Cancer; A 21st Century Plague

RP4

Group 4: Canine - Regression- Biophysics

Faris Abuain, Bryce Bear, Jaden Chagnon, Lucas Monter, Arham Saeed

April 04, 2024

ISCI 1A24B

Table of Contents

Abstract

It is ubiquitous in the scientific community to test a process, drug or procedure on an analogous system to humans prior to human trials. In this project, Canine Venereal Tumour Disease (CVTD) in canines is explored as an analogous system with close applications to leukemia, Wilms' tumour, and skin cancer. This two-sided study included a computational model and a biophysical lab proposal. Study 1 aimed to uncover the relationship between numerous exome genes and a decrease in tumour proportion, along with compounds which can activate or inactive the genes accordingly, and finally, how to predict the inhibitory capacity (PCI50) of said compounds. Study 2 aimed to explore the effect of different vincristine concentrations using Vincasar PFS® on tumour proportion and compare the experimental PCI50 with the computational model in Study 1. Study 1 benefited greatly from the public whole-exome genome canine dataset and a chemical compound database from the European Institute for Bioinformatics. Seven genes with a positive correlation to a decrease in tumour proportion were selected, and accompanying compounds were compiled into a dataset for the development of an ordinary least squares regression model, which excluded outliers using the interquartile range. A biophysical study is proposed to determine the efficacy of the chemotherapeutic drug vincristine (Vincasar PFS®) in the treatment of CVTD. In order to make this determination, 10 cell culture lines were prepared: three unmodified tumour cell samples from a canine afflicted with CVTD and seven tumour cell samples which were modified with CRISPR to either activate or inactivate key genes that were identified in Study 1. Each cell line will be cultured nine times each for a total of 90 cell culture mediums. Three dosages of Vincasar PFS® will be administered to each cell culture three times, and responsive tumour cell cultures will be identified through cytotoxicity assay. Subsequently, those culture mediums will be analyzed with MRI/MRS technology to collect tumour proportion data, calculate PCI50 for each genotype/cell culture line, and make determinations as to the general and relative efficacy of Vincasar PFS® as a treatment for CVTD. A multi-linear regression model was derived using nine chemical fingerprints to predict PCI50. Most notably, the number of rotatable bonds, AlogP, CXlogP, and CXlogD (all pertaining to the octanal/water partition coefficient) were strongly correlated to the PCI50 value. On training and testing data, the nine-variable linear model had an R^2 value of 0.93 and 0.89, an adjusted R^2 value of 0.84 and 0.57, and a mean squared error of 0.01 and 1.57, respectively. The results from Study 1 will facilitate and inform the results of Study 2. The application of

computational models to recapitulate a canine's response to Vincasar PFS® should provide invaluable insights into the potential biophysical responses of humans to pro-drugs, but there exists an irrefutable research gap between these studies. To vet the model more accurately and to investigate the displacement of vincristine as an earlier-order drug, the model's assumptions and salient evolutionary behaviour of canine cultured cells must be analyzed and in/validated thoroughly through a biophysical wet lab.

Keywords: PCI50, Canine Venereal Tumour Disease (CVTD), MRI, vincristine, ordinary least squares.

General Introduction & Background

The canine transmissible venereal tumour disease (CVTD) is a type of transmissible cancer that occurs naturally through allogeneic cellular transplants.[1] The mechanism for transmission of CVTD involves the blood of an afflicted individual coming into contact with another canine. Once the foreign blood has entered the new system, CVTD begins to take form within the organism. Since the cancer often originates in the genitals of canines that are sexually active, common methods of transmission include sexual activity, licking or sniffing[1]. These methods of transfer are all natural actions of a common canine, explaining why the disease can spread so rapidly across a population.

CVTD is only linked to mortality when the host is a newborn or has an impaired immune system.[2] In a properly developed canine, CVTD will experience a period of growth within the affected region of the organism, followed by a stage of equilibrium where the tumour does not experience growth. This is followed by a stage of regression, which involves the death of cells that comprise the mass. Regression is assisted by a mass flow of lymphocytes to the cancerous region, which kills the cancer cells.[2] There is no defined timeline for the stages of this cancer's existence as each stage varies between individuals based on the genetic background, as well as the strength of the individual's immune system.[2]

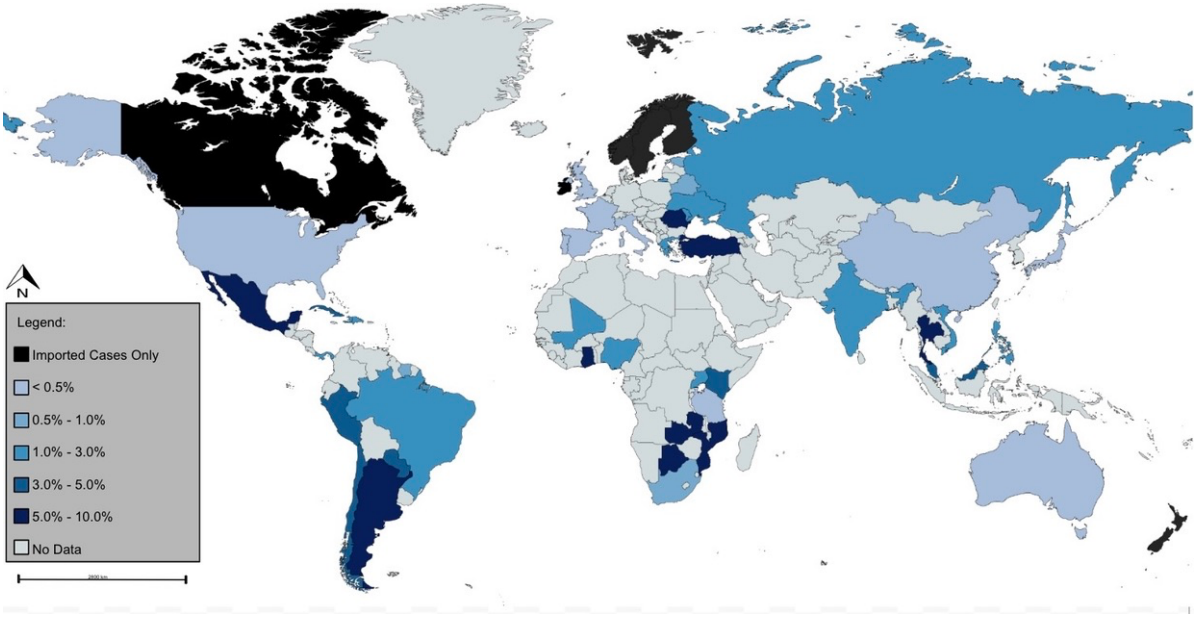
CVTD is first recognizable as small lumps, equating to roughly 1 centimetre in diameter. The tumours will grow to possess multiple lobes and have a total diameter of up to 10 centimetres.[2] These tumours commonly form and grow alongside connective tissues and blood vessels.[2] In terms of biological trends, this cancer does not favour sex in its development. As seen in Figure 1B, current data suggests an even spread in the development of CVTD across sex. The canine's health also has no notable link to the development of this cancer form, as 72% of canines with this cancer are otherwise healthy individuals (Figure 1C).[1]

GIS Application

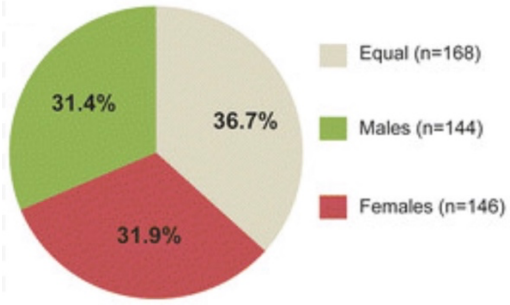
CVTD affects canines throughout the world but is more prevalent in impoverished countries (Figure 1A). This connection makes sense as developing countries have less medical funding and technology available in comparison to nations with high GDPs. Insufficient medical resources would be a reason that more dogs would attain this transferable cancer. Additionally, in many developing countries, canines are seen as wild animals rather than domestic counterparts. For this reason, it is unlikely that the health of a dog would be a priority for people in a developing country.

Canine health is heavily related to the state of humans in the said region. An additional connection to the likelihood of CVTD development in a region is how susceptible countries are to natural disasters. Such disasters include earthquakes, floods, tsunamis, landslides, volcanic eruptions, hurricanes, etc. Many countries highlighted as high to very high in terms of their disaster risk (Figure 1D) also have high rates of CVTD. From a sociological perspective, this makes sense, as people experiencing a natural disaster would be less likely to focus on the health of a canine.

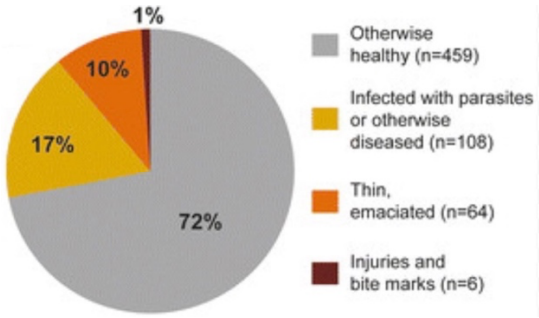
A)


N
Legend:
Imported Cases Only
< 0.5%
0.5% - 1.0%
1.0% - 3.0%
3.0% - 5.0%
5.0% - 10.0%
No Data

B)


31.4%
36.7%
31.9%
Equal (n=168)
Males (n=144)
Females (n=146)

C)


1%
10%
17%
72%
Otherwise healthy (n=459)
Infected with parasites or otherwise diseased (n=108)
Thin, emaciated (n=64)
Injuries and bite marks (n=6)

D)

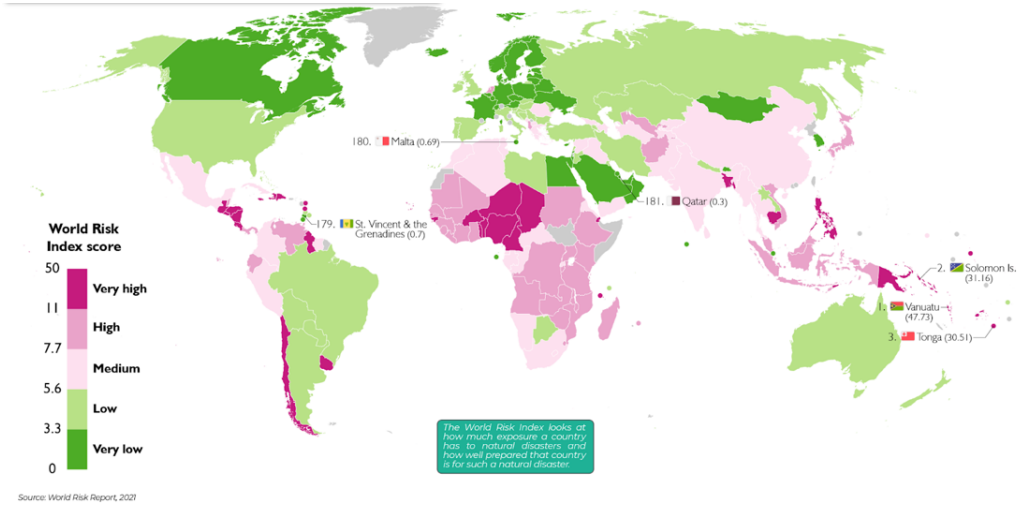

World Risk Index score
50
11
7.7
5.6
3.3
0
Very high
High
Medium
Low
Very low
180. Malta (0.69)
181. Qatar (0.3)
179. St. Vincent & the Grenadines (0.7)
2. Solomon Is. (31.16)
1. Vanuatu (47.73)
3. Tonga (30.51)
The World Risk Index looks at how much exposure a country has to natural disasters and how well prepared that country is for such a natural disaster.
Source: World Risk Report, 2021

Figure 1*: (A) This map made by us displays the prevalence of CVTD in each country, ranging from 0% (imported cases only) to 10%.*[1] *We used data found from multiple sources combined with canine population records to create this mapped data.*[3] *(B) This figure displays the relation between developing CVTD and the sex of canines. The data was collected from various respondents who stated whether they saw the cancer form more in males, females of if it was equal. The data displayed shows no trend in terms of sex.*[1] *(C) In the same study as Figure 1B, respondents were surveyed on the overall health of dogs diagnosed with CVTD. Data collected states that 72% of canines with CVTD were healthy individuals meaning there is no link between the health of the canine and the development of CVTD.*[1] *(D) a map that displays the risk index score of each country. A high score relates to a high level of natural disasters occurring. A low score relates to a low number of natural disasters occurring.*[4] *Notably, many countries with a high-risk index scores have a high prevalence of CVTD.*[5]

Chemotherapeutic Background

Contrary to radiation and surgical treatments that target a specific cancer-induced site in a patient's body, chemotherapeutics control the progression and/or kill cancer cells throughout the body without a specific anatomical target site.[6] The most common type of chemotherapeutics are alkylating agents; these compounds react with electron-rich atoms that occur in biological species to form covalent bonds.[7] The chemotherapeutic that will be used in our study is the vincristine (*Figure 2*) compound: Vincasar PFS®. Vincristine is a type of vinca alkaloid which is obtained from the Madagascar periwinkle plant, and is the second most used drug to treat cancer.[8] Other types of vinca alkaloids that have properties similar to vincristine include vinblastine, vinorelbine, and vindesine.

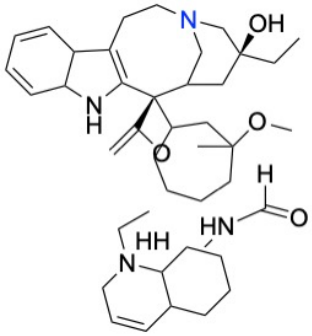


Figure 2: *Vincristine chemical structure C46H56N4O10. H2SO4 (created by authors).*[9]

Vinca alkaloids work via the interaction and disruption of microtubule function. They do this by connecting to one of 17 high-affinity binding sites on the tubulin. When the vincristine binds to the desired site, it causes an interruption in the microtubule congregation (used in cell structural support, pathway for transport, and generates force in cell division). Additionally, this process decreases the growth rate and shortens the assembly at the end of the microtubule. This causes the production of a kinetic cap that suppresses function. The location that is primarily affected in the cell from this is the miotic spindle, as its interference with the miotic spindle results in meta-phase arrest, which is the blockage of the pathway that signals the transition from metaphase to anaphase.[10] This is important as they comprise the cytoskeleton. The cytoskeleton is comprised of three main proteins: actin filaments, intermediate filaments, and microtubules. The vincristine specifically targets the microtubules in a cell. A microtubule is comprised of two types of proteins: alpha-tubulin and beta-tubulin.[10] These combine to form long proliferations of alternating proteins known as protofilaments. Thirteen of these long proliferation chains come together to form a singular microtubule.[10]

These microtubules are essential for cell division as they are the building blocks for centrosomes used in metaphase. During metaphase, chromosomes align in the middle of the cell, the chromosomes consist of a pair of sister chromatids that have a centrosome at its center. Once every component of the cell is aligned the centrosome sends out spindle fibers towards the spindle poles that are built from microtubules.

Vincristine binds to the heterodimer tubulin subunits alpha and beta of microtubules. These are responsible for the elongation of the microtubules. Thus, when vincristine binds to the microtubules it interferes with the addition of new tubulin subunits proliferating into the protofilaments. The interference with microtubule binding causes the cell's chromosomes to fail to align and separate properly, as the microtubules form the mitotic spindles that are used to segregate chromosomes into daughter cells, resulting in metaphase arrest.[10] When metaphase arrest has occurred for a prolonged period of time, it decreases cell proliferation from the incompletion of mitosis, triggering the cell's response for apoptosis. In cancerous cells, the cells are replicating at a greater rate than healthy cells. Thus, the microtubules are always functioning, causing an overall decrease in CVTD as the cancer cells will be unable to replicate.

As vincristine affects mitosis in cells that are rapidly dividing throughout the body, other areas that have a high affinity for the targeting site of vincristine include gastronomical epithelium, and hair follicles. In human patients, it additionally causes neurotoxicity, most commonly in the form of peripheral neuropathy; this is a result of the disruptions of the microtubule polymerization in neural axons that can impact motor, sensory and autonomic nerves. This causes the body to experience a loss of feeling in the fingers and toes, orthostatic hypotension, and constipation. In canines, common side effects include vomiting, anorexia, and diarrhea due to the strong affinity of vincristine to target the gastrointestinal system. Hematological side effects are also common, including leucopenia, neutropenia and thrombocytopenia.[11]

Vincasar PFS® is most commonly administered via intravenous injections; this results in the chemotherapeutic agent targeting all cells in the body, not only the cancerous ones, but Vincasar PFS® has the highest affinity for the cancerous ones. This means that it does not need to be directly injected into the cancerous region. The dosage, duration, and quantity are case-dependent; however, it is most typically administered at 0.025 mg/kg once a week for a month, that is diluted in 250 mL of normal saline, and administered slowly over half an hour.

Motivations & Study Approach

Motivations

The motivation for this project was cemented on the premise of a three-pillar informed and sustainable perspective. Firstly, the desire to minimize human impact by capitalizing on already existing conditions, i.e., using canines already infected with CVTD and already susceptible to certain environmental carcinogens similar to humans. Furthermore, the acknowledgment of and optimization of CVTD as a forgiving and curable cancer for experimental treatment.[12] The unique characteristic of immunity to reinfection and complete regression of CVTD (allowing for a novel life after diagnosis) allows scientists to experimentally administer a variety of dose concentrations (from null/subtherapeutic to excess/toxic), which may be unethical on their human counterparts for the sake of learning.[12] Lastly, the encouragement of pro-drug thinking for potential human applications. Prodrugs are compounds with optimal physiochemical properties and decreased toxicity with great specificity. Currently, only 15 prodrugs/liposomes are FDA approved, with vincristine liposomes being FDA approved for human use in 2012,[13] but only as a third line of treatment.[14] The continual development of prodrugs/liposomes in lieu of mitotic inhibitors is important to reduce side effects on healthy tissues.[15] The execution of study 2 will rely on the homeostatic and genomic similarities between *Canis lupus famillaris* and *Homo sapiens* to inform the possible displacement of vincristine liposome (and other biochemically similar drugs) as an earlier-line drug in human cancers. This will be substantiated using an MRI monitor.

Study 1 Questions

1. Given a dataset with 207 genes, which genes are most positively associated with a decrease in tumour proportion?
2. How can each of the genes' chemical fingerprints influence their inhibitory capacity?
3. How can we predict the inhibitory capacity of the tumour (using a multi-linear regression model) to determine a treatment suitable for a canine? How is this method transferable to human cancer?

Study 2 Questions

1. How effective is vincristine in treating CVTD, and what is the minimum effective dose, i.e. PCI50?
2. How well does the equation derived from the dataset in Study 1 predict experimental results of drugs not included in the original data set (assess predictive power in novel situations)?
3. How does Vincristine compare to a typical (i.e. the unknown) chemotherapeutic treatment course?
4. How significantly does the genotype of the tumour impact the efficacy of vincristine?

Study 1 and 2 Methodological Graphical Abstract

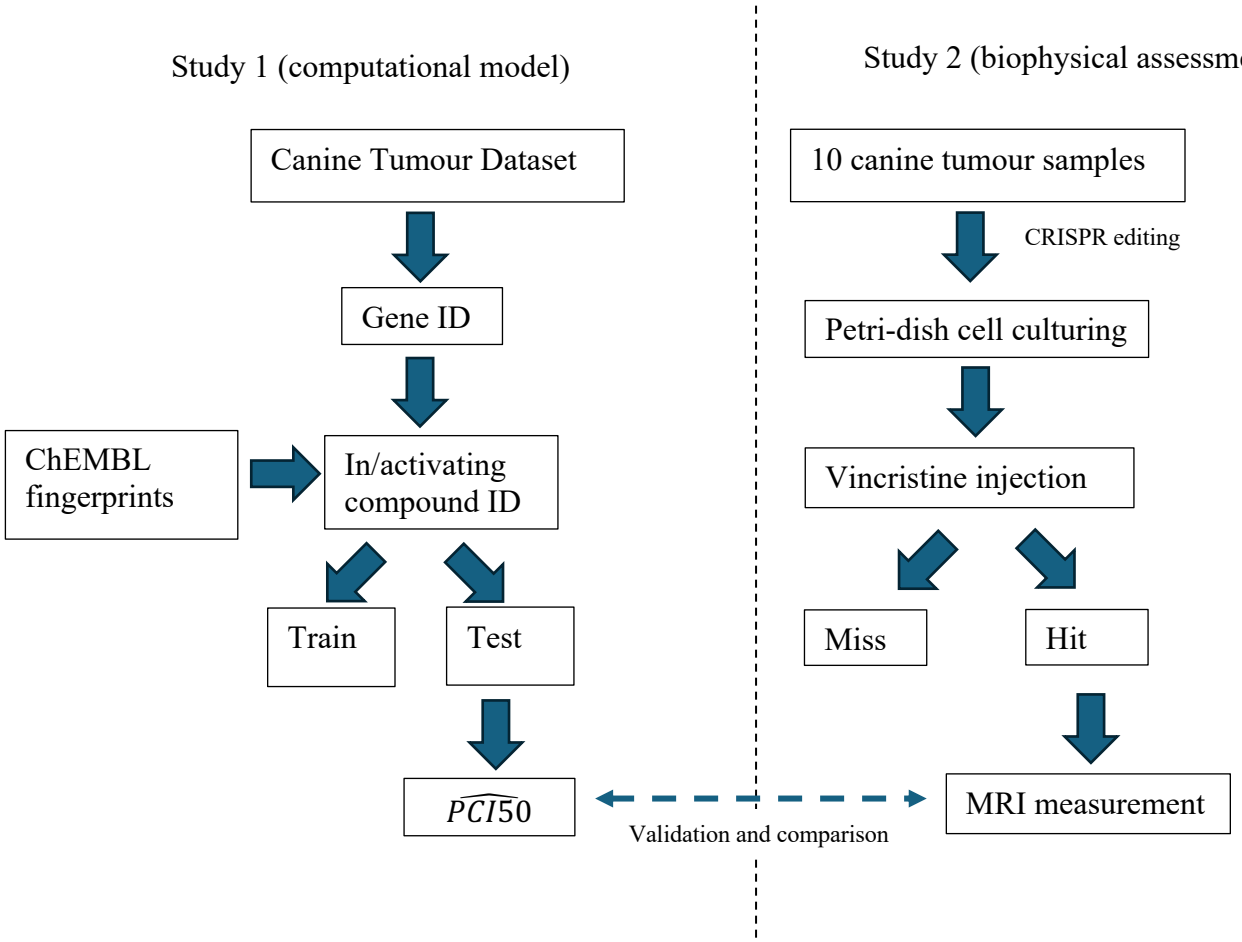


***Figure 3**: The methodology demonstrates the integration of Study One and study Two to answer the proposed research questions.*

Study 1 Methods

Study one's methodology can be categorized into three primary sections. Firstly, data collection, followed by data analysis and finally the development of a predictive regression model using an ordinary least square (OLS) package.

Data Collection

Canine whole exome sequencing (WES) and whole genome sequencing (WGS) data were downloaded from Alsaihati et. al, 2021.[16] Data cleaning was performed in two ways. Firstly, genes associated with a mutation type of "pathway" and a corresponding gene of "cell cycle" were dropped from the dataset because the chemical compounds associated with the inhibition of the cell cycle are mitotic inhibitors, which have intolerable toxicities on prolific normal tissues.[15] Rather than using mitotic inhibitors, cancer-specificity can be achieved by the development of pro-drugs (compounds with optimal physiochemical properties with decreased toxicity)[17]; to be explored in this project. Furthermore, data wherein the $Gene_1$ was not the same as $Gene_n$ were removed to simplify the model. This comprised only three of 207 genes in the dataset. The canine WES and WGS data were assumed to use the targeted panel sequencing of 207 genes from one individual canine infected with CVTD, who received six chemotherapeutic administrations of one chemically unknown chemotherapeutic drug.

Data Analysis

The tumour proportion for each of the remaining 204 genes was calculated at the end of six chemotherapeutic trials and fitted with a polynomial regression model. The numerical data after six chemotherapeutic trials was assumed to consider all the tumour's heterogenous subclones and return a weighted final proportion. This assumption could be validated by performing several biopsies. The ten genes with the lowest tumour proportion after six chemotherapeutic administrations were noted, and associated compounds and their activities were compiled into a secondary dataset downloaded from the European Bioinformatics Institute (a subsidiary of the European Molecular Biology Laboratory).[18] The dataset downloaded included 153 relevant

compounds and 23 chemical fingerprints. The correlation between all 23 chemical fingerprints (CFPs) was computed. Nine of the 22 CFPs had a strong correlation with the pChEMBL value, i.e., the PCI50 value scaled by a constant value $k = 9$.[18] The PCI50 value represents the half-maximal inhibitory concentration of an exogenous substance to inhibit tumour growth.[19]

Predictive Modeling

A traditional linear regression model explores the relationship between two or more variables by means of developing a first order equation to predict an outcome. Multi-linear regression models are of the form the predicted value equals the coefficient for one independent variable (β_n) multiplied by that independent variable (X_n), plus a y-intercept (β_0), $for\ n\ terms,$ plus an error coefficient (ε), as seen in Equation 1.

$$y = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n + \varepsilon \quad (1)$$

The CFP dataset was split into 80% training and 20% testing data. This 80/20 split is based on the Pareto Principle, which states that approximately 80% of all effects come from approximately 20% of the causes.[20] Using the nine CFPs, a multi-linear regression model was trained to remove outliers not meeting the inter-quartile range and assessed using the R squared value, adjusted R squared score, standard deviation, and F-statistics. A similar analysis was performed for the testing data. It is necessary to evaluate the model from many perspectives, using various metrics to avoid over-relying on one metric, e.g., R^2, which may lead to overfitting.

Study 1 Results

Analyzing the WES and WSG data, the ten genes associated with the lowest proportion of tumour cells after the six consecutive chemotherapeutic administrations were selected. These genes included PIK3CA, CNTN6, SAV1, SFRP2, MDM2, NOTCH1, PIK3CB, HRAS CDKN2A and SFRP1. Using an external database, the CFPs of 153 compounds associated with the genes aforesaid were extracted and analyzed.

Table 1: *Of the ten genes most positively associated with a decrease in tumour proportion over six chemotherapeutic trials, seven were associated with one or more chemical compounds. A brief description for each of the seven genes is provided, along with the number of associated compounds.*[18]

Gene	Description	Number of associated compounds
PIK3CA	PIK3CA is a protein coding gene that codes for p110 alpha. This p110 alpha is a subunit for the enzyme Phosphatidylinositol 3-kinase (PI3K). PI3K phosphorylates signaling molecules that transmit chemical messages throughout the cell. PI3K also plays a large role in cell proliferation, migration, and production of new proteins. When PI3KCA is mutated, the signaling and coding of PI3K is increased, downstream signaling such as AKT is stimulated (protein kinase B).[21]	2
SAV1	SAV1 plays an important role in the regulation of a wide variety of cellular functions such as protein degradation, transcription, and RNA splicing. From the suppression of tumour growth in colorectal cancer, SAV1 is known to enhance the effects of Hippo kinase in the Hippo pathway.[22]	126
MDM2	MDM2 is both an oncogene and a tumour suppressor simultaneously, it is a critical negative regulation of p53 protein which acts as a tumour suppressor. It decreases protein levels and	4

	function by controlling its activity of transcriptional processing, protein stabilization and nuclear localization.[23]	
NOTCH1	NOTCH1 is connected to various signaling pathways throughout the body. This causes the gene to be affected by various forms of cancer throughout the body such as mammary cancer, breast cancer and leukemia. Note that there are clinical implications of NOTCH1. Specific cancers have high levels of NOTCH1. Targeting this gene is a strategy that can be implicated when fighting certain cancers. Additionally, increasing the concentration of NOTCH1 in a region can help to combat cancer as the presence of the gene can increase the effects of chemotherapy on the tumour.[24]	1
PIK3CB	PIK3CB assists in the growth of cancer throughout the body. This gene is able to regulate cell adhesion to collagen I.[25]	1
HRAS	HRAS is a gene found in canines that can affect the likelihood of developing cancers if said gene is mutated. Whole exome sequencing of canine acanthomatous ameloblastoma has identified a pattern between HRAS mutations and the development of this cancer.[26]	4
CDKN2A	The purpose of the CDKN2A gene is to generate instructions to certain cells to develop proteins. Certain proteins linked to this gene are INK4A and ARF. Both proteins work as tumour suppressors and ensure that cell growth does not become uncontrolled.[27]	15

Exploring the correlation between all 22 CFPs and pChEMBL, nine CFPs were selected for further analysis (Figure 2). The CFPs of interest include molecular weight, AlogP (the lipophilicity of a molecule as the partition coefficient of a solute between octanol and water), the number of rotatable bonds, the CX LogP (the octanal/water partition coefficient) and CX LogD (the octanal/water partition coefficient at pH = 7.4), the inorganic flag (a binary marker to determine if the chemical compound has any carbon to hydrogen bonds), the number of heavy atoms (the number of non-hydrogen atoms in the molecule), the monoisotopic molecular weight

(the sum of masses of most abundant isotopes in the compound),[18] and the NP likeness score (the similarity of a molecule to a natural i.e., an endogenous secondary metabolite).[28]

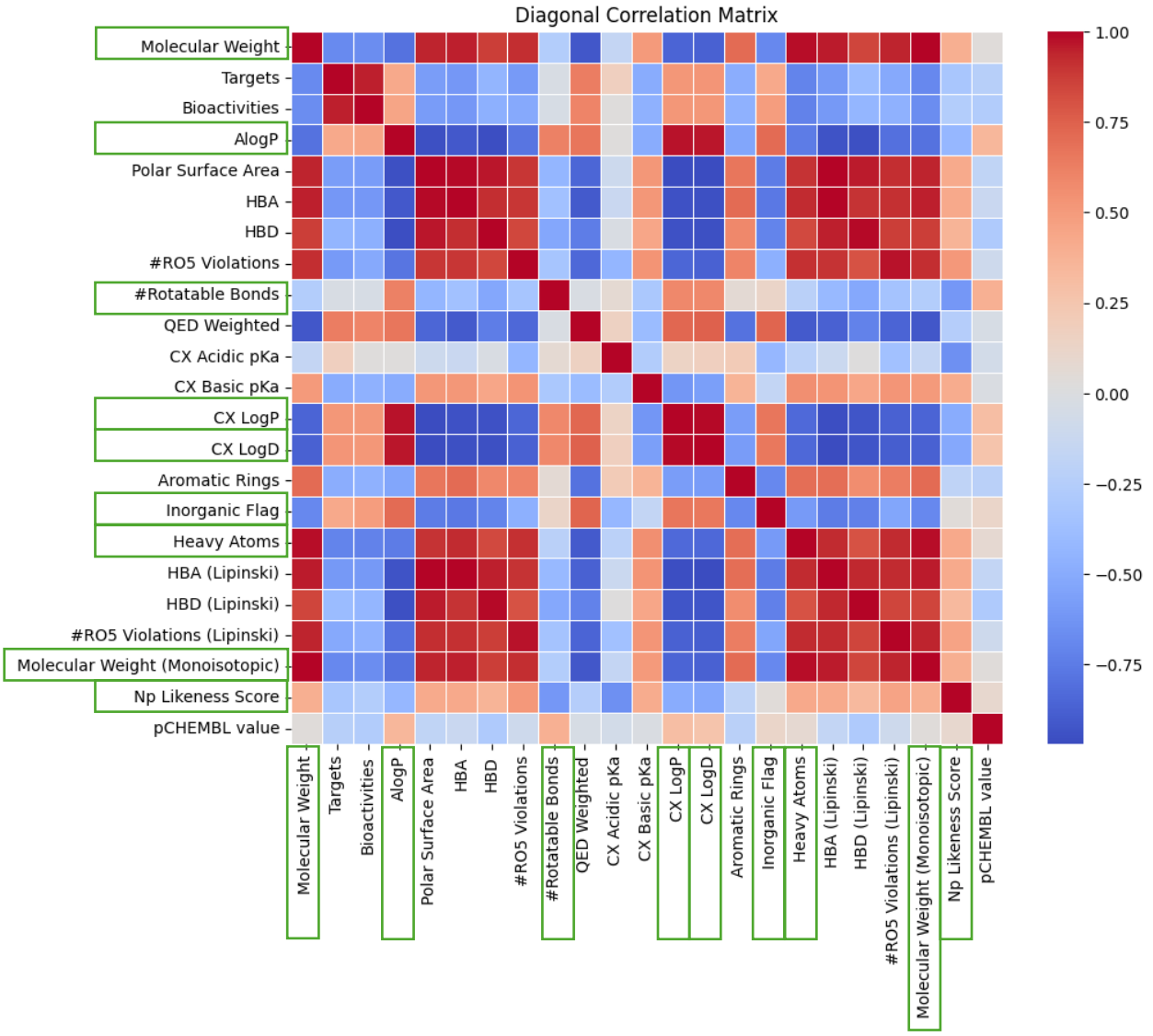


Figure 4*: The matrix heatmap for the relationship between all 23 CFPs. Nine of the 23 CFPs with the greatest correlation to the pChEMBL value (a PCI50 value scaled by K=9) were selected to create a multi-linear regression model; illustrated by the outlining green boxes. Heatmap made using the Pearson's correlation coefficient; a value closer to 1.00 showing a strong correlation, a value closer to -1.00 representing a negative correlation, and a value around 0.00 representing no correlation between independent variables.*

Using the ordinary least square (OLS) package from statsmodel.api, a multi-linear regression model was computed.[29] The OLS model minimizes the sum of squared errors between the predicted and actual output. The constant value ($\hat{\beta}_0$) is computed by subtracting the coefficient of the first independent variable multiplied by the mean of the first independent variable from the mean of the dependent variable (y, i.e., pChEMBL), as seen in equation 2.[30]

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{2}$$

Solving for the coefficient (slope) of each of the nine independent variables requires further algebraic expansion. Using molecular weight as $\hat{\beta}_1$, i.e., the slope coefficient for molecular weight can be computed by the summation (from i = 1 to N = 20 in the case of our dataset) of the compound's molecular weight minus the mean molecular weight multiplied by pChEMBL value for the i^{th} chemical compound minus the mean pChEMBL for all of the iterations. These are divided by the summation from i = 1 to N = 20 of the i^{th} chemical compounds molecular weight minus the mean molecular weight of all compounds from the dataset, all squared, as seen in equation 3.[30]

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N=20}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N=20}(x_i - \bar{x})^2} \tag{3}$$

Table 2A. *Python OLS output summary for the nine CFPs. The coefficients were extracted and used in the pChEMBL i.e., PCI50 value scaled. The standard error representing the standard deviation for a sample population. A T-value average above 2.00 portraying a high confidence rating of the coefficient as a predictor. P-value less than 0.05 demonstrating a statistically significant correlation between the variable and PCI50 value. 95% upper and lower confidence boundaries shown, i.e., the OLS is 95% confident that the coefficient falls within this range.*

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4.2460	0.7016	6.0519	0.0005	2.5870	5.9049
Molecular Weight	-1.5193	0.2827	-5.3745	0.0010	-2.1877	-0.8508
AlogP	0.3534	0.0820	4.3069	0.0035	0.1594	0.5474
#Rotatable Bonds	0.0856	0.1031	0.8307	0.4336	-0.1581	0.3294
CX LogP	0.1995	0.0721	2.7668	0.0278	0.0290	0.3700
CX LogD	-0.2731	0.0520	-5.2489	0.0012	-0.3962	-0.1501
Inorganic Flag	-0.1965	0.2892	-0.6793	0.5188	-0.8803	0.4874
Heavy Atoms	-0.3139	0.0732	-4.2855	0.0036	-0.4871	-0.1407
Molecular Weight (Monoisotopic)	1.5402	0.2867	5.3717	0.0010	0.8622	2.2182
Np Likeness Score	-0.2201	0.1390	-1.5831	0.1574	-0.5489	0.1087

$$C(pChEMBL) = -1.52(MW) + 0.35(AlogP) + 0.09(\#RB) + 0.20(CXlogP) - 0.27(CXlogD) - 0.19(IF) - 0.31(HA) + 1.54(MWM) - 0.22(NP) + 4.25 \quad (4)$$

Using the coefficients from the OLS multi-linear regression model, a nine-variable equation was computed, using testing and training data, and is summarized in Table 2B.

Table 2B. *Regression Statistics for testing and training data of the pChEMBL dataset.*

Metric	Train	Test
R Square	0.9303	0.8930
Adjusted R Square	0.8406	0.5740
Mean Squared Error (MSE)	0.0087	1.5684

To simply visualize the accuracy of the multi-linear regression model, the actual pChEMBL, i.e., scaled PCI50 values and the predicted PCI50 values were plotted. The absolute difference in pChEMBL between the true and precited values is the residual, which can be represented using the error term. Outliers were determined using the inter-quartile range. Outliers are pictorially observed to have large absolute residuals. The absolute residual of the model can be fine-tuned by adjusting the number of CFPs used in the equation for pChEMBL (equation 4). For example, using all 22 CFPs in relation to pChEMBL yields an R squared value of 1.00, thus decreasing the residuals vector, whereas using only the most correlated CFP (i.e., molecular weight), the R^2 value is only 0.059, and the adjusted R^2 value is -0.003 (underfitting). Finding a middle ground relative to the number of CFPs used was important for the development of our model to avoid overfitting. Overfitting occurs when the model is so closely fitted to the training data that the model is unable to make accurate predictions from any data other than the training data.[31] Using a more comprehensive database would increase the number of training and testing iterations, which would improve the accuracy of the network. Furthermore, the breadth of the CFP database was limited to those recognized by the European Bioinformatics Institute; by expanding the database to include other exogenous compounds, it is possible to observe novel trends and patterns.

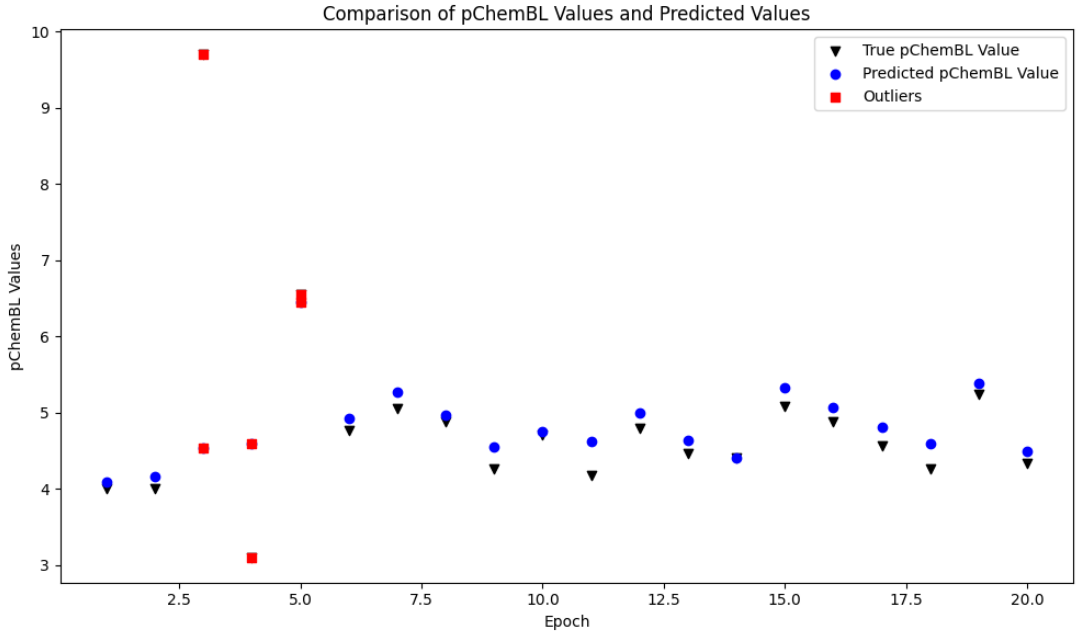


Figure 5*: A scatterplot of the predicted (blue) and actual (black) PCI50 values for 21 different epochs. Outlying epochs are shown in red but did not contribute to the regression statistics.*

Study 2 Motivations

Summary Of Relevant Study One Outcomes

In Study 1, data on the tumour proportion remaining after treatment with one chemotherapeutic drug was recorded and analyzed using regression analysis in Python. By observing the most effective trials in the dataset, in terms of the lowest tumour proportion remaining after trial, we determined which genetic mutations had the greatest effect on the tumour and, thus, which genes to target with analysis. The data from 7 of these 207 specific genes was selected for further investigation. Subsequently, using chemical data from the ChEMBL database for compounds associated with these specific genes, linear regression was applied to find how each of the 27 chemical fingerprints for a given compound relates to its pChEMBL value, which is equivalent to the scaled pCI50: the negative log of the compound concentration at which the target biological activity is decreased by 50%. Based on this analysis, we obtained a multivariable linear equation, observing 9 specific chemical fingerprints which, together, strongly predicted the pChEMBL, or efficacy, of a given compound. This equation was developed computationally, not experimentally, from the dataset in Study 1, based on those genes which were most effective in reducing tumour proportion in order to uncover which chemical features of a compound have the greatest therapeutic impact in the treatment of CVTD.

Our Goals And Motivations For Study Two

Study 2 will be a biophysical wet-laboratory study of the efficacy of the chemotherapeutic drug Vincasar PFS® as a treatment for Canine Venereal Tumour Disease (CVTD). To elucidate the efficacy of the drug, three different dosages of Vincasar PFS® will be tested in vitro on 10 configurations of CVTD tumour cell culture. Specifically, the 10 configurations will consist of 3 unmodified cell cultures, to act as a control, and the 7 genotypes identified as the most responsive to the chemotherapeutic from the original dataset used in Study 1. Based on a similar study conducted by Haque and Vaiselbuh wherein they tested Vincasar PFS® in different dosages on leukemia cells, in a 96-well plate, an ideal dosage of Vincasar PFS® to administer to each cell culture ranges between 0.1 and 4.0 μM.[32] Thus, we will administer 0.5, 1.0, and 1.5 μM

doses in this specific study. Motivated by our Study 1 results, we will observe the effect of each of the three dosages of Vincasar PFS® on the proportion of tumour cells remaining after 24 hours of treatment, using MRI and MRS measurements of tumour cell proportion, with three trials of each dosage, for each of the 10 cell cultures. Thus, at the end of the study, we will have tested the Vincasar PFS® on a total of 90 different cell cultures. Prior to MRI measurement, cell cultures will be sorted to exclude those wherein the Vincasar PFS® had minimal, or no effect, which can result due to cell cultures with immunity to the drug. This is vital since immunity to Vincasar PFS® occurs in approximately 10% of canines with CVTD in the wild, and each of our cell cultures will be derived from a different afflicted canine.[33] This pre-measurement sorting will be done by performing a cytotoxicity assay on 90 wells in the 90-well plate and choosing only cell cultures in which there is an indication that a significant number of cells were damaged or died (Figure 7). Only cell cultures which experience a "hit" will be transferred to an MRI machine for further analysis. We will then collect data on the proportion of tumour cells remaining in each of these cell cultures after treatment. The data we collect will then be analyzed to determine the half-maximal inhibitory concentration, or PCI50 value, which is the concentration of a drug or other bioactive compound at which the target biological function has its activity reduced by 50%. In this case, we will approximate the inhibition of the target function's activity as the rate of cell death, or the reduction in tumour proportion as observed with our biophysical tools. The PCI50 values will be determined by plotting the concentration of the drug, or dosage, against the proportion of tumour cells remaining for each of the cultures. We will use curve-sketching software to find the IC50 for each cell culture line, based on up to 9 samples for each (3 trials for each dosage). We will subsequently calculate the PCI50 for each genotype and compare it to the predicted PCI50/pChEMBL value that we would obtain based on the equation formulated in Study 1, from the original dataset. This serves two purposes: firstly, we determine the effective dosage for half-maximal inhibition of each genotype of the tumour with Vincasar PFS® by finding the PCI50; and secondly, we validate our Study 1 results by demonstrating that the equation for predicted pChEMBL is generalizable to novel therapeutic treatments. These results will also be compared against the unidentified chemotherapeutic used in the original data, which was analyzed in Study 1 and, from these results, draw conclusions as to the efficacy of Vincasar PFS® relative to another chemotherapeutic in treating CVTD.

Human Connection

In relation to human-centric applications of the use of vincristine to treat various kinds of cancer, we can use the data we generated from our model of the canines to compare its efficacy to what we would expect from human cancers. This can be used as an accurate comparison due to the mechanism of action of Vincasar PFS®. The efficacy of vincristine binding to the β Tubuli regulates the efficiency that vincristine has on the tumour regression. The structure of the catharanthine alkaloid, a part of the binding sector of the vincristine, is integral in this efficiency as catharanthine acts as a cAMP inducer.[34] As a cAMP inducer, it suppresses a particular cell type: HepG2. HepG2 is an epithelial-like cell that regulates cell proliferation in regards to the metabolism of drugs.[35] This subsequently reduces the transcription factor of Nuclear factor erythroid 2-related factor 2 (NRF2), which regulates the defence mechanism of the expression of genes that are involved in drug detoxification.[36]

Thus, the efficiency of the binding results in a response that affects the rate of cancer cell proliferation, survival, and migration and promotes immune evasion by reprogramming the emerging tumour microenvironment. As the mechanism in which vincristine acts is the same for humans and canines, we can compare the two, to determine how well Vincasar PFS® is suited to treat canines from human models, and how well it is suited to treat humans via canine models. This can be done by comparing experiments that have already been trialled to generate representative data that will give an idea of how well vincristine works on humans in comparison to canines. From comparing these experiments, we are given a baseline on how efficient it is in standardized dosages that we can then compare to our dosages to see how well our data would hypothetically human trials. The data that has been collected from multiple trials has the overall 5-year survival rates from the following cancers when treated with vincristine: lymphoblastic leukemia, neuroblastoma, Wilms' tumour, and CVTD. The overall survival percentage for each cancer was 85%,[37] 76.26%,[38] 90%,[39] 92%[37] respectively. These values were then plotted in coding software R to analyze the similarities between the survival rates of human cancer and CVTD *(Figure 6)*.

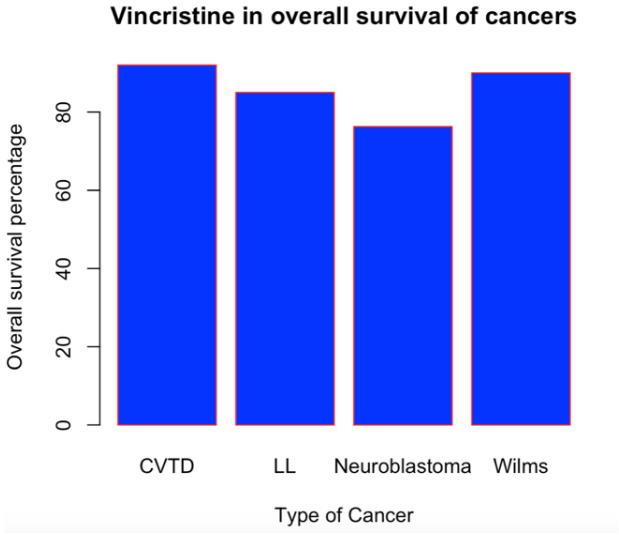


Figure 6*: Data that has been taken from four unique trials with similar dosages for three different human cancers: lymphoblastic leukemia (LL), Neuroblastoma, Wilms' Tumour (Wilms), and CVTD. This gives insight into the upside that treating human cancers with vincristine has; as this models the similarities between each cancer, they allude that the binding efficiency would be similar for each, especially for Wilms' tumour. With this knowledge scientists can use CVTD in canines as a model for Wilms' tumour, with the potential for further testing of new synthesizes based on the canines' responses (created by authors).*

Research Questions

To summarize our goals, we seek to answer four key research questions in Study 2:

1. Firstly, how effective is Vincasar PFS® in the treatment of CVTD, and what is the minimum effective dosage? In other words, what is the half-maximal inhibitory concentration (IC50), or the concentration of drug required to inhibit the process of tumour growth by half?
2. How well does the equation derived from the dataset in Study 1 predict the experimental PCI50 / pChEMBL for novel therapeutics?
3. How does Vincasar PFS® as a treatment for CVTD compare to the unidentified chemotherapeutic used in Study 1's dataset? This will be identified by comparing the raw tumour proportion data we obtained for the 7 tested genotypes to the tumour proportion data for these genotypes int he original dataset.
4. How significantly does the genotype of the tumour impact the efficacy of Vincasar PFS® as a treatment? We can answer this by simple comparison: which genotypes were the most responsive to the treatment? Is there a significant difference in the experimental tumour proportion data between each of the different genotypes, or between a given genotype and the control(s)?

Study 2 Methods

CRISPR For Modification Of Certain Tumour Cell Cultures

We will be using CRISPR and cell culture to create and plate petri dish samples of the seven most impactful tumour genotypes we found in our Study 1 analysis. We will begin by taking samples of tumour cells from 10 different canines afflicted with CVTD and culturing those cells in 9 different cell culture mediums for later testing. In order to activate and inactivate the seven key genes to obtain each genotype from the original dataset, we will make use of the CRISPR/dCas9 system. CRISPR, which stands for Clustered Regularly Interspaced Short Palindromic Repeats, refers to repeating sequences of DNA associated with Cas9, a protein found in prokaryotes which is used for adaptive immunity by recognizing and cleaving

bacteriophage DNA.[40] This tool has been adapted by researchers for genomic editing which encompasses both clinical-therapeutic and basic research purposes. The CRISPR/dCas9 system we will be using is derived from CRISPR/Cas9, but instead consists of a catalytically inactive, or "dead" Cas9 (dCas9) protein complex, which is unable to cleave DNA; a single guide RNA (sgRNA) that is complementary to the promoter region of the target gene and, thus, "guides" the dCas9 to the gene; and, depending on the form, either a transcriptional activator (CRISPRa) or transcriptional repressor/interference factor (CRISPRi).[41] The dCas9 – sgRNA – transcriptional effector complex binds to the promoter sequence of the target gene, with the target gene either being overexpressed due to the presence of CRISPRa, the prevention of gene expression due to interference from CRISPRi.[42] This technology improves functional genomics by reducing cost and simplifying the process of targeted gene activation and gene repression.[41] After culturing 9 cell culture mediums from each tumour cell sample, we will use this technology to selectively activate or inactivate specific genes in seven of the ten cell culture lines, creating nine tumour cell cultures of each genotype. The effect of the subsequent administration of a specific dose of Vincasar PFS® to these in vitro tumour cell cultures will be observed with in vitro cytotoxicity assays. The cytotoxic activity of the tumour cells will be qualitatively observed, as described below.

In Vitro Cytotoxicity Assay Based On LDH Detection

After administering Vincasar PFS® treatment in vitro to our tumour cell cultures, we will need a way of identifying cell cultures wherein the treatment has some effect. This can be done using so-called in vitro cytotoxicity assays, which measure the level of cell death or damage in a culture medium, either quantitatively or qualitatively. Specifically, for the purposes of this study, we will be using a lactose dehydrogenase (LDH) assay. LDH is an enzyme which is rapidly released when the plasma membrane of a cell is damaged, as a result of apoptosis (programmed cell death), necrosis, and other forms of cellular damage.[43] In the case of treatment with Vincasar PFS®, which inhibits the ability of tumour cells to replicate, the reduction in tumour proportion is a result of apoptosis of the inhibited tumour cells. Thus, an LDH assay provides an excellent way to directly observe the effects of Vincasar PFS® on the CVTD cultures and a direct approximation of the reduction in tumour proportion. When LDH is released by damaged and

necrotic cells, it promotes the oxidation of lactate into pyruvate, producing NADH as a result of the concurrent reduction of NAD^+.[44] The assay will be performed through the addition of a drop of the compound INT ($C_{19}H_{13}ClIN_5O_2$), a tetrazolium salt with a yellow hue, to each well in the 90-well plate *(Figure 7)*.

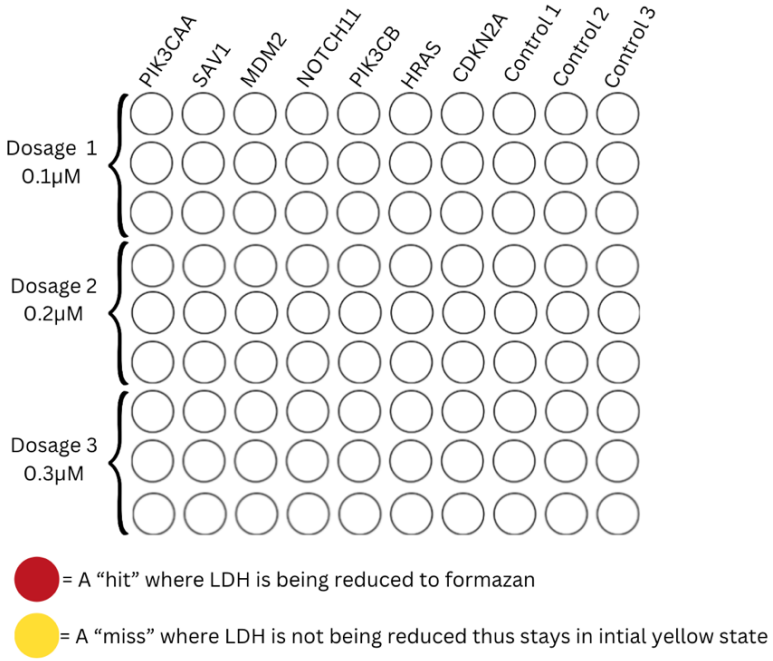


***Figure** 7: is the assay that is used to cull samples to mitigate the 10% of samples that are vincristine resistant. A hit is indicated when the well turns a crimson red as the LDH that is present in the cells is reduced into formazan, while the cells that do not experience necrosis will stay in the initial state of yellow from the INT tetrazolium salt present in the well. The three control groups are healthy canine cells, while columns one through seven are the key genes that were diagnosed in Study 1 analysis (created by authors).*

NADH reduces INT into a soluble formazan-class red dye, providing a basis for the qualitative identification of the presence of LDH in each of our cell cultures.[43] Thus, we can identify cell cultures in which the Vincasar PFS® has a somewhat significant effect on the proportion of tumour cells remaining after treatment. These cell cultures will subsequently be transferred to an MRI machine for quantitative measurement of tumour proportion, as described below.

Biophysical Interactions

In this specific experiment, studying the physics of the tumour and how the chemotherapeutic, vincristine affects different gene types by either activating or inactivating them is essential. Although not all of this data has an immediate purpose, it will soon be used to help inform further research and understanding of CVTD tumours, which can also be applied to human cancers. There are countless biophysical properties we would like to asses through our experiment, including cell mechanical states, adhesion properties, intracellular mechanics and proliferation.[45] For example, one can use a computational model to demonstrate intracellular interactions and mechanical pressure's impact on cell proliferation with respect to cancer.[45] Through our experiment, this can now be confirmed with respect to CVTD tumour analysis.

MRI/MRS Implementation

This data will be collected using Magnetic Resonance Imaging (MRI), which is needed to assess the biophysical mechanisms of cancer cells. As of the 21st century, MRI is an extremely common form of imaging found globally in countless healthcare and research institutions.[46] MRI is perfect for the analysis of the progression of the tumour, as the chemotherapeutic is administered to tumour cells of differing genotypes. MRI will provide the physical manifestation and data of the experiment's results, acting as the visual form of tumour cell analysis. Unless the tumour is separated from the canine, MRS will be avoided due to its danger to DNA segments, which are highly vulnerable in canine genital areas, where the CVTD tumour is found.[47] However, through a biopsy, we can remove the tumour (or a portion of it) and assess it through MRS, giving us detailed data on the chemical and environmental factors of the tumour, which can be used for a biophysical analysis.[45] Either way, through MRS and MRI, we can gather extremely valuable data from CVTD tumours, which can be applied to human cancers by using biophysical mechanisms.

Curve-Fitting Software For Analysis Of Tumour Cell Proportion Data

Based on MRI and MRS measurements of tumour proportion, as discussed above, we can obtain the IC50 value of Vincasar PFS® for each genotype, in addition to our unmodified controls. This is done by plotting the raw tumour proportion data for cell cultures in which we observed a "hit" against the concentration of the chemotherapeutic (i.e. the specific dosage used in that trial), and then using either curve-fitting statistical software or point-to-point analysis to determine the IC50 for the drug. We will be using curve-fitting statistical software to simplify the analysis in Study 2. Dosage is our independent variable and will thusly be plotted on the x-axis, while tumour cell proportion data is dependent and will be plotted on the y-axis. There will be 10 curves, one for each genotype and three separate curves for the unmodified tumour cell culture. Since we will be testing three unmodified cultures in each trial, plotting them separately avoids differences that will arise since each control is plated from a separate original sample, thus controlling for differences in genotype in the unmodified cell culture lines. Therefore, there will be a maximum of nine data points, since there are three trials, with three different dosages, for each culture. Less data point may be obtained than expected, depending on the proportion of "hits" or positive LDH cytotoxicity assays.

Imaging and Spectroscopy

MRI is a medical imaging test that uses magnetism and radiation to generate images of the inner human body.[46] Organs, bones, tissues, muscles, and even blood vessels and nerves can be imaged with the precision and accuracy of MRI technology.[46] MRI for humans can typically be used by sliding a patient on a tray into a cylindrical tube, which creates a magnetic field around the patient.[46] The magnetic field orients all the patient's atoms in the same direction, which are moved out of position by the induced radio waves.[46] The radio waves pass through the patient for only a moment, so when the atoms move back into their original magnetic alignment, they return a radio signal received by the computers in the cylindrical tube, which is subsequently formatted into an image or short clip.[46]

This process of inducing the magnetic field on the human body is similar to the phenomenon that occurs when an electrical charge is brought into the proximity of an insulator: polarization (*Figure 8a)*. Although no net displacement of an insulator's atoms occurs, the orientation of electrons and protons match their surrounding electric field. Similarly, MRI relies on the orientation of a patient's atoms to align in a specific direction, based on an induced magnetic field, to generate an image. However, instead of relying on the electrical properties of an atom, MRI technology relies on an atom's natural magnetic properties. Electrons have a property called *electron spin;* a term used to describe their magnetic dipoles. The direction of their magnetic dipole moment is considered North. As these electrons orbit the nucleus in a fashion similar to a current in a loop, an overall magnetic dipole is created for the atom. An MRI induces an external magnetic field to align all of these atoms and electrons by applying torque to rotate them. Every atom has a magnetic north and south pole, and the induced magnetic field in an MRI orients these atoms accordingly. Hence, every atom maximizes stability and minimizes potential energy, as displayed in Figure 8b.[48]

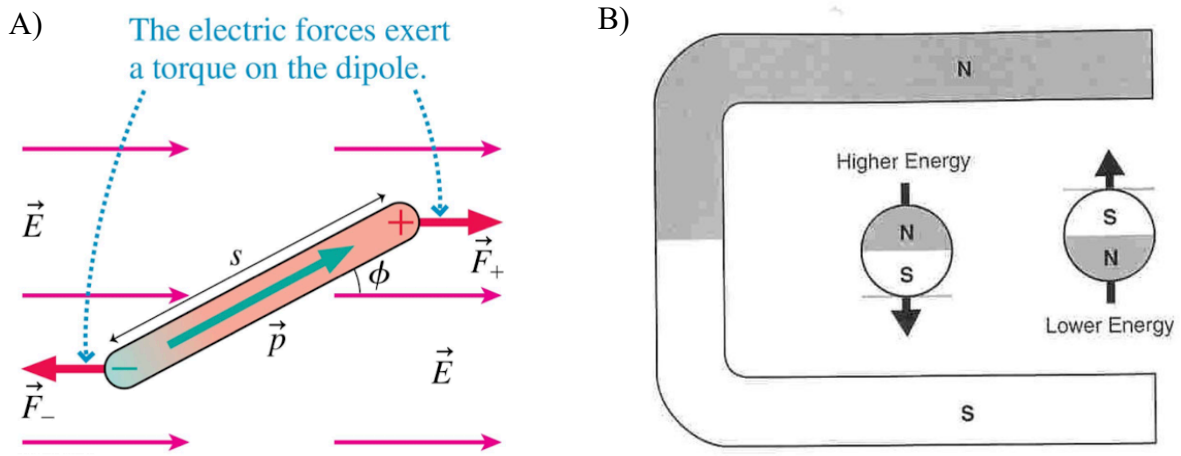


Figure 8*: On the left is Figure 8a, which shows how a diploe orients itself in the presence of an electrical field, similar to how an atom orients itself in a magnetic field. On the right is Figure 8b, which shows how an atom would minimize energy and maximize stability by orienting itself appropriately in the presence of a magnetic field.*[48]

Moreover, note the magnet used in the MRI's cylindrical tube is powered by electricity; therefore, it's an electromagnet. Specifically, a superconducting electromagnet cooled by liquid helium which helps keep the magnet at the requisite temperature.[49] It does this by decreasing the coil's total resistance to almost 0. Heat is generated and dissipated as a result of resistance, which poses a dangerous hazard when running high amounts of voltage and current through multiple coiled wires. This is mitigated through liquid helium cooling, resulting in heat being minimized, making the MRI safe for patients. However, this comes with the cost of a constant upkeep of the systems of an MRI, which can prove to be expensive and tedious.[49] An electric current is passed through the coils within the helium-filled cylindrical chamber, generating a magnetic field.[50] Equation 5 below can be used to calculate the magnetic field generated by circular coils.

$$B = \frac{\mu_0 N i}{2r} \quad (5)$$

B = Magnetic Field (Tesla)

μ_0 = *4 x* 10^{-7} *(*$\frac{Tesla \cdot Metre}{Ampere}$*)*

N = Number of turns in coil

i = Current running through coil (Amperes)

r = radius of loop formed by coils (meters)

Equation 5*: The equation above derived from Biot-Savart Law gives the magnetic field (B) for a current running through a wire that's wrapped in circles, forming a coil.*[51]

Similarly, Magnetic Resonance Spectroscopy, or MRS, is another form of magnetism and radiation-influenced imaging.[52] The MRS process and machinery are the same as an MRI.[53] However, the MRS generates images using radio waves transmitted from the atomic nuclei of a patient's body.[52] The MRS also serves a slightly different purpose than the MRI, as the imaging reflects the concentrations of certain chemicals in a certain parts of the body (e.g. neurotransmitters).[52] Nuclear Magnetic Resonance (NMR) and MRS are the exact same thing; NMR has been the term used in the physics industry for decades, whereas MRS is used in biological and medical contexts.[52]

MRI has crucial implications in healthcare and biological research, specifically cancer and imaging tumours. In case study 1, we used a chemotherapeutic on a variety of genes, and the results of whether they are effective or not on the canine tumour were used to produce preliminary (theoretical) results. In case study 2, an MRI is needed to observe the CVTD cell regression after treatment with Vincasar. As different experimentations in the lab take place regarding the bio-physical mechanisms of the tumour, its progress and changes on a cellular level can be monitored and analyzed through MRI technology.

Conclusions and Perspectives

Cancer, a 21st-century plague, claims countless lives annually. Yet, the grim reality of losing loved ones to this disease can be altered. This proposal delves into the potential findings that could revolutionize cancer treatments. As of now, there is no supported evidence that suggests that humans are able to attain CVTD from canines.[54] However, it is important to note cancer's ability to adapt to a drug, making the treatment ineffective.[55] The takeaway from this is that it is not unreasonable to extrapolate that a mutation could cause the transmissible cancer to eventually be passed on to humans. Especially given that canines and humans have a similar structure in terms of their skin as well as the regular contact between humans and canines.[56] The funding required to propel this research in Study 2 (the biophysics lab) is rooted in Study 1, where the theoretical data and derivations served as the driving force behind this research proposal.

In Study 1, linear regression was the primary methodology used to analyze 2 datasets, which provided information motivating Study 2. One chemically unknown chemotherapeutic was applied to 207 genes, with the first goal being to identify the primary genes, and proteins associated with a decrease in tumour proportion. Through a data analysis, 10 genes were discovered to have the most positive decreasing tumour proportions. Seven genes were found to have associated compounds, 156 in total. These 156 compounds comprised a secondary dataset, which tended to insight into 23 chemical fingerprints, of which nine were used to create a multi-linear regression model for predicting PCI50 with a high degree of certainty. The predicted PCI50 value allows us to explore the relationship between chemotherapeutic

potency/concentration and a decrease in tumour proportion. This theoretical data, modelling, and information motivated the consequent research proposal, which uses the funding available to pursue the biophysical lab research being proposed in Study 2.

In Study 2, our experiment aims to obtain data on the half-maximal inhibitory concentration (IC50) of Vincasar PFS® for the unmodified tumour and seven modified genotypes selected based on the Study 1 dataset. As discussed previously, this data will answer the key research questions and thus inform us regarding the efficacy of Vincasar PFS® as a treatment for CVTD. By testing Vincasar PFS® on different genotypes and comparing the raw tumour cell proportion data between the genotypes and the control, as well as with the original data from Study 1, we can also make determinations as to the impact of genotype on treatment efficacy and the relative efficacy of Vincasar PFS® compared to the unidentified chemotherapeutic from Study 1.

Ultimately, this proposal can only live up to its potential if given the attention needed, which comes in the form of funding, so what's been proposed can be pursued. Going through this biophysical lab can generate a lot of data about biophysical mechanisms, effective chemotherapeutic treatments, and their subsequent effects on different genotypes, which are all based on canine CVTD but can be applied to humans. With cancer research as a primary focus of the healthcare industry, the proposed research in study 2 is but a considerable contribution to this sector, acting as a source of information for future cancer solutions and treatments.

References

1. Strakova A, Murchison EP. The changing global distribution and prevalence of canine transmissible venereal tumour. *BMC Vet Res*. 2014;10(1):168. doi:10.1186/s12917-014-0168-9

2. Siddle HV, Kaufman J. Immunology of naturally transmissible tumours. *Immunology*. 2015;144(1):11-20. doi:10.1111/imm.12377

3. Gibson DN, Singleton DA, Brant B, Radford AD, Killick DR. Temporospatial distribution and country of origin of canine transmissible venereal tumours in the UK. *Veterinary Record*. 2021;189(12):e974. doi:10.1002/vetr.974

4. Aleksandrova M, Balasko S, Kaltenborn M, et al. *World Risk Report 2021*. https://weltrisikobericht.de/wp-content/uploads/2021/09/WorldRiskReport_2021_Online.pdf

5. Landgeist. Risk to Natural Disasters. Landgeist. Published October 19, 2021. Accessed April 4, 2024. https://landgeist.com/2021/10/19/risk-to-natural-disasters/

6. Mollaei M, Hassan ZM, Khorshidi F, Langroudi L. Chemotherapeutic drugs: Cell death- and resistance-related signaling pathways. Are they really as smart as the tumor cells? *Translational Oncology*. 2021;14(5):101056. doi:10.1016/j.tranon.2021.101056

7. Colvin M. Alkylating Agents. In: *Holland-Frei Cancer Medicine. 6th Edition*. BC Decker; 2003. Accessed April 4, 2024. https://www.ncbi.nlm.nih.gov/books/NBK12772/

8. Moudi M, Go R, Yien CYS, Nazre M. Vinca alkaloids. *Int J Prev Med*. 2013;4(11):1231-1235.

9. Škubník J, Pavlíčková VS, Ruml T, Rimpelová S. Vincristine in Combination Therapy of Cancer: Emerging Trends in Clinics. *Biology (Basel)*. 2021;10(9):849. doi:10.3390/biology10090849

10. Ray S, Kundu LM, Goswami S, et al. Metaphase arrest and delay in cell cycle kinetics of root apical meristems and mouse bone marrow cells treated with leaf aqueous extract of *Clerodendrum viscosum* Vent. *Cell Proliferation*. 2013;46(1):109-117. doi:10.1111/cpr.12011

11. Hantrakul S, Klangkaew N, Kunakornsawat S, et al. Clinical Pharmacokinetics and Effects of Vincristine Sulfate in Dogs with Transmissible Venereal Tumor (TVT). *The Journal of Veterinary Medical Science*. 2014;76(12):1549-1553. doi:10.1292/jvms.14-0180

12. Fassati A. What a dog transmissible tumor can teach us about cancer regression. *Molecular & Cellular Oncology*. 2018;5(4):e1472059. doi:10.1080/23723556.2018.1472059

13. Rahim MA, Jan N, Khan S, et al. Recent Advancements in Stimuli Responsive Drug Delivery Platforms for Active and Passive Cancer Targeting. *Cancers*. 2021;13(4):670. doi:10.3390/cancers13040670

14. O'Brien S, Schiller G, Lister J, et al. High-Dose Vincristine Sulfate Liposome Injection for Advanced, Relapsed, and Refractory Adult Philadelphia Chromosome–Negative Acute Lymphoblastic Leukemia. *JCO*. 2013;31(6):676-683. doi:10.1200/JCO.2012.46.2309

15. Yan VC, Butterfield HE, Poral AH, et al. Why Great Mitotic Inhibitors Make Poor Cancer Drugs. *Trends in Cancer*. 2020;6(11):924-941. doi:10.1016/j.trecan.2020.05.010

16. Alsaihati BA, Ho KL, Watson J, et al. Canine tumor mutational burden is correlated with TP53 mutation across tumor types and breeds. *Nat Commun*. 2021;12(1):4670. doi:10.1038/s41467-021-24836-9

17. Jornada D, Dos Santos Fernandes G, Chiba D, De Melo T, Dos Santos J, Chung M. The Prodrug Approach: A Successful Tool for Improving Drug Solubility. *Molecules*. 2015;21(1):42. doi:10.3390/molecules21010042

18. Assay and Activity Questions | ChEMBL Interface Documentation. Published February 29, 2024. Accessed March 23, 2024. https://chembl.gitbook.io/chembl-interface-documentation/frequently-asked-questions/chembl-data-questions

19. Aykul S, Martinez-Hackert E. Determination of half-maximal inhibitory concentration using biosensor-based protein interaction analysis. *Analytical Biochemistry*. 2016;508:97-103. doi:10.1016/j.ab.2016.06.025

20. Sachowski J. Determine Collection Requirements. In: *Implementing Digital Forensic Readiness*. Elsevier; 2016:73-83. doi:10.1016/B978-0-12-804454-4.00007-1

21. Manning BD, Toker A. AKT/PKB Signaling: Navigating the Network. *Cell*. 2017;169(3):381-405. doi:10.1016/j.cell.2017.04.001

22. SAV1 salvador family WW domain containing protein 1 [Homo sapiens (human)] - Gene - NCBI. Accessed March 31, 2024. https://www.ncbi.nlm.nih.gov/gene/60485#:~:text=SAV1%20suppresses%20tumor%20growth%20in,points%20in%20the%20Hippo%20pathway

23. Manfredi JJ. The Mdm2–p53 relationship evolves: Mdm2 swings both ways as an oncogene and a tumor suppressor. *Genes Dev*. 2010;24(15):1580-1589. doi:10.1101/gad.1941710

24. Gharaibeh L, Elmadany N, Alwosaibai K, Alshaer W. Notch1 in Cancer Therapy: Possible Clinical Implications and Challenges. *Mol Pharmacol*. 2020;98(5):559-576. doi:10.1124/molpharm.120.000006

25. Qu J, Zheng B, Ohuchida K, et al. PIK3CB is involved in metastasis through the regulation of cell adhesion to collagen I in pancreatic cancer. *Journal of Advanced Research*. 2021;33:127-140. doi:10.1016/j.jare.2021.02.002

26. Saffari PS, Vapniarsky N, Pollack AS, et al. Most canine ameloblastomas harbor HRAS mutations, providing a novel large-animal model of RAS-driven cancer. *Oncogenesis*. 2019;8(2):11. doi:10.1038/s41389-019-0119-1

27. CDKN2A gene: MedlinePlus Genetics. Accessed March 31, 2024. https://medlineplus.gov/genetics/gene/cdkn2a/

28. Sorokina M, Steinbeck C. NaPLeS: a natural products likeness scorer—web application and database. *J Cheminform*. 2019;11(1):55. doi:10.1186/s13321-019-0378-z

29. Josef Perktold, Skipper Seabold, Kevin Sheppard, et al. statsmodels/statsmodels: Release 0.14.1. Published online December 14, 2023. doi:10.5281/ZENODO.593847

30. Emerick K. Derivation of OLS Estimator. Published online September 1, 2011. Accessed March 18, 2024. https://are.berkeley.edu/courses/EEP118/current/derive_ols.pdf

31. Mutasa S, Sun S, Ha R. Understanding artificial intelligence based radiology studies: What is overfitting? *Clinical Imaging*. 2020;65:96-99. doi:10.1016/j.clinimag.2020.04.025

32. Haque S, Vaiselbuh SR. Vincristine and prednisone regulates cellular and exosomal miR-181a expression differently within the first time diagnosed and the relapsed leukemia B cells. *Leukemia Research Reports*. 2020;14:100221. doi:10.1016/j.lrr.2020.100221

33. Setthawongsin C, Teewasutrakul P, Tangkawattana S, Techangamsuwan S, Rungsipipat A. Conventional-Vincristine Sulfate vs. Modified Protocol of Vincristine Sulfate and L-Asparaginase in Canine Transmissible Venereal Tumor. *Front Vet Sci*. 2019;6:300. doi:10.3389/fvets.2019.00300

34. Heidari-Kalvani N, Alizadeh-Fanalou S, Yarahmadi S, et al. Investigation of the effects of catharanthine and Q10 on Nrf2 and its association with MMP-9, MRP1, and Bcl-2 and apoptosis in a model of hepatocellular carcinoma. *Naunyn-Schmiedeberg's Arch Pharmacol*. 2024;397(4):2507-2522. doi:10.1007/s00210-023-02767-0

35. Donato MT, Tolosa L, Gómez-Lechón MJ. Culture and Functional Characterization of Human Hepatoma HepG2 Cells. In: Vinken M, Rogiers V, eds. *Protocols in In Vitro Hepatocyte Research*. Vol 1250. Springer New York; 2015:77-93. doi:10.1007/978-1-4939-2074-7_5

36. He F, Ru X, Wen T. NRF2, a Transcription Factor for Stress Response and Beyond. *IJMS*. 2020;21(13):4777. doi:10.3390/ijms21134777

37. Terwilliger T, Abdul-Hay M. Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood Cancer J*. 2017;7(6):e577-e577. doi:10.1038/bcj.2017.53

38. Tu Y, Cheng S, Zhang S, Sun H, Xu Z. Vincristine induces cell cycle arrest and apoptosis in SH-SY5Y human neuroblastoma cells. *International Journal of Molecular Medicine*. 2013;31(1):113-119. doi:10.3892/ijmm.2012.1167

39. Metzger ML, Dome JS. Current Therapy for Wilms' Tumor. *The Oncologist*. 2005;10(10):815-826. doi:10.1634/theoncologist.10-10-815

40. Questions and Answers about CRISPR. Broad Institute. Published December 17, 2014. Accessed April 4, 2024. https://www.broadinstitute.org/what-broad/areas-focus/project-spotlight/questions-and-answers-about-crispr

41. Karlson CKS, Mohd-Noor SN, Nolte N, Tan BC. CRISPR/dCas9-Based Systems: Mechanisms and Applications in Plant Sciences. *Plants*. 2021;10(10):2055. doi:10.3390/plants10102055

42. Heidersbach AJ, Dorighi KM, Gomez JA, Jacobi AM, Haley B. A versatile, high-efficiency platform for CRISPR-based gene activation. *Nat Commun*. 2023;14(1):902. doi:10.1038/s41467-023-36452-w

43. Kumar P, Nagarajan A, Uchil PD. Analysis of Cell Viability by the Lactate Dehydrogenase Assay. *Cold Spring Harb Protoc*. 2018;2018(6):pdb.prot095497. doi:10.1101/pdb.prot095497

44. LDH-Glo™ Cytotoxicity Assay | LDH Assay | LDH Release | J2380 | Promega. Accessed April 4, 2024. https://www.promega.com/products/cell-health-assays/cell-viability-and-cytotoxicity-assays/ldh-glo-cytotoxicity-assay/

45. Gorfe AA. Biophysics of cancer. *Biophysical Journal*. 2022;121(19):E1-E2. doi:10.1016/j.bpj.2022.09.017

46. Magnetic Resonance Imaging (MRI). Published August 29, 2023. Accessed April 1, 2024. https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/magnetic-resonance-imaging-mri

47. Bottomley PA. NMR in medicine. *Computerized Radiology*. 1984;8(2):57-77. doi:10.1016/0730-4862(84)90065-9

48. MRI BASICS. Accessed April 1, 2024. https://www.sas.upenn.edu/~wwalsh/MRI%20BASICS.html

49. MRI instrumentation and safety: MRI main magnet | e-MRI. IMAIOS. Accessed April 1, 2024. https://www.imaios.com/en/e-mri/mri-instrumentation-and-mri-safety/main-magnet

50. How MRI Works. HowStuffWorks. Published October 25, 2010. Accessed April 1, 2024. https://science.howstuffworks.com/mri.htm

51. Reese C. Magnetic Fields of coils. University of California, Santa Barbra. Accessed March 20, 2024. https://web.physics.ucsb.edu/~lecturedemonstrations/Composer/Pages/68.46.html#:~:text=If%20we%20bend%20the%20wire,the%20number%20of%20turns%20in

52. Neil J, Ackerman JJH. Magnetic Resonance (MR); Overview. In: *Encyclopedia of the Neurological Sciences*. Elsevier; 2014:971-972. doi:10.1016/B978-0-12-385157-4.00198-6

53. Magnetic Resonance (MR) spectroscopy. Published April 2018. Accessed March 27, 2024. https://mayfieldclinic.com/pe-mrspectroscopy.htm

54. Edward W, Kirpensteijn J. Can Humans Get Cancer from Their Pet? World Small Animal Veterinary Association. Accessed March 17, 2024. https://wsava.org/wp-content/uploads/2020/01/Clinician-s-Brief-October-2017_TP-Humans-Cancer.pdf

55. Why Do Cancer Treatments Stop Working? - NCI. Published December 21, 2016. Accessed April 4, 2024. https://www.cancer.gov/about-cancer/treatment/research/drug-combo-resistance

56. DevAccount. Skin - the difference between canine and human skin. Vetwest Veterinary Clinics. Published August 22, 2022. Accessed April 4, 2024. https://www.vetwest.com.au/pet-library/skin-the-difference-between-canine-and-human-skin/

Appendix A – Python

Note that the csv files can are found in excel format can be viewed via this McMaster SharePoint link: https://mcmasteru365-my.sharepoint.com/:x:/r/personal/montel2_mcmaster_ca/Documents/Year%201/ISCI1A24/RP%27s/RP4/RP4%20excel%20(not%20csv)%20for%20intuition.xlsx?d=w5f54bbe95dd441278fb5f980ead539c1&csf=1&web=1&e=h3ofHn

```
import pandas as pd
import statsmodels.api as sm
from google.colab import files

# Upload the CSV dataset
uploaded = files.upload()

# Load the CSV dataset
data = pd.read_csv('work3.csv')

# Calculate the quartiles and IQR of pChEMBL
q1 = data['pChEMBL value'].quantile(0.25)
q3 = data['pChEMBL value'].quantile(0.75)
iqr = q3 - q1

# Calculate the upper and lower bounds for outliers
upper_bound = q3 + 1.5 * iqr
lower_bound = q1 - 1.5 * iqr

# Exclude outliers
data = data[(data['pChEMBL value'] >= lower_bound) & (data['pChEMBL value'] <= upper_bound)]

X_indices = [1, 4, 9, 13, 14, 16, 17, 21, 22]  # see dataset for important varaiables which have been selecetd for
Y_index = 23           # pChEMBL value

X = data.iloc[:, X_indices]
Y = data.iloc[:, Y_index]

# Add a constant i.e., y intercpet
X = sm.add_constant(X)
```

```
# Fit the OLS model
model = sm.OLS(Y, X).fit()

# results
print(model.summary())
```

```
import pandas as pd
import statsmodels.api as sm
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error

# dataset
data = pd.read_csv('work3.csv')

# Calculate the quartiles and IQR of pChEMBL
q1 = data['pChEMBL value'].quantile(0.25)
q3 = data['pChEMBL value'].quantile(0.75)
iqr = q3 - q1

# Calculate the upper and lower bounds for outliers
upper_bound = q3 + 1.5 * iqr
lower_bound = q1 - 1.5 * iqr

# Exclude outliers
data = data[(data['pChEMBL value'] >= lower_bound) & (data['pChEMBL value'] <= upper_bound)]

X_indices = [1, 4, 9, 13, 14, 16, 17, 21, 22]  # see above
Y_index = 23           # pChEMBL value

# Select X and Y using iloc
X = data.iloc[:, X_indices]
Y = data.iloc[:, Y_index]

# Split the dataset into training and testing (see Dr. Cousin's regression primer)
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, random_state=4)
```

```
# predicting the constant
X_train = sm.add_constant(X_train)
X_test = sm.add_constant(X_test)

# Fit the OLS model on the training data (80% of the dataset)
model = sm.OLS(Y_train, X_train).fit()

# Predict on the testing data (20% of the dataset)
Y_pred = model.predict(X_test)

# Calculate Mean Squared Error
mse = mean_squared_error(Y_test, Y_pred)
print("Mean Squared Error:", mse)

# results
print(model.summary())

#Figure 3. (these #s were calculated in excel, and added as lists for ease).
import matplotlib.pyplot as plt

# Data from excel sheet 3
epoch = list(range(1, 21))
pChEMBL_values = [4, 4, 9.7, 3.1, 6.55, 4.77, 5.05, 4.88, 4.27, 4.71, 4.18, 4.8, 4.47, 4.4, 5.08, 4.88, 4.56,
4.26, 5.24, 4.34] # actual
predicted_values = [4.0969, 4.157207, 4.542234, 4.590586, 6.45727, 4.920587, 5.276781, 4.974315,
4.544098, 4.746628, 4.628456, 4.998786, 4.639023, 4.408185, 5.320906, 5.065532, 4.813632, 4.593288,
5.381124, 4.497947] # predicted

# Plotting actual and predicted
plt.figure(figsize=(10, 6)) #fits the page well
plt.scatter(epoch, pChEMBL_values, color='k', marker = 'v', label='True pChEMBL Value') # actual PCI50
plt.scatter(epoch, predicted_values, color='b', label='Predicted pChEMBL Value') # predicted

# we want to visualize the outliers
outlier_epochs = [3, 4, 5] # form excel
```

```
for outlier_epoch in outlier_epochs:
    index = outlier_epoch - 1
    plt.scatter(epoch[index], pChEMBL_values[index], color='r', marker = 's', label='Outliers' if outlier_epoch
== 3 else None)
    plt.scatter(epoch[index], predicted_values[index], color='r', marker = 's')

# Plotting formalities
plt.xlabel('Epoch')
plt.ylabel('pChEMBL Values')
plt.title('Comparison of pChEMBL Values and Predicted Values')
plt.legend()
plt.show()
```

```
#CONFUSION MATRIX RP4 (green boxes added in word) --> SEE green boxes for most corelated CFPs
#readers guide: look at the very bottom row (pChEMBL) the shades closer to red are the ones used in the
equation since they have a stronger pearsons correlation to pChEMBL =PCI50 (half maximal inhibitory)

#libraries
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
from google.colab import files #i found this method worked best for my computer

#define the csv + upload it
uploaded = files.upload()
df=pd.read_csv('PLEASEWORK2.csv')
df.head (7)

#making the confusion matrix
correlation_matrix = df.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, cmap='coolwarm', square=True, linewidths=0.5)
plt.title('Confusion Matrix')
```

```
plt.show()
```

Appendix B- R Code

```
> C <- c(92, 85, 76.26, 90)
> N <- c("CVTD", "LL", "Neuroblastoma", "Wilms")
> barplot(C,names.arg=N,xlab="Type of Cancer",ylab= "Overall survival percentag
e",col="blue",main="Vincristine in overall survival of cancers",border="red")
>
```