Network potential identifies therapeutic *miRNA* cocktails in Ewing sarcoma

Davis T. Weaver[1,2,*], Kathleen I. Pishas[3,*], Drew Williamson[4,*], Jessica Scarborough[1,2], Stephen L. Lessnick[5], Andrew Dhawan[2,6,†], and Jacob G. Scott[1,2,7,†]

[1]Case Western Reserve University School of Medicine, Cleveland, OH, 44106, USA
[2]Translational Hematology Oncology Research, Cleveland Clinic, Cleveland OH, 44106, USA
[3]Peter MacCallum Cancer Centre, Melbourne, Australia, 3000
[4]Department of Pathology, Brigham & Women's Hospital, Boston, MA, 02144
[5]Nationwide Children's Hospital, Columbus, Ohio, 43205
[6]Division of Neurology, Cleveland Clinic, Cleveland Ohio, 44195
[7]Department of Physics, Case Western Reserve University, Cleveland, OH, 44106, USA
[*]contributed equally
[†]dhawana@ccf.org, scottj10@ccf.org

ABSTRACT

MicroRNA (miRNA)-based therapies are an emerging class of targeted therapeutics with many potential applications. Ewing Sarcoma patients could benefit dramatically from personalized miRNA therapy due to inter-patient heterogeneity and a lack of druggable (to this point) targets. However, because of the broad effects miRNAs may have on different cells and tissues, trials of miRNA therapies have struggled due to severe toxicity and unanticipated immune response. In order to overcome this hurdle, a network science-based approach is well-equipped to evaluate and identify miRNA candidates and combinations of candidates for the repression of key oncogenic targets while avoiding repression of essential housekeeping genes. We first characterized 6 Ewing sarcoma cell lines using mRNA sequencing. We then estimated a measure of tumor state, which we term network potential, based on both the mRNA gene expression and the underlying protein-protein interaction network in the tumor. Next, we ranked mRNA targets based on their contribution to network potential. We then identified miRNAs and combinations of miRNAs that preferentially act to repress mRNA targets with the greatest influence on network potential. Our analysis identified TRIM25, APP, ELAV1, RNF4, and HNRNPL as ideal mRNA targets for Ewing sarcoma therapy. Using predicted miRNA-mRNA target mappings, we identified miR-3613-3p, let-7a-3p, miR-300, miR-424-5p, and let-7b-3p as candidate optimal miRNAs for preferential repression of these targets. Ultimately, our work, as exemplified in the case of Ewing sarcoma, describes a novel pipeline by which personalized miRNA cocktails can be designed to maximally perturb gene networks contributing to cancer progression.

Author Summary

Precision medicine in cancer aims to find the right treatment, for the right patient, at the right time. Substantial variation between patient tumors, even of the same disease site, has limited the application of precision medicine in the clinic. In this study, we present novel computational tools for the identification of targets for cancer therapy using widely available sequencing data. We used a network-science based approach that leveraged multiple types of 'omic data to identify functionally relevant disease targets. Further, we developed algorithms to identify potential miRNA-based therapies that inhibit these predicted disease targets. We applied this pipeline to a novel Ewing Sarcoma transcriptomics data-set as well as publicly available patient data from the St. Jude Cloud. We identified a number of promising therapeutic targets for this rare disease, including EWSR1, the proposed driver of Ewing Sarcoma development. These novel data and methods will provide researchers with new tools for the development of precision medicine treatments in a variety of cancer systems.

Conflict of Interest Statement: The authors have no conflicts of interest to disclose.

Introduction

Ewing sarcoma is a rare malignancy arising from a gene fusion secondary to rearrangements involving the EWS gene[1]. There are 200-300 reported cases each year in the United States, disproportionately affecting children[2]. High levels of inter-tumor heterogeneity are observed among Ewing sarcoma patients despite a shared EWS gene fusion initiating event[3]. Ewing sarcoma

is also extremely prone to developing resistance to available chemotherapeutics [4]. These features make it an ideal system to develop personalized therapies for resistant tumors or to avoid the development of resistance altogether.

MicroRNA (miRNA)-based therapeutics, including anti-sense oligonucleotides, are an emerging class of cancer therapy[5]. Recent work has highlighted the critical importance of miRNAs in the development and maintenance of the cancer phenotype[4–6]. MiRNA dysregulation has been implicated in the development of each of the hallmark features of cancer[7], and restoration of expression of some of these critical downregulated miRNAs has been studied as a potential treatment for several different cancers[6,8,9]. In particular, in the past decade, anti-sense oligonucleotide inhibitors of the STAT3 transcription factor have shown promise in the settings of lymphoma[10,11] and neuroblastoma[12]. MiR-34 has shown to be effective in pre-clinical studies for treatment of both lung cancer[13–15] and prostate cancer[16]. Finally, miR-34 and let-7 combination therapy has been shown to be effective in pre-clinical studies of lung cancer[15].

MiRNAs have been recognized as potential high-value therapeutics in part due to their ability to cause widespread changes in a cell-signaling network[5]. A single miRNA molecule can bind to and repress multiple mRNA transcripts[6,17–19], a property that can be exploited when designing therapy to maximally disrupt a cancer cell signaling network. This promiscuity of miRNA

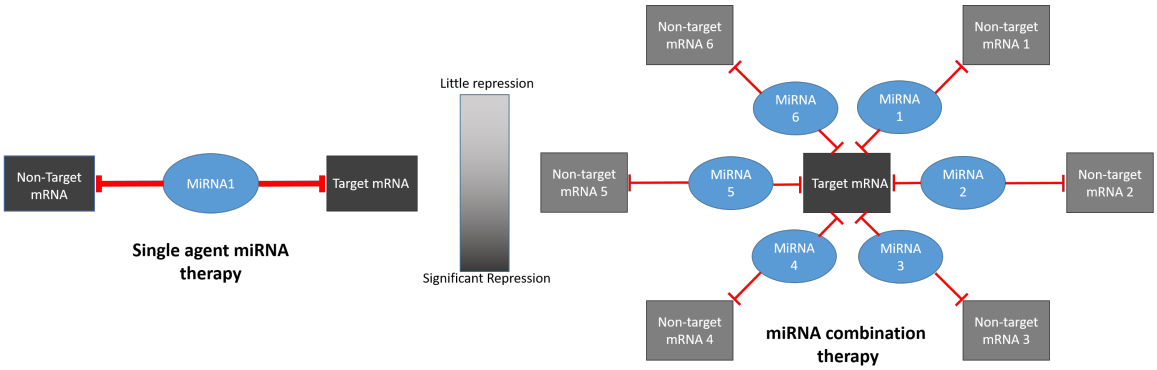

Figure 1. Cartoon describing rationale for focusing on miRNA combination therapy. With single-agent therapy, both target mRNA and non-target mRNA are inhibited an equal amount, potentially resulting in toxicity due to off-target effects. With miRNA combination therapy, the common target mRNA is inhibited to a greater degree than any individual non-target miRNA.

binding may also increase the risk of off-target effects and toxicity (Figure 1). For example, miR-34 was effective in pre-clinical studies for the treatment of a variety of solid tumors[13–16], only to fail in a phase I clinical trial due to "immune-related serious adverse events"[20]. To capitalize on the promise of miRNA-based cancer therapy while limiting potential toxicity, we developed a systematic, network-based approach to evaluate miRNA cocktails. We focused on miRNA cocktails rather than single miRNA therapeutics due to the potential for miRNA cocktails to minimize toxicity compared to single miRNA regimens[21] (**Figure 1**).

In this work, we build on previous studies applying thermodynamic measures to cell signaling networks in the field of cancer biology[22–24], as well as works that describe a method to use gene homology to map miRNAs to the mRNA transcripts they likely repress[6,17,18]. Reitman et al. previously described a metric of cell state analogous to Gibbs free energy that can be calculated using the protein-protein interaction network of human cells and corresponding transcriptomic data[22]. Gibbs free energy has been correlated with a number of cancer-specific outcomes, including cancer grade and patient survival[23]. Additionally, Reitman et al. leveraged Gibbs and other network measures to identify personalized protein targets for therapy in a dataset of low-grade glioma patients from The Cancer Genome Atlas (TCGA)[22]. Previous work has also highlighted the critical importance of miRNAs to maintenance and development of the oncogenic phenotype, and demonstrated the utility of applying miRNA-mRNA mappings. [6] In this work, we developed and applied a computational pipeline that leverages these network principles to identify miRNA cocktails for the treatment of Ewing sarcoma.

1 Methods

1.1 Overview

We characterized six previously described Ewing sarcoma cell lines in triplicate [25] – A673, ES2, EWS502, TC252, TC32, and TC71 – using paired miRNA and mRNA sequencing. By evaluating 6 distinct cell lines, we aimed to assess the heterogeneity inherent to Ewing sarcoma *in-vitro*. We also utilized mRNA sequencing data for 15 ewing sarcoma patient tumor samples made available on the St. Jude Cloud[26]. We then defined a measure of tumor state, which we term network potential (**Equation 1**), based on both mRNA gene expression and the underlying protein-protein interaction (PPI) network. Next, we ranked mRNA

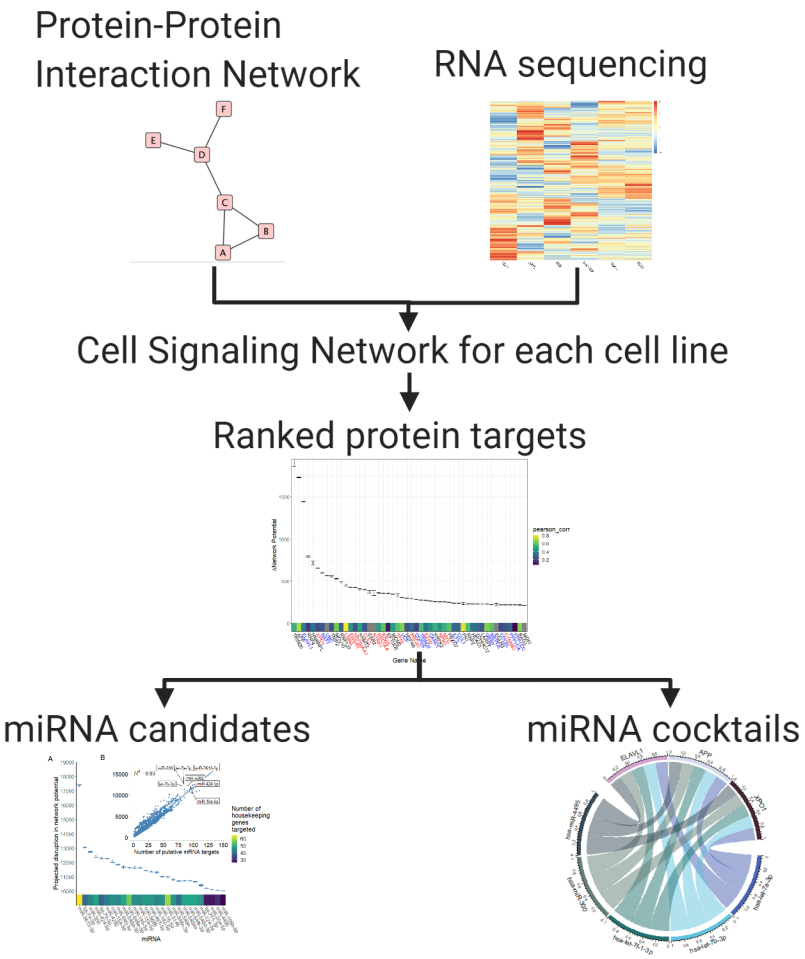

Figure 2. Simplified schematic of our computational pipeline. We defined a measure of tumor state, which we term network potential (Equation 1), based on both mRNA gene expression and the underlying protein-protein interaction (PPI) network. Next, we ranked mRNA targets based on their contribution to network potential of each cell line, aiming to approximate the relative importance of each mRNA to network stability. After identifying these mRNA targets, we then identified miRNA and miRNA cocktails that preferentially acted to repress the most influential of the ranked mRNA targets, with the aim of defining synthetic miRNA-based therapy for down-regulation of these targets.

targets based on their contribution to network potential of each cell line, aiming to approximate the relative importance of each mRNA to network stability. Relative importance of each mRNA to network stability was determined by calculating the change in network potential of each network before and after *in silico* repression of each mRNA (ΔG, described in **Section 1.5**). After identifying these mRNA targets, we then identified miRNA and miRNA cocktails that preferentially acted to repress the most influential of the ranked mRNA targets, with the aim of defining synthetic miRNA-based therapy for down-regulation of these targets. Our computational pipeline is schematized in **Figure 2**.

1.2 Data sources

We utilized three data sources to develop our Ewing sarcoma cell signaling networks: the BioGRID protein-protein interaction database[27], mRNA expression data from 6 Ewing sarcoma cell lines, which are available on GEO (accession GSE98787), and mRNA expression data from 15 Ewing sarcoma patient samples, which are available on the St. Jude Cloud[26].

Protein Protein Interaction Databases The BioGRID interaction database contains curated data detailing known interactions between proteins for a variety of different species, including Homo sapiens. The data were generated by manual curation of the biomedical literature to identify documented interactions between proteins[27]. To assist in manual curation, the BioGRID project uses a natural language processing algorithm that analyzes the scientific literature to identify manuscripts likely to contain information about novel PPIs. The dataset is therefore limited to protein interactions that are reliably reported in the

scientific literature. As new research accumulates, substantial changes to the PPI network may occur. For example, between 2016 and 2018, the number of documented PPIs in *Homo sapiens* grew from 365,547 to 449,842. The 449,842 documented interactions in 2018 were identified through curation of 27,631 publications[27]. Importantly, the PPI network is designed to represent normal human tissue. To assess the importance of the specific PPI used to our results, we repeated much of our analysis using stringdb, another publicly available PPI with millions of documented interactions. To maintain some consistency with biogrid, we modulated the provided "interaction score" until the resulting network had around the same number of edges as the biogrid network. Interaction score is a number ranging from 0 to 1000 that describes the likelihood that a given protein protein interaction is biologically relevant. For our analysis, we included only interactions with an interaction score ≥ 700.

Ewing sarcoma transcriptomics Second, we utilized mRNA expression data from *in vitro* experiments conducted on six Ewing sarcoma cell lines (3 biological replicates per cell line). RNA/miRNA extraction was performed with a Qiagen kit with on-column DNase digestion. These mRNA and miRNA expression data were then normalized to account for between sample differences in data processing and further adjusted using a regularized log (Rlog) transformation[28,29]. In order to extend our study to patient samples, we repeated our analysis on 15 patient tumors from the St. Jude Cloud for which RNA sequencing data was available. The St. Jude Cloud is a comprehensive, cloud-based data-sharing ecosystem that provides genomic data on thousands of samples from patients with pediatric cancer[26].

Notably, methods for calculating network potential from this type of data require protein concentrations rather than mRNA transcript concentrations. For the purposes of this analysis, we assumed that concentration of protein in an Ewing sarcoma tumor was equivalent to the concentration of the relevant mRNA transcript. A large body of work suggests that mRNA levels are the primary driver of protein levels in a cell under steady state conditions (i.e. not undergoing proliferation, response to stress, differentiation etc)[30–33]. However, recent work in a 375 cancer cell lines has shown that mRNA expression may not be predictive of protein expression in the setting of malignancy[34]. For this reason, we included the protein-mRNA correlations from their experiments alongside some of our key findings to provide needed context.

1.3 Network development

We first developed a generic network to represent human cell signaling networks using the BioGRID interaction database[27]. The BioGRID protein-protein interaction network can be downloaded as a non-linear data structure containing ordered pairs of proteins and all the other proteins with which they interact. This data structure can be represented as an undirected graph, with vertex set $\mathscr{V}$, where each vertex represents a protein, and edge set ($\mathscr{E}$) describes the interactions between proteins.

Using mRNA sequencing data from 6 Ewing sarcoma cell lines in triplicate, we then ascribed mRNA transcript concentration for each gene as an attribute to represent the protein concentration for each node in the graph. Through this process, we developed networks specific to each cell line and replicate in our study (18 total samples). We then repeated the same process to develop networks specifics to the 15 patient tumor samples.

1.4 Network potential calculation

Using the cell signaling network with attached cell line and replicate number specific normalized mRNA expression data, we defined a measure of tumor state following Reitman et al.[22], which we term network potential. Our Network potential metric was inspired by Gibbs free energy in physics or chemistry. We first calculate the network potential of the i-th node in the graph:

$$G_i = C_i \ln\left[\frac{C_i}{\sum C_j + C_i}\right]. \tag{1}$$

where G_i is equal to the network potential of an individual node of the graph, C_i is equal to the concentration of protein corresponding to node G_i, and C_j is the concentration of protein of the j-th neighbor of G_i. Because of the natural log transformation, G_i will always return a negative number. Total network potential (G) of the network can then be calculated as the sum over all nodes:

$$G = \sum_i G_i. \tag{2}$$

where G is equal to the total network potential for each biological replicate of a given cell line. We then compared total network potential across cell lines and biological replicates. More negative network potentials were interpreted as being "larger" in the absolute sense.

1.5 Ranking of protein targets

After calculating network potential for each node and the full network, we simulated "repression" of every node in each network by reducing their expression (computationally) to zero, individually[35]. Clinically, this would be akin to the application of a drug that perfectly inhibited the protein/mRNA of interest. Next, we re-calculated network potential for the full network and calculated the change in network potential (ΔG) by subtracting the new network potential value for the network potential value of the "unrepressed" network. Because the network potential of each node was negative, systematic repression of a given node always drove the total network potential to be less negative. As a result, this approach will always return a positive ΔG. We then ranked each node in the network according to ΔG for further analysis.

We also evaluated the top predicted genes by ΔG against a null model of ΔG to evaluate the likelihood that these observed disruptions were due to random chance. To construct our null distribution of ΔG, we repeated the following process 1000 times for each sample under study:

1. We constructed a random graph that preserves the original degree distribution for the underlying protein-protein interaction network by iteratively swapping edges. For each random graph, we performed n*100 swaps where n is the number of nodes in the original graph.

2. We then constructed a cell signaling network using the random graph and the mRNA expression data for that sample. mRNA expression was left unchanged for each random graph.

3. We computed the ΔG for the top 50 proteins under study on the new random graph.

We then calculated the average and standard deviation ΔG from all 1000 iterations of the above process to compute a bootstrapped null distribution of ΔG. We then computed confidence intervals for ΔG, employing the bonferroni correction to account for multiple hypothesis testing.

Our pipeline was designed to make use of parallel computing on the high-performance cluster (HPC) at Case Western Reserve University.

1.6 Identification of miRNA cocktails

To generate miRNA-mRNA mappings, we implemented a protocol described previously[36]. Briefly, we identified all predicted mRNA targets for each miRNA in our dataset using the miRNAtap database in R, version 1.18.0, as implemented through the Bioconductor targetscan org.Hs.eg.db package, version 3.8.2[17]. We used all five possible databases (default settings): DIANA version 5.061[19], Miranda 2010 release62[37], PicTar 2005 release63[38], TargetScan 7.164[39] and miRDB 5.065[18], with a minimum source number of 2, and the union of all targets found was taken as the set of targets for a given miRNA. Through this mapping, we identified a list of mRNA transcripts that are predicted to be repressed by a given miRNA. Our code and processed data files are available on Github at: https://github.com/DavisWeaver/MiR_Combo_Targeting/.

Using this mapping, as well as our ranked list of promising gene candidates for repression from our network analysis, we were able to identify a list of miRNA that we predict would maximally disrupt the Ewing sarcoma cell signaling network when introduced synthetically. To rank miRNA targets, we first identified all the genes on the full target list that a given miRNA was predicted to repress (described in Section 1.5). Next, we summed the predicted ΔG when each of these genes was repressed *in silico* to generate the maximum potential disruption that could be achieved if a given miRNA were introduced synthetically into an Ewing sarcoma tumor. We then ranked miRNA candidates in descending order of the maximum predicted network disruption (Figure 5).

Given the documented cases of systemic toxicities associated with miRNA-based therapies, the miRNA that inhibits the most targets might not necessarily be the best drug target. We therefore sought to identify combinations of miRNAs that individually repressed key drug targets, while avoiding repression of housekeeping genes that may lead to toxicity. We defined housekeeping genes using a previously described gene set[40]. In this study, housekeeping genes were identified by evaluating RNA sequencing data from a large number of normal tissue samples. Genes that are consistently expressed in all or nearly all tissue types were assumed to be so-called housekeeping genes. Our hypothesis is that by giving a cocktail of miRNAs with predicted activity against one or multiple identified drug targets, each individual miRNA could be given at a low dose such that only the mRNA transcripts that are targeted by multiple miRNAs in the cocktail are affected (Figure 1). Also with an eye towards limiting toxicity, we restrained our search to endogenous miRNAs rather than broadening to engineered exogenous miRNA mimics. To that end, we designed a loss function (see equation below) to balance the the effects of repressing the housekeeping gene set I as well as the target gene set J:

$$L(\mu) = \sum_{i,j} A(c)(G_i) - A(c)(G_j) \quad \text{for } i \in I \textit{ and } j \in J, \tag{3}$$

$$A(c) = \begin{cases} 0, & \text{if } c \leq 1 \\ 0.2c, & \text{if } 1 < c < 6 \\ 1, & \text{otherwise} \end{cases} \tag{4}$$

Where $A(c)$ determines the degree of repression as a function of the number of times, c, that a given gene, i, is targeted by a given miRNA cocktail, μ. We recognize that assuming each additional miRNA additively represses 20% of a given gene is somewhat arbitrary. Future work could improve on this miRNA cocktail optimization approach by more formally addressing miRNA repression in different contexts.

We first transformed the projected change in network potential for each gene such that housekeeping genes exerted a positive change in network potential and the top 10 predicted targets exerted a negative change in network potential. We then ranked 3-miRNA combinations according to their projected effect on network potential, where more negative changes in network potential were interpreted as most effective for maximizing on-target effects while minimizing off-target effects. As a further constraint, a gene had to be targeted by 2 or more miRNA in a given cocktail to be considered repressed. Each miRNA was assumed to downregulate a given gene by 20%, such that genes targeted by 2 miRs were assumed to have their expression decreased by 40%, and genes targeted by 3 miRs were assumed to have their expression decreased by 60%. We repeated our analysis, varying between 10% and 50% repression to assess the impact of this assumption on our predicted miRNA cocktails. Rather than evaluate every potential 3-miRNA combination, we limited our analysis to miRNA that target at least 2 of our 10 target genes. These constraints were defined *a priori*. We repeated this analysis to identify cocktails that target larger or smaller groups of mRNA (the top 5 or 15 mRNA targets) in order to assess the stability of the predicted cocktail to changing conditions.

2 Results

2.1 Network Overview

We calculated the network potential, a unitless measure of cell state, for each protein in the cell signaling networks for each of the six Ewing sarcoma cell lines in our experiment. An overview comparing network potential to normalized mRNA expression can be found in **Figure 3**. An additional overview of the total network potential for each cell line and biological replicate compared to total mRNA expression is presented in **Figure S1**. The histograms of network potential and mRNA expression demonstrate markedly different distributions (**Figure 3**), indicating that network potential describes different features of a cell signaling network compared to mRNA expression alone. Notably, network potential and mRNA expression for these cell lines are stable across different biological replicates, as demonstrated by the low interquartile range (**Figure S1C** and **S1D**). There were larger differences in both mean expression and network potential across cell lines (**Figure S1C** and **S1D)** when compared to between-replicate differences. The global average network potential across all samples was -3.4×10^5 with a standard deviation of 1605.

The included patient samples (beginning with SJEWS on **Figure S1**) demonstrate substantially more variation in both mRNA expression and network potential. Much of what we are capturing here can likely be attributed to batch effects, as these patient samples may have been sequenced years apart on machines with dramatically different capabilities.

2.2 Identification of Protein Targets

We identified TRIM25, APP, ELAV1, RNF4, and XPO1 as top 5 targets for therapy for each of the 6 cell lines based on the degree of network disruption induced following *in silico* repression of each gene. Of these 5 genes, only XP01 has been previously implicated in oncogenesis[41], while only ELAVL1 is a known essential housekeeping gene. There was a high degree of concordance between cell lines among the top predicted targets (**Table S1**). Of the top ten predicted targets, all 10 targets are conserved for all 6 cell lines. The top 50 protein targets are presented in **Figure 4**. Many of the top identified genes fell within the 99.99% confidence interval of the computed null distribution (Figure 4), suggesting that these genes are highly connected hub genes that are likely to score high in ΔG regardless of the tumor-specific RNA-sequencing information provided. As a sensitivity analysis, we repeated our protein identification pipeline using stringdb as the PPI network provider rather than biogrid. Using stringdb, we identified broadly similar targets, with the majority of identified genes being either essential housekeeping genes or genes causally implicated in cancer. 13 out of the 50 identified targets were shared between stringdb and biogrid. In contrast to our main results using biogrid, the vast majority of top targets identified with stringdb fell within the 99.99% confidence interval of the computed null distribution (**Figure S3**) In addition, the distribution of network potential across all genes looks extremely similar regardless of which PPI was used (**Figures 3** and **S3**).

Surprisingly, there was very little overlap between the top predicted targets between the cell line data and the patient samples (**Figure 4**). The genes that appeared most frequently among the top projected targets for the 15 patient tumor samples were

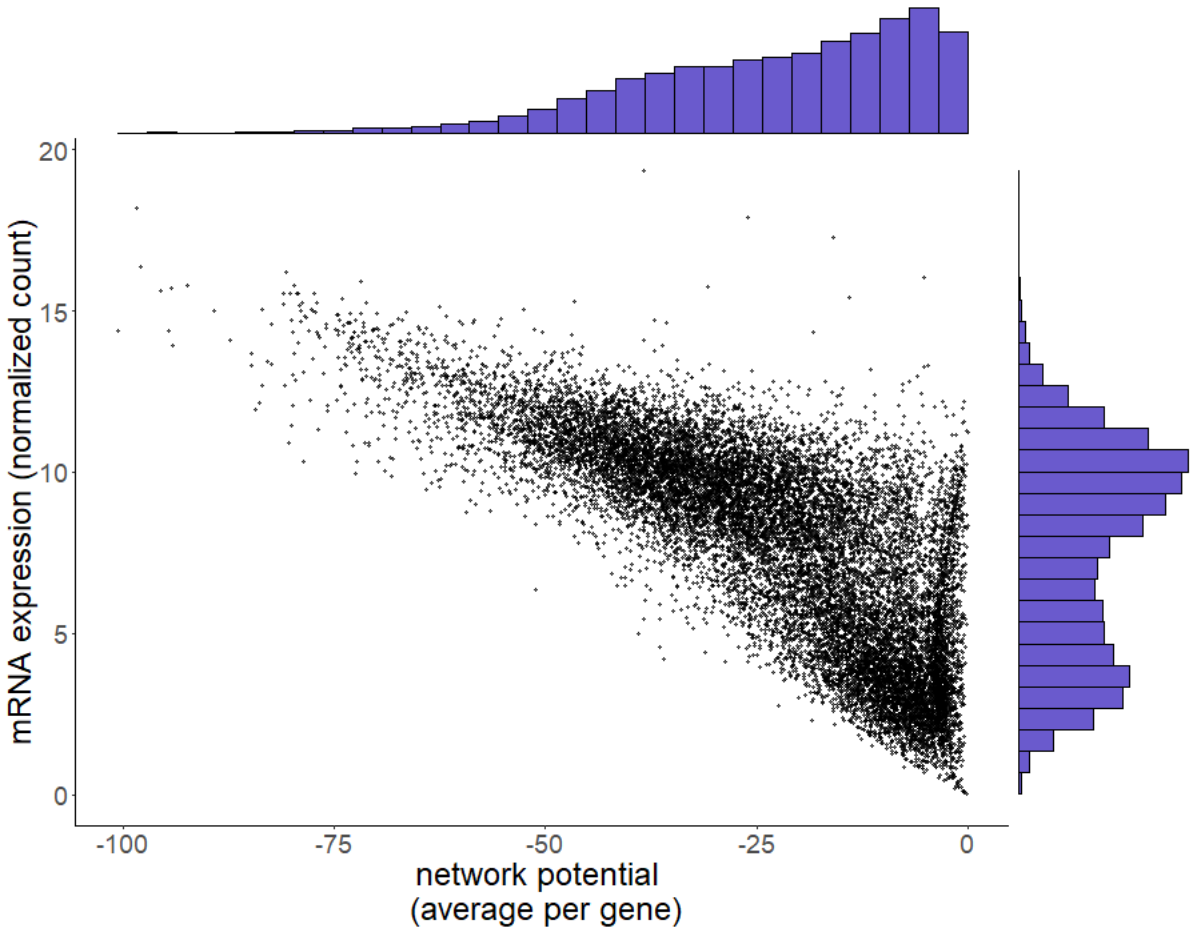

Figure 3. Network potential demonstrates a different distribution compared to mRNA expression. Main panel: scatterplot comparing mRNA expression and network potential for all genes in our 18 Ewing Sarcoma cell lines. For each gene, we averaged across all samples for both mRNA expression and network potential. Unlike network potential (top axis histogram), mRNA expression (right axis histogram) has a bimodal distribution.

SLC24A1, ARTN, DHRSX, TEX261, and FRMD8. Compared to the cell line samples, fewer of the most frequent projected targets were cancer-associated or defined housekeeping genes (**Figure 4**)

Some of these identified genes in the cell line data are likely essential housekeeping genes highly expressed in all or most cells in the body, making them inappropriate drug targets (**Figure 4**). TRIM25, and ELAV1, for example, are involved in protein modification and RNA binding, respectively[42]. We therefore repeated this analysis, limiting our search to gene targets that have been causally implicated in cancer[41]. With this limitation in place, we identified XPO1, LMNA, EWSR1, HSP90AA1, and CUL3 as the top 5 targets for therapy when ΔG was averaged for all cell lines. The top 10 cancer-related targets for each cell line can be found in (**Table S2**). The top 50 protein targets (limited to those causally implicated in cancer) can be found in **Figure S2**.

We also conducted gene set enrichment analysis for the all the genes represented in our cell signaling network (averaged across all samples). We ranked genes by network potential (averaged across all samples) and compared our gene set to the "hallmarks" pathways set, downloaded from the Molecular Signatures Database (MSigDB)[43,44]. This analysis was conducted using the fGSEA package in R, which uses the Benjamini - Hochberg procedure to correct the false discovery rate[45,46]. Our gene set was enriched (adjusted p-value < 0.05) in 24 of the 50 pathways included in the hallmarks set; including apoptosis, DNA repair, mTOR signaling, MYC signaling, and WNT β-catenin signaling. Our gene set was also highly enriched (normalized enrichment score $= 1.73$) in the miRNA bio-genesis pathway. The full results are presented in Table S3.

2.3 Identification of miRNA Cocktails

We identified several miRNAs that were predicted to dramatically disrupt the Ewing sarcoma cell signaling network (**Figure 5**). When averaging all cell lines, we identified miR-3613-3p, let-7a-3p, miR-300, miR-424-5p, and let-7b-3p as the ideal miRs

for preferential repression of proteins predicted to be important for Ewing sarcoma signaling network stability. miR-3613-3p, let-7a-3p, miR-300, miR-424-5p, and let-7b-3p were predicted to cause an average network network potential increase (driving

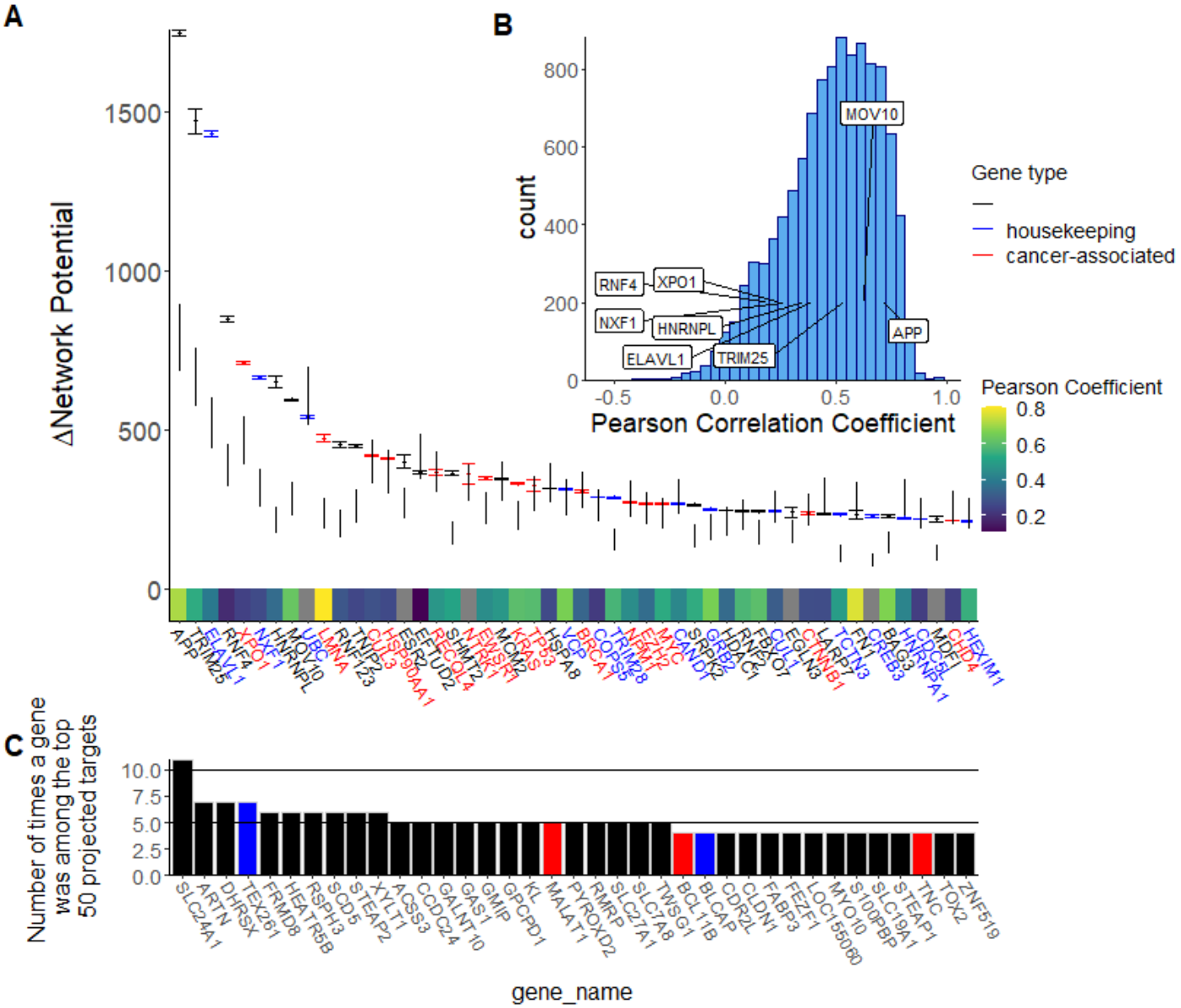

Figure 4. TRIM25, APP, ELAVL1, AND RNF4, and XPO1 are the top protein targets ranked by predicted disruption following *in silico* repression. Panel A: Box and whisker plot describing the change in network potential following *in silico* repression for each of the top 50 proteins. 99.99% confidence interval from the permutation test are displayed alongside the box and whisker plots. It is notable that EWSR1, the kinase associated with Ewing sarcoma development, is considered highly influential in the cell signaling network by this method, even in comparison to the computed null distribution. Genes that have previously been causally implicated in cancer according to the Cosmic database are highlighted in red[41]. Essential housekeeping genes (excluding those that are causally implicated in cancer) are highlighted in blue. The heat-map on the x-axis corresponds to the protein-mRNA correlation of each gene in the Cancer Cell Line Encyclopedia[34]. **Panel B:** Histogram depicting the distribution of Pearson correlation between mRNA expression and protein expression from the Cancer Cell Line Encyclopedia for all nodes included in our final Ewing sarcoma cell signaling networks. Proteins that were ranked particularly highly in panel A were labeled in panel B. **Pancel C:** Bar chart describing the most frequently observed genes among the top projected targets for the 15 patient tumor samples we analyzed. There was very little overlap in top projected targets between the cell line and patient data, reflecting the transcriptional heterogeneity present in Ewing Sarcoma.

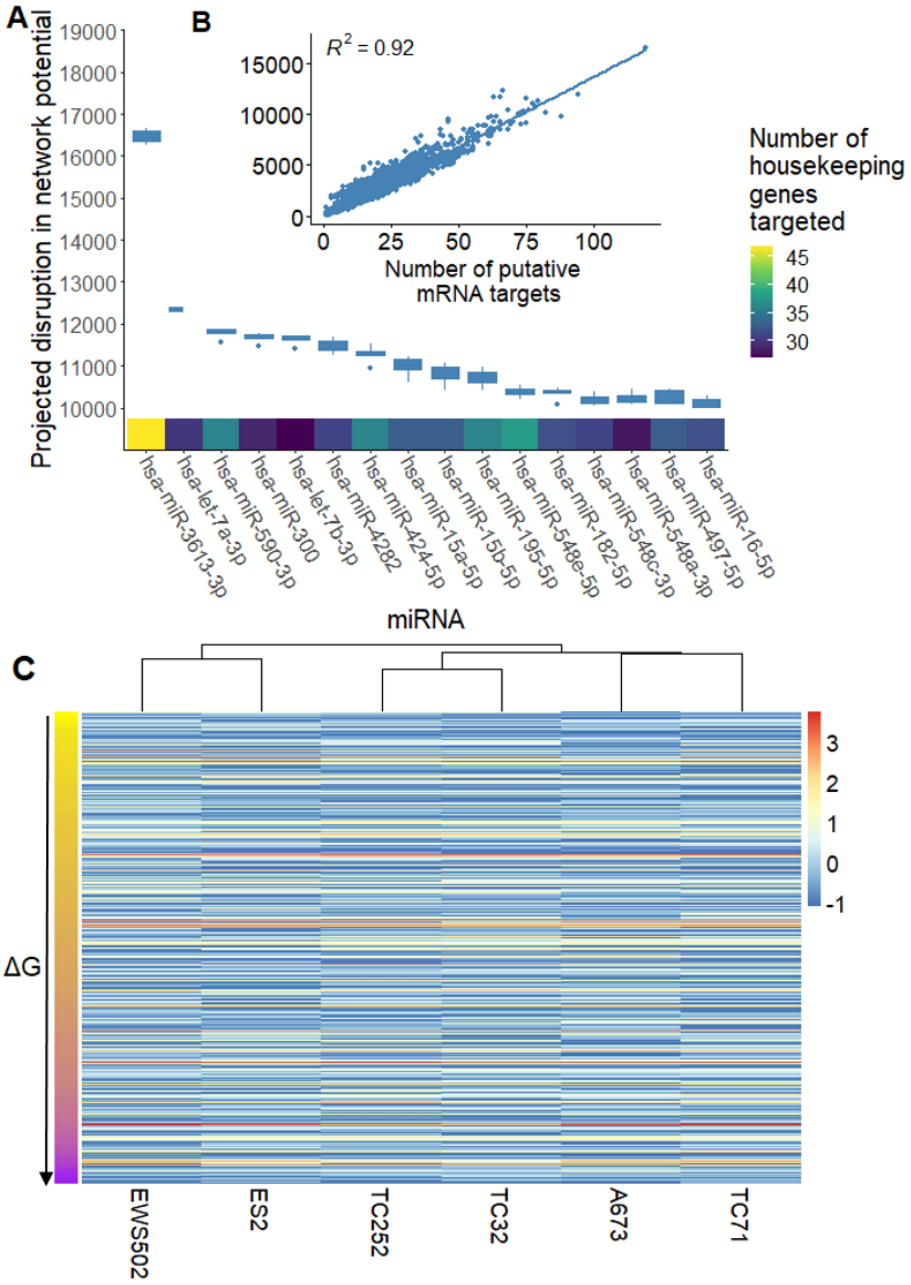

Figure 5. Many of the most promising miRNA candidates repress large numbers of essential housekeeping genes. We identified the top miRNA for treatment of Ewing sarcoma, ranked by their predicted disruption of the Ewing sarcoma cell signaling network. **A:** Boxplot showing the projected disruption in network potential for the top miRNA candidates (averaged across all samples). The heatmap on the x-axis describes the number of essential housekeeping genes that each miRNA is predicted to target. **B:** Scatterplot showing the relationship between projected network disruption and the number of putative mRNA targets for a given miRNA. **C:** Heatmap showing z-score normalized miRNA expression for 622 of the evaluated miRNA for the 6 cell lines under study. The Y axis is clustered by the projected ΔG associated with a given miRNA. There doesn't seem to be a clear pattern of miRNA expression based on projected ΔG.

the system less negative) of 17382, 13034, 12746, 12364 and 12280, respectively (see **Figure 5**). It should also be noted that we were able to identify a substantial number of miRNAs with potential activity against the Ewing sarcoma cell signaling network. We identified 27 miRNAs with an average predicted network potential disruption of greater than 10,000. For comparison, the largest network change in network potential that could be achieved with a single gene repression across all cell lines was just

2064 (TRIM25). miRNA sequencing of the 6 cell lines under study did not reveal any clear pattern of miRNA expression based on predicted network potential disruption.

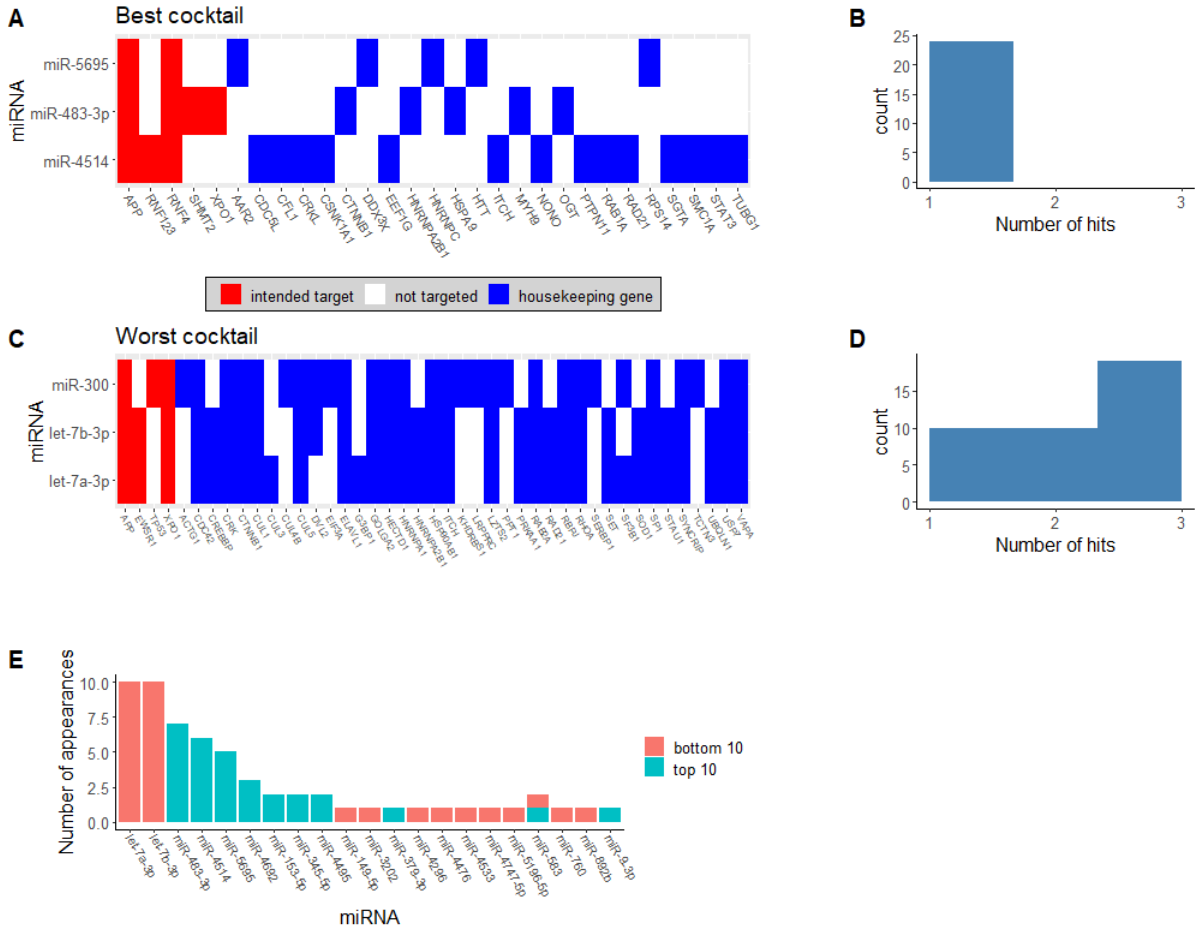

Figure 6. We identified miR-483-3p, miR-5695, and miR-4514s as the optimal 3-miRNA cocktail for Ewing Sarcoma therapy. We identified cocktails that are predicted to maximally downregulate target genes (red shading on the figure), while avoiding downregulation of essential housekeeping genes to limit toxicity (blue shading on the figure). **Panel A.** shows the targeting heatmap for the best predicted cocktail for cell line A673. The miRNA that make up the cocktail are presented on the y-axis. Putative gene targets are highlighted on the x-axis. Lines that span multiple miRNAs occur when a gene is downregulated by 2 or more miRNAs in the cocktail. **Panel B.** shows a histogram of the number of microRNA that target a given housekeeping gene in the best cocktail. **Panel C.** displays the targeting heatmap for the worst-performing cocktail for cell line A673 among those tested for reference. **Panel D.** shows a histogram of the number of microRNA that target a given housekeeping gene in the worst predicted cocktail. **Panel E** shows a bar graph showing the miRNA that most frequently appear in either the bottom or top 10 predicted cocktails (averaged across cell lines) for Ewing Sarcoma therapy.

These individual miRNAs target large numbers of transcripts in the cell and therefore may be difficult to administer as single-agents due to extreme toxicity. For example, the top miR candidate, miR-3613-3p, was predicted to repress 144 distinct mRNA transcripts in the full target set. We therefore sought to identify cocktails of miRNA that could cooperatively down-regulate key non-housekeeping genes while avoiding cooperative down-regulation of housekeeping genes that may be associated with toxicity. When targeting the top 10 predicted proteins from our *in silico* repression experiments, a 3 miRNA cocktail of miR-483-3p, miR-379-3p, and miR-345-5p was predicted to be the most optimal across all cell lines (**Figure 6A** and **Figure 6B**). Under the same conditions, a 3-miR cocktail of miR-300, let-7b-3p, and let-7a-3p was predicted to be the least optimal among 16,215 tested combinations (**Figure 6C** and **Figure 6D**). Notably, the most and least optimal miRNA combinations had similar activity against the 10 targets (**Figure 6A** and **Figure 6C**). The worst cocktail was defined by high levels of cooperative downregulation of housekeeping genes rather than lack of efficacy against putative targets (**Figure 6C** and **Figure 6D**). Let-7b-3p and let-7a-3p were heavily represented in the least optimal cocktails tested, appearing in 10 of the

10 worst 3 miRNA cocktails (**Figure 6E**). These highly promiscuous miRNA target large numbers of housekeeping genes, limiting their therapeutic utility alone or in combination (**Figure 5B**).

Notably, many of the most promising miRNA when considering only their total predicted network disruption tend to appear in the least optimal cocktails (**Figure 5**). This likely occurs because these miRNA tend to target large numbers of housekeeping genes and large numbers of genes overall. In contrast, the best miRNA cocktails tend to be composed of miRNA that target relatively few genes overall but exhibit some degree of target specificity. Put another way, they target the desired target genes while repressing relatively few essential housekeeping genes. An extreme example of this is the case of miR-483-3p. MiR-483-3p is in the bottom 50% of all miRNA when ranked by predicted network disruption, and is only predicted to repress 10 different transcripts. However, because it selectively targets several of our targets of interest, this relative small total projected network disruption is actually an attractive feature that makes it easy to build effective cocktails that include miR-483-3p. As a result, miR-483-3p appears in 7 of the top 10 predicted 3 miRNA cocktails. To assess the stability of our results, we repeated this analysis, focusing on the top 5 or top 15 predicted protein targets. We also repeated this analysis, assuming 10 % and 50% repression per miRNA that target a given mRNA. The top and bottom predicted cocktails were similar across these conditions and across all six cell lines. We have included the full ranked list of all miRNA cocktails tested across all conditions on Github

3 Discussion

In this work, we described a novel methodology for the identification of potential miRNA cocktails for Ewing sarcoma therapy. First, we performed paired miRNA and mRNA sequencing on six Ewing sarcoma cell lines (GEO accession GSE98787). We then defined a metric of cell state, network potential, based on mRNA expression and signaling network topology. Using *in silico* repression and change in network potential, we identified the most important proteins in the cell signaling network for each of the 6 cell lines. We then repeated this process for 15 patient tumor samples derived from the St. Jude Cloud[26]. Notably, this set of proteins was enriched in 24 of the 50 pathways included in the "halmarks" gene set[43,44]. The ranked protein set was also enriched for genes involved in the canonical miRNA biogenesis pathway[6]. We then evaluated more than 16000 3-miRNA cocktails (per cell line) based on predicted ability to disrupt key proteins in the Ewing Sarcoma cell signaling network while avoiding cooperative down-regulation of essential housekeeping genes. We ranked these 3-miRNA cocktails to identify promising miRNA combinations for therapy of Ewing Sarcoma.

The protein targets and miRNA candidates we identified in our dataset are consistent with the literature on Ewing sarcoma and cancer cell signaling, suggesting biological plausibility of our methodology. Of the top 50 protein targets that we identified, 15 were previously causally implicated in cancer[41], including EWSR1, the proposed driver of Ewing sarcoma development. In addition, our network-based approach suggests that known oncogenic hub genes such as KRAS and MYC are prime targets for disruption in cancer cells. We also identified a number of plausible targets that were not previously implicated in cancer, such as MOV10. MOV10 is an RNA helicase involved in the RNA-induced silencing complex (RISC), a key complex involved in epigenetic signaling by miRNA[47]. As mentioned previously, our findings suggest that the miRNA biogenesis pathway is enriched in the setting of Ewing Sarcoma. The central role of MOV10 in the EWS cell signaling network provides further evidence for the importance of miRNA signaling in EWS oncogenesis.

Many of the miRNA we identified as potential therapeutic candidates have been previously studied due to their association with cancer outcomes, including members of the let-7 family, miR-300, miR-424-5p, miR-4282, miR-15a-5p, and miR-590-3p. Loss of expression of the let-7 family of miRNA has been widely implicated in cancer development[48–51]. In Ewing sarcoma specifically, low levels of let-7 family miRNA have been correlated with disease progression or recurrence[48]. The let-7 family of miRNA have also been studied as treatment for non-small cell lung cancer in the pre-clinical setting[15]. Loss of miR-300 has been previously correlated with development and aggressiveness of hepatocellular carcinoma[52] as well as in oncogenesis of pituitary tumors[53]. Reduced expression of miR-424-5p and miR-4282 have each been implicated in the development of basal-like breast cancer[54,55]. MiR-15a-5p has been shown to have anti-melanoma activity[56]. In addition, miR-590-3p has been show to suppress proliferation of both breast cancer[57], and hepatocellular carcinoma[58]. The broad literature linking many of our proposed miRNA candidates for Ewing sarcoma treatment to the development and maintenance of cancer highlights the ability of our computational pipeline to identify potentially promising therapeutic candidates in this setting. Prior to application of these findings for treatment of Ewing sarcoma or any other disease, specific *in vitro* and *in vivo* validation is needed.

The process by which putative miRNA targets were selected was based on sequence homology rather than direct experimental validation. As a result, it is possible that we included false positive miRNA targets in our analysis. For this study we relied on a protein-protein interaction network presumably curated from analyzing normal human cells. It is possible that the derangements observed in cancer cells could change the underlying interaction network of a tumor cell. In the future, it may be possible to utilize protein-protein interaction networks specific to cancer or even specific to the cancer type under study. In addition, we did not consider specific binding sites that these miRNA may use to repress target mRNA. Certain miRNA may share binding sites on their target mRNA (i.e. the let-7 family of miRNA), which may make our assumption of linear additive miRNA effects invalid. We also used mRNA concentration as a surrogate for protein concentration in designing our cell signaling network.

While this is not true in all cases, it is likely a reasonable approximation under steady state conditions[30–33] (see Section 1.2 for more details). In addition, protein-mRNA correlations in the cancer cell line atlas for the top proteins identified by our pipeline were fairly good, ranging from 0.07 to 0.8 for the top 50 identified protein targets.[34] (Fig S2).

Despite these limitations, our findings may facilitate the development of novel therapies for patients suffering from Ewing Sarcoma. To this point, severe toxicity has limited the translation of miRNA-based cancer therapies to the clinical setting. Our pipeline may enable the development of better miRNA therapies that clear this hurdle and open up this promising avenue of therapy for patients suffering from cancer. In addition, this novel method can facilitate the rapid identification of key proteins in any cancer cell signaling network for which mRNA sequencing data is available. This may facilitate more rapid drug discovery and assist in the discovery of proteins and miRNA that play a significant role in the cancer disease process.

Acknowledgments

We acknowledge experimental support from Julia Selich-Anderson. JGS thanks the NIH for their support through NIH R37CA244613, their Loan Repayment Program, and the Paul Calabresi Career Development Award for Clinical Oncology (NIH K12CA076917). KIP acknowledges financial support from the NHMRC CJ Martin Overseas Biomedical Fellowship (APP1111032) and Alex's Lemonade Stand Young Investigator Grant (APP37138). DTW acknowledges financial support from the Cancer Center Trainee Award for Cancer Research from the Case Comprehensive Cancer Center. This work made use of the High Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University. JS acknowledges financial support from the NIH for their support through NIH T32GM007250.

Code and data availability

All of the software we developed for this project can be found on Github. (https://github.com/DavisWeaver/MiR_Combo_Targeting). All processed data needed for reproduction of the results of the paper are available in the same repository. We have also included an Rmarkdown version of the pre-print to aid in reproducibility. All raw data files were published on GEA (accession number GSE98787).

References

1. Mackintosh, C., Madoz-Gúrpide, J., Ordóñez, J. L., Osuna, D. & Herrero-Martín, D. The molecular pathogenesis of Ewing's sarcoma. *Cancer biology & therapy* **9**, 655–67 (2010).
2. Esiashvili, N., Goodman, M. & Marcus, R. B. Changes in Incidence and Survival of Ewing Sarcoma Patients Over the Past 3 Decades. *J. Pediatr. Hematol.* **30**, 425–430, DOI: 10.1097/MPH.0b013e31816e22f3 (2008).
3. Lawlor, E. R. & Sorensen, P. H. Twenty Years On – What Do We Really Know About Ewing Sarcoma And What Is The Path Forward? *Critical reviews oncogenesis* **20**, 155 (2015).
4. Riggi, N. *et al.* EWS-FLI-1 modulates miRNA145 and SOX2 expression to initiate mesenchymal stem cell reprogramming toward Ewing sarcoma cancer stem cells. *Genes & development* **24**, 916–32, DOI: 10.1101/gad.1899710 (2010).
5. Rupaimoole, R. & Slack, F. J. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nat. Rev. Drug Discov.* **16**, 203–222, DOI: 10.1038/nrd.2016.246 (2017).
6. Dhawan, A., Scott, J. G., Harris, A. L. & Buffa, F. M. Pan-cancer characterisation of microRNA across cancer hallmarks reveals microRNA-mediated downregulation of tumour suppressors. *Nat. Commun.* **9**, 5228, DOI: 10.1038/s41467-018-07657-1 (2018).
7. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *cell* **100**, 57–70 (2000).
8. LIN, X., YANG, Z., ZHANG, P. & SHAO, G. miR-154 suppresses non-small cell lung cancer growth in vitro and in vivo. *Oncol. Reports* **33**, 3053–3060, DOI: 10.3892/or.2015.3895 (2015).
9. Lin, K. *et al.* Loss of MIR15A and MIR16-1 at 13q14 is associated with increased TP53 mRNA, de-repression of BCL2 and adverse outcome in chronic lymphocytic leukaemia. *Br. J. Haematol.* **167**, 346–355, DOI: 10.1111/bjh.13043 (2014).
10. Reilley, M. J. *et al.* STAT3 antisense oligonucleotide AZD9150 in a subset of patients with heavily pretreated lymphoma: results of a phase 1b trial. *J. for ImmunoTherapy Cancer* **6**, 119, DOI: 10.1186/s40425-018-0436-5 (2018).
11. Hong, D. *et al.* AZD9150, a next-generation antisense oligonucleotide inhibitor of STAT3 with early evidence of clinical activity in lymphoma and lung cancer. *Sci. Transl. Medicine* **7**, 185–314, DOI: 10.1126/scitranslmed.aac5272 (2015).
12. Odate, S. *et al.* Inhibition of STAT3 with the Generation 2.5 Antisense Oligonucleotide, AZD9150, Decreases Neuroblastoma Tumorigenicity and Increases Chemosensitivity. *Clin. Cancer Res.* **23**, 1771–1784, DOI: 10.1158/1078-0432.CCR-16-1317 (2017).
13. Kasinski, A. L. & Slack, F. J. miRNA-34 prevents cancer initiation and progression in a therapeutically resistant K-ras and p53-induced mouse model of lung adenocarcinoma. *Cancer research* **72**, 5576–87, DOI: 10.1158/0008-5472.CAN-12-2001 (2012).
14. Wiggins, J. F. *et al.* Development of a Lung Cancer Therapeutic Based on the Tumor Suppressor MicroRNA-34. *Cancer Res.* **70**, 5923–5930, DOI: 10.1158/0008-5472.CAN-10-0655 (2010).
15. Stahlhut, C. & Slack, F. J. Combinatorial Action of MicroRNAs let-7 and miR-34 Effectively Synergizes with Erlotinib to Suppress Non-small Cell Lung Cancer Cell Proliferation. *Cell Cycle* **14**, 2171–2180, DOI: 10.1080/15384101.2014.1003008 (2015).
16. Liu, C. *et al.* The microRNA miR-34a inhibits prostate cancer stem cells and metastasis by directly repressing CD44. *Nat. medicine* **17**, 211–5, DOI: 10.1038/nm.2284 (2011).
17. Maciej Pajak, T. I. S. miRNAtap: miRNAtap: microRNA Targets— Aggregated Predictions., DOI: 10.18129 (2019).
18. Wong, N. & Wang, X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* **43**, D146–D152, DOI: 10.1093/nar/gku1104 (2015).
19. Maragkakis, M. *et al.* DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acids Res.* **37**, W273–W276, DOI: 10.1093/nar/gkp292 (2009).
20. O'Neill, V. A Multicenter Phase I Study of MRX34, MicroRNA miR-RX34 Liposomal Injection - Full Text View - ClinicalTrials.gov, DOI: NCT01829971.
21. Lai, X., Eberhardt, M., Schmitz, U. & Vera, J. Systems biology-based investigation of cooperating microRNAs as monotherapy or adjuvant therapy in cancer, DOI: 10.1093/nar/gkz638 (2019).
22. Rietman, E. A., Scott, J. G., Tuszynski, J. A. & Klement, G. L. Personalized anticancer therapy selection using molecular landscape topology and thermodynamics. *Oncotarget* **8**, 18735–18745, DOI: 10.18632/oncotarget.12932 (2017).

23. Rietman, E. A., Platig, J., Tuszynski, J. A. & Lakka Klement, G. Thermodynamic measures of cancer: Gibbs free energy and entropy of protein-protein interactions. *J. biological physics* **42**, 339–50, DOI: 10.1007/s10867-016-9410-y (2016).

24. Rietman, E. & Tuszynski, J. A. Using Thermodynamic Functions as an Organizing Principle in Cancer Biology. 139–157, DOI: 10.1007/978-3-319-74974-7{_}8 (Springer, Cham, 2018).

25. Pishas, K. I. *et al.* Therapeutic targeting of KDM1A/LSD1 in ewing sarcoma with SP-2509 engages the endoplasmic reticulum stress response. *Mol. Cancer Ther.* **17**, 1902–1916, DOI: 10.1158/1535-7163.MCT-18-0373 (2018).

26. McLeod, C. *et al.* St. Jude Cloud: A Pediatric Cancer Genomic Data-Sharing Ecosystem. *Cancer Discov.* **11**, 1082–1099, DOI: 10.1158/2159-8290.cd-20-1230 (2021).

27. Oughtred, R. *et al.* The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* **47**, D529–D541, DOI: 10.1093/nar/gky1079 (2019).

28. Zwiener, I., Frisch, B. & Binder, H. Transforming RNA-Seq Data to Improve the Performance of Prognostic Gene Signatures. *PLoS ONE* **9**, e85150, DOI: 10.1371/journal.pone.0085150 (2014).

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

30. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* **165**, 535–50, DOI: 10.1016/j.cell.2016.03.014 (2016).

31. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232, DOI: 10.1038/nrg3185 (2012).

32. Li, J. J., Bickel, P. J. & Biggin, M. D. System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ* **2**, e270, DOI: 10.7717/peerj.270 (2014).

33. Jovanovic, M. *et al.* Dynamic profiling of the protein life cycle in response to pathogens. *Science* **347**, 1259038–1259038, DOI: 10.1126/science.1259038 (2015).

34. Nusinow, D. P. *et al.* Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* **180**, 387–402.e16, DOI: 10.1016/j.cell.2019.12.023 (2020).

35. Albert, R., Jeong, H. & Barabási, A. L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382, DOI: 10.1038/35019019 (2000).

36. Dhawan, A. *et al.* Collateral sensitivity networks reveal evolutionary instability and novel treatment strategies in ALK mutated non-small cell lung cancer. *Sci. Reports* **7**, 1232, DOI: 10.1038/s41598-017-00791-8 (2017).

37. Enright, A. J. *et al.* MicroRNA targets in Drosophila. *Genome Biol.* **5**, R1, DOI: 10.1186/gb-2003-5-1-r1 (2003).

38. Lall, S. *et al.* A Genome-Wide Map of Conserved MicroRNA Targets in C. elegans. *Curr. Biol.* **16**, 460–471, DOI: 10.1016/j.cub.2006.01.050 (2006).

39. Friedman, R. C., Farh, K. K.-H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105, DOI: 10.1101/gr.082701.108 (2008).

40. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes, revisited, DOI: 10.1016/j.tig.2013.05.010 (2013).

41. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947, DOI: 10.1093/nar/gky1015 (2019).

42. Stelzer, G. *et al.* The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. In *Current Protocols in Bioinformatics*, vol. 54, 1–1, DOI: 10.1002/cpbi.5 (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2016).

43. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740, DOI: 10.1093/bioinformatics/btr260 (2011).

44. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst.* **1**, 417–425, DOI: 10.1016/j.cels.2015.12.004 (2015).

45. Korotkevich, G., Sukhov, V. & Sergushichev, A. A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *bioRxiv* 060012, DOI: 10.1101/060012 (2016).

46. Benjamini, Y., Yoav ; Hochberg. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 1995.pdf. *J. Royal Stat. Soc. Ser. B (Methological)* **57**, 289–300, DOI: 10.2307/2346101 (1995).

47. Goodier, J. L., Cheung, L. E. & Kazazian, H. H. MOV10 RNA Helicase Is a Potent Inhibitor of Retrotransposition in Cells. *PLoS Genet.* **8**, 1002941, DOI: 10.1371/journal.pgen.1002941 (2012).

48. Hameiri-Grossman, M. *et al.* The association between let-7, RAS and HIF-1 in Ewing Sarcoma tumor growth. *Oncotarget* **6**, 33834–48, DOI: 10.18632/oncotarget.5616 (2015).

49. Guo, M., Zhao, X., Yuan, X., Jiang, J. & Li, P. MiR-let-7a inhibits cell proliferation, migration, and invasion by down-regulating PKM2 in cervical cancer. *Oncotarget* **8**, 28226–28236, DOI: 10.18632/oncotarget.15999 (2017).

50. Tang, R. *et al.* MiR-let-7a inhibits cell proliferation, migration, and invasion by down-regulating PKM2 in gastric cancer. *Oncotarget* **7**, 5972–84, DOI: 10.18632/oncotarget.6821 (2016).

51. Pan, X.-M. *et al.* A let-7 KRAS rs712 polymorphism increases colorectal cancer risk. *Tumor Biol.* **35**, 831–835, DOI: 10.1007/s13277-013-1114-3 (2014).

52. Wang, R. *et al.* miR-300 regulates the epithelial-mesenchymal transition and invasion of hepatocellular carcinoma by targeting the FAK/PI3K/AKT signaling pathway. *Biomed. & Pharmacother.* **103**, 1632–1642, DOI: 10.1016/J.BIOPHA.2018.03.005 (2018).

53. Liang, H.-q. *et al.* The PTTG1-targeting miRNAs miR-329, miR-300, miR-381, and miR-655 inhibit pituitary tumor cell tumorigenesis and are involved in a p53/PTTG1 regulation feedback loop. *Oncotarget* **6**, 29413–27, DOI: 10.18632/oncotarget.5003 (2015).

54. Wang, J., Wang, S., Zhou, J. & Qian, Q. miR-424-5p regulates cell proliferation, migration and invasion by targeting doublecortin-like kinase 1 in basal-like breast cancer. *Biomed. & Pharmacother.* **102**, 147–152, DOI: 10.1016/J.BIOPHA.2018.03.018 (2018).

55. J. Zhao, G.-Q. J. MiR-4282 inhibits proliferation, invasion and metastasis of human breast cancer by targeting Myc. *Eur. review for medical pharmacological sciences* DOI: 10.26355/eurrev{_}201812{_}16643 (2018).

56. Alderman, C. & Yang, Y. The anti-melanoma activity and oncogenic targets of hsa-miR-15a-5p. *RNA & disease (Houston, Tex.)* **3** (2016).

57. Rohini, M., Gokulnath, M., Miranda, P. & Selvamurugan, N. miR-590–3p inhibits proliferation and promotes apoptosis by targeting activating transcription factor 3 in human breast cancer cells. *Biochimie* **154**, 10–18, DOI: 10.1016/J.BIOCHI.2018.07.023 (2018).

58. Ge, X. & Gong, L. MiR-590-3p suppresses hepatocellular carcinoma growth by targeting TEAD1. *Tumor Biol.* **39**, 101042831769594, DOI: 10.1177/1010428317695947 (2017).

59. Lin, S. & Gregory, R. I. MicroRNA biogenesis pathways in cancer. DOI: 10.1038/nrc3932 (2015).

1. **Figure S1: Network potential describes different features of a cell signaling network compared to mRNA expression alone. Panel A:** Histogram of mRNA expression for each gene (averaged across all samples). **Panel B:** Histogram of the network potential for each gene (averaged across all samples) mRNA transcripts with an expression level of zero were excluded from both histograms to better visualize the distribution of genes that are expressed. **Panel C:** Box plot showing the total mRNA expression for each cell line and patient sample (patient samples begin with SJEWS). **Panel D:** Box plot showing the total network potential for each cell line and patient sample.

2. **Figure S2: Protein targets ranked by contribution to network stability.** When averaging across cell lines, XPO1, LMNA, EWSR1, HSP90AA1, and CUL3 were identified as the most important proteins in the Ewing sarcoma cell signaling network (when limiting our analysis to proteins causally implicated in cancer[41]). When each protein was simulated as completely repressed *in silico*, network potential was increased by 654, 456, 429, 425, and 399, respectively. The heatmap at the bottom of the plot describes the protein-mRNA correlation for each gene in the cancer cell line atlas. Grey indicates no data was available. It is reassuring that EWSR1, the kinase associated with Ewing sarcoma development, is identified as highly influential in the cell signaling network by this method.

3. **Figure S3: Panel A:** Scatterplot with marginal histograms comparing mRNA expression to network potential. **Panel B:** Box and whisker plot showing the change in network potential for the top 50 genes, as well as 99.99% confidence intervals from the permutation test. We also show a histogram comparing the top 50 genes identified by our pipeline using stringdb compared to biogrid as the protein-protein interaction network.

4. **Supplemental Results:** Additional methods and results referenced in the main body of the text