Integration of Biodynamic Imaging and RNA-seq predicts chemotherapy response in canine diffuse large B-cell lymphoma

Jonathan Fine[1], Deepika Dhawan[2], Sagar Utturkar[3], Phillip San Miguel[4], Gaurav Chopra[1,3,5], John Turek[3,6], David Nolte[3,7], Michael O. Childress[2,3], and Nadia A. Lanman[3,5,8]

[1]Department of Chemistry, Purdue University, West Lafayette, IN
[2]Department of Veterinary Clinical Sciences, Purdue University, West Lafayette, IN
[3]Purdue University Center for Cancer Research, Purdue University, West Lafayette, IN
[4]Purdue University Genomics Core, Purdue University, West Lafayette, IN
[5]Integrative Data Science Initiative, Purdue University, West Lafayette, IN
[6]Department of Basic Medical Sciences, Purdue University, West Lafayette, IN
[7]Department of Physics, Purdue University, West Lafayette, IN
[8]Department of Comparative Pathobiology, West Lafayette, IN

Running Title: BDI and RNA-seq predict DLBCL CHOP response

Key Points

- Combined intracellular Doppler spectra and RNA-seq classify DLBCL samples as sensitive or resistance to CHOP chemotherapy
- Key dynamic features were identified that can be used to classify dogs with naturally-occurring DLBCL as CHOP-sensitive or -resistant

Abstract

Diffuse large B-cell lymphoma (DLBCL) is a common, aggressive cancer of notorious genotypic and phenotypic heterogeneity. A major challenge is predicting response to drug treatment that has typically been done using genomic tools alone with little success. A novel method that incorporates phenotypic profiling for predicting the effectiveness of therapy for individual patients is desperately needed. BioDynamic Imaging (BDI) is a technique for measuring time-dependent fluctuations in back-scattered light through living tumor tissues to identify critical changes in intracellular dynamics that are associated with phenotypic response to drugs. In this study, BDI and RNA sequencing (RNA-seq) data were collected on tumor samples from dogs with naturally occurring DLBCL, an animal model of increasingly recognized relevance to the human disease. BDI and RNA-seq data were combined to identify correlations between gene co-expression modules and linear combinations of biomarkers to provide biological mechanistic interpretations of BDI biomarkers. Using regularized multivariate logistic regression, we combined RNA-seq and BDI data to develop a novel machine learning model to accurately predict the clinical response of canine DLBCL to combination chemotherapy (i.e. CHOP). Our model incorporates data on the expression of 4 genes and 3 BDI-derived phenotypic biomarkers, capturing changes in transcription, microtubule related processes, and apoptosis. These results suggest that multi-scale genomic and phenotypic data integration can identify patients that respond to a given treatment *a priori* in a disease that has been difficult to treat. Our work provides an important framework for future development of strategies and treatments in precision cancer medicine.

Introduction

Diffuse large B-cell lymphoma (DLBCL) is a common, aggressive form of non-Hodgkin lymphoma diagnosed in approximately 25,000 human patients each year, 1/3 of whom will die from the disease(1, 2). Diffuse large B-cell lymphoma (DLBCL) is characterized by molecular and biochemical heterogeneity that have confounded the use of targeted drugs to improve cure rates from conventional chemoimmunotherapy (3). The current standard of care chemotherapy regimen for DLBCL is a combination therapy that combines rituximab, cyclophosphamide, doxorubicin hydrochloride, vincristine sulfate (Oncovin®), and prednisone (i.e. R-CHOP). Chemotherapy often fails due to drug resistance (4), and no targeted therapy has been developed which significantly improves survival (5-7). Challenges in predicting treatment efficacy for individual patients, coupled with a lack of understanding in the development of R-CHOP resistance mechanisms, is a significant cause of this difficulty. The International Prognostic Index, the most commonly used prognostic indicator, is based on simple clinical attributes (8, 9). While there are molecular prognostic schemes based on gene expression (10), there is no available method for predicting response to chemotherapy regimens in individual patients with DLBCL. Thus far, genetic analysis alone is insufficient to predict the response of individual cases of DLBCL to drug therapy (11).

In spite of the importance of developing personalized therapy in DLBCL, progress has been slow (4). Locus heterogeneity, an outbred population, and poor clinical documentation lead to difficulties in

human cancer gene mapping and predictive model development (12). While murine models have led to important breakthroughs in DLBCL research (13, 14), these models have limited translational application to precision cancer treatment (15, 16). Dogs with DLBCL have been proposed as a valuable model in which to develop novel personalized medicine strategies for humans with this cancer. Pet dogs develop DLBCL at a high rate, and the disease has similar clinical features to those seen in humans (12, 17). Pet dogs with DLBCL are treated with CHOP and exhibit heterogeneity in their response to CHOP chemotherapy as seen in humans' response to R-CHOP, making the dog model especially appropriate for informing precision medicine strategies to treat human DLBCL (17, 18).

BioDynamic Imaging (BDI) has been proposed as an accurate method to discriminate between DLBCL that will respond favorably versus unfavorably to treatment (19). BDI is an optical imaging technology that records phenotypic responses of fresh, living, three-dimensional tumor tissues to chemotherapeutic drugs in the *ex vivo* environment (5-7, 11, 19, 20). BDI quantifies how treatments change intracellular dynamics by using Doppler spectroscopy to measure the response to compounds inside viable *ex vivo* three-dimensional biopsies with intact tumor microenvironments. These responses are then statistically associated with clinical outcomes such as objective tumor response or survival time. Although preliminary results show that BDI predicts chemosensitivity of naturally-occurring DLBCL in dogs (19), the relationship of BDI data to molecular processes underlying a tumor's phenotypic drug response has yet to be defined.

We hypothesized that transcriptomic profiling could be correlated with BDI biomarkers, allowing us to characterize BDI biomarkers based on biological processes. Furthermore, we hypothesized that combined analysis of subcellular motion and gene expression patterns would enable the development of a machine learning model that could accurately classify DLBCL biopsies as sensitive or resistant to CHOP chemotherapy

In this study, we use intracellular dynamics data obtained from BDI and combine it with gene expression data from RNA sequencing (RNA-seq), defining the Doppler spectroscopic signatures recorded by BDI in terms of discrete biological processes. We also show that this integration creates an improved classifier of clinical chemotherapy response in canine DLBCL. Due to the success of machine learning applications across chemistry and biology (21-23), we explore multiple machine learning methodologies used to develop this classifier and discuss the implications of this model's predictions for the biology of these tumors. These studies provide validation for the use of pet dogs as a relevant model for human DLBCL and provide an excellent rationale for using multi-scale modeling to predict tumor response to chemotherapeutic regimens.

Methods

Assessment of Clinical Endpoints

Clinical management of study animals is detailed in Supplemental Methods. The primary clinical endpoints were objective response to chemotherapy and progression-free survival (PFS). The metric for assessing objective response was a caliper-based measurement measurements of a minimum of 1 and maximum of 5 peripheral lymph nodes. Objective response was classified according to established criteria (24) and is detailed in **Table S1**. Complete remission (CR) was defined as complete absence of detectable cancer following CHOP treatment. Partial remission (PR) was defined as >30% (but <100%) reduction in the sum of the longest diameters of up to 5 peripheral lymph nodes. Progressive disease (PD) was defined as >20% increase in the sum of the longest diameters of up to 5 peripheral lymph nodes, or the appearance of new lesions. Stable disease (SD) was defined as measurable tumor burden not meeting the criteria for PR or PD. Progression-free survival was defined as the time in days from initiation of chemotherapy to detection of PD or to death from any cause, whichever came first. All dogs in this study underwent incisional wedge biopsy or surgical extirpation of a peripheral lymph node at the

time of study entry (i.e. prior to initiation of chemotherapy) to provide tissue for histopathologic confirmation of DLCBL. Portions of these biopsy samples were reserved and processed for BDI and additional tumor biopsy portions were frozen in liquid nitrogen or homogenized in TRIzol reagent (ThermoFisher) within 30 minutes of harvesting for collection of tumoral RNA.

RNA-seq and analysis

All solid tumor specimens were stored in liquid nitrogen, while RNA samples were stored at -80°C until the time of RNA extraction. RNA extraction was performed using the RNeasy kit (Qiagen). Poly A selection and library preparation was performed using the TruSeq Stranded kit (Illumina) followed by sequencing on a NovaSeq6000. Read trimming of the 2x150 reads was performed using Trimmomatic v.0.32 (25) followed by mapping to the CanFam3.1 genome using STAR v.2.5.4b (26) and counting using HTSeq v.0.7.0 (27). A differential expression analysis was performed using the DESeq2 (28) package and later edgeR (29-31) to select intersecting differentially expressed genes (DEGs). The Benjamini-Hochberg method(32) corrected for multiple testing and genes with FDR < 0.05 were denoted as significant. IPA (Qiagen, CA, USA) and ClusterProfiler (33) were used to perform pathway and enrichment analyses and a pre-ranked GSEA was performed (34). DisGeNET v7.0 was used to probe disease-gene associations (35, 36). Enrichment and associations were deemed statistically significant at an adjusted p-value of 0.05. Data are available at GEO accession number GSE156220.

BioDynamic Doppler Imaging and Analysis

Complete details of the physics and optical engineering behind BioDynamic Imaging (BDI) can be found in (37, 38) and details of the application of BDI for drug screening can be found in (6, 11). The procedures for using BDI for measuring the response to chemotherapy, including sample handling and stabilization procedures, can be found in (19). Fresh tumor biopsy samples are processed into multiple small tissue samples of approximately $1mm^3$ volume, then immobilized in single wells of a 96-well plate before application of *ex vivo* drug treatments. Drug treatments consisted of a DMSO control, CHOP combination chemotherapeutic agent, as well as the individual components of CHOP in RPMI 1640 medium containing 0.1% DMSO. The CHOP therapy drugs were prepared containing doxorubicin (10µM), 4-hydroxycyclophosphamide (5µM), vincristine (60nM), and prednisolone (0.6µM). The longitudinal time course consists of 18 Loops over 12 hours: 6 Loops prior to drug through 12 Loops after drug. During the measurement of a sample in a single well 2000 digital holography (DH) camera frames are acquired of each sample at 25 fps. Each digital hologram is converted to the image domain through spatial 2D fast Fourier transforms (FFTs) to create a single image-domain section, termed an optical coherence image (OCI). A stack of 2000 image-domain sections (optical coherence images) constitute a time series for each pixel. The sample is masked from the background by the choice of an intensity threshold. The pixel-based time series is converted to a fluctuation power spectrum through temporal FFTs. The resulting pixel-based spectra are averaged into a single spectrum for a given sample on a given Loop. The baseline spectrum $S(\omega,0)$ is defined as the average of the last 4 Loops spectra prior to the treatment. The drug-response spectrogram is defined as

$$D(\omega,t) = log S(\omega,t) - log S(\omega,0)$$

where the time index *t* represents the Loop number. A spectrogram is generated for a given dog and a given treatment averaged over the wells. A data quality figure is assigned to each well based on multiple quality control criteria. The data quality has a base value of unity and is reduced by a factor of 2 for each criterion that is not satisfied by the well data. The spectrograms are averaged over the replicate wells for the treatment weighted by the data quality per well. Each dog's response is thus represented by an average spectrogram for each treatment. This time-frequency representation next must be converted into feature vectors, which form the basis of BDI-derived biomarkers of a tissue's phenotypic response to a given drug. Information on BDI biomarker definition is provided in the Supplemental Methods (**Tables S2-S4**).

Modeling of chemotherapy response

Models were trained on data and feature selection was performed by selecting BDI biomarkers with greater than a 0.80 area under the precision recall curve (AUPRC) and high confidence DEGs with greater than a 0.95 AUPRC. The Cohen Kappa statistic (39) was used to quantify model success based on inter-model reliability due to the unbalanced nature of the data. The resistant tumor samples were designated as the positive case and the sensitive tumor samples designated as the negative case. Kappa values of 0.80-0.90 are interpreted as strong, and above 0.90 is near perfect predictive ability (40, 41). The Caret package (42) was used to train models using leave-one-out cross validation (LOOCV) for hyperparameter tuning. Leave one out testing (LOOT) was used to enable testing of the final model.

Cross-correlation of BDI and transcriptomic variables

A network analysis was performed, identifying modules of co-expressed genes using WGCNA (43). Log(FPKM) values were used after filtering genes with <0.3 FPKM in 20% of samples. A soft thresholding power of 14 was chosen, which was the lowest power at which the scale-free topology had a fit R^2 of greater than 0.85 (44, 45). A signed adjacency matrix was computed and hierarchical clustering followed by a dynamic tree-cut algorithm identified modules with a minimum size of 30 genes. The first principal component of each of the modules (eigengenes), were used to correlate the modules with BDI biomarkers.

Results

Transcriptional changes associated with chemotherapy resistance mirror those seen in human DLBCL patients

Initially, we set out to determine whether the canine model is an appropriate system for studying chemotherapy treatment response and resistance in human DLBCL, identifying gene expression changes that accompany CHOP resistance in canine DLBCL in the process. RNA-seq and BDI were performed on tumors from nineteen pet dogs with confirmed diagnoses of DLBCL (**Figure 1**, **Table S1**). Of these tumors, 6 were categorized as resistant and 13 were considered sensitive to CHOP. BDI yielded 81 biomarkers from 5 drug treatments (CHOP along with individual monotherapy components). The canine RNA-seq data shows a high degree of variability between patients (**Figure 2S**), a characteristic also observed in human DLBCL (46).

A total of 70 differentially expressed genes (DEGs) were identified between sensitive and resistant dogs, 27 of which are downregulated, and 43 are upregulated in resistant tumors (**Figure 2A**). The enrichment analysis shows upregulation of the MAPK cascade in resistant tumors, a finding that is in common between this study and human DLBCL (47) as well as gene ontology (GO) terms, gene set enrichment analysis (GSEA) (34), and disease associations related to immune responses (**Figure 2A-D, Table S6-7**). In addition to immune related diseases, lymphoid neoplasms and drug responses are enriched and involve *ABCB1*, which is upregulated in resistant tumors. *ABCB1* encodes an ATP-binding cassette (ABC) transporter that is associated with multidrug resistance in both humans and dogs, including resistance to CHOP in canine DLBCL (48). The Jak-Stat pathway is activated in the CHOP-resistant tumors and has previously been shown to be activated in R-CHOP resistant tumors in human patients, drawing yet another similarity between human and canine DLBCL (49). The GSEA also highlights that *MYC* targets have lower expression in resistant samples. These findings fit with the pathogenic role that constitutively activated MYC plays in human B-cell lymphomas (50). Ultimately, a number of gene expression changes were identified that relate to the development of CHOP resistance. The results herein mirror those identified in studies of R-CHOP resistance in human DLBCL patients, validating the canine model system as an appropriate model for studying molecular features associated with human DLBCL.

BioDynamic Imaging captures phenotypic changes associated with CHOP treatment and resistance

We used BDI to capture phenotypic differences in treatment response that discriminate between dogs whose tumors respond favorably versus unfavorably to CHOP treatment. Time-frequency spectrograms were captured, measuring changes in intracellular motion in response to CHOP combination and monotherapy treatment. The time-frequency spectrograms are converted into feature vectors with elements associated with parts or patterns of the spectrograms. In addition to spectrogram-based features, there are also preconditions (such as light-scattering brightness and spectral density dynamic range) as well as drug-induced changes in these preconditions. All the raw biomarkers are defined in **Table S2**. The time-frequency decomposition is approached globally and locally. Global patterns are generated as low-order Legendre polynomials. These polynomials are taken as an inner product over the spectrograms to generate Legendre coefficients that represent the global features of the spectrograms. Only orders 0, 1 and 2 are used along the frequency and time axes to generate 9 global features. Local patterns are simply low, mid, and high-frequency bands with average, linear and quadratic time dependence, which generates 9 local features. The preconditions consist of normalized standard deviation (NSD), backscatter brightness (BSB), number of pixels in the sample mask (NCNT), the spectral dynamic range (DR), the Nyquist floor (NY), the knee frequency (KNEE), the half-width (HW), the spectral slope (S) and the linear slope (SF). Each precondition is changed by the drug treatment, providing additional features that are the changes in the preconditions from baseline to the endpoint of the assay. There are 27 drug-response features: 18 are based on spectrograms and 9 are drug-induced changes in preconditions. These 27 features are concatenated for each drug to create a feature vector of $27*N_{drug}$ elements (**Table S9)**.

A number of significant relationships are observed between biomarkers (**Figure 3A**, **Figure 3B**) and in many cases, biomarkers cluster by drug treatment. Therefore, a principal component analysis (PCA) was used to generate a set of orthologous features for each drug treatment. Within all drug treatments, the first 3 principal components contain the vast majority of the variability in the data, thus generating 15 features. The top 3 PCs contain 93.2% of the variability in the CHOP-, 88% in cyclophosphamide-, 79.3% in doxorubicin-, 90.6% in prednisone-, and 86.2% of the variability in vincristine-treated samples. For each of the five drug treatments, the loadings were used to determine the contribution of the BDI biomarkers to each principal component (**Figures 3C-G**). Ultimately, biomarkers describing phenotypic variation in treatment response were computed, providing quantitative measures of CHOP treatment effects on subcellular motion in live tissue samples.

Multi-scale data integration and modeling accurately predicts chemotherapy response

After BDI and RNA-seq data were collected from tumors, the goal was to determine whether a multi-scale model could be fit that would accurately classify tumor samples as sensitive or resistant to CHOP chemotherapy from pre-treatment data alone. The focus on pre-treatment data alone in classifying samples because of the high impact such a model could have, in which a clinical outcome could be predicted before patients are needlessly subjected to chemotherapy that does not improve survival time. A total of 45 high confidence protein coding DEGs were combined with BDI biomarkers to yield a total input space of 126 features.

Traditionally, evaluating a machine learning model requires one to partition the data into three parts: a training set, a validation set, and a testing set. The training set is directly introduced into the machine learning model and the weights and other parameters of the model are identified from this set. This model is then applied to predict the results of the validation set and the model is allowed to be retrained multiple times using different hyperparameters to optimize performance. A final model is created using a combination of the training and validation sets. Note that the validation set is allowed to change multiple times so that it encompasses all of the data in the training set in a scheme referred to as *cross validation*, which allows the model hyperparameters to be adjusted for a large amount of data.

The final model is then used to predict the results of the test set to obtain final statistics for the performance of the model. In addition to LOOCV, an evaluation paradigm called Leave One Out Testing (LOOT) was used to test models. Details of LOOT and a visual diagram of this process is given in **Figure S2**.

In these studies we use absolute shrinkage and selection operator LASSO (L1) and ridge (L2) regularized logistic regression, which are established machine learning techniques that generalize well to data not seen in training or validation sets (51, 52). To increase generalization, feature selection was employed based on variable Area-Under-the-Precision-Recall-Curve (AUPRC) (**Table S11-12**). BDI biomarkers with an AUPRC > 0.90 and high confidence RNA-seq variables (**Table S10**) with an AUPRC> 0.95 were selected. The 3 BDI biomarkers ($SDIP1_{dox}$, $LOF0_{chop}$, and $ALLF1_{pred}$) yielded a model with a kappa of 0.43 and an accuracy of 73.68% (**Table S13**). It should be noted that many of the BDI biomarkers are exceptionally strong predictors and the three chosen are not the only biomarkers that are strong predictors. Rather, the low kappa value suggests that the model using 81 BDI biomarkers is overfit and thus aggressive feature selection methods were employed in the current setting. The RNA variables selected are ENSCAFG00000029984, *KIAA1217*, *SH2D4A*, and *FGFR4* (**Figure 4A**). ENSCAFG00000029984 encodes a zinc-finger protein 92 homolog (E-value= $1.66x10^{-76}$ in BlastP(53)), *SH2D4A* encodes a potential tumor-suppressor protein (54) and *KIAA1217* encodes a protein implicated in lung cancer progression (55). *FGFR4* encodes a receptor for fibroblast growth factors that is involved in cell proliferation, differentiation and migration(56) and is associated with poor prognosis in lymphoma (57) and in the development of drug resistance (58, 59).

There is no significant correlation between the RNA and the BDI biomarkers selected and a Principal Component Analysis shows that these 7 variables separate the 19 patients (**Figure 4B-C**). A model trained on these 7 variables yielded a kappa of 1.00 with an accuracy of 100% (**Table 1**), showing that the selected variables enable prediction of poor clinical response to CHOP therapy (**Tables S14**). Hence, the integration of gene expression and phenotypic data in a regularized multivariate regression model enables accurate classification of response to CHOP from pre-treatment data only and significantly outperforms models that incorporate data describing only gene expression or phenotypic changes alone.

BDI biomarkers correlate with biological processes ranging from small to large-scale intracellular movements

Despite the apparent utility of BDI, the biological meaning of the majority of BDI spectral signatures was previously unknown due to the technical difficulty of associating spectrogram biomarkers with discrete molecular processes. To address this, co-expression networks were identified in the RNA-seq data and correlated to BDI data. A total of 16 modules of co-expressed genes were identified (**Figure 5A**), with sizes ranging from 44 to 2270 genes (**Table 2**). The networks were annotated with biological processes that are enriched in the genes grouped together in the modules (**Table S15**). To infer relationships between co-expression modules and BDI biomarkers, eigengenes were correlated with raw BDI biomarkers **(Table 2**, **Figure 5B**) and the top 3 principal components of each treatment (**Table 2, Figure 5C**), which are simply linear combinations of the raw BDI biomarkers, as determined by PCA loadings.

Precondition changes were identified to be related to large intracellular movements and shape changes in cells. The change in average intracellular speeds (DHW) and the change in overall activity (DNSD) under cyclophosphamide treatment and DNY markers describing the change in the Nyquist floor (fast organelle processes) for cyclophosphamide, doxorubicin, and prednisone are all related to large intracellular movements and microtubule related processes (Module 8, **Table 2**).

The change in the dynamic range of spectral density (DDR), change in intracellular speeds (DHW) and the change in the spectral slope (DSF) are strongly correlated with changes in cellular shape (Modules 12 and 13, **Table 2**) through the linear combinations of biomarkers under CHOP, vincristine, cyclophosphamide, and prednisone treatments.

Golgi to endosome transport (Module 15, **Table 2**) is correlated with the global biomarker describing high frequency regions of the spectra across all times (HI0). The correlation of the HI0 BDI biomarkers with a module enriched for golgi to endosome transport fits with previous observations showing that high frequency regions of the spectrograms are indicative of fine movements, such as vesicle transport (20).

The biomarker $HI2_{dox}$, which captures quadratic time dependencies at high frequencies as well as $LOF0_{dox}$, which captures low frequency regions of spectrograms across all times, are correlated with apoptosis (Module 12, **Table 2**), fitting with prior observations (11). The 3 BDI biomarkers ($SDIP1_{dox}$, $LOF0_{chop}$, and $ALLF1_{pred}$) from logistic regression, were determined to be related to transcriptional/translational processes, microtubule related processes, and apoptosis, respectively. This work enables interpretation of BDI biomarkers in terms of molecular processes, thus enhancing the utility of BDI and suggesting that BDI can provide a means by which to draw biological conclusions about treatment effects in live *ex vivo* tissue biopsies.

Discussion

This study illustrates the value of the canine model to study chemotherapy resistance to DLBCL and captures the response to therapy as well as the development of CHOP resistance in terms of both phenotypic and transcriptional responses. While a caveat of these studies is the small sample size, a number of known factors in human DLBCL were nonetheless observed amongst the DEGs. BDI data captured subcellular movement in response to treatment and as a result of this study, biological interpretations of BDI biomarkers are now available. BDI biomarkers that capture response to apoptosis, cellular movement, shape changes, and vesicle trafficking were identified. Future studies with larger sample sizes and on a variety of types of cancer and treatments are likely to refine the biological annotations of BDI biomarkers.

We show that the integration of multiscale data describing cellular and molecular dynamics in a machine learning model predicts chemotherapeutic response in a relevant animal model of DLBCL. While the BDI biomarkers are strong predictors, the small sample size in this study necessitated aggressive feature selection. The three BDI biomarkers that we included in the most accurate models are not to be taken as the only important BDI biomarkers. Likewise, because different BDI biomarkers are associated with different biological processes, the BDI biomarkers with the most powerful predictive capacity may change depending on the treatment and cancer type assayed. Future studies on targeted therapies as well as the cytotoxic drugs studied here should be performed with larger samples sizes.

We show that BDI is a powerful technology for predicting treatment response and provide biological interpretations of the BDI spectrograms. As this technology is a powerful method for characterizing cellular response to environments in living tissue, this resource is expected to aid in future interpretation of these data and the phenotypic responses accordingly. While genetic and transcriptomic profiles can be used to identify signatures for classifying samples, phenotypic variability often obfuscates the clinical validity of such methods. The studies here make a strong case for building predictive multivariate models that incorporate phenotypic and transcriptomic measures of drug response, such BDI, on live *ex vivo* tissue. Such approaches are needed in order to move personalized medicine forward and to predict therapeutic efficacy for individual patients.

Acknowledgements

These studies were supported by the Walther Cancer Foundation, the Purdue University Center for Cancer Research (NIH grant P30 CA023168), the IU Simon Cancer Center (NIH grant P30 CA082709) and the Ralph W. and Grace M. Showalter Research Trust, and The Purdue University Challenge award. We gratefully acknowledge support of the Purdue University Genomics Core Facility, the Collaborative Core for Cancer Bioinformatics, and the staff at the Purdue University School of Veterinary Medicine. Additional funds in part by the Integrative Data Science Institute award at Purdue University and the National Institutes of Health, National Center for Advancing Translational Sciences ASPIRE Design Challenge awards are also acknowledged. We also greatly appreciate Purdue University Research Computing and Purdue ECN for continued computational and IT support. We are grateful to the clients at the Purdue University School of Veterinary Medicine, specifically those whom consented to have their beloved pets enrolled in this study.

Authorship Contributions

Contribution: J.T., D.N., M.O.C., and N.A.L. conceived and designed experiments; J.F., P.S.M, J.T., D.N., M.O.C., and N.A.L. designed methodology; D.D, J.T, D.N., M.O.C. acquired the data, managed patients, and provided facilities; J.F., S.U., G.C., D.N., M.O.C., and N.A.L. analyzed and interpreted the data J.F., D.N., S.U., and N.A.L wrote the paper and prepared figures; J.T., D.N., M.O.C., G.C, S.U., and N.A.L. reviewed/revised manuscript

Correspondence: Nadia A. Lanman, Purdue University, 201 S. University Street, West Lafayette, IN 47907. Phone: 765-494-2635**;** email: natallah@purdue.edu and Michael O. Childress, Purdue University, 625 Harrison Street, West Lafayette, IN 47907. Phone: 765-494-1107; email: mochildr@purdue.edu

Conflicts of Interest Disclosures

David Nolte and John Turek have a financial interest in Animated Dynamics, Inc. that is seeking to commercialize biodynamic technology with a intellectual property license from Purdue University.

References

1. Teras LR, DeSantis CE, Cerhan JR, Morton LM, Jemal A, Flowers CR. 2016 US lymphoid malignancy statistics by World Health Organization subtypes. CA Cancer J Clin. 2016;66(6):443-59. Epub 2016/09/13. doi: 10.3322/caac.21357. PubMed PMID: 27618563.
2. Roschewski M, Staudt LM, Wilson WH. Diffuse large B-cell lymphoma-treatment approaches in the molecular era. Nat Rev Clin Oncol. 2014;11(1):12-23. Epub 2013/11/13. doi: 10.1038/nrclinonc.2013.197. PubMed PMID: 24217204.
3. Younes A, Ansell S, Fowler N, Wilson W, de Vos S, Seymour J, Advani R, Forero A, Morschhauser F, Kersten MJ, Tobinai K, Zinzani PL, Zucca E, Abramson J, Vose J. The landscape of new drugs in lymphoma. Nat Rev Clin Oncol. 2017;14(6):335-46. Epub 2016/12/30. doi: 10.1038/nrclinonc.2016.205. PubMed PMID: 28031560; PMCID: PMC5611863.
4. Intlekofer AM, Younes A. Precision therapy for lymphoma--current state and future directions. Nat Rev Clin Oncol. 2014;11(10):585-96. Epub 2014/08/20. doi: 10.1038/nrclinonc.2014.137. PubMed PMID: 25135367.
5. Merrill D, An R, Sun H, Yakubov B, Matei D, Turek J, Nolte D. Intracellular Doppler Signatures of Platinum Sensitivity Captured by Biodynamic Profiling in Ovarian Xenografts. Sci Rep. 2016;6:18821. Epub 2016/01/07. doi: 10.1038/srep18821. PubMed PMID: 26732545; PMCID: PMC4702146.
6. Sun H, Merrill D, An R, Turek J, Matei D, Nolte DD. Biodynamic imaging for phenotypic profiling of three-dimensional tissue culture. J Biomed Opt. 2017;22(1):16007. Epub 2017/03/17. doi: 10.1117/1.JBO.22.1.016007. PubMed PMID: 28301634; PMCID: PMC5221565.
7. An R, Wang C, Turek J, Machaty Z, Nolte DD. Biodynamic imaging of live porcine oocytes, zygotes and blastocysts for viability assessment in assisted reproductive technologies. Biomed Opt Express. 2015;6(3):963-76. Epub 2015/03/24. doi: 10.1364/BOE.6.000963. PubMed PMID: 25798318; PMCID: PMC4361448.
8. International Non-Hodgkin's Lymphoma Prognostic Factors P. A predictive model for aggressive non-Hodgkin's lymphoma. N Engl J Med. 1993;329(14):987-94. Epub 1993/09/30. doi: 10.1056/NEJM199309303291402. PubMed PMID: 8141877.
9. Sehn LH, Berry B, Chhanabhai M, Fitzgerald C, Gill K, Hoskins P, Klasa R, Savage KJ, Shenkier T, Sutherland J, Gascoyne RD, Connors JM. The revised International Prognostic Index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. Blood. 2007;109(5):1857-61. Epub 2006/11/16. doi: 10.1182/blood-2006-08-038257. PubMed PMID: 17105812.
10. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Jr., Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000;403(6769):503-11. Epub 2000/02/17. doi: 10.1038/35000501. PubMed PMID: 10676951.
11. An R, Merrill D, Avramova L, Sturgis J, Tsiper M, Robinson JP, Turek J, Nolte DD. Phenotypic profiling of Raf inhibitors and mitochondrial toxicity in 3D tissue using biodynamic imaging. J Biomol Screen. 2014;19(4):526-37. Epub 2013/12/24. doi: 10.1177/1087057113516674. PubMed PMID: 24361645; PMCID: PMC4028716.
12. Davis BW, Ostrander EA. Domestic dogs and cancer research: a breed-based genomics approach. ILAR J. 2014;55(1):59-68. doi: 10.1093/ilar/ilu017. PubMed PMID: 24936030; PMCID: PMC4158346.
13. Takizawa M, Tolarova H, Li Z, Dubois W, Lim S, Callen E, Franco S, Mosaico M, Feigenbaum L, Alt FW, Nussenzweig A, Potter M, Casellas R. AID expression levels determine the extent of cMyc oncogenic

translocations and the incidence of B cell tumor development. J Exp Med. 2008;205(9):1949-57. doi: 10.1084/jem.20081007. PubMed PMID: 18678733; PMCID: PMC2526190.
14. Pasqualucci L, Bhagat G, Jankovic M, Compagno M, Smith P, Muramatsu M, Honjo T, Morse HC, 3rd, Nussenzweig MC, Dalla-Favera R. AID is required for germinal center-derived lymphomagenesis. Nat Genet. 2008;40(1):108-12. doi: 10.1038/ng.2007.35. PubMed PMID: 18066064.
15. Moreno L, Pearson AD. How can attrition rates be reduced in cancer drug discovery? Expert Opin Drug Discov. 2013;8(4):363-8. doi: 10.1517/17460441.2013.768984. PubMed PMID: 23373702.
16. Jackson SJ, Thomas GJ. Human tissue models in cancer research: looking beyond the mouse. Dis Model Mech. 2017;10(8):939-42. doi: 10.1242/dmm.031260. PubMed PMID: 28768734; PMCID: PMC5560067.
17. Richards KL, Motsinger-Reif AA, Chen HW, Fedoriw Y, Fan C, Nielsen DM, Small GW, Thomas R, Smith C, Dave SS, Perou CM, Breen M, Borst LB, Suter SE. Gene profiling of canine B-cell lymphoma reveals germinal center and postgerminal center subtypes with different survival times, modeling human DLBCL. Cancer Res. 2013;73(16):5029-39. doi: 10.1158/0008-5472.CAN-12-3546. PubMed PMID: 23783577; PMCID: PMC3755352.
18. Childress MO, Ramos-Vara JA, Ruple A. Retrospective analysis of factors affecting clinical outcome following CHOP-based chemotherapy in dogs with primary nodal diffuse large B-cell lymphoma. Vet Comp Oncol. 2018;16(1):E159-E68. doi: 10.1111/vco.12364. PubMed PMID: 29152834.
19. Choi H, Li Z, Sun H, Merrill D, Turek J, Childress M, Nolte D. Biodynamic digital holography of chemoresistance in a pre-clinical trial of canine B-cell lymphoma. Biomed Opt Express. 2018;9(5):2214-28. Epub 2018/05/16. doi: 10.1364/BOE.9.002214. PubMed PMID: 29760982; PMCID: PMC5946783.
20. Custead MR, An R, Turek JJ, Moore GE, Nolte DD, Childress MO. Predictive value of ex vivo biodynamic imaging in determining response to chemotherapy in dogs with spontaneous non-Hodgkin's lymphomas: a preliminary study. Converg Sci Phys Oncol. 2015;1(1). Epub 2015/01/01. doi: 10.1088/2057-1739/1/1/015003. PubMed PMID: 27280042; PMCID: PMC4894334.
21. Fine J, Rasjashekar A, Chopra G. Accurate and Automated de novo Identification of Molecular Functional Groups Using Deep Learning Architectures. ChemRxiv. 2019.
22. Brylinski M, Skolnick J. Cross-reactivity virtual profiling of the human kinome by X-react(KIN): a chemical systems biology approach. Mol Pharm. 2010;7(6):2324-33. Epub 2010/10/21. doi: 10.1021/mp1002976. PubMed PMID: 20958088; PMCID: PMC2997910.
23. Ashtawy HM, Mahapatra NR. A Comparative Assessment of Predictive Accuracies of Conventional and Machine Learning Scoring Functions for Protein-Ligand Binding Affinity Prediction. IEEE/ACM Trans Comput Biol Bioinform. 2015;12(2):335-47. Epub 2015/09/12. doi: 10.1109/TCBB.2014.2351824. PubMed PMID: 26357221.
24. Vail DM, Michels GM, Khanna C, Selting KA, London CA, Veterinary Cooperative Oncology G. Response evaluation criteria for peripheral nodal lymphoma in dogs (v1.0)--a Veterinary Cooperative Oncology Group (VCOG) consensus document. Vet Comp Oncol. 2010;8(1):28-37. Epub 2010/03/17. doi: 10.1111/j.1476-5829.2009.00200.x. PubMed PMID: 20230579.
25. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20. Epub 2014/04/04. doi: 10.1093/bioinformatics/btu170. PubMed PMID: 24695404; PMCID: PMC4103590.
26. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21. Epub 2012/10/30. doi: 10.1093/bioinformatics/bts635. PubMed PMID: 23104886; PMCID: PMC3530905.
27. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-9. Epub 2014/09/28. doi: 10.1093/bioinformatics/btu638. PubMed PMID: 25260700; PMCID: PMC4287950.

28. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. Epub 2014/12/18. doi: 10.1186/s13059-014-0550-8. PubMed PMID: 25516281; PMCID: PMC4302049.
29. Dai Z, Sheridan JM, Gearing LJ, Moore DL, Su S, Wormald S, Wilcox S, O'Connor L, Dickins RA, Blewitt ME, Ritchie ME. edgeR: a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens. F1000Res. 2014;3:95. Epub 2014/05/27. doi: 10.12688/f1000research.3928.2. PubMed PMID: 24860646; PMCID: PMC4023662.
30. Nikolayeva O, Robinson MD. edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology. Methods Mol Biol. 2014;1150:45-79. Epub 2014/04/20. doi: 10.1007/978-1-4939-0512-6_3. PubMed PMID: 24743990.
31. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40. Epub 2009/11/17. doi: 10.1093/bioinformatics/btp616. PubMed PMID: 19910308; PMCID: PMC2796818.
32. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med. 1990;9(7):811-8. Epub 1990/07/01. doi: 10.1002/sim.4780090710. PubMed PMID: 2218183.
33. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-7. Epub 2012/03/30. doi: 10.1089/omi.2011.0118. PubMed PMID: 22455463; PMCID: PMC3339379.
34. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. Bioinformatics. 2007;23(23):3251-3. Epub 2007/07/24. doi: 10.1093/bioinformatics/btm369. PubMed PMID: 17644558.
35. Pinero J, Bravo A, Queralt-Rosinach N, Gutierrez-Sacristan A, Deu-Pons J, Centeno E, Garcia-Garcia J, Sanz F, Furlong LI. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. Nucleic Acids Res. 2017;45(D1):D833-D9. Epub 2016/12/08. doi: 10.1093/nar/gkw943. PubMed PMID: 27924018; PMCID: PMC5210640.
36. Pinero J, Ramirez-Anguita JM, Sauch-Pitarch J, Ronzano F, Centeno E, Sanz F, Furlong LI. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020;48(D1):D845-D55. Epub 2019/11/05. doi: 10.1093/nar/gkz1021. PubMed PMID: 31680165; PMCID: PMC7145631.
37. Jeong K, Turek JJ, Nolte DD. Fourier-domain digital holographic optical coherence imaging of living tissue. Appl Opt. 2007;46(22):4999-5008. Epub 2007/08/07. doi: 10.1364/ao.46.004999. PubMed PMID: 17676107.
38. Li Z, Sun H, Turek J, Jalal S, Childress M, Nolte DD. Doppler fluctuation spectroscopy of intracellular dynamics in living tissue. J Opt Soc Am A Opt Image Sci Vis. 2019;36(4):665-77. Epub 2019/05/03. doi: 10.1364/JOSAA.36.000665. PubMed PMID: 31044988; PMCID: PMC6791373.
39. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull. 1968;70(4):213-20. Epub 1968/10/01. doi: 10.1037/h0026256. PubMed PMID: 19673146.
40. Brennan P, Silman A. Statistical methods for assessing observer variability in clinical measures. BMJ. 1992;304(6840):1491-4. Epub 1992/06/06. doi: 10.1136/bmj.304.6840.1491. PubMed PMID: 1611375; PMCID: PMC1882212.
41. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276-82. Epub 2012/10/25. PubMed PMID: 23092060; PMCID: PMC3900052.
42. M K. Building Predictive Models in R Using the caret Package. J Stat Software, Artic. 2008;28(5):1–26.
43. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. Epub 2008/12/31. doi: 10.1186/1471-2105-9-559. PubMed PMID: 19114008; PMCID: PMC2631488.

44. Barabasi AL, Bonabeau E. Scale-free networks. Sci Am. 2003;288(5):60-9. Epub 2003/04/19. doi: 10.1038/scientificamerican0503-60. PubMed PMID: 12701331.
45. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005;4:Article17. Epub 2006/05/02. doi: 10.2202/1544-6115.1128. PubMed PMID: 16646834.
46. Autio M, Leivonen SK, Bruck O, Mustjoki S, Jorgensen JM, Karjalainen-Lindsberg ML, Beiske K, Holte H, Pellinen T, Leppa S. Immune cell constitution in the tumor microenvironment predicts the outcome in diffuse large B-cell lymphoma. Haematologica. 2020. Epub 2020/02/23. doi: 10.3324/haematol.2019.243626. PubMed PMID: 32079690.
47. Vega GG, Aviles-Salas A, Chalapud JR, Martinez-Paniagua M, Pelayo R, Mayani H, Hernandez-Pando R, Martinez-Maza O, Huerta-Yepez S, Bonavida B, Vega MI. P38 MAPK expression and activation predicts failure of response to CHOP in patients with Diffuse Large B-Cell Lymphoma. BMC Cancer. 2015;15:722. Epub 2015/10/18. doi: 10.1186/s12885-015-1778-8. PubMed PMID: 26475474; PMCID: PMC4609122.
48. Tomiyasu H, Tsujimoto H. Comparative Aspects of Molecular Mechanisms of Drug Resistance through ABC Transporters and Other Related Molecules in Canine Lymphoma. Vet Sci. 2015;2(3):185-205. Epub 2015/08/12. doi: 10.3390/vetsci2030185. PubMed PMID: 29061940; PMCID: PMC5644633.
49. Rui L, Drennan AC, Ceribelli M, Zhu F, Wright GW, Huang DW, Xiao W, Li Y, Grindle KM, Lu L, Hodson DJ, Shaffer AL, Zhao H, Xu W, Yang Y, Staudt LM. Epigenetic gene regulation by Janus kinase 1 in diffuse large B-cell lymphoma. Proc Natl Acad Sci U S A. 2016;113(46):E7260-E7. Epub 2016/11/02. doi: 10.1073/pnas.1610970113. PubMed PMID: 27799566; PMCID: PMC5135360.
50. Hemann MT, Bric A, Teruya-Feldstein J, Herbst A, Nilsson JA, Cordon-Cardo C, Cleveland JL, Tansey WP, Lowe SW. Evasion of the p53 tumour surveillance network by tumour-derived MYC mutants. Nature. 2005;436(7052):807-11. Epub 2005/08/12. doi: 10.1038/nature03845. PubMed PMID: 16094360; PMCID: PMC4599579.
51. Allam A, Nagy M, Thoma G, Krauthammer M. Neural networks versus Logistic regression for 30 days all-cause readmission prediction. Sci Rep. 2019;9(1):9277. Epub 2019/06/28. doi: 10.1038/s41598-019-45685-z. PubMed PMID: 31243311; PMCID: PMC6595068.
52. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12-22. Epub 2019/02/15. doi: 10.1016/j.jclinepi.2019.02.004. PubMed PMID: 30763612.
53. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10. Epub 1990/10/05. doi: 10.1016/S0022-2836(05)80360-2. PubMed PMID: 2231712.
54. Quagliata L, Andreozzi M, Kovac M, Tornillo L, Makowska Z, Moretti F, Heim MH, Heinimann K, Piscuoglio S, Terracciano LM. SH2D4A is frequently downregulated in hepatocellular carcinoma and cirrhotic nodules. Eur J Cancer. 2014;50(4):731-8. Epub 2013/12/10. doi: 10.1016/j.ejca.2013.11.018. PubMed PMID: 24315626.
55. Lee MS, Kim RN, I H, Oh DY, Song JY, Noh KW, Kim YJ, Yang JW, Lira ME, Lee CH, Lee MK, Kim YD, Mao M, Han J, Kim J, Choi YL. Identification of a novel partner gene, KIAA1217, fused to RET: Functional characterization and inhibitor sensitivity of two isoforms in lung adenocarcinoma. Oncotarget. 2016;7(24):36101-14. Epub 2016/05/07. doi: 10.18632/oncotarget.9137. PubMed PMID: 27150058; PMCID: PMC5094986.
56. Prieto-Dominguez N, Shull AY, Teng Y. Making way for suppressing the FGF19/FGFR4 axis in cancer. Future Med Chem. 2018;10(20):2457-70. Epub 2018/10/17. doi: 10.4155/fmc-2018-0099. PubMed PMID: 30325210.

57. Gao L, Feng Z, Li Q, Li L, Chen L, Xiao T. Fibroblast growth factor receptor 4 polymorphism is associated with increased risk and poor prognosis of non-Hodgkin's lymphoma. Tumour Biol. 2014;35(4):2997-3002. Epub 2013/11/20. doi: 10.1007/s13277-013-1386-7. PubMed PMID: 24248544.
58. Turkington RC, Longley DB, Allen WL, Stevenson L, McLaughlin K, Dunne PD, Blayney JK, Salto-Tellez M, Van Schaeybroeck S, Johnston PG. Fibroblast growth factor receptor 4 (FGFR4): a targetable regulator of drug resistance in colorectal cancer. Cell Death Dis. 2014;5:e1046. Epub 2014/02/08. doi: 10.1038/cddis.2014.10. PubMed PMID: 24503538; PMCID: PMC3944229.
59. Lang L, Teng Y. Fibroblast Growth Factor Receptor 4 Targeting in Cancer: New Insights into Mechanisms and Therapeutic Strategies. Cells. 2019;8(1). Epub 2019/01/13. doi: 10.3390/cells8010031. PubMed PMID: 30634399; PMCID: PMC6356571.

Table 1: Summary of Leave One Out Testing. Details are given in **Tables (S20 – S)**

Model	Accuracy	Precision	Recall	F1	TEST SET Kappa	MEAN Validation Kappa
ALL BDI BIOMARKERS	0.6316	0.4	0.3333	0.3636	0.1074	0.1894
ALL RNA VARIABLES	0.6842	0.5	0.1667	0.25	0.1094	0.3906
TOP 3 BDI BIOMARKERS	0.7368	0.57	0.6667	0.6154	0.4172	0.4272
TOP 5 RNA VARIABLES	0.8947	0.75	1	0.8571	0.7765	0.7671
BDI AND RNA VARIABLES	1	1	1	1	1	0.9922

[a] Best is defined as the top BDI and RNA variables as measured by their respective AUPRC values

Table 2. Modules of co-expressed genes, functions and associated BDI biomarkers.

Module number	Module Color	Number of genes in the cluster	Enriched biological processes	Associated Individual BDI Biomarkers	Associated BDI PCs
0	grey	6110	Metabolism, apoptosis, ligase activity, interpret with caution, genes in this network are not correlated with each other	**All CDIP0(-)** except vinc and cyclop, CDIP1pred(-), CDIP2chop(+), CDIP2pred(+), CDIP2vinc(+), **all DHW(-),**	0
1	cyan	102	Protein assembly, ribosomal related processes	CDIP1dox(-), DSFpred(+), LOF0pred(+)	Pred2(+)

2	green	410	Lipid transport, transport	AllF0dox(-), CDIP0chop(-), CDIP0doc(-), DHWdox(-), DNSDcyclop(+), DNYcyclop(+), **all MID0 (-)**	Dox1(-), dox3(-), chop2(+), cyclop2(-)
3	lightcyan	44	Protein synthesis and elongation, growth, genes encoding ribosomal proteins (these proteins are involved in protein synthesis but also involved in cellular stress response)	CDIP0dox (-), CDIP1dox(-),CDIP1cyclop, CDIP1pred(-), DSBdox(+), DDRdox(+), DDRpred(+), DHWdox(-), DSFpred(+), **all MID0(-)**	Dox1(-), dox3(-),
4	brown	637	Regulation of apoptosis, degredation, metabolism, ATP related metabolic processes, protein degredation	CDIP1cyclop (+), CDIP2pred(+), DNSDdox(+), DSFpred(-), HI2dox(-), LOF0dox(+), LOF0pred(+), SDIP0chop(+)	Pre2(-)
5	salmon	104	Immune related/inflammation related processes, response to stimulus	AllF1dox(+), AllF1vinc(-), AllF2vinc(+), CDIP1dox(+), DSFpred(-), LOF0dox(+), MID0dox(+)	Dox1(+), dox3(+), pred2(+)
6	yellow	600	Splicing, transcriptional regulation, some miRNA related gene sets	CDIP0dox(+), CDIP0pred(+), DSBdox(-), DDRdox(-), DDRpred(-)	Dox2(-), pred1(+)
7	greenyellow	147	Mitosis, cell cycle, spindle formation, proliferation, organelle fission	**AllF1vinc(-),** CDIP2chop(+),CDIP2pred(+), DHWpred(-)	Pred2(-)
8	red	246	Microtubule related processes, many genes that encode components of the cytoplasmic dynein motor protein complex	AllF0chop(-), CDIP0chop(-), CDIP0dox(-), CDIP0pred(-), CDIP1pred(-), CDIP2chop(+), CDIP2pred(+), DDRdox(+), DHWpred(-), DNYpred(-), MID0chop(-), MID1pred(-), MID1pred(-)	Dox2(+), chop2(+), pred1(-)
9	black	230	Cell migration/adhesion, locomotion	DHWcyclop(-), DNYpred(-),	0
10	blue	727	miRNA regulation, post transcriptional/translational regulation	AllF0dox(-), AllF1dox (-), CDIP0cyclop(-), DNSDdox(-), DNYdox(-), DNYdox(-), DSFpred(+), LOF0pred(+), LOF1dox(-), MID0dox(-), MID2dox(+), SDIP0chop(-), SDIP1dox(+), SDIP2pred(+)	Dox1(-)
11	turquoise	2270	Chromatin organization, RNA biosynthesis	SDIP2vinc(+), CDIP0cyclop(-), CDIP1cyclop(-), DHWvinc(-), DNSDdox(-), DNYdox(-), all HI0(-), HI2dox(+), SDIP0chop(-)	Chop3(+)
12	midnight blue	44	Actin cytoskeleton related, fibril formation	SDIP2chop(+), SDIP2vinc(+)	Chop3(+), vinc2(+),cyclop1(-)

13	purple	205	EMT, extracellular matrix organization,	AllF1chop(+), all CDIP0(-), all CDIP2(+), all DBSB(+), DHWcyclop(-), DHWpred(-), DHWvinc(-), DSFcyclop(+), H1cyclop(+)	Pred1(-), cyclop1(-)
14	magenta	208	Taxis, chemotaxis	CDIP0dox(-), CDIP2pred(-), DNYpred(-)	0
15	pink	225	Golgi to endosome transport, ribosomal proteins	CDIP0cyclop(-),CDIP1dox(+), **all HI0(-)**, SDIP0pred(-), SDIP2vinc(+)	Dox2(-), chop3(+)
16	tan	146	Immune activation	SDIP1vinc(-), DSBdox(+), DNYpred(-),all HI0(-),SDIP0chop(-)	Chop3(+)

Figure Legends

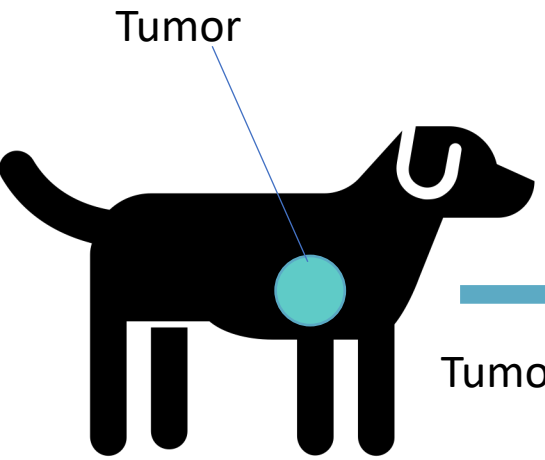
Figure 1. Schematic of the current study design. RNA-seq and BDI was performed on tumors from nineteen pet dogs with confirmed diagnoses of spontaneous diffuse large B-cell lymphoma. Tumor biopsies were taken from each dog before and after chemotherapy. These biopsies were then subjected to RNA sequencing (RNA-seq) as well as BioDynamic Imaging (BDI) on the standard-of-care CHOP combination chemotherapy as well as individual monotherapies. BDI measures subcellular motion using Doppler spectroscopy in live 3-dimensional tissue samples. Dogs were later placed on a CHOP regimen and then were classified as sensitive or resistant to chemotherapy based on progression-free survival times. Gene expression and subcellular motion changes in response to CHOP treatment and differences between sensitive and resistant dogs were identified. A network analysis identified modules of co-expressed genes which were correlated with BDI biomarkers. Together, RNA-seq and BDI were integrated into a regularized logistic regression model that accurately predicts chemotherapy response from pre-treatment data alone.

Figure 2. RNA-seq differential expression and enrichment results. A. Volcano plot with statistically significantly differentially expressed genes colored orange if they have higher expression in resistant samples and pink if they have higher expression in sensitive samples. **B.** Bar chart showing the pathways that are statistically significantly differentially expressed (adjusted p-value <0.05 **C.** A gene concept map showing a number of hub genes that are differentially expressed between sensitive and resistant DLBCL samples. Grey nodes indicate diseases that are statistically significantly enriched (padj <0.05) amongst the genes altered in resistant samples from DisGeNET, with size depicting the number of differentially expressed genes associate with the disease. Colored nodes show hub genes colored by the $\log_2$(fold-change) in resistant/sensitive samples, with yellow depicting higher expression of the gene in resistant samples and purple showing higher expression of the gene in sensitive samples **D.** Barchart showing the normalized enrichment score (NES) for the gene set enrichment analysis (GSEA) for significantly enriched gene sets. Positive enrichment scores indicate that the genes in the gene set have higher expression in resistant samples than sensitive samples. Negative enrichment scores indicate that the genes in the gene set have higher expression in the sensitive samples.

Figure 3. Linear relationships exist amongst BDI variables. A. Hierarchical clustering based on Euclidean distance between BDI variables. Red depicts a low dissimilarity (distance), white depicts moderate dissimilarity, and blue represents high dissimilarity between variables. **B.** A spearman correlation matrix shows linear relationships exist between many BDI variables. Red depicts a high positive correlation between markers, white depicts a lack of correlation, and blue depicts a negative correlation. **C.** Loadings for the top three principal components in BDI biomarker PCA analysis under cyclophosphamide treatment. PC1 is termed "cyclop1", PC2 is termed "cyclop2, and PC3 is termed "cyclop3" **D.** Loadings for the top three principal components in BDI biomarker PCA analysis under prednisone treatment. **E.** Loadings for the top three principal components in BDI biomarker PCA analysis under doxycycline treatment. **F.** Loadings for the top three principal components in BDI biomarker PCA analysis under CHOP treatment. **G.** Loadings for the top three principal components in BDI biomarker PCA analysis under vincristine treatment.

Figure 4. A. Boxplots of normalized RNA variable counts for features used in the model. Counts per million shown in tumors that are sensitive or resistant to chemotherapy. **B.** Spearman Correlations between the top 20 BDI variables and statistically significant RNA-seq variables are given in (a) where green boxes show the decision tree selected variables used to build the logistic regression model. **C.** A principal component analysis plot is given in (b) to show how the selected variables separate the resistant versus sensitive lymphoma tumors. Additional correlation plots for the Spearman and Kendall correlations are provided in the supporting information.

Figure 5. Network analysis and correlation with BDI biomarkers. A. Gene dendrogram generated by average linkage hierarchical clustering on log transformed FPKM pre-CHOP values. The color row underneath the dendrogram shows the module assignment determined by the Dynamic Tree Cut. **B.** Spearman correlations between identified modules and raw BDI biomarkers. **C.** Spearman correlations between identified co-expressed modules of genes and linear combinations of BDI biomarkers, generated by taking the first 3 principal components within each drug treatment.



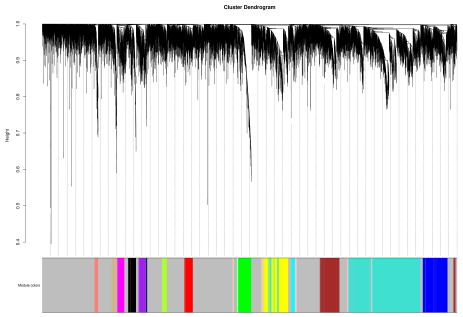
Tumor biopsy

BDI

RNA-seq

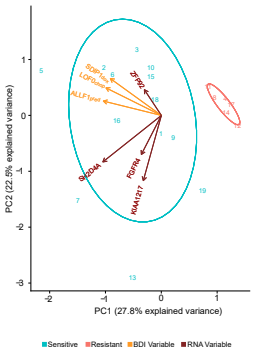
Drug treatment

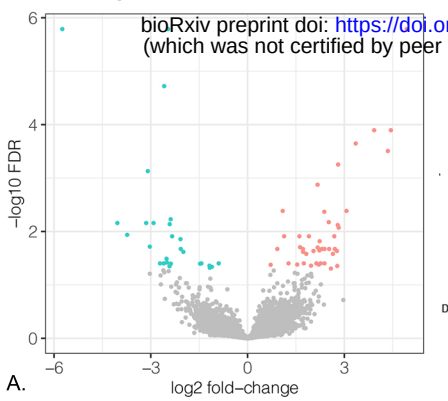
Measure light scattering on treated samples



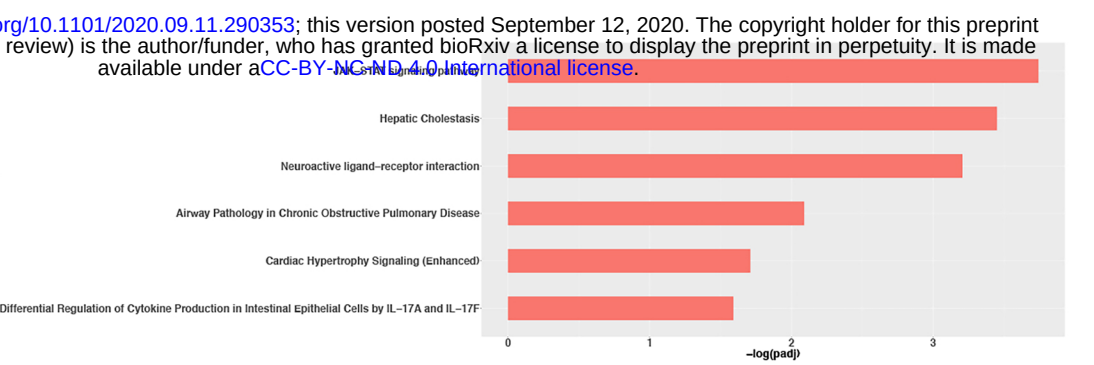
Network analysis

Cross-correlation and machine learning

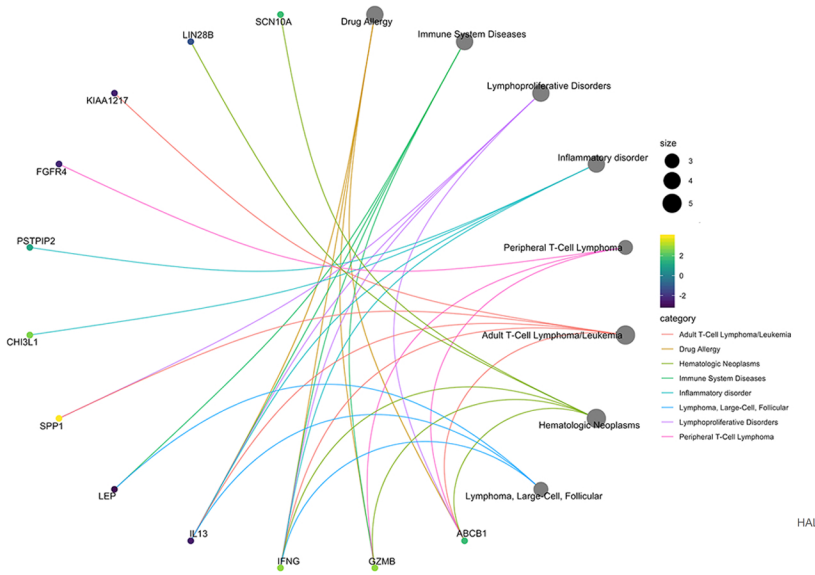




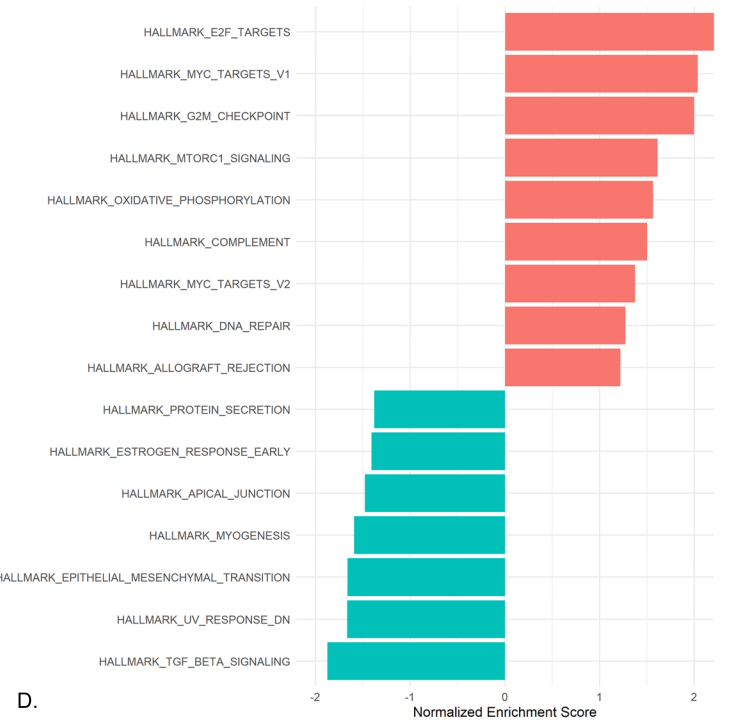
A.



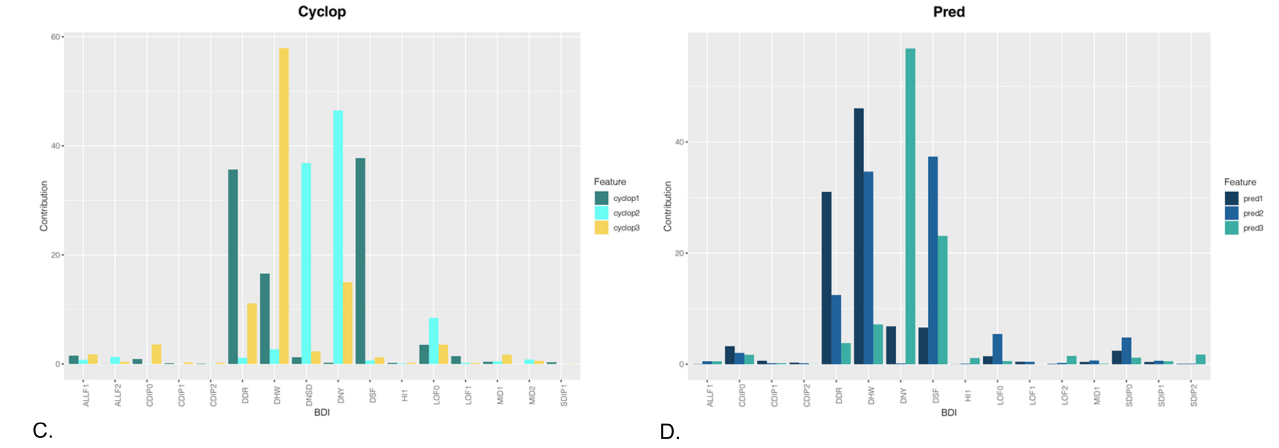
B.

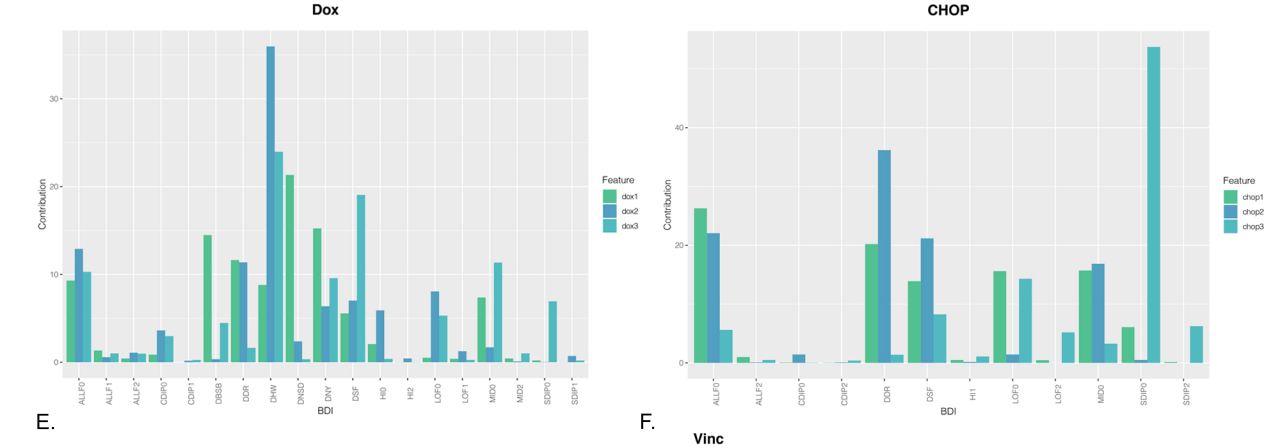


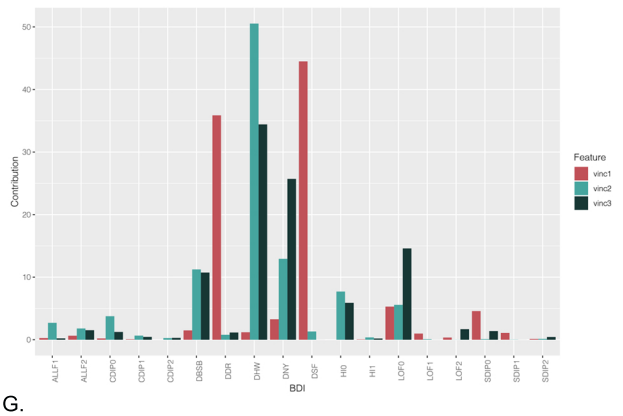
C.

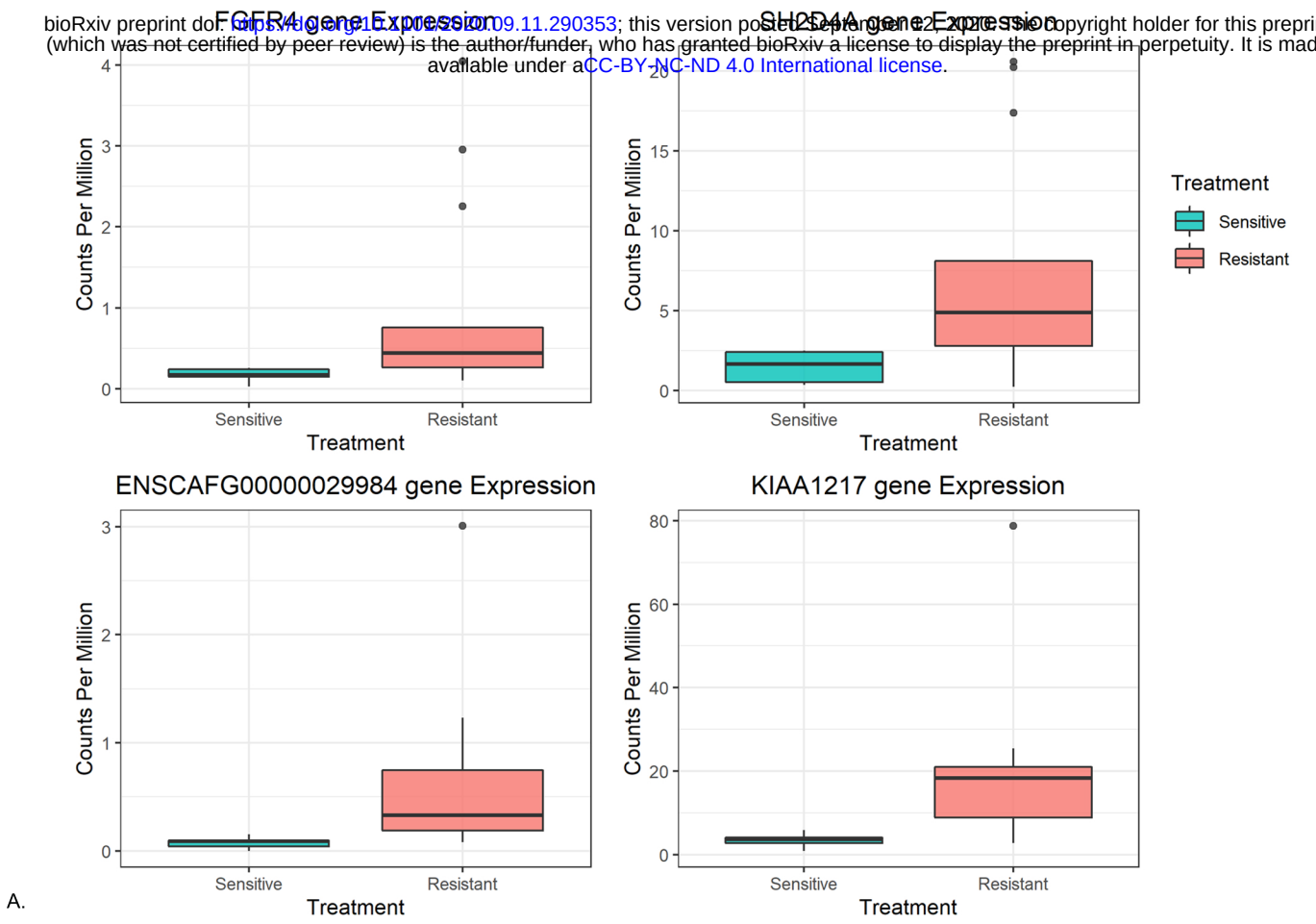


D.

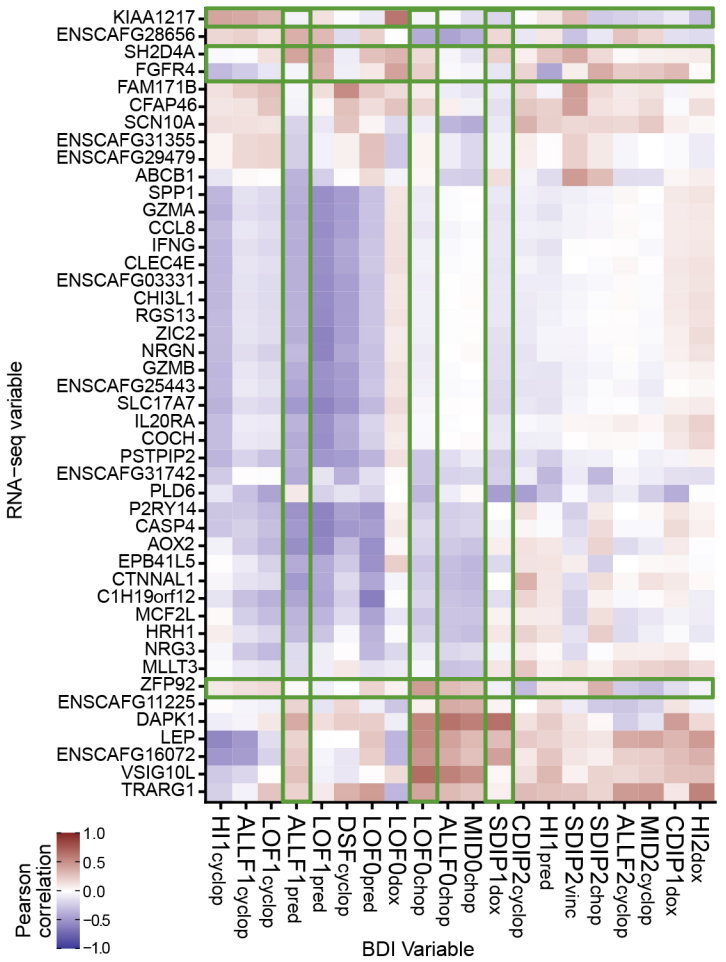




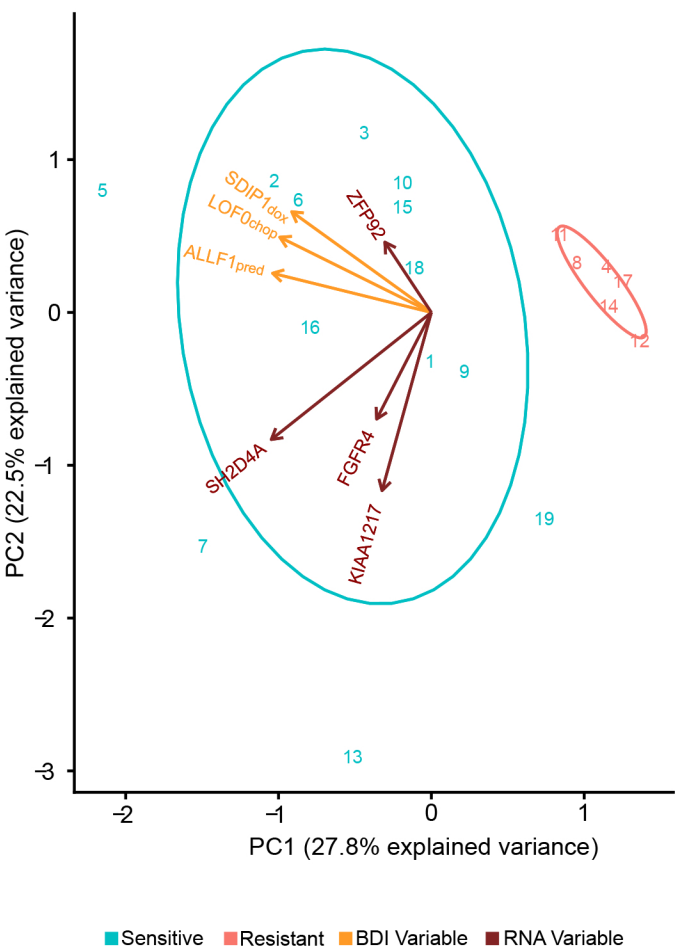




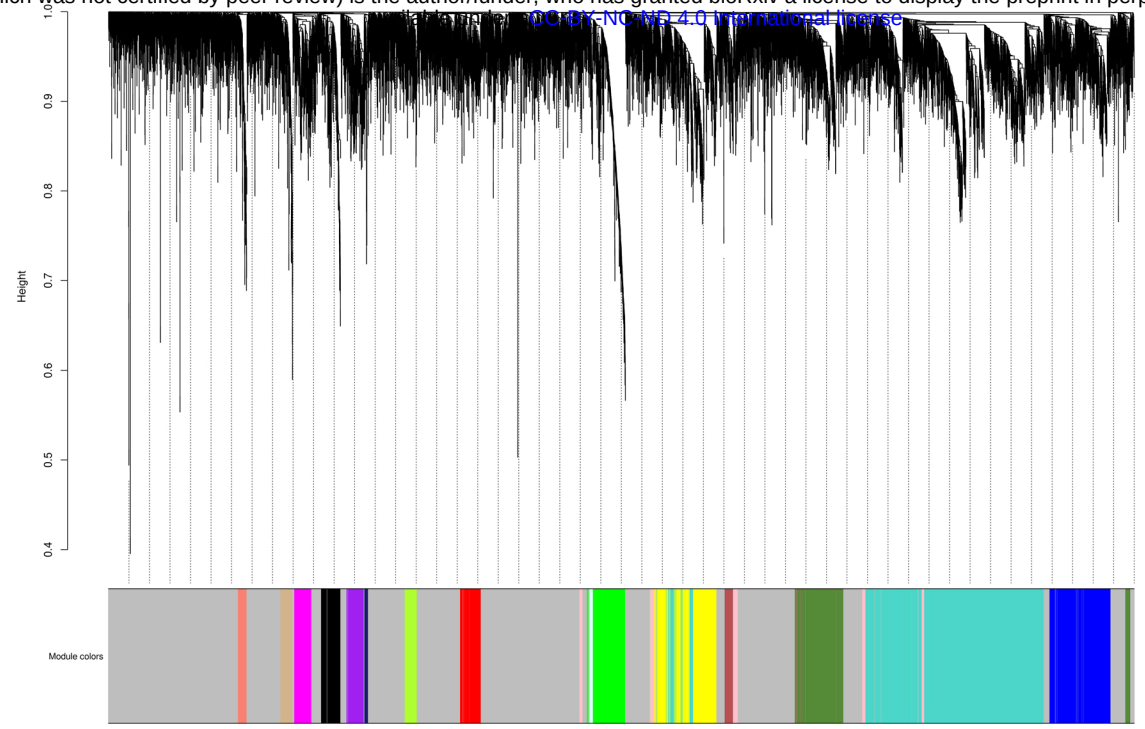
A.



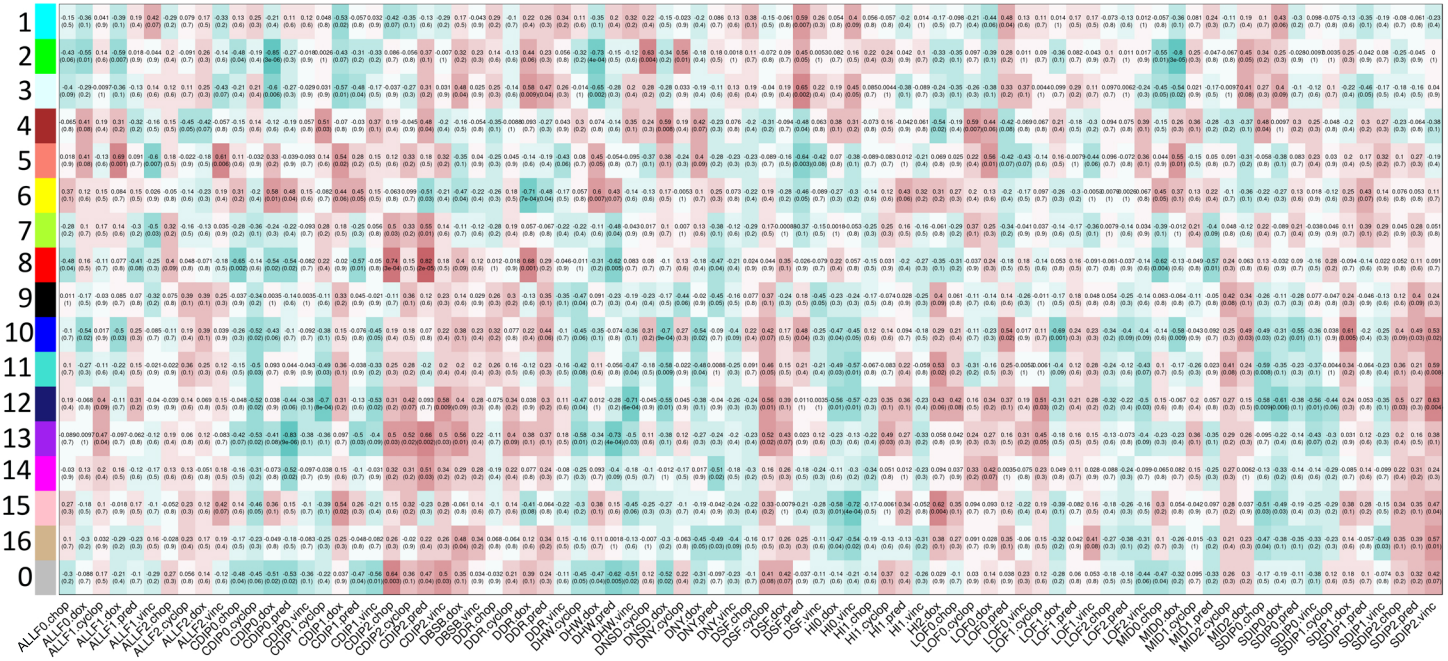
B.



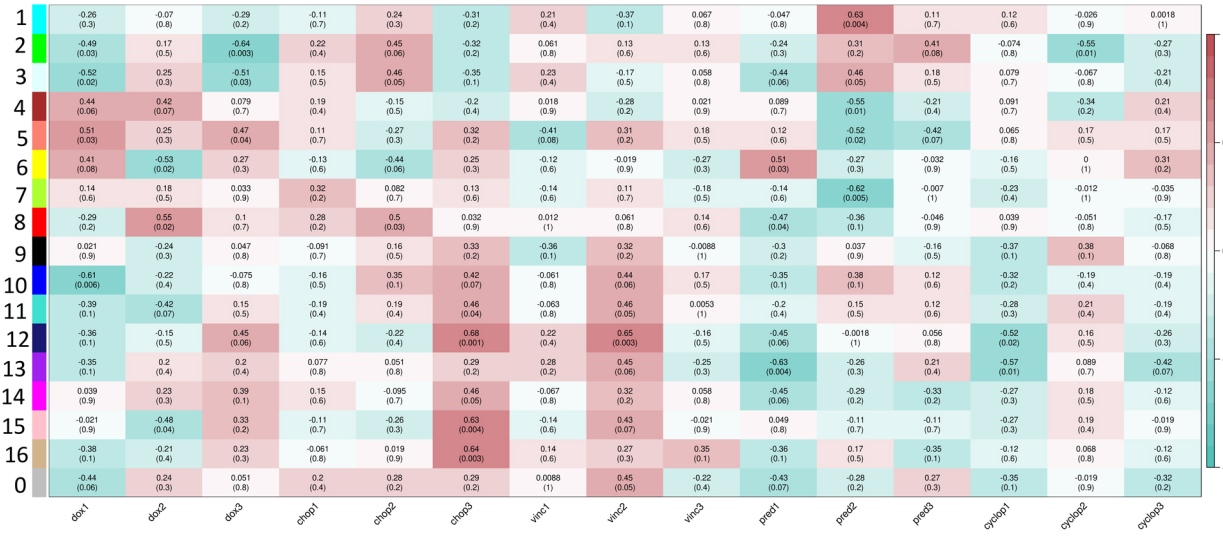
C.



A.



B.



C.