Title

Identifying Markers of Drug Sensitivity in Cancer Cells with Focal Amplifications

Author

Jayaraman, Rachana

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Identifying Markers of Drug Sensitivity in Cancer Cells with Focal Amplifications

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Bioinformatics

by

Rachana Jayaraman

2022

ABSTRACT OF THE THESIS

Identifying Markers of Sensitivity in

Cancer Cells with Focal Amplifications

by

Rachana Jayaraman

Master of Science in Bioinformatics

University of California, Los Angeles, 2022

Professor Thomas Graeber, Chair

Focal amplifications (FAs) are important contributors to genomic instability within cancer cells and can influence their response to targeted therapies. FAs can take on two different forms: double minutes (DMs), which are circular segments of extrachromosomal DNA (ecDNA), or homogeneously staining regions (HSRs), which are intrachromosomal regions of amplification. DMs, or ecDNAs, are particularly known for enrichment of oncogenes and their contribution to cell plasticity, especially in conjunction with BRD4 which promotes overexpression of the oncogenes. However, while FAs confer advantages upon the cells, there are also weaknesses that are introduced due to the increased instability of the genome. This project aims to identify weak points in cells with FAs through large-scale computational analysis of drug screens. The area under the dose response curve (AUC) was chosen as a measure of sensitivity for this analysis since a smaller AUC indicates that the sample responded more quickly to the drug. We performed linear regressions for hundreds of drugs to identify those with a smaller AUC value when treating samples with FA compared to samples without FA. The targets associated with

these drugs were then compared to differential expression analysis and network correlation analysis modules to identify if they were differentially expressed or part of a larger network specific to cells with DMs and HSRs. We found several targets that appeared to be consistently associated with our criteria for sensitivity including AKT1, BRD2, BRD3, BRD4, DNMT3A, FLT3, KIT, and PDGFRB. These targets appeared even when removing samples with known driver genes and reducing tissue-specific effects on the linear regressions. A subset of these targets has been demonstrated to be associated with DNA repair and genomic instability in cancer cells from previous studies. Most importantly, AKT1 and FLT3 were found to be effective targets for joint inhibition with BRD4 in previous research, which introduces potential for targeting ecDNA and DM hubs. Transcriptome (RNA-seq) data found that numerous genes scoring in the drug sensitivity analysis were differentially regulated. Network analysis of co-expressed gene modules identified by network correlation analysis showed interactions between identified drug targets and differentially expressed genes and support the existence of gene interactions that were not apparent from the sensitivity analysis alone. These results provide more insight into potential targeted therapies for cells with FA and the potential weak points within the cells.

The thesis of Rachana Jayaraman is approved.

Hilary A. Coller

Jingyi Li

Thomas G. Graeber, Committee Chair

University of California, Los Angeles

2022

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am grateful for the guidance of my advisor Dr. Thomas Graeber and my mentor Kai Song, who has worked with me on this project over the last three years. I would also like to thank my family and friends for their continued support.

CHAPTER 1

Introduction

The progression of tumors is dependent upon the dysregulation of cellular pathways and mechanisms that control cell division and proliferation, allowing them to rapidly divide even in cellular conditions that would normally trigger senescence or apoptosis [29]. This dysregulation can be attributed to the accumulation and selection of beneficial genomic abnormalities over time. For instance, tumor cells often contain amplifications of genes and chromosomes that provide a growth advantage under certain conditions or in conjunction with specific mutations [32]. Amplifications in cancer cells can occur at large scales; whole genome doubling is commonly observed in advanced tumors and is an example of polyploidy, or the existence of extra sets of chromosomes [3]. On the other hand, amplifications can also exist as smaller regions of the genome which are enriched for specific sets of genes. These smaller amplifications, known as focal amplifications (FAs), can exist in two different forms. The first mode of FA is double minutes (DMs), which are small circular segments of extrachromosomal DNA (ecDNA) that arise from a chromosomal segment (Figure 1). They can also exist as intrachromosomal regions of amplifications, also known as homogeneously staining regions (HSRs), seen as a large uniform band (Figure 1) [30].

ecDNAs are of particular interest because of their independent nature within the cell. According to a study conducted by Turner et al., ecDNA was detected in nearly 40 percent of the cancer samples they studied, but their presence varied greatly across cells within the same sample since they segregate randomly and unevenly into daughter cells [32]. Sequencing of ecDNAs also demonstrated enrichment of oncogenes such as EGFR, MYC, CDK4, and MDM2 [33]. The copy numbers of these oncogenes tend to vary across ecDNAs, creating more

heterogeneity in the genome of the cells within the cancer sample. The genes on the ecDNAs have also been shown to have higher expression, as quantified by transcript counts, and there appears to be higher chromatin accessibility due the reduced compaction of nucleosomes [33]. To promote oncogene overexpression, clusters of 10 to 100 ecDNAs, known as ecDNA hubs, form to increase enhancer-gene interactions. These ecDNAs are tethered by BRD4, a BET protein, and it has been shown that disrupting BRD4 with JQ1, a BET inhibitor, reduces oncogene expression [13]. Furthermore, ecDNAs have been shown to be able to re-integrate into chromosomes and form HSRs [32]. All these characteristics of ecDNAs together show the selective advantages and plasticity they provide to cancer cells, particularly due to their ability to adapt and increase oncogene expression.

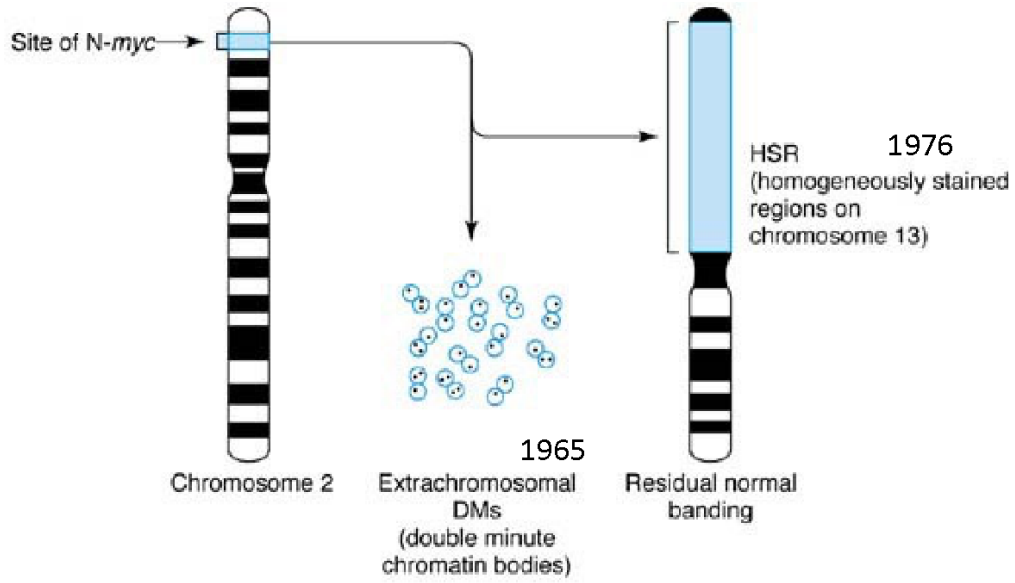


Figure 1.1: Two modes of focal amplification (FA). Bottom: Double minutes, extrachromosomal circular segments containing N-Myc. Right: Homogeneously staining regions, uniform stretch of chromosome with gene amplified in tandem.

While gene amplifications can promote tumor growth through selection of oncogenes, they can also unintentionally lead to genomic vulnerabilities in less obvious ways. For instance, a 2021 study by Cohen-Sharir et al. found that aneuploid cells are sensitive to the inhibition of the spindle assembly checkpoint. This was due to the increased accumulation of defects, which lead to the activation of cell death [5]. Another study by Quinton et al. on cells with whole genome doubling (WGD) found similar dependencies on signals from the spindle-assembly checkpoint, DNA replication factors, and proteasome function. In particular, it was found that the mitotic kinesin protein KIF18A was required for the cells with whole genome doubling to survive since its loss led to many mitotic errors [26]. These findings have important implications for treating cells with large-scale duplication, however not all cancer cells have aneuploidy or whole genome doubling. In many cases they have DMs or HSRs, which may not lead to mitotic errors as easily since they are smaller segments. However, the amplification of genes, especially as ecDNA, can still lead to unintended consequences by disrupting the activity of important pathways and are a point of interest for drug sensitivity studies. By identifying genes that contribute to weak points in the tumor cells, they may serve as more effective therapeutic targets for cells with FA compared to the oncogenes.

In this study, we performed a large-scale comparison of drug response between cancer cell lines with and without FA to identify markers of sensitivity and potential drug targets. Using copy number data for CCLE cell lines, we classified samples as FA positive and negative based on the highest amount of amplification present. We then performed two linear regressions to identify drugs with a negative trend between the area under the curve (AUC), an indicator of sensitivity, and the FA term: one with a control for tissue type and one for each tissue-drug pair. Differential expression analysis of RNA-seq data from melanoma cells with DMs and HSRs

further showed the extensive dysregulation in cells with FA, including differential expression of identified targets.

CHAPTER 2

Methods

2.1 Drug Sensitivity Analysis

2.1.1 Data

For this analysis, the following data was obtained from the DepMap portal's 21Q4 release: Cancer Cell Line Encyclopedia (CCLE) cell line sample information, CCLE segment-level DNA copy number data [9], and drug screen data from three resources: GDSC, CTRP, and PRISM 19Q4 [6, 14, 22, 28].

2.1.2 Gene-level Copy Number Calculation

a b

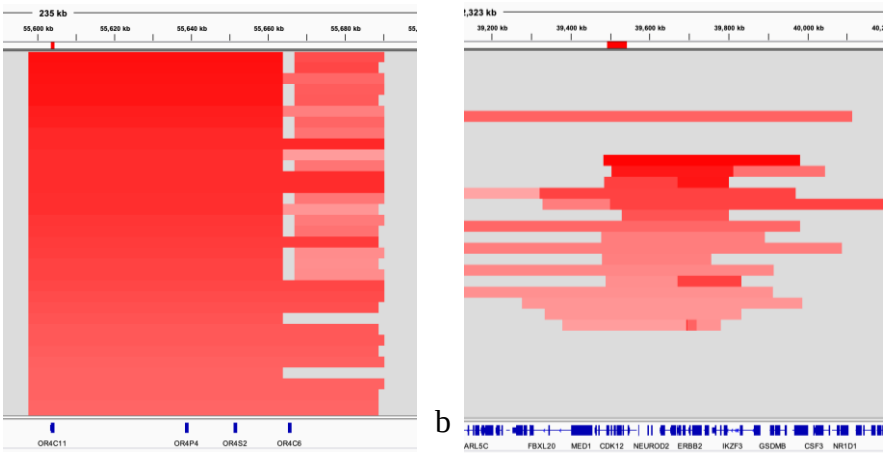


Figure 2.1: CCLE copy number data. a. Segmentation artifact, identified by uniform amplicons in the majority of samples, b. Example of a normal FA copy number distribution

The segment-level copy number data was visually inspected in a genome viewer (IGV) to check for sequencing artifacts. These were identified by looking for uniform amplicon blocks across samples, typically defined by harsh borders and a consistent color (Figure 2.1.a). Segments

identified as artifacts were removed from the copy number data prior to calculating the weighted copy number for each gene.

$$weighted_i = \frac{\sum_i p_i s_i}{\sum_i p_i} \quad (2.1.1)$$

$$i \in [segments\ overlapping\ gene\ A], p_i: percent\ overlap, s_i: segment\ mean$$

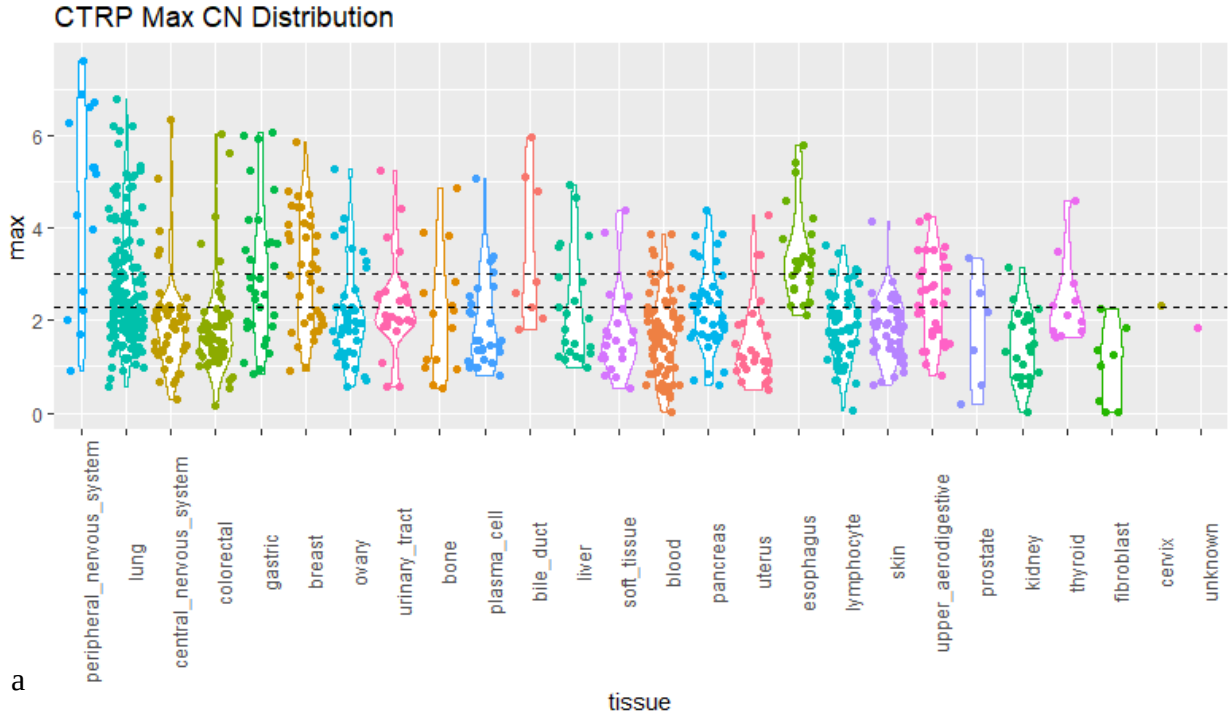
For each CCLE sample, GenomicRanges' findOverlaps function was used to identify segments in the copy number data that overlapped with genes in the reference GRCh38 genome. The percent overlap for the segments was then calculated by using pintersect to identify the overlapping region between the gene and the segment. For each gene, the weighted copy number was calculated using the percent overlap and segment mean for the overlapping segments (Equation 2.1.1). If a gene was wholly contained within a segment, then the segment mean was used on its own. The weighted copy number for each gene was then log2 transformed to compress the values.

2.1.3 Focal Amplification Classification

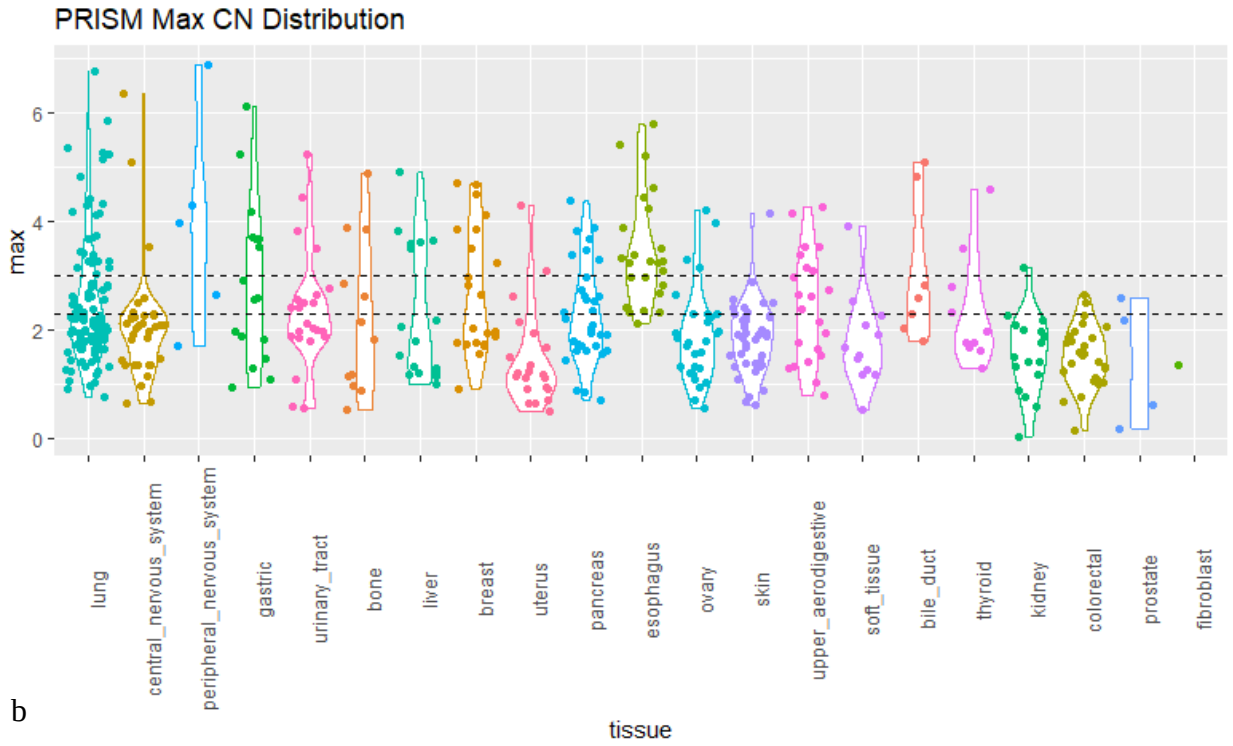
For this analysis, samples were classified as FA high (FA+) or FA low (FA-) only based on the maximum log2 copy number (log2CN), irrespective of the gene that was amplified in the sample. To classify samples as FA positive (FA+) or FA negative (FA-), several thresholds were used: (2.3, 3), (2,3), (1.8, 3), (1.5, 3), (2.3, 3.5), (2, 3.5), (1.5, 3.5), (1.8, 3.5), (2.3, 4), (2, 4), (1.8, 4), and (1.5, 4). In the notation for the thresholds, the first number served as the upper bound for the maximum log2CN of FA- samples, and the second number served as the lower bound for the maximum log2CN of FA+ samples.

The maximum log2CN for each sample was compared to each threshold, and the sample was classified as FA- if it fell below the first number or FA+ if it fell above the second number. Any samples with a log2CN between the two bounds were considered neither FA- or FA+ and were removed from further analysis. This created a binary classification in which a sample either had enough amplification at any locus in the genome to be considered focally amplified, or the sample had little to no amplification compared to the normal genome.

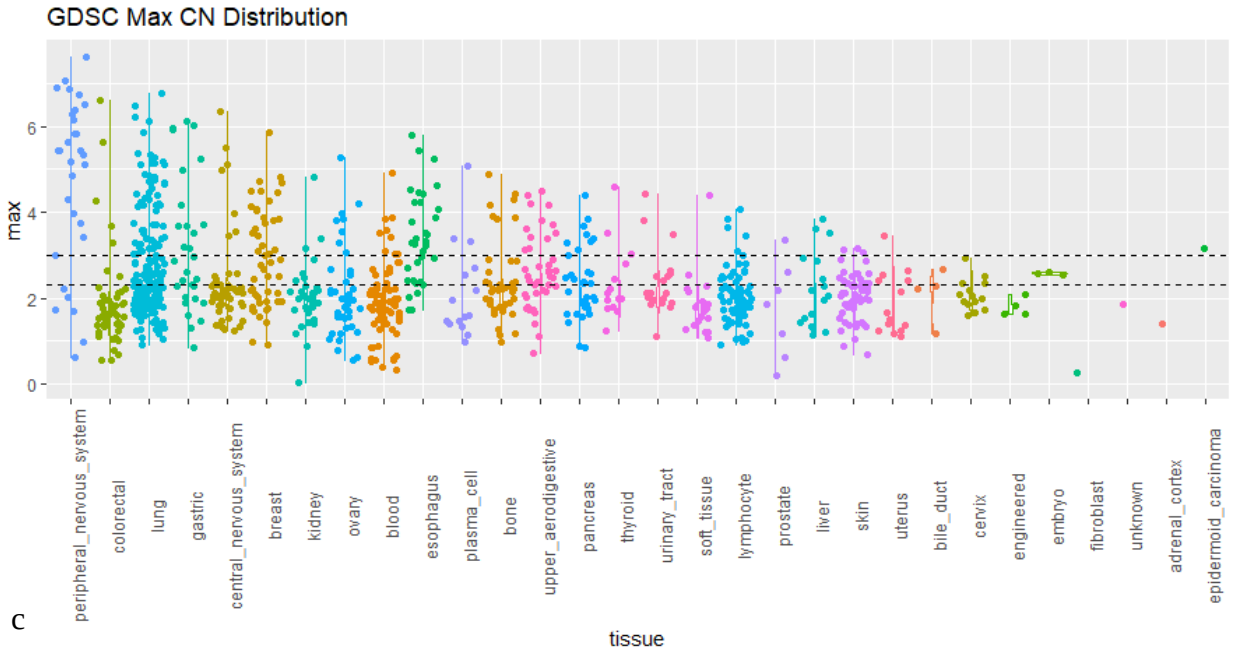
To ensure there were enough samples for comparison within each tissue type, only tissues with at least three FA+ and at least three FA- samples were kept for further analysis. For instance, in PRISM there were insufficient FA+ samples for the last seven tissues in the plot so all samples within those tissue types were removed (Figure 2.2.b). The samples from uterus and skin tissues were also removed (Figure 2.2.b). This filtering and classification process was performed for each set of thresholds and each dataset.



a



b



c

Figure 2.2: Maximum gene-level copy number for each CCLE sample, by tissue type, a) samples in CTRP, b) samples in PRISM, c) samples in GDSC

2.1.4 Linear Models

Sample (cell line)	Drug ID	AUC	Tissue	FA
ACH-000004	1788	14.709	Blood	low
ACH-000074	1788	14.056	Blood	low
ACH-000122	1788	14.279	Blood	low
…	…	…	…	…
ACH-000146	1788	14.299	Blood	high
ACH-000326	1788	12.489	Blood	high
ACH-000406	1788	14.968	Blood	high
…	…	…	…	…
ACH-000035	1788	13.838	Lung	low
ACH-000086	1788	15.379	Lung	low
ACH-000121	1788	13.096	Lung	low
…	…	…	…	…
ACH-000012	1788	14.201	Lung	high
ACH-000309	1788	14.361	Lung	high
ACH-000335	1788	13.204	Lung	high
…	…	…	…	…

Table 2.1: Example of data frame for linear regression, taken from drug 1788 in CTRP. Multiple tissue types, FA low and FA high samples within each tissue.

For each dataset, the following data was aggregated for each drug: sample, drug id, AUC, tissue of sample, and FA classification. The table above shows an example of the data collected for the threshold (2.3, 3) within the CTRP dataset. Among all measures of drug response provided in the screens, AUC was chosen as a measure of sensitivity for this analysis since it reflects the

maximum dosage and the IC50 values, both of which are important for understanding how responsive cancer cells are to a given drug [24]. A sample that is more sensitive will have a smaller AUC value since it responds more quickly to the drug and has a narrower dose response curve.

The FA values were converted to factors for the linear regression and took on binary values: 0 for FA- and 1 for FA+. The tissue values were also converted to factors, with the levels assigned alphabetically. The data was further filtered to remove drug-sample combinations with missing AUC values. Any drug that failed to meet the following criteria was then removed from the analysis: at least two tissue types, at least 3 FA+ samples within each tissue, and at least 3 FA- samples within each tissue.

$$auc = \beta_0 + \beta_1 * FA + \beta_2 * Tissue + \varepsilon \quad (2.1.2)$$

Equation 2.1.2 shows the first regression run for each drug within each data set. This model aimed to identify the amount of change in the AUC value for each drug that could be attributed to the presence of FA while controlling for tissue-specific differences. Given that the FA term takes on a binary classification, the beta estimate for the FA term, denoted β_{FA}, is the predicted change in AUC between the FA- samples and FA+ samples. For each drug, the summary function was used to obtain the beta estimates and the p-values for each term in the model as well as the intercept.

$$auc = \beta_0 + \beta_1 * FA + \varepsilon \quad (2.1.3)$$

Another linear regression modeled after equation 2.1.3 was then performed for each drug-tissue combination to look at tissue-specific changes in AUC between FA- and FA+ samples. The

same summary statistics were obtained for each drug-tissue combination. Within each regression analysis, the false discovery rate was calculated using the BH correction.

2.2 Transcriptome Analysis

2.2.1 Data

Our lab generated transcriptome (RNA-seq) data for 15 melanoma samples from the m249_vsr line, and they fell under three karyotypic categories: five DM-&HSR-, five DM+&HSR-, and five DM-&HSR+. Given that most genes within FAs are those that code for proteins, the counts data was filtered for only protein coding genes in the GRCh38 assembly.

2.2.2 Differential Gene Expression

To identify up- and down-regulation of genes, we performed differential gene expression analysis and looked for changes that can be attributed to the presence of DMs or HSRs. The DESeq2 package in R [17] was used on raw counts to perform three separate comparisons: DM+&HSR- versus DM-&HSR+, DM+&HSR- versus DM-&HSR-, and DM-&HSR+ versus DM-HSR-. For each comparison, we filtered genes that had a log2-transformed fold change (log2FC) in expression that was greater than 0.5 or less than -0.5. The genes were also filtered for a BH-corrected p-value less than 0.05 to ensure that the differential expression was significant. To identify the biological role of the differentially expressed (DE) genes, the filtered DE genes from each comparison were run through Gene Ontology [2].

2.2.3 Network Correlation Analysis

To further explore the specific gene interactions and protein networks, Weighted Gene Correlation Network Analysis (WGCNA) [15] was performed using log2 counts per million (CPM) data of all differentially expressed genes identified earlier. As with the differentially expressed genes, we performed gene ontology analysis of the modules created by WGCNA.

Sample	Karyotype	DM-&HSR-	DM+&HSR-	DM-&HSR+
Sample 1	DM-&HSR-	1	0	0
Sample 2	DM-&HSR-	1	0	0
Sample 3	DM+&HSR-	0	1	0
Sample 4	DM-&HSR+	0	0	1
Sample 5	DM-&HSR+	0	0	1
…	…	….	…	…

ME
0.7
-0.1
0.5
1.2
0.1
…

Table 2.2: Analysis design for correlating karyotype with module eigengene (ME). Left half, binary karyotype assignment for each sample: 1 if it falls under the karyotype and 0 if it does not. Right: example of module eigengenes (MEs) used to correlate with each vector on left.

To look at the relationship between modules and focal amplifications, we created a binary list for each karyotypic category and assigned a 1 if the sample had that karyotype and 0 if it did not (Table 2.2). Each of these karyotypic lists were then correlated with each of the module eigengenes (MEs). The ME is the first principal component of the expression matrix for the module and serves as a representative expression profile of each sample within the module [15].

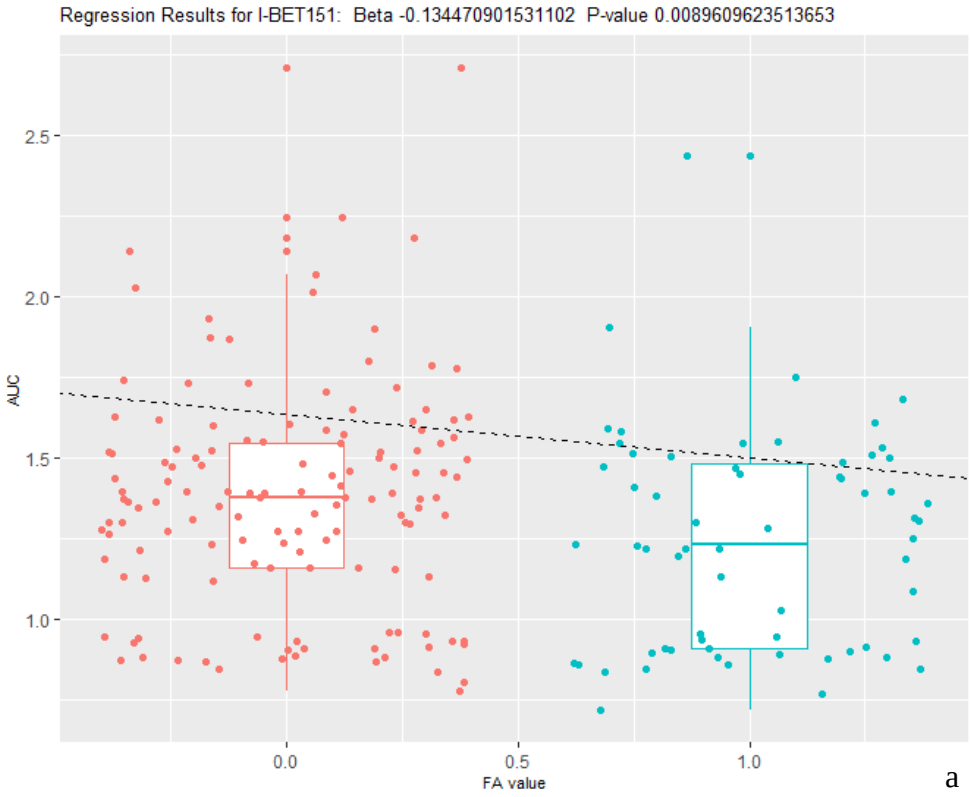
The strongest correlations will show the modules that are most associated with a given karyotype. For each correlation, the correlation coefficient and the p-value were calculated to quantify the strength and significance.

CHAPTER 3

Drug Sensitivity Analysis Results

3.1 Threshold-Specific Results

After running the linear regression, drugs were filtered based on two criteria: a negative β_{FA} and a p-value or FDR below 0.05. A negative β_{FA} indicates that the AUC for FA+ samples was lower compared to the AUC for FA- samples, suggesting that a smaller dosage of the drug is needed to treat FA+ samples.



a

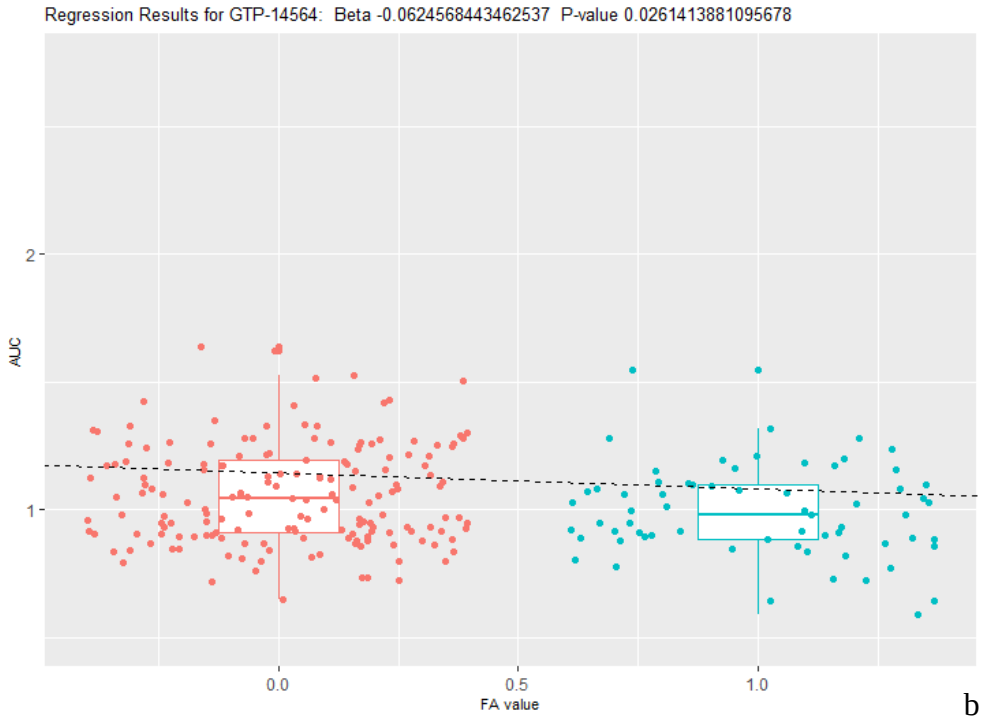


Figure 3.1: Visualization of linear regression results for two drugs. FA value of 0 represents FA- samples and FA value of 1 represents FA+ samples, binary value used in regression. Box plot shows distribution of AUC values for each group. Dashed line is estimated regression line with slope equal to β_{FA}, listed at the top, and estimated intercept is the AUC value at FA = 0. a) Distribution and regression for I-BET151, a BRD inhibitor. b) Distribution for GTP-14564, a FLT3 inhibitor

This is observed in Figure 3.1, which demonstrates that on average FA+ samples had lower AUC values compared to FA- samples when treated with I-BET151 and GTP-14564. The regression lines for both drugs captured this difference given they both had negative slopes, which were quantified by the β_{FA} estimates listed at the top. It is also apparent from the distributions in Figure 3.1 that the difference in the mean AUC between FA+ and FA- samples was larger for I-BET151 compared to GTP-14564. This was captured by the β_{FA} estimates for the two regression lines since the β_{FA} estimate for I-BET151 had a larger magnitude. The p-value or FDR below 0.05 serves as the criteria for statistical significance and provides confidence that the FA term predicts the change in AUC. By using both criteria for filtering

drugs, we have confidence that the associated targets for the drugs are points of weakness and contribute to the observed differences in drug response for FA+ samples.

In PRISM, the threshold (1.8, 4) had the greatest number of drugs that met the two criteria, and the second most was under the threshold (2, 4). This suggests that the most significant differences in drug response were observed between FA- samples with a maximum log2CN less than 2 and FA+ samples with a maximum log2CN greater than 4. Among all thresholds in PRISM, only (1.8, 3.5) had a drug with an FDR value below 0.05. The drug was GTP-14564 which is a FLT3 and tyrosine kinase inhibitor. Similarly, in CTRP the thresholds (1.8, 4) and (2, 4) had the most drugs that met the two criteria, but there were more drugs with an FDR value below 0.05. The drug purmorphamine, an activator of SMO, appeared to have a significant FDR for the following thresholds: (2.3, 3), (2, 3), (1.8, 3), (1.5, 3), (2, 3.5), and (1.8, 3.5). The threshold (2, 3) also had the following drugs: MI-1, an MEN1 inhibitor; CI-976, which inhibits ACAT1; and GW-405833, which inhibits CNR2. The threshold (1.5, 3) also had CI-976 in addition to ML203, which activates PKM. ML203 appeared to have a significant FDR for the thresholds (2.3, 3.5), (1.8, 3.5), and (1.5, 3.5) as well. The following two drugs both had significant FDRs for the threshold (2.3, 3.5), (2, 3.5), and (1.8, 3.5): axitinib, a VEGFR, c-KIT, and PDGFR inhibitor, and canertinib, an EGFR and ERBB2 inhibitor. MK-2206, an AKT1 inhibitor, had significant FDRs for the thresholds (2.3, 3.5), (2.3, 4), and (2, 4). The final drug with a significant FDR was MGCD-265 under the threshold (2.3, 4), which is also a c-MET and VEGFR inhibitor.

Looking across drugs filtered in both PRISM and CTRP, the total frequency of each gene target was counted for each threshold. In particular, there are several genes associated with FA that were of interest when looking at the distribution: TSC1, TSC2, RHEB, TBC1D7,

AKT1, RPTOR, MTOR, RICTOR, YWHAE, DDIT4, MLST8, AKT3, STAT1, PDE4, IFI16, RCOR1, HDAC1, HTT, SIN3A, CTDSP1, KDM1A, SIN3B, HIST1H2BD, HIST1H2BJ, HIST1H2BH, ATR, H2AX, H2AZ, H2AB1, DDX, RBP, BRD2, BRD3, BRD4, BRDT, PARG, VDAC2/3, SLC17A11, NEDD4, NEDD4L, PNKP, DNMT3A, DNMT3B, and TERT.

3.1.1 Threshold (2.3, 3)

Gene	PRISM	CTRP	Total
ADRA1A/B	4	0	4
BCL2	1	3	4
FLT3	2	2	4
HTR2A/C	4	0	4
KIT	2	1	3
PDGFRB	2	1	3
BCL2L1/2	0	3	3
EGFR/ERBB2	0	3	3
BRD2/3/4	3	0	3
DNMT3A	1	0	1
AKT1	0	1	1
TERT	0	1	1
PIK3CD	0	1	1
PIK3CG	0	1	1

Table 3.1: Frequency of top targets among filtered drugs for threshold (2.3, 3). First column is frequency among drugs in PRISM regression analysis, second column is frequency among drugs in CTRP regression analysis

The genes of interest, based on our pre-defined list of genes associated with FA, that appeared under this threshold are BRD2, BRD3, BRD4, DNMT3A, AKT1, and TERT. BRD2, BRD3, and BRD4 appeared together three times in PRISM with the following drugs: I-BET151, OTX015, and bromosporine. Of these BRD inhibitors, I-BET151 and OTX015 had the largest β_{FA} estimates across all the filtered drugs. This suggests that among the statistically significant results, BRD inhibition elicited the biggest change in drug response for FA+ samples. DNMT3A appeared once with the drug azacitidine in PRISM, and TERT appeared once under CTRP with the drug BIBR-1532. AKT1 appeared once under CTRP as well with the drug MK-2206, which had a significant FDR in other thresholds but not this one.

BCL2, FLT3, KIT, and PDGFRB are the few genes that appeared across both datasets. BCL2 appeared under venetoclax in PRISM and under three combinations of navitoclax in CTRP. BCL2L1 and BCL2L2 were also listed as targets for these three drugs in CTRP, accounting for the frequency seen in the table above. In both PRISM and CTRP, the genes FLT3, KIT, and PDGFRB were grouped together under the same drugs. In PRISM, the three genes appeared under GTP-14564 and crenolanib. In CTRP, FLT3 appeared alone under MGCD-265, and all three genes appeared under axitinib. MGCD-265 also targets FLT1, KDR, and MET, and axitinib targets FLT1, KDR, and PDGFRA in addition to the other three targets. EGFR had some of the largest β_{FA} estimates for CTRP, appearing under canertinib, neratinib, and afatinib. Of these three drugs, canertinib had the largest estimate and second largest estimate across all the filtered drugs. PIK3CD and PIK3CG, known to be associated with the AKT pathway, appeared under PIK-93 and CAL-101 respectively.

PRISM and CTRP did not appear to have overlapping drugs for any of the targets listed above. GDSC also did not appear to have an overlap with PRISM, but it did have navitoclax,

which was observed in CTRP. However, many of the drugs identified in GDSC did have similar targets to those observed in the other two datasets. Osimertinib is an EGFR inhibitor, and savolitinib is a c-MET inhibitor. Amuvatinib is a KIT, PDGFRA, and FLT3 inhibitor, following the grouping seen in the other datasets. One target that appeared only in GDSC was HDAC, which was under the drug panobinostat.

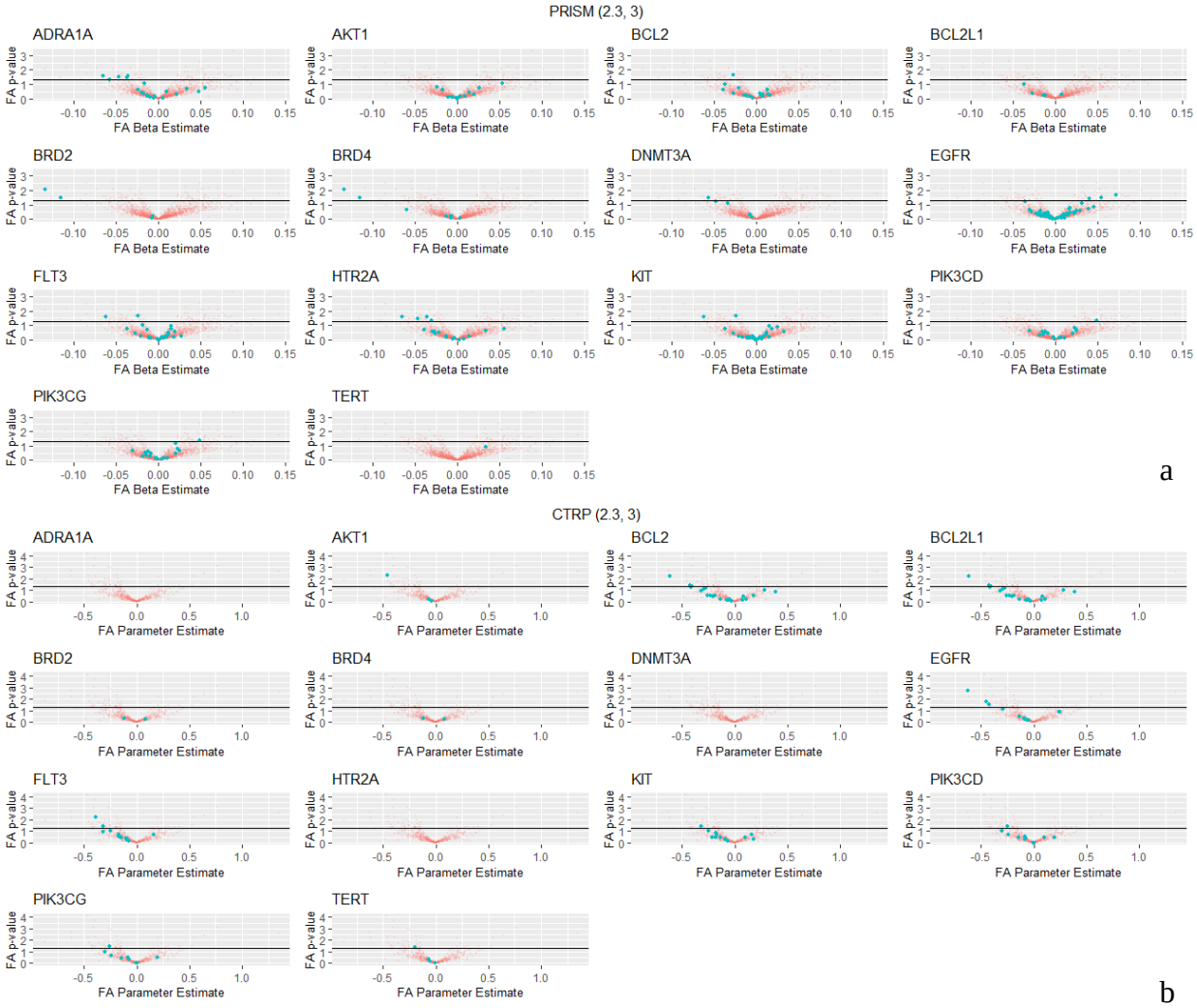


Figure 3.2: Distribution of regression results for threshold (2.3, 3). X-axis: β_{FA}, Y-axis: -log10 p-value for FA term. Red dot: all drugs in the dataset, Blue dot: drug that targets the gene or protein listed above the plot. Solid line is the 0.05 p-value cut-off for significance with smaller p-values above the line. a) Distribution for PRISM drugs, b) Distribution for CTRP drugs.

Looking at the distribution of drugs for each of the targets in Figure 3.2, DNMT3A, BRD2, and BRD4 have the strongest skews towards the negative β_{FA} estimates. With the exception of one drug in CTRP and PRISM, all drugs for BRD2 and BRD4 appear to have negative β_{FA} estimates. DNMT3A, which is observed only in PRISM, appears to have only negative β_{FA} estimates as well. Other genes appear to have a slightly mixed distribution based on the dataset. AKT1 appears to have a roughly symmetric distribution in PRISM, with the lowest p-value falling on the positive β_{FA} side of the violin plot, however in CTRP all the estimates for AKT1 are negative. Similarly, for FLT3 and KIT, the distribution in PRISM is roughly symmetric for positive and negative estimates, but there is more of a skew towards negative estimates in CTRP. However, in both PRISM and CTRP, there are only significant p-values for negative β_{FA} estimates for both genes, providing more confidence in the skew towards negative estimates.

EGFR also appears to have a more symmetric distribution for positive and negative estimates in PRISM, however the only significant p-values appear on the positive estimate side (Figure 3.2.a). In contrast, CTRP has primarily negative estimates for EGFR drugs, and there are only significant p-values for negative estimates (Figure 3.2.b). Similar patterns were seen for PIK3CD and PIK3CG, with one drug attaining a significant p-value in PRISM on the positive estimate side.

3.1.2 Threshold (2, 3)

Gene	PRISM	CTRP	Total
ADRA1A/B	4	0	4
FLT3	2	2	4

EGFR/ERBB2	0	4	4
DNMT1/3A	3	0	3
HTR2A/C	3	0	3
KIT/PDGFRB	2	1	3
BCL2/2L1/2L2	0	3	3
BRD2/3/4	3	0	3
AKT1	0	1	1
HDAC1/2	0	1	1
TERT	0	1	1
PIK3CG	0	1	1

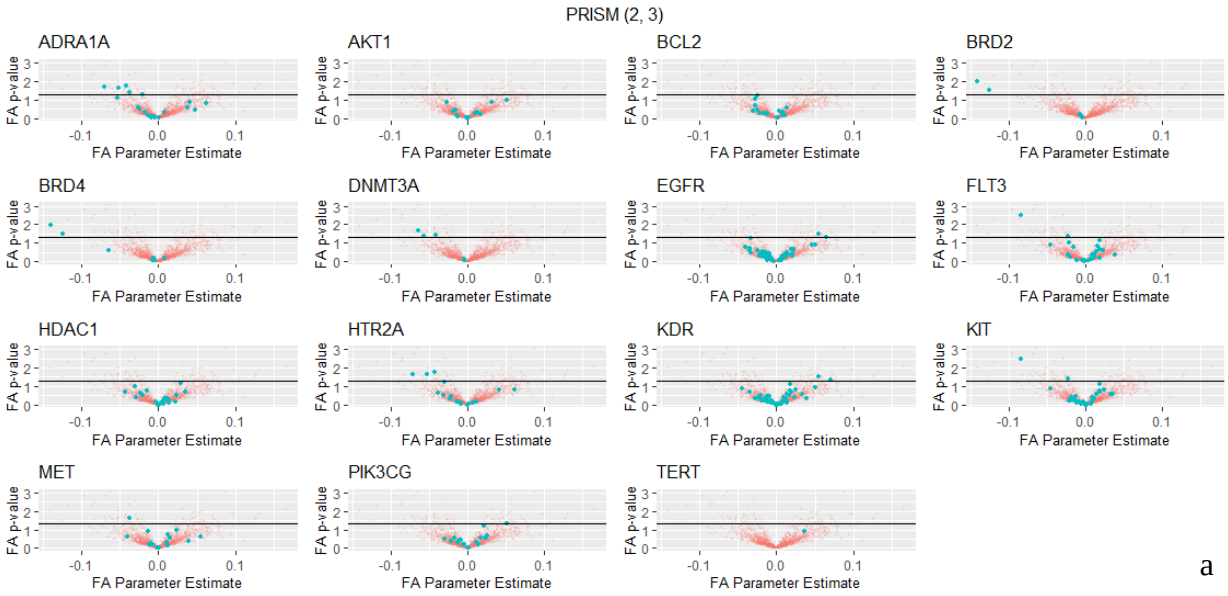
Table 3.2: Frequency of top targets among filtered drugs for threshold (2, 3)

For the threshold (2,3), the following genes of interest appeared: BRD2, BRD3, BRD4, DNMT3A, AKT1, TERT, and HDAC1. BRD2, BRD3, and BRD4 appeared together under the same three drugs as the previous threshold in PRISM, and I-BET151 and OTX015 had the largest β_{FA} estimates again. AKT1 and TERT appeared in CTRP under the same drugs as well. DNMT3A appeared under azacitidine again in PRISM, and it also appeared twice under decitabine, which was tested in two batches. HDAC1 was the new target to appear in this threshold and was seen in CTRP with the broad ID BRD-K80183349.

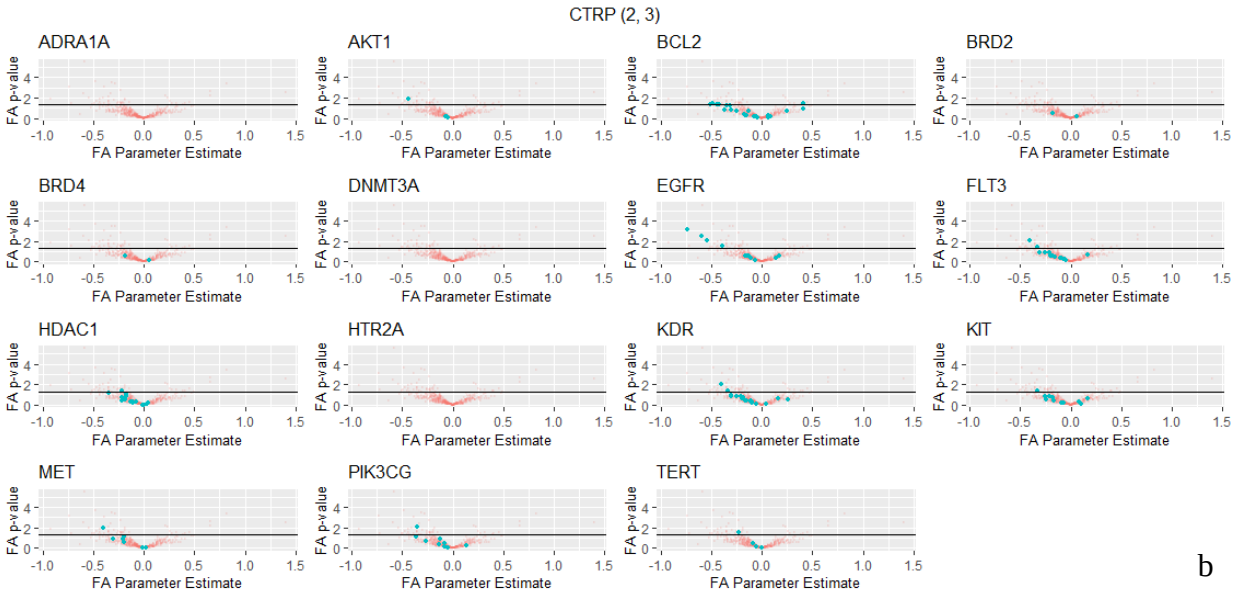
Like the last threshold, the targets FLT3, KIT, and PDGFRB were seen together in PRISM and CTRP under the same drugs. BCL2 did not appear in PRISM for this threshold, but it did appear under two of the three navitoclax combinations from the last threshold in CTRP. The third observation of BCL2 in CTRP for this threshold was under navitoclax on its own. EGFR had the same three drugs that were observed in the last threshold, in addition to lapatinib.

PIK3CG also appeared under PIK-93 again in CTRP for this threshold. GDSC had panobinostat again under this threshold, which appeared to be the only significant result.

Looking at the distribution of drugs in Figure 3.3, BRD2, BRD4, and DNMT3A have the strongest skews towards negative β_{FA} estimates. Again, with the exception of one drug in CTRP and PRISM, both of which have very weak p-values, all estimates for BRD2 and BRD4 are negative DNMT3A also appears to have only negative β_{FA} estimates in PRISM, with no data in CTRP to compare with. Three of the four drugs for DNMT3A also have a significant p-value. The distributions for AKT1, FLT3, and KIT appear to be the same as before, with all three having a stronger skew towards negative estimates in CTRP and only significant p-values for negative β_{FA} estimates (Figure 3.3.b). HDAC1 and MET also appeared to follow a similar distribution. EGFR also has the same contrast between PRISM and CTRP as in the previous threshold, with PRISM primarily having significant p-values for positive β_{FA} estimates and CTRP having only significant p-values for negative estimates. This pattern was observed again for PIK3CG as well. BCL2 appeared to have no strong skew, especially in CTRP where there were significant p-values for both positive and negative estimates (Figure 3.3.b).



a



b

Figure 3.3: Distribution of regression results for threshold (2, 3). X-axis: β_{FA}, Y-axis: -log10 p-value for FA term. Red dot: all drugs in the dataset, Blue dot: drug that targets the gene or protein listed above the plot. Solid line is the 0.05 p-value cut-off for significance with smaller p-values above the line. a) Distribution for PRISM drugs, b) Distribution for CTRP drugs.

3.1.3 Threshold (1.8, 3)

Gene	PRISM	CTRP	Total
EGFR	1	4	5
FLT3	2	3	5
KIT/PDGFRB	3	2	5
FLT1/KDR	1	3	4
PDGFRA	2	2	4
ERBB2	0	4	4
ADRA1A/D	3	0	3
BRD2/3/4	3	0	3
ADRA1B	2	0	2
HTR2A/C	2	0	2

MET	1	1	2
BCL2/2L1	0	2	2
DNMT3A	1	0	1
PIK3CA	1	0	1
PIK3CG	0	1	1

Table 3.3: Frequency of top targets among filtered drugs for threshold (1.8, 3)

For the threshold (1.8,3), the following genes of interest appeared: BRD2, BRD3, BRD4 and DNMT3A, which is fewer than the last two thresholds. BRD2, BRD3, and BRD4 appeared under the same three drugs as the previous two thresholds in PRISM, and I-BET151 and OTX015 had the largest β_{FA} estimates as observed the last two times. DNMT3A appeared only once under decitabine for this threshold. PIK3CA, a gene involved in AKT signaling, appeared for the first time in PRISM under the drug taselisib. Like the last threshold, the targets FLT3, KIT, and PDGFRB were seen together in PRISM and CTRP under the same drugs. Lenvatinib appeared for the first time under this threshold in CTRP, and it accounted for the increased frequencies of FLT1, KDR, and PDGFRA compared to the last threshold. MGCD-265 also appeared for the first time in PRISM, marking the first overlap between CTRP and PRISM. Another drug that appeared in PRISM under this threshold is semaxanib, which targets FLT1, KDR, KIT, PDGFRA, and PDGFRB.

BCL2 did not appear under this threshold in PRISM again, however it did appear under the drugs TW-37 and navitoclax in CTRP, marking a departure from the navitoclax combinations that appeared in previous thresholds. EGFR had the same four drugs that were observed for the last threshold in CTRP, and rociletinib appeared in PRISM, marking the first

EGFR inhibitor in the dataset to meet both criteria. GDSC did not appear to have any significant results for these targets under this threshold.

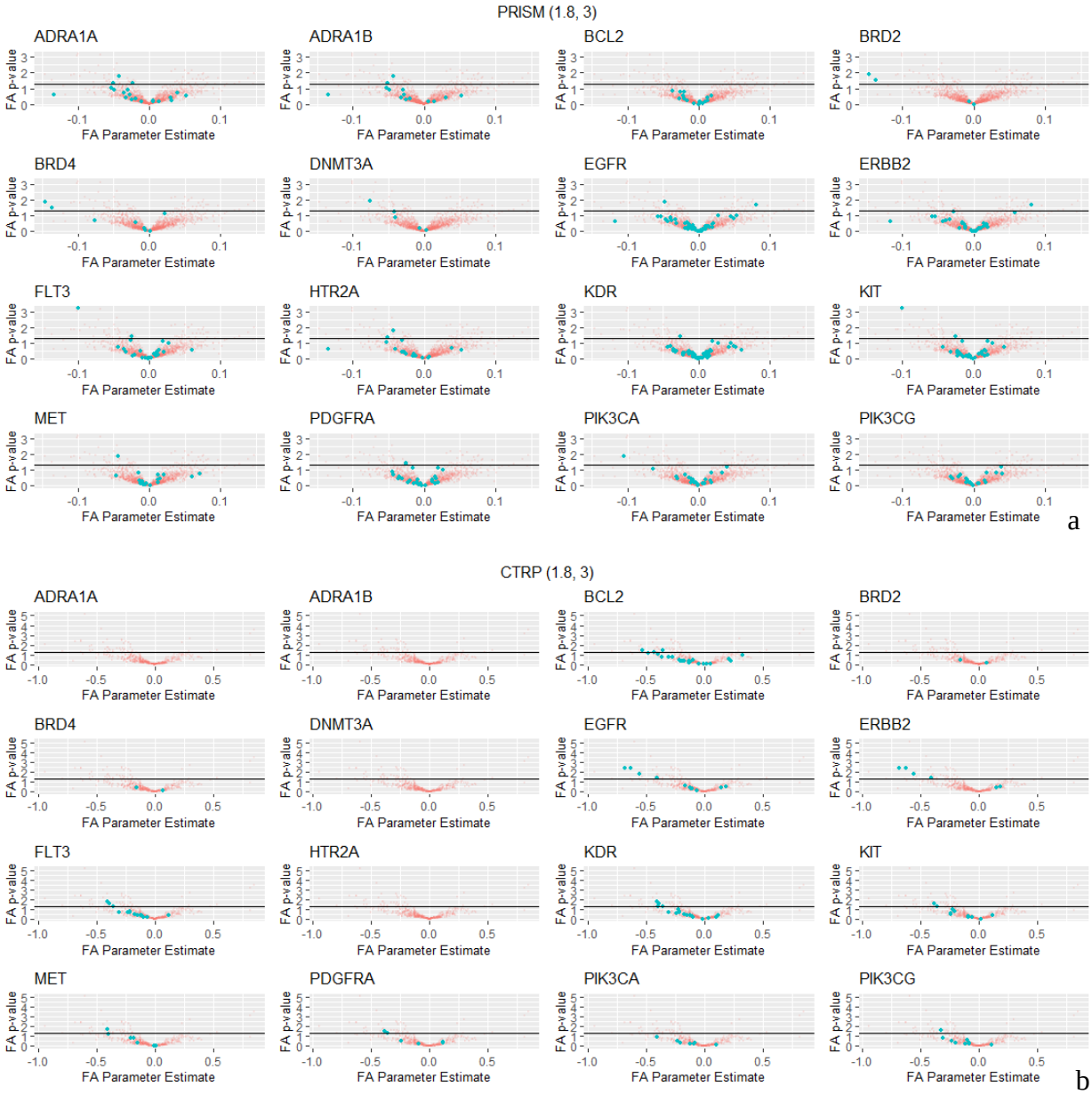


Figure 3.4: Distribution of regression results for threshold (1.8, 3). X-axis: β_{FA}, Y-axis: -log10 p-value for FA term. Red dot: all drugs in the dataset, Blue dot: drug that targets the gene or protein listed above the plot. Solid line is the 0.05 p-value cut-off for significance with smaller p-values above the line a) Distribution for PRISM drugs, b) distribution for CTRP drugs

Like before, BRD2, BRD4, and DNMT3A have the strongest skews towards negative β_{FA} estimates. The remaining genes – BCL2, EGFR, ERBB2, FLT3, KIT, KDR, MET, PDGFRA, PIK3CA, and PIK3CG - appear to have a roughly symmetric distribution for both positive and negative estimates in PRISM but a skew towards negative estimates in CTRP (Figure 3.4). In particular, the distribution looks more even for both sides of the x-axis compared to previous thresholds. However, the following genes only had significant p-values for negative estimates in PRISM: FLT3, KDR, KIT, PDGFRA, MET, and PIK3CA (Figure 3.4.a).

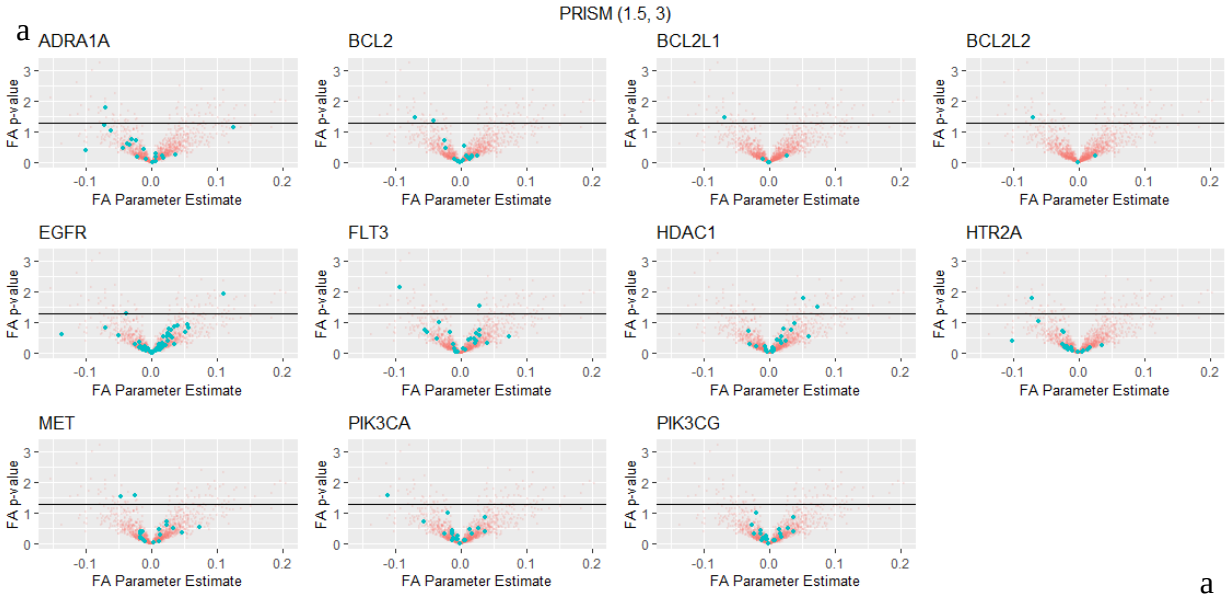
3.1.4 Threshold (1.5,3)

Gene	**PRISM**	**CTRP**	**Total**
BCL2	2	3	5
BCL2L1	1	3	4
BCL2L2	1	2	3
MET	2	0	2
EGFR/ERBB2	0	2	2
ADRA1A/B/D	1	0	1
FLT3	1	0	1
HTR2A/C	1	0	1
PIK3CA	1	0	1
HDAC1/2/3/6/8	0	1	1
PIK3CG	0	1	1

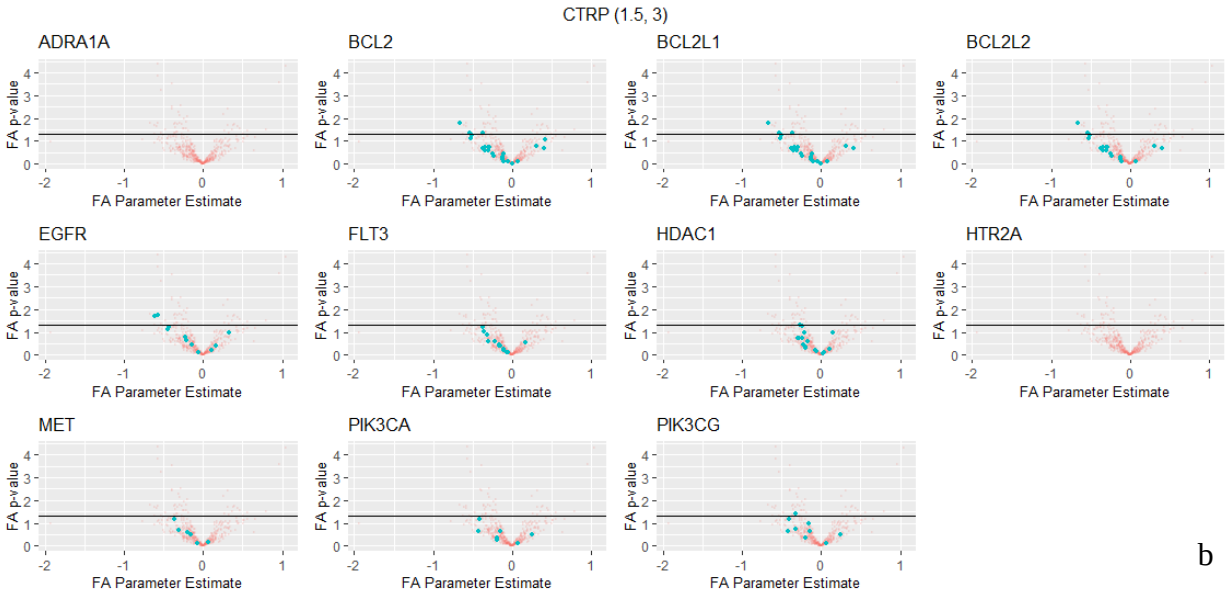
Table 3.4: Frequency of top targets among filtered drugs for threshold (1.5, 3)

The only notable gene of interest from the pre-defined list above is HDAC1, which appeared in CTRP with a vorinostat-carboplatin combination. This marks the first threshold in which BRD2, BRD3, BRD4 and DNMT3A did not appear under the filtered drugs list in PRISM. FLT3 appeared again under GTP-14565 in PRISM, and PIK3CA appeared under taselisib again as well. MET appeared under MGCD-265 in PRISM, as well as tivantinib, which has not appeared in previous thresholds. BCL2 appeared under navitoclax for both PRISM and CTRP in this threshold, and PRISM also had gambogic-acid. The other two BCL2 drugs in CTRP were TW-37, which appeared in the last threshold, and a navitoclax combination that was seen in the first threshold. GDSC did not appear to have any significant results for this threshold.

Compared to previous thresholds, there are no targets in PRISM that appear to have a skew towards negative estimates. HDAC1 appears to have a slight skew towards positive estimates, with only significant p-values on the positive side of the x-axis (Figure 3.5.a). FLT3 also has a significant p-value for a positive estimate in this threshold. However as before, most drugs in CTRP have a skew towards negative estimates for these targets (Figure 3.5.b).



a



b

Figure 3.5: Distribution of regression results for threshold (1.5, 3). X-axis: β_{FA}, Y-axis: -log10 p-value for FA term. Red dot: all drugs in the dataset, Blue dot: drug that targets the gene or protein listed above the plot. Solid line is the 0.05 p-value cut-off for significance with smaller p-values above the line. a) Distribution for PRISM drugs, b) Distribution for CTRP drugs.

3.1.5 Threshold (2.3, 3.5)

Gene	PRISM	CTRP	Total
EGFR	1	4	5
BCL2	1	3	4
ERBB2	0	4	4
FLT3	1	2	3
BCL2L1/2	0	3	3
PDGFRB	1	1	2
FLT1/KDR	0	2	2
PIK3CD	0	2	2
PIK3CG	0	2	2
HDAC3	0	2	2

AKT1	0	1	1
HDAC1/2/6/8	0	1	1
MET	0	1	1
PIK3CB	0	1	1

Table 3.5: Frequency of top targets among filtered drugs for threshold (2.3, 3.5)

For this threshold, the genes of interest that appeared were: AKT1 and HDAC1, each of which appeared once in CTRP. AKT1 appeared under MK-2206 again and HDAC1 appeared under tacedinaline again. All four drugs for EGFR and ERBB2 in CTRP were the same as for previous thresholds, but gefitinib was the new drug for EGFR in PRISM. As for the remaining targets seen in PRISM, GTP-14564 accounted for both FLT3 and PDGFRB, and venetoclax accounted for BCL2, both of which were observed in previous thresholds. Most of the drugs observed in CTRP were seen in previous thresholds as well. BCL2 and BCL2L1 were observed under navitoclax and two navitoclax combinations again. The drugs axitinib and MGCD-265 also appeared again in CTRP, accounting for FLT3, FLT1, KDR, PDGFRB, and MET. PIK3CG and HDAC1 appeared under PIK-93 and tacedinaline respectively as well. The new drugs present in this threshold were Repligen 136, ZSTK474, and CAL-101 which target HDAC3, PIK3CB/D/G, and PIK3CD respectively.

There were many drugs identified in GDSC that target some of the genes listed above. In particular, navitoclax and lapatinib overlap with CTRP. Other notable drugs include CP-724714, osimertinib, and dacomitinib, which target ERBB2 and EGFR. ACY-1215, trichostatin-a, and panobinostat were also the notable HDAC inhibitors for this threshold. Amuvatinib and savolitinib were the c-MET inhibitors, and BIBR-1532 targets TERT.

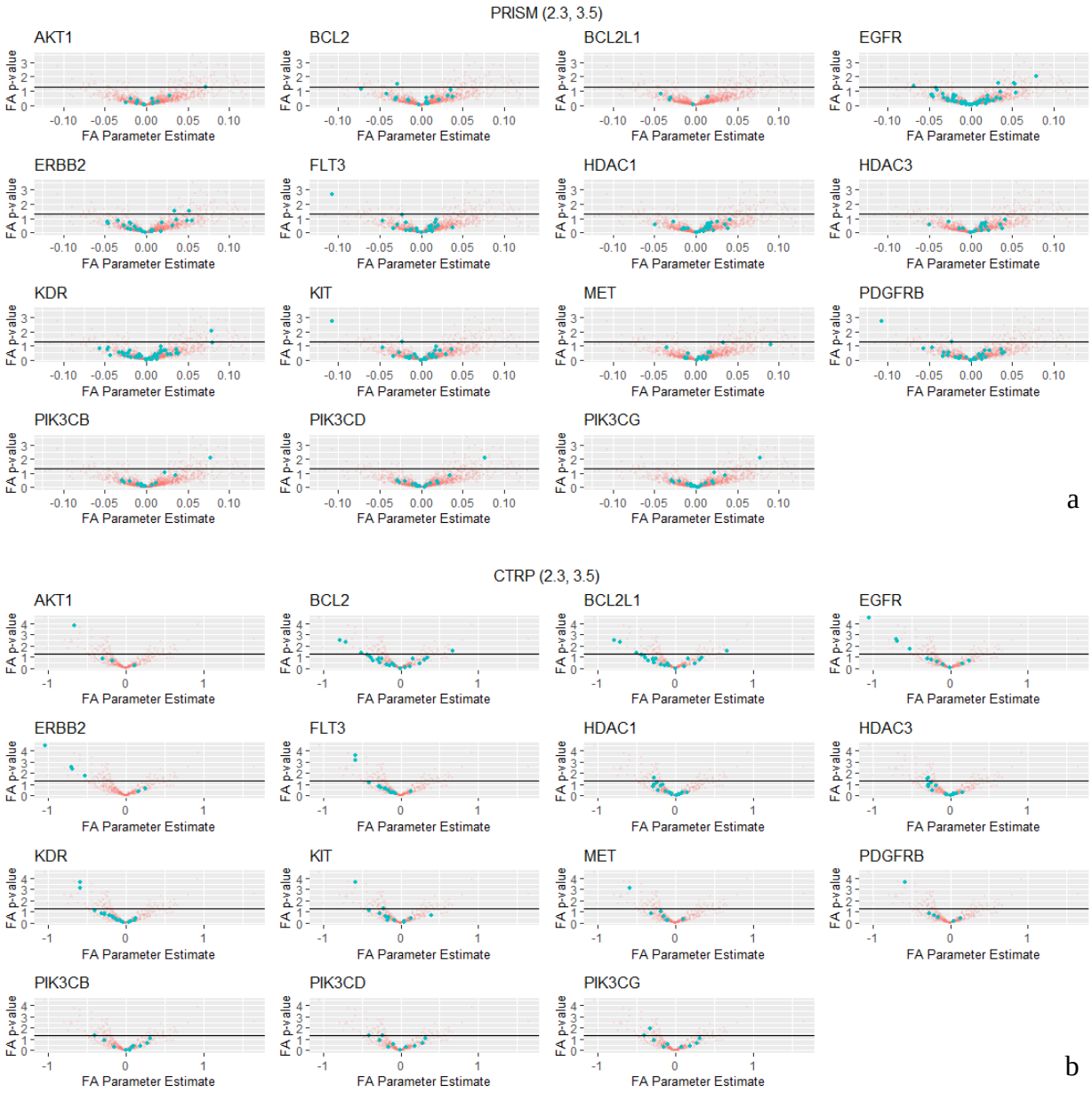


Figure 3.6: Distribution of regression results for threshold (2.3, 3.5). X-axis: β_{FA}, Y-axis: -log10 p-value for FA term. Red dot: all drugs in the dataset, Blue dot: drug that targets the gene or protein listed above the plot. Solid line is the 0.05 p-value cut-off for significance with smaller p-values above the line. a) Distribution for PRISM drugs, b) Distribution for CTRP drugs.

None of the targets appear to be skewed towards the negative estimates in PRISM and compared to other thresholds, there appears to be more positive estimates with a significant p-value (Figure 3.6.a). HDAC1 and HDAC3 appear to be slightly skewed towards the positive

estimates. In CTRP, AKT1, EGFR, ERBB2, FLT3, MET, and PDGFRB appear to have the strongest skews towards the negative estimates. These targets all have at most two drugs with a positive estimate, and all the positive estimates appear to have relatively weak p-values (Figure 3.6.b). HDAC1, HDAC3, KIT and KDR also appear to have a skew towards negative estimates, however they do not appear to be as strongly skewed as the other six targets since they have more drugs with positive estimates. The PI3K genes, BCL2, and BCL2L1 don't appear to have a skew in any direction. Unlike PRISM, however, all drugs with a significant p-value have a negative estimate, with the exception of one drug that targets BCL2 and BCL2L1 (Figure 3.6.b).

3.1.6 Thresholds (2,3.5), (1.5, 3.5), and (1.8, 3.5)

	Threshold								
	2, 3.5			**1.8, 3.5**			**1.5, 3.5**		
Gene	PRISM	CTRP	**Total**	PRISM	CTRP	**Total**	PRISM	CTRP	**Total**
EGFR	2	4	**6**	2	4	**6**	1	4	**5**
ERBB2	0	4	**4**	0	4	**4**	0	4	**4**
FLT3	1	2	**3**	1	2	**3**	1	2	**3**
BCL2	0	3	**3**	0	3	**3**	0	2	**2**
BCL2L1/2	0	3	**3**	0	3	**3**	0	2	**2**
BRD2/3/4	0	0	**0**	2	0	**2**	0	0	**0**
KIT/ PDGFRB	1	1	**2**	1	1	**2**	1	1	**2**
FLT1/ KDR	0	2	**2**	0	2	**2**	0	2	**2**
PIK3CG	0	2	**2**	0	1	**1**	0	1	**1**

ADRA1/B/D	1	0	**1**	1	0	**1**	2	0	**2**
HTR2A/C	1	0	**1**	1	0	**1**	1	0	**1**
PIK3CA	1	0	**1**	1	0	**1**	1	0	**1**
HDAC3	0	1	**1**	0	0	**0**	0	0	**0**
MET	0	1	**1**	0	1	**1**	1	1	**2**
MTOR	0	1	**1**	0	1	**1**	0	1	**1**
PDGFRA	0	1	**1**	0	1	**1**	0	1	**1**
PIK3CB	0	1	**1**	0	0	**0**	0	0	**0**
PIK3CD	0	1	**1**	0	0	**0**	0	0	**0**

Table 3.6: Combined frequency of top targets among filtered drugs for thresholds (2, 3.5), (1.8, 3.5), and (1.5, 3.5). Each set of three columns specific to the labeled threshold

Across all three thresholds, the only new drug was sirolimus in CTRP, which is a MTOR inhibitor. Other drugs observed under all three thresholds in CTRP include neratinib, afatinib, canertinib, lapatinib, MGDC-265, axitinib, navitoclax, and PIK-93. For the threshold (2, 3.5) in CTRP, PIK3CB, PIK3CD, and PIK3CG appeared under ZST474, and HDAC3 appeared under Repligen 136. In PRISM, the drugs GTP-14564 and taselisib appeared across all three thresholds. For EGFR in PRISM, gefitinib appeared in the thresholds (2, 3.5) and (1.8, 3.5), and rociletinib appeared in (1.8, 3.5) and (1.5, 3.5). BRD2, BRD3, and BRD4 appeared again after several thresholds under (1.8, 3.5) in PRISM with the drugs I-BET151 and OTX015. Bromosporine did not appear in the results for the threshold despite being observed in previous results with I-BET151 and OTX015. Under the threshold (1.5, 3.5), the drug tivantinib was observed with MET as its target.

Some of the notable drugs from the threshold (2, 3.5) in GDSC were ACY-1215, amuvatinib, panobinostat, osimertinib, and BIBR-1532 which were observed in the previous threshold. Pilaralisib, a PI3K inhibitor, appeared under this threshold as well. For the threshold (1.8, 3.5), the drugs amuvatinib, panobinostat, and osimertinib appeared again. Other ERBB2 and EGFR inhibitors under this threshold were CP-724714, dacomitinib, and lapatinib. Lapatinib overlapped with CTRP under this threshold. For the threshold (1.5, 3.5), afatinib was the only notable drug and overlapped with CTRP.

Like under previous thresholds, most of the targets do not appear to have a strong skew in any direction with PRISM (Figures 3.7.a, 3.8.a, 3.9.a). For the threshold (2, 3.5), FLT3, PIK3CA, and KIT only have a significant p-value on the negative estimate side. BRD2 and BRD4 have drugs that have nearly significant p-values and negative estimates, however, there does not appear to be as strong of a skew as observed in previous thresholds. For (1.8, 3.5), BRD2 and BRD4 have two significant drugs as seen earlier, and there is a slightly more noticeable skew for BRD4 since only two drugs appear to have a positive estimate. Like the previous threshold, FLT3, KIT, and PIK3CA only have significant p-values for drugs with negative estimates (Figure 3.8.a). The distributions of targets for the threshold (1.5, 3.5) were similar to those in (2, 3.5), with no strong skew for any of the targets, including BRD2 and BRD4. FLT3, KIT, MET, and PIK3CA were the three targets with significant p-values only for negative estimates (Figure 3.9.a).

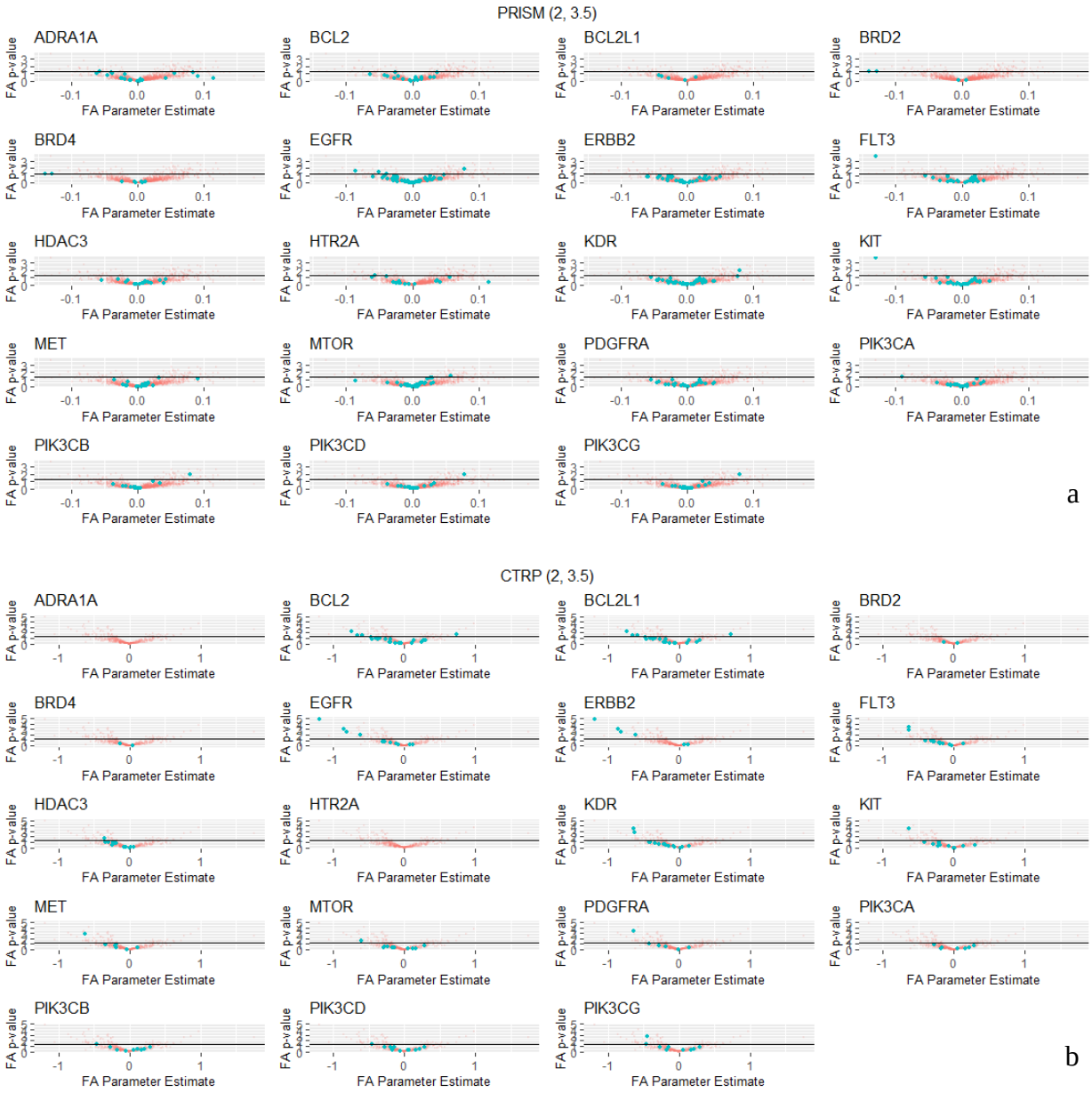


Figure 3.7: Distribution of regression results for threshold (2, 3.5). X-axis: β_{FA}, Y-axis: -log10 p-value for FA term. Red dot: all drugs in the dataset, Blue dot: drug that targets the gene or protein listed above the plot. Solid line is the 0.05 p-value cut-off for significance with smaller p-values above the line. a) Distribution for PRISM drugs, b) Distribution for CTRP drugs.

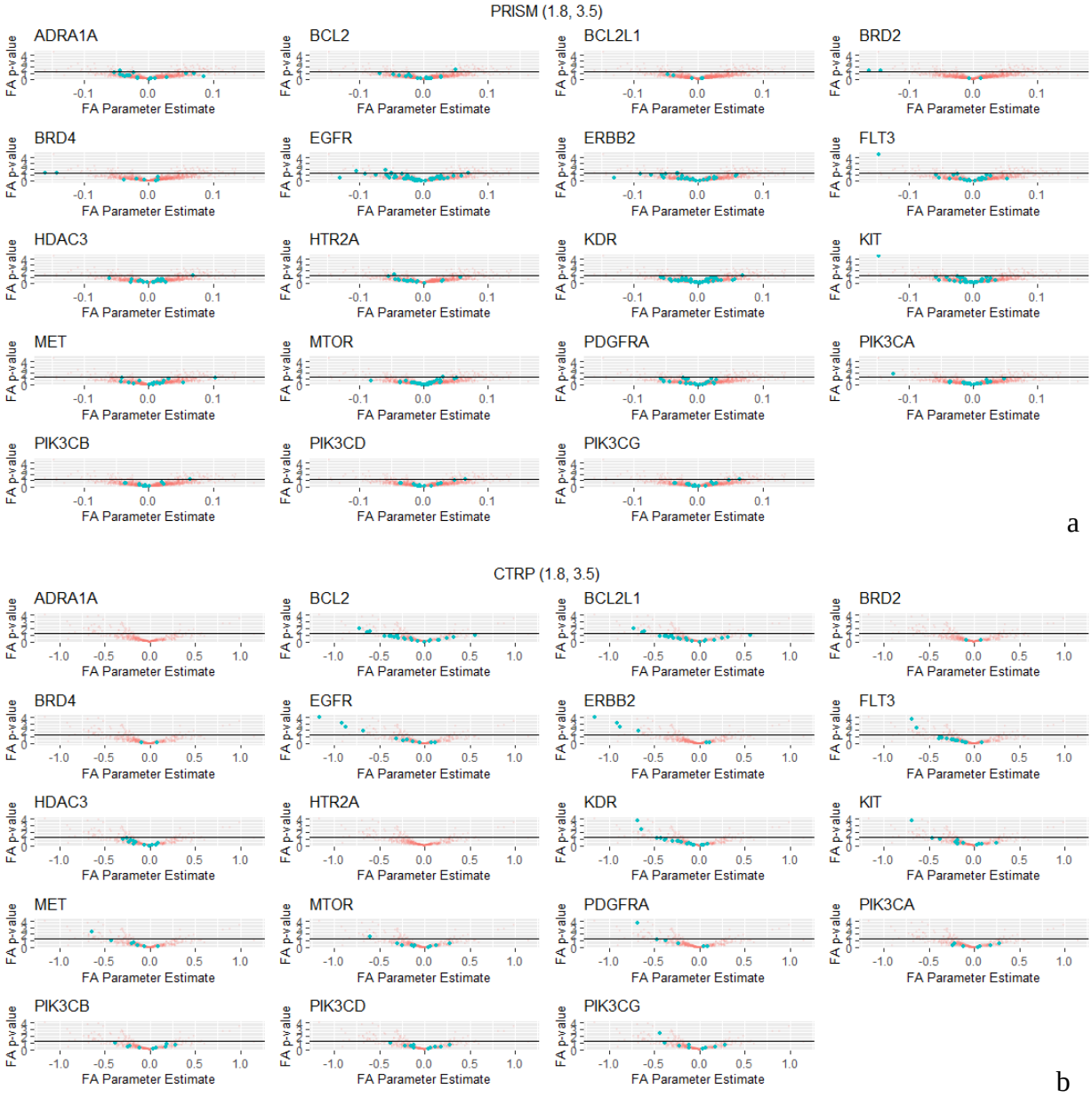


Figure 3.8: Distribution of regression results for threshold (1.8, 3.5). X-axis: β_{FA}, Y-axis: -log10 p-value for FA term. Red dot: all drugs in the dataset, Blue dot: drug that targets the gene or protein listed above the plot. Solid line is the 0.05 p-value cut-off for significance with smaller p-values above the line. a) Distribution for PRISM drugs, b) Distribution for CTRP drugs.

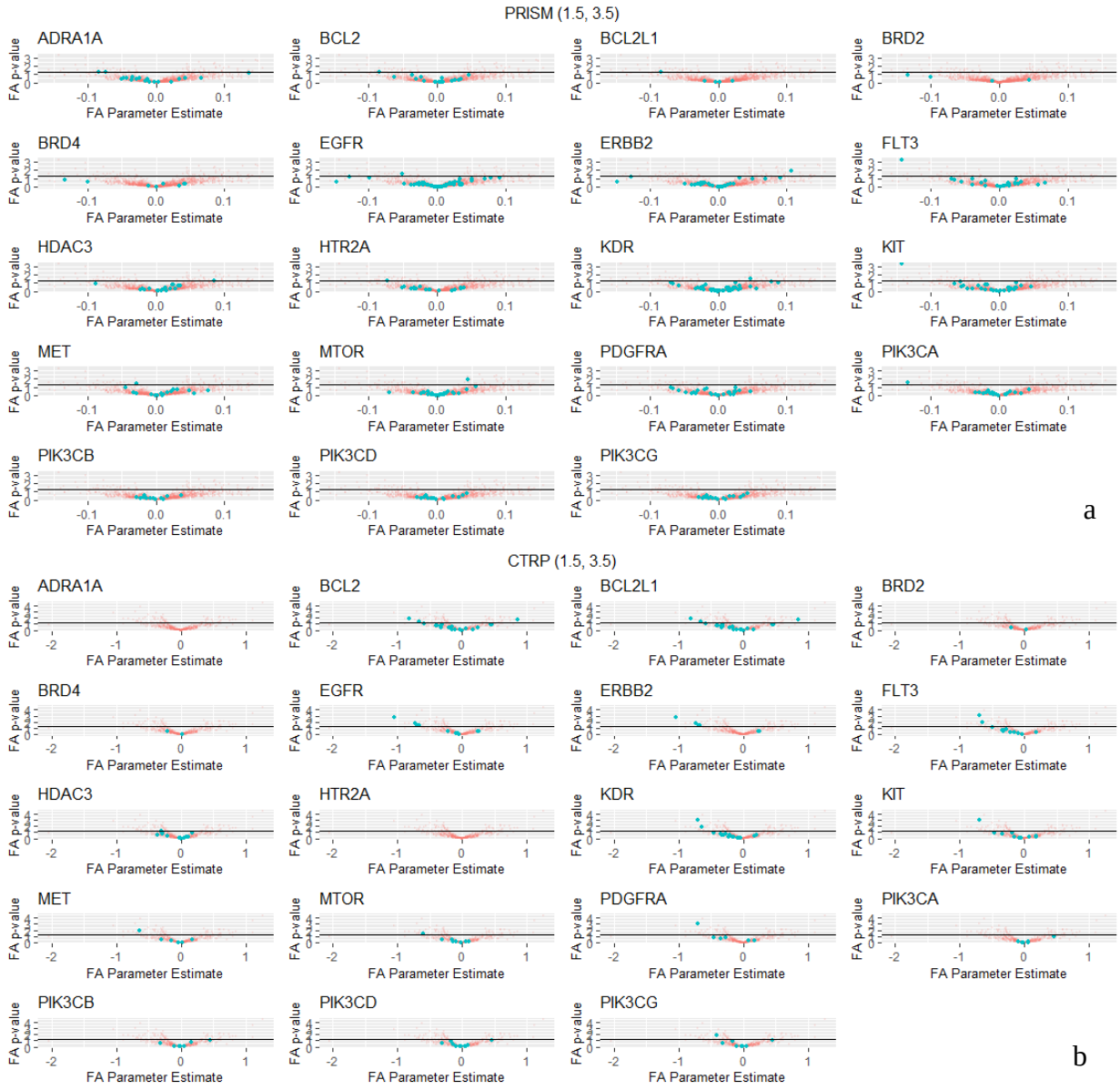


Figure 3.9: Distribution of regression results for threshold (1.5, 3.5). X-axis: β_{FA}, Y-axis: -log10 p-value for FA term. Red dot: all drugs in the dataset, Blue dot: drug that targets the gene or protein listed above the plot. Solid line is the 0.05 p-value cut-off for significance with smaller p-values above the line. a) Distribution for PRISM drugs, b) Distribution for CTRP drugs.

For CTRP, the targets EGFR, ERBB2, FLT3, HDAC3, KDR, MET, and PDGFRA had the strongest skews towards negative estimates, with only one or two drugs that had a positive parameter estimate (Figures 3.7.b, 3.8.b, 3.9.b). This was observed across all three thresholds, with the exception of HDAC3 under the threshold (1.5, 3.5). The other targets do not appear to

have a strong skew in any direction. Apart from BCL2 and BCL2L1, none of the targets had a positive estimate with a significant p-value in CTRP.

3.1.7 Thresholds (2.3, 4), (2,4), and (1.8, 4)

	Threshold								
	2.3, 4			**2,4**			**1.8, 4**		
Gene	PRISM	CTRP	**Total**	PRISM	CTRP	**Total**	PRISM	CTRP	**Total**
EGFR	11	4	**15**	10	4	**14**	9	4	**13**
ERBB2	7	3	**10**	7	3	**10**	6	3	**9**
HDAC3	3	5	**8**	4	6	**10**	3	5	**8**
HDAC1	3	4	**7**	4	5	**9**	3	4	**7**
HDAC6	3	4	**7**	4	5	**9**	3	4	**7**
HDAC8	3	4	**7**	4	5	**9**	3	4	**7**
HDAC2	2	4	**6**	2	5	**7**	2	4	**6**
FLT3	2	3	**5**	4	3	**7**	3	5	**8**
KDR	1	3	**4**	2	3	**5**	2	6	**8**
BCL2	1	2	**3**	1	2	**3**	2	0	**2**
FLT1	0	3	**3**	0	3	**3**	0	5	**5**
ADRA1A	2	0	**2**	2	0	**2**	2	0	**2**
HTR2A	2	0	**2**	2	0	**2**	2	0	**2**
MET	1	1	**2**	0	1	**1**	0	1	**1**
BCL2L1/2	0	2	**2**	0	2	**2**	0	0	**0**
KIT	0	2	**2**	1	3	**4**	1	4	**5**
PDGFRB	0	2	**2**	1	2	**3**	1	3	**4**

AKT1	1	2	**3**	1	2	**3**	1	0	**1**
ATR	1	0	**0**	1	0	**0**	2	0	**2**
DNMT3A	1	0	**1**	2	0	**2**	2	0	**2**
MTOR	1	0	**1**	1	2	**3**	1	1	**2**
PIK3CG	1	0	**1**	1	1	**2**	1	1	**2**
PDE4A	1	0	**1**	1	0	**1**	0	0	**0**
AKT3	0	1	**1**	0	1	**1**	0	1	**1**
PDGFRA	0	1	**1**	1	1	**2**	1	3	**4**
PIK3CA	0	0	**0**	0	1	**1**	0	1	**1**
PIK3CB	0	0	**0**	0	1	**1**	0	1	**1**

Table 3.7: Combined frequency of top targets among filtered drugs for thresholds (2.3, 4), (2, 4), and (1.8, 4). Each set of three columns specific to the labeled threshold

The genes of interest that appeared in these thresholds were HDAC1, AKT1, ATR, DNMT3A, MTOR, PDE4A, and AKT3. All HDAC targets in PRISM were under the drugs belinostat, trichostatin-a, and resminostat, none of which appeared under previous thresholds. Belinostat also appeared for HDAC1 in CTRP, along with the drugs tacedinaline, vorinostat, and entinostat. AKT1 appeared in PRISM for all thresholds under the drug A-674563. For CTRP, AKT1 was observed under the thresholds (2.3, 4) and (2, 4) with the drugs MK-2206 and AT7867, which also targets AKT3. MTOR appeared in PRISM under the drug VE-822, which also targets ATR and PIK3CG, and in CTRP with the drugs sirolimus and temsirolimus. In addition to VE-822, ATR was observed under the drug VE-821 in PRISM for the threshold (1.8, 4). Similar to previous thresholds, azacitidine appeared in PRISM for DNMT3A in addition to

SGI-1027 for the thresholds (2, 4) and (1.8, 4). PDE4A also appeared in PRISM under the drug GSK256066.

Under these thresholds, there was a large increase in the number of EGFR inhibitors in PRISM that had a negative β_{FA} estimate and a p-value less than 0.05. Rociletinib and gefitinib were the two drugs that overlapped with previous thresholds. The new ones to appear in PRISM were XL-647, pelitinib, lapatinib, osimertinib, poziotinib, dacomitinib, WZ8040, and afatinib. Of these drugs, half of them were observed in CTRP and GDSC. Most instances of ERBB2 fell with the EGFR inhibitors lapatinib, poziotinib, dacomitinib, afatinib, and XL-647. Tucatinib, an ERBB2 inhibitor, was a new drug that appeared in PRISM across all three thresholds. In CTRP, the drugs neratinib, canertinib, and lapatinib appeared again for EGFR and ERBB2 across all three thresholds. The fourth instance of EGFR was WZ8040, which overlapped with GDSC in addition to lapatinib.

FLT3 also appeared under different drugs in PRISM compared to previous thresholds: TG-02, pacritinib, and ENMD-2076. GTP-14564 appeared under the thresholds (2, 4) and (1.8, 4) as well. The drugs XL-647 and ENMD-2076 also target KDR, accounting for the frequency observed across the thresholds. FLT3 appeared under axitinib, lenvatinib, and MGCD-265 in CTRP as observed under previous thresholds, but additional drugs also appeared for these three thresholds: pazopanib and tivozanib. BCL2 appeared across all three thresholds in PRISM with the drug venetoclax, and it appeared under navitoclax combinations in CTRP.

The results for GDSC were similar to other thresholds. Under (2.3, 4), the drugs ACY-1215, afatinib, BIBR-1532, entinostat, gefitinib, osimertinib, navitoclax, and lapatinib were identified. Under (2,4), the following drugs were identified: ACY-1215, amuvatinib, AT7867, BIBR-1532, CP-724714, dacomitinib, entinostat, gefitinib, lapatinib, navitoclax, osimertinib,

panobinostat, trichostatin-a, and vorinostat. The results under (1.8,4) were a combination of drugs from both thresholds.

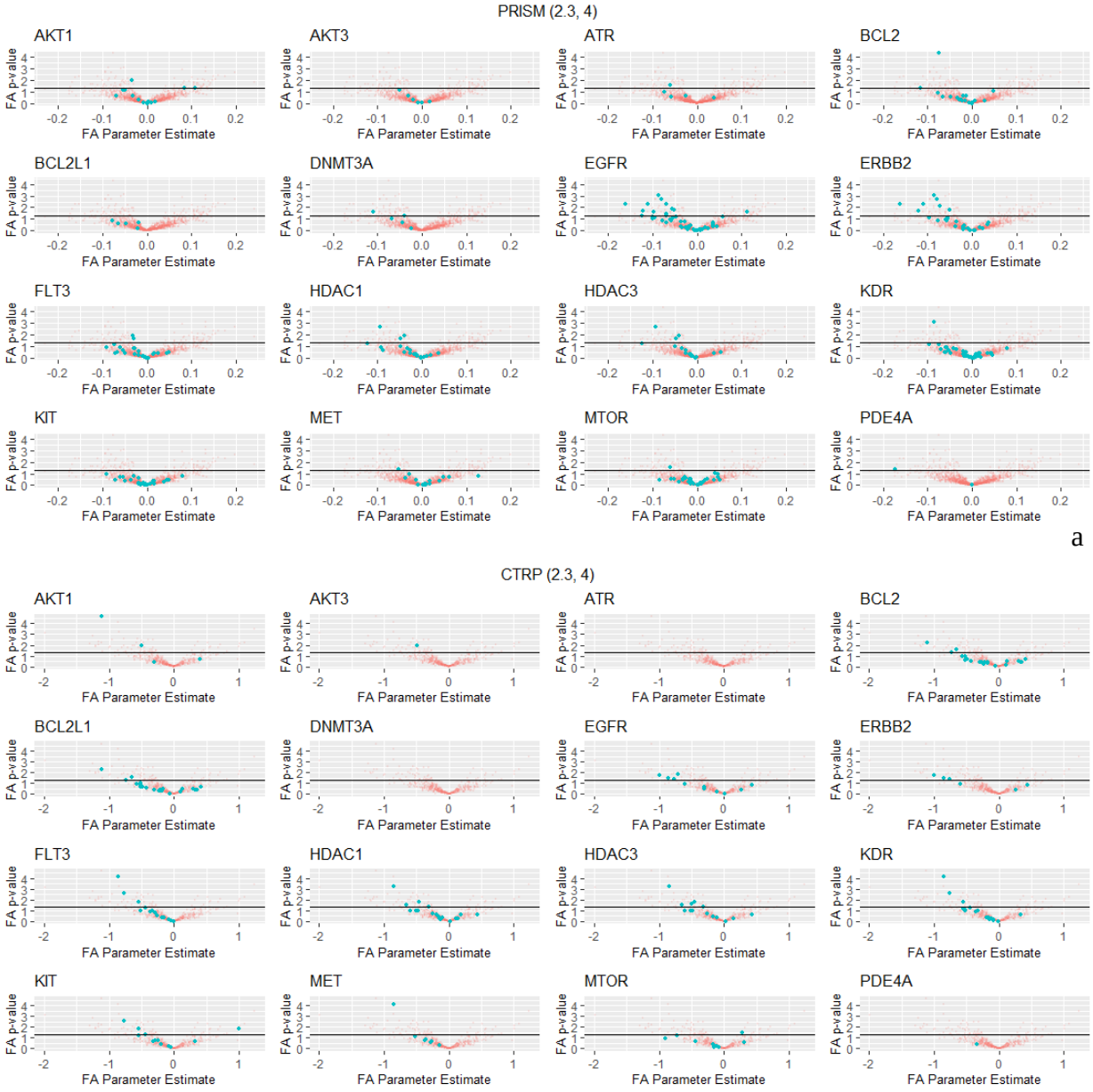


Figure 3.10: Distribution of regression results for threshold (2.3, 4). X-axis: β_{FA}, Y-axis: -log10 p-value for FA term. Red dot: all drugs in the dataset, Blue dot: drug that targets the gene or protein listed above the plot. Solid line is the 0.05 p-value cut-off for significance with smaller p-values above the line. a) Distribution for PRISM drugs, b) Distribution for CTRP drugs.

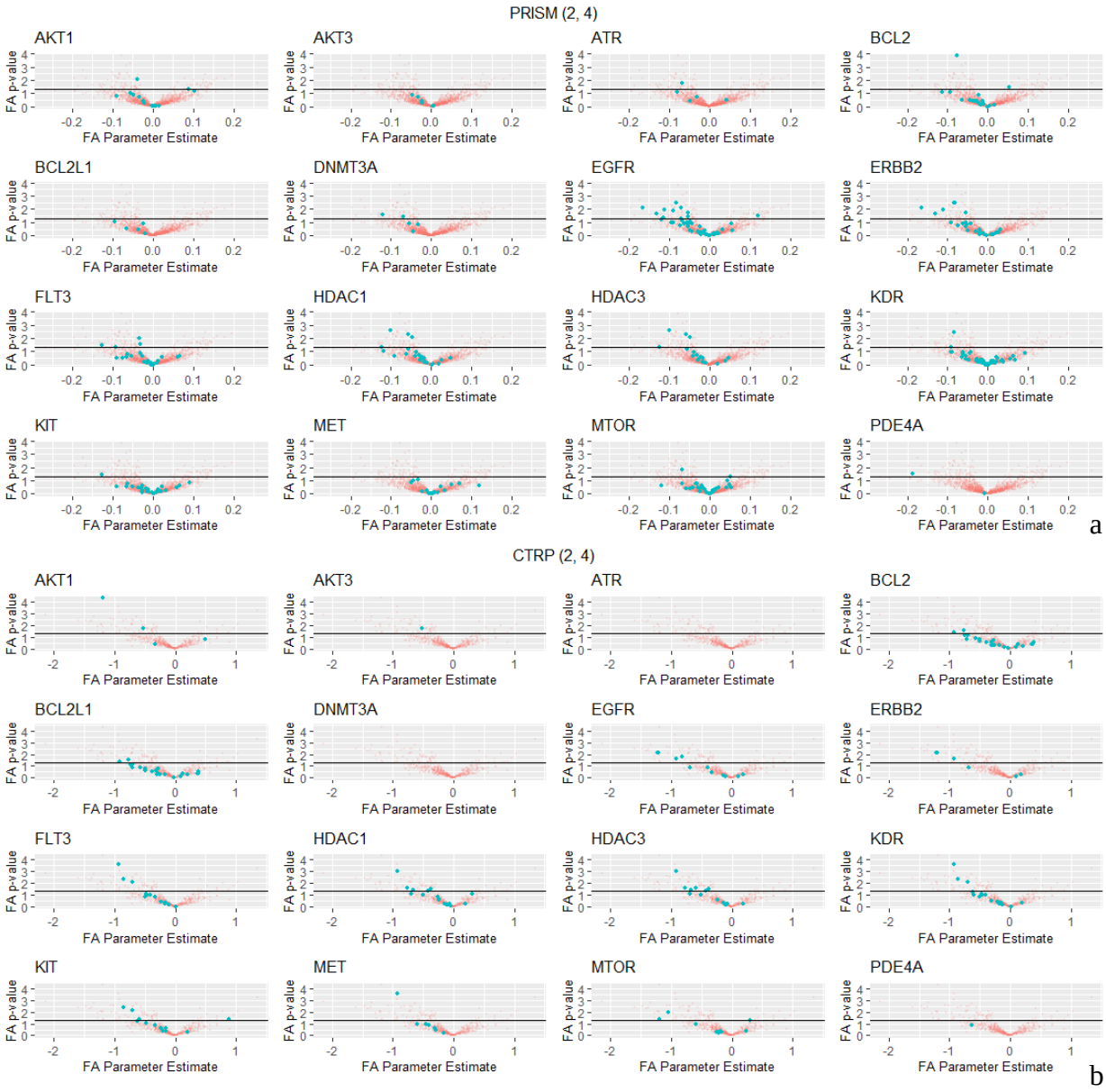


Figure 3.11: Distribution of regression results for threshold (2, 4). X-axis: β_{FA}, Y-axis: -log10 p-value for FA term. Red dot: all drugs in the dataset, Blue dot: drug that targets the gene or protein listed above the plot. Solid line is the 0.05 p-value cut-off for significance with smaller p-values above the line. a) Distribution for PRISM drugs, b) Distribution for CTRP drugs.

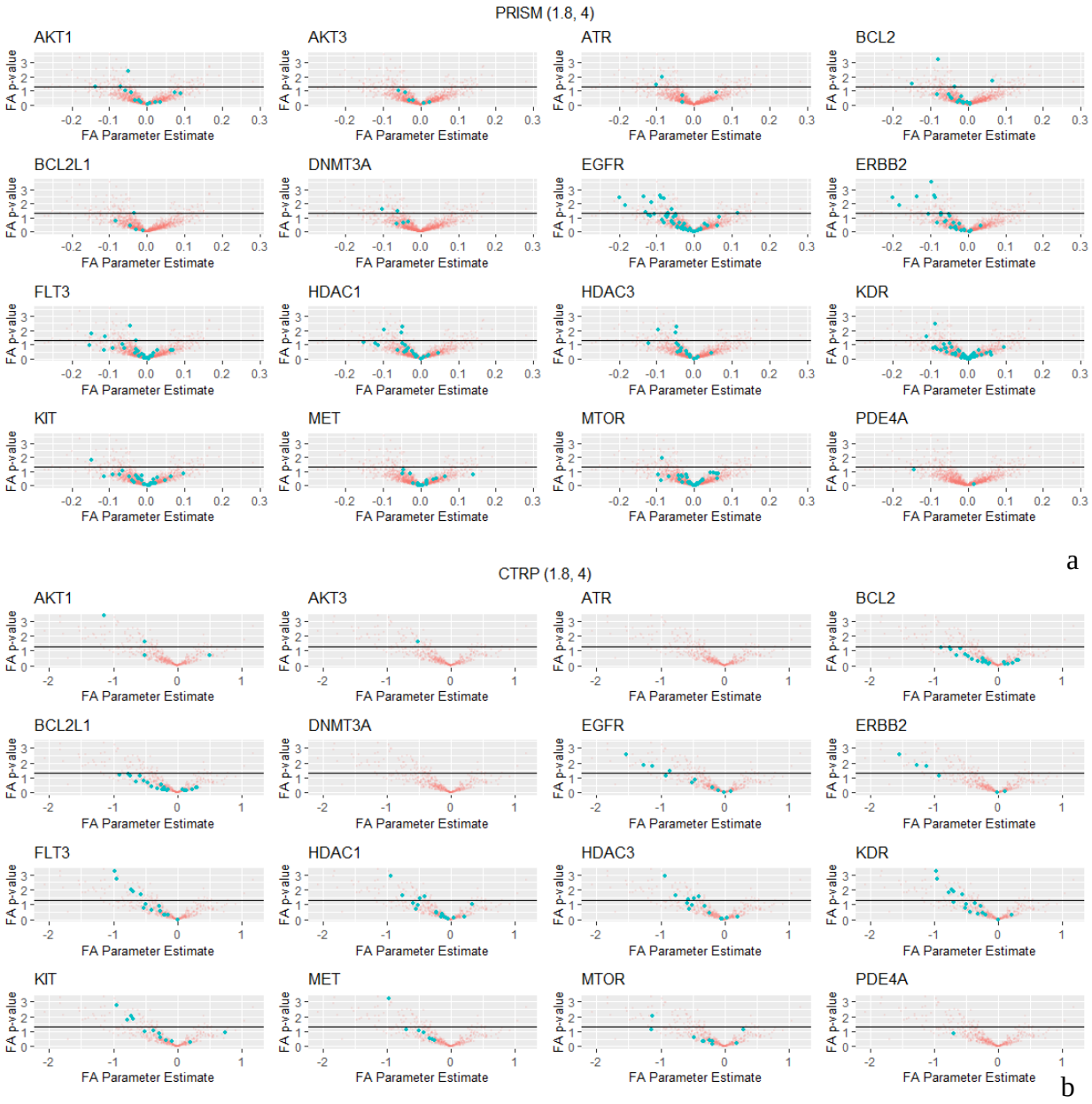


Figure 3.12: Distribution of regression results for threshold (1.8, 4). X-axis: β_{FA}, Y-axis: -log10 p-value for FA term. Red dot: all drugs in the dataset, Blue dot: drug that targets the gene or protein listed above the plot. Solid line is the 0.05 p-value cut-off for significance with smaller p-values above the line. a) Distribution for PRISM drugs, b) Distribution for CTRP drugs.

Unlike previous thresholds, there appears to be a stronger skew for the following targets in PRISM: AKT3, ATR, DNMT3A, and HDAC3 (Figures 3.10.a, 3.11.a, 3.12.a). For all these targets, all drugs, with the exception of two or three, had negative estimates and there were only significant p-values for negative estimates. These targets had a consistent skew across all three

thresholds. AKT1 also appeared to have a strong skew towards negative estimates, however there were two drugs with significant, or nearly significant, p-values and positive estimates for the thresholds (2.3, 4) and (2, 4) (Figures 3.10.a, 3.11.a). Other targets with a less-strongly defined skew towards negative estimates were: BCL2, EGFR, ERBB2, FLT3, and HDAC1. Compared to the other targets, these have more drugs with a positive estimate, but roughly three fourths of the drugs appear to have negative estimates. BCL2 and EGFR also appear to have one or two drugs with a significant p-value for a positive estimate (Figures 3.10.a, 3.11.a, 3.12.a). The remaining targets KDR, KIT, MET, and MTOR have a roughly even split between positive and negative estimates. In CTRP, most targets appeared to have a strong skew towards negative estimates as seen in previous thresholds. Among all targets, BCL2 and BCL2L1 had the weakest skew since they had the most drugs with a positive estimate. The remaining targets had around two drugs at most with a positive estimate. KIT and MTOR are the only two targets in CTRP that appear to have a significant p-value for a positive estimate (Figures 3.10.b, 3.11.b).

3.1.8 Threshold (1.5,4)

Gene	PRISM	CTRP	Total
EGFR	5	2	7
KDR	2	5	7
ERBB2	4	2	6
FLT1	1	4	5
FLT3	1	4	5
KIT	1	4	5

BCL2	2	2	4
HDAC1/2/3/6/8	1	3	4
PDGFRA/B	1	3	4
MET	1	1	2
MTOR	1	1	2
PIK3CG	1	1	2
BCL2L1/2	0	2	2
AKT1	1	1	2
ATR	1	0	1
BRD2/3/4	1	0	1
TERT	0	1	1

Table 3.8: Frequency of top targets among filtered drugs for threshold (1.5, 4)

For the threshold (1.5, 4), AKT1 appeared once in CTRP under MK-2206 and once under A-674563 in PRISM. HDAC appeared once in PRISM with trichostatin-a and three times in CTRP under tacedinaline, entinostat, and belinostat. ATR and MTOR appeared once in PRISM with VE-822, and MTOR appeared once in CTRP under sirolimus. BRD2, BRD3, and BRD4 reappeared under this threshold in PRISM with I-BET-762, marking the first threshold in which OTX015 was not observed with it. TERT appeared once in CTRP with the drug combination piperlongumine:MST-312 (1:1 mol/mol).

EGFR appeared five times in PRISM with the drugs XL-647, pelitinib, rociletinib, poziotinib, and dacomitinib, all of which were observed in previous thresholds. CTRP only had two drugs for EGFR, neratinib and lapatinib, which also target ERBB2. FLT3 appeared under the drugs MGCD-265, pazopanib, axitinib, and lenvatinib in CTRP, and most of these drugs

also target KDR, KIT, and PDGFRB. Pazopanib also appeared in PRISM along with TG-02 for FLT3. Similar to previous thresholds, navitoclax appeared in CTRP for BCL2 and venetoclax appeared in PRISM. Under GDSC, the following drugs were identified as observed in previous thresholds: ACY-1215, afatinib, gefitinib, lapatinib, and vorinostat. One of the new drugs that did not show up previously across any dataset was JQ1, which is a BRD4 inhibitor.

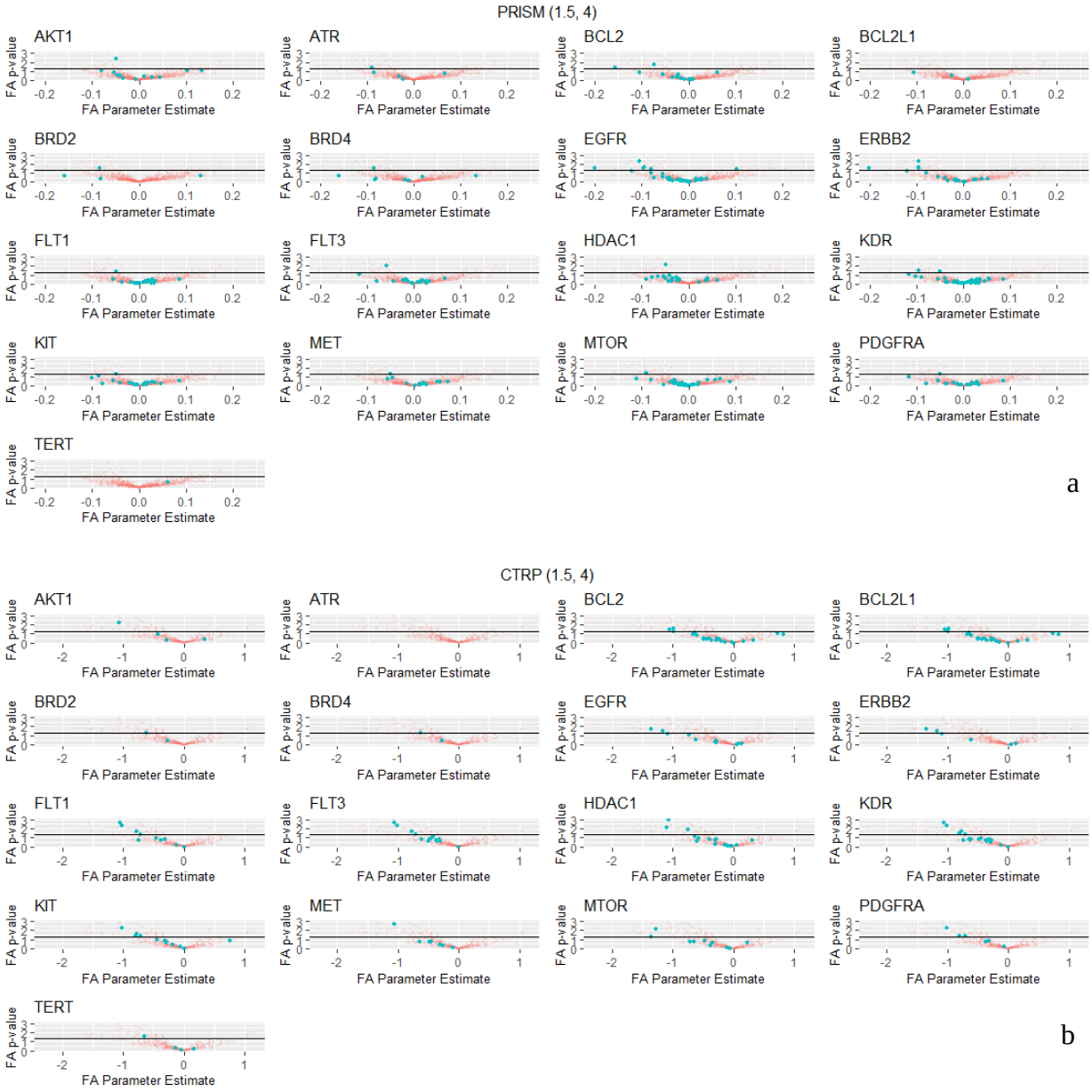


Figure 3.13: Distribution of regression results for threshold (1.5, 4). X-axis: β_{FA}, Y-axis: -log10 p-value for FA term. Red dot: all drugs in the dataset, Blue dot: drug that targets the gene or

protein listed above the plot. Solid line is the 0.05 p-value cut-off for significance with smaller p-values above the line. a) Distribution for PRISM drugs, b) Distribution for CTRP drugs.

In PRISM, ATR, BCL2, BCL2L1, BRD2, and BRD4 have the strongest skews towards negative estimates, with all but one or two of the drugs having negative estimates (Figure 3.13.a). None of these targets appeared to have a significant p-value for a positive estimate. EGFR, ERBB2, and HDAC1 also appear to have a skew towards negative estimates, but they have more drugs with positive estimates in comparison to the other five targets, and EGFR has one significant p-value for a positive estimate (Figure 3.13.a). The remaining drugs appear to have a roughly even split between positive and negative estimates, with some targets having a slight skew towards positive estimates. As seen in previous thresholds, almost all targets, with the exception of BCL2 and BCL2L1, have a strong skew towards negative estimates. Only one to two drugs have a positive estimate for most of these targets and there appear to be no positive estimates with a significant p-value (Figure 3.13.b).

3.2 Effect of Driver Genes and Tissue-specific Effects

3.2.1 Tissue-specific Associations

Across all thresholds, there appeared to be drugs that had a significant p-value and a negative β_{FA} estimate within specific tissue types. These results were generated by linear regressions modeled after equation 2.1.3, which was run for each drug-tissue combination. For example, in PRISM under the thresholds (2.3, 3), (2, 3), (1.8, 3), and (1.8, 3.5), the BRD inhibitors OTX015, I-BET151, and bromosporine had negative β_{FA} estimates in lung samples. Venetoclax also had negative estimates in lung, pancreas, and breast samples under the thresholds (2.3, 3) and (2.3, 3.5). GTP-14564 had negative estimates in bone and liver samples under the first eight

thresholds. DNMT3A inhibitors had more varying associations. Under (2.3, 3) and (2, 3), azacitidine had a negative estimate in the lung samples. Decitabine had negative estimates in the pancreas and gastric samples for (2, 3) and (1.8, 3). Across all thresholds, EGFR and HDAC inhibitors primarily had negative estimates in lung samples. AKT1 and ATR/MTOR inhibitors A-674563 and VE-822 also had negative estimates in lung samples.

CTRP also had numerous drug-tissue associations. The AKT1 inhibitor MK-2206 had negative estimates in lymphocyte, colorectal, and lung samples for the thresholds (2.3, 3) and (2, 3). This was also observed for MK-2206 with colorectal and lung samples under (2, 3.5) and (1.8, 3.5). Across the thresholds, the EGFR inhibitors neratinib, canertinib, afatinib, and lapatinib had numerous associations with breast, lung, and urinary tract samples. Under the first two thresholds, neratinib had negative estimates in urinary tract, breast, lung, uterus, and lymphocyte tissues. Canertinib and afatinib also had negative estimates in breast and urinary tract tissues. Sirolimus, the MTOR inhibitor, had negative estimates in colorectal, central nervous system, and lung samples for the threshold (1.5, 3.5). Negative estimates for sirolimus were also observed in colorectal and lung samples under (1.8, 3.5), (2, 4), (1.8, 4), and (1.5, 4). Axitinib and MGCD-265, the FLT3 inhibitors, also had many associations. For instance, under the threshold (1.8, 3.5), axitinib had negative estimates in colorectal, lung, central nervous system, pancreas, and urinary tract samples. MGCD-265 had negative estimates in central nervous system, gastric, and lung samples. The association between gastric samples and MGCD-265 appeared across the last eight thresholds from (2.3, 3.5) and onwards.

3.2.2 Impact of Driver Genes

Looking at the copy number distribution for these targets, many do not appear to contribute to the maximum copy number of FA+ samples and are likely not the driving force behind the

difference in drug response. For PRISM, the genes with more than two log2 copy numbers greater than 3 are ERBB2 and EGFR, both of which are known driver genes (Figure 3.14). In CTRP, ERBB2 and MET appear to have more samples with a log2 copy number greater than 3. EGFR has a few as well, but not as much compared to the other two genes (Figure 3.15).

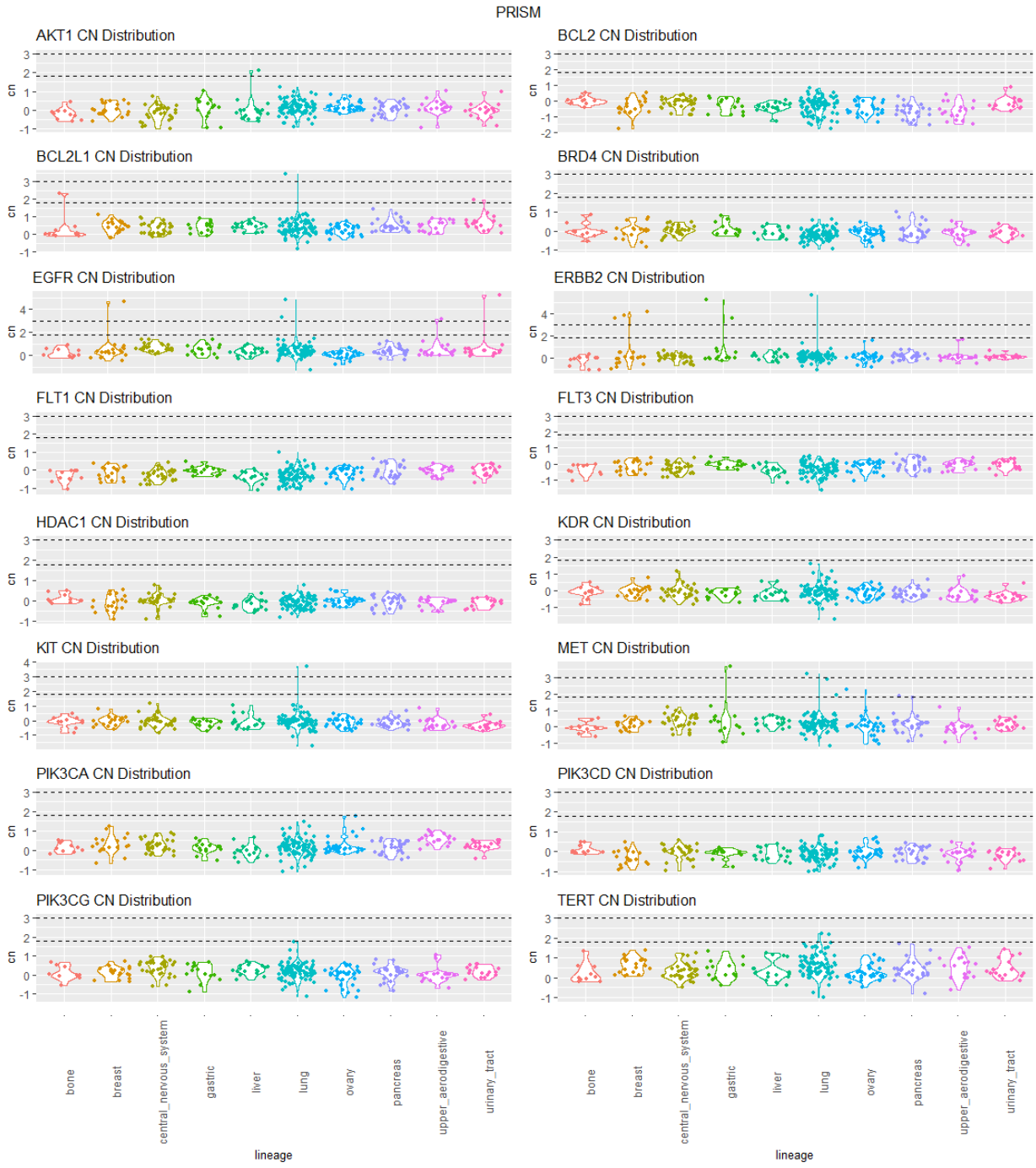


Figure 3.14: Distribution of the log2CN for genes observed in first linear regression. PRISM samples, broken down by tissue type.

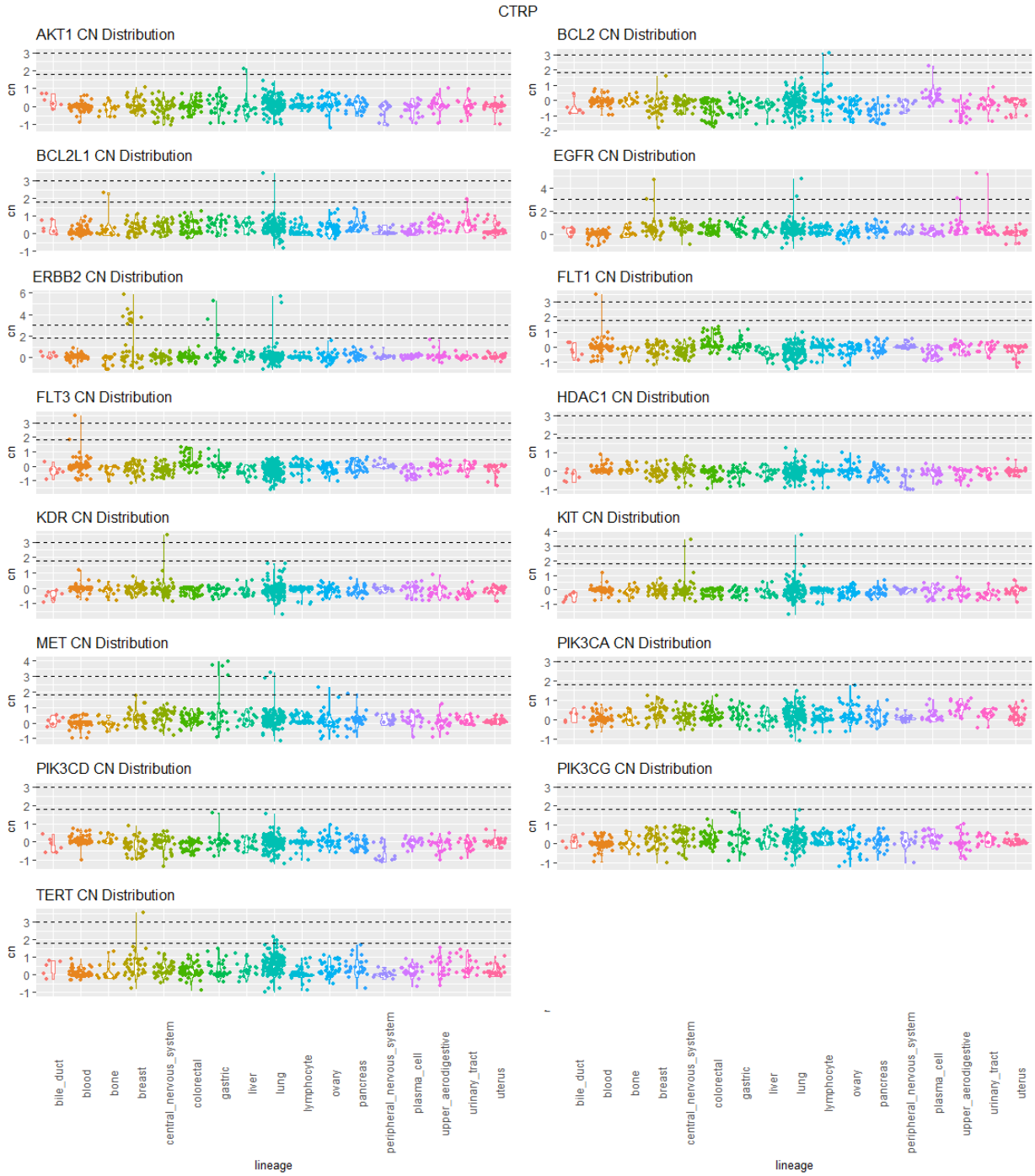


Figure 3.15: Distribution of the log2CN for genes observed in first linear regression. CTRP samples, broken down by tissue type.

Apart from EGFR and ERBB2, other driver genes could contribute to the differences within specific tissues. For PRISM, there are a few liver samples with high amplifications of CCND1 and a significant number of lung samples with highly amplified MYC (Figure 3.16). For CTRP, the following gene-tissue pairs may have an impact on the results: CCND1 in the liver, MYC in the lung, MYCL in the lung, and MYCN in the lung and nervous system (Figure 3.17). As noted earlier, ERBB2 may also affect the results in breast samples (Figure 3.17). One notable pattern is the association of BRD2, BRD3, BRD4, EGFR, and DNMT3A in lung tissue within PRISM. MYC is one of the most highly amplified genes in lung samples and is known to have interactions with BRD proteins and DNMT3A. Across lung samples in PRISM, and all samples in PRISM, ACH-000481 has the highest maximum log2CN of approximately 6.78 due to a MYC amplification.

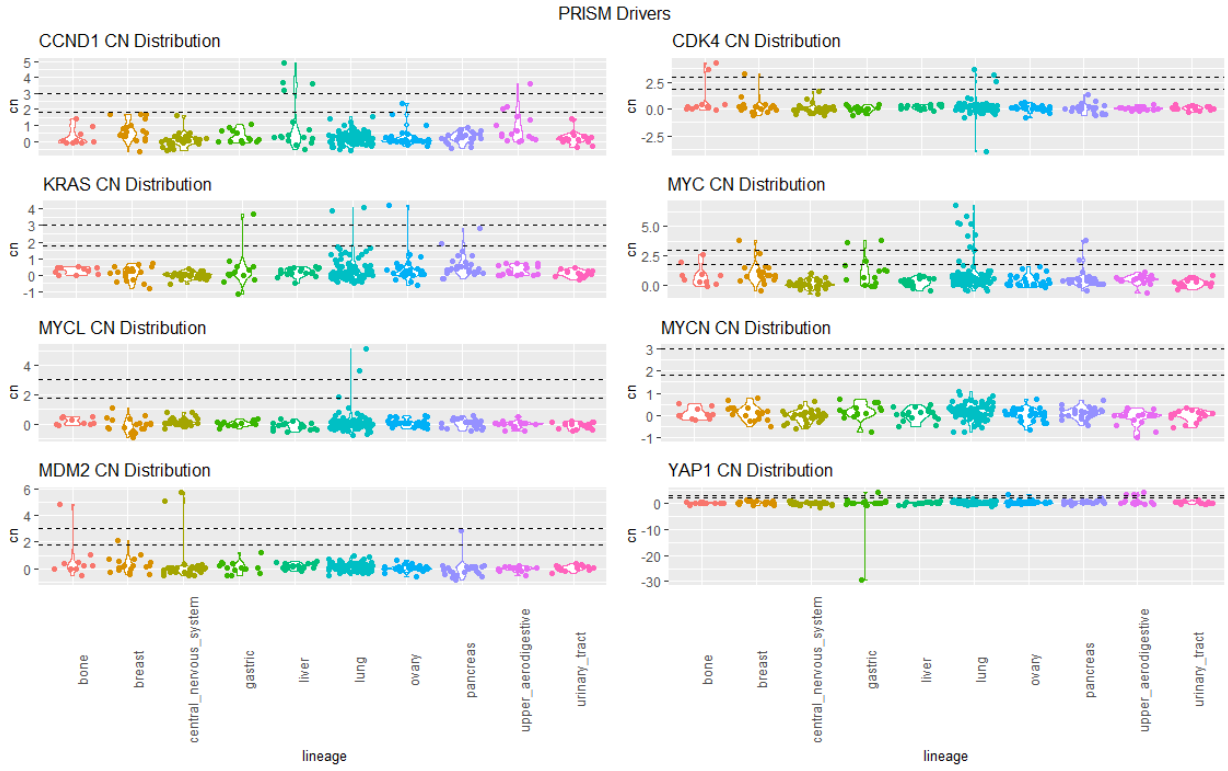


Figure 3.16: Distribution of the log2CN for known driver genes. PRISM samples, broken down by tissue type.

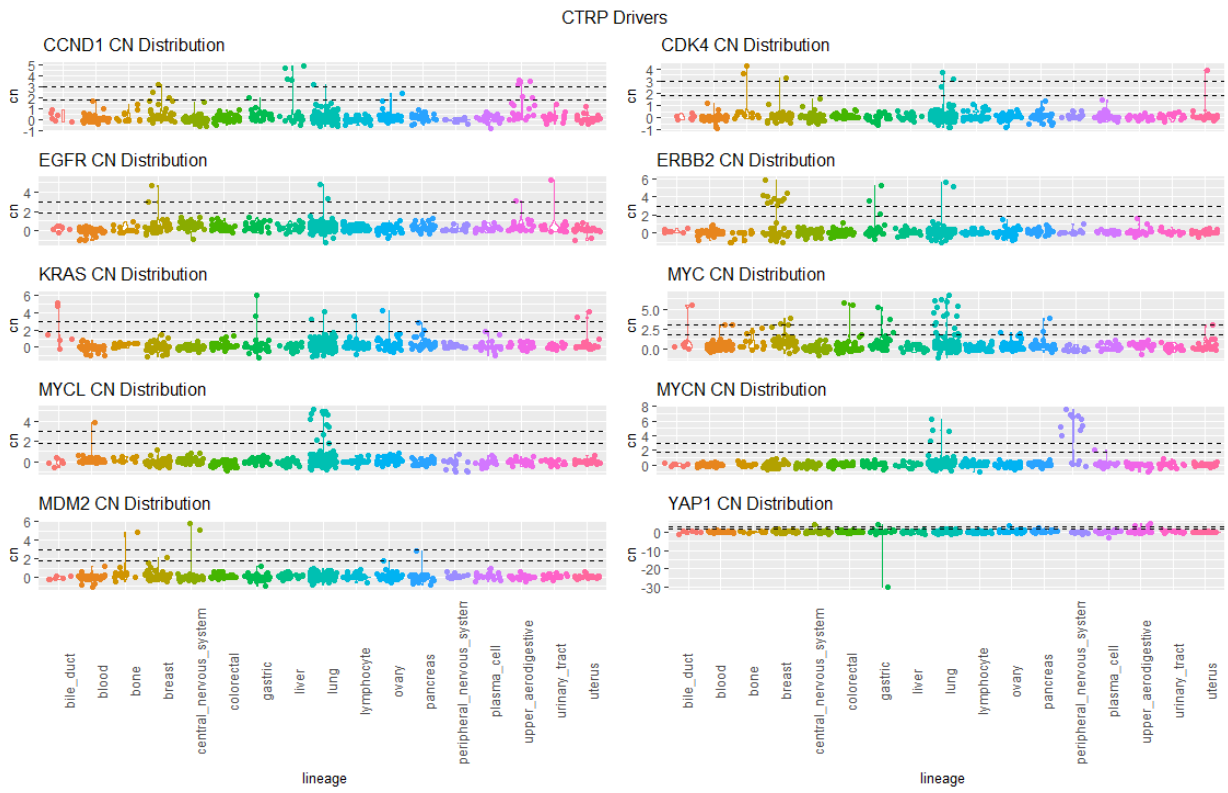


Figure 3.17: Distribution of the log2CN for known driver genes. CTRP samples, broken down by tissue type.

To check if the MYC amplification is driving many of the associations observed above, the second linear model was used to compare lung samples with MYC amplification greater than 3 and lung samples with a maximum log2CN less than 1.8. These samples were classified as FA+ and FA- respectively. As predicted, the BRD inhibitors I-BET151, OTX015, and bromosporine displayed significant negative β_{FA} estimates for the comparison between high-MYC CN lung samples and FA- lung samples. This aligns with the observation of the negative estimate for all three drugs in lung samples. As the Table 3.9 demonstrates, the β_{FA} was larger for the MYC-specific comparisons for the three drugs, and OTX015 and bromosporine's β_{FA} estimates had significant p-values. Similarly, rociletinib and taselisib's β_{FA} estimates became larger for the MYC-specific comparison, however their p-values were just below 0.1 and not 0.05. Decitabine had the least significant change among the drugs, with a p-value of roughly 0.5

for this comparison, and the pan-cancer estimate was larger compared to the MYC-specific comparison. Apart from EGFR, BRD4, and DNMT3A, the two FLT3 inhibitors crenolanib and GTP-14564 also had larger β_{FA} estimates for the MYC-specific comparison, and both had significant p-values around 0.01.

Drug	Target	Pan-cancer	MYC-specific
Rociletinib	EGFR	-0.04832583	-0.1183139
I-BET151	BRD2/3/4	-0.1464473	-0.2652055
OTX015	BRD2/3/4	-0.1367176	-0.4044675
Bromosporine	BRD	-0.09642211	-0.3362152
Decitabine	DNMT3A	-0.07524976	-0.0535432
Taselisib	PIK3CA	-0.1058979	-0.1954229
Crenolanib	FLT3/KIT/PDGFRB	-0.02430631	-0.07150869
GTP-14564	FLT/KIT/PDGFRB	-0.06245684	-0.20615271

Table 3.9: Change in estimate for β_{FA} between pan-cancer linear regression and MYC-specific linear regression for drugs with significant negative β_{FA} within lung samples. Pan-cancer model was the first linear regression across all tissue types for these drugs. Myc-specific model was linear regression for lung samples, with MYC log2CN > 3 as FA high and max log2CN < 1.8 as FA low.

For the comparison between high-CCND1 CN liver samples and FA- liver samples in PRISM, GTP-14564 was the main drug that appeared to have a significant negative estimate for β_{FA} within the liver tissues. Under the threshold (1.8, 3), the β_{FA} estimate was -0.2137, which was

larger than the estimate of -0.06245684 from the pan-cancer analysis. This aligns with the observation of GTP-14564 having a negative estimate across multiple thresholds for liver samples.

For the threshold (1.8, 3) in CTRP, the drugs neratinib, MGCD-265, and PIK-93 all had negative β_{FA} estimates and p-values below 0.05 for the regression between highly amplified MYC lung samples and FA- lung samples. As observed with MYC in PRISM, the β_{FA} estimates in this model were larger than the estimates for the pan-cancer analysis. Of these three drugs, MGCD-265 and neratinib had negative β_{FA} estimates between the FA term and AUC within lung samples for this threshold. Neratinib also had a negative β_{FA} estimate in breast samples for the regression between highly amplified ERBB2 samples and FA- samples, with a parameter estimate of -6.633 and a p-value of 3e-8. The drugs canertinib, lapatinib, and afatinib also had negative β_{FA} estimates for the regression of highly amplified ERBB2 samples, each with a parameter estimate between -5 and -6 and a p-value less than 1e-4.

For the MYCL-specific analysis, the following drugs had a negative β_{FA} estimate in lung samples with a p-value less than 0.05: axitinib, navitoclax, and lenvatinib. Axitinib, however, did not have a significant negative β_{FA} estimate across all lung samples under the threshold (1.8, 3). MYCN did not appear to have any drugs with a negative association in lungs and a p-value less than 0.05. There appeared to be insufficient samples with low FA for the peripheral nervous system comparison of MYCN across the drugs.

Under (1.8, 3), the drugs navitoclax, TW-37, and PIK-93 demonstrated significant negative β_{FA} estimates within the liver between highly-amplified CCND1 samples and FA- samples. The estimates for the FA term in this analysis were significantly greater across all three drugs, each with an estimate less than -1 and an estimate less than -5 for navitoclax. The

final comparison of highly amplified MET gastric samples demonstrated that MGCD-265 had a significant negative β_{FA} estimate, which aligns with the consistent observation of an association between MGCD-265 and liver samples across multiple thresholds in the second round of linear regressions. Lenvatinib, while having a significant p-value for the MET analysis, showed a positive association.

All high-MYC amplified samples were removed from the PRISM data set, and the first model was rerun to see how the results change without the driving amplifications. Of the drugs seen previously in the threshold (1.8, 3), the following still had a significant association without the MYC samples: GTP-14564, rociletinib, I-BET151, decitabine, and MGCD-265. OTX015 and bromosporine, the BRD inhibitors with a significant association for the MYC samples, no longer appeared in the results. I-BET151 also remained in the thresholds (2.3, 3) and (2, 3) without the highly amplified MYC samples. Under (2.3, 3), the drugs for DNMT3A and FLT3 no longer appeared, however they were still observed under (2, 3). Across all thresholds, targets like BCL2, EGFR, HDAC, FLT3, and PIK3CA still appeared at similar frequencies as previously observed.

Removing highly amplified EGFR samples did not drop rociletinib, which was the EGFR inhibitor identified in PRISM for the threshold (1.8,3). Across other thresholds with higher frequencies of EGFR inhibitors in the first regression analysis, EGFR inhibitors persisted in addition to most other drugs. For instance, under (2.3, 4), the drugs lapatinib, poziotinib, dacomitinib, afatinib, and XL-647 still appeared. Removing highly amplified CCND1 samples dropped semaxinib, a VEGFR inhibitor.

To test the effects of all three driver genes, the model was rerun with only samples that did not have a high amplification for any of the genes. The following drugs were dropped from

the results for the threshold (1.8, 3): rociletinib, OTX015, bromosporine, taselisib, MGCD-265, and semaxanib. Within these results, the following targets still appeared: ADRA1A, BRD2, BRD3. BRD4, FLT3, KIT, PDGFRB, and DNMT3A under one drug each. The following targets, however, no longer appeared for (1.8, 3): EGFR, PDGFRA, MET, PIK3CA, and PIK3CG. I-BET151 appeared in all three thresholds it was initially observed in (2.3, 3), (2, 3), and (1.8, 3), and decitabine still appeared in the thresholds (2, 3) and (1.8, 3). GTP-14564, XL-647, lapatinib, tucatinib, resminostat, dacomitinib, venetoclax, and afatinib still appeared in at least one threshold. GTP-14564 was not observed for threshold (2.3, 4) to (1.5, 4), which is when the EGFR and BCL2 inhibitors began to re-appear. Another target that appeared consistently in the first eight thresholds was CHEK1 with the drug CHIR-124.

The process was repeated with the following driver genes for CTRP: CCND1, ERBB2, MET, MYC, MYCN, and MYCL. When highly amplified CCND1 samples were removed, the following drugs no longer had a significant negative estimate for (1.8, 3): axitinib, navitoclax, TW-37, and lenvatinib. Dropping ERBB2 led to a loss of significant associations for all four EGFR and ERBB2 inhibitors: afatinib, neratinib, canertinib, and lapatinib. Only MGCD-265 dropped when highly-amplified MET samples were removed, and no drugs of interest were dropped with the removal of highly-amplified MYCN samples. Dropping MYC led to a loss of significant associations for axitinib, TW-37, and lenvatinib as observed with CCND1, and dropping MYCL also affected TW-37 and lenvatinib.

After dropping all driver genes for CTRP, the following drugs still appeared to have a significant negative estimate under (1.8, 3): canertinib, afatinib, neratinib, and PIK-93. One new drug that appeared was BRD-K24690302, which is an HDAC1 inhibitor. The number of targets decreased significantly between the initial model and the model without drivers for each

threshold, however the targets of interest did appear across various thresholds. PIK-93 and canertinib appeared across a few thresholds. MK-2206, the AKT1 inhibitor, still appeared across (2.3, 3), (2.3, 3.5), (2.3, 4), (2, 4), and (1.8, 4). AT7867 also appeared in a few thresholds with MK-2206. Axitinib appeared under (1.8, 3.5), marking the only FLT3 drug that remained across all thresholds.

CHAPTER 4

RNA-Seq Results

4.1 DE Analysis and Gene Ontology Results

RNA-seq of melanoma cell lines was used to identify changes occurring at the molecular level in the presence of focal amplifications. When comparing cells with DMs to cells without FA, there were a total of 561 genes upregulated with a log2FC greater than 1. Gene ontology analysis of these results showed two-fold enrichment of multiple biological processes: action potential, extracellular matrix organization, axonogenesis, regulation of transmembrane transport, regulation of plasma membrane bounded cell projection organization, and regulation of ion transport. Other GO terms that were enriched included cell adhesion, positive regulation of cell communication, positive regulation of signaling, negative regulation of cellular process, RNA processing, ncRNA metabolic process, immune effector process, and adaptive immune response. The last four GO terms had the least enrichment, under 1-fold. Roughly 75 genes made up the enrichment of the positive regulation of signaling and cell communication GO terms, including several transmembrane signal receptor proteins such MET, RET and NTRK2. G-protein coupled receptors like ADRA2A and HTR2B, which appeared in PRISM for the drug sensitivity analysis, also contributed to the term.

For the same gene expression comparison, there were 563 genes downregulated with a log2FC less than -1. Many processes were enriched in the gene ontology analysis, with negative regulation of plasminogen activation having the highest enrichment of 17.93. Other top GO terms were negative regulation of chemokine-mediated signaling pathway, negative regulation of oligodendrocyte differentiation, positive regulation of keratinocyte differentiation, positive

regulation of TGF-B production, and substrate-dependent cell migration. Some GO terms had less enrichment, 5-fold or less, but appeared to be pertinent: cellular response to VEGF stimulus, negative regulation of the ERK1 and ERK2 cascade, positive regulation of MAPK activity, and regulation of epithelial to mesenchymal transition. There were also numerous terms about negative regulation of cell growth and proliferation, and one term was positive regulation of apoptotic processes. FLT1, FLT4, EGF, and KDR were all within this DE list, which contributed to numerous GO terms such as regulation of endothelial cell proliferation.

In comparison, cells with HSRs had 301 genes upregulated with a log2FC greater than 1 compared to parental cells. Gene ontology had fewer processes enriched for these differentially expressed genes: serine family amino acid biosynthesis, nephron epithelium development, and generation of neurons. Serine family amino acid biosynthesis had the highest enrichment with a 16.88-fold enrichment.

As for downregulated genes in cells with HSRs, there were 636 that had a log2FC less than -1 compared to parental cells. Similar to cells with DMs, the downregulated genes produced an expansive list of GO terms. Many of the top hits were related to axons such as synapse assembly, axon regeneration, and axon extension. Terms that were observed in the DM down regulation enrichment and in this enrichment analysis were: negative regulation of oligodendrocyte differentiation, cellular response to VEGF stimulus, negative regulation of ERK1 and ERK2 cascade, positive regulation of MAPK activity, and regulation of endothelial cell proliferation along with other terms. New GO terms for these downregulated genes include cellular response to UV-A, negative regulation of fibroblast growth factor receptor signaling pathway, and mesoderm formation. Some of the key genes that contribute to the term

endothelial cell proliferation are FLT4, EGF, KDR, FLT1, and STAT1. These genes, in addition to PDGFRA, also contribute to VEGF response.

Comparing the gene expression patterns of cells with DMs versus the cells with HSRs, there were 149 genes with a log2FC greater than 1 and 12 genes with a log2FC less than -1. The up-regulated genes had three-fold or more enrichment of the GO terms: cGMP-mediated signaling, positive regulation of synapse assembly, cranial nerve development, calcium-mediated signaling, synapse organization, axon development, cell adhesion, and tube morphogenesis. Other GO terms with roughly two-fold enrichment were regulation of cellular component movement, negative regulation of signal transduction, positive regulation of signal transduction, and tissue development. The down-regulated genes however did not have any GO enrichment.

4.2 Network Correlation Modules Analysis

We then performed network correlation (WGCNA) analysis with all the differentially expressed genes and obtained 8 distinct modules. To look at the relationship between the modules and karyotypes, we plotted the transformed count data as a heat map between the karyotype and module; red indicates low log2CPM and yellow indicates high log2CPM (Figure 4.1). Visually inspecting the heatmap, there is some similarity in expression across a karyotype for specific modules. For instance, there is a substantial number of genes in the green-yellow module and dark olive green module that have high log2CPM values among the DM-&HSR- samples. Conversely, many of the genes in the orange module display low log2CPM values for those samples (Fig. 4.1).

To further look at the relationship between the karyotypes and the modules, we performed three separate correlations between the module eigengenes and the karyotypic assignments for the samples. We found many strong and significant correlations between the karyotypes and modules, confirming some of the patterns seen in the heatmap. Particularly, DM+&HSR- samples show positive correlations with several modules and DM-&HSR+ samples show negative correlations with those same modules: midnight blue, sienna, white, and yellow green (Figure 4.2). These four modules show opposite correlations between DM+&HSR- and DM-&HSR+ samples, but the strength and significance of the correlation differs across the modules. In particular, the sienna module does not have a strong correlation with DM-&HSR+ samples, and the white module does not have a strong correlation with DM+&HSR- samples, but the correlation is strong for the other karyotype.

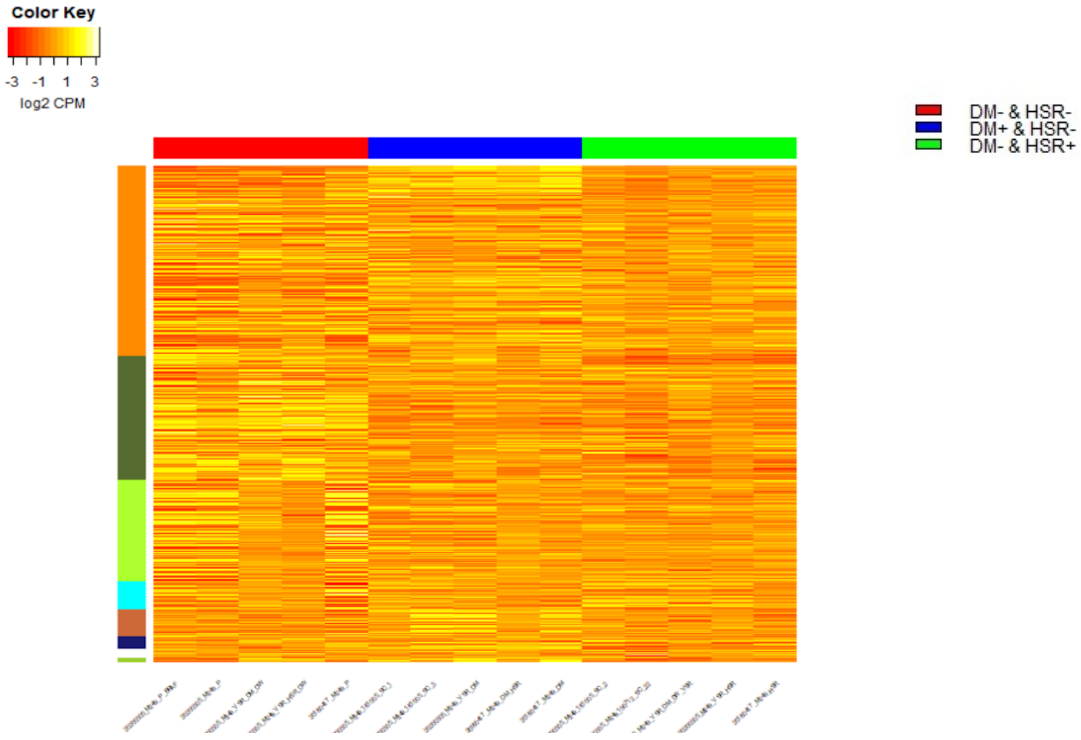


Figure 4.1: Heatmap of log2CPM transformed RNASeq data, with genes sorted by WCGNA modules, color indicated on left axis, and samples sorted by karyotype, indicated on top axis.

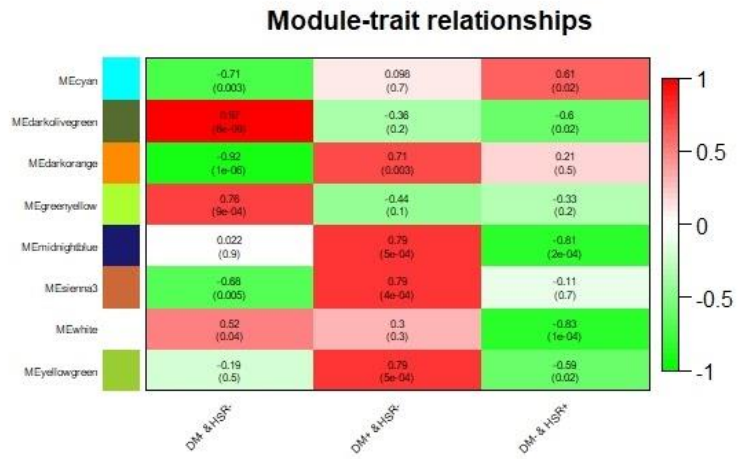


Figure 4.2: Each square represents the correlation between binarized karyotype assignment for the column's karyotype and the module eigengene for the given row's module, with p-value in parentheses.

To understand the processes represented by the modules, we performed gene ontology enrichment on them. The largest module created is the orange module, which had 1456 genes. Gene ontology analysis resulted in numerous enriched biological processes: regulation of platelet aggregation, regulation of fatty acid oxidation, positive regulation of gliogenesis, regeneration, myeloid leukocyte activation, and regulation of angiogenesis. Those with under 2-fold enrichment include positive regulation of GTPase activity, response to peptide, transmembrane receptor protein tyrosine kinase signaling, and regulation of DNA-binding transcription factor activity. Looking at Figure 4.2, these processes are positively correlated with DM+&HSR- samples and negatively correlated with DM-&HSR- samples. It also coincides with enriched processes seen in the gene ontology analysis for upregulated genes in samples with DMs such as cell adhesion and regulation of plasma membrane bounded cell projection organization.

The second largest module from WGCNA was the dark olive-green module with 954 genes, and it appears to capture differences between DM- & HSR- samples and DM- & HSR+

samples (Figure 4.2). Gene ontology analysis of the module showed enrichment of peripheral nervous system axon regeneration, leucine transport, positive regulation of receptor clustering, and negative regulation of pathway-restricted SMAD protein phosphorylation. Other notable GO terms include intrinsic apoptotic signaling pathway in response to DNA damage, negative regulation of protein processing, positive regulation of epithelial to mesenchymal transition, TGF-Beta receptor signaling pathway, and positive regulation of endothelial cell migration. EGF, MET, and HDAC9 were some of the genes that contributed to the last GO term. DDIT4, one of the genes associated with FAs, was mapped for the intrinsic apoptotic signaling pathway term.

The third largest module from WGCNA was the green yellow module, which contained 776 genes. Gene Ontology only identified two processes with at least one-fold enrichment: columnar/cuboidal epithelial cell development and negative regulation of cellular process. Along with the limited GO results, the module only showed significant correlation with the samples lacking focal amplifications (Figure 4.2). The remaining five modules, while having strong correlations with the karyotypic assignments, also lacked significant Gene Ontology enrichment.

4.3. Comparing Results From Drug Sensitivity Analysis

Of the genes associated with FAs and the targets identified in the drug sensitivity analysis, the following were downregulated in DM+&HSR- samples with a log2FC below -0.5 compared to parental cells: FLT1, FLT4, HDAC2, HDAC8, IFI16, KDM1B, KDR, MYCL, PDE4B, RCOR2, STAT1, and YWHAE. Those that were upregulated in DM+&HSR- samples with a log2FC greater than 0.5 were ADRA2A, BRAF, BCL2, BCL2L11, CTDSP1, CTDSP2, DDIT4,

DNMT3A, ERBB2, HDAC5, HDAC10, HDAC11, HTR2B, MET, and MYC. The following genes were downregulated in DM-&HSR+ samples with a log2FC of less than -0.5 compared to parental cells: FLT1, FLT4, HDAC2, HDAC9, KDR, PDE4B, PDGFRA, PNKP, RCOR2, RCOR3, and STAT1. Those that were upregulated with a log2FC greater than 0.5 include BRAF, BRCA2, DDIT4, MET, MYC, and STAT4. For the DM+&HSR-vs DM+&HSR- comparison, the genes ADRA2A, BCL2L11, HDAC5, HTR2B, PDGFRA, and STAT2 were upregulated with a log2FC greater than 0.5. No genes from the drug sensitivity analysis were downregulated.

The genes associated with FAs and the identified drug targets were also distributed across the different modules. For the dark olive green module, the genes PDGFRA, MET, FLT1, HDAC9, IFI16, BRAF, DDIT4, and PNKP were identified. Other genes such as FLT4, EGF, and RCOR2 were not immediate genes of interest but were part of the same gene families. The largest module, orange, encompassed the following genes: ADRA2A, HDAC5, STAT2, BCL2, BCL2L11, CTDSP2, ERBB2, HDAC11, KDM1B, PDE4B, STAT1, YWHAE. The green yellow module encompassed the genes CTDSP1, DNMT3A, HDAC2, HDAC8, KDR, and RCOR3.

The remaining four modules did not appear to contain as many identified drug targets or genes associated with FAs. However, looking at protein interaction networks, there appears to be connections between the genes within the modules and genes associated with FAs. There also appears to be connections between the genes in the modules and targets identified in the drug sensitivity analysis. For instance, in the midnight blue module there is an interaction between JMJD6, which was in the module, and BRD4, which then connects to RCOR1 and is linked to a larger protein network. Other genes like RAD54L, SMC5, RMI1, and POLE2

connect to ATR and TERT, which also link back to the denser portions of the network. TXNIP also appears to be connected to three drugs associated with FA: HDAC1, DDIT4, and NEDD4L. Similarly in the sienna module, there is a large cluster on the top containing genes that link to ATR, which then links to other targets from the drug sensitivity analysis. Of the four modules, the white module appears to have the most continuous connections among all the genes observed in the network, however it has fewer genes specific to the module. The yellow green module, which is the smallest of the four, has very few genes with high-confidence interactions or associations with genes of interest from the drug sensitivity analysis. The few that do include GSN, PRKCQ, HRK, SATB1, PDGFA, ADRB2, and CACNA1C.

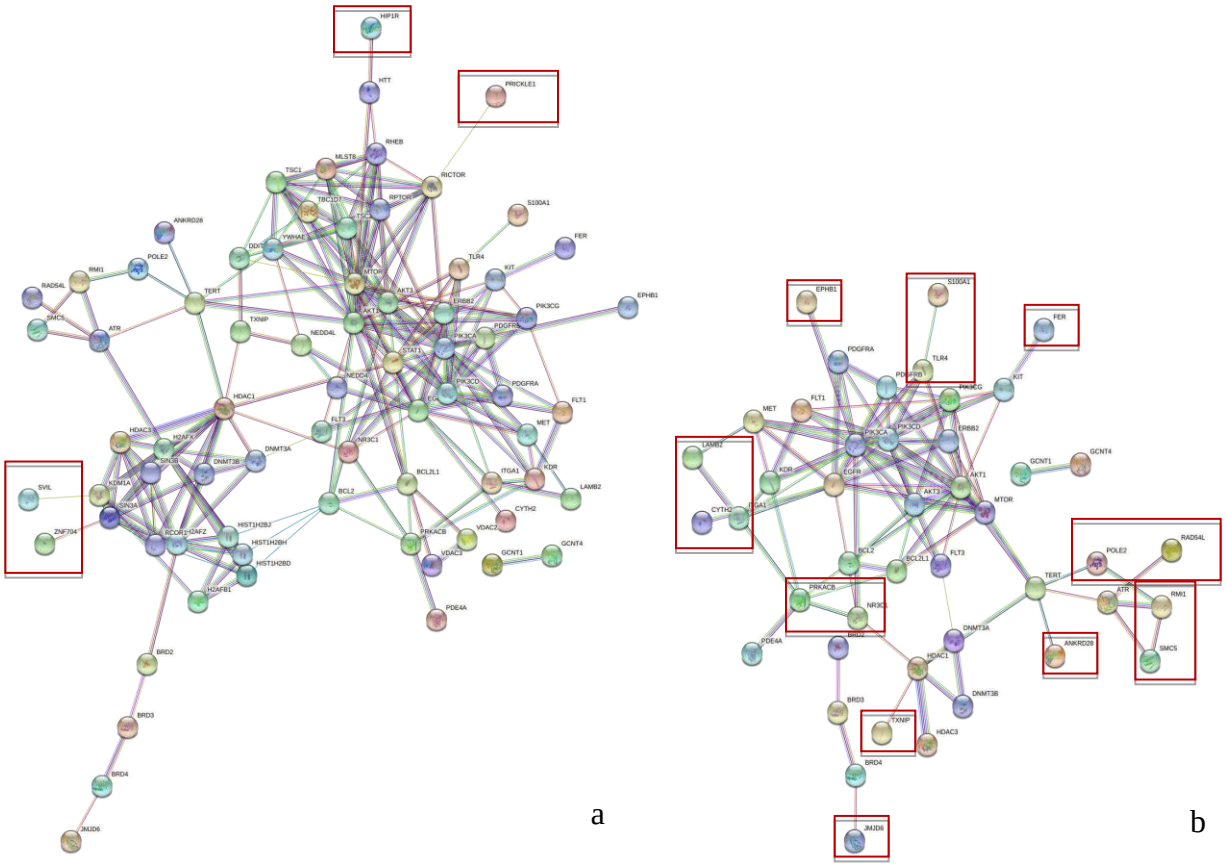


Figure 4.3: STRING-DB Network for genes in ‘Midnight blue’ module, filtered for only connected nodes and confidence of at least 0.700. Red boxes indicate genes within the module

and not from sensitivity analysis. a) Network looking at all genes of interest for FA, b) Network looking at genes of interest from drug sensitivity analysis.

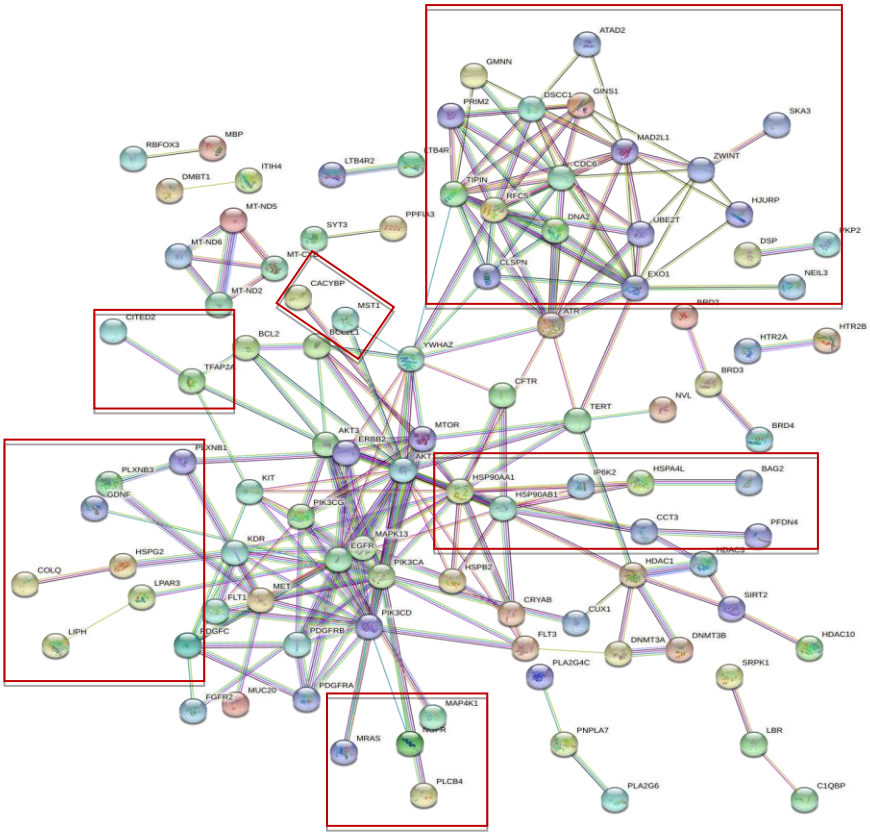


Figure 4.4: STRING-DB Network for genes in ‘Sienna’ module, filtered for only connected nodes and confidence of at least 0.700. Red boxes indicate genes within the module and not from sensitivity analysis, and connected with the main network instead of independent interactions (i.e. not like SRPK1, LBR, C1Q8P). Network looking at genes of interest from drug sensitivity analysis.

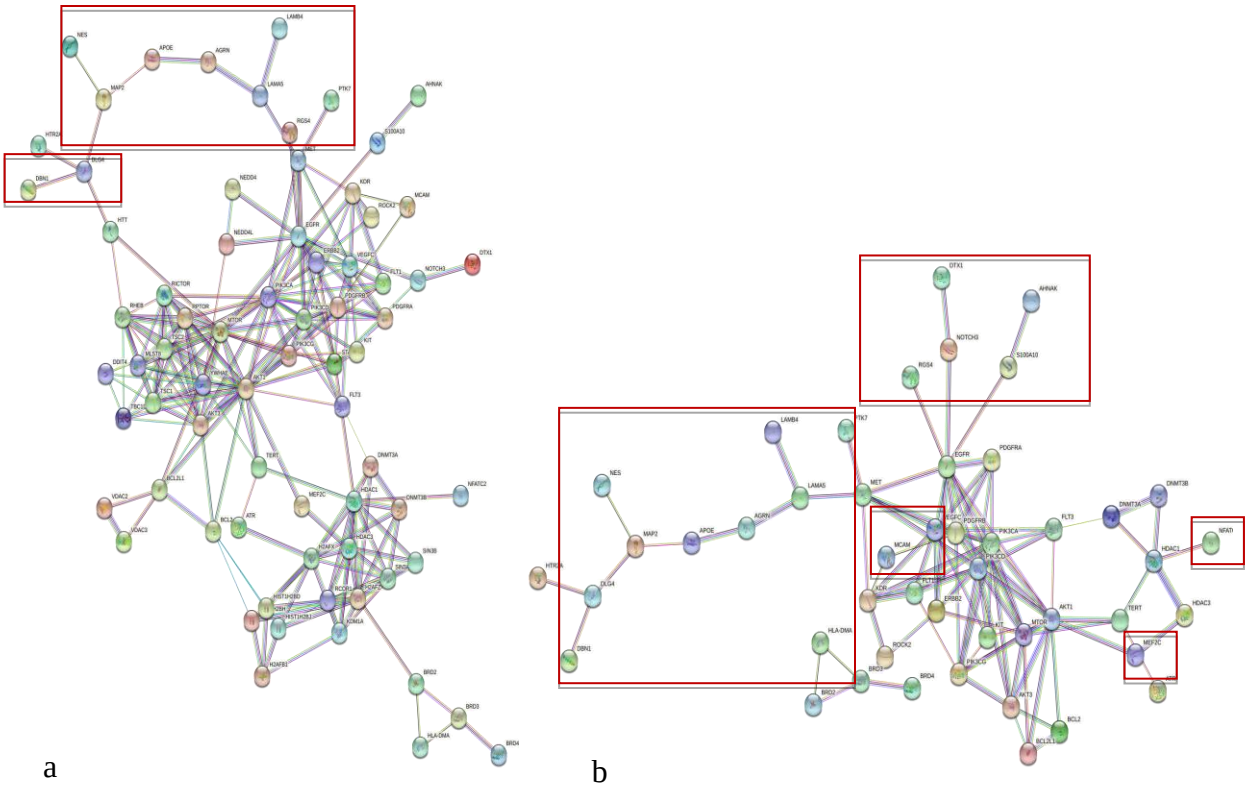


Figure 4.5: STRING-DB Network for genes in ‘White’ module, filtered for only connected nodes and confidence of at least 0.700. Red boxes indicate genes within the module and not in sensitivity analysis. a) Network looking at all genes of interest for FA, b) Network looking at genes of interest from drug sensitivity analysis.

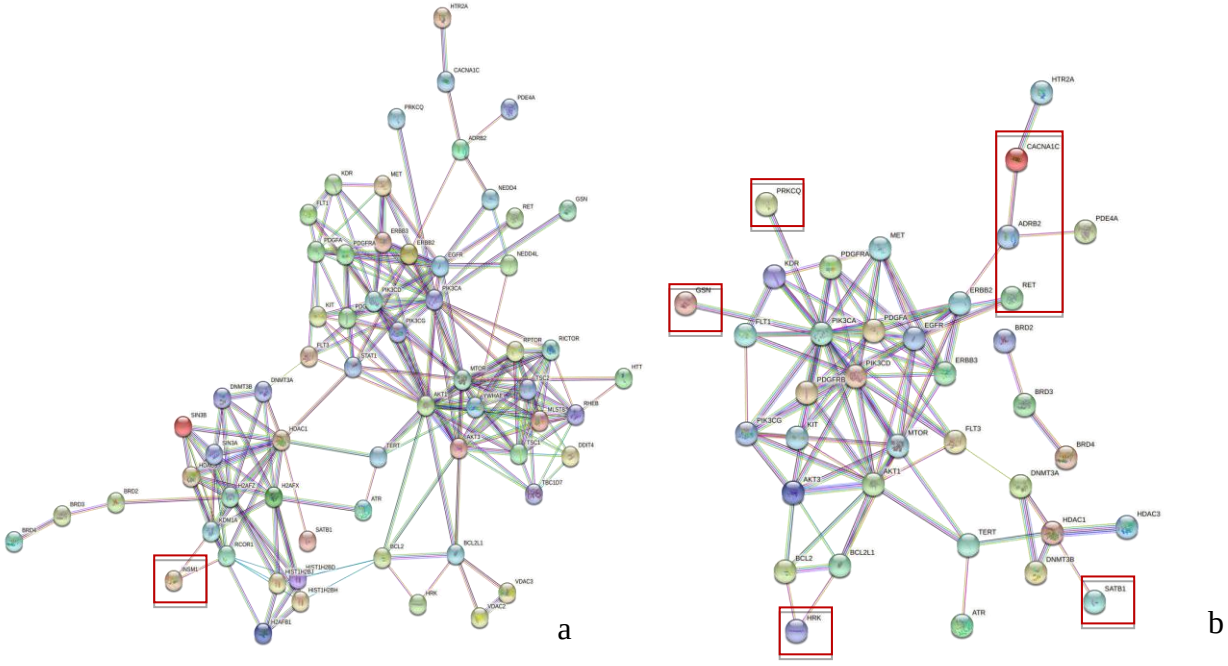


Figure 4.6: STRING-DB Network for genes in ‘Yellowgreen’ module, filtered for only connected nodes and confidence of at least 0.700. Red boxes indicate genes within the module and not in sensitivity analysis. a) Network looking at all genes of interest for FA, b) Network looking at genes of interest from drug sensitivity analysis.

CHAPTER 5

Discussion

The drug sensitivity analysis was able to identify candidate targets for treating cells with focal amplifications, many of which have been shown to be associated with ecDNA and focal amplifications in previous literature. By running the first linear regression across multiple thresholds of FA levels, we were able to capture a wide range of potential targets, including several that are implicated in genomic instability or focal amplifications from previous research: AKT1, ATR, BRD2, BRD3, BRD4, DNMT3A, HDAC1, MTOR, STAT1, PDE4B, and TERT. Other targets that appeared consistently across analyses were BCL2, EGFR, FLT3, PDGFRB, KIT, and PI3K. However, not all targets appeared at the same frequency, and some did not appear to have a skew towards negative β_{FA} estimates. Rather, it appeared that there were more drug-specific patterns of sensitivity in cells with FA as opposed to the target itself contributing to the response. For instance, targets such as EGFR, ERBB2, BCL2, and KDR had a roughly even split between positive and negative β_{FA} across most thresholds in PRISM. Another factor was the presence of driver amplifications that could affect the results of the pan-cancer model. Removing samples with MYC, EGFR, and CCND1 amplifications significantly reduced the number of drugs that met the filtering criteria along with the number of potential targets. Considering all these factors, the following targets remained skewed towards negative estimates and significant even without the presence of highly amplified driver genes: AKT1, PIK3CG, DNMT3A, FLT3 along with KIT and PDGFRB, and most importantly BRD2, BRD3, and BRD4. EGFR and BCL2 still appeared as well.

Past research and literature provide some insight into how cancer cells can become dependent on the targets identified above as well as their interactions with BRD4, which is

known to play a role in ecDNA tethering as mentioned previously. For instance, AKT1 is known to contribute to genomic instability by repressing homologous recombination when overexpressed or constitutively expressed [23]. The study suggests that AKT1 activation results in BRCA1 retention in the cytoplasm, thus preventing double stranded repair and allowing the mutations to accumulate. JQ1, a drug that blocks BRD4 and inhibits c-Myc, has been shown to downregulate the AKT pathway [25]. After treatment of PTEN-positive cells with JQ1, the authors observed an increase in PTEN activity and decrease in the PI3K/AKT function. These cells were found to be more sensitive to JQ1 compared to cells with a PTEN knockdown and PI3K/AKT upregulation. Another association with BRD4 is the inhibition of the FOXO3a-BRD4 complex or CDK6 to reduce AKTi resistance [16].

FLT3 is another gene that is known to impact genomic instability. In patients with acute myeloid leukemia (AML), there are often internal tandem duplications (ITDs) of the FLT3 juxtamembrane domain. A study showed that the presence of FLT3 with ITDs leads to cells using an alternative nonhomologous end joining pathway, which allows double stranded breaks to be repaired but in a more error-prone manner [7] . In doing so, cells continue to survive but the consequences of genomic instability increase over time. FLT3 has also been observed to be harbored on a double minute in a patient with chronic myelomonocytic leukemia [36], and FLT3 amplifications have been observed in metastatic colorectal cancer patients [10]. High FLT3 amplifications were typically not observed with amplification of other driver genes and were believed to act like a RAS mutation by signaling-based promotion of proliferation [10]. Another study showed that the combined use of JQ1 and an anti-FLT3 drug like quizartinib induced apoptosis in AML cells with FLT3-ITD. It is believed that joint inhibition of FLT3 and BRD4 decreased the levels of oncogenes like c-MYC, BCL2, and CDK4 in the cells [8].

While MTOR and TERT were not as robust targets after looking at the skew of β_{FA} estimates and the impact of driver genes, they still have notable impacts on cancer cells and are known to be associated with BRD4. TERT is often a dependency for survival in cancer cells. When TERT is overexpressed, the telomeres in the tumor genome are stabilized and this allows cancer cells to avoid senescence [35]. Furthermore, TERT amplifications have been found on ecDNAs, which can contribute to overexpression [4]. Inhibition of BRD4 has been shown to reduce TERT expression in neuroblastoma cell lines, slowing down the cell growth [12]. This was demonstrated to be caused by BRD4 activating and upregulating TERT expression in cooperation with transcription factors in the E2Fs and WNT pathway.

MTOR, while not typically amplified or found in double minutes, can be crucial for replication and ensuring proper segregation of chromosomes. MTOR regulates DNA duplication, promotes the progression of the DNA replication fork, and stabilizes the fork [11]. If there is a stalled fork and accompanying replication stress, MTOR may be recruited. It has been observed that inhibiting MTOR causes errors during chromosome segregation, thus dysregulation of MTOR in cancer cells can create more instability. In particular, this appears to happen in conjunction with CTDNEP1, which typically counteracts the phosphorylation of lipin-1 by MTOR and allows for lipid homeostasis in the ER [19]. With the stabilization of lipid levels, error correction is facilitated due to the lower viscosity of the cytoplasm. In contrast, if CTDNEP1 is knocked out, MTOR is not inhibited and the viscosity of the cytoplasm increases, preventing error correction and allowing micronuclei to form [19]. A study on renal cell carcinomas showed that BRD4 was a resistance factor against a PI3K/MTOR dual inhibitor and that cotreatment with BRD4 inhibitors increased apoptosis in the cells [34].

In addition to the genes above, EGFR and BCL2 are oncogenes known to be crucial to cell growth and proliferation. In particular, EGFR has been found on double minutes and in focal amplifications, along with amplifications of other oncogenes like MYC, CCND1, and ERBB2 [32]. A computational screening of millions of compounds indicated that a EGFR-BRD4 joint inhibitor may be valuable for treating cancers [1].

The sensitivity analysis as a whole was able to identify multiple targets that have shown to contribute to instability in cancer cells when dysregulated. These targets also appear to be associated with BRD4 in numerous ways, often providing additional benefit when jointly inhibited. Our findings provide potential guidance on approaches for treating cells with ecDNA and focal amplifications, as well as indicating that replication and division stress may be a significant vulnerability in FA cells, similar to implications of previous studies on other forms of genomic instability such as aneuploidy and whole genome doubling. With inhibition of these candidate target genes, there may be less tolerance to ecDNA/DMs where replication and division do not occur properly since copy number imbalances and mutations can be exacerbated to a fatal point.

One downside of the drug sensitivity analysis here is the lack of full consistency in results among the datasets, especially with JQ1 and other drugs targeting BRD2, BRD3, and BRD4. Across the different thresholds, there are very few overlapping results observed, and the frequency of targets varies greatly despite having many common drugs in the different datasets. There is furthermore a difference in the distribution of the β_{FA} estimates for PRISM and CTRP, with CTRP skewing towards more negative estimates. This difference makes it difficult to conclusively interpret whether there is an inclination for cancer samples to be sensitive to a particular gene's inhibition.

The transcriptome analysis did not show a clear signal for DMs and HSRs in terms of differential expression and gene ontology (GO) enrichment results. However, many of the individual genes from the drug sensitivity analysis, and other genes within the same gene family, appeared to be differentially expressed between FA cells and parental cells, with some overlap between the comparisons of DM+&HSR- vs DM-&HSR- and DM-&HSR+ vs DM-&HSR-. For instance, FLT1, FLT4, KDR, and STAT1 were down-regulated in both DM+&HSR- and DM-&HSR+ lines compared to parental cells, and BRAF and DDIT4 were up-regulated. However, there was no clear consistency of presence of differentially genes within the different network correlation (WGCNA) modules, and the GO results for the modules did not produce any terms that appeared relevant to FAs. In particular, the last four modules did not have any notable results from gene ontology enrichment analysis despite appearing to capture the expression differences between DM+&HSR- and DM-&HSR+ lines. However, a look into the potential protein networks for the modules does lend some evidence that targets play a role in DM and HSR biology. Across the last four modules, there are interactions with fairly high confidence between targets identified in the sensitivity analysis and the genes listed in the modules that correspond to differences between DMs and HSRs. This bolsters the support that the targets are needed to efficiently maintain cell states and proliferation in the context of genomic instability.

In conclusion, this drug sensitivity analysis project provides valuable insight into the vulnerabilities of cancer cells with focal amplifications. The application of linear regression to data from multiple drugs screens enabled the identification of drugs associated with sensitivity by quantifying the difference in drug response between FA+ and FA- samples as well as the statistical significance of such differences. The targets for these drugs encompassed numerous

genes known to be associated with FAs from previous studies and provided new candidate markers of vulnerability and sensitivity in FA+ cells. Further analysis of tissue-specific and gene-specific drug responses allowed us to narrow down the candidate markers of sensitivity to a robust set of targets, many of which have been linked to genomic instability in previous research and that tend to contribute to replication and division stress when inhibited. A subset of these targets has also been demonstrated to have effective joint inhibition with BRD4, thus strengthening their candidacy as targetable vulnerabilities in cells with DMs. Network analysis of co-expressed gene modules in cells with FAs provided additional insight into gene interactions involving the identified drug targets that could maintain the cell state in the presence of FAs and the resulting genomic instability. The results from this project lay the groundwork for additional exploration into the targetable vulnerabilities of FA+ cells. This is an important area of research given that FAs, and the resulting instability, are unique to cancer cells and can have a significant influence on drug response, as demonstrated in this project. One benefit of identifying vulnerabilities of FA+ cells is that the treatment does not have to be tailored specifically to the oncogenes that are amplified. Rather, the treatment can depend on the inherent state of genomic instability in the cell. This can potentially help minimize the risk of developing drug resistance via inducing replication stress or conditions that trigger apoptosis due to the presence of FAs, as opposed to directly targeting oncogenes and creating selection pressures. Furthermore, studying the interactions between candidate drug targets and other genes can uncover additional gene networks that function to maintain FAs, particularly DMs, and provide more insight into mechanisms that regulate genomic instability in cancer cells. Overall, our understanding of vulnerabilities in FA+ cells has expanded through this project.

Our results combined with further analysis can better inform therapeutic approaches to exploiting the instability and vulnerabilities within these cells.

APPENDIX

Drug	PRISM Target Annotation	CTRP Target Annotation
Purmorphamine	-	SMO
MI-1	-	MEN1
CI-976	-	ACAT1
ML203	-	PKM
PIK-93	-	PIK3CG
CAL-101	-	PIK3CD
Savolitinib*	-	-
Amuvatinib*	-	-
BRD-K80183349	-	HDAC1, HDAC2
Bromosporine	Bromodomain inhibitor	-
Repligen 136	-	HDAC3
Pilaralisib*	-	-
Piperlongumine:MST-312	-	TERT
MK-2206	AKT1, AKT2, AKT3	AKT1
Canertinib	AKT1, EGFR, ERBB2, ERBB4	EGFR, ERBB2
A-674563	AKT1, PKIA, PRKACA	-
AT7867	AKT2, GSK3B, PKIA, PRKACA	AKT1, AKT2, AKT3, RPS6KB2
VE-822	ATM, ATR, MTOR, PIK3CG	-
VE-821	ATR	-
ENMD-2076	AURKA, FLT3, KDR, PDGFRA, PTK2, SRC	-
MGCD-265	AXL, MET	FLT1, FLT3, KDR, MET
Venetoclax	BCL2	-
Gambogic-acid	BCL2	-
Navitoclax	BCL2, BCL2L1, BCL2L2	BCL2, BCL2L1, BCL2L2
TW-37	BCL2, BCL2L1, MCL1	BCL2, BCL2L1
I-BET151	BRD2, BRD3, BRD4	BRD2, BRD3, BRD4
OTX015	BRD2, BRD3, BRD4	-
JQ1	BRD4, BRDT	BRDT
TG-02	CDK1, CDK2, CDK7, CDK9, FLT3, JAK2	-
GW-405833	CNR2	CNR2
Pazopanib	CSF1R, FGF1, FGFR1, FGFR3, FLT1, FLT4, ITK, KDR, KIT, PDGFRA, PDGFRB, SH2B3	FLT1, FLT3, KDR, KIT, PDGFRB
Axitinib	CSF1R, FLT1, FLT4, KDR, PLK4	FLT1, FLT3, KDR, KIT,

		PDGFRA, PGDFRB
Crenolanib	CSF1R, FLT3, KIT, PDGFRA, PDGFRB	-
GTP-14564	CSF1R, FLT3, KIT, PDGFRB	-
Azacitidine	DNMT1, DNMT3A	DNMT1
Decitabine	DNMT1, DNMT3A	DNMT1
SGI-1027	DNMT1, DNMT3A, DNMT3B	-
Osimertinib	EGFR	-
Rociletinib	EGFR	-
Gefitinib	EGFR	AKT1, EGFR
Pelitinib	EGFR	-
WZ8040	EGFR	EGFR
XL-647	EGFR, EPHB4, ERBB2, FLT4, KDR	-
Lapatinib	EGFR, ERBB2	EGFR, ERBB2
Afatinib	EGFR, ERBB2, ERBB4	EGFR, ERBB2
Dacomitinib	EGFR, ERBB2, ERBB4	-
Poziotinib	EGFR, ERBB2, ERBB4	-
Neratinib	EGFR, ERBB2, KDR	EGFR, ERBB2
CP-724714	ERBB2	-
Tucatinib	ERBB2	-
Sirolimus	FGF2, FKBP1A, MTOR	MTOR
Semaxanib	FGFR1, FLT1, KDR, KIT, PDGFRA, PDGFRB, RET	-
Pacritinib	FLT3, JAK1, JAK2, JAK3	-
Lenvatinib	FLT4, KDR	FLT1, FLT3, KDR, KIT, PDGFRA, PDGFRB
Tacedinaline	HDAC1	HDAC1, HDAC2, HDAC3, HDAC6, HDAC8
Belinostat	HDAC1 – HDAC11	HDAC1, HDAC2, HDAC3, HDAC6, HDAC8
Panobinostat	HDAC1, HDAC2, HDAC3, HDAC4, HDAC6, HDAC7, HDAC8, HDAC9	-
Vorinostat	HDAC1, HDAC2, HDAC3, HDAC5, HDAC6, HDAC8, HDAC9, HDAC10, HDAC11	HDAC1, HDAC2, HDAC3, HDAC6, HDAC8
ACY-1215	HDAC1, HDAC2, HDAC3, HDAC6, HDAC8	-
Entinostat	HDAC1, HDAC2, HDAC3, HDAC9	HDAC1, HDAC2, HDAC3, HDAC6, HDAC8

Resminostat	HDAC1, HDAC3, HDAC6, HDAC8	-
Trichostatin-a	HDAC1-HDAC10	-
Temsirolimus	MTOR	MTOR
GSK256066	PDE4A	-
Taselisib	PIK3CA	-
ZSTK-474	PIK3CB, PIK3CD, PIK3CG	PIK3CB, PIK3CD, PIK3CG
BIBR-1532	TERT	TERT
Drugs with a star * next to the name appeared only in GDSC and lacked annotation. From references online, the following targets were determined: Savolitinib: MET [19] Amuvatinib: c-Kit, c-MET, FLT3, PDGFRA [21] Pilaralisib: PI3K [22]		

Supplementary Table 1: Drugs observed in drug sensitivity analysis with a negative β_{FA} and p-value or FDR less than 0.05. First column shows target information included in PRISM drug screens. Second column shows target information included in CTRP drug screens. GDSC did not provide any target information

REFERENCES

[1] Allen, B., Mehta, S., Ember, S. *et al.* Large-Scale Computational Screening Identifies First in Class Multitarget Inhibitor of EGFR Kinase and BRD4. *Sci Rep* **5,** 16924 (2015). https://doi.org/10.1038/srep16924

[2] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* **25**(1), 25–29 (2000). https://doi.org/10.1038/75556

[3] Bielski, C. M., Zehir, A., Penson, A.V., Donoghue, M. *et al.* Genome doubling shapes the evolution and prognosis of advanced cancers. *Nature genetics* **50**(8), 1189–1195 (2018). https://doi.org/10.1038/s41588-018-0165-1

[4] Chapman, O.S., Wang, S., Luebeck, J., Garancher, A. *et al.* Abstract 95: The landscape of extrachromosomal circular DNA in medulloblastoma subgroups. *Cancer Res* **81,** 95 (2021). https://doi.org/10.1158/1538-7445.AM2021-95

[5] Cohen-Sharir, Y., McFarland, J.M., Abdusamad, M. *et al.* Aneuploidy renders cancer cells vulnerable to mitotic checkpoint inhibition. *Nature* **590,** 486–491 (2021). https://doi.org/10.1038/s41586-020-03114-6

[6] Corsello, S.M., Nagari, R.T., Spangler, R.D. *et al.* Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. *Nat Cancer* **1,** 235–248 (2020). https://doi.org/10.1038/s43018-019-0018-6

[7] Fan, J., Li, L., Small, D., & Rassool, F. Cells expressing FLT3/ITD mutations exhibit elevated repair errors generated through alternative NHEJ pathways:

implications for genomic instability and therapy. *Blood* **116**(24), 5298–5305 (2010). https://doi.org/10.1182/blood-2010-03-272591

[8] Fiskus, W., Sharma, S., Qi, J., Shah, B., Devaraj, S. G., Leveque, C., Portier, B. P., Iyer, S., Bradner, J. E., & Bhalla, K. N. BET protein antagonist JQ1 is synergistically lethal with FLT3 tyrosine kinase inhibitor (TKI) and overcomes resistance to FLT3-TKI in AML cells expressing FLT-ITD. *Molecular cancer therapeutics* **13**(10), 2315–2327 (2014). https://doi.org/10.1158/1535-7163.MCT-14-0258

[9] Ghandi, M., Huang, F.W., Jané-Valbuena, J. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569,** 503–508 (2019). https://doi.org/10.1038/s41586-019-1186-3

[10] Hasegawa, H., Taniguchi, H., Nakamura, Y. *et al*. FMS-like tyrosine kinase 3 (*FLT3*) amplification in patients with metastatic colorectal cancer. *Cancer Sci*. **112**, 314– 322 (2021). https://doi.org/10.1111/cas.14693

[11] He, Z., Houghton, P.J., Williams, T.M., & Shen, C. Regulation of DNA duplication by the mTOR signaling pathway. *Cell Cycle* **20**(8), 742-751 (2021). https://doi.org/10.1080/15384101.2021.1897271

[12] Huang, M., Zeki, J., Sumarsono, N., Coles, G. L., Taylor, J. S., Danzer, E., Bruzoni, M., Hazard, F. K., Lacayo, N. J., Sakamoto, K. M., Dunn, J., Spunt, S. L., & Chiu, B. Epigenetic Targeting of *TERT*-Associated Gene Expression Signature in Human Neuroblastoma with *TERT* Overexpression. *Cancer research* **80**(5), 1024–1035 (2020). https://doi.org/10.1158/0008-5472.CAN-19-2560

[13] Hung, K.L., Yost, K.E., Xie, L. *et al.* ecDNA hubs drive cooperative intermolecular oncogene expression. *Nature* **600,** 731–736 (2021). https://doi.org/10.1038/s41586-021-04116-8

[14] Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R. *et al*. A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* **166**(3), 740–754 (2016). https://doi.org/10.1016/j.cell.2016.06.017

[15] Langfelder, P., Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008). https://doi.org/10.1186/1471-2105-9-559

[16] Liu, J., Duan, Z., Guo, W. *et al.* Targeting the BRD4/FOXO3a/CDK6 axis sensitizes AKT inhibition in luminal breast cancer. *Nat Commun* **9,** 5200 (2018). https://doi.org/10.1038/s41467-018-07258-y

[17] Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15,** 550 (2014). https://doi.org/10.1186/s13059-014-0550-8

[18] Markham A. Savolitinib: First Approval. *Drugs* **81**(14), 1665–1670 (2021). https://doi.org/10.1007/s40265-021-01584-0

[19] Merta, H., Carrasquillo Rodríguez, J.W., Anjur-Dietrich, M.I., Granade, M.E. *et al.* A CTDNEP1-lipin 1-mTOR regulatory network restricts ER membrane biogenesis to enable chromosome motions necessary for mitotic fidelity. *Developmental Cell* **56**(24), P3364-3379.E10 (2021). https://doi.org/10.1101/2021.03.02.433553

[20] National Center for Biotechnology Information. PubChem Compound Summary for CID 11282283, Amuvatinib (2022). Retrieved from https://pubchem.ncbi.nlm.nih.gov/compound/Amuvatinib.

[21] National Center for Biotechnology Information (2022). PubChem Compound Summary for CID 56599306, Pilaralisib (2022). Retrieved from https://pubchem.ncbi.nlm.nih.gov/compound/Pilaralisib.

[22] Picco, G., Chen, E.D., Alonso, L.G. *et al.* Functional linkage of gene fusions to cancer cell fitness assessed by pharmacological and CRISPR-Cas9 screening. *Nat Commun* **10,** 2198 (2019). https://doi.org/10.1038/s41467-019-09940-1

[23] Plo, I., Laulier, C., Gauthier, L., Lebrun, F., Calvo, F., & Lopez, B.S. AKT1 Inhibits Homologous Recombination by Inducing Cytoplasmic Retention of BRCA1 and RAD51. *Cancer Res* **68,** 9404–9412 (2008). https://doi.org/10.1158/0008-5472.CAN-08-0861

[24] Pozdeyev, N., Yoo, M., Mackie, R., Schweppe, R. E., Tan, A. C., & Haugen, B. R. Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies. *Oncotarget* **7**(32), 51619–51625 (2016). https://doi.org/10.18632/oncotarget.10010

[25] Qiu, H., Li, J., Clark, L. H., Jackson, A. L., Zhang, L., Guo, H., Kilgore, J. E., Gehrig, P. A., Zhou, C., & Bae-Jump, V. L. JQ1 suppresses tumor growth via PTEN/PI3K/AKT pathway in endometrial cancer. *Oncotarget* **7**(41), 66809–66821 (2016). https://doi.org/10.18632/oncotarget.11631

[26] Quinton, R.J., DiDomizio, A., Vittoria, M.A. *et al.* Whole-genome doubling confers unique genetic vulnerabilities on tumour cells. *Nature* **590,** 492–497 (2021). https://doi.org/10.1038/s41586-020-03133-3

[27] Rees, M. G., Seashore-Ludlow, B., Cheah, J. H., Adams, D. J. *et al*. Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nature chemical biology* **12**(2), 109–116 (2016). https://doi.org/10.1038/nchembio.1986

[28] Seashore-Ludlow, B., Rees, M. G., Cheah, J. H., Cokol, M. *et al*. Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer discovery* **5**(11), 1210–1223 (2015). https://doi.org/10.1158/2159-8290.CD-15-0235

[29] Stewart, Z. A., Westfall, M. D., & Pietenpol, J. A. Cell-cycle dysregulation and anticancer therapy. *Trends in pharmacological sciences* **24**(3), 139–145 (2003). https://doi.org/10.1016/S0165-6147(03)00026-9

[30] Storlazzi, C. T., Lonoce, A., Guastadisegni, M. C., Trombetta, D. *et. al*. Gene amplification as double minutes or homogeneously staining regions in solid tumors: origin and structure. *Genome research* **20**(9), 1198–1206 (2010). https://doi.org/10.1101/gr.106252.110

[31] Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids research* **49**(D1), D605–D612 (2021). https://doi.org/10.1093/nar/gkaa1074

[32] Turner, K., Deshpande, V., Beyter, D. *et al.* Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543,** 122–125 (2017). https://doi.org/10.1038/nature21356

[33] Wu, S., Turner, K.M., Nguyen, N. *et al.* Circular ecDNA promotes accessible chromatin and high oncogene expression. *Nature* **575,** 699–703 (2019). https://doi.org/10.1038/s41586-019-1763-5

[34] Xu, M., Xu, L., Wang, Y., Dai, G., Xue, B., Liu, Y. Y., Zhu, J., & Zhu, J. BRD4 inhibition sensitizes renal cell carcinoma cells to the PI3K/mTOR dual inhibitor VS-5584. *Aging* **12**(19), 19147–19158 (2020). https://doi.org/10.18632/aging.103723

[35] Yuan, X., Larsson, C. & Xu, D. Mechanisms underlying the activation of *TERT* transcription and telomerase activity in human cancer: old actors and new players. *Oncogene* **38,** 6172–6183 (2019). https://doi.org/10.1038/s41388-019-0872-9

[36] Zhang, H., Wang, X., Li, S., Wang, X. *et al.* FLT3 Amplification as Double Minute Chromosomes in a Patient with Chronic Myelomonocytic Leukemia. *Disease Markers*, Article ID 99232837 (2021). https://doi.org/10.1155/2021/993283