAPI was used as a viability stain.

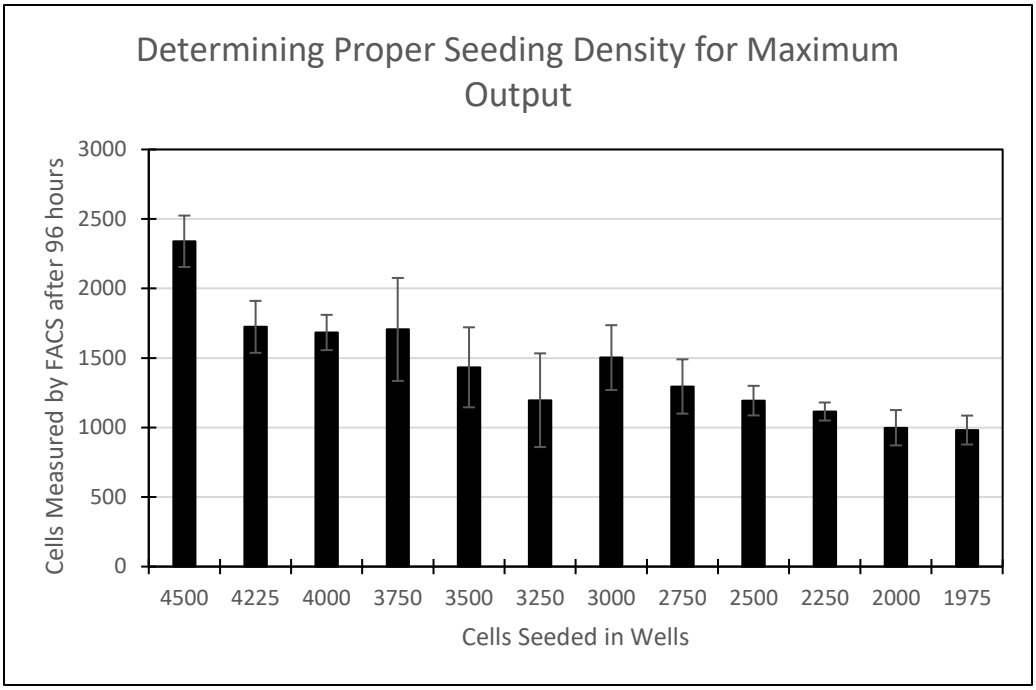
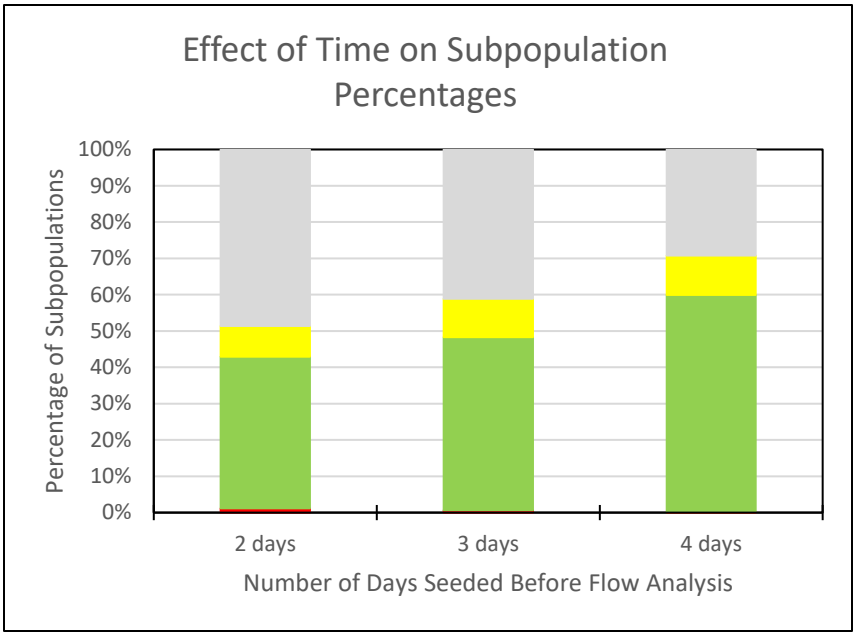

Determining Proper Seeding Density for Maximum Output
Cells Measured by FACS after 96 hours
3000
2500
2000
1500
1000
500
0
4500
4225
4000
3750
3500
3250
3000
2750
2500
2250
2000
1975
Cells Seeded in Wells

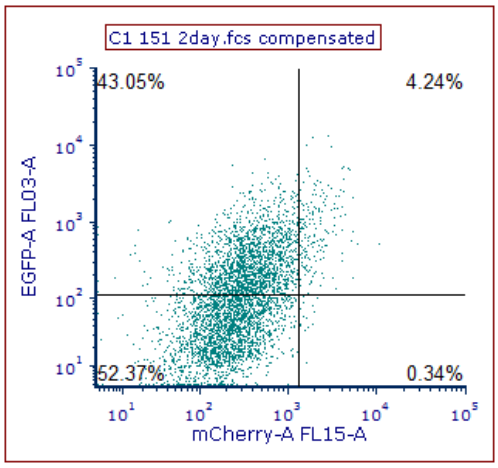
Figure 2.4 Cell density Titration and Testing of Cell Detachment Methods.

In a 96 well plate cells were seeded in a range from 4500 to 1975 cells, each in 225 cell increments (n=3). They were each dosed with RNAimax and allowed to grow for 96 hours following seeding. Cells were detached after aspirating media and by 50 uL of trypsin, and once detached, that was neutralized by 50 uL of DMEM Fluorobrite. Cells were counted by FACS analysis of DAPI-positive cells. Final values are means +/- standard deviation.

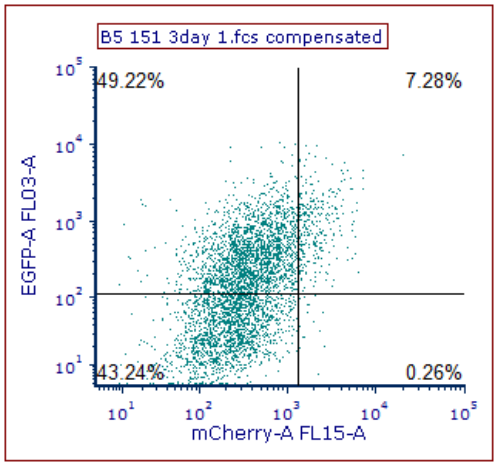
A)



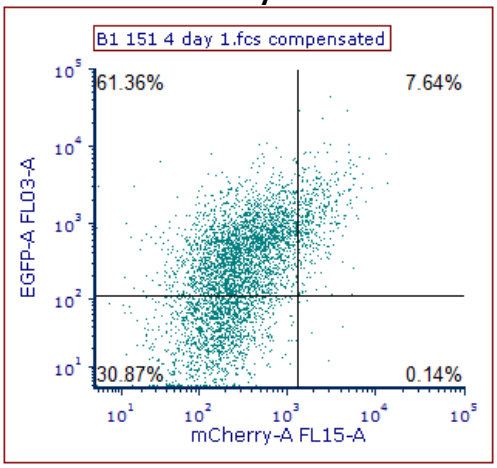
B)



2 days



4 days

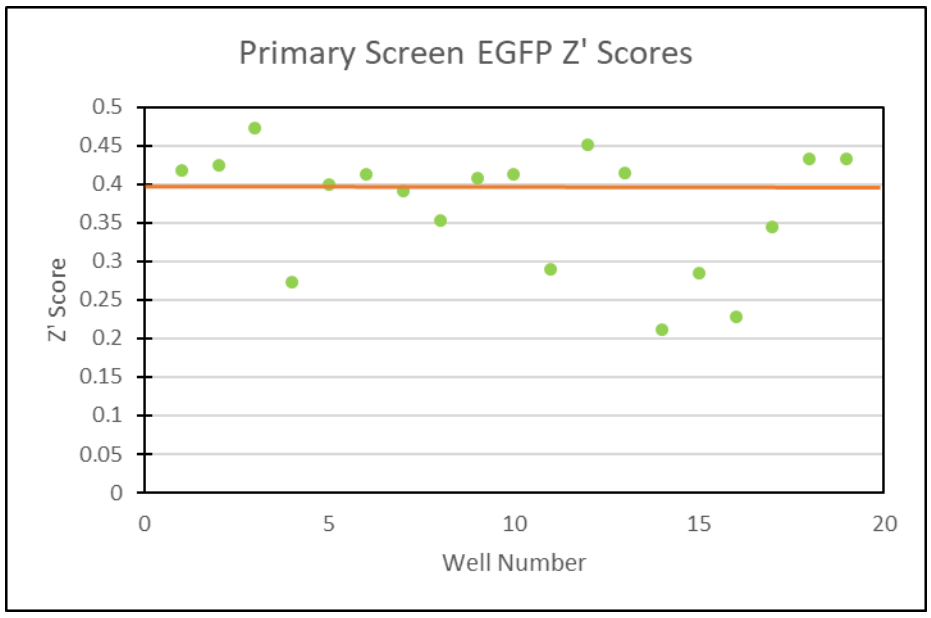


3 days

Figure 2.5 Analysis of Sub-Population Changes Over Time.

Fluorescent reporter cells were seeded at 500k cells in two 10cm plates per group. Cells were seeded 48, 72, and 96 hours before they were to be analyzed by FACS analysis. Cells were from the same passage source of cells. Cell percentages were in the correct proportion only after 96 hours total though there is a trend towards correct proportion as time continued. Panel A) depicts the final analyzed data. Panel B) depicts the trend of increasing epithelial-like cells (top left quadrant) with each day. Gray cells depict mCherry-EGFP-, yellow cells are mCherry+EGFP+, and Green cells are mCherry-EGFP+.

A)


Primary Screen EGFP Z' Scores
Z' Score
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
20
Well Number

B)

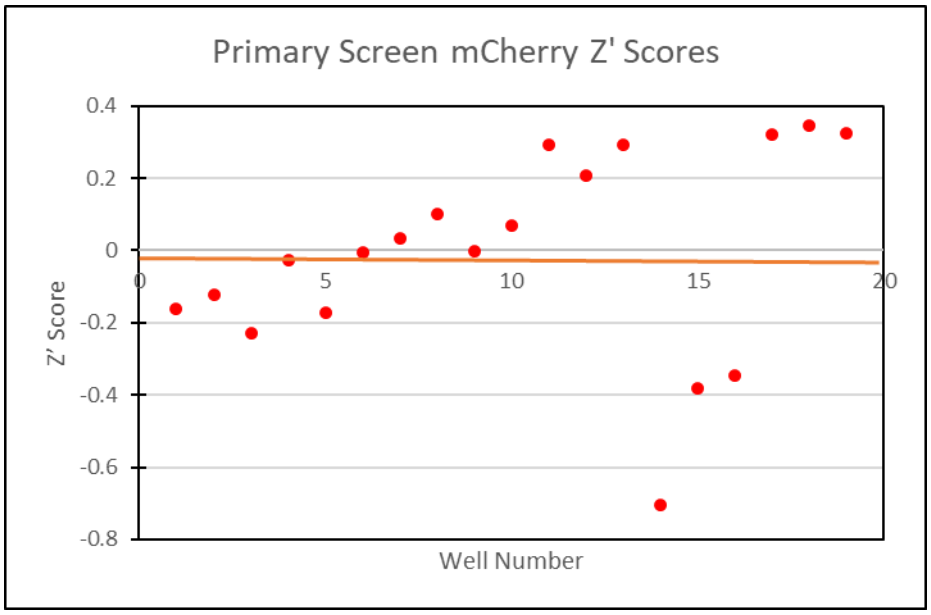
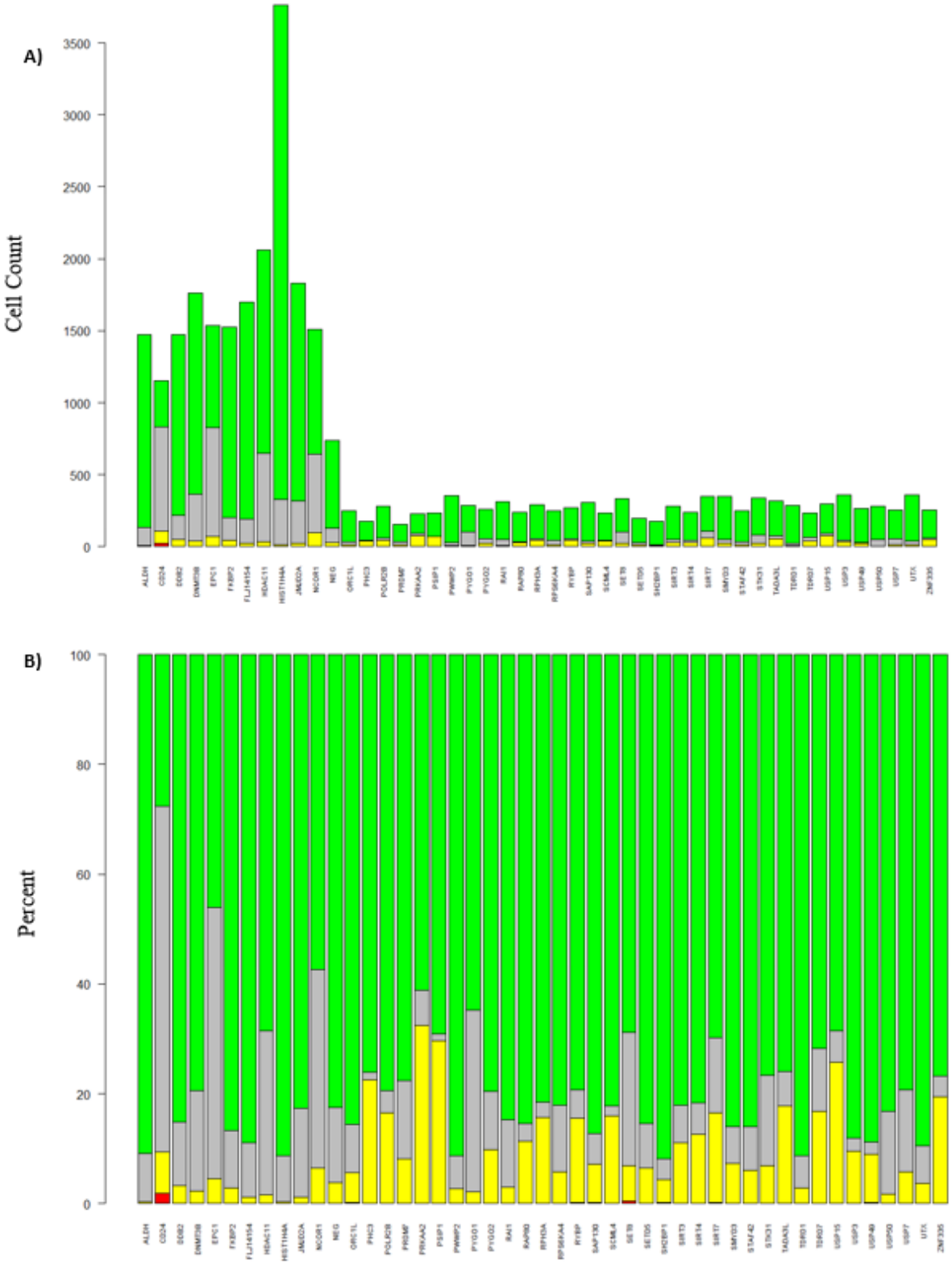
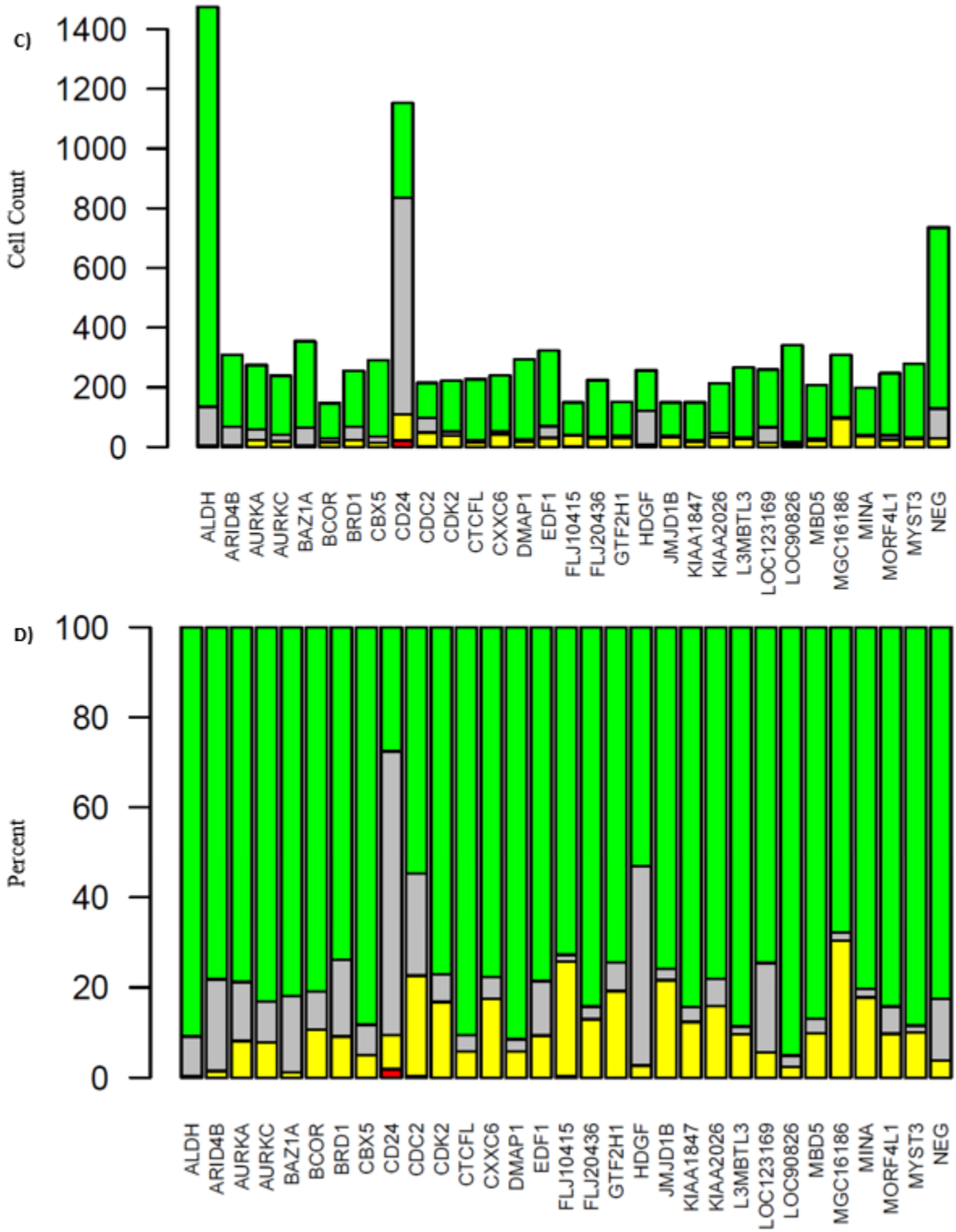

Primary Screen mCherry Z' Scores
Z' Score
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
0
5
10
15
20
Well Number

Figure 2.6 Determination of Z-prime Score by Fluorescent Output.

A 96 Well plate was seeded with ALDH1A3 and CD24 siRNA at 25 nM, and the results were calculated based on the effect had. These results are over several different days using fluorescent reporter cells at differing levels of density, and thus, likely different levels of A) EGFP and B) mCherry levels. Cells in green depict the EGFP+ cells and cells in red depict the mCherry+ cells. Scores were produced using the equation $1 - (3SD^{+} + 3SD^{-})/(|Ave^{+} - Ave^{-}|)$


A)
Cell Count
3500
3000
2500
2000
1500
1000
500
0
B)
Percent
100
80
60
40
20
0


C)
Cell Count
0
200
400
600
800
1000
1200
1400
ALDH
ARID4B
AURKA
AURKC
BAZ1A
BCOR
BRD1
CBX5
CD24
CDC2
CDK2
CTCFL
CXXC6
DMAP1
EDF1
FLJ10415
FLJ20436
GTF2H1
HDGF
JMJD1B
KIAA1847
KIAA2026
L3MBTL3
LOC123169
LOC90826
MBD5
MGC16186
MINA
MORF4L1
MYST3
NEG
D)
Percent
0
20
40
60
80
100
ALDH
ARID4B
AURKA
AURKC
BAZ1A
BCOR
BRD1
CBX5
CD24
CDC2
CDK2
CTCFL
CXXC6
DMAP1
EDF1
FLJ10415
FLJ20436
GTF2H1
HDGF
JMJD1B
KIAA1847
KIAA2026
L3MBTL3
LOC123169
LOC90826
MBD5
MGC16186
MINA
MORF4L1
MYST3
NEG

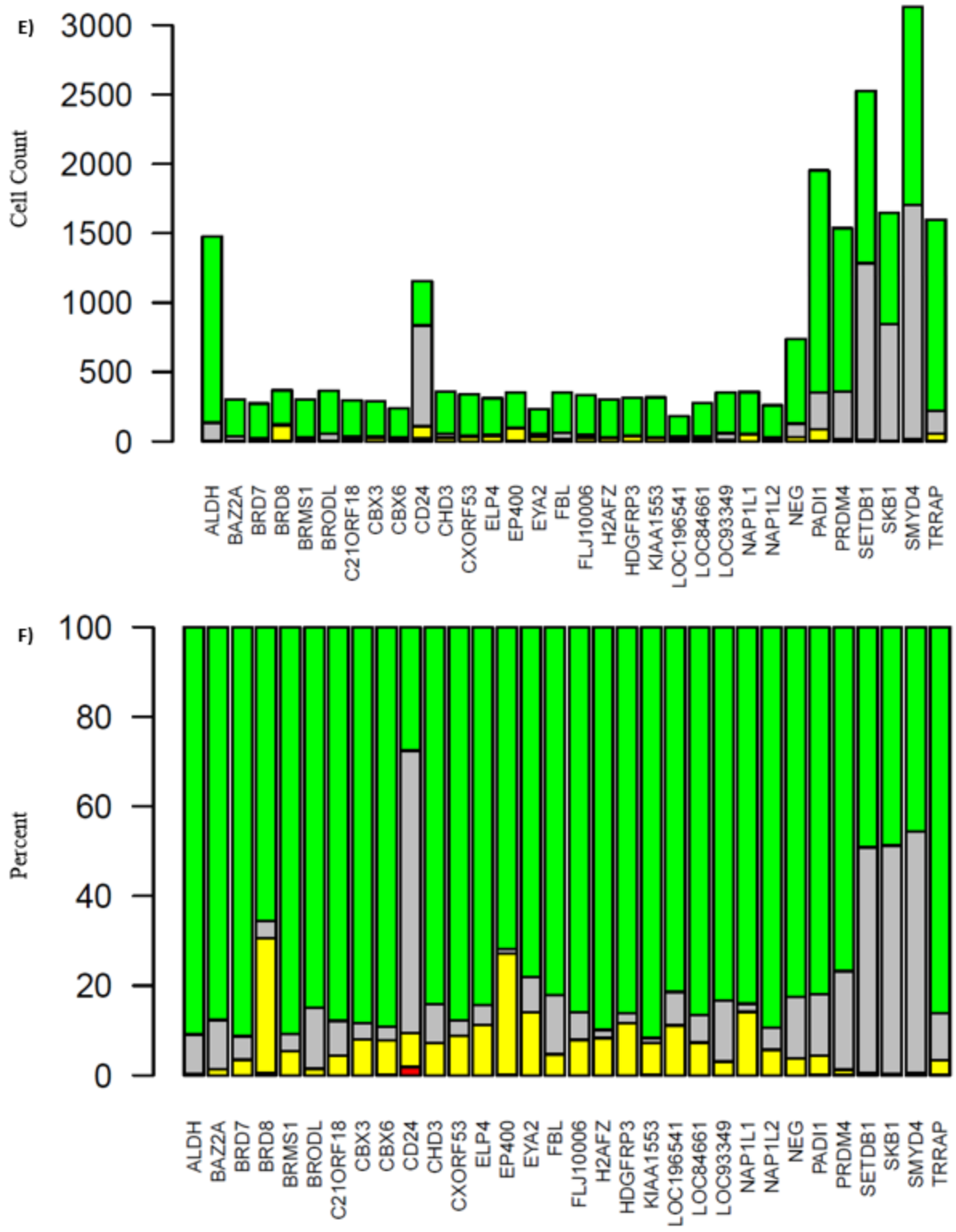
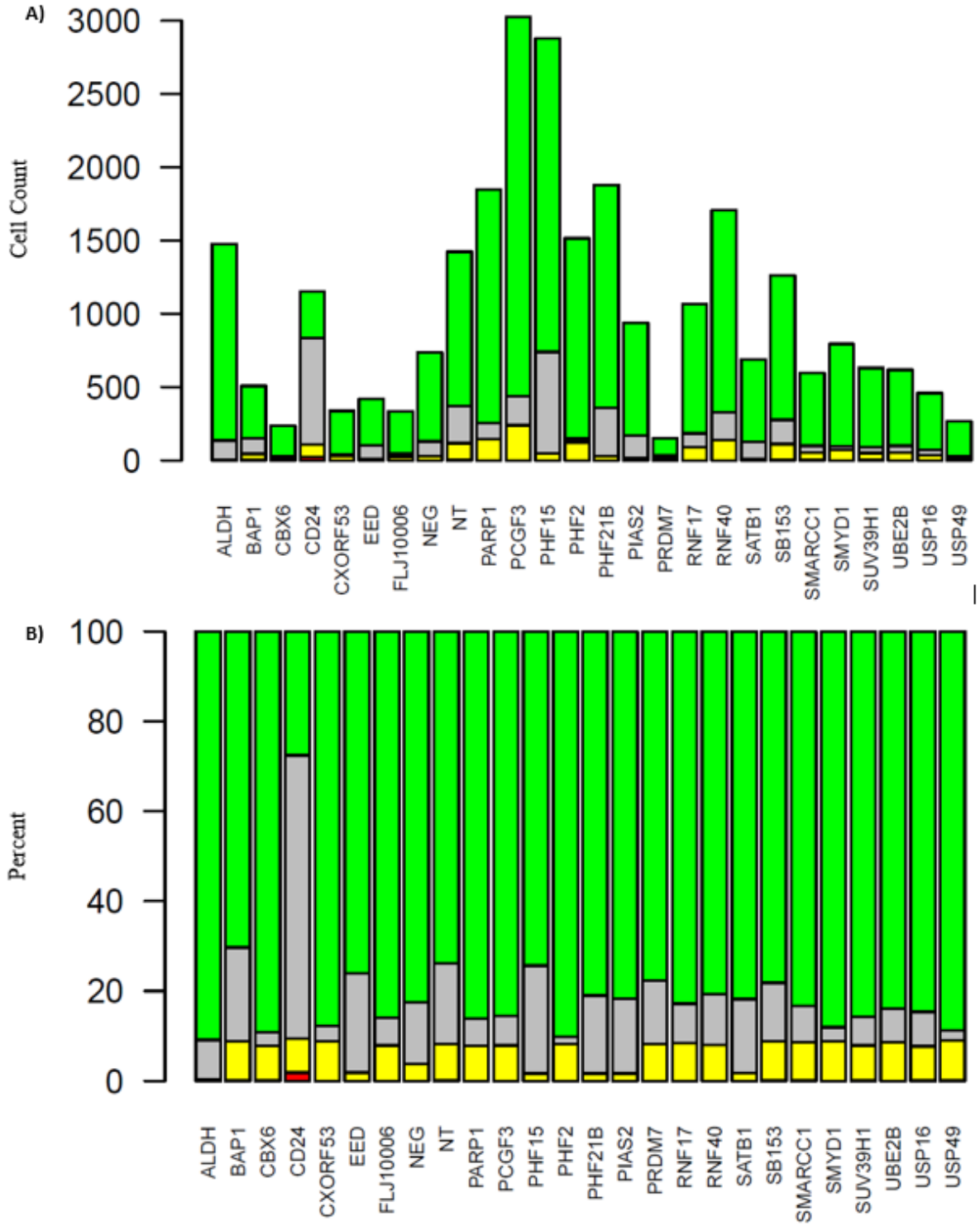
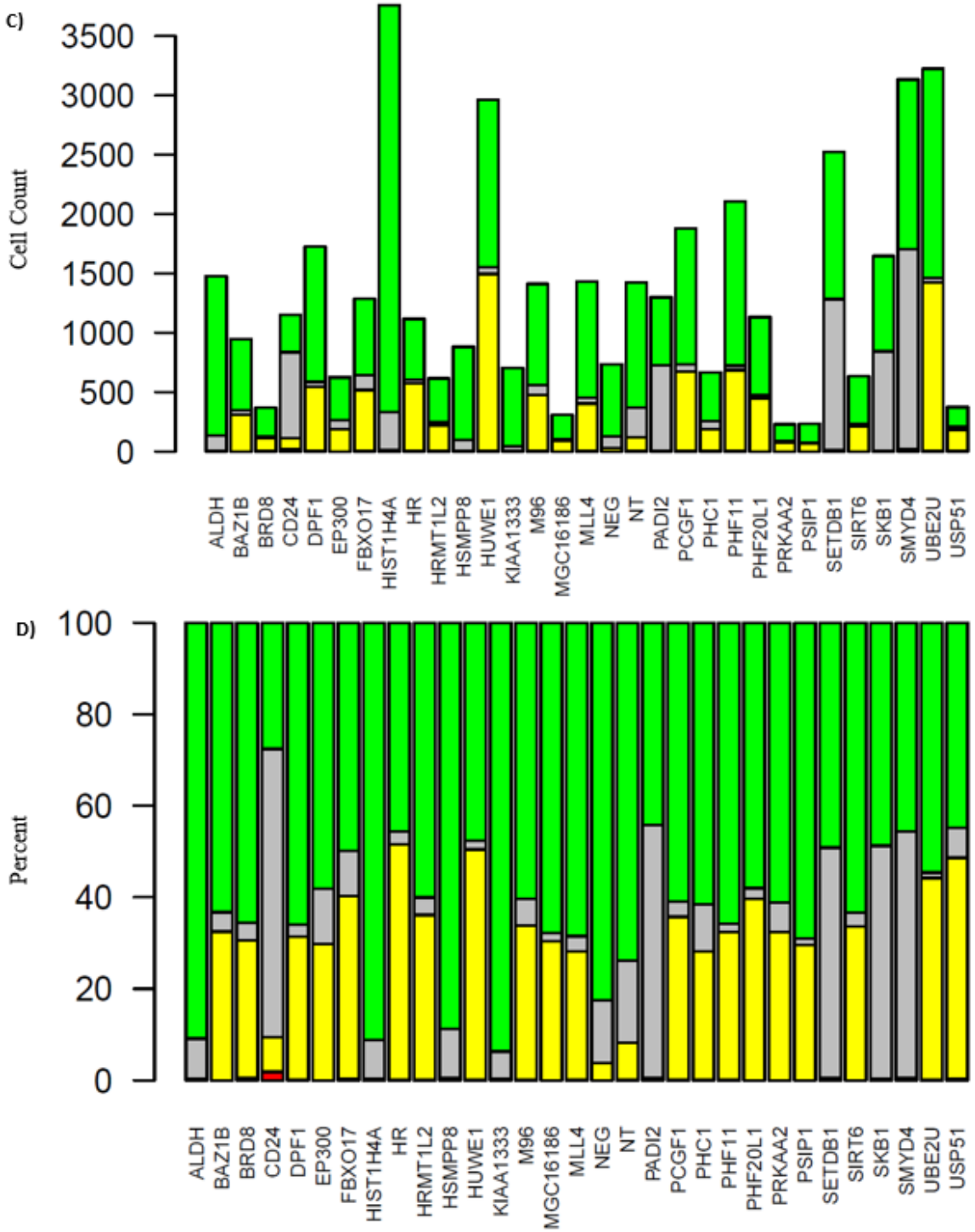
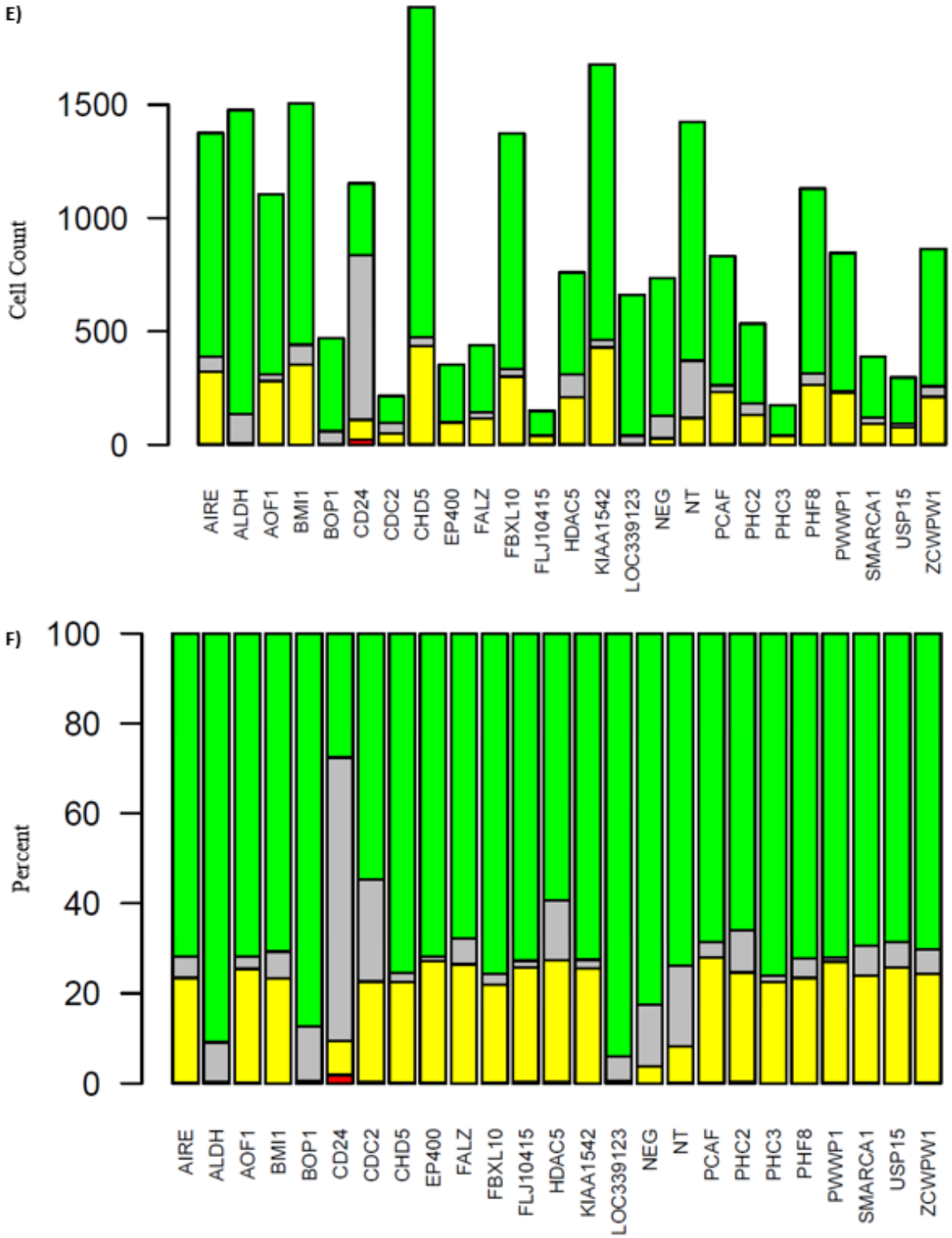

E)
Cell Count
3000
2500
2000
1500
1000
500
0
ALDH
BAZ2A
BRD7
BRD8
BRMS1
BRODL
C21ORF18
CBX3
CBX6
CD24
CHD3
CXORF53
ELP4
EP400
EYA2
FBL
FLJ10006
H2AFZ
HDGFRP3
KIAA1553
LOC196541
LOC84661
LOC93349
NAP1L1
NAP1L2
NEG
PADI1
PRDM4
SETDB1
SKB1
SMYD4
TRRAP
F)
Percent
100
80
60
40
20
0
ALDH
BAZ2A
BRD7
BRD8
BRMS1
BRODL
C21ORF18
CBX3
CBX6
CD24
CHD3
CXORF53
ELP4
EP400
EYA2
FBL
FLJ10006
H2AFZ
HDGFRP3
KIAA1553
LOC196541
LOC84661
LOC93349
NAP1L1
NAP1L2
NEG
PADI1
PRDM4
SETDB1
SKB1
SMYD4
TRRAP

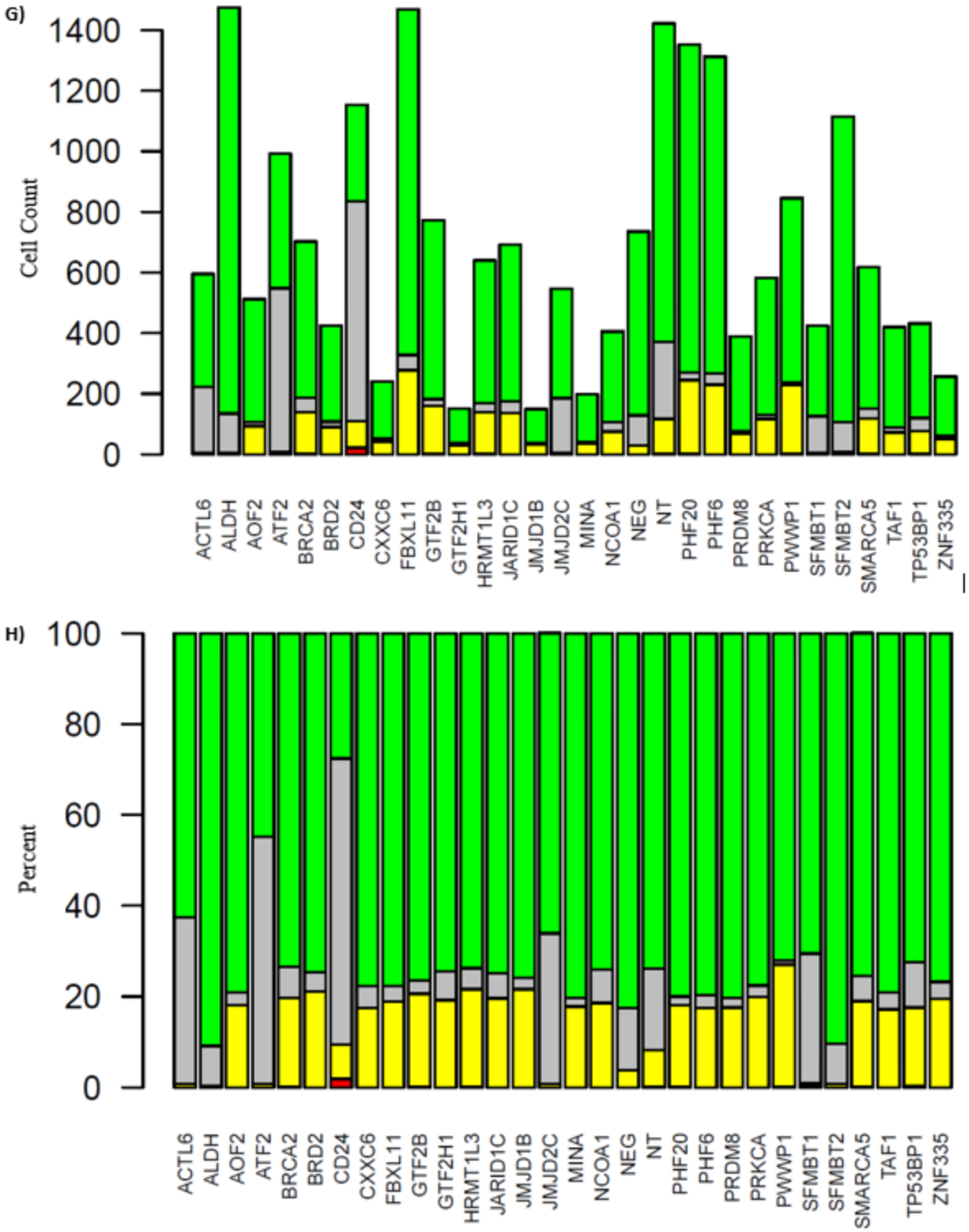
Figure 2.7 Viability Hits as Determined by Selection Criteria.

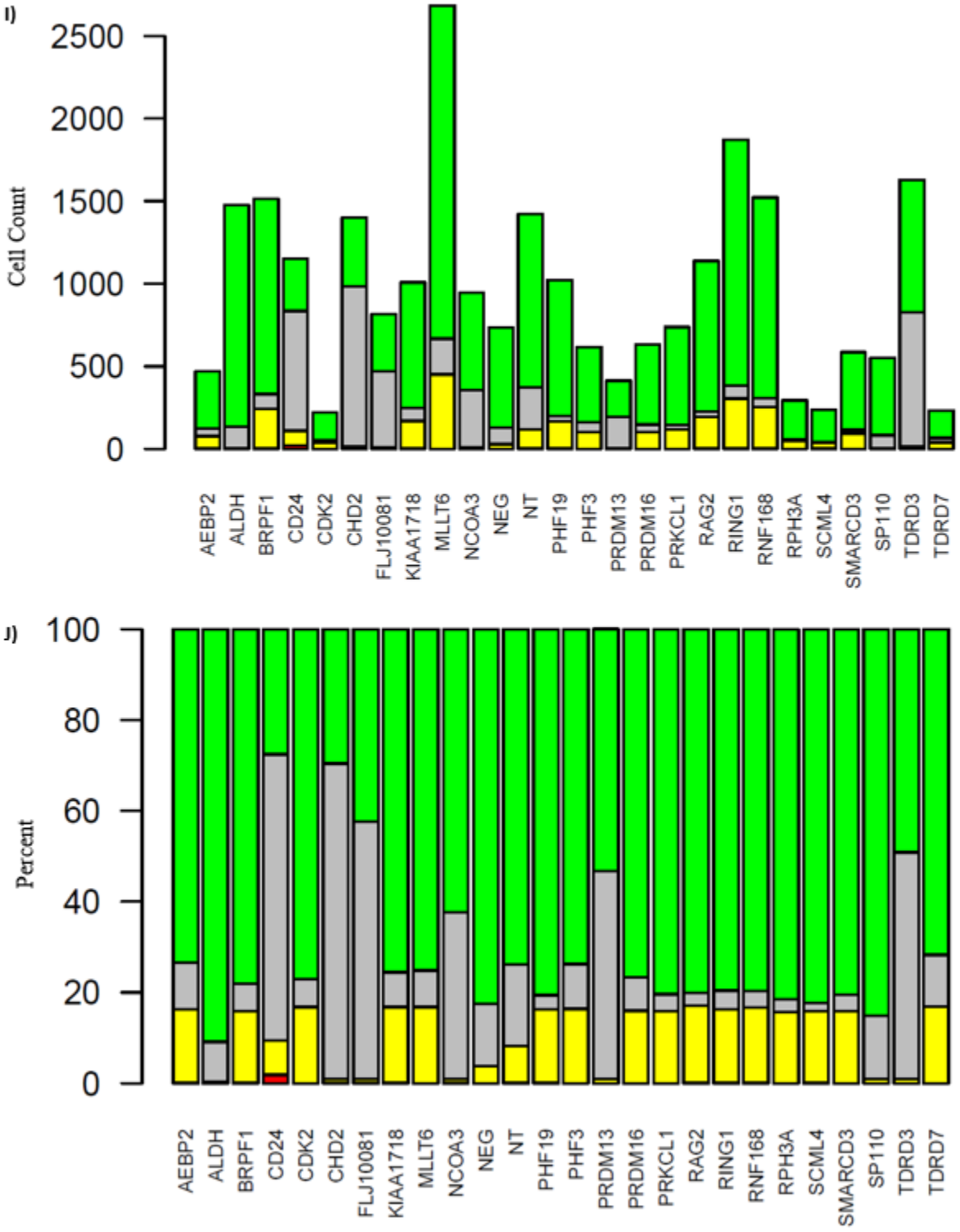
Panels A-F depict the hits selected according to selection protocols (Student's t-test, double-tailed, $p < 0.05$). Each hit is below the threshold compared with "NEG" groups. "NEG", "ALDH", and "CD24" are plate controls included in each screen. The first graph depicts total cell count, modified by those well's viability scores, and the second depicts percentage. Gray cells depict mCherry-EGFP-, yellow cells are mCherry+EGFP+, and Green cells are mCherry-EGFP+.

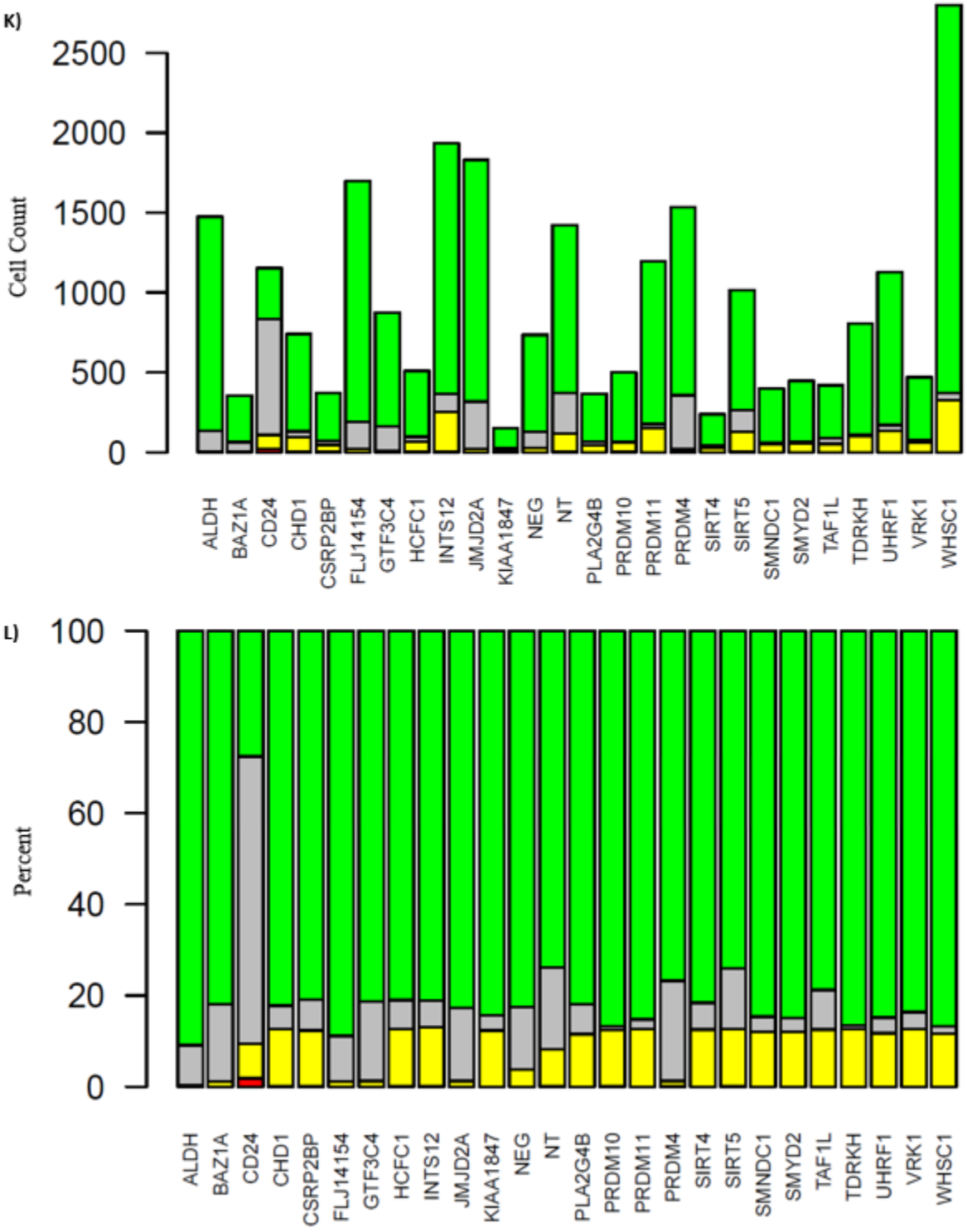

A)
Cell Count
3000
2500
2000
1500
1000
500
0
ALDH
BAP1
CBX6
CD24
CXORF53
EED
FLJ10006
NEG
NT
PARP1
PCGF3
PHF15
PHF2
PHF21B
PIAS2
PRDM7
RNF17
RNF40
SATB1
SB153
SMARCC1
SMYD1
SUV39H1
UBE2B
USP16
USP49
B)
Percent
100
80
60
40
20
0
ALDH
BAP1
CBX6
CD24
CXORF53
EED
FLJ10006
NEG
NT
PARP1
PCGF3
PHF15
PHF2
PHF21B
PIAS2
PRDM7
RNF17
RNF40
SATB1
SB153
SMARCC1
SMYD1
SUV39H1
UBE2B
USP16
USP49

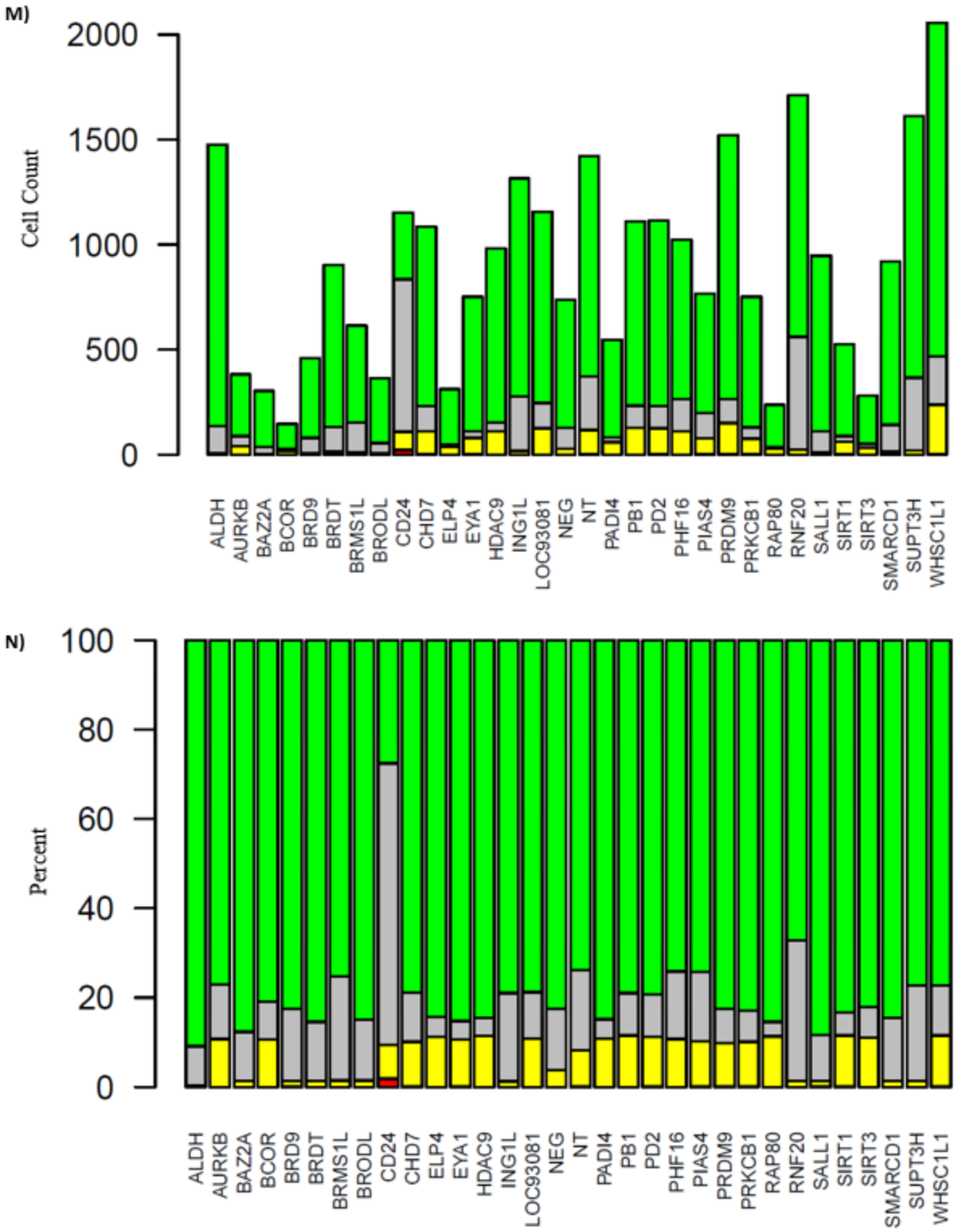

C)
Cell Count
3500
3000
2500
2000
1500
1000
500
0
ALDH
BAZ1B
BRD8
CD24
DPF1
EP300
FBXO17
HIST1H4A
HR
HRMT1L2
HSMPP8
HUWE1
KIAA1333
M96
MGC16186
MLL4
NEG
NT
PADI2
PCGF1
PHC1
PHF11
PHF20L1
PRKAA2
PSIP1
SETDB1
SIRT6
SKB1
SMYD4
UBE2U
USP51
D)
Percent
100
80
60
40
20
0
ALDH
BAZ1B
BRD8
CD24
DPF1
EP300
FBXO17
HIST1H4A
HR
HRMT1L2
HSMPP8
HUWE1
KIAA1333
M96
MGC16186
MLL4
NEG
NT
PADI2
PCGF1
PHC1
PHF11
PHF20L1
PRKAA2
PSIP1
SETDB1
SIRT6
SKB1
SMYD4
UBE2U
USP51

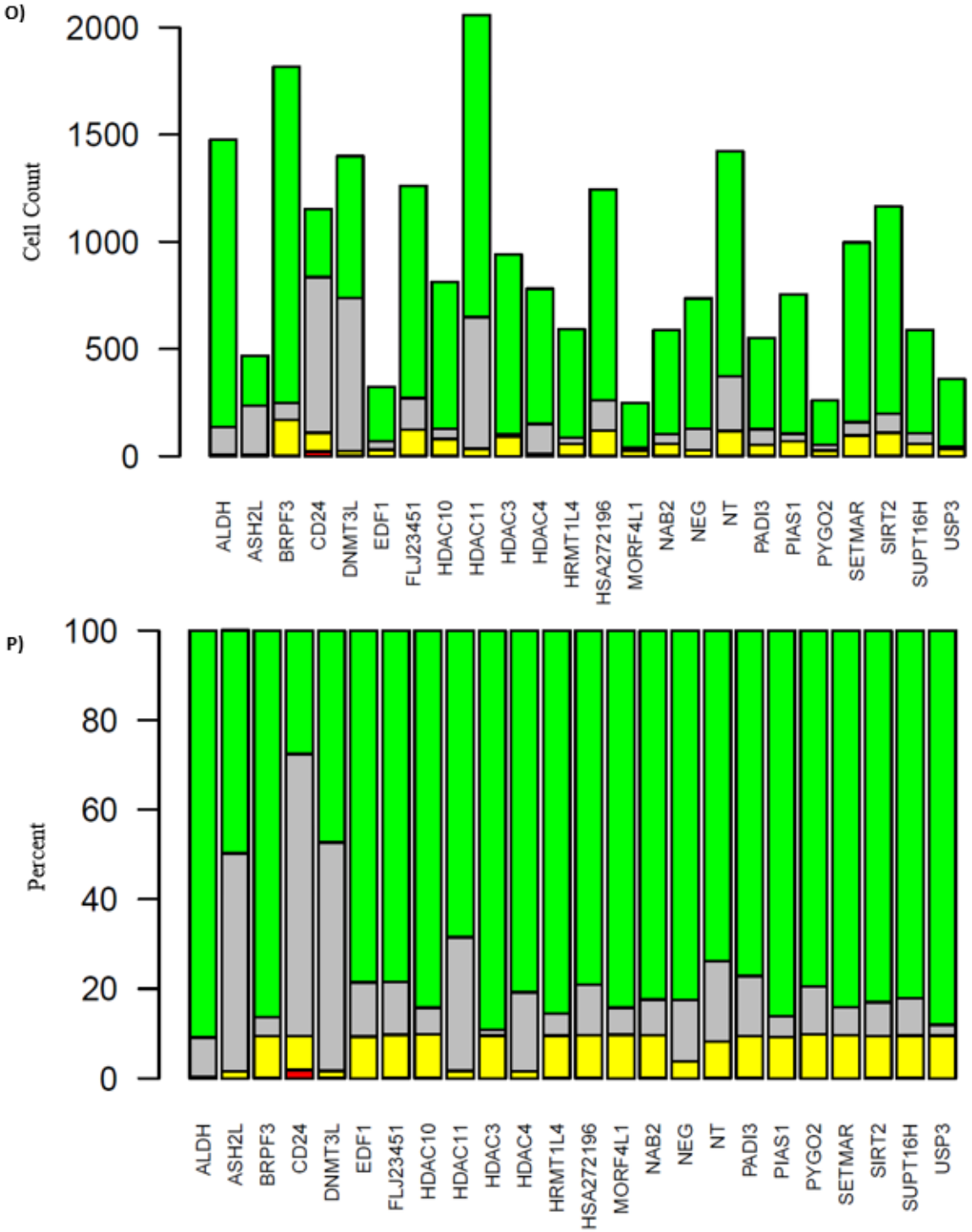

E)
Cell Count
0
500
1000
1500
AIRE
ALDH
AOF1
BMI1
BOP1
CD24
CDC2
CHD5
EP400
FALZ
FBXL10
FLJ10415
HDAC5
KIAA1542
LOC339123
NEG
NT
PCAF
PHC2
PHC3
PHF8
PWWP1
SMARCA1
USP15
ZCWPW1
F)
Percent
0
20
40
60
80
100
AIRE
ALDH
AOF1
BMI1
BOP1
CD24
CDC2
CHD5
EP400
FALZ
FBXL10
FLJ10415
HDAC5
KIAA1542
LOC339123
NEG
NT
PCAF
PHC2
PHC3
PHF8
PWWP1
SMARCA1
USP15
ZCWPW1


G)
Cell Count
0
200
400
600
800
1000
1200
1400
ACTL6
ALDH
AOF2
ATF2
BRCA2
BRD2
CD24
CXXC6
FBXL11
GTF2B
GTF2H1
HRMT1L3
JARID1C
JMJD1B
JMJD2C
MINA
NCOA1
NEG
NT
PHF20
PHF6
PRDM8
PRKCA
PWWP1
SFMBT1
SFMBT2
SMARCA5
TAF1
TP53BP1
ZNF335
H)
Percent
0
20
40
60
80
100
ACTL6
ALDH
AOF2
ATF2
BRCA2
BRD2
CD24
CXXC6
FBXL11
GTF2B
GTF2H1
HRMT1L3
JARID1C
JMJD1B
JMJD2C
MINA
NCOA1
NEG
NT
PHF20
PHF6
PRDM8
PRKCA
PWWP1
SFMBT1
SFMBT2
SMARCA5
TAF1
TP53BP1
ZNF335


I)
Cell Count
2500
2000
1500
1000
500
0
AEBP2
ALDH
BRPF1
CD24
CDK2
CHD2
FLJ10081
KIAA1718
MLLT6
NCOA3
NEG
NT
PHF19
PHF3
PRDM13
PRDM16
PRKCL1
RAG2
RING1
RNF168
RPH3A
SCML4
SMARCD3
SP110
TDRD3
TDRD7
J)
Percent
100
80
60
40
20
0
AEBP2
ALDH
BRPF1
CD24
CDK2
CHD2
FLJ10081
KIAA1718
MLLT6
NCOA3
NEG
NT
PHF19
PHF3
PRDM13
PRDM16
PRKCL1
RAG2
RING1
RNF168
RPH3A
SCML4
SMARCD3
SP110
TDRD3
TDRD7


K)
Cell Count
2500
2000
1500
1000
500
0
ALDH
BAZ1A
CD24
CHD1
CSRP2BP
FLJ14154
GTF3C4
HCFC1
INTS12
JMJD2A
KIAA1847
NEG
NT
PLA2G4B
PRDM10
PRDM11
PRDM4
SIRT4
SIRT5
SMNDC1
SMYD2
TAF1L
TDRKH
UHRF1
VRK1
WHSC1
L)
Percent
100
80
60
40
20
0
ALDH
BAZ1A
CD24
CHD1
CSRP2BP
FLJ14154
GTF3C4
HCFC1
INTS12
JMJD2A
KIAA1847
NEG
NT
PLA2G4B
PRDM10
PRDM11
PRDM4
SIRT4
SIRT5
SMNDC1
SMYD2
TAF1L
TDRKH
UHRF1
VRK1
WHSC1


M)
Cell Count
2000
1500
1000
500
0
ALDH
AURKB
BAZ2A
BCOR
BRD9
BRDT
BRMS1L
BRODL
CD24
CHD7
ELP4
EYA1
HDAC9
ING1L
LOC93081
NEG
NT
PADI4
PB1
PD2
PHF16
PIAS4
PRDM9
PRKCB1
RAP80
RNF20
SALL1
SIRT1
SIRT3
SMARCD1
SUPT3H
WHSC1L1
N)
Percent
100
80
60
40
20
0
ALDH
AURKB
BAZ2A
BCOR
BRD9
BRDT
BRMS1L
BRODL
CD24
CHD7
ELP4
EYA1
HDAC9
ING1L
LOC93081
NEG
NT
PADI4
PB1
PD2
PHF16
PIAS4
PRDM9
PRKCB1
RAP80
RNF20
SALL1
SIRT1
SIRT3
SMARCD1
SUPT3H
WHSC1L1


O)
Cell Count
2000
1500
1000
500
0
ALDH
ASH2L
BRPF3
CD24
DNMT3L
EDF1
FLJ23451
HDAC10
HDAC11
HDAC3
HDAC4
HRMT1L4
HSA272196
MORF4L1
NAB2
NEG
NT
PADI3
PIAS1
PYGO2
SETMAR
SIRT2
SUPT16H
USP3
P)
Percent
100
80
60
40
20
0
ALDH
ASH2L
BRPF3
CD24
DNMT3L
EDF1
FLJ23451
HDAC10
HDAC11
HDAC3
HDAC4
HRMT1L4
HSA272196
MORF4L1
NAB2
NEG
NT
PADI3
PIAS1
PYGO2
SETMAR
SIRT2
SUPT16H
USP3

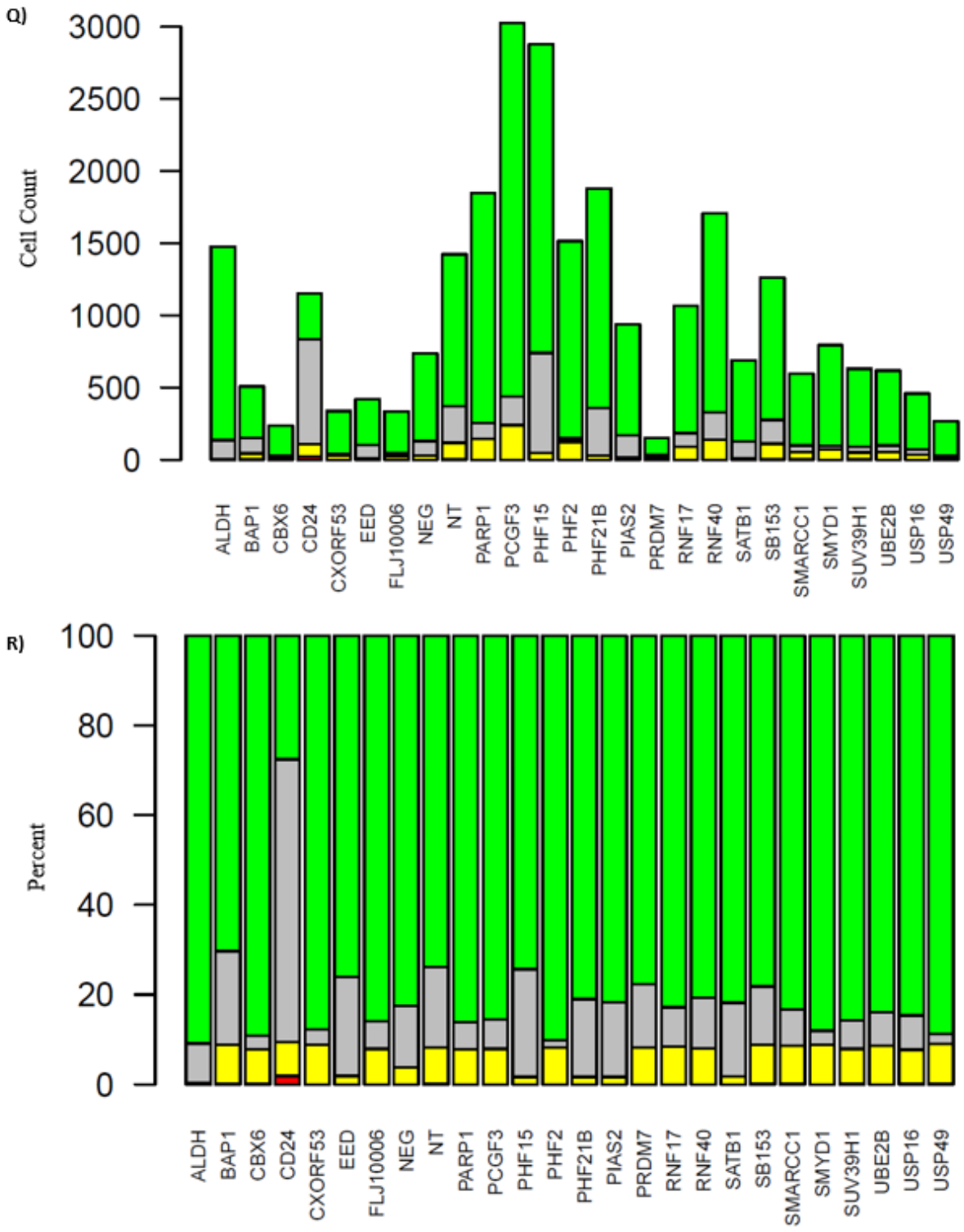
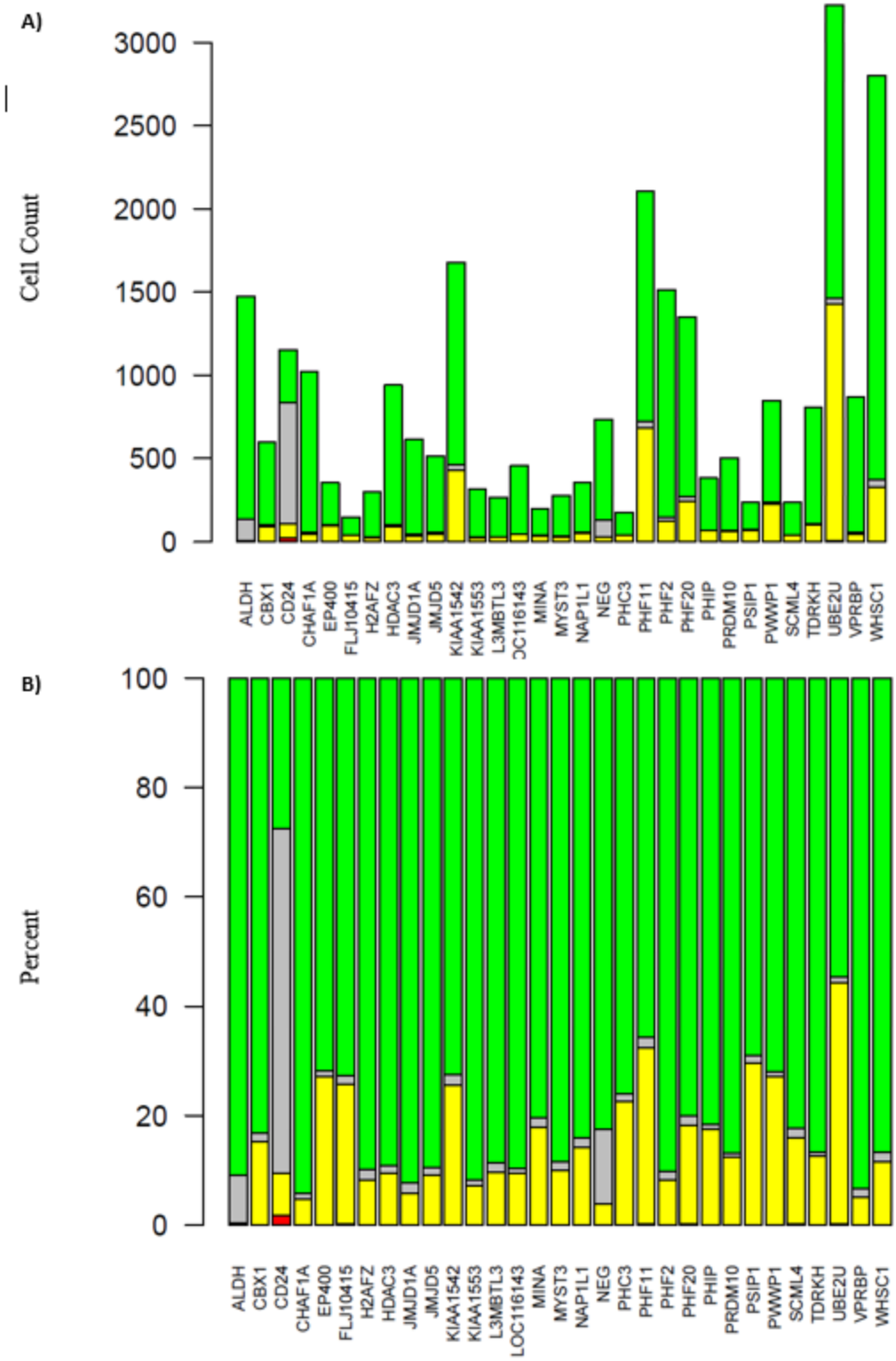
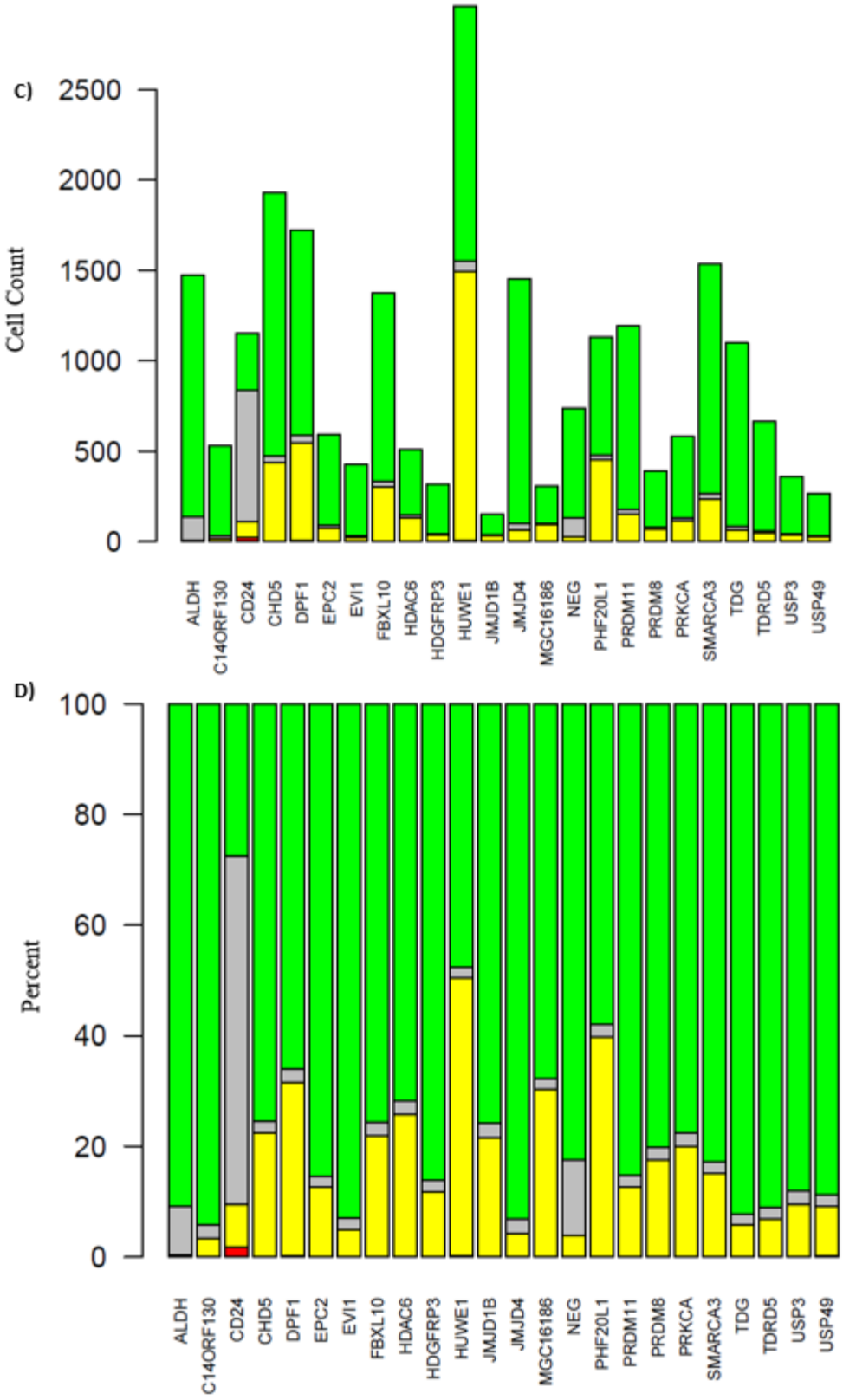
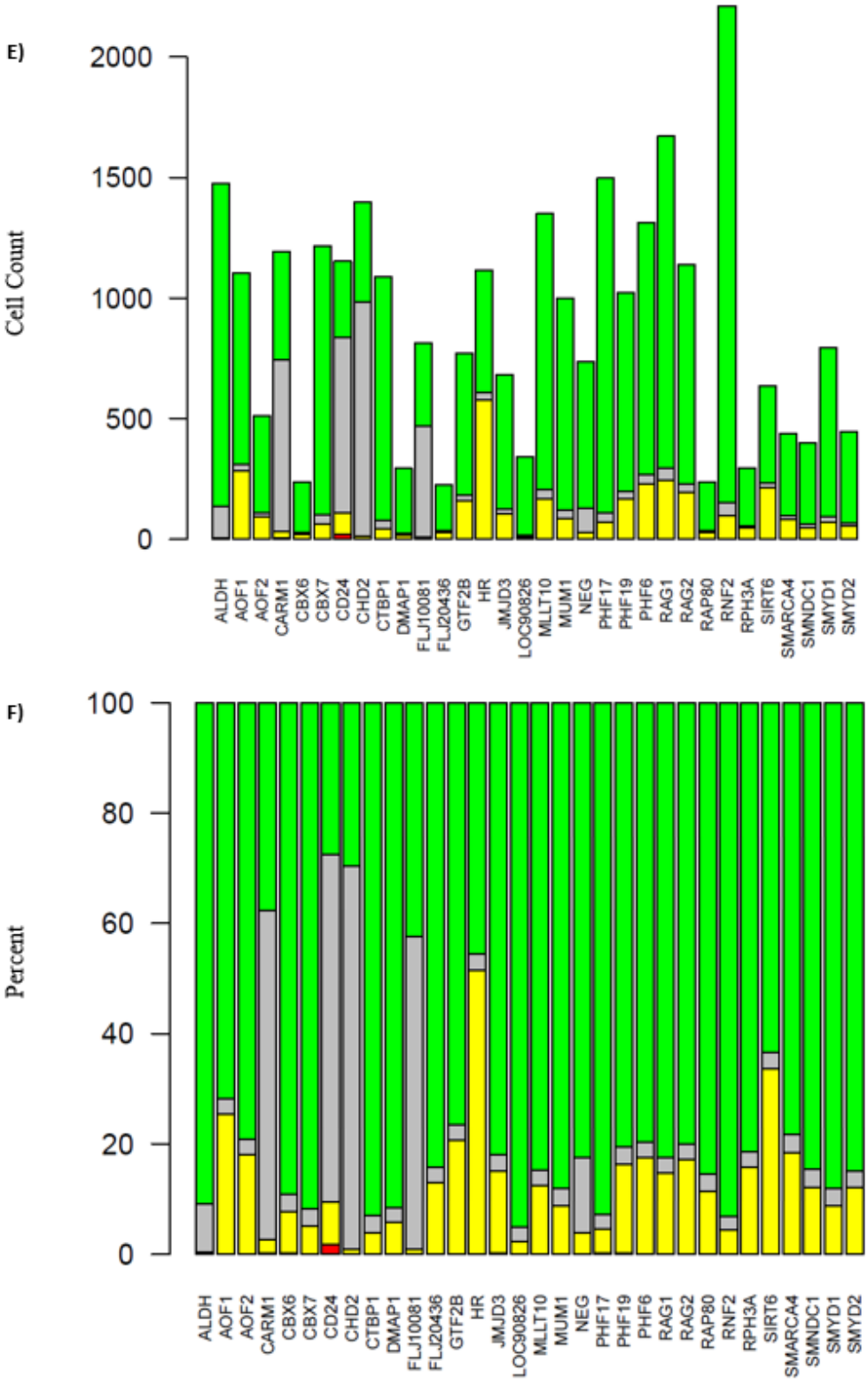

Q)
Cell Count
3000
2500
2000
1500
1000
500
0
ALDH
BAP1
CBX6
CD24
CXORF53
EED
FLJ10006
NEG
NT
PARP1
PCGF3
PHF15
PHF2
PHF21B
PIAS2
PRDM7
RNF17
RNF40
SATB1
SB153
SMARCC1
SMYD1
SUV39H1
UBE2B
USP16
USP49
R)
Percent
100
80
60
40
20
0
ALDH
BAP1
CBX6
CD24
CXORF53
EED
FLJ10006
NEG
NT
PARP1
PCGF3
PHF15
PHF2
PHF21B
PIAS2
PRDM7
RNF17
RNF40
SATB1
SB153
SMARCC1
SMYD1
SUV39H1
UBE2B
USP16
USP49

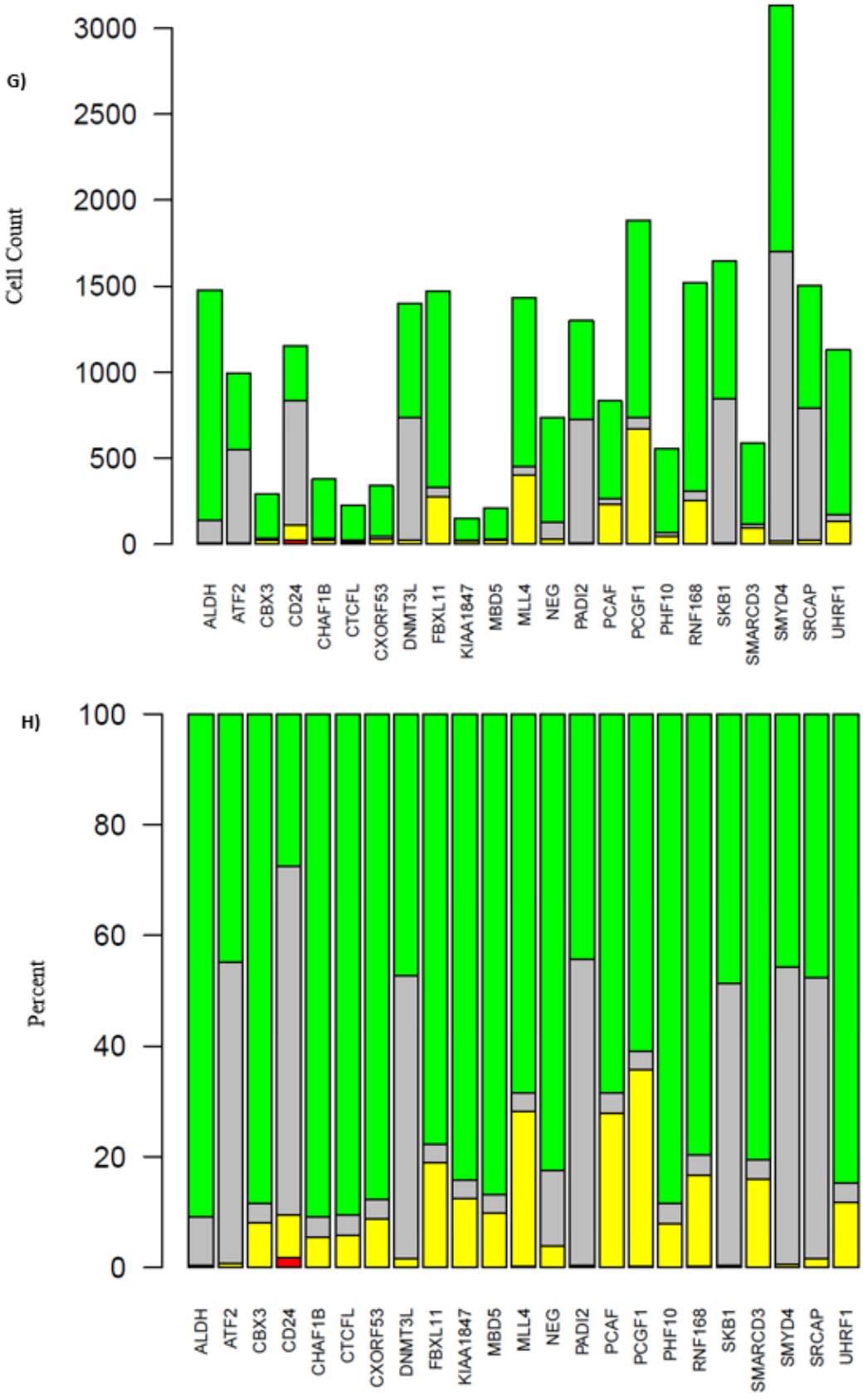
Figure 2.8 mCherry+EGFP+ Hits as Determined by Selection Criteria.

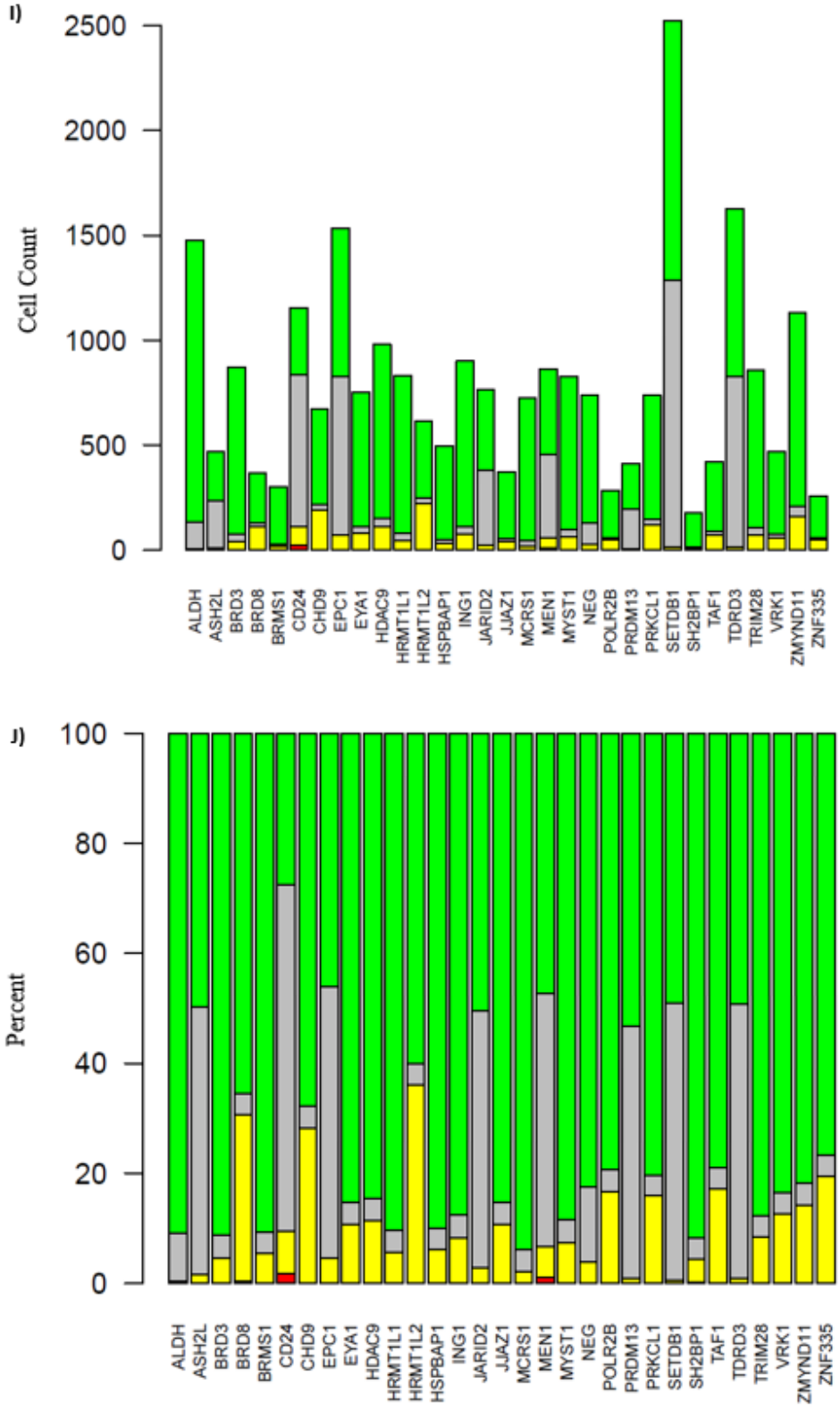
Panels A-R depict the hits selected according to selection protocols (Student's t-test, double tailed, $p < 0.05$). Each hit is below the threshold compared with "NEG" groups. "NEG", "ALDH", and "CD24" are plate controls included in each screen. The first graph depicts total cell count, modified by those well's viability scores, and the second depicts percentage. Gray cells depict mCherry-EGFP-, yellow cells are mCherry+EGFP+, and Green cells are mCherry-EGFP+.

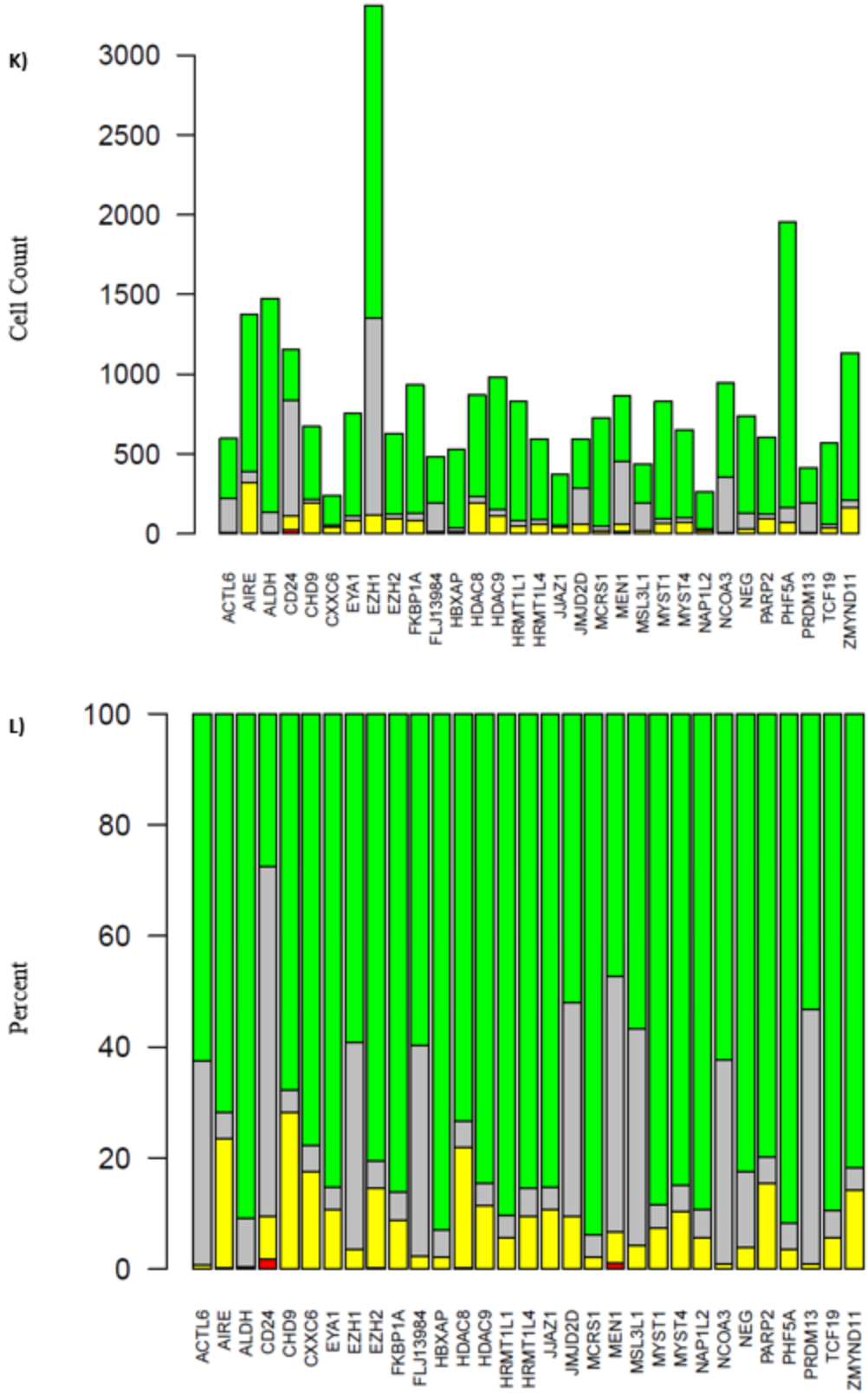

A)
Cell Count
3000
2500
2000
1500
1000
500
0
ALDH
CBX1
CD24
CHAF1A
EP400
FLJ10415
H2AFZ
HDAC3
JMJD1A
JMJD5
KIAA1542
KIAA1553
L3MBTL3
LOC116143
MINA
MYST3
NAP1L1
NEG
PHC3
PHF11
PHF2
PHF20
PHIP
PRDM10
PSIP1
PWWP1
SCML4
TDRKH
UBE2U
VPRBP
WHSC1
B)
Percent
100
80
60
40
20
0
ALDH
CBX1
CD24
CHAF1A
EP400
FLJ10415
H2AFZ
HDAC3
JMJD1A
JMJD5
KIAA1542
KIAA1553
L3MBTL3
LOC116143
MINA
MYST3
NAP1L1
NEG
PHC3
PHF11
PHF2
PHF20
PHIP
PRDM10
PSIP1
PWWP1
SCML4
TDRKH
UBE2U
VPRBP
WHSC1

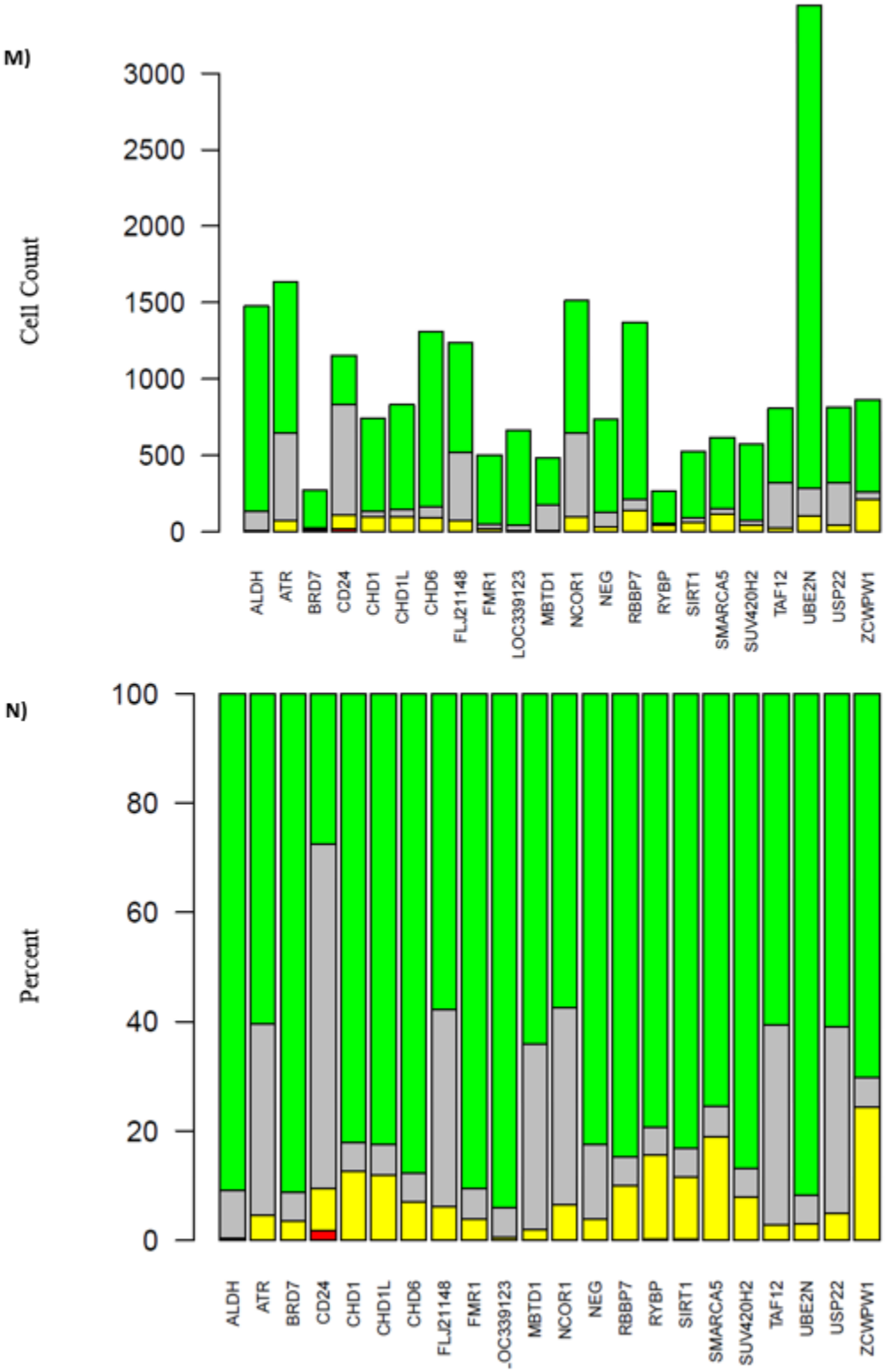

C)
Cell Count
2500
2000
1500
1000
500
0
ALDH
C14ORF130
CD24
CHD5
DPF1
EPC2
EVI1
FBXL10
HDAC6
HDGFRP3
HUWE1
JMJD1B
JMJD4
MGC16186
NEG
PHF20L1
PRDM11
PRDM8
PRKCA
SMARCA3
TDG
TDRD5
USP3
USP49
D)
Percent
100
80
60
40
20
0
ALDH
C14ORF130
CD24
CHD5
DPF1
EPC2
EVI1
FBXL10
HDAC6
HDGFRP3
HUWE1
JMJD1B
JMJD4
MGC16186
NEG
PHF20L1
PRDM11
PRDM8
PRKCA
SMARCA3
TDG
TDRD5
USP3
USP49

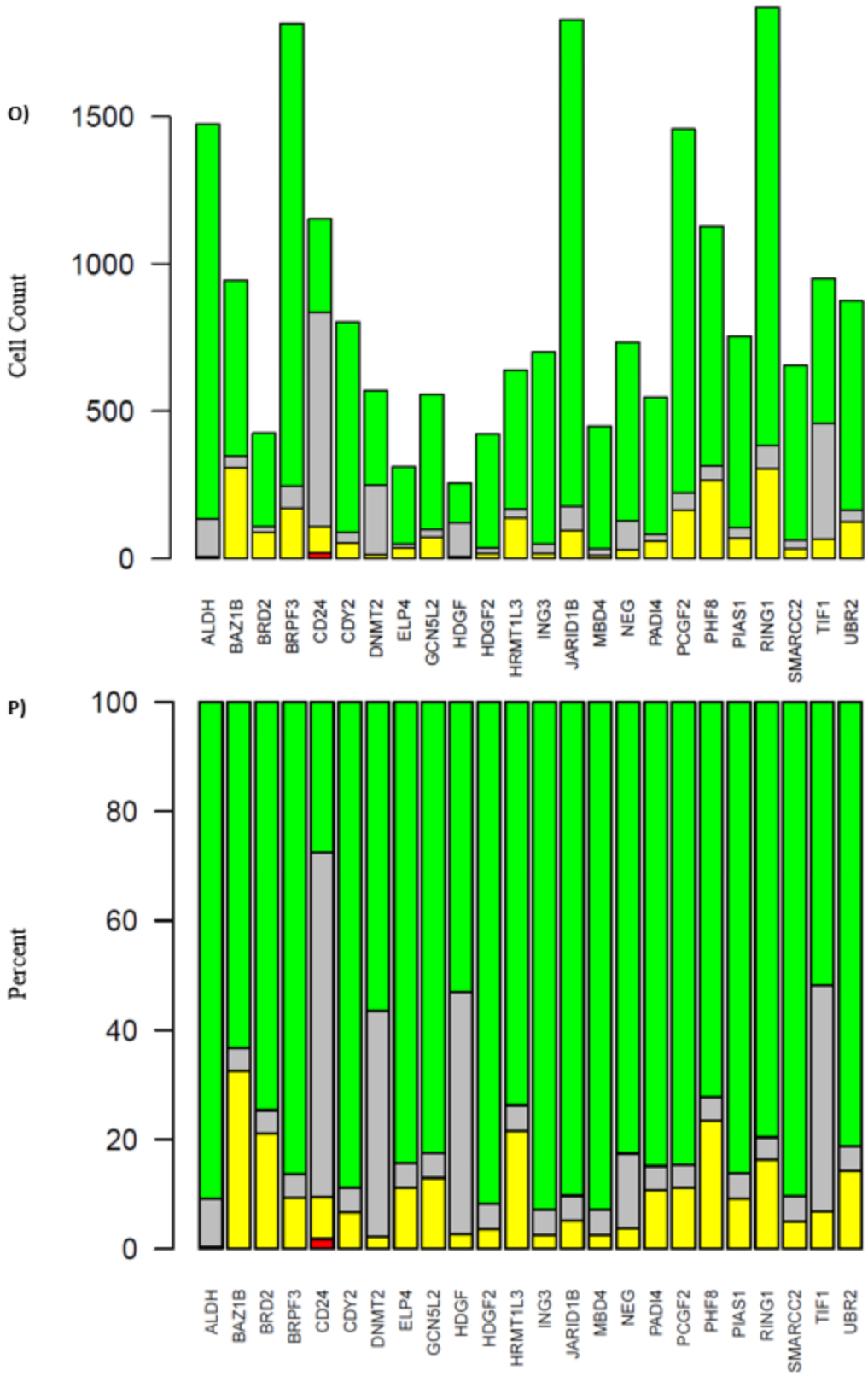

E)
Cell Count
2000
1500
1000
500
0
ALDH
AOF1
AOF2
CARM1
CBX6
CBX7
CD24
CHD2
CTBP1
DMAP1
FLJ10081
FLJ20436
GTF2B
HR
JMJD3
LOC90826
MLLT10
MUM1
NEG
PHF17
PHF19
PHF6
RAG1
RAG2
RAP80
RNF2
RPH3A
SIRT6
SMARCA4
SMNDC1
SMYD1
SMYD2
F)
Percent
100
80
60
40
20
0
ALDH
AOF1
AOF2
CARM1
CBX6
CBX7
CD24
CHD2
CTBP1
DMAP1
FLJ10081
FLJ20436
GTF2B
HR
JMJD3
LOC90826
MLLT10
MUM1
NEG
PHF17
PHF19
PHF6
RAG1
RAG2
RAP80
RNF2
RPH3A
SIRT6
SMARCA4
SMNDC1
SMYD1
SMYD2

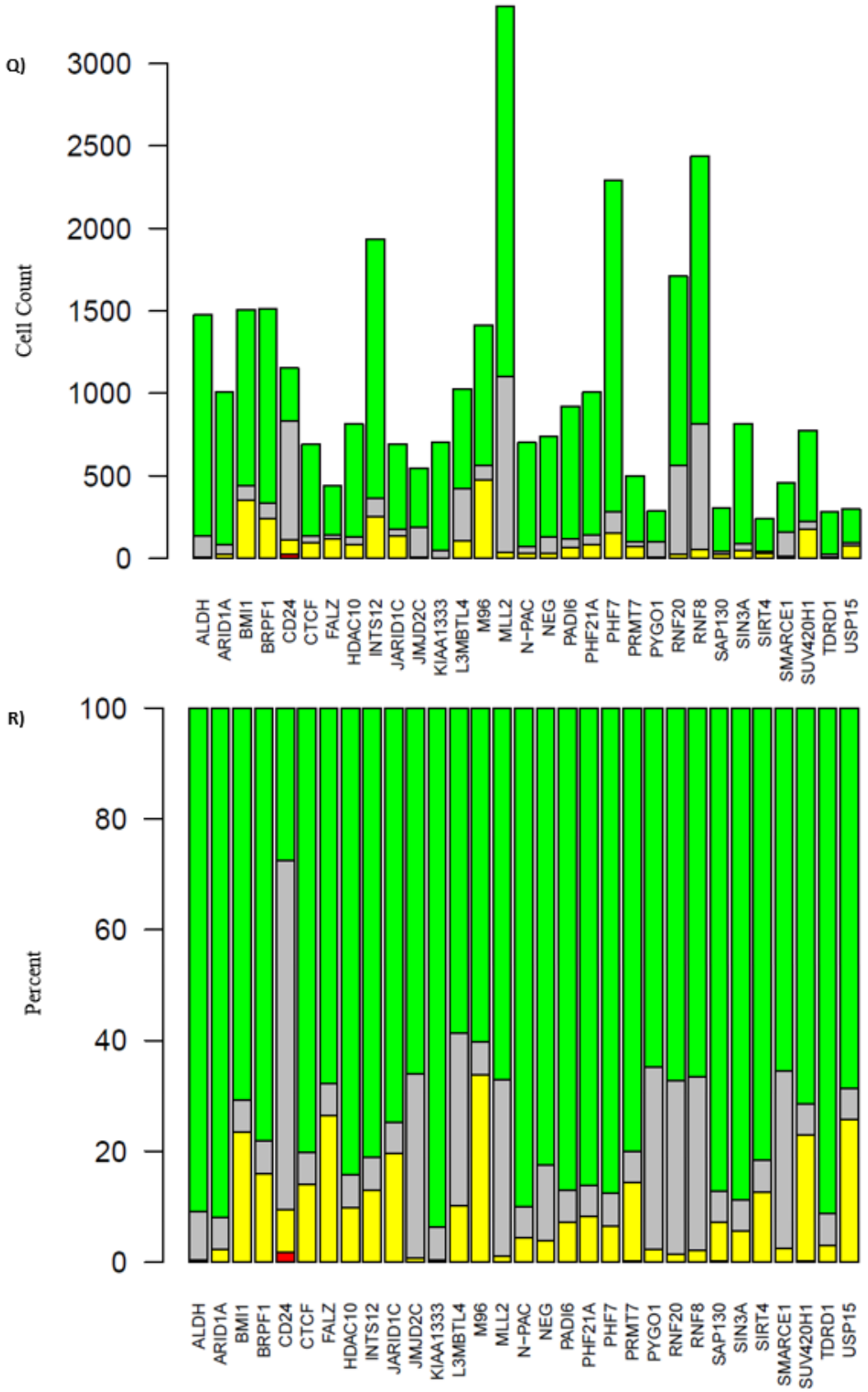

G)
Cell Count
3000
2500
2000
1500
1000
500
0
ALDH
ATF2
CBX3
CD24
CHAF1B
CTCFL
CXORF53
DNMT3L
FBXL11
KIAA1847
MBD5
MLL4
NEG
PADI2
PCAF
PCGF1
PHF10
RNF168
SKB1
SMARCD3
SMYD4
SRCAP
UHRF1
H)
Percent
100
80
60
40
20
0
ALDH
ATF2
CBX3
CD24
CHAF1B
CTCFL
CXORF53
DNMT3L
FBXL11
KIAA1847
MBD5
MLL4
NEG
PADI2
PCAF
PCGF1
PHF10
RNF168
SKB1
SMARCD3
SMYD4
SRCAP
UHRF1


I)
Cell Count
2500
2000
1500
1000
500
0
ALDH
ASH2L
BRD3
BRD8
BRMS1
CD24
CHD9
EPC1
EYA1
HDAC9
HRMT1L1
HRMT1L2
HSPBAP1
ING1
JARID2
JJAZ1
MCRS1
MEN1
MYST1
NEG
POLR2B
PRDM13
PRKCL1
SETDB1
SH2BP1
TAF1
TDRD3
TRIM28
VRK1
ZMYND11
ZNF335
J)
Percent
100
80
60
40
20
0
ALDH
ASH2L
BRD3
BRD8
BRMS1
CD24
CHD9
EPC1
EYA1
HDAC9
HRMT1L1
HRMT1L2
HSPBAP1
ING1
JARID2
JJAZ1
MCRS1
MEN1
MYST1
NEG
POLR2B
PRDM13
PRKCL1
SETDB1
SH2BP1
TAF1
TDRD3
TRIM28
VRK1
ZMYND11
ZNF335

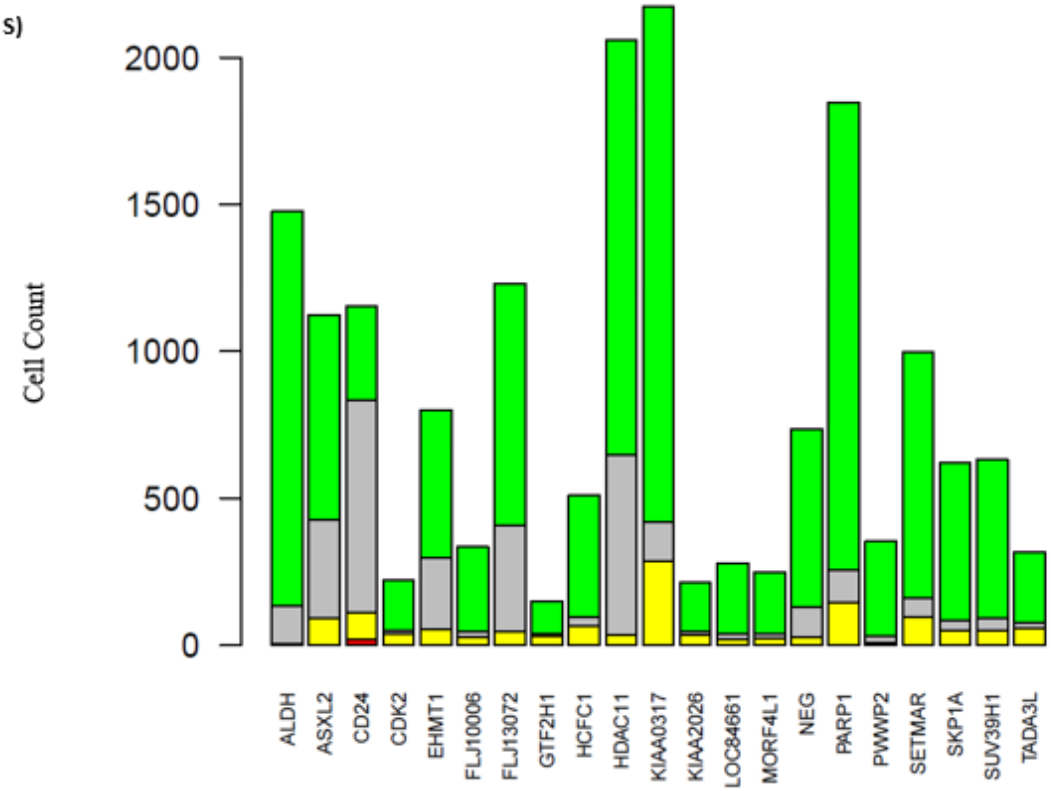

K)
Cell Count
3000
2500
2000
1500
1000
500
0
ACTL6
AIRE
ALDH
CD24
CHD9
CXXC6
EYA1
EZH1
EZH2
FKBP1A
FLJ13984
HBXAP
HDAC8
HDAC9
HRMT1L1
HRMT1L4
JJAZ1
JMJD2D
MCRS1
MEN1
MSL3L1
MYST1
MYST4
NAP1L2
NCOA3
NEG
PARP2
PHF5A
PRDM13
TCF19
ZMYND11
L)
Percent
100
80
60
40
20
0
ACTL6
AIRE
ALDH
CD24
CHD9
CXXC6
EYA1
EZH1
EZH2
FKBP1A
FLJ13984
HBXAP
HDAC8
HDAC9
HRMT1L1
HRMT1L4
JJAZ1
JMJD2D
MCRS1
MEN1
MSL3L1
MYST1
MYST4
NAP1L2
NCOA3
NEG
PARP2
PHF5A
PRDM13
TCF19
ZMYND11


M)
Cell Count
3000
2500
2000
1500
1000
500
0
ALDH
ATR
BRD7
CD24
CHD1
CHD1L
CHD6
FLJ21148
FMR1
LOC339123
MBTD1
NCOR1
NEG
RBBP7
RYBP
SIRT1
SMARCA5
SUV420H2
TAF12
UBE2N
USP22
ZCWPW1
N)
Percent
100
80
60
40
20
0
ALDH
ATR
BRD7
CD24
CHD1
CHD1L
CHD6
FLJ21148
FMR1
_OC339123
MBTD1
NCOR1
NEG
RBBP7
RYBP
SIRT1
SMARCA5
SUV420H2
TAF12
UBE2N
USP22
ZCWPW1

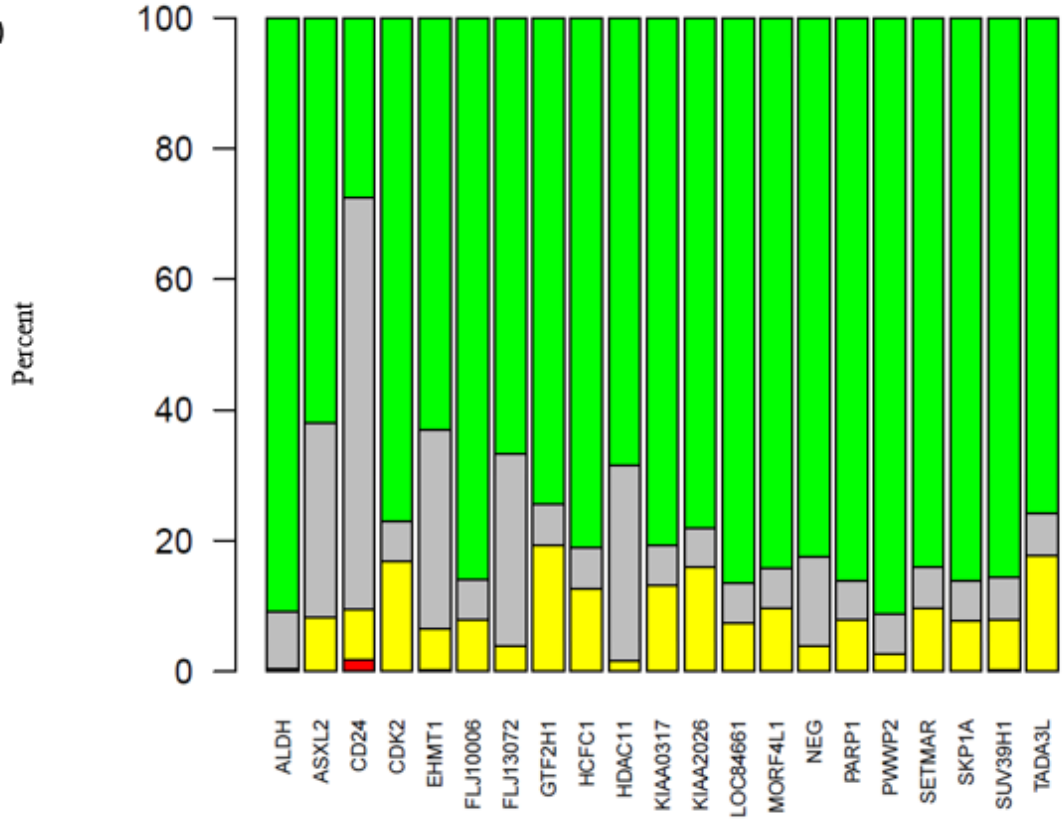

O)
Cell Count
1500
1000
500
0
ALDH
BAZ1B
BRD2
BRPF3
CD24
CDY2
DNMT2
ELP4
GCN5L2
HDGF
HDGF2
HRMT1L3
ING3
JARID1B
MBD4
NEG
PADI4
PCGF2
PHF8
PIAS1
RING1
SMARCC2
TIF1
UBR2
P)
Percent
100
80
60
40
20
0
ALDH
BAZ1B
BRD2
BRPF3
CD24
CDY2
DNMT2
ELP4
GCN5L2
HDGF
HDGF2
HRMT1L3
ING3
JARID1B
MBD4
NEG
PADI4
PCGF2
PHF8
PIAS1
RING1
SMARCC2
TIF1
UBR2


Q)
Cell Count
3000
2500
2000
1500
1000
500
0
ALDH
ARID1A
BMI1
BRPF1
CD24
CTCF
FALZ
HDAC10
INTS12
JARID1C
JMJD2C
KIAA1333
L3MBTL4
M96
MLL2
N-PAC
NEG
PADI6
PHF21A
PHF7
PRMT7
PYGO1
RNF20
RNF8
SAP130
SIN3A
SIRT4
SMARCE1
SUV420H1
TDRD1
USP15
R)
Percent
100
80
60
40
20
0
ALDH
ARID1A
BMI1
BRPF1
CD24
CTCF
FALZ
HDAC10
INTS12
JARID1C
JMJD2C
KIAA1333
L3MBTL4
M96
MLL2
N-PAC
NEG
PADI6
PHF21A
PHF7
PRMT7
PYGO1
RNF20
RNF8
SAP130
SIN3A
SIRT4
SMARCE1
SUV420H1
TDRD1
USP15

S)


Cell Count
2000
1500
1000
500
0
ALDH
ASXL2
CD24
CDK2
EHMT1
FLJ10006
FLJ13072
GTF2H1
HCFC1
HDAC11
KIAA0317
KIAA2026
LOC84661
MORF4L1
NEG
PARP1
PWWP2
SETMAR
SKP1A
SUV39H1
TADA3L

T)


Percent
100
80
60
40
20
0
ALDH
ASXL2
CD24
CDK2
EHMT1
FLJ10006
FLJ13072
GTF2H1
HCFC1
HDAC11
KIAA0317
KIAA2026
LOC84661
MORF4L1
NEG
PARP1
PWWP2
SETMAR
SKP1A
SUV39H1
TADA3L

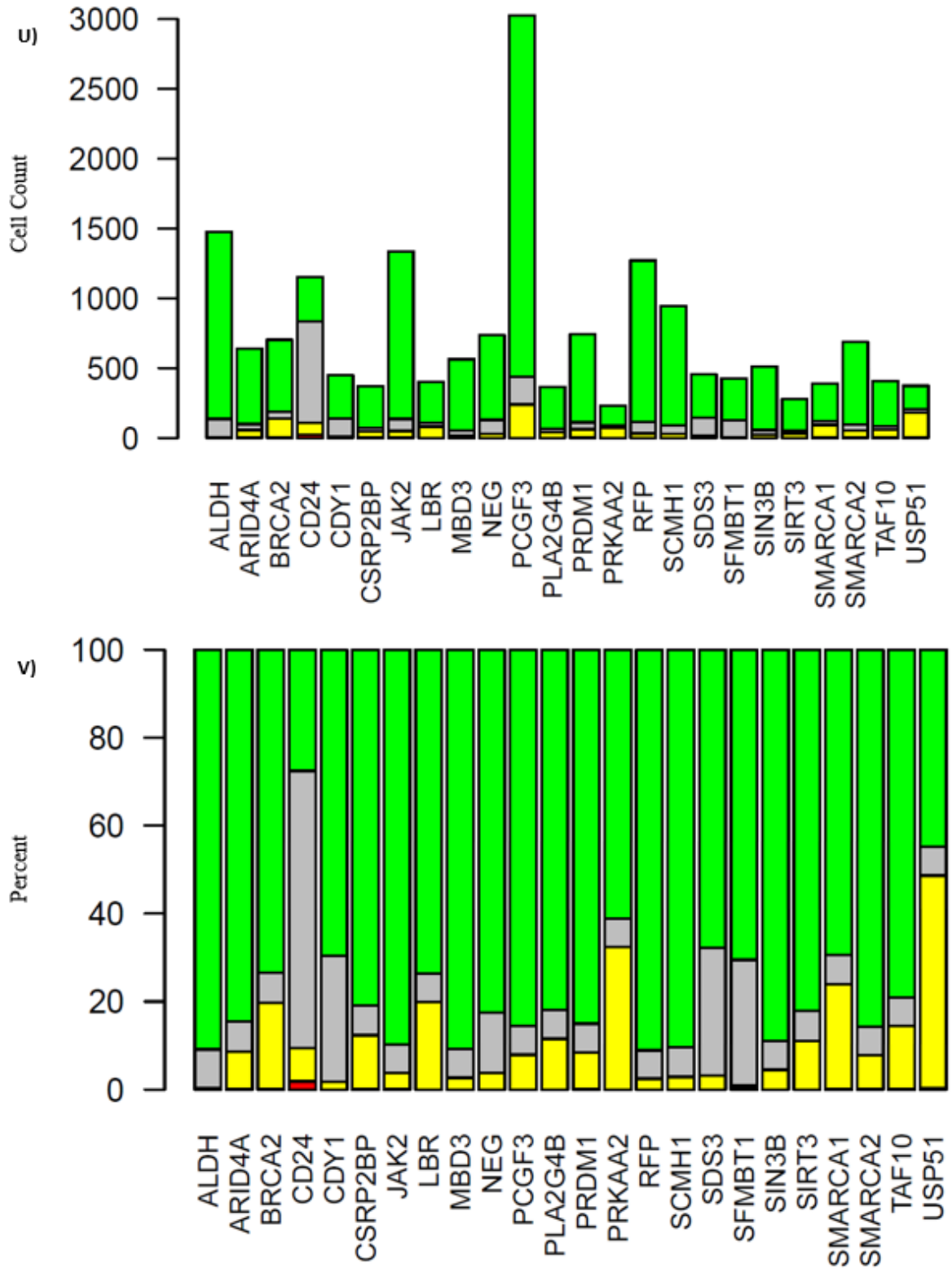

u)
Cell Count
3000
2500
2000
1500
1000
500
0
ALDH
ARID4A
BRCA2
CD24
CDY1
CSRP2BP
JAK2
LBR
MBD3
NEG
PCGF3
PLA2G4B
PRDM1
PRKAA2
RFP
SCMH1
SDS3
SFMBT1
SIN3B
SIRT3
SMARCA1
SMARCA2
TAF10
USP51
v)
Percent
100
80
60
40
20
0
ALDH
ARID4A
BRCA2
CD24
CDY1
CSRP2BP
JAK2
LBR
MBD3
NEG
PCGF3
PLA2G4B
PRDM1
PRKAA2
RFP
SCMH1
SDS3
SFMBT1
SIN3B
SIRT3
SMARCA1
SMARCA2
TAF10
USP51

Figure 2.9 mCherry-EGFP- Hits as Determined by Selection Criteria.

Panels A-V depict the hits selected according to selection protocols (Student's t-test, double tailed, $p < 0.05$). Each hit is below the threshold compared with "NEG" groups. "NEG", "ALDH", and "CD24" are plate controls included in each screen. The first graph depicts total cell count, modified by those well's viability scores, and the second depicts percentage. Gray cells depict mCherry-EGFP-, yellow cells are mCherry+EGFP+, and Green cells are mCherry-EGFP+.

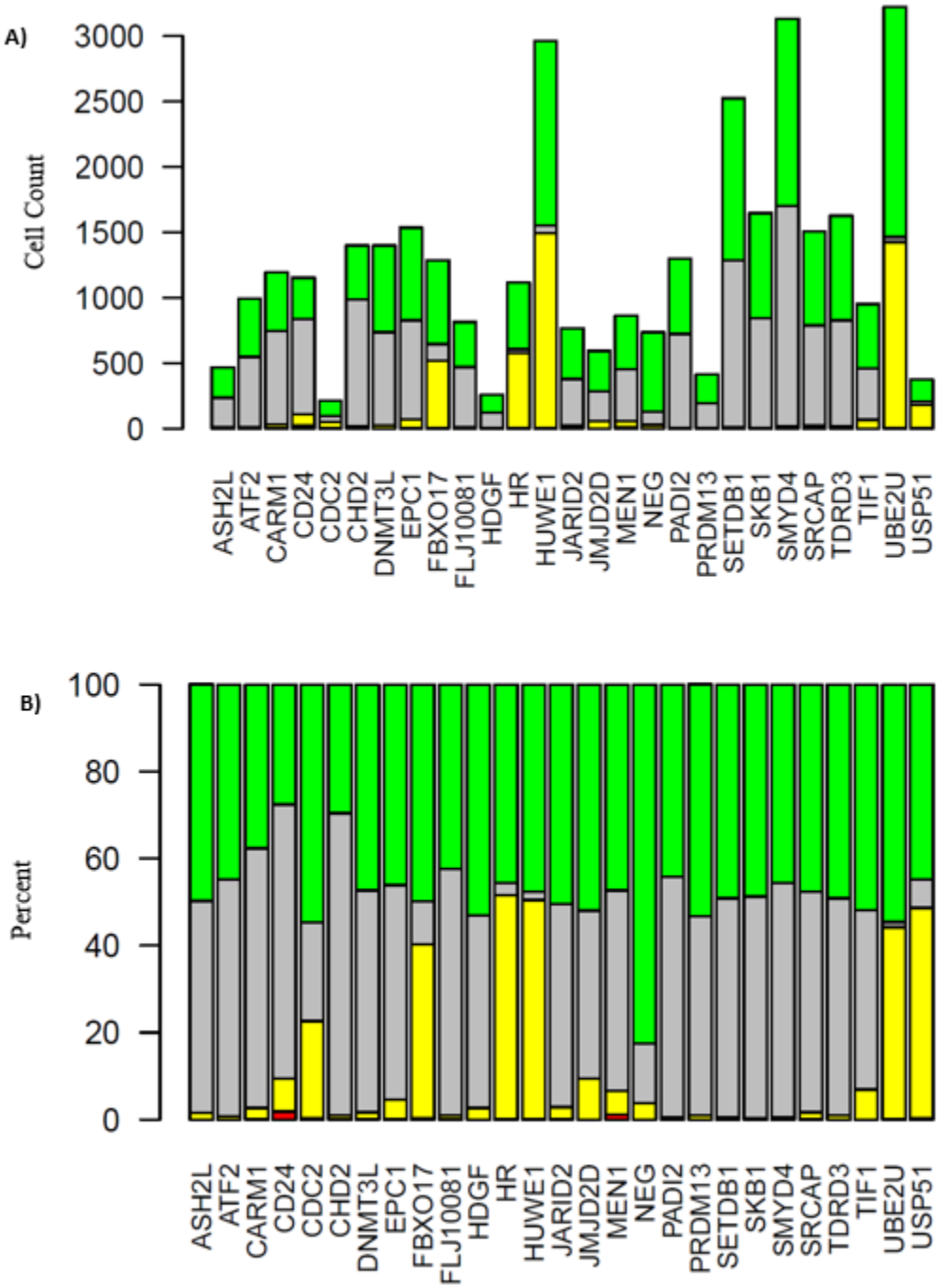

A)
Cell Count
3000
2500
2000
1500
1000
500
0
ASH2L
ATF2
CARM1
CD24
CDC2
CHD2
DNMT3L
EPC1
FBXO17
FLJ10081
HDGF
HR
HUWE1
JARID2
JMJD2D
MEN1
NEG
PADI2
PRDM13
SETDB1
SKB1
SMYD4
SRCAP
TDRD3
TIF1
UBE2U
USP51
B)
Percent
100
80
60
40
20
0
ASH2L
ATF2
CARM1
CD24
CDC2
CHD2
DNMT3L
EPC1
FBXO17
FLJ10081
HDGF
HR
HUWE1
JARID2
JMJD2D
MEN1
NEG
PADI2
PRDM13
SETDB1
SKB1
SMYD4
SRCAP
TDRD3
TIF1
UBE2U
USP51

Figure 2.10 mCherry-EGFP+ Hits as Determined by Selection Criteria.

Panels A-B depict the hits selected according to selection protocols (Student's t-test, double-tailed, $p < 0.05$). Each hit is below the threshold compared with "NEG" groups. "NEG", "ALDH", and "CD24" are plate controls included in each screen. The first graph depicts total cell count, modified by those well's viability scores, and the second depicts percentage. Gray cells depict mCherry-EGFP-, yellow cells are mCherry+EGFP+, and Green cells are mCherry-EGFP+.

Chapter 3

Assaying Top 120 Hits by Single siRNA Screening and Selection of TRIM24 and EPC2 for Further Testing

Abstract

Any high-throughput primary screen, regardless of method and targets, requires additional confirmation. Any data generated ought to be examined with suspicion and the knowledge that these most likely are false positives or have interfered with the assay output. Typically, a dose-response screening is conducted to eliminate these false-positives; however, siRNA screening requires a different approach. As seed-sequences can lead to consistent off-target effects in pooled siRNA, any siRNA must be separated and tested individually. Furthermore, if there are any additional safety features, those should be selected to ensure the highest quality data. From a primary screen conducted with siGenome pooled siRNA, 120 initial gene target hits were selected for additional follow-up. These gene targets had several connections to one another, whether that be chromatin modification, polycomb identity, or convergent histone modification. These 120 hits were tested in the same manner as the primary screen, and this data was used to confirm consistent. This analysis generated eighteen confirmed gene targets across selected outputs from the reporter cell line ($EGFP^+$ $mCherry^+$, $EGFP^+$ $mCherry^-$, $EGFP^-$ $mCherry^-$). After additional screening in the reporter cell line and other non-inflammatory triple-negative breast cancer cell lines (SUM149, MDA-MB-468, and BT-20), two proteins were confirmed for future investigation: E3 ubiquitin ligase TRIM24 and the NuA4 acetyltransferase protein EPC2. Both of these proteins were selected for their ability to affect $EGFP^+$ $mCherry^+$ (or $ALDH^+$ $CD44^+$ $CD24^+$) cells in distinct ways: TRIM24 decreased while EPC2 increased the proportion of these cells. In this way, these proteins can be further investigated as potential antagonists in subpopulation establishment and

maintenance. Additionally, dTRIM24, a TRIM24 degrader, was partially able to offset Docetaxel-induced expansion of undifferentiated cells. This result is critical since chemotherapies used in TNBC treatment increase this undifferentiated population of cells, potentially leading to the grim prognosis of this most-deadly of breast cancer subtypes. These results highlights a key translational avenue for patients in need.

Introduction

Although the primary goal of high throughput screens is to identify relevant hits for a given target or phenotype, these initial hits must be confirmed in subsequent assays (1). Thus, screening campaigns follow a general pattern: a primary screen, a secondary screen, and an orthogonal screen (2). The primary screen sifts through a small or large library of potential hits, whether compounds or interfering constructs, and determines which structures or targets are most likely worth investigating further. The secondary screen further sorts through these potential hits and removes potentials that might interfere in some way. In assays dependent upon some form of light, precipitation, quenching, autofluorescence, or general compound degradation can interfere with the assay itself; for these reasons, suspected hits are further tested in a dose-dependent manner, which can highlight which compounds are likely true hits and which are due to assay interference (3-5).

How hits are verified depends on the nature of the assay and the interfering or activating substances screened. In siRNA screening, there are six general stages to confirmation: 1) secondary screening consisting of multiple separate individual siRNA of higher quality than the primary screen 2) Confirm top rated hits by western blotting and by checking mRNA levels of targeted protein by qPCR 3) Rescue experiments using siRNA-resistant DNA delivered by lentiviral transduction 4) assay confirmation by orthogonal methods 5) and in other relevant cell lines (3, 4, 6).

Secondary Screening Protocols

With siRNA screening assays, there is little point to dose-dependence as it is assumed the initial siRNA levels are sufficient to affect a target. Thus, secondary assays involve using single siRNA strands rather than pooled siRNA that is often used in primary screens (4, 7). Rather than determining dose-dependency, the goal in these secondary screens is to determine a consistent

outcome with these different siRNA strands. For instance, 2/4 siRNA that repeat the initial primary screen result would be sufficient for labeling that a hit; any more siRNA that repeat further increase confidence in gene target relevance. Lastly, screening campaigns often involve further assays that require a different output. For instance, a primary assay using fluorescent polarization would be further confirmed by an assay that uses no light-based signals of activity, such as calorimetry, Octet Red, or crystal structures for especially promising candidates (8-11). For siRNA screening, however, confirmation follows a different path.

Secondary screening for siRNA HTS campaigns are more about determining consistency rather visualizing any issues with interfering compounds (4). siRNA strands, even those that have the same gene target, possess distinct features and shapes (12). These double-stranded siRNA are split by Dicer then are integrated within the RNA-induced silencing complex (RISC), where the target mRNA is degraded by nuclease activity (13, 14). However, these different strands have the potential for off-target effects, and these off-target effects may be the reason for phenotypic changes observed within the assay (3). Therefore, single strands of siRNA are used in conformational secondary screening when most of the candidates are excluded (5). Using these unique strands in individual wells allows researchers to observe which strands produce the same effect versus those that have no effect or a completely different effect compared to the primary screening. In this present assay, the same output was used in secondary screening as in primary.

Secondary Screening Verification

While secondary screening is critical for confirming initial hits, additional screening is necessary to confirm the results discovered, particularly due to the limited cell numbers that can be tested in a 96-well plate. To that end, top confirmed hits should be tested in larger cell numbers with the same individual siRNA strands used in the secondary screen to exclude any hits that were mistakenly identified due to sampling error or unforeseen assay interference (3, 5). After this subsequent confirmation, pooled siRNA can be used since the final effect will be produced by the most effective individual siRNA in the mixture (15, 16). However, individual siRNA will lead to no loss in quality; both approaches are acceptable.

At this stage in confirmation, siRNA is assumed to have decreased protein levels. However, protein level confirmation must be verified by western blotting techniques (17, 18). This verification can also assist in selecting a higher or lower siRNA dose depending on the knockdown efficacy at the

given dose and the mode of transfection, whether forward or reverse. Additionally, mRNA levels may need to be verified by PCR. Not all proteins decrease over this time scale and may be long lived, but the mRNA may be depleted by the siRNA (19, 20). In these rarer cases, western blot analysis may not accurately represent a successful knockdown (17).

Still Further Verification

However, even if multiple siRNA strands against a given gene target each have the same effect, that does not necessarily mean the phenotypic response is due to that protein target's reduced levels. Therefore, it is necessary to confirm that the phenotype correlates with decreased protein levels with western blotting as well as confirming reduced mRNA levels with qPCR. These confirmation experiments should be performed with any future experiments as an added confirmation of the protein's effect.

Confirmational cell line selection also requires consideration. Selected TNBC cell lines should be close in transcriptional expression without being too dissimilar. The critical CD44 and CD24 populations must be considered in addition to transcriptional identity. $CD44^{+}CD24^{+}$ and $CD44^{+}CD24^{-}$ cells should correspond to epithelial and mesenchymal cells, respectively. As was done in SUM149, expression of E-cadherin and other markers of the epithelial lineage should be high in $CD44^{+}CD24^{+}$ cells, and the expression of mesenchymal markers such as vimentin should be high in $CD44^{+}CD24^{-}$ cells.

However, endpoint verification is only one critical aspect of a campaign. There must be investigation into how these endpoints arise. There are several possibilities, 1) such as cell-state transitions, 2) cell-state arrest or increased replication, 3) selective sub-population death, and 4) sub-population death and increased replication of other sub-populations. Different methods can be used to decipher the processes involved that produce the end-result, such as apoptotic assays to determine cell-death at a fixed timepoint, MTS to determine the same over time, and cell-cycle analysis using DNA-staining reagents. So often are these experiments performed that assay kits are commercially available. However, the two-color reporter cells can be used in a way that allow us to generate all this data in one experiment through live-cell imaging.

To conduct these experiments, the cells can undergo siRNA interference and images are captured by a live-cell microscope every thirty minutes over seventy-two hours. This permits direct visualization of processes that lead to the endpoints observed via FACS analysis, whether that be

cell-transitions, cell-stasis, or cell death, or a combination of all three processes. This is due to expression (or lack) of mCherry and EGFP, individually or in combination, that permit the reconstruction and quantification of the stem cell hierarchy (21). Additionally, these results can be quantified using programs such as FIJI to gain a better idea of what processes are occurring rather than qualitative speculation on the processes involved.

Towards Translation

Docetaxel (trade name Taxotere®) is a derivative of taxol, which is derived from the Pacific yew tree and has been used for decades to treat numerous cancers, such as lung, ovarian, prostate, and importantly, late stage, metastatic breast cancer (22-26). It is one of the many taxol-related drugs given to late stage breast cancer patients. However, like many other chemotherapies in late stage breast cancer patients, it fails to completely eliminate the tumor (27). Additional research has found that docetaxel does kill most differentiated cells but increases the proportion of stem-like and progenitor-like cell populations (28-31). This process ensures that the cancer will not be completely eliminated and instead may recur in another metastatic site (31). However, efforts have continued to attempt to block or reverse this docetaxel-induced de-differentiation (30, 32-36). According to unpublished data, this increase in de-differentiated cells can occur in fewer than seventy-two hours, implying an epigenetic mechanism for cell alterations. Since some of the epigenetic hits have the ability to decrease progenitor cell populations, it may be of some interest to investigate whether any of those proteins may be able to prevent a de-differentiation event. While siRNA can be used, degraders can accomplish the same function over the same time span (37, 38).

Overall, I successfully conducted the necessary secondary screen with individual siRNA available from Dharmacon. Due to the flexibility of the R-program, it was possible to compare and contrast all data generated, from the primary screen to the secondary screen, and determine which gene targets merit future follow-up studies. For those experiments, I selected two proteins, TRIM24 and EPC2. At first, only repeated knockdown studies are required in the reporter-cells, but antibody and the Aldefluor™ assay must be conducted to confirm the selected effects in SUM149 but also BT20 and MDA-MB-468. I additionally tested for a TRIM24 degrader, dTRIM24, to attenuate docetaxel-induced expansion of stem-like cells.

Results

Well-optimized primary screens are very effective at identifying hits, but those hits, regardless of identity, must be confirmed by additional screening (39). The next stage in a high-throughput screening campaigns are known as secondary screening. This screening often involves dosing compounds in a range of concentrations to ascertain if the effect corresponds with the level of potential hit added. This process is irrelevant for confirming hits with siRNA dosed at lower levels (30 nM and below) as long as the siRNA concentration can produce an effect (18).

Single siRNA Further Refines List of Hits

The next stage of the screening campaign required both new siRNA, improved strand quality from the siGENOME line of siRNA, as well as testing each individual strand separately rather than pooled. From the overall hit pool, 120 gene targets were selected for secondary screening using 480 unique siRNA strands (Data tables 3.1-3.4). The siRNA selected was Dharmacon's ON TARGETplus single siRNA stock plates. From these stock plates the siRNA would be delivered to the test wells. Each stock plate was tested twice to ascertain if this effect is repeatable at the small scale and afforded increased opportunity to exclude false-positive hits. A successful hit for the secondary screening was defined as 1) Effecting significant decrease in the selected sub-population ($P < 0.05$) 2) 2/4 individual siRNA must produce this effect 3) This effect must be repeatable in multiple experiments. These standards follow standard practice for siRNA secondary screening.

Of the 480 siRNA strands (against 120 gene targets) 108 of those were hits via secondary screening (22.5%) (Figures 3.1-3.4). However, the hit criteria extend beyond being hits in the secondary screen. Any hit defined by this screening campaign must 1) Have 2/4 secondary screening siRNA be hits 2) These hits must be for the same category 3) these hits must reduce the population in question in both primary and secondary screens ($p < 0.05$). Another R script was produced that applied these criteria to the whole of the screening data, which produced 58 gene target hits (12.1% hit rate; 22 hits for viability, 5 for $mCherry^{+}EGFP^{+}$ cells, 25 for $mCherry^{-}EGFP^{-}$, and 6 for $mCherry^{-}EGFP^{+}$) (Figures 3.5-3.7, Panels A and B; Data Table 3.5, Panels A-D). These figures represent the top hits of each category by fold change. Following the trend established in primary screening, $mCherry^{-}EGFP^{-}$ cells have the highest number of hits compared to the other sub-

populations. Thus may imply it is difficult to decrease epithelial-like cells in a majority epithelial cell line. Additionally, there was a noticeable overlap between viability and mCherry^{-}EGFP^{-} sub-population hits with 15 shared between those two outputs (Data Table 3.5, Panel E). It may be in these instances that those gene targets decrease all cells to some degree but impact the double-negative cells most significantly, but further mechanism-based studies are needed to identify a specific process.

Integrated Data Analysis Identifies Top Hits in Each Sub-Population

The exception to this integrated data process were mCherry^{+}EGFP^{+} hits. While both mCherry^{-}EGFP^{-} and mCherry^{-}EGFP^{+} sub-populations both produced an appreciable number of initial and consistent secondary hits, the exception was mCherry^{+}EGFP^{+}. The same procedures that worked for the two differentiated populations returned zero hits that were consistent through two phases of the screening campaign. However, there were high quality hits that appeared in the secondary screen, and while they did not meet pre-established conditions, they would be further investigated to see if they confirm. One hit, TRIM24, stood out for its effect in primary screening.

Overall, 40 siRNA gene targets were found to have consistency across both primary and secondary screening, but not all these 40 targets had several siRNA equally effective against reducing the given subpopulation. As in the primary screen, the mCherry^{-}EGFP^{-} subpopulation had the highest potential hits at 23. Of these 23, 10 confirmed according to the previously detailed hit requirements (CTBP1, SMYD2, GTF2B, SIRT6, HUWE1, NSD2, HR, BRMS1, DCAF1, and EPC2) (Figure 3.6, Panels A and B). Next highest was mCherry^{-}EGFP^{+} at 12 potential hits, but only 5 were high-confidence (MEN1, EPC1, ATF2, ASH2L, ACTL6B) (Figure 3.7, Panels A and B). Unsurprisingly the lowest, the double positive population, mCherry^{+}EGFP^{+}, had the fewest potential hits at 5 and only one was high confidence (TRIM24) (Figure 3.5, Panels A and B).

While these results are encouraging, additional studies are required in cells beyond the reporter cell line. First, these gene targets should be followed up with the reporter cells but at a larger scale. Second, these targets should be tested in SUM149 using traditional methods (CD44 and CD24 antibody staining with FACS analysis plus the ALDEFLUOR™ assay). Several experiments were undertaken to test these targets for their ability to function in larger cell numbers (Figure 3.8, Panels A-C). While several targets recapitulate their effects in larger numbers of both reporter cells and in parental SUM149, TRIM24 and EPC2 stood out for their ability to affect the same

population of cells but in opposite ways: TRIM24 knockdown increased $mCherry^{-}EGFP^{-}$ cells and decreased $mCherry^{+}EGFP^{+}$ while EPC2 knockdown decreased $mCherry^{-}EGFP^{-}$ cells and increased $mCherry^{+}EGFP^{+}$ cells (Figures 3.9 Panel A).

Further Investigations with TRIM24 and EPC2

TRIM24 is implicated in both pro-cancerous roles by negatively regulating p53 and in its role in transcriptional activation, but it also behaves as a tumor suppressor by negatively regulating retinoic acid signaling (40-43). Thus, its activity is context-dependent and has different roles in different systems, and the heterogeneity of genetic and epigenetic dysregulation determines which role it takes on; however, the abundance of data indicates TRIM24 often takes on a pro-cancerous role, especially in breast cancer (44, 45). Owing to its ability to both decrease Progenitor-like cells ($mCherry^{+}EGFP^{+}$) yet increase mesenchymal cells ($mCherry^{-}EGFP^{-}$), I selected this protein for further analysis.

Conversely, the lysine-histone acetyltransferase associated EPC2 was selected for its ability to increase epithelial-like cells ($mCherry^{-}EGFP^{+}$), and epithelial-like progenitor cells ($mCherry^{+}EGFP^{+}$), making it the "anti-TRIM24" (46, 47).

In studies featuring SUM149, TRIM24 knockdown recapitulated its effect in SUM149 cells by the ALDEFLUOR™ assay but did not increase the $CD44^{+}/CD24^{-}$ population as in the reporter cells (Figure 3.9, Panel D and E). Additionally, it had similar effects in BT-20 and MDA-MB-468 but not in MDA-MB-231 (Figure 3.9 Panel F and G). Overall, TRIM24 knockdown appears to have similar effects in multiple cell lines, but not within the context of $CD44^{+}/24^{+}$ cells. This may be due to a specific clonal effect between SUM149 and its derived reporter cell line.

Conversely, EPC2 failed to increase ALDEFLUOR™ positive cells but did increase the $CD44^{+}/CD24^{+}$ percentage of cells in SUM149, an event that occurred in the reporter cell line as well (Figure 3.9, Panels A and B). Additionally, I attempted to select only $CD44^{+}/CD24^{+}$ cells and test for any difference in aldehyde dehydrogenase activity, but that showed no difference between control groups and EPC2 knockdown due to increased standard deviation from fewer cells in each group. Future studies will be required for further corroboration among other TNBC cell lines such as BT-20 and MDA-MB-468.

TRIM24 Degrader, dTRIM24, Attenuates Docetaxel-Induced Expansion of Undifferentiated Cells

The primary treatment for triple-negative breast cancer patients is chemotherapy, especially taxanes such as taxol, paclitaxel, docetaxel, and other variations with different formulations (22, 23, 48, 49). The goal with these drugs is to slow or stop cell expansion of the tumor by disabling mitotic spindle assembly (50). Despite initial positive responses to such therapy, ultimately most patients develop resistance, relapse, and subsequently die. Additionally, resistant populations arise or can predominate as treatment continues (51). Due to this continuing issue, researchers have investigated the reasons underlying this poor prognosis. They've discovered that docetaxel and other such taxanes expand stem-like and undifferentiated cell populations despite killing most differentiated cells (29, 52, 53). Various efforts have been undertaken to counteract this effect, but none have made a clinical impact (27, 30, 33, 34, 54).

One of the selected protein targets in my studies, TRIM24, significantly and consistently decreased de-differented cells in TNBC cell lines. To investigate if this effect may be relevant to TNBC treatment, I transfected reporter cells with Cell Signalling non-targeting negative siRNA (40 nM), TRIM24 siRNA (20 nM), KMT2D siRNA (20 nM), and TRIM24 plus KMT2D siRNA (40 nM total) and analyzed the cells after 72 hours had elapsed. While both TRIM24 and KMT2D had their expected effects (Figure 3.10, Panel A and B) (TRIM24 decreased epithelial progenitor-like cells, KMT2D dramatically increased that same population), the combination siRNA managed to keep that population relatively the same as negative siRNA while also maintaining the cytotoxicity previously observed with KMT2D knockdown (Figure 3.10 Panel C). These results gave further indication that TRIM24 may be a relevant target for further translational research.

Due to TRIM24 knockdown decreasing this very population, I investigated whether there were any chemical probes that might offset docetaxel's increase in these populations. One such compound is a degrader (dTRIM24) that conjugates TRIM24's non-functional bromodomain to the ubiquitin ligase VHL (38). The developing group further characterized dTRIM24 to have a maximal effect at 5 μM that wanes after 72 hours and requires redosing (38). It additionally disturbs TRIM24's ability to function as a transcriptional coactivator or repressor (38). Therefore, I conducted preliminary studies to investigate how this degrader might disrupt docetaxel's unfortunate consequences.

After treating the reporter cells for 72 hours with DMSO, dTRIM24 (5 μM), Docetaxel (5 nM), or dTRIM24 (5 μM) and Docetaxel (5 nM), I discovered that dTRIM24 significantly attenuates de-differentiation induced by Docetaxel (Figure 3.11, Panels A and B). This effect occurred in both stem-like and epithelial progenitor-like cells. Additionally, while there was no significant cytotoxic effect with dTRIM24 alone, dTRIM24 and Docetaxel in combination maintained Docetaxel's cytoxicity while significantly decreasing the expansion of progenitor and stem-like cells (Figure 3.11 Panel C). While encouraging, these very preliminary results deserve more experiments to verify the results as well as investigate optimal use of this relative new degrader.

Discussion

High-throughput screening campaigns are primarily focused on discovering hits against a given target, but equally as important is confirming those initial hits. Therefore, secondary screens are critical in differentiating between false and true positive hits. These secondary screens should be similar to the primary screen, save for a few details. In most compound screening campaigns, the secondary screen will feature varying levels of the compound tested against the protein target in an attempt to verify if the effect occurs in a dose-dependent manner. However, siRNA screening is instead focused on how repeatable the effect is with variations on the core siRNA strands. Thus, multiple siRNA are used independently of one another. Furthermore, usually higher quality siRNA are used for these experiments, which will further exclude false positives from the list of hits.

For present purposes, six Dharmacon's OnTarget Plus single siRNA plates were ordered that contained the top 120 gene targets. These were selected from the top hits by viability, $EGFP^{+}mCherry^{+}$ ("yellow"), $EGFP^{+}mCherry^{-}$ ("green"), and $EGFP^{-}mCherry^{-}$ ("grey"). Each of these groupings correspond to $CD24^{+}ALDH^{+}$, $CD24^{+}ALDH^{-}$, and $CD24^{-}ALDH^{-}$, respectively in SUM149, MDA-MB-468, and BT20. The secondary screen was conducted much in the same manner as the primary screen, save for an increase in the numbers of cells seeded due to the decreased cell growth of epithelial-majority cell lines. The raw FCS files were analyzed on FCS Express (version 7), and that data was further processed and prepared for use in R-programs, first to analyze the top hits and secondly to determine which hits demonstrated the same effect across all campaign screenings conducted. This is a critical step as this determines whether gene targets

have the same effect merely in the secondary screen or across multiple steps. Varying levels of success were found across the screening efforts.

Excepting a few cases, it was very difficult for any protein knockdown to consistently decrease less-differentiated cells. Only TRIM24, ACT6B, and ATF2 were found to consistently perform that function in the reporter cell line (Figure 3.8 Panel A-C; Figure 3.9 Panel A). Examples of $EGFP^+$ cells decreased were scarce with only six targets found to decrease this population (Data Table 3.5, Panel D). This may imply there was an unforeseen difficulty of decreasing the presence of epithelial-like cells in a majority epithelial cell line, which contradicts the assay-window hypothesis, that is, the higher the number of cells, the more likelihood of decreasing that population. Furthermore, there were far more examples of $EGFP^-$ hits despite that mesenchymal-like cells being one of the smaller populations (10-20% in SUM149). It may be that these cell populations are epigenetically poised to be epithelial cells, and thus examples of epigenetic gene proteins are far more likely to tip the balance further into primarily epithelial cells and reduce the $EGFP^-$ cell percentage. Overall, this trend matches what was observed in the primary screen.

Despite difficulty for some populations to undergo some effect, no protein managed to ablate any one population. Only reduction or cell shifting was found when knocking-down these epigenetic proteins. This incomplete effect is not unexpected for epigenetic proteins. While there may not be a single master regulator across multiple TNBC cell lines, it is likely that certain proteins will have a similar effect. The final hit selection will depend on how relevant the protein target is across several TNBC cell lines, and thus, a higher likelihood of future translational experiments. Two such proteins were found to be largely consistent: TRIM24 and EPC2.

TRIM24

As a member of the tripartite motif family, TRIM24 (TIF1-α) is a ubiquitin ligase and acts as a transcriptional coactivator dependent primarily on histone H3 modifications, namely H3K23ac (44, 55-63). It possesses three binding domains: one non-functional PHD-bromodomain, two B-box domains, and one RING-type zinc finger (44, 64-66). TRIM24 is correlated with negative prognosis in cancers such as breast, non-small cell lung, gliomas, head and neck squamous carcinoma, hepatocellular carcinoma, gastric prostate, ovarian, and cervical (44, 45, 63, 67-79). These negative correlations likely are due to TRIM24's key functions: p53 negative regulation,

nuclear receptor interaction, expression of pro-cancer immune signals, and role in Wnt/GSK3β/β-catenin (40, 43, 69, 76, 77, 80, 81).

TRIM24 behaves as an oncogenic factor by targeting p53 for degradation in human breast cancer cells and in general is a negative regulator of p53 function (40). Deactivation, degradation, or mutation of p53 leads to a higher likelihood of further cancer genetic mutation. Accordingly, TRIM24 is subject to autodegradation during the DNA damage response; however, TRIM28, another member of the tripartite motif family, can prevent TRIM24 degradation (82, 83).

TRIM24 can interact with several nuclear receptors. It interacts with ER and AR and is involved in detrimental hormone signaling in both prostate and breast cancer (43, 84-89). It can also interact with the retinoic acid receptor, which has been shown to influence cancer stem cells (41, 42, 90-92). Furthermore, TRIM24 increases the effect of STAT3 signaling, which is implicated in worse prognosis of some cancers (69). This is especially of interest owing to SUM149 being an inflammatory breast cancer (93).

From the data available, there is a clear association with de-differentiated cells, especially evident with its negative role in many cancer types. Pathways such as retinoic acid signaling, STAT, WNT, and its role negatively regulating p53 all likely play a role in the phenotype measured by screening and by follow-up studies in multiple cell lines.

Indeed, TRIM24 did appear to have a similar role in other cell-lines, such as MDA-MB-468 and BT-20, specifically with $ALDH^+$ cells (Figure 3.9 Panel F and G). These follow-up results build a strong case that TRIM24 has consistent effect in some triple-negative breast cancers. Testing other in other TNBC cell lines is critical since the origin cell line for the reporter cell is SUM149, which is an inflammatory breast cancer cell line (94). Inflammatory breast cancers are especially invasive and deadly, having a worse 5-year relative survival than TNBC (95). Thus, the effect should be tested in other TNBC cell lines that aren't as aggressive, such as BT-20 and MDA-MB-468. Positive results in this cell line indicate a common effect in TNBC.

However, it did not have an effect in MDA-MB-126 (Figure 3.9 Panel F and G). This likely indicates another issue with triple-negative breast cancers: not intratumoral heterogeneity but intertumoral heterogeneity. Since TNBC is diagnosed by process of elimination, there are likely highly dissimilar triple-negative breast cancers lumped together. It is therefore unlikely that a single target (or multiple targets) will be applicable across all breast cancers. While there is a

method to intrinsically diagnosis breast cancers as basal, that method does align with 76% of cancers diagnosed by ER, AR, and HER2 staining methods (96-99).

dTRIM24 and Docetaxel

In the era of targeted therapies ushered in by Gleevec and Tamoxifen, researchers must always have an eye towards translational aspects of their work. While there may never be one single target across the heterogenous landscape of triple-negative breast cancers, there will be many individual targets that are relevant to select populations of patients. TRIM24 knockdown and degradation result in the same effect, a decrease of undifferentiated cells, TRIM24 may be one of those targets. While undifferentiated cells are only a portion of the intratumoral heterogeneity, they are a critical factor in why current chemotherapies fail (29, 52, 53). Docetaxel, among others, increases this proportion of cells both *in vitro* and in the clinic (100, 101). Our unpublished data indicate that this effect is less selective cell killing and more due to those undifferentiated cells symmetrically dividing more quickly than their differentiated counterparts. At least partially, that process may be disrupted by dTRIM24 (Figure 3.11, Panels A-C). While dTRIM24 did offset Docetaxel-induced expansion of undifferentiated cells significantly, the effect was not as dramatic as previous efforts (30).

The key issue with this preliminary data is that both compounds were dosed in a single concentration. While Docetaxel use is far better reported, the TRIM24 degrader, dTRIM24, has only been used by the original authors and only within the context of leukemia to improve their chemical probe. This lack of use leads to open questions about dosing, frequency, and applicability in other indications, such as TNBC. As of yet, these experiments have not been performed by any individual or lab and will need to be conducted if dTRIM24 is to be of any experimental use in TNBC.

Additionally, TRIM24 itself may not be the best target for targeting these de-differentiated cells. As TRIM24 is a transcriptional coactivator/repressor and even has ubiqutinase activity, all those functions are context dependent (40, 43, 44, 56, 57, 82, 90). However, the net result of that activity did lead to the most dangerous of cell sub-populations both in the cell line in question and those in late-stage, metastatic TNBC. While these preliminary results are encouraging, there may be a regulated gene or genes that may be more easily targeted or may have a larger impact on this

population than TRIM24 alone. Thus, additional RNA sequencing studies may be useful to determine a downstream target that may have a more likely chance of therapeutic success.

EPC2

EPC2 is a largely uncharacterized polycomb group protein, known to primarily interact with the lysine acetyltransferase complex NuA4/TIP60 (known as KAT5 in humans) (46). Besides a study discovering that EPC2 and its paralogue EPC1 maintain the oncogeneticity of MLL-dependent leukemia cells, it is largely unstudied in mammalian contexts (102). However, this work in mice and drosophila has been mainly focused on EPC1 rather than EPC2 (103). Only in higher order mammals is EPC2 found (103).

EPC2 was identified as a paralogue of the more researched EPC1, which is a non-catalytic but necessary member of the NuA4/TIP60 nuclear acetylase complex (47, 104, 105). EPC1 is known to interact with other proteins, such as RFP (Ring finger protein/TRIM27), cell-cycle transcription factor E2F1 (106-108). In drosophila and yeast, E(Pc) was found to have roles in stem cell fate, DNA integrity, and differentiation (109, 110).

In mammals, EPC was found to repress hematological tumor initiation by negatively regulating the JAK/STAT pathway, in contrast to other research indicating the opposite in mammalian cells (102, 111). In general, EPC have been found to regulate germline differentiation and stability (46, 110, 112). Aside from these few hints in hematological cancers, there is little research on either EPC1 or EPC2 in their role in cancer development, differentiation, or regulation.

While Enhancer of Polycomb is conserved in drosophila, yeast, and mammals, but each with different interacting partners (113). The only known binding partner of EPC2 is EZH2, which both interacts and represses EPC2 gene expression (114).

Based upon available information, much of it focused on the paralogue EPC1, it is difficult to speculate on what role EPC2 may have within the cell, specifically triple-negative breast cancer cells. It was selected for its role in increasing $EGFP^+$ cells, yet there seems to be no reason why this increase should occur. While the majority of research occurred in *Drosophila melanogaster* on the homologue E(Pc), the applicability to EPC2 is unclear. Additionally, future work was primarily conducted on the paralogue EPC1, which has a number of interacting partners and roles in DNA stability, differentiation, and germline stability in undifferentiated cells. While EPC1 and

EPC2 are similar (EPC1 with 836 bp and EPC2 with 807), it is unclear if EPC2 plays any role in these functions (103).

However, EPC1, EPC2, and KAT5 (known as Tip60 in mice) were all top hits in the primary screen. EPC1 was classified as a $mCherry^{-}EGFP^{+}$ hit and KAT5 as a $mCherry^{+}EGFP^{+}$ hit. While both EPC1 and EPC2 increased $EGFP^{+}$ cells, KAT5 had the added effect of decreasing the $mCherry^{+}EGFP^{+}$ positive cells. Thus, it may be possible EPC1 and EPC2 may perform similar roles in human cells, but that requires further biochemical investigation.

EPC2 siRNA replicated its activity in SUM149 by increasing CD44+CD24+ cells relative to control but demonstrated no increase in ALDEFLUOR™ activity (Figure 3.9, Panels B and C). Selecting for only for $CD44^{+}/24^{+}/ALDH^{+}$ cells, cells that mirror the identity of mCherry+EGFP+ cells, does not show a significant increase with EPC2 knockdown compared to negative control. Therefore, there seems to be a consistent effect in $CD44^{+}/24^{+}$ positive cells, irrespective of differentiation level. There must be some biochemical reason why this occurs, obscured by the lack of knowledge concerning EPC2.

While there is significantly more literature available for its paralogue EPC1, it remains difficult to discern how relevant this EPC1 data is. The most likely connection would be to KAT5, a lysine acetyltransferase. However, that connection has not been verified. However, it has been demonstrated that EPC2 and the lysine methyltransferase EZH2 associate with one another (114). EZH2 has primarily been investigated for his participation in the repressive PRC2 complex, but it also possesses non-canonical roles as a transcriptional activator (115-118). In the primary screen, both EZH2 and EPC2 were both selected for secondary screening for their effect on mCherry-EGFP- cells, so it may be more likely that EZH2 may have a role in producing this observed phenotype following EPC2 knockdown.

Mechanistic Analysis of how TRIM24 and EPC2 Effect their Sub-Population Changes

In the reporter cell line, TRIM24 and EPC2 knockdown had opposite effects: TRIM24 decreased undifferentiated cells (both stem-like and progenitor) and EPC2 increased the population of progenitor-like cells. This is curious since both TRIM24 and EPC2 have roles related to histone acetylation status. In TRIM24's case, it associates with H3Kme0 and H3K23Ac via its c-terminal

bromodomain (119). EPC2 is a member of the NuA4, which is a histone acetyltransferase that hyperacetylates H4 and H2A while they are in the nucleosome (46, 120). EPC2's role is to recognize any peptide chain during H2A acetylation, which is H4's (121). This acetylation increases likelihood of transcription (122). More effort will be needed to form a hypothesis of what may be occurring with these opposing roles.

TRIM24 (TIFα) is a protein that has diverse roles, namely as a transcriptional activator, negative regulator of both retinoic acid signaling and p53, an E3 ubiquitin ligase, a positive regulator of WNT/β-catenin signaling, of STAT3 signaling, and promotes nuclear receptor amplification of transcription (41, 44, 56, 58, 60, 65, 69, 73, 82, 85, 86). TRIM24 can associate with other nuclear receptors by associations with fellow coactivators CARM1 and GRIP1, indicating that the numerous binding domains TRIM24 possesses (RING, two B-boxes, Coiled-coil, PHD, and Bromodomain) may be function in binding other cofactors for increased gene expression by additional epigenetic modulation (123).

Therefore, nuclear receptors should be likely binding partners for TRIM24. Although estrogen, progesterone, and androgen receptors are some of the most well known of nuclear receptors, TNBCs are by definition negative for each of those receptors. Of potential interest is NR1D1 (Nuclear receptor subfamily 1 Group D Member 1), which is significantly correlated with positive response to chemotherapy in TNBC patients (124). However, this protein is understudied, especially in TNBC.

Two questions arise from TRIM24's association with nuclear receptors: 1) TRIM24 can bind a non-canonical histone signature to increase gene expression 2) This binding is enhanced by other transcriptional cofactors and nuclear receptors. The question is what cofactors and what nuclear receptors does TRIM24 bind and how does TRIM24 siRNA deplete undifferentiated cell types? One validated culprit appears: RXRA, also known as Retanoid X Receptor-alpha. However, TRIM24 has been previously shown to repress retinoic acid signaling (41, 90). This signaling is critical for cancer stem cell maintenance (42). It is then of some surprise that decreased levels of TRIM24 lead to decreased levels of undifferentiated cells in multiple cell lines. It may be that TRIM24 maintains this population by other mechanisms.

TRIM24 and TRIM33 (TIF-γ) form a heterodimer through their coiled-coiled regions (57). Together, this heterodimer has been demonstrated to have tumor-suppressive effects (57). Additionally, TRIM33 expression has been shown to be tumor suppressive, and low expression

levels are associated with poor outcome in operable TNBC patients (125, 126). Does TRIM24 knockdown release TRIM33 to perform other functions beyond those it conducts with TRIM24? Does a "freed" TRIM33 regulate other genes that decrease these undifferentiated sub-populations? Since TRIM33 has ubiquitin ligase activity, what proteins are then degraded by TRIM33? Does TRIM24 knockdown decrease WNT/B-Catenin/SOX2/SOX4 levels, thus disrupting CSC maintenance? These are all possibilities, and will require future work.

It may not be that the mechanical cause of undifferentiated cell decreases occur directly through lack of some process that TRIM24 performs but the release of other proteins, such as TRIM33, to have amplified, tumor suppressive effects that leads to the phenotype we see following TRIM24 siRNA use.

EPC2 role in NuA4 Histone Acetylase Complex

Unfortunately, there is a dearth of information on EPC2 specifically in the literature. Ostensibly, it has been identified as a member of the Tip60/NuA4 acetyltransferase complex but goes unnoticed in many papers, which list only the more studied EPC1. EPC2 lacks 29 amino acids that EPC1 possesses, but it is unknown what role those amino acids could play and is also unknown whether EPC1 can replace EPC2 or if EPC2 is even necessary for NuA4 histone acetylase activity (103). Additionally, EPC1/2's recognition of H4 is largely non-specific, and replacement of the H4 tail with H3 leads to the same acetylation of the H3-tail (121). Thus, NuA4 may be a non-specific, non-target acetylase complex.

One possibility is that reduction of EPC2 leads to further reduction of the NuA4 complex, but that assumes EPC1 cannot replace EPC2 within the complex. Considering that EPC1/2 is a recognition protein for H4, there may be less NuA4 associated with nucleosomes, and thus, less acetylation.

TRIM24 and EPC2 likely have different regions of regulations, even if TRIM24 and NuA4 regulate some of the same processes (127). In all likelihood, the inability for TRIM24 to bind H3K4 with any methylation likely excludes this protein from any actively transcribed regions. NuA4 likely is involved in maintaining actively transcribed regions rather than turning on rarer genes. Thus, TRIM24 activity may be more focused by its c-terminal binding region and by other associations with other proteins and receptors, and may be an alternate method to transcribe genes

that lack H3Kme3 while EPC2 (and NuA4) may be more closely associated with traditional transcription methods.

Broader Implications for Cancer Research

At the end of this screening campaign, there was only one group that had developed a reporter cell system in triple negative breast cancer. Their system relied upon coexpression of stem cell transcription factor NANOG with green fluorescence protein (128). However, no such siRNA screening campaign has been conducted with that single-marked reporter cell. Thus, this present work marks the first time a fluorescent reporter cell system has been used for any siRNA screening, let alone for epigenetic targets.

However, this approach should not be limited only to TNBC but can be applied to other types of aggressive and highly heterogeneous cancers, such as glioblastoma multiforme (GBM), which has a five-year relative survival rate of 5% (129). While TNBC is characterized by treatment-induced selection of resistant clones, GBM is characterized by being primary, developing *de novo,* or as secondary, where GBM develops from lower grade gliomas (51, 130). The transition from lower-grade to GBM occurs mainly due to MGMT methylation and mutations in the isocitrade dehydrogenase 1/2 genes (131, 132). The focus has been on the genetic heterogeneity of GBM types rather than examining the origins of that heterogeneity, which may also be epigenetics. While focusing on one protein subtype, different cancer types have different gene expression programs that can change with cancer induction. Any siRNA search ought to be rationally designed to focus on the most likely causes of heterogeneity, such as DNA repair mechanisms, metabolism, kinases, phosphatases, or other protein classes. Epigenetics would likely be of less interest since that is far more validated in TNBC than in GBM.

Such screening, not necessarily siRNA, can also be used to identify potential targets for therapeutic impact. For example, if a GBM fluorescent reporter protein is developed, it could be used to perform FDA-approved drugs to determine which compound affected which population and to what extent. These could guide future drug discovery and clinical trial efforts.

These highly-heterogeneous cancers retain their high mortality rates and their poor prognosis, mainly due to the current drug discovery and development model not favoring the approach that

would lead to improved outcomes for patients. Rather than focusing on synergy of combination therapies of the whole population, the focus should be on additive effects on all sub-populations. Such an approach may not lead to a cure, but any improvement in outcome is needed.

Materials and Methods

Cell line maintenance

SUM149 and its derivative cell lines were cultured in a 37° C, 95% humidity, 5% CO_2 environment in Ham's F-12 medium, (Invitrogen, Carlsbad, CA) supplemented with 5% fetal bovine serum (Fisher Scientific, Pittsburgh, PA), 1% antibiotic-antimycotic (Invitrogen, Carlsbad, CA), 5 μg/ml insulin (Sigma-Aldrich, St Louis, MO), 1 μg/ml hydrocortisone (Sigma-Aldrich, St Louis, MO), and 4 μg/ml gentamicin (Invitrogen, Carlsbad, CA).

The MDA-MB-468 cell line was cultured in a 37° C, 95% humidity, 5% CO_2 environment in DMEM media (Invitrogen, Carlsbad, CA) supplemented with 5% fetal bovine serum (Fisher Scientific, Pittsburgh, PA), 1% antibiotic/antimycotic (Invitrogen, Carlsbad, CA), 1% Glutamax (Invitrogen, Carlsbad, CA), 1% Sodium Pyruvate (Invitrogen, Carlsbad, CA).

The BT-20 cell line was cultured in a 37° C, 95% humidity, 5% CO_2 environment in 1640 RPMI (Corning, Dublin, OH) supplemented with 10% fetal bovine serum (Fisher Scientific, Pittsburgh, PA) and 1% antibiotic-antimycotic (Invitrogen, Carlsbad, CA).

Single siRNA Plate sets for Secondary Screening

Six siRNA plates were ordered that contained sets of four 0.1 nmol siRNA in a 96 well format (Horizon Discovery and Dharmacon, Lafayette, CO). These plates were centrifuged to ensure all siRNA were at the bottom of the wells before the siRNA was suspended in 1x siRNA buffer (Horizon Discovery and Dharmacon, Lafayette, CO) produced from 5x buffer and ultrapure water (Invitrogen, Carlsbad, CA). These stock plates served as the siRNA source for secondary screening efforts.

High-throughput siRNA Secondary Screening

4000 cells/well were seeded in clear-bottomed 96 well plates (Corning, Corning, NY) seeded with siRNA from stock plates (Human ON-TARGETplus by Horizon Discovery and Dharmacon, Lafayette, CO) for 20 nM final. To each well was added antibiotic free Ham's F-12 medium, (Invitrogen, Carlsbad, CA) supplemented with 5% fetal bovine serum (Fisher Scientific, Pittsburgh, PA), 5 μg/ml insulin (Sigma-Aldrich, St Louis, MO), and 1 μg/ml hydrocortisone (Sigma-Aldrich, St Louis, MO), 0.6 uL of Lipofectamine RNAimax Transfection Reagent (ThermoFischer Scientific, Carlsbad, CA), and Opti-Mem™ (Fisher Scientific, Waltham, MA). Cells were incubated for 96 hours. After that elapsed time, the media was aspirated and 50 uL of Trypsin-EDTA (0.25%) (ThermoFisher Scientific, Carlsbad, CA). Cells were returned to incubator for 10 minutes, after which each well was mixed by a multichannel pippetteman fifty total times to further displace the cells. The cells were allowed to incubate for several more minutes. Trypsin was neutralized with 50 uL of FluoroBrite™ DMEM Media (ThermoFisher Scientific, Carlsbad, CA) containing 2% fetal bovine serum (Fisher Scientific, Pittsburgh, PA), 5 μg/ml insulin (Sigma-Aldrich, St. Louis, MO), 1 μg/ml hydrocortisone (Sigma-Aldrich, St Louis, MO), and 1 ug/mL DAPI (Sigma-Aldrich, St. Louis, MO). Cells were pipetted again to ensure a single cell suspension. Plate was analyzed on a Zeti 5 flow cytometer (Bio-Rad, Hercules, CA).

Flow Cytometry

Flow cytometry was performed using anti-CD44-PE, anti-CD24-APC, and corresponding isotype antibodies (Biolegend, San Diego, CA) on a ZETI5 flow cytometer (Bio-Rad, Hercules, CA). The ALDEfluor™ Assay Kit was performed according to manufacturer's (StemCell Technologies, Vancouver, Canada) instructions with a single modification: the 37° incubation was performed in a water bath. Each sample was normalized to a matched N,N-diethylaminobenzaldehyde (DEAB) control for sample-by-sample gating. Analysis was performed using FCS Express (Version 7; De Novo Software, Pasadena, CA).

Confirmational Sub-population Analysis

500k SUM149, MDA-MB-468, or BT-20 were seeded in triplicate in 10 cm plates (Corning, Corning, NY) and dosed with 20 nM siRNA using Opti-Mem™ (Fisher Scientific, Waltham, MA) and Lipofectamine RNAimax Transfection Reagent (ThermoFischer Scientific, Carlsbad, CA) in antibiotic-antimycotic-free Ham's F12 media (Invitrogen, Carlsbad, CA). After a total of 96 hours had passed, cells were detached using Trypsin-EDTA (0.25%) (ThermoFisher Scientific, Carlsbad, CA) and neutralized with complete Ham's F12 media. Cells were counted, stained with anti-CD44-PE, anti-CD24-APC, and corresponding isotype antibodies (Biolegend, San Diego, CA) and the ALDEfluor™ assay (StemCell Technologies, Vancouver, Canada) was performed as previously described. Cells were reseeded in ALDEfluor™ buffer (StemCell Technologies, Vancouver, Canada) with 1 ug/mL DAPI (Sigma-Aldrich, St. Louis, MO). Each sample was normalized to a matched N,N-diethylaminobenzaldehyde (DEAB) control for sample-by-sample gating. Analysis was performed using FCS Express (Version 7; De Novo Software, Pasadena, CA). Data are means of each of the three samples by $CD44^{+}/CD24^{+}$ and percentage of $ALDH^{hi}$ signal. Error bars are the standard deviation of the DEAB-free samples.

MTS assay

SUM149 cells were seeded in a 96 well plate at 2000 cells/well with 20 nM of TRIM24 siRNA (Horizon Discovery and Dharmacon, Lafayette, CO), Opti-Mem™ (Fisher Scientific, Waltham,

MA), and 0.6 uL of Lipofectamine RNAimax Transfection Reagent (ThermoFischer Scientific, Carlsbad, CA) in antibiotic-antimycotic-free Ham's F12 media (Invitrogen, Carlsbad, CA). Every 24 hours, the MTS assay (Promega, Madison, WI) was conducted according to manufacturer's instructions. Fluorescent data was collected with a Cytation 5 plate reader (BioTek, Winooski, VT). Data represents the mean +/- standard deviation of the three samples following data normalization.

Combination treatment with dTRIM24 and Docetaxel

Reporter cells were seeded at $1x10^6$ cells per 10 cm dish in triplicate in 10 mL complete F12 media. These cells were treated with DMSO (5 µL), dTRIM24 (5 µM) (Tocris Biosciences, Bristol, UK), Docetaxel (5 nM) (LKT Laboratories, St. Paul, MN), or dTRIM24 (5 µM) and Docetaxel (5 nM). Cells were treated for 72 hours before being analyzed on a ZETI5 flow cytometer (Bio-Rad, Hercules, CA). Data was processed in FCS Express (Version 7; De Novo Software, Pasadena, CA).

Statistical analysis

Statistical comparisons used a single-tailed Student's t-test unless otherwise noted.

References

1. Pereira DA, Williams JA. Origin and evolution of high throughput screening. Br J Pharmacol. 2007;152(1):53-61.

2. Archer JR. History, evolution, and trends in compound management for high throughput screening. Assay and drug development technologies. 2004;2(6):675-81.

3. Echeverri CJ, Beachy PA, Baum B, Boutros M, Buchholz F, Chanda SK, et al. Minimizing the risk of reporting false positives in large-scale RNAi screens. Nat Methods. 2006;3(10):777-9.

4. Chen M, Du Q, Zhang HY, Wang X, Liang Z. High-throughput screening using siRNA (RNAi) libraries. Expert Rev Mol Diagn. 2007;7(3):281-91.

5. Henderson MC, Azorsa DO. High-throughput RNAi screening for the identification of novel targets. Methods in molecular biology (Clifton, NJ). 2013;986:89-95.

6. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. Br J Pharmacol. 2011;162(6):1239-49.

7. Calegari F, Haubensak W, Yang D, Huttner WB, Buchholz F. Tissue-specific RNA interference in postimplantation mouse embryos with endoribonuclease-prepared short interfering RNA. Proc Natl Acad Sci U S A. 2002;99(22):14236-40.

8. Parker GJ, Law TL, Lenoch FJ, Bolger RE. Development of High Throughput Screening Assays Using Fluorescence Polarization: Nuclear Receptor-Ligand-Binding and Kinase/Phosphatase Assays. Journal of biomolecular screening. 2000;5(2):77-88.

9. Maun HR, Wen X, Lingel A, de Sauvage FJ, Lazarus RA, Scales SJ, et al. Hedgehog pathway antagonist 5E1 binds hedgehog at the pseudo-active site. J Biol Chem. 2010;285(34):26570-80.

10. Nji E, Chatzikyriakidou Y, Landreh M, Drew D. An engineered thermal-shift screen reveals specific lipid preferences of eukaryotic and prokaryotic membrane proteins. Nat Commun. 2018;9(1):4253.

11. Newby ZE, O'Connell JD, 3rd, Gruswitz F, Hays FA, Harries WE, Harwood IM, et al. A general protocol for the crystallization of membrane proteins for X-ray structural investigation. Nat Protoc. 2009;4(5):619-37.

12. Hamilton AJ, Baulcombe DC. A species of small antisense RNA in posttranscriptional gene silencing in plants. Science. 1999;286(5441):950-2.

13. Novina CD, Sharp PA. The RNAi revolution. Nature. 2004;430(6996):161-4.

14. Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. 2001;15(2):188-200.

15. Parsons BD, Schindler A, Evans DH, Foley E. A direct phenotypic comparison of siRNA pools and multiple individual duplexes in a functional assay. PLoS One. 2009;4(12):e8471-e.

16. Smith C. Sharpening the tools of RNA interference. Nature Methods. 2006;3(6):475-86.

17. Sigoillot FD, King RW. Vigilance and validation: Keys to success in RNAi screening. ACS Chem Biol. 2011;6(1):47-60.

18. Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M, et al. Expression profiling reveals off-target gene regulation by RNAi. Nat Biotechnol. 2003;21(6):635-7.

19. Doench JG, Petersen CP, Sharp PA. siRNAs can function as miRNAs. Genes Dev. 2003;17(4):438-42.

20. Semizarov D, Frost L, Sarthy A, Kroeger P, Halbert DN, Fesik SW. Specificity of short interfering RNA determined through gene expression signatures. Proceedings of the National Academy of Sciences. 2003;100(11):6347.

21. Liu S, Cong Y, Wang D, Sun Y, Deng L, Liu Y, et al. Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. Stem Cell Reports. 2014;2(1):78-91.

22. Ringel I, Horwitz SB. Studies with RP 56976 (taxotere): a semisynthetic analogue of taxol. Journal of the National Cancer Institute. 1991;83(4):288-91.

23. Taguchi T, Furue H, Niitani H, Ishitani K, Kanamaru R, Hasegawa K, et al. [Phase I clinical trial of RP 56976 (docetaxel) a new anticancer drug]. Gan To Kagaku Ryoho. 1994;21(12):1997-2005.

24. Einzig AI, Schuchter LM, Recio A, Coatsworth S, Rodriquez R, Wiernik PH. Phase II trial of docetaxel (Taxotere) in patients with metastatic melanoma previously untreated with cytotoxic chemotherapy. Med Oncol. 1996;13(2):111-7.

25. Miyoshi Y, Uemura H, Nakamura M, Hasumi H, Sugiura S, Makiyama K, et al. Treatment of androgen-independent, hormone-refractory prostate cancer with docetaxel in Japanese patients. Int J Clin Oncol. 2005;10(3):182-6.

26. Piccart MJ, Gore M, Ten Bokkel Huinink W, Van Oosterom A, Verweij J, Wanders J, et al. Docetaxel: an active new drug for treatment of advanced epithelial ovarian cancer. Journal of the National Cancer Institute. 1995;87(9):676-81.

27. Kassam F, Enright K, Dent R, Dranitsaris G, Myers J, Flynn C, et al. Survival Outcomes for Patients with Metastatic Triple-Negative Breast Cancer: Implications for Clinical Practice and Trial Design. Clinical Breast Cancer. 2009;9(1):29-33.

28. Aulmann S, Waldburger N, Penzel R, Andrulis M, Schirmacher P, Sinn HP. Reduction of CD44(+)/CD24(-) breast cancer cells by conventional cytotoxic chemotherapy. Hum Pathol. 2010;41(4):574-81.

29. Lee HE, Kim JH, Kim YJ, Choi SY, Kim SW, Kang E, et al. An increase in cancer stem cell population after primary systemic therapy is a poor prognostic factor in breast cancer. Br J Cancer. 2011;104(11):1730-8.

30. Burnett JP, Lim G, Li Y, Shah RB, Lim R, Paholak HJ, et al. Sulforaphane enhances the anticancer activity of taxanes against triple negative breast cancer by killing cancer stem cells. Cancer Lett. 2017;394:52-64.

31. Creighton CJ, Li X, Landis M, Dixon JM, Neumeister VM, Sjolund A, et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proc Natl Acad Sci U S A. 2009;106(33):13820-5.

32. Kulsum S, Sudheendra HV, Pandian R, Ravindra DR, Siddappa G, R N, et al. Cancer stem cell mediated acquired chemoresistance in head and neck cancer can be abrogated by aldehyde dehydrogenase 1 A1 inhibition. Mol Carcinog. 2017;56(2):694-711.

33. Chen H, Shien K, Suzawa K, Tsukuda K, Tomida S, Sato H, et al. Elacridar, a third-generation ABCB1 inhibitor, overcomes resistance to docetaxel in non-small cell lung cancer. Oncol Lett. 2017;14(4):4349-54.

34. Qiu S, Deng L, Bao Y, Jin K, Tu X, Li J, et al. Reversal of docetaxel resistance in prostate cancer by Notch signaling inhibition. Anticancer Drugs. 2018;29(9):871-9.

35. Sánchez BG, Bort A, Vara-Ciruelos D, Díaz-Laviada I. Androgen Deprivation Induces Reprogramming of Prostate Cancer Cells to Stem-Like Cells. Cells. 2020;9(6).

36. Pegram MD, Konecny GE, O'Callaghan C, Beryt M, Pietras R, Slamon DJ. Rational combinations of trastuzumab with chemotherapeutic drugs used in the treatment of breast cancer. Journal of the National Cancer Institute. 2004;96(10):739-49.

37. Yang CY, Qin C, Bai L, Wang S. Small-molecule PROTAC degraders of the Bromodomain and Extra Terminal (BET) proteins - A review. Drug Discov Today Technol. 2019;31:43-51.

38. Gechijian LN, Buckley DL, Lawlor MA, Reyes JM, Paulk J, Ott CJ, et al. Functional TRIM24 degrader via conjugation of ineffectual bromodomain and VHL ligands. Nat Chem Biol. 2018;14(4):405-12.

39. Sachse C, Krausz E, Krönke A, Hannus M, Walsh A, Grabner A, et al. High-throughput RNA interference strategies for target discovery and validation by using synthetic short interfering RNAs: functional genomics investigations of biological pathways. Methods Enzymol. 2005;392:242-77.

40. Allton K, Jain AK, Herz HM, Tsai WW, Jung SY, Qin J, et al. Trim24 targets endogenous p53 for degradation. Proc Natl Acad Sci U S A. 2009;106(28):11612-6.

41. Khetchoumian K, Teletin M, Tisserand J, Herquel B, Ouararhni K, Losson R. Trim24 (Tif1 alpha): an essential 'brake' for retinoic acid-induced transcription to prevent liver cancer. Cell cycle (Georgetown, Tex). 2008;7(23):3647-52.

42. Gudas LJ, Wagner JA. Retinoids regulate stem cell differentiation. J Cell Physiol. 2011;226(2):322-30.

43. Groner AC, Cato L, de Tribolet-Hardy J, Bernasocchi T, Janouskova H, Melchers D, et al. TRIM24 Is an Oncogenic Transcriptional Activator in Prostate Cancer. Cancer Cell. 2016;29(6):846-58.

44. Tsai WW, Wang Z, Yiu TT, Akdemir KC, Xia W, Winter S, et al. TRIM24 links a non-canonical histone signature to breast cancer. Nature. 2010;468(7326):927-32.

45. Chambon M, Orsetti B, Berthe ML, Bascoul-Mollevi C, Rodriguez C, Duong V, et al. Prognostic significance of TRIM24/TIF-1α gene expression in breast cancer. Am J Pathol. 2011;178(4):1461-9.

46. Galarneau L, Nourani A, Boudreault AA, Zhang Y, Héliot L, Allard S, et al. Multiple links between the NuA4 histone acetyltransferase complex and epigenetic control of transcription. Mol Cell. 2000;5(6):927-37.

47. Selleck W, Fortin I, Sermwittayawong D, Côté J, Tan S. The Saccharomyces cerevisiae Piccolo NuA4 histone acetyltransferase complex requires the Enhancer of Polycomb A domain and chromodomain to acetylate nucleosomes. Mol Cell Biol. 2005;25(13):5535-42.

48. Valero V. Docetaxel as single-agent therapy in metastatic breast cancer: clinical efficacy. Semin Oncol. 1997;24(4 Suppl 13):S13-1-s-8.

49. Robinson DM, Keating GM. Albumin-bound Paclitaxel: in metastatic breast cancer. Drugs. 2006;66(7):941-8.

50. Lavelle F, Bissery MC, Combeau C, Riou JF, Vrignaud P, André S. Preclinical evaluation of docetaxel (Taxotere). Semin Oncol. 1995;22(2 Suppl 4):3-16.

51. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, et al. Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell. 2018;173(4):879-93.e13.

52. Li X, Lewis MT, Huang J, Gutierrez C, Osborne CK, Wu MF, et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. Journal of the National Cancer Institute. 2008;100(9):672-9.

53. Hirschmann-Jax C, Foster AE, Wulf GG, Nuchtern JG, Jax TW, Gobel U, et al. A distinct "side population" of cells with high drug efflux capacity in human tumor cells. Proc Natl Acad Sci U S A. 2004;101(39):14228-33.

54. Hashida S, Yamamoto H, Shien K, Miyoshi Y, Ohtsuka T, Suzawa K, et al. Acquisition of cancer stem cell-like properties in non-small cell lung cancer with acquired resistance to afatinib. Cancer Sci. 2015;106(10):1377-84.

55. Zhang J, Luo H, Liu H, Ye W, Luo R, Chen HF. Synergistic Modification Induced Specific Recognition between Histone and TRIM24 via Fluctuation Correlation Network Analysis. Sci Rep. 2016;6:24587.

56. Kanno Y, Kure Y, Kobayashi S, Mizuno M, Tsuchiya Y, Yamashita N, et al. Tripartite Motif Containing 24 Acts as a Novel Coactivator of the Constitutive Active/Androstane Receptor. Drug Metab Dispos. 2018;46(1):46-52.

57. Herquel B, Ouararhni K, Davidson I. The TIF1α-related TRIM cofactors couple chromatin modifications to transcriptional regulation, signaling and tumor suppression. Transcription. 2011;2(5):231-6.

58. Torres-Padilla ME, Zernicka-Goetz M. Role of TIF1alpha as a modulator of embryonic transcription in the mouse zygote. The Journal of cell biology. 2006;174(3):329-38.

59. Teyssier C, Ou CY, Khetchoumian K, Losson R, Stallcup MR. Transcriptional intermediary factor 1alpha mediates physical interaction and functional synergy between the coactivator-associated arginine methyltransferase 1 and glucocorticoid receptor-interacting protein 1 nuclear receptor coactivators. Mol Endocrinol. 2006;20(6):1276-86.

60. Remboutsika E, Lutz Y, Gansmuller A, Vonesch JL, Losson R, Chambon P. The putative nuclear receptor mediator TIF1alpha is tightly associated with euchromatin. J Cell Sci. 1999;112 (Pt 11):1671-83.

61. Lv D, Jia F, Hou Y, Sang Y, Alvarez AA, Zhang W, et al. Histone Acetyltransferase KAT6A Upregulates PI3K/AKT Signaling through TRIM24 Binding. Cancer Res. 2017;77(22):6190-201.

62. Appikonda S, Thakkar KN, Shah PK, Dent SYR, Andersen JN, Barton MC. Cross-talk between chromatin acetylation and SUMOylation of tripartite motif-containing protein 24 (TRIM24) impacts cell adhesion. J Biol Chem. 2018;293(19):7476-85.

63. Ma L, Yuan L, An J, Barton MC, Zhang Q, Liu Z. Histone H3 lysine 23 acetylation is associated with oncogene TRIM24 expression and a poor prognosis in breast cancer. Tumour Biol. 2016;37(11):14803-12.

64. Heery DM, Kalkhoven E, Hoare S, Parker MG. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. Nature. 1997;387(6634):733-6.

65. Remboutsika E, Yamamoto K, Harbers M, Schmutz M. The bromodomain mediates transcriptional intermediary factor 1alpha -nucleosome interactions. J Biol Chem. 2002;277(52):50318-25.

66. Liu J, Li F, Bao H, Jiang Y, Zhang S, Ma R, et al. The polar warhead of a TRIM24 bromodomain inhibitor rearranges a water-mediated interaction network. Febs j. 2017;284(7):1082-95.

67. Cui Z, Cao W, Li J, Song X, Mao L, Chen W. TRIM24 overexpression is common in locally advanced head and neck squamous cell carcinoma and correlates with aggressive malignant phenotypes. PLoS One. 2013;8(5):e63887.

68. Liu X, Huang Y, Yang D, Li X, Liang J, Lin L, et al. Overexpression of TRIM24 is associated with the onset and progress of human hepatocellular carcinoma. PLoS One. 2014;9(1):e85462.

69. Lv D, Li Y, Zhang W, Alvarez AA, Song L, Tang J, et al. TRIM24 is an oncogenic transcriptional co-activator of STAT3 in glioblastoma. Nat Commun. 2017;8(1):1454.

70. Offermann A, Roth D, Hupe MC, Hohensteiner S, Becker F, Joerg V, et al. TRIM24 as an independent prognostic biomarker for prostate cancer. Urol Oncol. 2019;37(9):576.e1-.e10.

71. Wang H, Xue W, Jiang X. Overexpression of TRIM24 Stimulates Proliferation and Glucose Metabolism of Head and Neck Squamous Cell Carcinoma. Biomed Res Int. 2018;2018:6142843.

72. Zhou HE, Pan SS, Han H. TRIM24 aggravates the progression of ovarian cancer through negatively regulating FOXM1 level. Eur Rev Med Pharmacol Sci. 2019;23(24):10647-56.

73. Fang Z, Deng J, Zhang L, Xiang X, Yu F, Chen J, et al. TRIM24 promotes the aggression of gastric cancer via the Wnt/β-catenin signaling pathway. Oncol Lett. 2017;13(3):1797-806.

74. Lin L, Zhao W, Sun B, Wang X, Liu Q. Overexpression of TRIM24 is correlated with the progression of human cervical cancer. Am J Transl Res. 2017;9(2):620-8.

75. Zhu Y, Zhao L, Shi K, Huang Z, Chen B. TRIM24 promotes hepatocellular carcinoma progression via AMPK signaling. Exp Cell Res. 2018;367(2):274-81.

76. Zhang LH, Yin AA, Cheng JX, Huang HY, Li XM, Zhang YQ, et al. TRIM24 promotes glioma progression and enhances chemoresistance through activation of the PI3K/Akt signaling pathway. Oncogene. 2015;34(5):600-10.

77. Zhang LH, Yin YH, Chen HZ, Feng SY, Liu JL, Chen L, et al. TRIM24 Promotes Stemness and Invasiveness of Glioblastoma Cells via Activating SOX2 Expression. Neuro Oncol. 2020.

78. Li H, Sun L, Tang Z, Fu L, Xu Y, Li Z, et al. Overexpression of TRIM24 correlates with tumor progression in non-small cell lung cancer. PLoS One. 2012;7(5):e37657.

79. Chang YC, Chi LH, Chang WM, Su CY, Lin YF, Chen CL, et al. Glucose transporter 4 promotes head and neck squamous cell carcinoma metastasis through the TRIM24-DDX58 axis. J Hematol Oncol. 2017;10(1):11.

80. Li C, Xin H, Shi Y, Mu J. Knockdown of TRIM24 suppresses growth and induces apoptosis in acute myeloid leukemia through downregulation of Wnt/GSK-3β/β-catenin signaling. Hum Exp Toxicol. 2020;39(12):1725-36.

81. Yu T, Gan S, Zhu Q, Dai D, Li N, Wang H, et al. Modulation of M2 macrophage polarization by the crosstalk between Stat6 and Trim24. Nat Commun. 2019;10(1):4353.

82. Jain AK, Allton K, Duncan AD, Barton MC. TRIM24 is a p53-induced E3-ubiquitin ligase that undergoes ATM-mediated phosphorylation and autodegradation during DNA damage. Mol Cell Biol. 2014;34(14):2695-709.

83. Fong KW, Zhao JC, Song B, Zheng B, Yu J. TRIM28 protects TRIM24 from SPOP-mediated degradation and promotes prostate cancer progression. Nat Commun. 2018;9(1):5007.

84. Kikuchi M, Okumura F, Tsukiyama T, Watanabe M, Miyajima N, Tanaka J, et al. TRIM24 mediates ligand-dependent activation of androgen receptor and is repressed by a bromodomain-containing protein, BRD7, in prostate cancer cells. Biochim Biophys Acta. 2009;1793(12):1828-36.

85. Thenot S, Charpin M, Bonnet S, Cavailles V. Estrogen receptor cofactors expression in breast and endometrial human cancer cells. Mol Cell Endocrinol. 1999;156(1-2):85-93.

86. Thénot S, Bonnet S, Boulahtouf A, Margeat E, Royer CA, Borgna JL, et al. Effect of ligand and DNA binding on the interaction between human transcription intermediary factor 1alpha and estrogen receptors. Mol Endocrinol. 1999;13(12):2137-50.

87. Eng FC, Barsalou A, Akutsu N, Mercier I, Zechel C, Mader S, et al. Different classes of coactivators recognize distinct but overlapping binding sites on the estrogen receptor ligand binding domain. J Biol Chem. 1998;273(43):28371-7.

88. Sathya G, Yi P, Bhagat S, Bambara RA, Hilf R, Muyan M. Structural regions of ERalpha critical for synergistic transcriptional responses contain co-factor interacting surfaces. Mol Cell Endocrinol. 2002;192(1-2):171-85.

89. Thénot S, Henriquet C, Rochefort H, Cavaillès V. Differential interaction of nuclear receptors with the putative human transcriptional coactivator hTIF1. J Biol Chem. 1997;272(18):12062-8.

90. Khetchoumian K, Teletin M, Tisserand J, Mark M, Herquel B, Ignat M, et al. Loss of Trim24 (Tif1alpha) gene function confers oncogenic activity to retinoic acid receptor alpha. Nat Genet. 2007;39(12):1500-6.

91. Lovat PE, Annicchiarico-Petruzzelli M, Corazzari M, Dobson MG, Malcolm AJ, Pearson AD, et al. Differential effects of retinoic acid isomers on the expression of nuclear receptor co-regulators in neuroblastoma. FEBS Lett. 1999;445(2-3):415-9.

92. Carrier M, Lutzing R, Gaouar S, Rochette-Egly C. TRIM24 mediates the interaction of the retinoic acid receptor alpha with the proteasome. FEBS Lett. 2018;592(8):1426-33.

93. Chavez KJ, Garimella SV, Lipkowitz S. Triple negative breast cancer cell lines: one tool in the search for better treatment of triple negative breast cancer. Breast Dis. 2010;32(1-2):35-48.

94. Lacerda L, Debeb BG, Smith D, Larson R, Solley T, Xu W, et al. Mesenchymal stem cells mediate the clinical phenotype of inflammatory breast cancer in a preclinical model. Breast Cancer Res. 2015;17(1):42.

95. Charafe-Jauffret E, Ginestier C, Iovino F, Tarpin C, Diebel M, Esterni B, et al. Aldehyde dehydrogenase 1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer. Clin Cancer Res. 2010;16(1):45-55.

96. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res. 2007;13(8):2329-34.

97. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences. 2001;98(19):10869-74.

98. Banerjee S, Reis-Filho JS, Ashley S, Steele D, Ashworth A, Lakhani SR, et al. Basal-like breast carcinomas: clinical outcome and response to chemotherapy. J Clin Pathol. 2006;59(7):729-35.

99. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al. Immunohistochemical and Clinical Characterization of the Basal-Like Subtype of Invasive Breast Carcinoma. Clinical Cancer Research. 2004;10(16):5367-74.

100. Tanei T, Morimoto K, Shimazu K, Kim SJ, Tanji Y, Taguchi T, et al. Association of Breast Cancer Stem Cells Identified by Aldehyde Dehydrogenase 1 Expression with Resistance to Sequential Paclitaxel and Epirubicin-Based Chemotherapy for Breast Cancers. Clinical Cancer Research. 2009;15(12):4234.

101. Creighton CJ, Li X, Landis M, Dixon JM, Neumeister VM, Sjolund A, et al. Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proceedings of the National Academy of Sciences. 2009;106(33):13820.

102. Huang X, Spencer GJ, Lynch JT, Ciceri F, Somerville TD, Somervaille TC. Enhancers of Polycomb EPC1 and EPC2 sustain the oncogenic potential of MLL leukemia stem cells. Leukemia. 2014;28(5):1081-91.

103. Searle NE, Pillus L. Critical genomic regulation mediated by Enhancer of Polycomb. Current genetics. 2018;64(1):147-54.

104. Searle NE, Torres-Machorro AL, Pillus L. Chromatin Regulation by the NuA4 Acetyltransferase Complex Is Mediated by Essential Interactions Between Enhancer of Polycomb (Epl1) and Esa1. Genetics. 2017;205(3):1125-37.

105. Chittuluru JR, Chaban Y, Monnet-Saksouk J, Carrozza MJ, Sapountzi V, Selleck W, et al. Structure and nucleosome interaction of the yeast NuA4 and Piccolo-NuA4 histone acetyltransferase complexes. Nat Struct Mol Biol. 2011;18(11):1196-203.

106. Tezel G, Shimono Y, Murakumo Y, Kawai K, Fukuda T, Iwahashi N, et al. Role for O-glycosylation of RFP in the interaction with enhancer of polycomb. Biochem Biophys Res Commun. 2002;290(1):409-14.

107. Wang Y, Alla V, Goody D, Gupta SK, Spitschak A, Wolkenhauer O, et al. Epigenetic factor EPC1 is a master regulator of DNA damage response by interacting with E2F1 to silence death and activate metastasis-related gene signatures. Nucleic Acids Res. 2016;44(1):117-33.

108. Shimono Y, Murakami H, Hasegawa Y, Takahashi M. RET finger protein is a transcriptional repressor and interacts with enhancer of polycomb that has dual transcriptional functions. J Biol Chem. 2000;275(50):39411-9.

109. Qi D, Jin H, Lilja T, Mannervik M. Drosophila Reptin and other TIP60 complex components promote generation of silent chromatin. Genetics. 2006;174(1):241-51.

110. Feng L, Shi Z, Chen X. Enhancer of polycomb coordinates multiple signaling pathways to promote both cyst and germline stem cell differentiation in the Drosophila adult testis. PLoS Genet. 2017;13(2):e1006571.

111. Bailetti AA, Negrón-Piñeiro LJ, Dhruva V, Harsh S, Lu S, Bosula A, et al. Enhancer of Polycomb and the Tip60 complex repress hematological tumor initiation by negatively regulating JAK/STAT pathway activity. Dis Model Mech. 2019;12(5).

112. Feng L, Shi Z, Xie J, Ma B, Chen X. Enhancer of polycomb maintains germline activity and genome integrity in Drosophila testis. Cell Death Differ. 2018;25(8):1486-502.

113. Stankunas K, Berger J, Ruse C, Sinclair DA, Randazzo F, Brock HW. The enhancer of polycomb gene of Drosophila encodes a chromatin protein conserved in yeast and mammals. Development. 1998;125(20):4055-66.

114. Guil S, Soler M, Portela A, Carrère J, Fonalleras E, Gómez A, et al. Intronic RNAs mediate EZH2 regulation of epigenetic targets. Nature Structural & Molecular Biology. 2012;19(7):664-70.

115. Gan L, Yang Y, Li Q, Feng Y, Liu T, Guo W. Epigenetic regulation of cancer progression by EZH2: from biological insights to therapeutic potential. Biomark Res. 2018;6:10.

116. Yan J, Li B, Lin B, Lee PT, Chung T-H, Tan J, et al. EZH2 phosphorylation by JAK3 mediates a switch to noncanonical function in natural killer/T-cell lymphoma. Blood. 2016;128(7):948-58.

117. Kim KH, Roberts CW. Targeting EZH2 in cancer. Nat Med. 2016;22(2):128-34.

118. Kirmizis A, Bartley SM, Kuzmichev A, Margueron R, Reinberg D, Green R, et al. Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. Genes Dev. 2004;18(13):1592-605.

119. Tsai W-W, Wang Z, Yiu TT, Akdemir KC, Xia W, Winter S, et al. TRIM24 links a non-canonical histone signature to breast cancer. Nature. 2010;468(7326):927-32.

120. Doyon Y, Selleck W, Lane WS, Tan S, Côté J. Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. Mol Cell Biol. 2004;24(5):1884-96.

121. Xu P, Li C, Chen Z, Jiang S, Fan S, Wang J, et al. The NuA4 Core Complex Acetylates Nucleosomal Histone H4 through a Double Recognition Mechanism. Molecular Cell. 2016;63(6):965-75.

122. Nourani A, Doyon Y, Utley RT, Allard S, Lane WS, Côté J. Role of an ING1 growth regulator in transcriptional activation and targeted histone acetylation by the NuA4 complex. Mol Cell Biol. 2001;21(22):7629-40.

123. Teyssier C, Ou C-Y, Khetchoumian K, Losson R, Stallcup MR. Transcriptional intermediary factor 1alpha mediates physical interaction and functional synergy between the coactivator-associated arginine methyltransferase 1 and glucocorticoid receptor-interacting protein 1 nuclear receptor coactivators. Molecular endocrinology (Baltimore, Md). 2006;20(6):1276-86.

124. Na H, Han J, Ka NL, Lee MH, Choi YL, Shin YK, et al. High expression of NR1D1 is associated with good prognosis in triple-negative breast cancer patients treated with chemotherapy. Breast Cancer Res. 2019;21(1):127.

125. Pommier RM, Gout J, Vincent DF, Alcaraz LB, Chuvin N, Arfi V, et al. TIF1γ Suppresses Tumor Progression by Regulating Mitotic Checkpoints and Chromosomal Stability. Cancer Res. 2015;75(20):4335-50.

126. Kassem L, Deygas M, Fattet L, Lopez J, Goulvent T, Lavergne E, et al. TIF1γ interferes with TGFβ1/SMAD4 signaling to promote poor outcome in operable breast cancer patients. BMC Cancer. 2015;15:453.

127. Judes G, Rifaï K, Ngollo M, Daures M, Bignon Y-J, Penault-Llorca F, et al. A bivalent role of TIP60 histone acetyl transferase in human cancer. Epigenomics. 2015;7(8):1351-63.

128. Thiagarajan PS, Hitomi M, Hale JS, Alvarado AG, Otvos B, Sinyuk M, et al. Development of a Fluorescent Reporter System to Delineate Cancer Stem Cells in Triple-Negative Breast Cancer. Stem Cells. 2015;33(7):2114-25.

129. Ostrom QT, Bauchet L, Davis FG, Deltour I, Fisher JL, Langer CE, et al. The epidemiology of glioma in adults: a "state of the science" review. Neuro Oncol. 2014;16(7):896-913.

130. Ohgaki H, Kleihues P. The definition of primary and secondary glioblastoma. Clin Cancer Res. 2013;19(4):764-72.

131. Watanabe T, Nobusawa S, Kleihues P, Ohgaki H. IDH1 mutations are early events in the development of astrocytomas and oligodendrogliomas. Am J Pathol. 2009;174(4):1149-53.

132. Hegi ME, Diserens A-C, Gorlia T, Hamou M-F, de Tribolet N, Weller M, et al. MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. New England Journal of Medicine. 2005;352(10):997-1003.

Gene	Gene ID	Function
ACTL6B	51412	Chromatin remodeling
ARID4B	51742	Subunit of SIN3A
ATF2	1386	anti-apoptosis, cell growth, and DNA damage response
CARM1	10498	Methylates arginine in proteins
CHAF1A	10036	Core component of the CAF-1 complex-- chromatin assembly in DNA replication and DNA repair.
CHD2	1106	DNA-binding helicase that specifically binds to the promoter of target genes
CTBP1	1487	Corepressor
VPRBP	9730	Substrate recognition component of E3 ubiquitin-protein ligase complexes
FBXO17	115290	Substrate-recognition component of the SCF (SKP1-CUL1-F-box protein)-type E3 ubiquitin ligase complex
GTF2H1	2965	Component of the transcription and DNA repair factor IIH (TFIIH) core complex
CDK1	983	Kinase essential for G1/S and G2/M phase transitions
BAZ1B	9031	Chromatin remodeling, acts as a transcription regulator, and DNA damage response
FBXO44	93611	Involved in chromatin-dependent regulation of transcription
JADE3	9767	Member of a family of large proteins containing PHD-type zinc fingers.
KDM4D	55693	Histone demethylase (H3k9)
KDM7A	80853	Demethylase (H3K9me2, H3K27me2, H4K20Me1)
CUL4B	8450	E3 ubiquitin ligase and catalyzes the polyubiquitination of specific protein substrates in the cell.
PHF2	5253	Lysine demethylase (histones and non-histone proteins)
USP49	25862	Deubiquitase (H2BK120Ub, a marker of transcriptional activiation).
PHF19	26147	Polycomb group (PcG)protein, binds (H3K36me3) and recruits the PRC2 complex.
CDK2	1017	Serine/threonine kinase. Critical during the G1 to S phase transition
CBX3	11335	Transcriptional silencing. Recognizes and binds H3K9me1/2/3
BRMS1	25855	Transcriptional repressor. Down-regulates transcription activation by NF-kappa-B
PRKAA2	5563	Catalytic subunit of AMP-activated protein kinase (AMPK), an energy sensor protein kinase
TDRD7	23424	Contains conserved Tudor domains and LOTUS domains. RNA processing.
NCOA3	8202	nuclear receptor coactivator that enhances with nuclear hormone receptors
CHD5	26038	Chromodomain helicase DNA-binding protein family member.
SIRT6	51548	Sirtuin family of NAD-dependent enzymes that are implicated in cellular stress resistance, genomic stability, aging and energy homeostasis.
WDR92	116143	WD40 domain protein.
KDM6B	23135	Lysine demethylase (H3K27me2 or H3K27me3).

Table 3.1 Viability Hits Selected for Secondary Screening.

Primary hits ordered for secondary screening. Please note that hits in this section were selected by ability to increase or decrease the population to a significant value by a Student's two-tailed T-Test ($P < 0.05$). This table represents viability hits, which decreased all populations. All information concerning protein function was summarized from Pubmed gene sites for each protein and does not reflect the full scope of activities and protein-protein interactions (Accessed October 10th, 2020)

Gene	Gene ID	Function
HDGFL1	154150	Cellular proliferation and differentiation.
HR	55806	Transcriptional corepressor of nuclear receptors. Connected with deacetylases
HUWE1	10075	E3 ubiquitin ligase. Degrades for Mcl1 degradation and p53
JADE2	23338	E3 ubiquitin-protein ligase. Chromatin organization.
JMJD8	339123	Positive regulator of TNF-induced NF-kappa-B signaling
KANSL3	55683	Chromatin organization. Histone acetyltransferase activity (H4k5 and H4k16).
KAT5	10524	Histone acetyltransferase. Chromatin remodeling and transcription.
KDM3B	51780	Lysine Demethylase 3B. Chromatin organization. Related to ATP and metabolic pathways
KDM5C	8242	Histone demethylase (H3k4me). Transcription and chromatin remodeling
KDM8	79831	Histone lysine demethylase. Tumor suppressor
G2E3	55632	G2/M-Phase Specific E3 Ubiquitin Protein Ligase). Essential in early embryonic development to prevent apoptotic death
ING5	84289	Tumor suppressor protein that inhibits cell growth and induces apoptosis. It interacts with tumor suppressor p53.
SMYD4	114826	Might be a methyltransferase
JADE2	23338	E3 ubiquitin-protein ligase. Chromatin organization.
SFMBT2	57713	DNA binding and histone binding
BOP1	23246	rRNA processing in the nucleus and cytosol and Gene Expression
ATF2	1386	(HAT) that specifically acetylates histones H2B and H4. May be involved in DNA damage response
KMT2D	8085	Histone Methyltransferase (H3k4). Transcriptional regulation via association with ASCOM
BRWD3	254065	WD40 repeat protein. Believed to be involved in chromatin modification
RNF8	9025	Ring-finger protein. Shown to interact with several class II ubiquitin-conjugating enzymes (E2), including UBE2E1/UBCH6, UBE2E2, and UBE2E3.Depletion of this protein causes cell growth inhibition and cell cycle arrest
MECOM	2122	Transcriptional regulator and oncoprotein that may be involved in hematopoiesis, apoptosis, development, and cell differentiation and proliferation. Interacts with CTBP1, SMAD3, CREBBP, KAT2B, MAPK8, and MAPK9
KDM3A	55818	Zinc-finger protein. transcriptional activation. Histone demethylase (H3k9)
PRMT2	3275	Arginine Methyltransferase. Associated with Breast Cancer. Coactivator of nuclear hormone receptors
HDAC10	83933	Histone deacetylase. Uncharacterized.
JMJD4	65094	xoglutarate and iron-dependent C4-lysyl hydroxylation of ETF1 at 'Lys-63' thereby promoting the translational termination e
PHF20L1	51105	Methyllysine-binding protein, component of the MOF histone acetyltransferase protein complex.
GTF2B	2959	This gene encodes the general transcription factor IIB, one of the ubiquitous factors required for transcription initiation by RNA polymerase II.
NAP1L2	4674	NAP1L2 (Nucleosome Assembly Protein 1 Like 2) chromatin binding and histone binding.
NSD2	7468	Expressed ubiquitously in early development. Histone methyltransferase (H3k27me3)
SMNDC1	10285	Necessary for spliceosome assembly. Overexpression causes apoptosis

Table 3.2 mCherry+EGFP+ Hits Selected for Secondary Screening

Primary hits ordered for secondary screening. Please note that hits in this section were selected by ability to increase or decrease the population to a significant value by a Student's two-tailed T-Test ($P < 0.05$). This table represents double-positive hits, which increased or decreased the double-positive sub-population. All information concerning protein function was summarized from Pubmed gene sites for each protein and does not reflect the full scope of activities and protein-protein interactions (Accessed October 10th, 2020)

Gene	Gene ID	Function
KMT2D	8085	Histone methyltransferase (H3k4) Lys-4 position of histone H3. Part of a large protein complex called ASCOM
L3MBTL1	26013	Polycomb group (PcG) protein that reads mono- and dimethyllysine residues (PTMs). Transcriptional repressor via heterochromatin maintenance. Required for normal mitosis.
METTL21C	196541	Protein-lysine methyltransferase
PADI2	11240	Peptidyl arginine deiminase family of enzymes, which catalyze the post-translational deimination of targets, including vimentin
PCMT1	5110	Type II of protein carboxyl methyltransferase enzymes. The encoded enzyme plays a role in protein repair
PHC3	80012	Component of a Polycomb group (PcG) multiprotein PRC1-like complex, a complex class required to maintain heterochromatin.
PHF3	23469	Contains a PHD finger. May function as a transcription factor
PHIP	55023	Stimulates cell proliferation through regulation of cyclin transcription and has an anti-apoptotic activity through AKT1 phosphorylation and activation.
PRDM10	56980	Transcription factor. It contains a positive regulatory domain
PRDM7	11105	May function as a histone methyltransferase; Among its related pathways are Lysine degradation and Metabolism.
CBX1	10951	The protein may play an important role in the epigenetic control of chromatin structure and gene expression by interacting with both heterochromatin and centromeres
SMARCA4	6597	SWI/SNF member. Helicase and ATPase activities. Regulate transcription by chromatin alteration. Regulates CD44 expression
KDM1B	221656	Histone demethylase (H3Kme1/2). Required for de novo DNA methylation of a subset of imprinted genes during oogenesis.
H2AFZ	3015	This gene encodes a replication-independent member of the histone H2A family that is distinct from other members of the family.
EZH2	2146	Polycomb-group (PcG) family. Catalytic component of PRC2 complex (H3K27me3). Transcriptional coactivator.
PHRF1	57661	PHD And Ring Finger Domains 1
RBBP7	5931	Ubiquitous WD-40 repeat protein. Member of many epigenetic complexes (PRC2, SIN3A, DREAM, NuRD, NURF, CAF-1) .
KAT6A	7994	Histone acetyltransferases (H3kAc) and co-activator for several transcription factors.
ING1	3621	This gene encodes a tumor suppressor protein that can induce cell growth arrest and apoptosis. Direct binder of p53 and critical component in p53-signalling
PRDM11	56981	PR/SET Domain 11; Gene Ontology (GO) annotations related to this gene include nucleic acid binding and methyltransferase activity
DPY30	84661	Integral subunit of the SET1/MLL family (H3K4 methyltransferases). Directly controls cell-cycle regulators
KMT5C	84787	H4 'Lys-20' trimethylation represents a specific tag for epigenetic transcriptional repression. Mainly functions in pericentric heterochromatin regions, thereby playing a central role in the establishment of constitutive heterochromatin in these regions.
CHD1	1105	The CHD family of proteins is characterized by the presence of chromo (chromatin organization modifier) domains and SNF2-related helicase/ATPase domains. CHD genes alter gene expression possibly by modification of chromatin structure thus altering access of the transcriptional apparatus to its chromosomal DNA template
BRCA2	675	Mutations in this protein confer increased lifetime risk of developing breast or ovarian cancer. Maintenance of genetic stability and DNA repair
EP400	57634	Component of the NuA4 histone acetyltransferase complex-- transcriptional activation by acetylation of nucleosomal histones H4 and H2A.
CHD6	84181	DNA-dependent ATPase that increases accessibility of chromatin in a non-sliding manner.
USP51	158880	Among its related pathways are Ubiquitin-Proteasome Dependent Proteolysis.
PRDM8	56978	Histone methyltransferases that predominantly act as negative regulators of transcription. The encoded protein contains an N-terminal Su(var)3-9, Enhancer-of-zeste, and Trithorax (SET) domain and a double zinc-finger domain.
EPC2	26122	Polycomb group (PcG) family. The encoded protein is a component of the NuA4 histone acetyltransferase complex and can act as both a transcriptional activator and repressor. The encoded protein has been linked to apoptosis, DNA repair, differentiation, and gene silencing
ELP4	26610	Component of the six subunit elongator complex, a histone acetyltransferase complex that associates directly with RNA polymerase II during transcriptional elongation

Table 3.3 mCherry-EGFP- Hits Selected for Secondary Screening.

Primary hits ordered for secondary screening. Please note that hits in this section were selected by ability to increase or decrease the population to a significant value by a Student's two-tailed T-Test ($P < 0.05$). This table represents negative hits, which increased or decreased the negative sub-population. All information concerning protein function was summarized from Pubmed gene sites for each protein and does not reflect the full scope of activities and protein-protein interactions (Accessed October 10th, 2020)

Gene	Gene ID	Function
PRMT5	10419	Arginine methyltransferase. Targets histones, TEFs, and p53. Attentuates MAPK1 signalling
RAG2	5897	Required for RAG complex activity. It probably acts as a sensor of chromatin state that recruits the RAG complex to H3K4me3
RNF20	56254	E3 ubiquitin ligase. monoubiquitinating histone H2B. Likely tumor suppressor. Positively regulates the p53. Suppresses protooncogenes and
SCML4	256380	SCMH1 (Scm Polycomb Group Protein Homolog 1) is a Protein Coding gene. Among its related pathways are Metabolism of proteins and SUMOylation. Gene Ontology (GO) annotations related to this gene include DNA-binding transcription factor activity and sequence-specific DNA binding
SFMBT1	51460	Histone-binding protein. Recruits corepressor complexes--chromatin compaction.
TDRKH	11022	Involved in primary piRNA biogenesis pathway. Represses transposable elements and prevents their mobilization.
TRIM24	8805	E3 Ubiquitin LigaseTranscriptional coactivator. Interacts with nuclear receptors, coactivators, and modifies gene transcription. Has higest affinity for H3k4me0 and H3kAC. Promotes ubiquitination and proteasomal degradation of p53/TP53, thus plays a role in apoptosis. Up-regulates . Modulates transcription activation by retinoic acid (RA) receptors.
TRIM28	10155	Nuclear corepressor for KRAB zinc finger proteins. Recruits deacetylation (NuRD) complex, and SETDB1 (H3K9me) to KRAB-target genes. Inhibits E2F1 activity.
UBE2U	148581	Ubiquitin Ligase.
UBR2	23304	E3 ubiquitin-protein ligase which is a component of the N-end rule pathway, which leads to protein degradation. Chromatin inactivation. Controls cell growth by negatively regulating mTOR pathway.
KDM4C	23081	Jumonji domain 2 (JMJD2) family. Demethylase (H3kme2-->H3Kme2).
ASH2L	9070	Transcriptional regulator. Involved in Set1/ASH2 histone methyltransferase complex (H3kme1). Involved in Mll1 complex (H3Kme2). Stimulates activity of HMT activities of KMT2A, KMT2B, KMT2C, KMT2D, SETD1A and SETD1B (PubMed:21220120, PubMed:22266653)
RNF20	56254	E3 ubiquitin ligase. monoubiquitinating histone H2B. Likely tumor suppressor. Positively regulates the p53. Suppresses protooncogenes and growth-related genes related to EGF signalling. Component of the RNF20/40 E3 ubiquitin-protein ligase complex (H2BK120ub1--activation mark). H2BK120ub1 is a prerequisite ffor H3K4me and H3K79me
TDRD3	81550	Scaffolding protein. Binds dimethylarginine-containing proteins. Coactivator: recognizes and binds asymmetric dimethylation (H3R17me2a and H4R3me2a) and recruits proteins.
SETDB1	9869	Trimethylase (H3k9me3) s-9' of histone H3. Mainly functions in euchromatin regions, thereby playing a central role in the silencing of euchromatic genes. However, can be repressive by TRIM28/TIF1B, a factor recruited by KRAB zinc-finger proteins.
KANSL3	55683	KANSL3 (KAT8 Regulatory NSL Complex Subunit 3). Chromatin organizer. Likely involved with HATs.
BRMS1L	84312	Involved in HDAC1-stranscritional repression. This protein is a component of the SIN3a family of histone deacetylase complexes
SRCAP	10847	Core catalytic component of the multiprotein chromatin-remodeling SRCAP complex. Necessary for H2A.Z incorporation. Can function as a transcriptional activator in Notch-mediated, CREB-mediated and steroid receptor-mediated transcription
DNMT3L	29947	Catalytically inactive regulatory factor of DNA methyltransferases that can either promote or inhibit DNA methylation depending on the context. Prevents additional H3K27me3 marks by binding to PRC2 complex.
ACTL6B	51412	Chromatin Remodeler. Component of SWI/SNF chromatin remodeling complexes
SMYD2	56950	Protein-lysine N-methyltransferase that methylates both histones and non-histone proteins, including p53/TP53 and RB1. Trimethylates
PHF8	23133	Histone lysine demethylase. (H3K9me1/2, H3k20me1, H3k27me2). Thus, transcriptional activator. Controls cell cycle progression
MCRS1	10445	Modulates the transcription repressor activity of DAXX by recruiting it to the nucleolus (PubMed:11948183). As part of the NSL complex it may be involved in acetylation of nucleosomal histone H4 on several lysine residues.
KAT2A	2648	Protein lysine acyltransferase that can act as a acetyltransferase, glutaryltransferase or succinyltransferase, depending on the context (PubMed:29211711). Catalyzes activating succinyltransferation (H3k79succ) with high frequency around transcription start sites.
EPC1	80314	Polycomb group (PcG) family. The encoded protein is a component of the NuA4 histone acetyltransferase complex and can act as both a transcriptional activator and repressor. The encoded protein has been linked to apoptosis, DNA repair, differentiation, and gene silencing.
MEN1	4221	Scaffold protein that functions in histone modification and epigenetic gene regulation. Essential component of a MLL/SET1 histone methyltransferase (HMT) complex (H3Kme3), a complex that specifically methylates 'Lys-4' of histone H3 (H3K4). Binds to the TERT promoter and represses telomerase expression
KMT2D	8085	The protein encoded by this gene is a histone methyltransferase that methylates the Lys-4 position of histone H3. The encoded protein is part of a large protein complex called ASCOM, which has been shown to be a transcriptional regulator of the beta-globin and estrogen receptor genes
TRIM24	8805	E3 Ubiquitin LigaseTranscriptional coactivator. Interacts with nuclear receptors, coactivators, and modifies gene transcription. Has higest affinity for H3k4me0 and H3kAC. Promotes ubiquitination and proteasomal degradation of p53/TP53, thus plays a role in apoptosis. Up-regulates . Modulates transcription activation by retinoic acid (RA) receptors.
DCAF1	9730	Substrate recognition for E3-ubiq-prot-ligases. Ubiquitination activity recruited by RAG1. Ubuiquitates TERT. Represses transcription by phosphorylation of (H2AT120P). Involved in cell cycle
UBE2U	148581	Ubiquitin Ligase.

Table 3.4 mCherry-EGFP+ Hits Selected for Secondary Screening.

Primary hits ordered for secondary screening. Please note that hits in this section were selected by ability to increase or decrease the population to a significant value by a Student's two-tailed T-Test ($P < 0.05$). This table represents epithelial hits, which increased or decreased the epithelial sub-population. All information concerning protein function was summarized from Pubmed gene sites for each protein and does not reflect the full scope of activities and protein-protein interactions (Accessed October 10th, 2020)

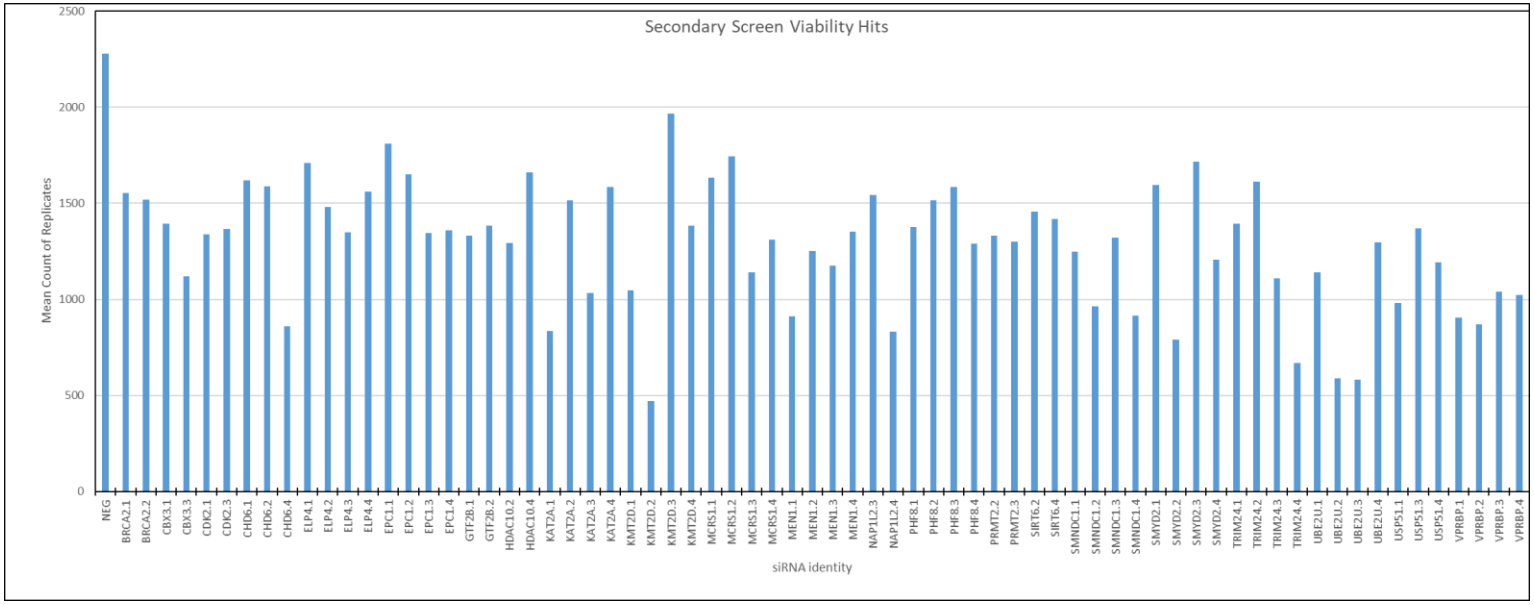

Secondary Screen Viability Hits
Mean Count of Replicates
2500
2000
1500
1000
500
0
siRNA identity
NEG
BRCA2.1
BRCA2.2
CBX3.1
CBX3.3
CDK2.1
CDK2.3
CHD6.1
CHD6.2
CHD6.4
CHD6.4
ELP4.1
ELP4.2
ELP4.3
ELP4.4
EPC1.1
EPC1.2
EPC1.3
EPC1.4
GTF2B.1
GTF2B.2
HDAC10.2
HDAC10.4
KAT2A.1
KAT2A.2
KAT2A.3
KAT2A.4
KMT2D.1
KMT2D.2
KMT2D.3
KMT2D.4
MCRS1.1
MCRS1.2
MCRS1.3
MCRS1.4
MEN1.1
MEN1.2
MEN1.3
MEN1.4
NAP1L2.3
NAP1L2.4
PHF8.1
PHF8.2
PHF8.3
PHF8.4
PRMT2.2
PRMT2.3
SIRT6.2
SIRT6.4
SMNDC1.1
SMNDC1.2
SMNDC1.3
SMNDC1.4
SMYD2.1
SMYD2.2
SMYD2.3
SMYD2.4
TRIM24.1
TRIM24.2
TRIM24.3
TRIM24.4
UBE2U.1
UBE2U.2
UBE2U.3
UBE2U.4
USP51.1
USP51.3
USP51.4
VPRBP.1
VPRBP.2
VPRBP.3
VPRBP.4

Figure 3.1 Viability Hits Identified from Secondary Screening Alone.

These are the hits that would be selected by secondary screening alone. Each meets the criteria of a decrease in population ($p < 0.05$) by a single tailed Student's t-test and at least two single siRNA possessing the same effect. Three replicates per data group. No consideration was given to previous effect in the primary screen. Y-axis measures the total number of cells measured by the mean of the gene target.

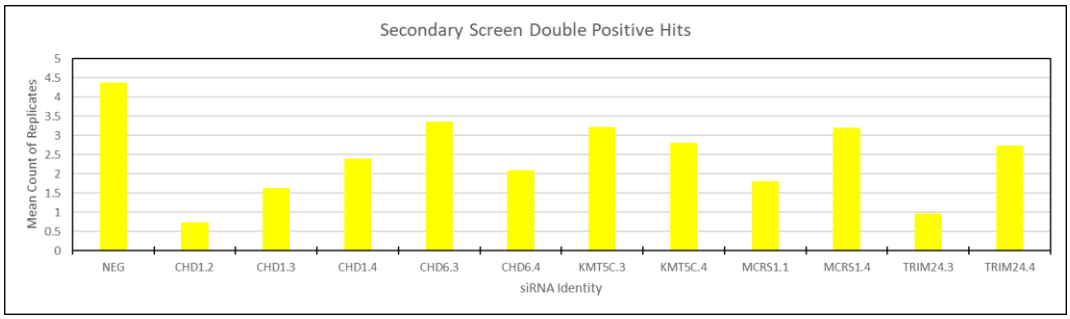

Secondary Screen Double Positive Hits
Mean Count of Replicates
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
NEG
CHD1.2
CHD1.3
CHD1.4
CHD6.3
CHD6.4
KMT5C.3
KMT5C.4
MCRS1.1
MCRS1.4
TRIM24.3
TRIM24.4
siRNA Identity

Figure 3.2 Double Positive Hits Identified from Secondary Screening Alone.

These are the hits that would be selected by secondary screening alone. Each meets the criteria of a decrease in population ($p < 0.05$) by a single tailed Student's t-test and at least two single siRNA possessing the same effect. Three biological replicates per group. No consideration was given to previous effect in the primary screen.
The y-axis represents the average sub-population percentage of each of the three wells for each siRNA (shown on the x-axis)

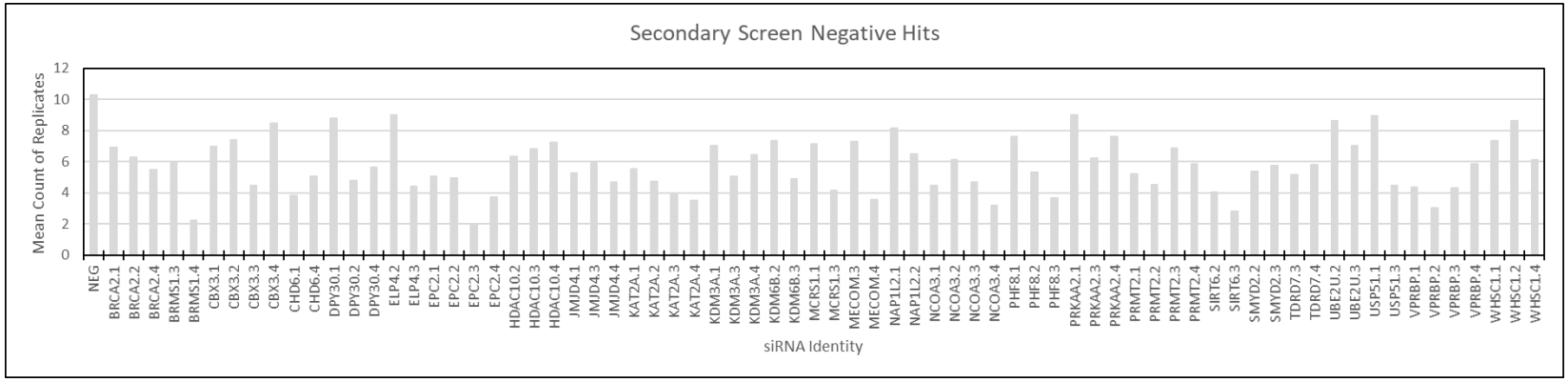

Secondary Screen Negative Hits
Mean Count of Replicates
0
2
4
6
8
10
12
NEG
BRCA2.1
BRCA2.2
BRCA2.4
BRMS1.3
BRMS1.4
CBX3.1
CBX3.2
CBX3.3
CBX3.4
CHD6.1
CHD6.4
DPY30.1
DPY30.2
DPY30.4
ELP4.2
ELP4.3
EPC2.1
EPC2.2
EPC2.3
EPC2.4
HDAC10.2
HDAC10.3
HDAC10.4
JMJD4.1
JMJD4.3
JMJD4.4
KAT2A.1
KAT2A.2
KAT2A.3
KAT2A.4
KDM3A.1
KDM3A.3
KDM3A.4
KDM6B.2
KDM6B.3
MCRS1.1
MCRS1.3
MECOM.3
MECOM.4
NAP1L2.1
NAP1L2.2
NCOA3.1
NCOA3.2
NCOA3.3
NCOA3.4
PHF8.1
PHF8.2
PHF8.3
PRKAA2.1
PRKAA2.3
PRKAA2.4
PRMT2.1
PRMT2.2
PRMT2.3
PRMT2.4
SIRT6.2
SIRT6.3
SMYD2.2
SMYD2.3
TDRD7.3
TDRD7.4
UBE2U.2
UBE2U.3
USP51.1
USP51.3
VPRBP.1
VPRBP.2
VPRBP.3
VPRBP.4
WHSC1.1
WHSC1.2
WHSC1.4
siRNA Identity

Figure 3.3 Negative Hits Identified from Secondary Screening Alone.

These are the hits that would be selected by secondary screening alone. Each meets the criteria of a decrease in population ($p < 0.05$) by a single tailed Student's t-test and at least two single siRNA possessing the same effect. Three biological replicates per gene target. No consideration was given to previous effect in the primary screen.
The y-axis represents the average sub-population percentage of each of the three wells for each siRNA (shown on the x-axis)

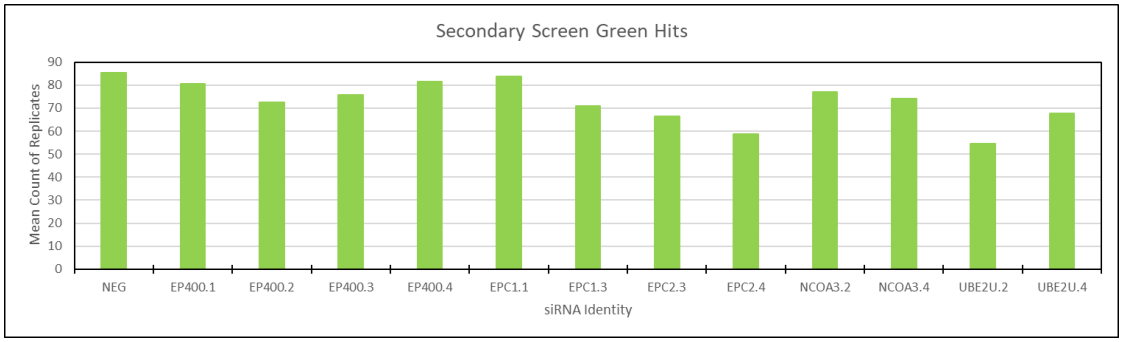
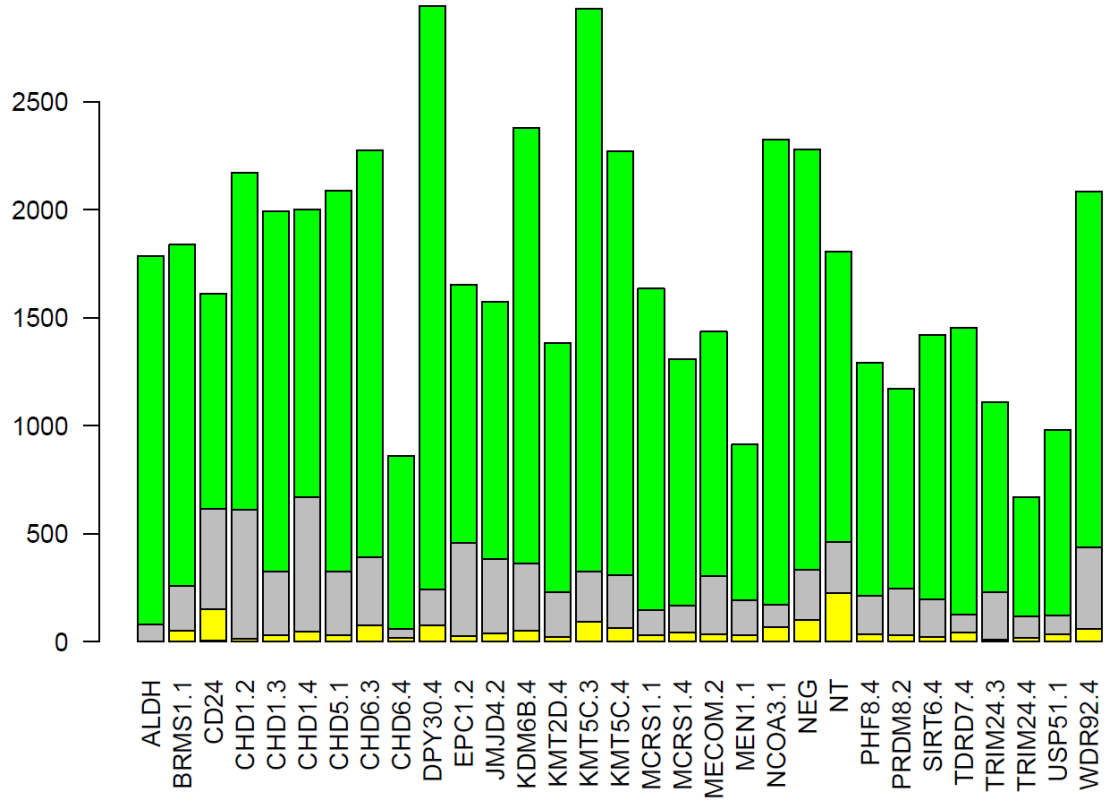

Secondary Screen Green Hits
Mean Count of Replicates
90
80
70
60
50
40
30
20
10
0
NEG
EP400.1
EP400.2
EP400.3
EP400.4
EPC1.1
EPC1.3
EPC2.3
EPC2.4
NCOA3.2
NCOA3.4
UBE2U.2
UBE2U.4
siRNA Identity

Figure 3.4 Epithelial-like Hits Identified from Secondary Screening Alone.

These are the hits that would be selected by secondary screening alone. Each meets the criteria of a decrease in population ($p < 0.05$) by a single tailed Student's t-test and at least two single siRNA possessing the same effect. Three biological replicates per gene target. No consideration was given to previous effect in the primary screen.
The y-axis represents the average sub-population percentage of each of the three wells for each siRNA (shown on the x-axis)

A


2500
2000
1500
1000
500
0
ALDH
BRMS1.1
CD24
CHD1.2
CHD1.3
CHD1.4
CHD5.1
CHD6.3
CHD6.4
DPY30.4
EPC1.2
JMJD4.2
KDM6B.4
KMT2D.4
KMT5C.3
KMT5C.4
MCRS1.1
MCRS1.4
MECOM.2
MEN1.1
NCOA3.1
NEG
NT
PHF8.4
PRDM8.2
SIRT6.4
TDRD7.4
TRIM24.3
TRIM24.4
USP51.1
WDR92.4

B

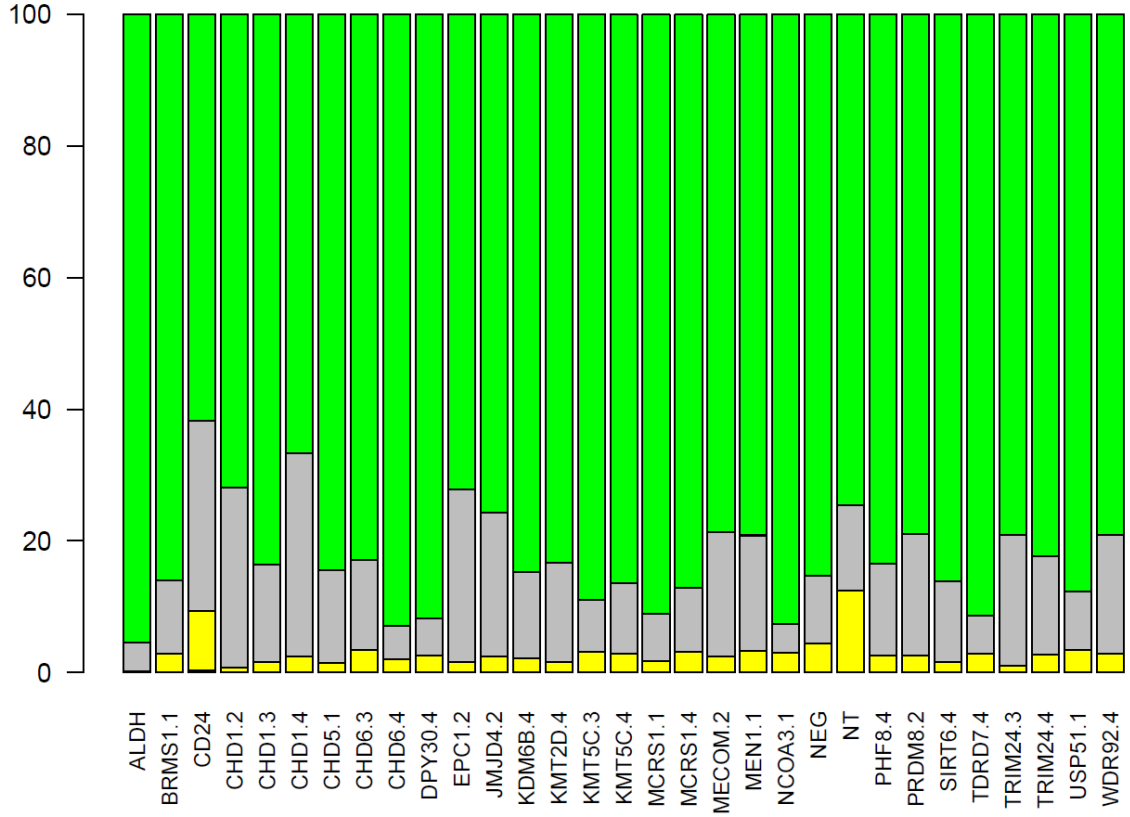
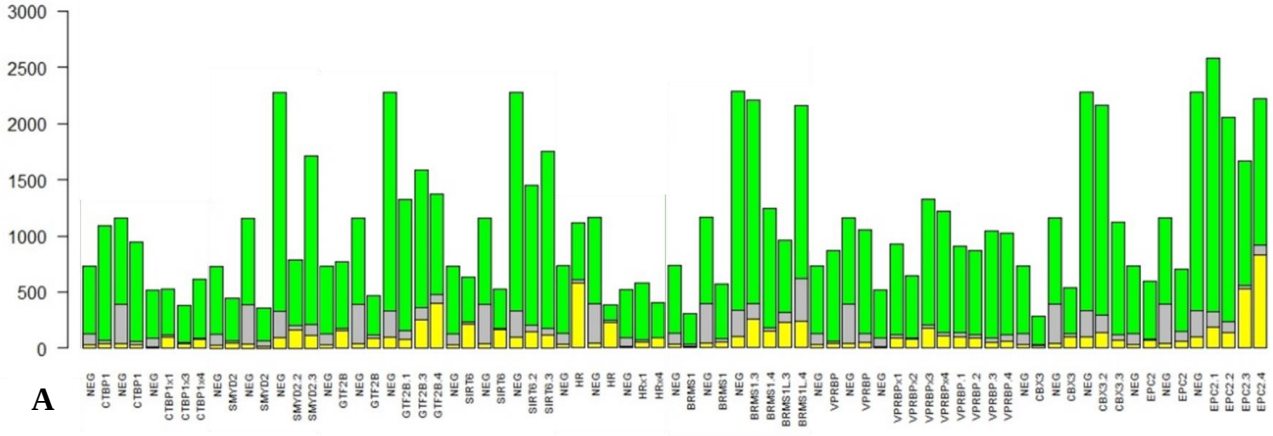

100
80
60
40
20
0
ALDH
BRMS1.1
CD24
CHD1.2
CHD1.3
CHD1.4
CHD5.1
CHD6.3
CHD6.4
DPY30.4
EPC1.2
JMJD4.2
KDM6B.4
KMT2D.4
KMT5C.3
KMT5C.4
MCRS1.1
MCRS1.4
MECOM.2
MEN1.1
NCOA3.1
NEG
NT
PHF8.4
PRDM8.2
SIRT6.4
TDRD7.4
TRIM24.3
TRIM24.4
USP51.1
WDR92.4

Figure 3.5 Highest Quality Hits of mCherry+EGFP+ Hits in Secondary Screening.

This double-positive data set is the exclusion to other analysis used for negative and epithelilal-like. Only secondary screening data are used for these graphs. Sub-numerals following the gene name denote which specific siRNA hit. These selected figures represent high confidence hits, which represent the top performing hits. None of the above gene targets met the $p < 0.05$ target by a single-tailed student t-test. Panel A represents the cell percentages when total cell count (as measured by DAPI-negative cells) is considered. Panel B considers only the percentages of each cell sub-population. Gray cells depict mCherry-EGFP-, yellow cells are mCherry+EGFP+, and Green cells are mCherry-EGFP+.


3000
2500
2000
1500
1000
500
0
A
NEG
CTBP1
NEG
CTBP1
NEG
CTBP1x1
CTBP1x3
CTBP1x4
NEG
SMYD2
NEG
SMYD2
NEG
SMYD2.2
SMYD2.3
NEG
GTF2B
NEG
GTF2B
NEG
GTF2B.1
GTF2B.3
GTF2B.4
NEG
SIRT6
NEG
SIRT6
NEG
SIRT6.2
SIRT6.3
NEG
HR
NEG
HR
NEG
HRx1
HRx4
NEG
BRMS1
NEG
BRMS1
NEG
BRMS1.3
BRMS1.4
BRMS1L.3
BRMS1L.4
NEG
VPRBP
NEG
VPRBP
NEG
VPRBPx1
VPRBPx2
VPRBPx3
VPRBPx4
VPRBP.1
VPRBP.2
VPRBP.3
VPRBP.4
NEG
CBX3
NEG
CBX3
NEG
CBX3.2
CBX3.3
NEG
EPC2
NEG
EPC2
NEG
EPC2.1
EPC2.2
EPC2.3
EPC2.4

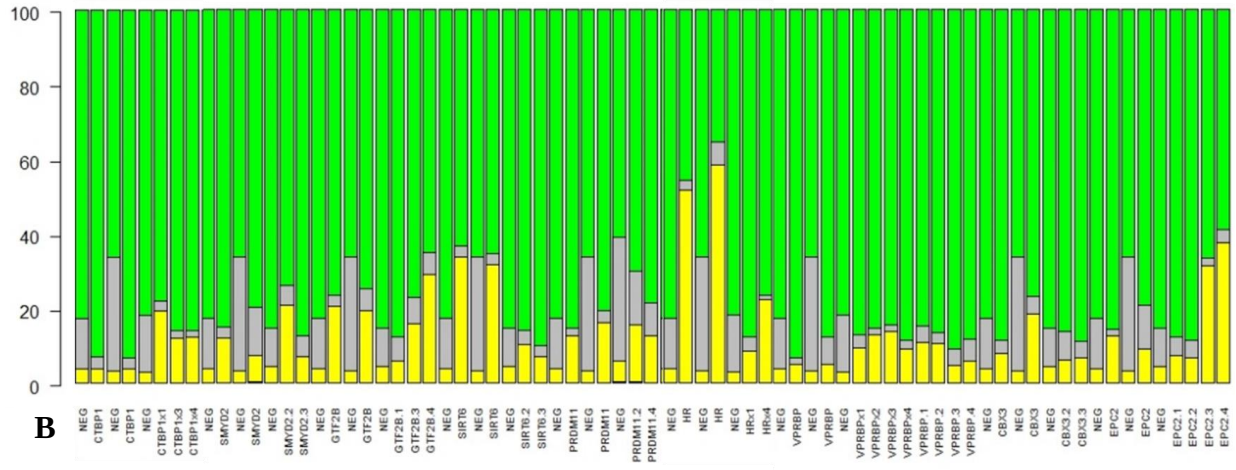

100
80
60
40
20
0
B
NEG
CTBP1
NEG
CTBP1
NEG
CTBP1x1
CTBP1x3
CTBP1x4
NEG
SMYD2
NEG
SMYD2
NEG
SMYD2.2
SMYD2.3
NEG
GTF2B
NEG
GTF2B
NEG
GTF2B.1
GTF2B.3
GTF2B.4
NEG
SIRT6
NEG
SIRT6
NEG
SIRT6.2
SIRT6.3
NEG
PRDM11
NEG
PRDM11
NEG
PRDM11.2
PRDM11.4
NEG
HR
NEG
HR
NEG
HRx1
HRx4
NEG
VPRBP
NEG
VPRBP
NEG
VPRBPx1
VPRBPx2
VPRBPx3
VPRBPx4
VPRBP.1
VPRBP.2
VPRBP.3
VPRBP.4
NEG
CBX3
NEG
CBX3
NEG
CBX3.2
CBX3.3
NEG
EPC2
NEG
EPC2
NEG
EPC2.1
EPC2.2
EPC2.3
EPC2.4

Figure 3.6 Integrated mCherry-EGFP- Hits Across Screening Campaign.

Each data set (Primary, Primary Repeat, and Secondary Screening) was combined into one R program and analyzed for continuity in effect. The order of each siRNA are Primary, Primary Repeat, and Secondary Screening hits in that order. Sub-numerals denote which specific siRNA hit. These selected figures represent high confidence hits, which represent the top performing hits by a one-tailed Student's t-test ($p < 0.05$). Panel A represents the cell percentages when total cell count (as measured by DAPI-negative cells) is considered. Panel B considers only the percentages of each cell sub-population. Gray cells depict mCherry-EGFP-, yellow cells are mCherry+EGFP+, and Green cells are mCherry-EGFP+.

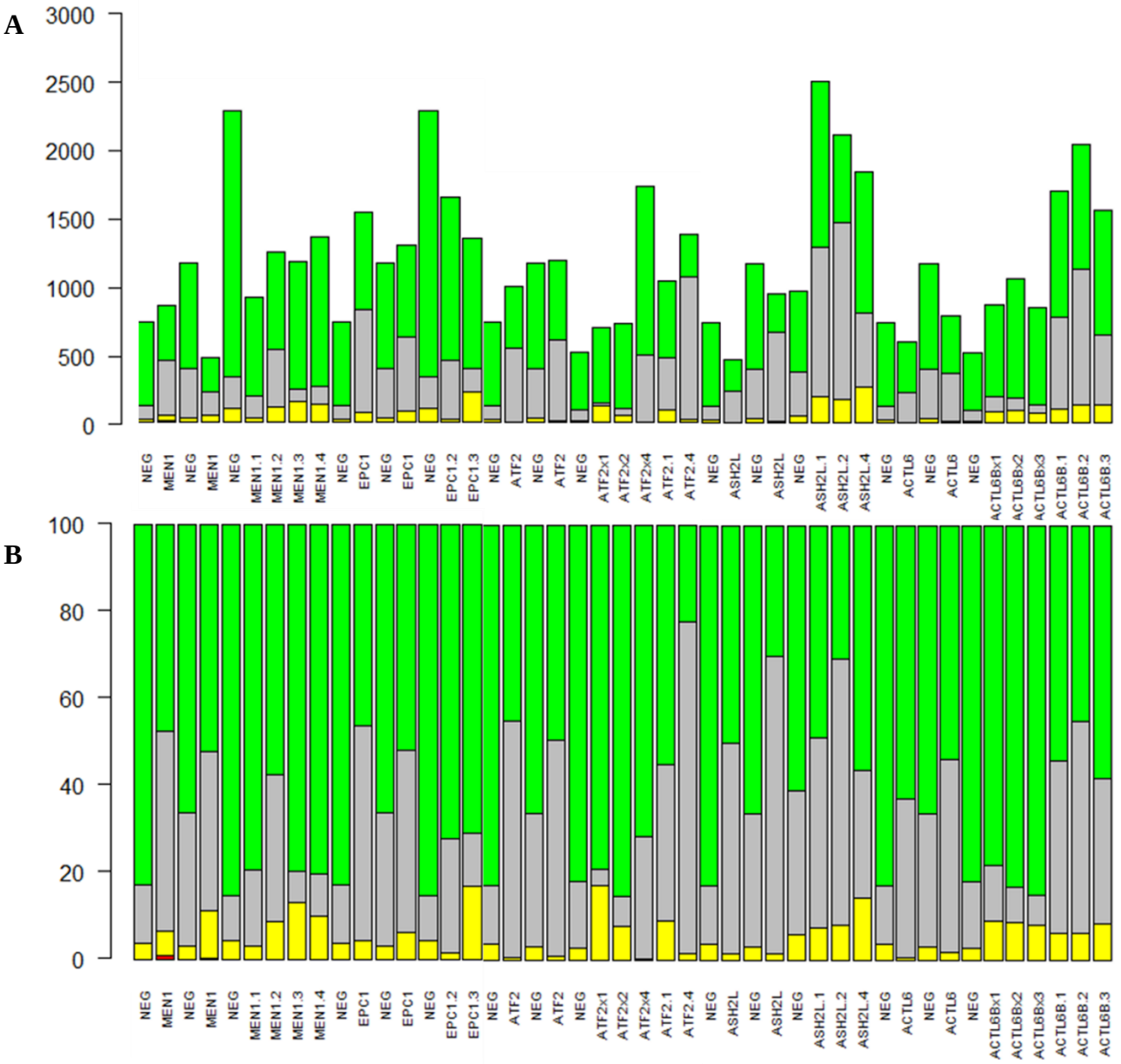

A
3000
2500
2000
1500
1000
500
0
B
100
80
60
40
20
0
NEG
MEN1
NEG
MEN1
NEG
MEN1.1
MEN1.2
MEN1.3
MEN1.4
NEG
EPC1
NEG
EPC1
NEG
EPC1.2
EPC1.3
NEG
ATF2
NEG
ATF2
NEG
ATF2×1
ATF2×2
ATF2×4
ATF2.1
ATF2.4
NEG
ASH2L
NEG
ASH2L
NEG
ASH2L.1
ASH2L.2
ASH2L.4
NEG
ACTL6
NEG
ACTL6
NEG
ACTL6B×1
ACTL6B×2
ACTL6B×3
ACTL6B.1
ACTL6B.2
ACTL6B.3

Figure 3.7 Integrated mCherry-EGFP+ Hits Across Screening Campaign.

Each data set (Primary, Primary Repeat, and Secondary Screening) was combined into one R program and analyzed for continuity in effect. The order of each siRNA are Primary, Primary Repeat, and Secondary Screening hits in that order. Sub-numerals denote which specific siRNA hit. These selected figures represent high confidence hits, which represent the top performing hits by a one-tailed Student's t-test ($p < 0.05$). Panel A represents the cell percentages when total cell count (as measured by DAPI-negative cells) is considered. Panel B considers only the percentages of each cell sub-population. Gray cells depict mCherry-EGFP-, yellow cells are mCherry+EGFP+, and Green cells are mCherry-EGFP+.

A

Gene Identity	Postive siRNA
BRCA2	2
CBX3	2
CDK2	2
CHD6	2
ELP4	4
EPC1	4
GTF2B	2
HDAC10	2
KAT2A	4
KMT2D	4
MCRS1	4
MEN1	4
NAP1L2	2
PHF8	4
PRMT2	2
SIRT6	2
SMNDC1	4
SMYD2	4
TRIM24	4
UBE2U	4
USP51	3
VPRBP	4

B

Gene Identity	Postive siRNA
BRCA2	3
CBX3	4
CHD6	2
DPY30	3
ELP4	2
EPC2	4
HDAC10	3
JMJD4	3
KAT2A	4
KDM3A	3
KDM6B	2
MCRS1	2
MECOM	2
NAP1L2	2
NCOA3	4
PHF8	3
PRKAA2	3
PRMT2	4
SIRT6	2
SMYD2	2
TDRD7	2
UBE2U	2
USP51	2
VPRBP	4
WHSC1	3

C

Gene Identity	Postive siRNA
CHD1	2
CHD6	2
KMT5C	2
MCRS1	2
TRIM24	2

D

Gene Identity	Postive siRNA
EP400	4
EPC1	2
EPC2	2
KAT2A	2
NCOA3	2
UBE2U	2

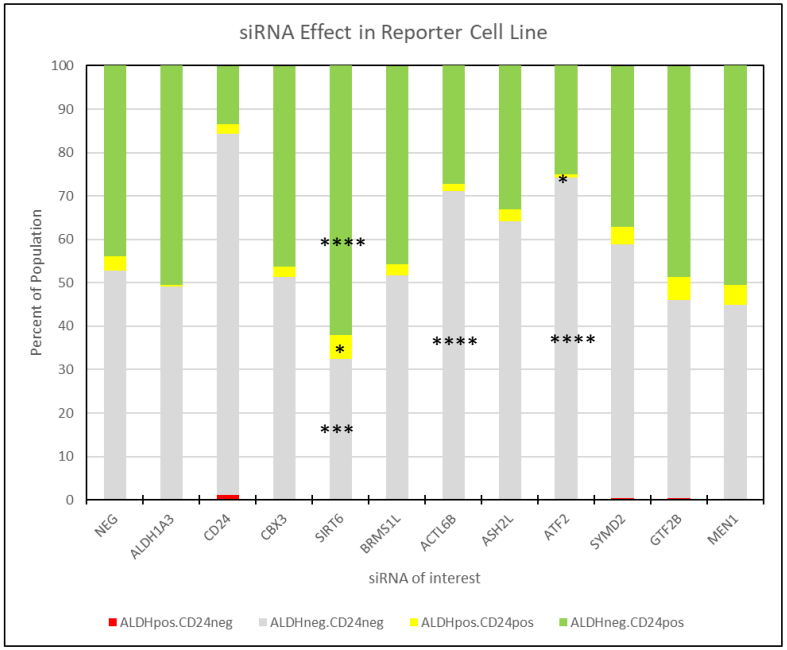
E

Identity	Viability	Negative	CD24	Double-positive
BRCA2	via	neg		
CBX3	via	neg		
CDK2	via			
CHD1				double
CHD6	via	neg		double
DPY30		neg		
ELP4	via	neg		
EP400			cd24	
EPC1	via		cd24	
EPC2		neg	cd24	
GTF2B	via			
HDAC10	via	neg		
JMJD4		neg		
KATA2A	via	neg	cd24	
KDM3A		neg		
KDM6B		neg		
KMT2D	via			
KMT5C				double
MCRS1	via	neg		double
MECOM		neg		
MEN1	via			
NAP1L2	via	neg		
NCOA3		neg	cd24	
PHF8	via	neg		
PRKAA2		neg		
PRMT2	via	neg		
SIRT6	via	neg		
SMNDC1	via			
SMYD2	via	neg		
TDRD7		neg		
TRIM24	via			double
UBE2U	via	neg	cd24	
USP51	via	neg		
VPRBP	via	neg		
WHSC1		neg		

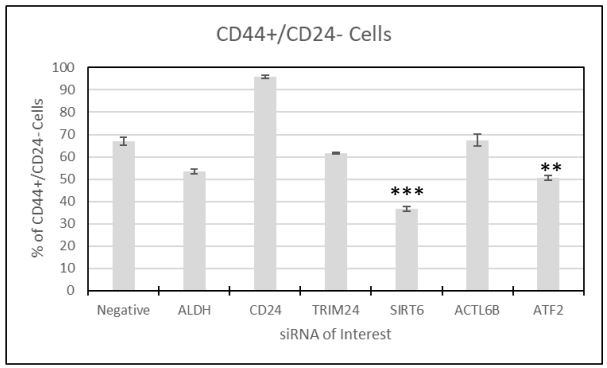
Table 3.5 Final Integrated Hit List Arranged by Most Impacted Sub-Population.

Panel A represents the Viability hits, Panel B represents the double positive hits, Panel C the negative hits, Panel D the Green hits. Each of these tables includes the number of positive siRNA in the secondary screening. Panel E represents the same data but also includes the multiple categories that each siRNA gene target affects. Blue cells depict viability hits, gray cells depict mCherry-EGFP-, yellow cells are mCherry+EGFP+, and Green cells are mCherry-EGFP+.

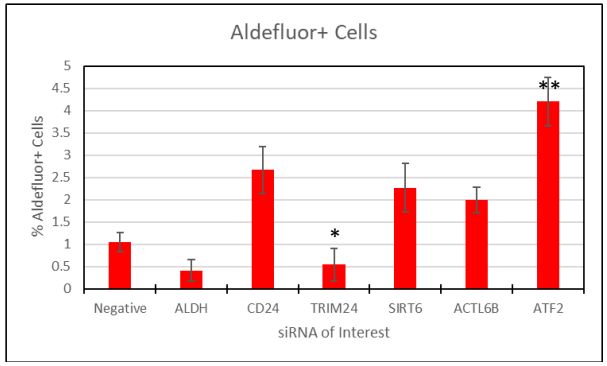
A



B



C



* p < 0.05
** p < 0.01
*** p < 0.001
**** p < 0.0001

Figure 3.8 Retesting of Select Hits from Integrated Hit List.

Panel A represents all chosen siRNA selected for retesting by fold change of effect. Those that were selected for future studies (SIRT6, ACTL6B, and ATF2) were further tested by traditional CD44, CD24, and Aldefluor assay along with TRIM24, shown in Panels B and C, respectively. Statistical significance is marked by asterisk(s) by a one-tailed student t-test ($p < 0.05$). Gray cells depict mCherry-EGFP-, yellow cells are mCherry+EGFP+, and Green cells are mCherry-EGFP+.

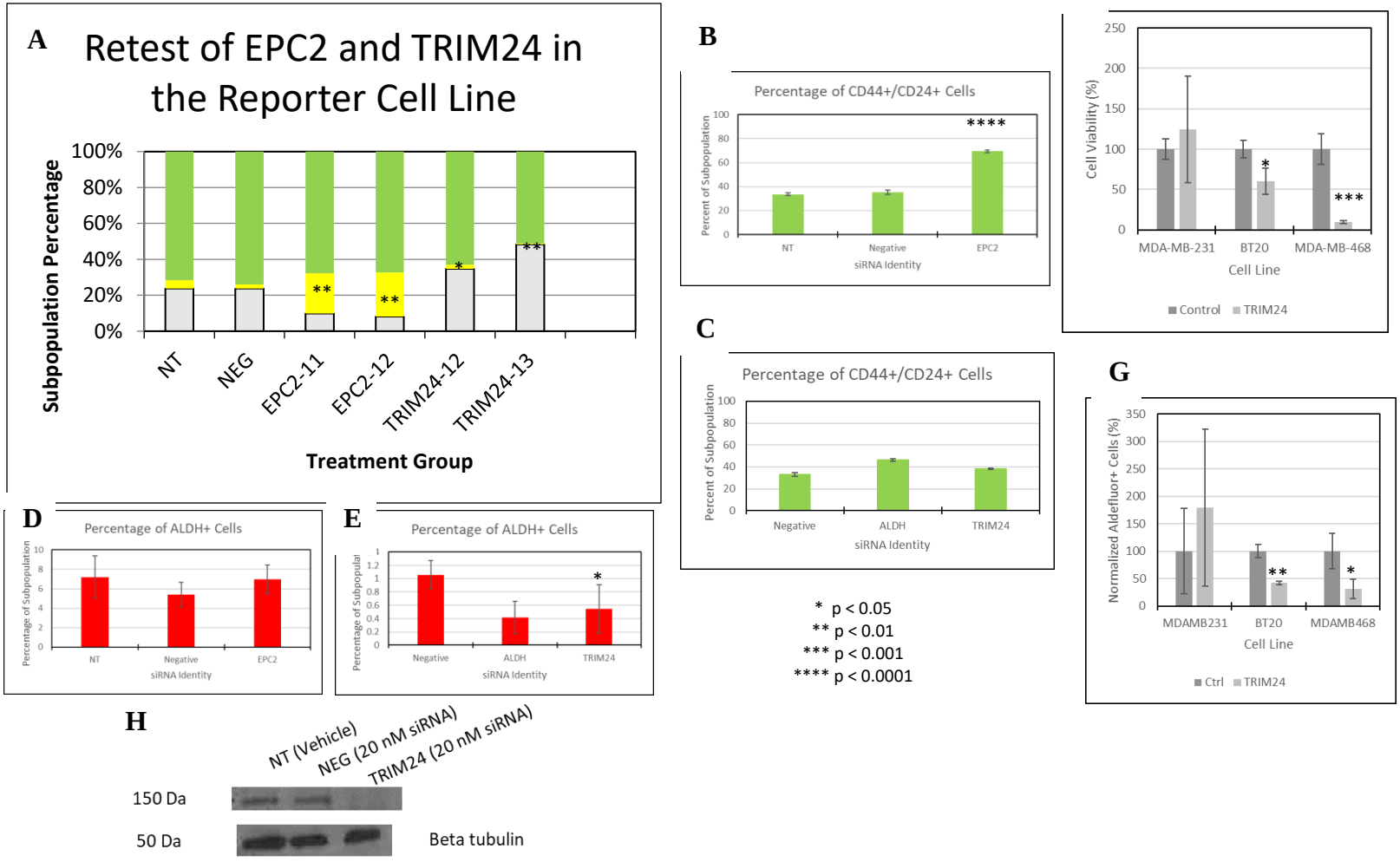
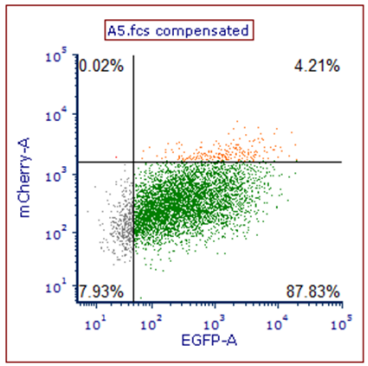

A
Retest of EPC2 and TRIM24 in the Reporter Cell Line
Subpopulation Percentage
100%
80%
60%
40%
20%
0%
NT
NEG
EPC2-11
EPC2-12
TRIM24-12
TRIM24-13
Treatment Group
B
Percentage of CD44+/CD24+ Cells
Percent of Subpopulation
NT
Negative
EPC2
siRNA Identity
C
Percentage of CD44+/CD24+ Cells
Percent of Subpopulation
Negative
ALDH
TRIM24
siRNA Identity
D
Percentage of ALDH+ Cells
Percentage of Subpopulation
NT
Negative
EPC2
siRNA Identity
E
Percentage of ALDH+ Cells
Percentage of Subpopulation
Negative
ALDH
TRIM24
siRNA Identity
F
Cell Viability (%)
MDA-MB-231
BT20
MDA-MB-468
Cell Line
Control
TRIM24
G
Normalized Aldefluor+ Cells (%)
MDAMB231
BT20
MDAMB468
Cell Line
Ctrl
TRIM24
* p < 0.05
** p < 0.01
*** p < 0.001
**** p < 0.0001
H
NT (Vehicle)
NEG (20 nM siRNA)
TRIM24 (20 nM siRNA)
150 Da
50 Da
Beta tubulin

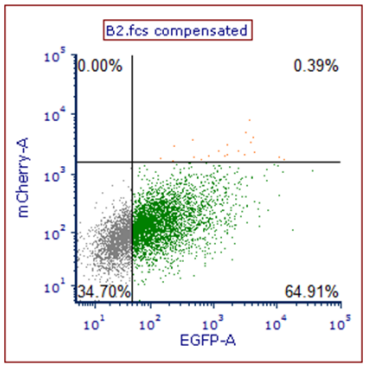
Figure 3.9 Retesting of TRIM24 and EPC2 in Expanded Cell Numbers and Multiple Cell Lines.

Panel A depicts EPC2 and TRIM24 knockdown using multiple siRNA as compared to no treatment and to negative siRNA. Panels B and C depict the effect EPC2 knockdown had in SUM149 using CD44 and CD24 antibody staining and using the Aldefluor™ assay kit. EPC2 did not recapitulate the expansion of undifferentiated cells. Panels D and E depict the effect TRIM24 knockdown had in SUM149 using CD44, CD24, and Aldefluor assay kit. It significantly reduced undifferentiated cells, but had little effect on differentiated epithelial cells. TRIM24 also had impact on cell viability and Aldefluor+ cells in both BT-20 and MDA-MB-468 TNBC cells but not MDA-MB-231 (Panels F and G). Panel H is a western blot showing TRIM24 knockdown in SUM149.

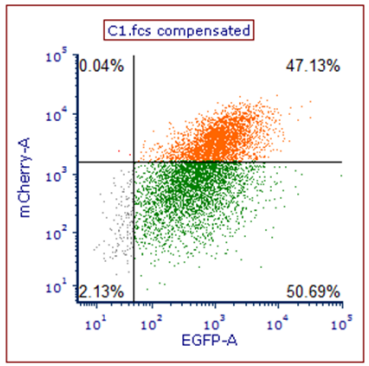
A



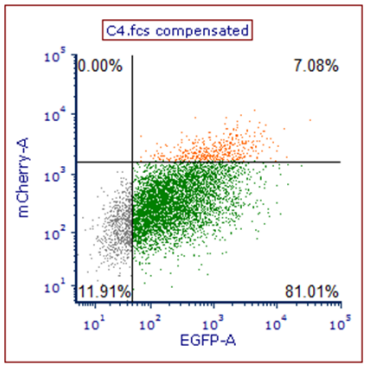
Neg siRNA



TRIM24 siRNA

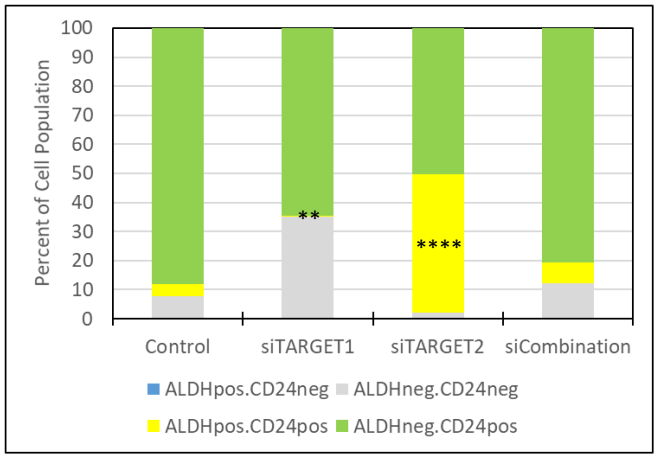


KMT2D siRNA

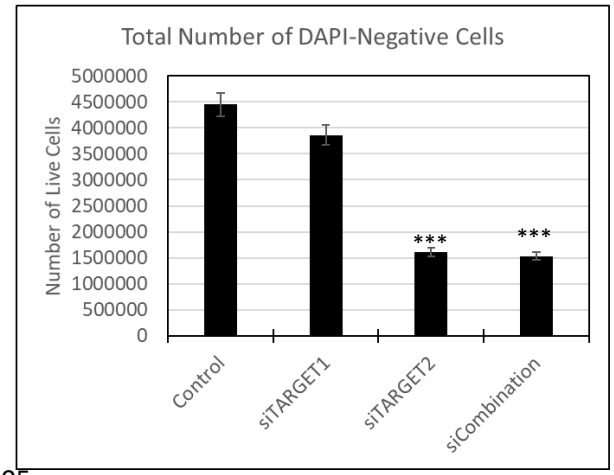


TRIM24 + KMT2D siRNA

B



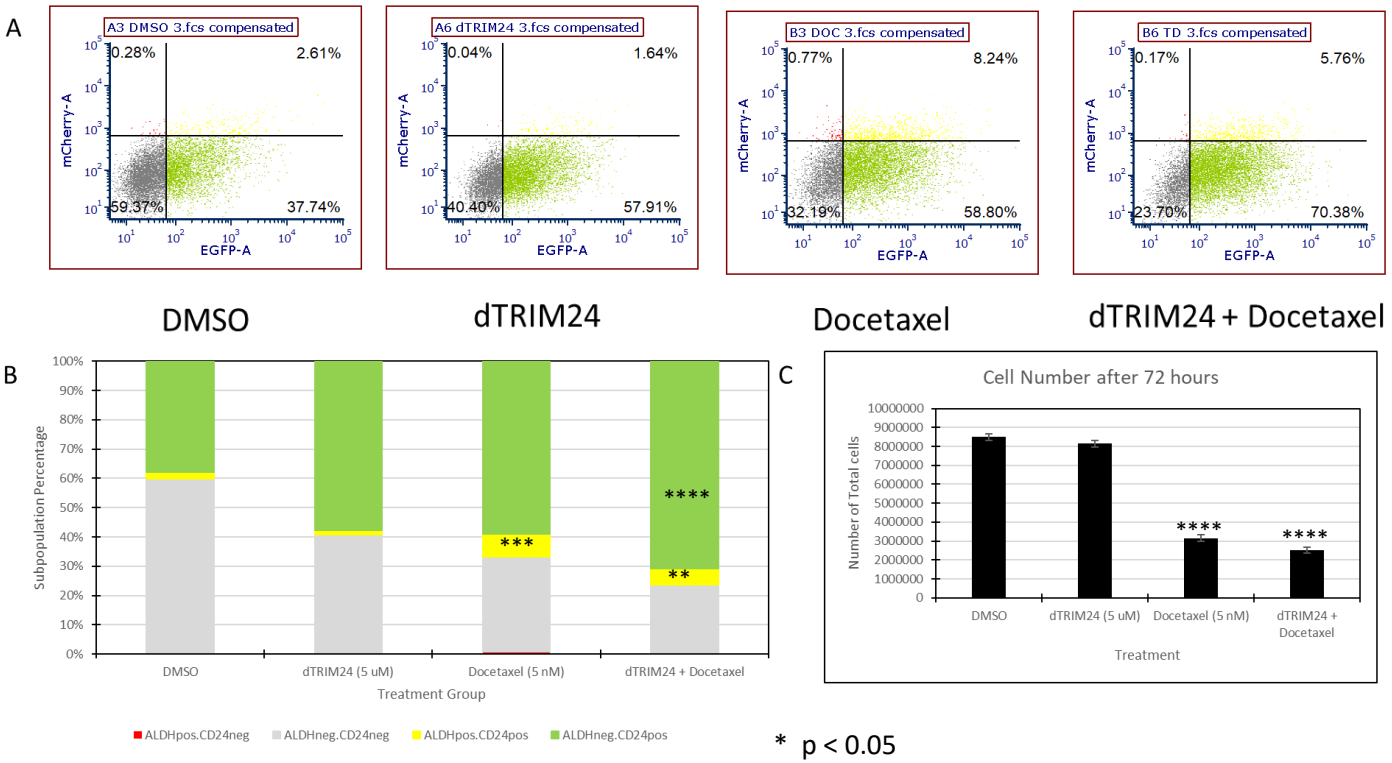
C



* p < 0.05
** p < 0.01
*** p < 0.001
**** p < 0.0001

Figure 3.10 Exploring Ability of TRIM24 siRNA to Decrease Induced Progenitor Cells.

Panel A depicts four representative treatments. The first plot is treated with Cell Signalling Non-Targeting siRNA (40 nM), the second has been treated with TRIM24 siRNA (20 nM), the third with KMT2D siRNA (20 nM), and the fourth has been treated with both TRIM24 and KMT2D (both 20 nM). Panel B depicts the average results of all relevant groups. Panel C depicts that although TRIM24 decreases progenitor cell expansion, it does not decrease the cytotoxic effect of KMT2D knockdown in the reporter cell line.



* p < 0.05
** p < 0.01
*** p < 0.001
**** p < 0.0001

Figure 3.11 Preliminary Studies into whether dTRIM24 can Attenuate Docetaxel-Induced Expansion of both Progenitor and Stem-like Cells.

Panel A depicts four representative treatments. The first plot is treated with DMSO, the second has been treated with dTRIM24 (5 μM), the third with Docetaxel (5 nM), and the fourth has been treated with both dTRIM24 and Docetaxel (5 μM and 5 nM respectively). Cells were treated for 72 hours. Panel B depicts the average results of all relevant groups. Panel C depicts that although dTRIM24 attenuates stem and progenitor-like cell expansion, it does not diminsh the cytotoxic effect of Docetaxel treatment in the reporter cell line.

Chapter 4

Summary of Findings and Future Directions

Heterogeneity remains a critical issue in many late-stage cancers, namely Glioblastoma multiforme (GBM) and triple negative breast cancer (TNBC). Glioblastoma multiforme features high levels of different cell types that arise by before detection (1, 2). So too is there variance based upon the original cell type or types from which the cancer developed, and these genetic differences influence the heterogeneity that arises (3-5). With GBM, there is no discussion among caregivers of cures, only a treatment course and certainty: resection if possible, then radiation and chemotherapy, followed by chemotherapy alone (6, 7). The unfortunate fate of those with late-stage GBM is death, and those diagnosed with metastatic triple-negative breast cancer (TNBC) meet that same fate. Due to their being no single target, chemotherapies are used (8-10). A portion of these patients are cured, a portion experience a decline in the tumor size, but ultimately the cancer will return and it will return stronger and more resistant (11, 12).

In TNBC and other late-stage cancers, cellular heterogeneity is deadly for several reasons. One, not all cells may respond to treatment the same (13-15). Two, variations in gene expression may permit different metastatic sites, such as bone, brain, lung, or liver (16, 17). Three, fully resistant cells may be present within the morass and serve as the vanguard of a renewed tumor (18-20). While there is of course heterogeneity between patients (intertumoral heterogeneity), there is also heterogeneity within the tumor (intratumoral heterogeneity) (21-25). Underlying intratumoral heterogeneity are three sources: genetic, epigenetic, and the microenvironment.

Genetic heterogeneity can determine any future hierarchy and mutational programs within the final, and evolving, tumor (26). Already due to transcriptional changes and chromosomal

abnormalities, additional genetic mutations are all but certain (18). However, initial gene expression programs are not the total of heterogeneity. Epigenetic mechanisms determine which genes are expressed within a given timeframe. These mechanisms can silence, overexpress, and otherwise alter these gene expression programs (27). These epigenetic proteins also have capabilities beyond canonical functions and can behave as transcriptional activators or repressors by protein-protein interactions (28, 29). Both genetic and epigenetic programs ultimately determine the transcriptional and translational profiles of all cells within the tumor, further enhancing the ability of these cells to evade and overcome treatment attempts (30, 31). Lastly, there is heterogeneity due to the microenvironment, which may be further classified by tumor-associated immune cells, by supporting non-cancerous fibroblasts that send growth and survival signals, or by other tumor cells themselves (32, 33).

Because intratumoral heterogeneity has been an ongoing problem, many groups have sought to investigate the sources or patterns of these sources, whether genetic or by the microenvironment (18, 24, 31, 34-37). However, there have been no efforts to discover the epigenetic origin of this heterogeneity. Such efforts would be critical since epigenetic processes likely lead to cancer cell adaptation and survival in the short term. In the literature and unpublished data, cancer cells not only can survive first line chemotherapies but also actually de-differentiate into stem-like and progenitor-like cells that have increased replication in this environment (13, 19). While ultimately genetic mutations will continue to arise and establish a resistant breed of tumor, in the short term these cells survive due to short term translational alterations due to epigenetics (14). While this de-differentiation is the method seen in lab, other transitions to more resistant strains may occur in patients, whether that be vertically with de-differentiation or horizontally between differentiated states.

Therefore, I optimized and executed a high-throughput siRNA screen against 505 epigenetic gene targets. Because it is difficult to screen for heterogeneity changes using traditional antibody methods, I used a fluorescent reporter cell line that modeled this same heterogeneity in real time. This cell line coexpresses mCherry with ALDH1A3 and EGFP with CD24, the former being a marker of de-differentiated cells and the latter an epithelial marker (38, 39). I optimized each facet of the screen as with any high-throughput project (positive and negative controls, cell seeding, transfection reagent, time in well, mode of transfection, mode of data analysis, and mode of selecting hits). Both primary and secondary screenings were conducted, the former using pooled

siRNA and the latter using single siRNA for the top-120 targets from the primary screen. Throughout the screen, thirty-five siRNA had a consistent and significant effect in at least one of the sub-populations studied, whether differentiated epithelial or mesenchymal, epithelial progenitors, or viability. Of these, nineteen had at least two effects, and of these three had three effects. These results exemplify why it is difficult to target a given sub-population by targeting a key epigenetic protein—the results will affect all cells within the population and affect them differently. The preponderance of viability effects (22/35) illustrates that fact, where there was a great deal of overlap with each of the other populations. Only four gene targets had a significant effect on viability without also decreasing another sub-population. It may be that the process of transfection always carries a cost beyond which negative siRNA can control. It may also be disrupted epigenetic processes, whether acetylation or deacetylation, ubiquitination, or the addition or removal of methylation, cell viability will be compromised to some degree. However, the current effort isn't focused on decreasing cell numbers but rather affecting what those cells currently are.

Of the seven gene targets that managed to repeat in follow-up studies, two were selected for future investigation: TRIM24 and EPC2. TRIM24, an E3 ubiquitin ligase, both increased EGFP- cells and decreased $mCherry^{+}EGFP^{+}$ cells in the reporter cell line; however, the former effect did not repeat in any other genetic context (40). It also had an appreciable effect on cellular proliferation. EPC2 was selected due to its effect in increasing $EGFP^{+}$ cells, which repeated in SUM149 as $CD44^{+}/CD24^{+}$ cells.

With an eye towards translational impact, I focused on what combination of siRNA could replicate what occurs with docetaxel and other chemotherapies in TNBC. They increase the number of stem-like cells while killing most other cells, and this increase is due to increased symmetrical division of those cells. For this effect, I used KMT2D, a lysine methyltransferase, which increases $mCherry^{+}EGFP^{+}$ cells and affects viability dramatically. To counter that, I used TRIM24, which decreases $mCherry^{+}EGFP^{+}$ cells. In combination, these siRNA decreased viability without a massive increase in the progenitor-like populations.

I then investigated if the same process could be true for docetaxel. Docetaxel and other TNBC-specific chemotherapies enrich for stem-like cells following treatment, the same effect seen after KMT2D siRNA (13). For TRIM24 disruption, I instead used a previously developed TRIM24 degrader, dTRIM24 (41). In the reporter cell line, I found that dTRIM24 attenuated the docetaxel-

induced increase in less differentiated cells, both stem-like and epithelial-progenitor like. In this preliminary study, not only was there increased cytotoxicity, the less-differentiated cells did not expand to the extent that docetaxel alone cells did. While targeting TRIM24's non-functional bromodomain provides valuable information, there may be better targets that TRIM24 to prevent this rise in stem-like cells due to chemotherapy in TNBC, targets that TRIM24 may regulate. Future RNAseq or ssRNAseq may provide further information on what target or targets might specifically disrupt the expansion of less-differentiated cells. TRIM24's oncogenicity has been broadly documented against cancer types, and similar roles should be investigated within the context of TNBC (40, 42-55). Of the available data, TRIM24's role in consistently regulating TNBC stem-like and epithelial progenitor-like cells likely encompasses all its functions, including retinoic acid signaling, transcriptional coactivation and corepression, WNT/β-catenin signaling, and E3-ubiquitinase activity have a role to play in the revealed phenotype seen in three TNBC cell lines (43, 45, 51, 52, 54, 56, 57). Since TRIM24 knockdown consistently reduces the most dangerous cell sub-type in TNBC, further research is critical for what may be a solution to the ongoing problem posed by chemotherapy treatment in TNBC.

Future Studies

Lentiviral Construction to Confirm Protein Function

Validating protein knockdown requires less extensive effort than fully characterizing a novel and important protein. Typically in high-throughput screening, assay design grants some measure of certainty that a given siRNA will have a given effect consistently (58). For instance, this campaign's secondary screening used four high-quality siRNA strands against each of the 120 gene targets and necessary for each hit was that at least two strands must have the same effect, and furthermore, that effect must be consistent with the effect observed in the primary screen. Additionally, the primary screen used pooled siRNA but of a different class, siGenome from Dharmacon. That means by the time final integration of all screening data is conducted, there is reasonable confidence that protein knockdown can be successfully performed repeatedly using the

tools and assay protocols. However, consistent results with different tools only provides so much confidence.

Thus, western blots should be produced using monoclonal antibodies against the target. The procedure has been common and well-known practice for decades to visualize a given proteins levels in cellular lysate (59).

For both TRIM24 and EPC2, western blots will be produced to verify siRNA effect after 72 hours. In both cases, the siRNA is expected to significantly reduce protein levels within the chosen samples compared to controls. Additionally, this result adds further confidence to the correlation between protein levels and sub-population effects demonstrated by flow cytometry analysis. This experiment suggests that TRIM24's presence and activity is necessary to maintain stem-like cells in the reporter cells, SUM149, BT-20, and MDA-MB-468. Thus, removing TRIM24 from the cellular milieu has the opposite effect: that cell sub-population significantly decreases, comparable to ALDH1A3 siRNA in SUM149. Likewise, EPC2 appears to discourage the development of epithelial populations in SUM149 and the reporter cell line, and thus removing EPC2 increases the epithelial sub-population.

TRIM24 and EPC2 Lentivirus Production

However, visual display of protein levels can only confirm so much. Rather than a simple claim that TRIM24 siRNA decreases de-differented cell populations, the claim should be far stronger: TRIM24 is critical in stem-cell regulation and maintenance. In order for that much stronger claim to be made, additional experiments must be conducted using yet-to-be-made tools. Previously, a lack of protein produced a population-specific effect, such as a decrease in de-differentiated cells. The use of dTRIM24 assisted in making this conclusion in that it had the same effect as TRIM24 siRNA in decreasing undifferentiated cells. However, the stronger claim requires protein overexpression that should produce an effect opposite to protein knockdown (i.e. TRIM24 overexpression should increase the levels of de-differentiated cells). To accomplish that goal, the University of Michigan's Vector Core will produce two third generation lentiviruses to assay the effect of TRIM24 and EPC2 in TNBC cell lines and verify their effects in specific sub-populations.

Both lentiviruses share the same basic structure. Both are third generation lentiviruses derived from HIV-1 but have decreased ability to self-replicate. Additionally, three vectors are used, one containing the transgene, another the envelope (which also contains trophism marker VSVG), and another the packaging vectors. Each of these vectors were added to 293T cells in order to produce lentivirus into the supernatant. However, both lentiviruses possess different selection markers: L-TRIM24 with a near-IR protein, and L-EPC2 with a puromycin resistance gene. For TRIM24, the positive cells will be sorted by FACS by expression of the protein irFP713. Those cells positive for this selection marker will be used to establish a cell line augmented with TRIM24 overexpression. For EPC2, the construct contains a puromycin-resistance gene, and thus the cells will be subjected to several rounds of puromycin to kill off the non-transduced cells. Empty-vector controls, having no transgene but containing the selection vector, will be used in conjunction with both transgene-containing lentiviral constructs.

Regardless of selection method, the primary goal will be to verify protein function at the phenotypic level. Since the cell line to be transduced is SUM149, sub-population determination will be performed with antibodies against CD44 and CD24 plus the ALDEFLUOR™ assay kit as was previously done in confirmation experiments. If there is a sufficient amount of lentivirus remaining, the SUM149 reporter cell line will be transduced, and potentially BT-20 and or MDA-MB-468. The more cell lines that can be transduced with these transgenes, the more certain any positive or negative results will be.

L-TRIM24 is expected to perform the reverse function that TRIM24 siRNA performed. De-differentiated cells are expected to increase with increasing levels of TRIM24, and overall cell growth is expected to increase relative to non-targeting control. This protein level increase will need to be verified by affinity western blotting experiments.

For L-EPC2, mesenchymal-like cells are expected to increase in response to EPC2 overexpression. As preliminary data indicates that EPC2 knockdown decreased cell viability slightly, overexpression could result in no change or a slight increase in cell proliferation. This latter result can investigated by the MTS assay. As with TRIM24, overexpression will need to be verified by western blotting. Examination of stem cell pathways (such as WNT/NOTCH/β-catenin) can also be investigated by western blot as well, which should clarify how TRIM24 effects its alterations, either with increased or decreased protein levels.

These lentiviral-augmented cell lines will serve a purpose beyond verifying protein function. These modified cell lines will serve as a valuable elements for future experiments exploring the "why" underlying these sub-population alterations, such as single-cell RNAseq. This experiment is especially critical for TRIM24 as it is a known modifier of gene expression. While targeting TRIM24 for therapeutic intervention is unlikely using currently available tools, there may be a target regulated by TRIM24 that may be more amenable for further investigation.

Future Experiments with TRIM24

TRIM24 has varied roles as an E3 ubiquitin ligase, a transcriptional activator/repressor, and a regulator of key stem cell pathways . Accordingly, knockdown of TRIM24 decreases the levels of undifferentiated cells in multiple cell lines . While it is tempting to speculate on why this is due to previously published research, further investigation should be performed to determine how TRIM24 knockdown (or lentiviral-induced overexpression) alters gene expression patterns. Indeed, there has been no effort to discover which gene pathways TRIM24 regulates on a broad scale. Therefore, I propose to discover what mRNA are expressed by individual cells by using single cell RNA sequencing (ssRNAseq) following TRIM24 knockdown. Previously, ssRNAseq was performed on both the reporter cell line and parental SUM149 (unpublished data). Using this previous gene expression data, alterations can be attached to different sub-populations. Alternatively, flow cytometry can be used to sort out distinct populations and regular RNAseq can be performed on these groups to grant a deeper view of gene expression changes in all sub-populations following TRIM24 knockdown. These experiments can be expanded by the inclusion of groups that feature fellow TIF/TRIM member TRIM33 and TRIM24+TRIM33 protein knockdown. These proteins have been demonstrated to form heterodimers that regulate distinct gene targets (60). It is important to investigate whether this heterodimer regulates a different set of genes than either TRIM24 or TRIM33 alone. Regardless of sequencing experimental approach, any of these experiments can highlight both known stem cell regulating pathways (WNT/β-catenin/STAT3/retinoic acid signaling) that TRIM24 has previously been implicated in regulating in various cancer backgrounds (40, 52, 56, 57, 61).

Additionally, TRIM24 has been shown to negatively regulate p53 by its E3 ubiquitin ligase activity (62). However, there have been no other studies to determine what other proteins are negatively regulated by this ligase activity, and it is doubtful that p53 is the only target of TRIM24. Therefore,

I propose to discover protein targets of TRIM24's E3 ubiquitin ligase activity by protein MS/MS with or without protein knockdown. The absence of ubiquitin will be the marker for TRIM24's activity, but will need further verification to determine if it is ubiquitinated by TRIM24. Alternatively, point mutations may be introduced to the zinc-conjugating B-boxes on TRIM24, which has been shown to abrogate ubiquitin ligase activity in other TRIM family members, including TRIM33 (63, 64). Since there have been mutation studies regarding TRIM24 specifically, homology can be used to identify which cysteine residues should be converted to alanines.

Future Experiments with EPC2

EPC2 is an understudied polycomb protein and a likely member of the human NuA4 histone acetyl-ligase complex . Its main role is establishing hyperacetylated regions by targeting nucleosomes, specifically H4 and H2A . However, most research into NuA4 has focused on the paralogue EPC1, of which EPC2 lacks 29 amino acids on the C-terminus. Beyond that, however, EPC1 and EPC2 share a conserved N-terminus that interacted with NuA4 and other member proteins .

Considering EPC2 and TRIM24 have diverging phenotypes, yet both have some interaction with acetylated regions, more effort should be done in further characterizing the function of EPC2 within the context of NuA4. The lack of certain amino acids may impinge on some associations that EPC1 possesses, such as with a large part of the NuA4 complex and can only associate with a portion of the sub-units in the piccolo-NuA4 complex, which contains Esa1, the acetyl-ligase, and other member proteins Yng2 and Eaf6 . This indicates there may be target specificity involved between EPC1 and EPC2 within the context of NuA4 and piccolo-NuA4 activity. Both EPC1 and EPC2 are critical for target recognition . Therefore, there is likely some targeting difference based on differing amino acid composition. Therefore, I propose to discover the specific targeting regions of EPC2 that are acetylated by piccolo-NuA4 and not by EPC1. This will require sequencing using ATAC-seq following EPC1, EPC2, or EPC1 and EPC2 knockdown. Briefly, ATAC-seq is a process that detects open regions of chromatin by integration of the Tn5 transposase into open regions (65). This data can be correlated with sub-population data following EPC1, EPC2, and EPC1 plus EPC2 knockdown to identify what genes lead to the changes previously discovered.

Additionally, some epigenetic complexes have been known to degrade if a key member is degraded or reduced in number, such as EED when SUZ12 and EZH2 are knocked down (66). Do piccolo-NuA4 members degrade with reduced levels of EPC2 or can EPC1 compensate and stabilize? Therefore, I propose to measure protein levels of YnG2 and Eaf6 following EPC2, EPC1, or EPC1/2 knockdown using western blotting methods. This data can contextualize how necessary EPC2 is for piccolo-NuA4 function or if EPC1 can effectively replace EPC2 in stabilizing piccolo-NuA4 members.

Materials and Methods

Lentivirus Preparation

Both EPC2 and TRIM24 lentiviruses were produced at the University of Michigan Vector Core (Ann Arbor, MI). Briefly, cDNA (Horizon Discovery) for TRIM24 and EPC2 were ligated into a plasmid, each flanked by long terminal repeats. The transgene plasmid (TRIM24 or EPC2), envelope plasmid (EF1a-VSVG), and packaging plasmids (GAG, POL, TAT) were added to pre-seeded 293T cells. After 48 hours, the supernatant was collected and concentrated by centrifugation. At that time, the assembled plasmids were sequenced to ensure proper integration of transgene and promoters.

Lentivirus Transduction

SUM149 cells were seeded at $0.50x10^6$ cells/well in a six well plate and allowed to settle over 24 hours. The media was removed and replaced with antibiotic/antimycotic free media. 150 uL of viral supernatant (UM Vector Core, Ann Arbor, MI) was added to each well (control or lentivirus). Polybrene (Santa Cruz Biotechnology, Dallas, TX) was added for a final concentration of 8 ug/mL. Plates were gently mixed to ensure equal distribution of additives. Cells were incubated for 24 hours at 37 C. After that time, media was aspirated and replaced with fresh media.

Antibiotic selection

3 ug/mL of Puromycin (Sigma Aldrich, St. Louis, MO) was added to each well, previously added to media. Media containing puromycin was added and removed every two days. Cells were further grown to establish a lentiviral-augmented cell line.

irFP713 selection

Transduced cells were allowed to grow for several more days before being transferred from a 6-well plate to 10 cm plates. The cells were then selected for near-IR fluorescence by sorting on a MoFlo Astrios EQ Sorter (Beckman Coulter) and seeded once more in F-12 media. The cells were once more allowed to expand before being sorted and collected once more.

Live Cell Image Capture and Analysis

Live-cell image data were acquired on an Olympus IX83 P2ZF microscope, with a motorized IX3-CBH stage, and IX3 halogen lamp using the cellSens Dimensions software (Olympus). Time-lapse fluorescent microscopy images were acquired every 30 minutes for 72 hours in a 96-well plate using a 10X objective. FIJI was used for image processing. Background fluorescence was subtracted by a rolling ball radius of 50px, the cell of interest were selected by construction regions of interest (ROIs). Variability in cell size was compensated by multiplying the background mean fluorescence by ROI area, which was subtracted from each cell's total intensity. Every cell at t=0 and t=72 were created as ROIs and the compensated mean mCherry and EGFP fluorescence was measured and compared between timepoints. Representative samples were selected from these timepoints.

Cell-Cycle analysis

Cells were trypsinized, washed, and pelleted. While vortexing, ice-cold 70% ethanol was added dropwise to the sample. Samples were stored at 4C for 1 hour. Ethanol was removed by centrifugation at 1000G for 10 minutes and cells were washed 2x times in PBS. Ribonuclease was added (5 ug) (ThermoFIsher Scientific, Waltham, MA) and the cells were incubated for fifteen

minutes. After that time elapsed, 200 uL of 50 ug/mL Propidium Iodide (ThermoFIsher Scientific, Waltham, MA) was added to each cell sample. Cells were analyzed for PI levels, and a histogram was constructed for each sample. Plots were integrated to analyze what stage of cell cycle the populations were predominantly within.

Western Blot

Cells were lysed with 100 uL 1x RIPA buffer containing 1x Protease Inhibitor (Sigma Aldrich, St. Louis, MO) and 1x Phosphatase Inhibitors I and II (Sigma Aldrich, St. Louis, MO). Additionally, they were briefly sonicated ten times. Cellular debris was removed by centrifugation, and the total protein content was determined by using the BCA assay according to manufacturer's instructions (ThermoFisher, Waltham, MA). Protein content was normalized, and the samples were denatured with a combination of 5x SDS and boiling for 10 minutes.

Denatured protein samples were loaded into 8-15% precast SDS-Page gels, running buffer (ingredients) was added, and proteins were separated using 120 V (constant voltage) for one hour. Gel was then wet transferred onto nitrocellulose membranes at 0.22 AMP for 1 hour (constant amps). After transferring, the nitrocellulose was washed and stained with PONCEAU (Sigma Aldrich, St. Louis, MO) to verify protein transfer. A scan was taken for future records. Nitrocellulose (ThermoFisher, Waltham, MA) was then cut according to protein size and blocked with 5% milk in PBS. Nitrocellulose was then treated with antibodies overnight in 1% milk in PBST at 4C. The next morning, the blots were washed 3x in PBST then exposed to anti-rabbit goat secondary antibody linked with HRP (AbCam, Cambridge, United Kingdom) in 1% milk in PBST. Samples were incubated for two hours then washed with PBST. Samples were treated with Pierce ECL Substrate (ThermoFisher, Waltham, MA). The chamber, film (Thermo Scientific CL XPosure Film), and exposed nitrocellulose was taken to a dark room. The film was developed using a Konica Film Processor (Tokyo, Japan).

References

1. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014;344(6190):1396-401.

2. Sottoriva A, Spiteri I, Piccirillo SG, Touloumis A, Collins VP, Marioni JC, et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc Natl Acad Sci U S A. 2013;110(10):4009-14.

3. Stommel JM, Kimmelman AC, Ying H, Nabioullin R, Ponugoti AH, Wiedemeyer R, et al. Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. Science. 2007;318(5848):287-90.

4. Francis JM, Zhang CZ, Maire CL, Jung J, Manzo VE, Adalsteinsson VA, et al. EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. Cancer Discov. 2014;4(8):956-71.

5. Meyer M, Reimand J, Lan X, Head R, Zhu X, Kushida M, et al. Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity. Proc Natl Acad Sci U S A. 2015;112(3):851-6.

6. Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. The New England journal of medicine. 2005;352(10):987-96.

7. Wen PY, Kesari S. Malignant gliomas in adults. The New England journal of medicine. 2008;359(5):492-507.

8. Slamon DJ, Neven P, Chia S, Fasching PA, De Laurentiis M, Im SA, et al. Overall Survival with Ribociclib plus Fulvestrant in Advanced Breast Cancer. The New England journal of medicine. 2020;382(6):514-24.

9. Valero V. Docetaxel as single-agent therapy in metastatic breast cancer: clinical efficacy. Semin Oncol. 1997;24(4 Suppl 13):S13-1-s-8.

10. Robinson DM, Keating GM. Albumin-bound Paclitaxel: in metastatic breast cancer. Drugs. 2006;66(7):941-8.

11. Carey LA, Dees EC, Sawyer L, Gatti L, Moore DT, Collichio F, et al. The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. Clin Cancer Res. 2007;13(8):2329-34.

12. Kennedy WR, Tricarico C, Gabani P, Weiner AA, Altman MB, Ochoa LL, et al. Predictors of Distant Metastases in Triple-Negative Breast Cancer Without Pathologic Complete Response After Neoadjuvant Chemotherapy. J Natl Compr Canc Netw. 2020;18(3):288-96.

13. Lee HE, Kim JH, Kim YJ, Choi SY, Kim SW, Kang E, et al. An increase in cancer stem cell population after primary systemic therapy is a poor prognostic factor in breast cancer. Br J Cancer. 2011;104(11):1730-8.

14. Karaayvaz M, Cristea S, Gillespie SM, Patel AP, Mylvaganam R, Luo CC, et al. Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. Nature Communications. 2018;9(1):3588.

15. Masuda H, Baggerly KA, Wang Y, Zhang Y, Gonzalez-Angulo AM, Meric-Bernstam F, et al. Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes. Clin Cancer Res. 2013;19(19):5533-40.

16. Wu Q, Li J, Zhu S, Wu J, Chen C, Liu Q, et al. Breast cancer subtypes predict the preferential site of distant metastases: a SEER based study. Oncotarget. 2017;8(17):27990-6.

17. Echeverria GV, Powell E, Seth S, Ge Z, Carugo A, Bristow C, et al. High-resolution clonal mapping of multi-organ metastasis in triple negative breast cancer. Nature Communications. 2018;9(1):5079.

18. Kim C, Gao R, Sei E, Brandt R, Hartman J, Hatschek T, et al. Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell. 2018;173(4):879-93.e13.

19. Li X, Lewis MT, Huang J, Gutierrez C, Osborne CK, Wu MF, et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. Journal of the National Cancer Institute. 2008;100(9):672-9.

20. Echeverria GV, Ge Z, Seth S, Zhang X, Jeter-Jones S, Zhou X, et al. Resistance to neoadjuvant chemotherapy in triple-negative breast cancer mediated by a reversible drug-tolerant state. Sci Transl Med. 2019;11(488).

21. Chiu AM, Mitra M, Boymoushakian L, Coller HA. Integrative analysis of the inter-tumoral heterogeneity of triple-negative breast cancer. Sci Rep. 2018;8(1):11807-.

22. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature. 2012;486(7403):395-9.

23. Perou CM. Molecular Stratification of Triple-Negative Breast Cancers. The Oncologist. 2011;16(suppl 1):61-70.

24. Wang D-Y, Jiang Z, Ben-David Y, Woodgett JR, Zacksenhaus E. Molecular stratification within triple-negative breast cancer subtypes. Sci Rep. 2019;9(1):19107.

25. Sachs N, de Ligt J, Kopper O, Gogola E, Bounova G, Weeber F, et al. A Living Biobank of Breast Cancer Organoids Captures Disease Heterogeneity. Cell. 2018;172(1-2):373-86.e10.

26. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61-70.

27. Biswas S, Rao CM. Epigenetic tools (The Writers, The Readers and The Erasers) and their implications in cancer therapy. European journal of pharmacology. 2018;837:8-24.

28. Chaffer CL, Marjanovic ND, Lee T, Bell G, Kleer CG, Reinhardt F, et al. Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. Cell. 2013;154(1):61-74.

29. Granit RZ, Gabai Y, Hadar T, Karamansha Y, Liberman L, Waldhorn I, et al. EZH2 promotes a bi-lineage identity in basal-like breast cancer cells. Oncogene. 2013;32(33):3886-95.

30. Choi JD, Lee J-S. Interplay between Epigenetics and Genetics in Cancer. Genomics Inform. 2013;11(4):164-73.

31. Risom T, Langer EM, Chapman MP, Rantala J, Fields AJ, Boniface C, et al. Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer. Nature Communications. 2018;9(1):3815.

32. Xiao Y, Ma D, Zhao S, Suo C, Shi J, Xue MZ, et al. Multi-Omics Profiling Reveals Distinct Microenvironment Characterization and Suggests Immune Escape Mechanisms of Triple-Negative Breast Cancer. Clin Cancer Res. 2019;25(16):5002-14.

33. Ma XJ, Dahiya S, Richardson E, Erlander M, Sgroi DC. Gene expression profiling of the tumor microenvironment during breast cancer progression. Breast Cancer Res. 2009;11(1):R7.

34. Bareche Y, Venet D, Ignatiadis M, Aftimos P, Piccart M, Rothe F, et al. Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. Ann Oncol. 2018;29(4):895-902.

35. Lanning NJ, Castle JP, Singh SJ, Leon AN, Tovar EA, Sanghera A, et al. Metabolic profiling of triple-negative breast cancer cells reveals metabolic vulnerabilities. Cancer Metab. 2017;5:6-.

36. Karaayvaz M, Cristea S, Gillespie SM, Patel AP, Mylvaganam R, Luo CC, et al. Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. Nature communications. 2018;9(1):3588-.

37. Hollestelle A, Nagel JHA, Smid M, Lam S, Elstrodt F, Wasielewski M, et al. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. 2010;121(1):53-64.

38. Akashi T, Shirasawa T, Hirokawa K. Gene expression of CD24 core polypeptide molecule in normal rat tissues and human tumor cell lines. Virchows Arch. 1994;425(4):399-406.

39. Marcato P, Dean CA, Pan D, Araslanova R, Gillis M, Joshi M, et al. Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis. Stem Cells. 2011;29(1):32-45.

40. Jain AK, Allton K, Duncan AD, Barton MC. TRIM24 is a p53-induced E3-ubiquitin ligase that undergoes ATM-mediated phosphorylation and autodegradation during DNA damage. Mol Cell Biol. 2014;34(14):2695-709.

41. Gechijian LN, Buckley DL, Lawlor MA, Reyes JM, Paulk J, Ott CJ, et al. Functional TRIM24 degrader via conjugation of ineffectual bromodomain and VHL ligands. Nat Chem Biol. 2018;14(4):405-12.

42. Yuan X, Zhou Y, Casanova E, Chai M, Kiss E, Gröne HJ, et al. Genetic inactivation of the transcription factor TIF-IA leads to nucleolar disruption, cell cycle arrest, and p53-mediated apoptosis. Mol Cell. 2005;19(1):77-87.

43. Khetchoumian K, Teletin M, Tisserand J, Mark M, Herquel B, Ignat M, et al. Loss of Trim24 (Tif1alpha) gene function confers oncogenic activity to retinoic acid receptor alpha. Nat Genet. 2007;39(12):1500-6.

44. Khetchoumian K, Teletin M, Tisserand J, Herquel B, Ouararhni K, Losson R. Trim24 (Tif1 alpha): an essential 'brake' for retinoic acid-induced transcription to prevent liver cancer. Cell cycle (Georgetown, Tex). 2008;7(23):3647-52.

45. Kikuchi M, Okumura F, Tsukiyama T, Watanabe M, Miyajima N, Tanaka J, et al. TRIM24 mediates ligand-dependent activation of androgen receptor and is repressed by a bromodomain-containing protein, BRD7, in prostate cancer cells. Biochim Biophys Acta. 2009;1793(12):1828-36.

46. Chambon M, Orsetti B, Berthe ML, Bascoul-Mollevi C, Rodriguez C, Duong V, et al. Prognostic significance of TRIM24/TIF-1α gene expression in breast cancer. Am J Pathol. 2011;178(4):1461-9.

47. Li H, Sun L, Tang Z, Fu L, Xu Y, Li Z, et al. Overexpression of TRIM24 correlates with tumor progression in non-small cell lung cancer. PLoS One. 2012;7(5):e37657.

48. Cui Z, Cao W, Li J, Song X, Mao L, Chen W. TRIM24 overexpression is common in locally advanced head and neck squamous cell carcinoma and correlates with aggressive malignant phenotypes. PLoS One. 2013;8(5):e63887.

49. Liu X, Huang Y, Yang D, Li X, Liang J, Lin L, et al. Overexpression of TRIM24 is associated with the onset and progress of human hepatocellular carcinoma. PLoS One. 2014;9(1):e85462.

50. Zhang LH, Yin AA, Cheng JX, Huang HY, Li XM, Zhang YQ, et al. TRIM24 promotes glioma progression and enhances chemoresistance through activation of the PI3K/Akt signaling pathway. Oncogene. 2015;34(5):600-10.

51. Groner AC, Cato L, de Tribolet-Hardy J, Bernasocchi T, Janouskova H, Melchers D, et al. TRIM24 Is an Oncogenic Transcriptional Activator in Prostate Cancer. Cancer Cell. 2016;29(6):846-58.

52. Fang Z, Deng J, Zhang L, Xiang X, Yu F, Chen J, et al. TRIM24 promotes the aggression of gastric cancer via the Wnt/β-catenin signaling pathway. Oncol Lett. 2017;13(3):1797-806.

53. Lin L, Zhao W, Sun B, Wang X, Liu Q. Overexpression of TRIM24 is correlated with the progression of human cervical cancer. Am J Transl Res. 2017;9(2):620-8.

54. Wang H, Xue W, Jiang X. Overexpression of TRIM24 Stimulates Proliferation and Glucose Metabolism of Head and Neck Squamous Cell Carcinoma. Biomed Res Int. 2018;2018:6142843.

55. Li C, Xin H, Shi Y, Mu J. Knockdown of TRIM24 suppresses growth and induces apoptosis in acute myeloid leukemia through downregulation of Wnt/GSK-3β/β-catenin signaling. Hum Exp Toxicol. 2020;39(12):1725-36.

56. Tisserand J, Khetchoumian K, Thibault C, Dembélé D, Chambon P, Losson R. Tripartite motif 24 (Trim24/Tif1α) tumor suppressor protein is a novel negative regulator of interferon (IFN)/signal transducers and activators of transcription (STAT) signaling pathway acting through retinoic acid receptor α (Rarα) inhibition. J Biol Chem. 2011;286(38):33369-79.

57. Lv D, Li Y, Zhang W, Alvarez AA, Song L, Tang J, et al. TRIM24 is an oncogenic transcriptional co-activator of STAT3 in glioblastoma. Nat Commun. 2017;8(1):1454.

58. Sigoillot FD, King RW. Vigilance and validation: Keys to success in RNAi screening. ACS Chem Biol. 2011;6(1):47-60.

59. Burnette WN. "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate--polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. Anal Biochem. 1981;112(2):195-203.

60. Herquel B, Ouararhni K, Davidson I. The TIF1α-related TRIM cofactors couple chromatin modifications to transcriptional regulation, signaling and tumor suppression. Transcription. 2011;2(5):231-6.

61. Carrier M, Lutzing R, Gaouar S, Rochette-Egly C. TRIM24 mediates the interaction of the retinoic acid receptor alpha with the proteasome. FEBS Lett. 2018;592(8):1426-33.

62. Allton K, Jain AK, Herz HM, Tsai WW, Jung SY, Qin J, et al. Trim24 targets endogenous p53 for degradation. Proc Natl Acad Sci U S A. 2009;106(28):11612-6.

63. Garcia-Barcena C, Osinalde N, Ramirez J, Mayor U. How to Inactivate Human Ubiquitin E3 Ligases by Mutation. Frontiers in cell and developmental biology. 2020;8:39-.

64. Li Y, Wu H, Wu W, Zhuo W, Liu W, Zhang Y, et al. Structural insights into the TRIM family of ubiquitin E3 ligases. Cell Res. 2014;24(6):762-5.

65. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10(12):1213-8.

66. Cao Q, Wang X, Zhao M, Yang R, Malik R, Qiao Y, et al. The central role of EED in the orchestration of polycomb group complexes. Nature Communications. 2014;5(1):3127.