Modeling cellular response in large-scale radiogenomic databases to advance precision radiotherapy

Venkata SK. Manem[1,3], Meghan Lambie[1,3], Petr Smirnov[1,3,9], Victor Kofia[1], Mark Freeman[1], Marianne Koritzinsky[1,4,5], Mohamed E. Abazeed[6,7], Benjamin Haibe-Kains[1,2,3,8,9*], Scott V. Bratman[1,3,4*]

[1] Princess Margaret Cancer Center, University Health Network, Toronto, ON, Canada
[2] Department of Computer Science, University of Toronto, Toronto, ON, Canada
[3] Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada
[4] Department of Radiation Oncology, University of Toronto, Toronto, ON, Canada
[5] Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada
[6] Department of Translational Hematology Oncology Research, Cleveland, USA
[7] Department of Radiation Oncology, Cleveland Clinic, Cleveland,USA
[8] Ontario Institute of Cancer Research, Toronto, ON, Canada
[9] Vector Institute, Toronto, ON, Canada

*Co-corresponding authors:
Benjamin Haibe-Kains, PhD
Senior Scientist and Associate Professor
Princess Margaret Cancer Centre
University Health Network
University of Toronto
(416)-581-7628
bhaibeka@uhnresearch.ca

Scott V. Bratman, MD PhD
Scientist and Assistant Professor
Princess Margaret Cancer Centre
University Health Network
University of Toronto
(416)-634-7077
scott.bratman@rmp.uhn.ca

ABSTRACT

Radiotherapy is integral to the care of a majority of cancer patients. Despite differences in tumor responses to radiation (radioresponse), dose prescriptions are not currently tailored to individual patients. Recent large-scale cancer cell line databases hold the promise of unravelling the complex molecular arrangements underlying cellular response to radiation, which is critical to novel predictive biomarker discovery. Here, we present RadioGx, a computational platform for integrative analyses of radioresponse using radiogenomic databases. We first used RadioGx to investigate the robustness of radioresponse assays and indicators. We then combined radioresponse and genome-wide molecular data with established radiobiological models to predict molecular pathways that are relevant for individual tissue types and conditions. We also applied RadioGx to pharmacogenomic data to identify several classes of drugs whose effects correlate with radioresponse. RadioGx provides a unique computational toolbox to advance preclinical research for radiation oncology and precision medicine.

INTRODUCTION

Radiotherapy is routinely used as curative therapy for cancer patients. Recent technological advances have considerably augmented the physical precision of radiotherapy, resulting in improved cure rates and less toxicity (Baumann et al., 2016; Bernier et al., 2004; Verellen et al., 2007). Biologically motivated improvements (such as the addition of radiosensitizing drugs) to radiotherapy delivery have not seen such dramatic improvements despite the known differences in radiation efficacy that exist among patients with a particular tumor type (Bentzen and Overgaard, 1994; Kozin et al., 2008; Krause et al., 2009). This is due in part to a lack of predictive biomarkers on which to stratify patients. Instead, the stratification of patients to different radiotherapy-containing regimens continues to be based primarily on clinical variables such as tumor stage and patient age (Baumann et al., 2016; Verellen et al., 2007).

The biological determinants of cellular response to radiation, referred to as radioresponse, are complex and include both genomically based cell-intrinsic and external microenvironmental factors (Baumann et al., 2016; Bernier et al., 2004; Steel et al., 1989; Verellen et al., 2007; West et al., 1993). Intrinsic radiosensitivity is thought to vary among individual tumors of the same type with implications for optimal radiotherapy dosing and curability. Measurement of intrinsic radiosensitivity in molecularly-characterized cancer cell lines could provide the radiogenomic data necessary to develop predictors of radioresponse. However, despite decades of research there remain no clinically utilized radiosensitivity biomarkers that have been discovered from cell culture radiogenomic studies. There are many reasons for this, including the need for clonogenic assays when measuring intrinsic radiosensitivity *in vitro* (Puck and Marcus, 1956), which are cumbersome and are not amenable to large screens or radiogenomic studies (Bristow et al., 2018; Yard et al., 2015). Furthermore, radiosensitivity varies with dose in a complex and highly individual manner, rendering measurements at multiple dose-levels a necessity.

Most short-term cytotoxicity assays amenable for high-throughput analysis of drug response have endpoints at 72 hours. These assays are inappropriate for measuring radiosensitivity because of the delayed cellular death by mitotic catastrophe that often occurs following ionizing radiation exposure (Brown and Wouters, 1999). To address this limitation, an extended duration (9-day) viability assay was developed as a surrogate for clonogenic survival that is amenable to high-throughput processing in microtiter plates (Abazeed et al., 2013). This assay was recently applied to 533 cancer cell lines across 17 histologies with multiple radiation dose levels (Yard et al., 2016), becoming the largest radioresponse dataset published by a significant margin. This increase in the scale of radioresponse data holds great potential to contribute to the discovery of robust predictive biomarkers that could someday be translated into clinical use. However, full utilization by the research community requires sophisticated analysis tools that can appropriately model cellular response to radiation and seamlessly integrate associated molecular and pharmacogenomic profiles of cell lines.

In this study, we performed a preclinical assessment of intrinsic radiosensitivity using large-scale radiogenomic datasets (Figure 1). We sought to (*i*) model dose-response data using

3

the linear-quadratic (LQ) model (Brenner, 2008; Dale, 1985; Fowler, 1989); (*ii*) integrate the modeled radioresponse profiles with transcriptomic data to determine pathway- and tissue-specific determinants of radioresponse; (*iii*) infer radioresponse under hypoxic conditions; and *(iv)* identify classes of drugs with cytotoxic effects that correlate with radioresponse. To facilitate these and other future analyses, we developed RadioGx, a new computational toolbox enabling comparative and integrative analysis of radiogenomic datasets. Our work provides a framework for future hypothesis generation and preclinical assessments of radioresponse using appropriate biological assays and indicators.

RESULTS

The RadioGx Platform

To realize the full potential of large-scale radiogenomics datasets for robust biomarker discovery, we developed the RadioGx software package (Supplementary Figure 1). RadioGx represents the first computational toolbox that integrates radioresponse data with radiobiologic modeling and molecular data from hundreds of cancer cell lines. Within RadioGx, datasets are standardized with comprehensive cell line annotations including the type of radioresponse assay (i.e., clonogenic assay and 9-day viability assay) and indicators used to generate dose-response data (i.e., SF2 and AUC). RadioGx enables fitting of dose-response data using established radiobiological models, quality control in order to investigate the consistency and biological plausibility of radioresponse assays and indicators, and integration of these data with other data types and radioresponse models (Figure 1).
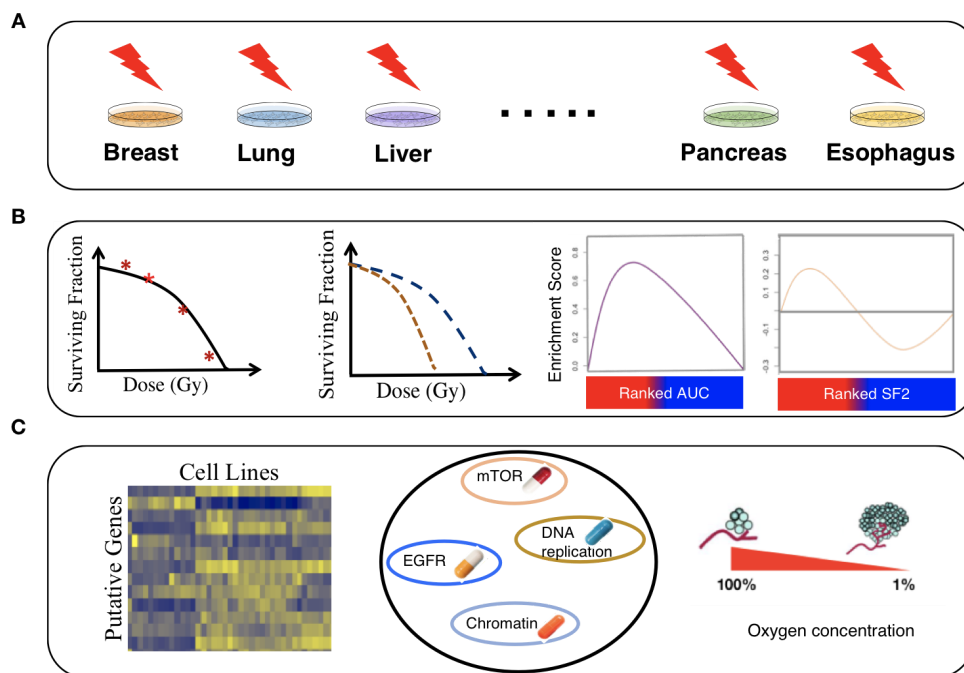

Figure 1: An overview of pre-clinical assessment of radiosensitivity using dose-response and molecular profiling data. (**A**) A schematic illustrating data collection, curation, processing of

4

radiogenomic data obtained from cancer cell lines of diverse histologies. (**B**) A schematic illustrating the process of radiobiological modeling using the linear-quadratic model, assessing the consistency of dose-response data across assays, and evaluating distinct radiation sensitivity indicators such as area under the survival curve (AUC) and surviving fraction after 2 Gy (SF2). (**C**) A schematic illustrating integrative analysis of modeled radioresponse from distinct cell lines with associated genomic/transcriptomic data, pharmacogenomic data, and hypoxia modeling.

Modeling radiation response within RadioGx

Multiple dose-response measurements from the same cell line can be incorporated into established radiobiological models to predict the effect of specific perturbations (e.g., radiotherapy fraction size or hypoxia) on radioresponse. Within RadioGx, we applied the commonly used linear-quadratic (LQ) model to fit 9-day viability assay data for 533 cancer cell lines (Yard et al., 2016) (Figure 2A). The LQ model goodness-of-fit was high for the majority of cell lines (median R^2 = 0.958; Supplementary Figure 2A). For 498/533 (93%) of cell lines, the model fit the data reasonably well ($R^2 \geq 0.6$); these cell lines were retained for subsequent analyses.
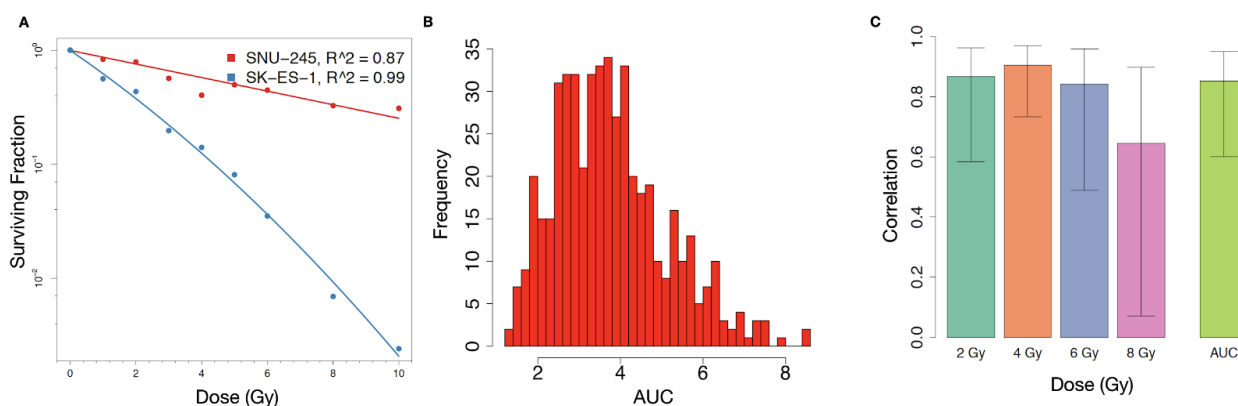

Figure 2: Fitting of dose-response data to the LQ model and concordance of radiation response across asays. (A) LQ model fit using RadioGx on the SNU-245 cholangiocarcinoma cell line (red) and SK-ES-1 Ewing sarcoma cell line (blue). The LQ model describes the fraction of cells predicted to survive (y-axis) a uniform radiation dose (x-axis) and is characterized by α and β components for each cell line. For SNU-245 and SK-ES-1, $\alpha = 0.14\,(Gy^{-1}), \beta\,(Gy^{-2}) = 0$ and $\alpha = 0.45\,(Gy^{-1}), \beta = 0.02\,(Gy^{-2})$, respectively. Solid curves indicate the model fit and points denote experimental data (Yard et al., 2016). **(B)** Histogram of AUC values calculated using the *computeAUC* function in RadioGx. **(C)** Correlation (Pearson R with standard deviation) of radioresponse results produced by the 9-day viability assay and the standard clonogenic assay according to the following indicators: SF2, SF4, SF6, SF8, and AUC. Primary data were obtained from Yard et al..

Using the LQ model for each cell line, we calculated the area under the survival curve (AUC) as a summary radioresponse indicator that is independent of a specific dose level. As expected, a range of radioresponse profiles were seen (Figures 2B). We next compared AUC and dose-specific survival data (SF2, SF4, SF6, and SF8) from the 9-day viability assay with clonogenic survival data generated by Yard et al. for a subset of cell lines (Figure 2C and Supplementary Figure 2B-F). We observed high Pearson correlation ($R \geq 0.8$) for AUC (n = 15),

5

SF2 (n = 12), SF4 (n = 15), and SF6 (n = 15), but SF8 showed only moderate correlation (R = 0.64; n = 11), consistent with prior observations suggesting poor reproducibility of survival assays following high doses of ionizing radiation (Nuryadi et al., 2018). Taken together, the 9-day viability assay provides a robust surrogate for clonogenic survival to ionizing radiation at a range of dose levels. Moreover, the LQ model within RadioGx allows for characterization of radioresponse and derivation of radioresponse indicators for the vast majority of cancer cell lines.

Comparison of radioresponse indicators

Summary indicators of radioresponse are useful for preclinical investigations. As radioresponse data within RadioGx has been fit to the LQ model, there is an opportunity to describe radioresponse through imputed survival across a range of dose levels (AUC) or at a specific dose level (e.g., SF2). There is currently no consensus regarding the optimal indicator for use across studies, with both AUC and SF2 frequently used (Deacon et al., 1984; de Jong et al., 2015; Hall et al., 2014; Torres-Roca et al., 2005). The use of SF2 as a radioresponse indicator has been bolstered by clinical observations that local tumor control following radiotherapy may be associated with SF2 measured from *ex vivo* tumor cells (Torres-Roca and Stevens, 2008). Moreover, SF2 is thought to differentiate between radiosensitive and radioresistant cell types (Fertil and Malaise, 1985). However, there is currently insufficient evidence to support the routine use of SF2 or AUC when probing the molecular determinants of radioresponse.

We compared AUC and SF2 across all cell lines within RadioGx. The values were well correlated (ρ = 0.92, 95%CI: 0.90 - 0.93, p=2.2e-16; Figure 3A); the strongest correlations were observed among the most radiosensitive cell lines, and the weakest correlations were observed among the most radioresistant cell lines, where cell death at higher doses likely contributes to the AUC value but has no bearing on SF2 (Figure 3B and Supplementary Figure 3). We then asked whether the biological processes that govern these two radioresponse indicators are the same. To achieve this, we correlated the basal level gene expression data from the Cancer Cell Line Encyclopedia (CCLE) (Barretina et al., 2012) with the radioresponse indicators (SF2 and AUC), and performed gene set enrichment analysis (GSEA) on the gene list ranked based on correlation estimates. For an FDR < 5%, 77 transcriptional pathways were enriched using AUC as the radioresponse indicator, out of which 41 and 36 pathways were positively and negatively correlated with AUC, respectively (Supplementary File 1, Supplementary Figure 4). Similarly, using SF2 as the radioresponse indicator, only 38 pathways were enriched, out of which 19 were positively correlated with the SF2 value. All but three of the pathways enriched using SF2 were enriched using AUC (Figures 3C and 3D).

The 17 pathways that were significantly correlated with radioresponse using the AUC indicator but not the SF2 indicator included biological processes known to impact radioresponse, suggesting stronger relevance for AUC. For instance, the *Nrf2-mediated oxidative stress response* pathway was positively associated with AUC but not with SF2 (Supplementary File 1). In conditions of oxidative stress, such as following radiation, degradation of Nrf2 is prevented, leading to its stabilization and translocation into the nucleus, where it activates expression of a

6

wide variety of downstream antioxidant targets (Espinosa-Diez et al., 2015); this pathway has previously been described as contributing to intrinsic radioresistance (Abazeed et al., 2013; Singh et al., 2010). In addition, progression through the cell cycle following radiation response is a known factor in determining cell survival vs. cell death via mitotic catastrophe. Three pathways directly related to cell cycle progression ([1] *cell cycle: G2/M DNA damage checkpoint regulation*, [2] *cell cycle: G1/S checkpoint regulation,* and [3] *cell cycle control of chromosomal replication*) were all seen exclusively when using AUC as the radioresponse indicator. Thus, as compared with SF2, AUC was able to capture more gene expression pathways putatively correlated with radioresponse including pathways with known mechanistic roles in mediating cellular radioresponse. Taken together, our analyses reveal AUC and SF2 as related radioresponse indicators with AUC providing for a more comprehensive characterization of the biological processes underpinning radioresponse. As a result of these findings, we exclusively used AUC as the radioresponse indicator for subsequent analyses.
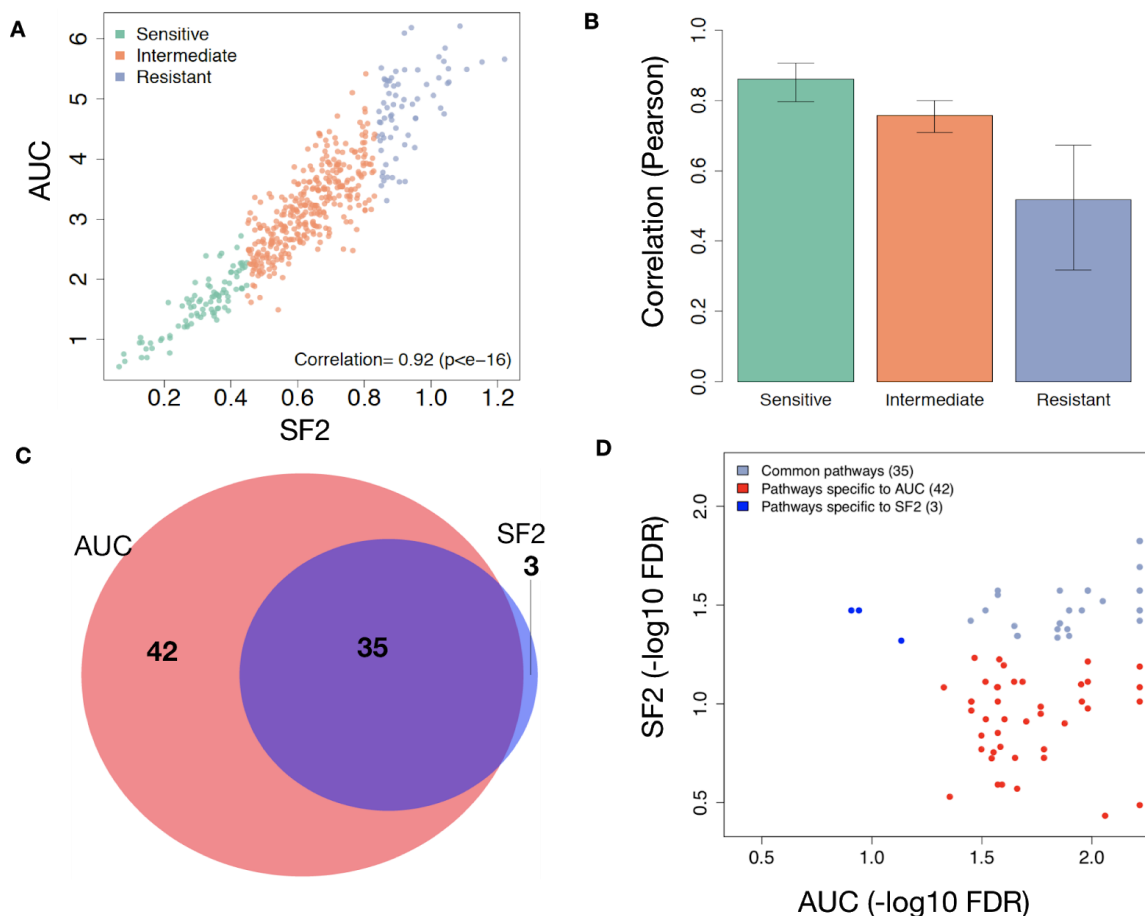

Figure 3: Concordance of SF2 and AUC. **(A)** Correlation between the radioresponse indicators, SF2 and AUC, across 498 cell lines. **(B)** Pearson correlation (with standard deviation) between SF2 and AUC across 498 cell lines based on tertiles. **(C)** Venn diagram illustrating the transcriptional pathways associated with radioresponse using SF2 or AUC as the response indicator. **(D)** False discovery rate (FDR) for each transcriptional pathway from (C) illustrating greater levels of statistical significance among pathways specific to AUC.

Radiobiological modeling to estimate impact of DNA repair on survival

The LQ model can be used to estimate the dependence of cellular survival on radiotherapy fraction size and DNA repair. The α and β values produced by the LQ model allow for comparisons among distinct cell lines or tumors, and in clinical practice the α/β ratio is used to predict cellular response to different radiotherapy fractionation schemes.

Using the LQ model, we derived the α/β ratio for cancer cell lines within RadioGx. A wide range of α/β values were observed (Figure 4A; median = 10.14; interquartile range = 4.49 - 28.07). As expected, the α component was strongly anti-correlated with AUC, whereas the β component displayed no significant association with AUC (Figure 4B). This result indicates that for the cell line data contained within RadioGx, dependence of cellular survival on radiotherapy fraction size is a distinct parameter that describes radioresponse and should therefore be considered alongside radiosensitivity (e.g., AUC or SF2) in preclinical investigations.

In order to examine the biological factors that underlie the differences between α, β and AUC, we identified transcriptional pathways that were significantly associated with each radioresponse metric. For an FDR of 5%, we found 14 pathways commonly associated with all 3 metrics (Figure 4C; Supplementary File 2). Supporting the biological relevance of these pathways, several known components of DNA damage response, signaling, and repair were represented among the 14 common pathways. For instance, pathways related to *mismatch repair in eukaryotes*, *role of BRCA1 in DNA damage response*, and *cell cycle control of chromosomal replication* were each present. These results, which are consistent with fundamental tenets of radiobiology, suggest that analysis of large cell line resources within RadioGx could be performed to generate novel hypotheses and could contribute to preclinical biomarker discovery.
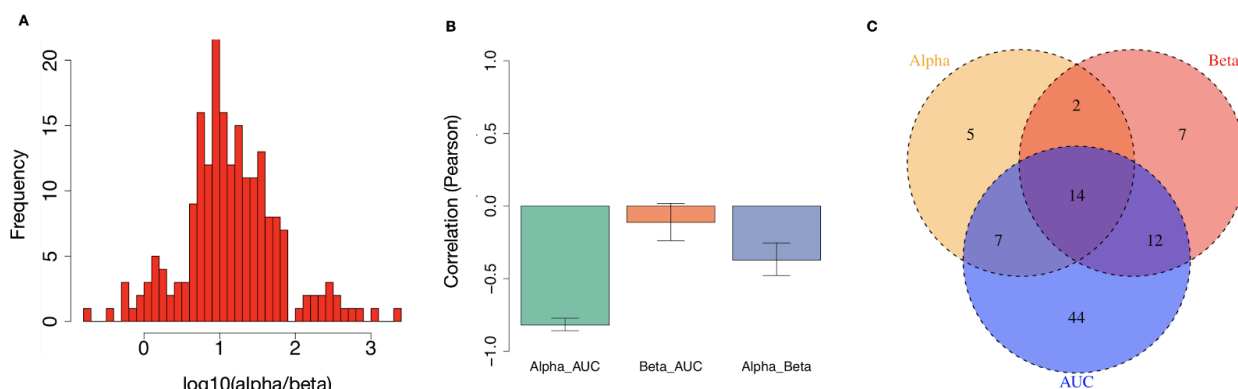

Figure 4: Distinct biological underpinnings of α/β **derived from the LQ model. (A)** Histogram of α/β (Gy) values obtained from the LQ model across all cell lines. **(B)** Pearson correlations (with standard deviation) between AUC and the α and β components of the LQ model. **(C)** Transcriptional pathways that are significantly associated with AUC, α, and/or β.

Modeling the effects of hypoxia on radioresponse

By integrating radioresponse and molecular data, RadioGx is meant to enable new biological insights and predictions. To further demonstrate the utility of RadioGx for this purpose, we next extended the radiobiological modeling to incorporate the putative effects of oxygen availability in the tumor microenvironment on radioresponse (Daşu et al., 2005).

Molecular oxygen is necessary to mediate the indirect effects of ionizing radiation to exert cell kill. Thus, cells become more resistant to radiation under oxygen-deficient conditions. We derived adjusted AUC values for the cancer cell lines within RadioGx at a range of oxygen partial pressures. As expected, reduced oxygen partial pressure resulted in a predicted increase in AUC (Figure 5A). Cell lines from distinct cancer histologies displayed consistent increases in AUC under hypoxic conditions (p<2.2e-16 for all, Wilcoxon test), but the magnitude of this increase differed between histologies (Figure 5B). The largest and smallest median differences in AUC were observed for cancer cell lines from the breast and large intestine, respectively. These differences reflect a non-linear relationship between oxygen availability and radioresponse that is dependent on α/β.
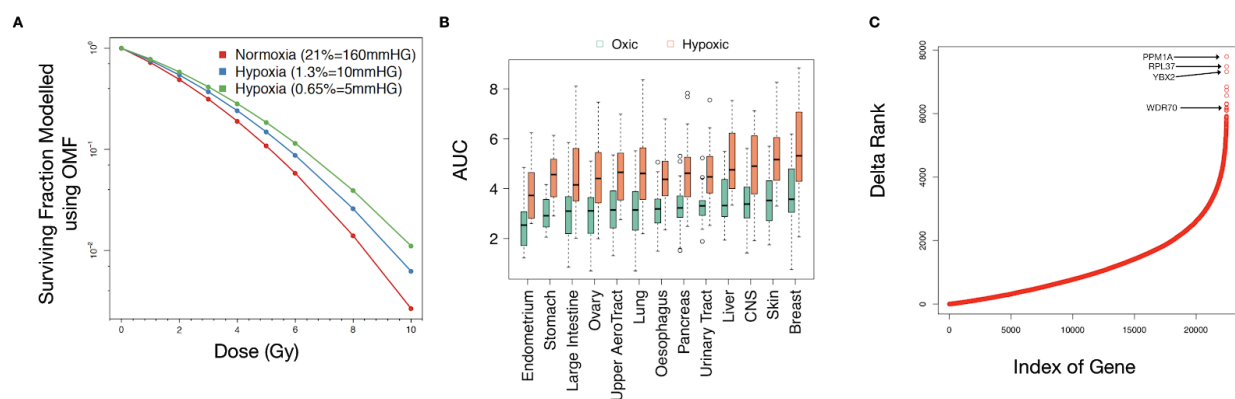

Figure 5: Integrative analysis of radiobiological model with transcriptomic data and prediction of radioresponse under hypoxia. (A) Hypothetical illustration of cancer cell surviving fraction according to dose and oxygen partial pressure, as modeled using RadioGx. Solid curves are modeled using Equation (3) (Methods). The computed AUC values are 2.41, 2.71, 2.97 for normoxia (160 mmHg), hypoxic condition with 10 mmHg, and hypoxic condition with 5 mmHg, respectively. **(B)** Changes in AUC by tissue type (with minimum of 15 cell lines within RadioGx) under normoxic (160 mmHg) or hypoxic (5 mmHg) conditions, ordered according to median AUC under normoxia. **(C)** The difference in ranks are shown between the strength of univariate association of each gene with AUC under normoxic (160 mmHg) vs. hypoxic (5 mm Hg) conditions across cancer cell lines within RadioGx.

Next, we evaluated the univariate association of gene expression levels measured under normoxic conditions with AUC values under normoxic and hypoxic conditions. For an FDR < 5%, the numbers of genes that were significantly associated with radioresponse were 1,825 and 2,395 under normoxic and hypoxic conditions, respectively (Supplementary File 3). Moreover, 1,375 genes were negatively associated with radioresponse under normoxic condition but

9

positively associated with radioresponse under hypoxic condition, and 471 genes were positively associated with radioresponse under normoxic condition but negatively associated with radioresponse under hypoxic condition (Supplementary Figure 5). In keeping with these effects, we observed large changes in the ranking of strength of correlation of specific genes with radioresponse under oxic and hypoxic conditions (Figure 5C). The gene with the greatest change, PPM1A, has been implicated in the regulation of cellular stress response and has previously been shown to have hypoxia-specific activity (Heikkinen et al., 2010). WDR70, a gene with known roles in DNA double strand break repair (Guo et al., 2016; Zeng et al., 2016), also displayed a large change in this analysis (Figure 5C). One might hypothesize based on our results that WDR70 could have previously uncharacterized hypoxia-specific activities and/or expression; these findings warrant further investigation.

Tissue specificity of radioresponse and repair
It is known that distinct tissues and tumor types respond differently to ionizing radiation exposure. Intrinsic radiosensitivity has been suggested as a major contributing factor to this differential response (Yard et al., 2016). We used RadioGx to interrogate radioresponse within tissue types represented by a minimum of 15 cell lines (Figure 6).

To examine the biological factors that may underlie suspected differences in radioresponse between tissue types, we identified 281 transcriptional pathways that were significantly associated with radioresponse within at least one tissue type (Figure 6A; Supplementary File 4). Of these 281 pathways, 123 were statistically significant only in one tissue type (Supplementary Figure 6). Overall, there were more statistically significant pathway associations with radiosensitivity than radioresistance (total across all tissue types: 437 and 226, respectively). Remarkably, we did not find any transcriptional pathways that were statistically significantly associated with radioresponse across all tissue types. We also observed variable α/β values among the tissue types within RadioGx (Figure 6B), suggesting heterogeneity of DNA repair and dependence on radiotherapy fraction size. Our findings support the use of tumor-specific as opposed to pan-cancer radioresponse biomarkers and radiosensitizing strategies.
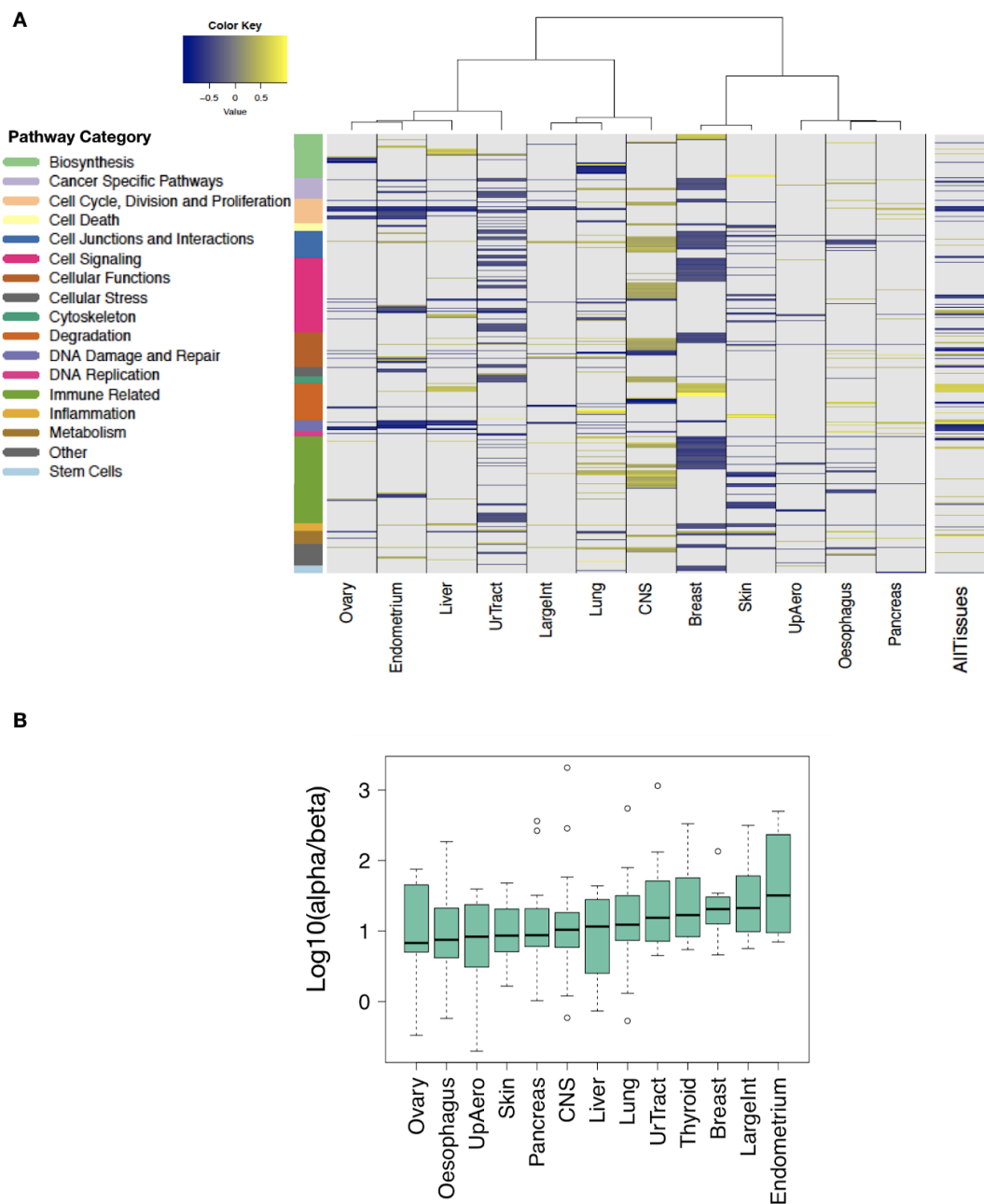
Figure 6: Tissue specificity of molecular determinants of radioresponse. (A) The tissue types (columns) represented by a minimum of 15 cancer cell lines were considered for analysis. A total of 281 pathways are depicted (rows) and are annotated by function. Colours designate pathways significantly associated with AUC (FDR < 5%). **(B)** Heterogeneity of α/β ratios across cancer cell lines derived from distinct tissue types ordered according to median values.

Common dependencies of therapeutic effects among radiotherapy and drugs

Datasets within RadioGx are standardized with regard to cell line annotations such that integrated analyses using other existing datasets can be easily conducted. For instance, our previously published tool, PharmacoGx (Smirnov et al., 2016), contains pharmacogenomic data

from multiple studies and enables meta-analysis of pharmacogenomic data. We wished to identify categories of drugs with cytotoxic effects that correlate with radioresponse, so we interrogated RadioGx to compare cellular responses to ionizing radiation and chemotherapeutic agents (n=545 distinct drugs). Drug responses were obtained from 480 cancer cell lines from the CTRPv2 pharmacogenomic dataset (Supplementary Table 1) that were in common between the datasets. We computed the correlation between drug response and radiation response across the cancer cell lines (Supplementary Figure 7) and then classified drugs according to pharmacological categories (i.e., by cellular targets and/or mechanisms of action). Drugs targeting the cytoskeleton, DNA replication, and mitosis displayed the strongest correlations with radioresponse (FDR < 5%) (Figure 7). Thymidylate synthetase inhibitors such as the known radiosensitizing drug, fluorourocil, also displayed cytotoxic effects that correlated with radioresponse but did not reach statistical significance. In addition to these anticipated and largely confirmatory findings, we also observed unexpected negative associations between radioresponse and cytotoxic effects of drugs targeting numerous cell signaling pathways (i.e., PI3K signalling, ERK MAPK signaling, WNT signalling, EGFR signalling, ABL signalling), although these were not statistically significant.
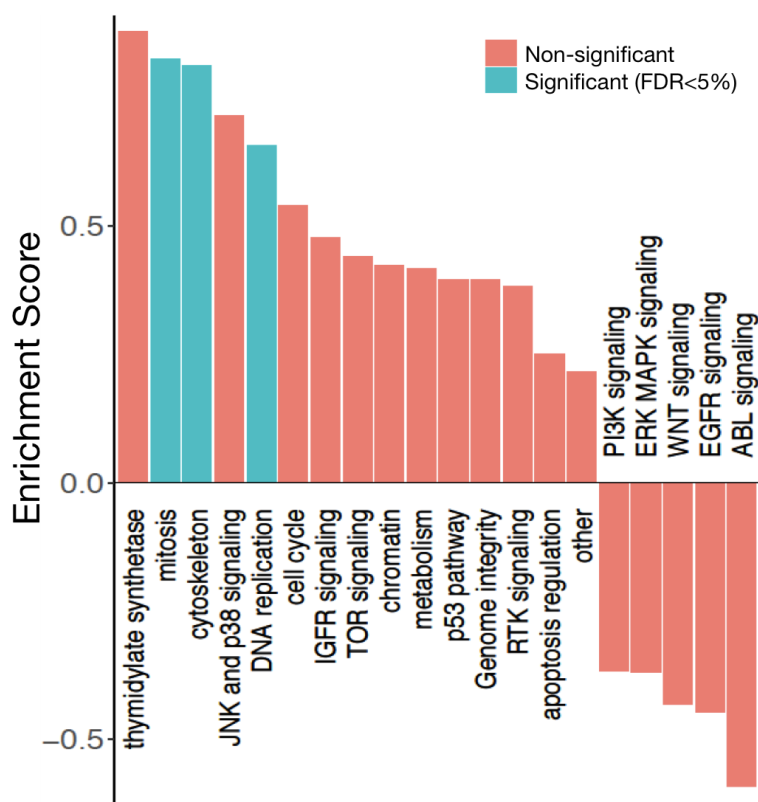

Figure 7: Identification of drugs and pharmacological classes with cytotoxic effects on cancer cell lines that correlate with radioresponse. Pharmacological enrichment analysis using radiation AUC as the radioresponse indicator. Pharmacological classes with statistically significant associations with radioresponse in cancer cell lines are indicated.

DISCUSSION

To date, the paradigm of precision medicine has primarily been applied to advanced incurable cancers. For early stage curable cancers for which radiotherapy is used with curative intent, there remains a need for more precise biologically-tailored radiotherapy delivery. For instance, there are currently no clinically implemented molecular biomarkers that are predictive of radioresponse. This also extends to predictive insights into the response of tumors to other therapeutic agents that may be administered in combination with radiotherapy. Although molecular diagnostic tools are making their way into clinical practice in other settings, the lack of equivalent molecular indicators in the field of radiobiology has impeded translation in this domain (Baumann et al., 2016; Bibault et al., 2013; Bristow et al., 2018).

Recently, large radioresponse and genomic datasets have been generated from hundreds of cancer cell lines, providing an opportunity to address this unmet need. We have developed RadioGx, an open-source software package that enables users to perform integrative analysis of radiogenomic datasets for preclinical evaluation of radioresponse determinants. RadioGx standardizes published nomenclature and annotations between datasets and integrates dose-response and molecular data.

We used RadioGx to evaluate the appropriateness of the 9-day viability assay for assessing radioresponse, the robustness of distinct radioresponse indicators, and the utility of applying established radiobiological models to the data for novel hypothesis generation. We confirmed the findings from Yard et al. that the 9-day viability assay, which is amenable to high-throughput processing and analysis, largely recapitulates the results of the more tedious clonogenic assay. We note that some prior putative intrinsic radiosensitivity gene expression signatures that were generated using cell line clonogenic survival data have failed to validate using independent sets of cancer cell lines (Bratman et al., 2017; Hall et al., 2014), highlighting the need for robust reproducible methodologies for future studies. Moreover, we found that AUC derived from the LQ model might provide a more complete characterization of the biological processes underpinning radioresponse as compared with the dose-specific SF2 indicator, particularly for relatively radioresistant cell types. Based on our findings, we suggest that AUC be the radioresponse indicator of choice for preclinical studies. While we found that the LQ model fit the radioresponse data for the vast majority of cancer cell lines within RadioGx, a small subset were not amenable to LQ modeling and should be excluded or used with caution in future radiogenomic analyses.

A major hurdle in the development of large-scale radioresponse datasets has been the technical and throughput challenges associated with the clonogenic assay. We demonstrated how existing data within RadioGx can be used to generate hypotheses and make predictions to inform future investigations. For instance, recognizing a dearth of large-scale radioresponse data under hypoxic conditions, we integrated radiobiological modeling with gene expression data from RadioGx, which allowed us to predict radioresponse under hypoxic conditions. Our findings suggest that the change in radioresponse under hypoxia is tissue-specific and that

13

specific genes are either differentially associated with radioresponse under normoxic and hypoxic conditions or may have expression levels or activity that are regulated by oxygen tension; these specific hypotheses generated by our analysis through RadioGx could be tested experimentally in future studies. In addition, by combining RadioGx with an existing pharmacogenomics analysis platform, we uncovered drugs with cytotoxic effects that are correlated or anti-correlated with radioresponse, suggestive of genomic/transcriptomic dependencies related to their mechanisms of action. We were able to confirm drug classes with therapeutic effects that overlap with ionizing radiation (e.g., mitotic inhibitors); moreover, this analysis proposed novel hypotheses regarding possible anticorrelated therapeutic effects with drugs targeting a number of cellular signaling pathways such as ABL and EGFR. Future studies may seek to examine whether members of these drug classes may make rational combination therapies with radiation as a result of reduced additive toxicity.

In summary, this study demonstrates the impact of combining radiogenomic datasets with established radiobiological models and other existing pharmacogenomic data. Future applications of RadioGx may include generation of biomarkers for intrinsic radiosensitivity and selection of novel combination therapies for preclinical testing. Thus, we envision that RadioGx will help to accelerate preclinical radiotherapeutic discovery pipelines and guide the selection of appropriate biological endpoints.

METHODS

Curation of dose-response and transcriptomic data
One of the major hurdles in genomic studies involving cell lines is the lack of standardized identifiers for cell lines. In order to overcome this, we assigned a unique identifier to each cell line and radiation therapy, and matched entities with the same unique identifier throughout the implementation. Moreover, there is a lack of standardization in annotating genomics features, i.e. annotating probe expression to gene expression across various microarray datasets. Hence, we have used the annotations from the BrainArray database, which reflect recent annotation of the human genome to perform the mapping from microarray probe sets to genomic expression data.

We implemented a RadioSet (also known as RSet) in the RadioGx package. This class is a data container storing radiation dose-response and molecular data along with experimental metadata (detailed structure provided in the Supplementary Materials). In addition, this class also enables efficient implementation of curated annotations for cell lines, and molecular features, which facilitates comparisons between across different datasets. We have implemented a unique set of functions that facilitates users to analyze radiogenomic datasets. One of the primary functions is the *downloadRSet* that allows users to download the *RadiationSet* (RSet) object. We have also incorporated a function, *linearQuadraticModel*, which plots the radiation cell survival curve using the standard radio-biological formalism, the linear-quadratic (LQ) model (see below). This function considers normal distribution to fit the errors as a default parameter, but the users also

14

have an option to choose Cauchy distribution. For a given dataset by the end user, this function fits the dataset with the LQ model, and returns radiobiological parameters alpha and beta along with the goodness of fit. To extract several features from this curve, we have implemented a function *computeAUC*, which enables users to compute area under the survival curve (AUC), *computeSF2* function, which returns the fraction of cells that survive a radiation dose of 2Gy, *computeD10* function, which returns the radiation dose at which only 10% of cells survive.

Supplementary Table 1 presents the sensitivity and transcriptomic datasets that are used in this study, and the functionality of the RadioGx package is presented in Table 1.

Function	Summary
linearquadraticmodel	Fit the dose response data using LQ model
computeAUC	Calculates the AUC of the LQ model fit
computeD10	Calculates the dose at which 10% of cells survive
computeSF2	Computes the SF2 for a given dose response data
doseresponsecurve	Plots the dose response curve

Table 1: Functionality of RadioGx package.

Radiobiological Model
A radiobiological model is a formulation that is used to allow comparisons of various clinically relevant radiotherapy treatment regimens. The most commonly used model in current clinical practice is the LQ model (Brenner, 2008; Dale, 1985; Fowler, 1989), which assumes that there are two components to cell killing induced by radiation: one that is proportional to dose (linear, α) and another that is proportional to the square of the dose (quadratic, β). The LQ model describes the fraction of cells that survived (S) a uniform dose D (Gy); the survival fraction of cells after irradiating with an acute dose D is given by:

$$S = exp(-\alpha D - \beta D^2) \qquad (1)$$

The ratio $\frac{\alpha}{\beta}$ varies by the cell population or tissue that is being irradiated, and reflects the response to different fractionation schemes. Cell populations or tissues with a high value are less sensitive to the effects of fractionation than those with a low value.

Radiobiological Modelling of Hypoxia
The LQ model can also be used to model the effect of hypoxia. Hypoxia is a hallmark of many solid malignant tumors and influences tumor progression, therapy resistance, development of metastases, clinical behavior, and response to conventional treatments like radiotherapy (Hall and Giaccia, 2012). The survival fraction of cells due to a given radiotherapy dose is given by

15

Equation (1) under well-oxygenated, or normoxic conditions. However, the surviving fraction of cells may vary depending on the amount of oxygen concentration in the tumor, as cells in the hypoxic region are considered to be more resistant to radiation therapy. This hypoxic effect can be incorporated into the LQ model using the, "*Oxygen Enhancement Ratio (OER)*", which can be normalized to yield the "*Oxygen Modification Factor (OMF)*" (Alper and Howard-Flanders, 1956; Daşu et al., 2005; Titz and Jeraj, 2008; Wouters and Martin Brown, 1997). OMF is defined as follows:

$$OMF \; = \; \frac{OER(O_2)}{OER_m} \; = \; \frac{1}{OER_m} \; \frac{(OER_m * O_2) + K_m}{O_2 + K_m} \quad (2)$$

where O_2 is the oxygen concentration in the system in mmHg, $K_m = 3$ mm Hg, defined as the oxygen at which half of the ratio is achieved, and $OER_m = 3$ is the maximum value at well-oxygenated condition. Therefore, the LQ model given in Equation (1) can be modified to include oxygen concentration as follows:

$$S = exp(-\,\alpha OMFD - \beta (OMFD)^2) \qquad (3)$$

In general, the OER can be a function of radiation dose. Some studies have suggested that the maximal oxygen enhancement varies in the range of 2.5-3 with differences in radiation dosage (Freyer et al., 1991; Palcic and Skarsgard, 1984; Skarsgard and Harrison, 1991). This can be simply included into the revised LQ model by considering different OERs for the parameters α and β, that is, OER_α and OER_β. However, since we consider the normalized OER (or, OMF), the introduction of these separate terms will not produce a significant difference in the final survival fraction. Thus, we assume $OER_\alpha = OER_\beta$ in our mathematical framework. We assume that the system is moderately hypoxic, i.e. approximately 5 mm HG for the present study.

Association with drug response and Pharmacological Enrichment Analysis
We used CTRPv2 dataset in PharmacoGx package (version 1.10.3) (Smirnov et al., 2016) that has 545 drugs to compute the association between radioresponse and drug response (defined by the Area under the curve of the Hill function). We also performed pharmacological enrichment analysis, an adaptation of the GSEA methodology. For this, we computed the correlation of radioresponse with each drug response, and a pharmacological set represents a gene set. Similar to the GSEA method, a running sum is calculated, starting with the first compound-level statistic to the last. The sum is increased if a compound-level statistic belongs to the pharmacological class of interest, otherwise, the sum is decreased. The enrichment score of the pharmacological class of interest is defined as the maximum deviation from zero of the running sum (Seashore-Ludlow et al., 2015) (Supplementary Figure 8).

Pathway Analysis
The pathway enrichment analysis on the gene expression data is carried out using the gene set enrichment analysis (GSEA) method (Subramanian et al., 2005) with pathways defined by QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity). Genes were ranked based on their coefficient of correlation

between the gene expressions and the IHC scores (core density or stromal retention ratio). GSEA was then used to compute the enrichment score for each pathway with statistical significance calculated using a permutation test (10,000 permutations) as implemented in the *piano* package (Väremo et al., 2013). Nominal p-values obtained for each pathway are corrected for multiple testing using the false discovery approach (FDR) (Benjamini and Hochberg, 1995).

Research Reproducibility

RadioGx is implemented in R. The code, documentation, and detailed tutorial describing how to run our pipeline and reproduce our analysis results are open-source and publicly available through the RadioGx GitHub repository (https://github.com/bhklab/RadioGx). A virtual machine reproducing the full software environment is available on Code Ocean. Our study complies with the guidelines outlined in (Gentleman, 2005; Sandve et al., 2013; Stroup et al., 2000). All the data are available in the form of RSet objects with associated digital object identifiers (DOI).

COMPETING INTERESTS

SVB is a co-inventor on a patent relating to circulating tumor DNA detection technology that has been licensed to Roche Molecular Diagnostics. BHK is a co-inventor on four patents related to the prediction of survival and drug response in breast cancer patients. All other authors declare no competing interests.

REFERENCES

Abazeed ME, Adams DJ, Hurov KE, Tamayo P, Creighton CJ, Sonkin D, Giacomelli AO, Du C, Fries DF, Wong K-K, Mesirov JP, Loeffler JS, Schreiber SL, Hammerman PS, Meyerson M. 2013. Integrative radiogenomic profiling of squamous cell lung cancer. *Cancer Res* **73**:6289–6298.

Alper T, Howard-Flanders P. 1956. Role of oxygen in modifying the radiosensitivity of E. coli B. *Nature* **178**:978–979.

Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jané-Valbuena J, Mapa FA, Thibault J, Bric-Furlong E, Raman P, Shipway A, Engels IH, Cheng J, Yu GK, Yu J, Aspesi P Jr, de Silva M, Jagtap K, Jones MD, Wang L, Hatton C, Palescandolo E, Gupta S, Mahan S, Sougnez C, Onofrio RC, Liefeld T, MacConaill L, Winckler W, Reich M, Li N, Mesirov JP, Gabriel SB, Getz G, Ardlie K, Chan V, Myer VE, Weber BL, Porter J, Warmuth M, Finan P, Harris JL, Meyerson M, Golub TR, Morrissey MP, Sellers WR, Schlegel R, Garraway LA. 2012. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**:603–607.

Baumann M, Krause M, Overgaard J, Debus J, Bentzen SM, Daartz J, Richter C, Zips D, Bortfeld T. 2016. Radiation oncology in the era of precision medicine. *Nat Rev Cancer* **16**:234–249.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Series B Stat Methodol* **57**:289–300.

Bentzen SM, Overgaard J. 1994. Patient-to-Patient Variability in the Expression of Radiation-Induced Normal Tissue Injury. *Semin Radiat Oncol* **4**:68–80.

Bernier J, Hall EJ, Giaccia A. 2004. Radiation oncology: a century of achievements. *Nat Rev Cancer* **4**:737–747.

Bibault J-E, Fumagalli I, Ferté C, Chargari C, Soria J-C, Deutsch E. 2013. Personalized radiation therapy and biomarker-driven treatment strategies: a systematic review. *Cancer Metastasis Rev* **32**:479–492.

Bratman SV, Milosevic MF, Liu F-F, Haibe-Kains B. 2017. Genomic biomarkers for precision radiation medicine. *Lancet Oncol* **18**:e238.

Brenner DJ. 2008. The Linear-Quadratic Model Is an Appropriate Methodology for Determining Isoeffective Doses at Large Doses Per Fraction. *Semin Radiat Oncol* **18**:234–239.

Bristow RG, Alexander B, Baumann M, Bratman SV, Brown JM, Camphausen K, Choyke P, Citrin D, Contessa JN, Dicker A, Kirsch DG, Krause M, Le Q-T, Milosevic M, Morris ZS, Sarkaria JN, Sondel PM, Tran PT, Wilson GD, Willers H, Wong RKS, Harari PM. 2018. Combining precision radiotherapy with molecular targeting and immunomodulatory agents: a guideline by the American Society for Radiation Oncology. *Lancet Oncol* **19**:e240–e251.

Brown JM, Wouters BG. 1999. Apoptosis, p53, and tumor cell sensitivity to anticancer agents. *Cancer Res* **59**:1391–1399.

Dale RG. 1985. The application of the linear-quadratic dose-effect equation to fractionated and protracted radiotherapy. *Br J Radiol* **58**:515–528.

Daşu A, Toma-Daşu I, Karlsson M. 2005. The effects of hypoxia on the theoretical modelling of tumour control probability. *Acta Oncol* **44**:563–571.

Deacon J, Peckham MJ, Steel GG. 1984. The radioresponsiveness of human tumours and the initial slope of the cell survival curve. *Radiother Oncol* **2**:317–323.

de Jong MC, Ten Hoeve JJ, Grénman R, Wessels LF, Kerkhoven R, Te Riele H, van den Brekel

MWM, Verheij M, Begg AC. 2015. Pretreatment microRNA Expression Impacting on Epithelial-to-Mesenchymal Transition Predicts Intrinsic Radiosensitivity in Head and Neck Cancer Cell Lines and Patients. *Clin Cancer Res* **21**:5630–5638.

Espinosa-Diez C, Miguel V, Mennerich D, Kietzmann T, Sánchez-Pérez P, Cadenas S, Lamas S. 2015. Antioxidant responses and cellular adjustments to oxidative stress. *Redox Biol* **6**:183–197.

Fertil B, Malaise EP. 1985. Intrinsic radiosensitivity of human cell lines is correlated with radioresponsiveness of human tumors: analysis of 101 published survival curves. *Int J Radiat Oncol Biol Phys* **11**:1699–1707.

Fowler JF. 1989. The linear-quadratic formula and progress in fractionated radiotherapy. *Br J Radiol* **62**:679–694.

Freyer JP, Jarrett K, Carpenter S, Raju MR. 1991. Oxygen enhancement ratio as a function of dose and cell cycle phase for radiation-resistant and sensitive CHO cells. *Radiat Res* **127**:297–307.

Gentleman R. 2005. Reproducible research: a bioinformatics case study. *Stat Appl Genet Mol Biol* **4**:Article2.

Guo L-D, Wang D, Yang F, Liang Y-J, Yang X-Q, Qin Y-Y, Ren L-F, Zeng M, Tang Z-Z, Wang X-J, Wang S, Liu C, Lou J-Y, Chen J. 2016. [Functional Analysis of DNA Damage Repair Factor WDR70 and Its Mutation in Ovarian Cancer]. *Sichuan Da Xue Xue Bao Yi Xue Ban* **47**:501–506.

Hall EJ, Giaccia AJ. 2012. Radiobiology for the Radiologist. Lippincott Williams & Wilkins.

Hall JS, Iype R, Senra J, Taylor J, Armenoult L, Oguejiofor K, Li Y, Stratford I, Stern PL, O'Connor MJ, Miller CJ, West CML. 2014. Investigation of radiosensitivity gene signatures in cancer cell lines. *PLoS One* **9**:e86329.

Heikkinen PT, Nummela M, Leivonen S-K, Westermarck J, Hill CS, Kähäri V-M, Jaakkola PM. 2010. Hypoxia-activated Smad3-specific dephosphorylation by PP2A. *J Biol Chem* **285**:3740–3749.

Kozin SV, Niemierko A, Huang P, Silva J, Doppke KP, Suit HD. 2008. Inter- and intramouse heterogeneity of radiation response for a growing paired organ. *Radiat Res* **170**:264–267.

Krause M, Gurtner K, Deuse Y, Baumann M. 2009. Heterogeneity of tumour response to combined radiotherapy and EGFR inhibitors: differences between antibodies and TK inhibitors. *Int J Radiat Biol* **85**:943–954.

Nuryadi E, Mayang Permata TB, Komatsu S, Oike T, Nakano T. 2018. Inter-assay precision of clonogenic assays for radiosensitivity in cancer cell line A549. *Oncotarget* **9**:13706–13712.

Palcic B, Skarsgard LD. 1984. Reduced oxygen enhancement ratio at low doses of ionizing radiation. *Radiat Res* **100**:328–339.

Puck TT, Marcus PI. 1956. Action of x-rays on mammalian cells. *J Exp Med* **103**:653–666.

Sandve GK, Nekrutenko A, Taylor J, Hovig E. 2013. Ten simple rules for reproducible computational research. *PLoS Comput Biol* **9**:e1003285.

Seashore-Ludlow B, Rees MG, Cheah JH, Cokol M, Price EV, Coletti ME, Jones V, Bodycombe NE, Soule CK, Gould J, Alexander B, Li A, Montgomery P, Wawer MJ, Kuru N, Kotz JD, -Y. Hon CS, Munoz B, Liefeld T, Dan ik V, Bittker JA, Palmer M, Bradner JE, Shamji AF, Clemons PA, Schreiber SL. 2015. Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discov* **5**:1210–1223.

Singh A, Bodas M, Wakabayashi N, Bunz F, Biswal S. 2010. Gain of Nrf2 function in non-small-cell lung cancer cells confers radioresistance. *Antioxid Redox Signal* **13**:1627–1637.

Skarsgard LD, Harrison I. 1991. Dose dependence of the oxygen enhancement ratio (OER) in

radiation inactivation of Chinese hamster V79-171 cells. *Radiat Res* **127**:243–247.

Smirnov P, Safikhani Z, El-Hachem N, Wang D, She A, Olsen C, Freeman M, Selby H, Gendoo DMA, Grossmann P, Beck AH, Aerts HJWL, Lupien M, Goldenberg A, Haibe-Kains B. 2016. PharmacoGx: an R package for analysis of large pharmacogenomic datasets. *Bioinformatics* **32**:1244–1246.

Steel GG, McMillan TJ, Peacock JH. 1989. The radiobiology of human cells and tissues. In vitro radiosensitivity. The picture has changed in the 1980s. *Int J Radiat Biol* **56**:525–537.

Stroup DF, Berlin JA, Morton SC, Olkin I, David Williamson G, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB, for the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) Group. 2000. Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting. *JAMA* **283**:2008–2012.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**:15545–15550.

Titz B, Jeraj R. 2008. An imaging-based tumour growth and treatment response model: investigating the effect of tumour oxygenation on radiation therapy response. *Phys Med Biol* **53**:4471–4488.

Torres-Roca JF, Eschrich S, Zhao H, Bloom G, Sung J, McCarthy S, Cantor AB, Scuto A, Li C, Zhang S, Jove R, Yeatman T. 2005. Prediction of radiation sensitivity using a gene expression classifier. *Cancer Res* **65**:7169–7176.

Torres-Roca JF, Stevens CW. 2008. Predicting response to clinical radiotherapy: past, present, and future directions. *Cancer Control* **15**:151–156.

Väremo L, Nielsen J, Nookaew I. 2013. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res* **41**:4378–4391.

Verellen D, De Ridder M, Linthout N, Tournel K, Soete G, Storme G. 2007. Innovations in image-guided radiotherapy. *Nat Rev Cancer* 71–71.

West CM, Davidson SE, Roberts SA, Hunter RD. 1993. Intrinsic radiosensitivity and prediction of patient response to radiotherapy for carcinoma of the cervix. *Br J Cancer* **68**:819–823.

Wouters BG, Martin Brown J. 1997. Cells at Intermediate Oxygen Levels Can Be More Important Than the "Hypoxic Fraction" in Determining Tumor Response to Fractionated Radiotherapy. *Radiat Res* **147**:541.

Yard B, Chie EK, Adams DJ, Peacock C, Abazeed ME. 2015. Radiotherapy in the Era of Precision Medicine. *Semin Radiat Oncol* **25**:227–236.

Yard BD, Adams DJ, Chie EK, Tamayo P, Battaglia JS, Gopal P, Rogacki K, Pearson BE, Phillips J, Raymond DP, Pennell NA, Almeida F, Cheah JH, Clemons PA, Shamji A, Peacock CD, Schreiber SL, Hammerman PS, Abazeed ME. 2016. A genetic basis for the variation in the vulnerability of cancer to DNA damage. *Nat Commun* **7**:11428.

Zeng M, Ren L, Mizuno K 'ichi, Nestoras K, Wang H, Tang Z, Guo L, Kong D, Hu Q, He Q, Du L, Carr AM, Liu C. 2016. CRL4(Wdr70) regulates H2B monoubiquitination and facilitates Exo1-dependent resection. *Nat Commun* **7**:11364.
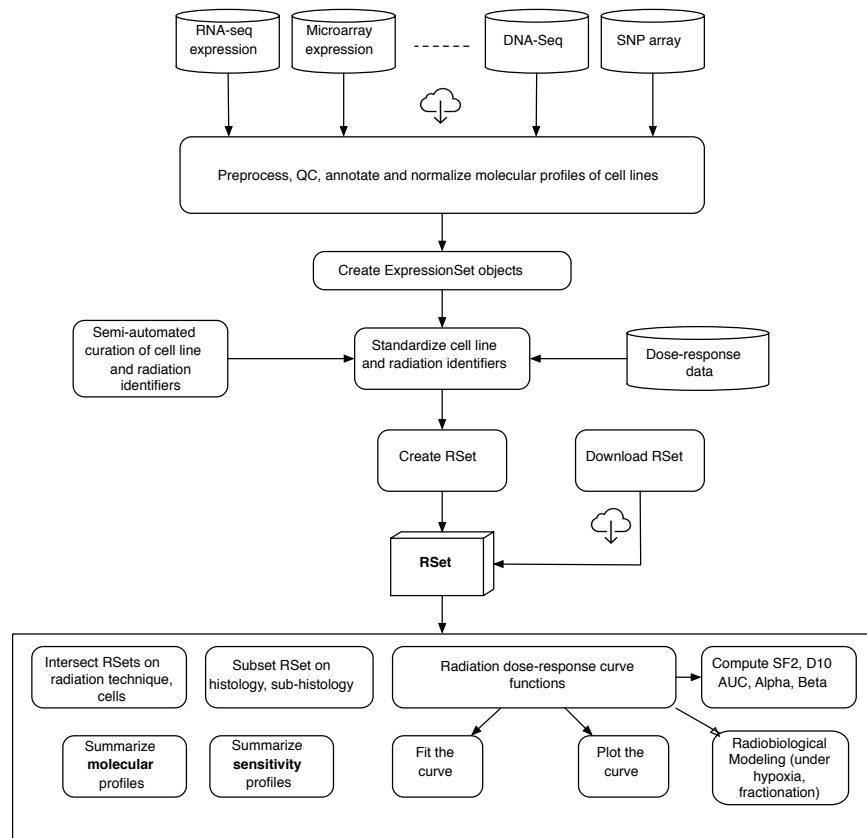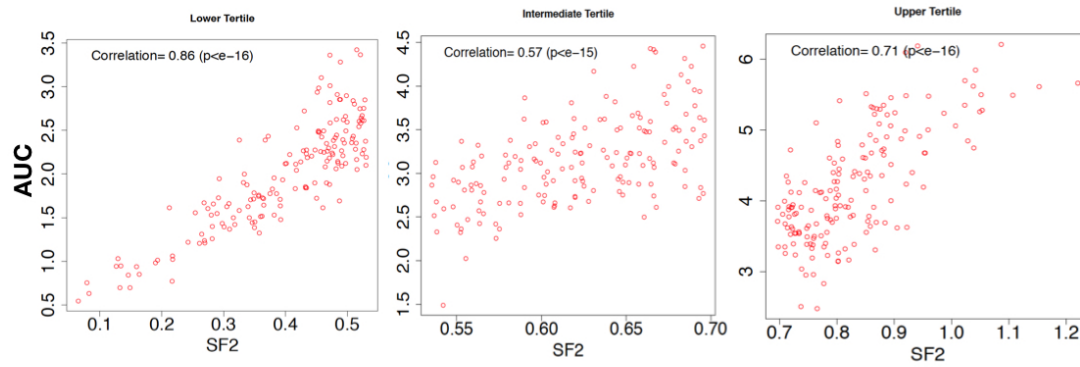
Supplementary Figure 1: Design of RadioGx platform.
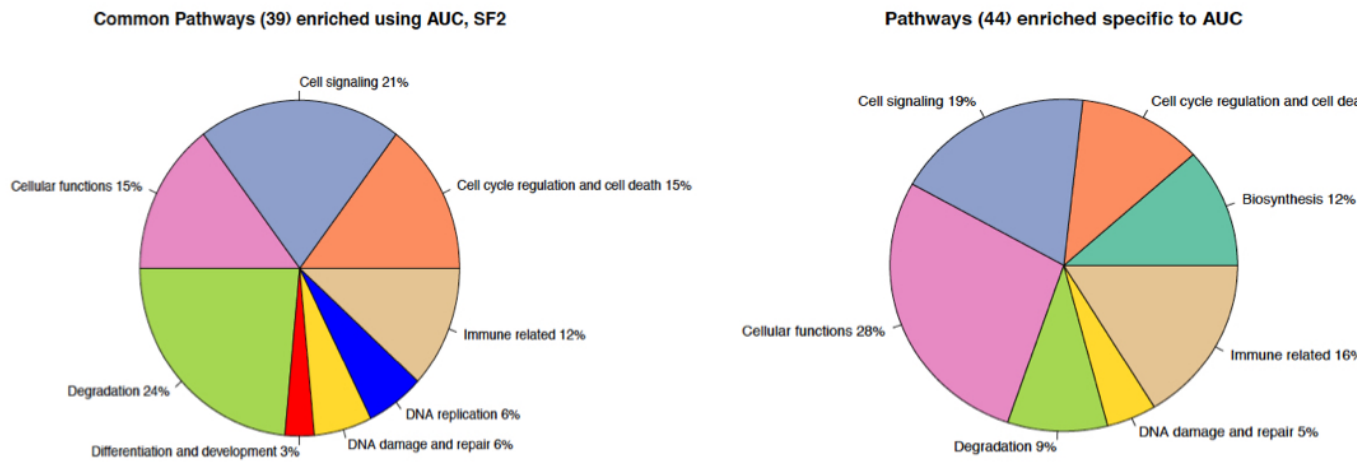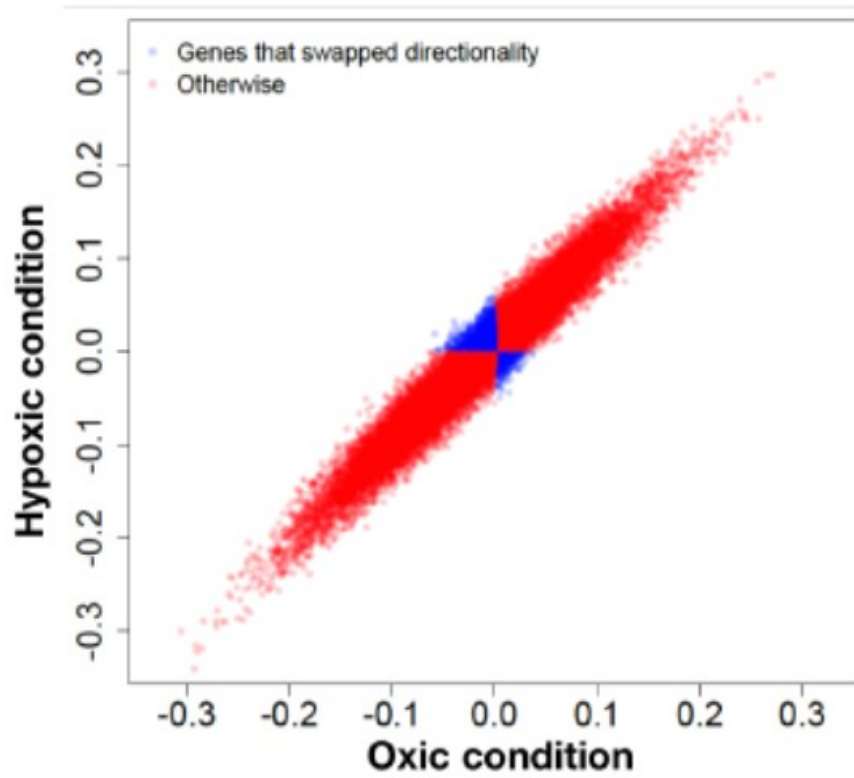
Goodness of fit by the LQ model
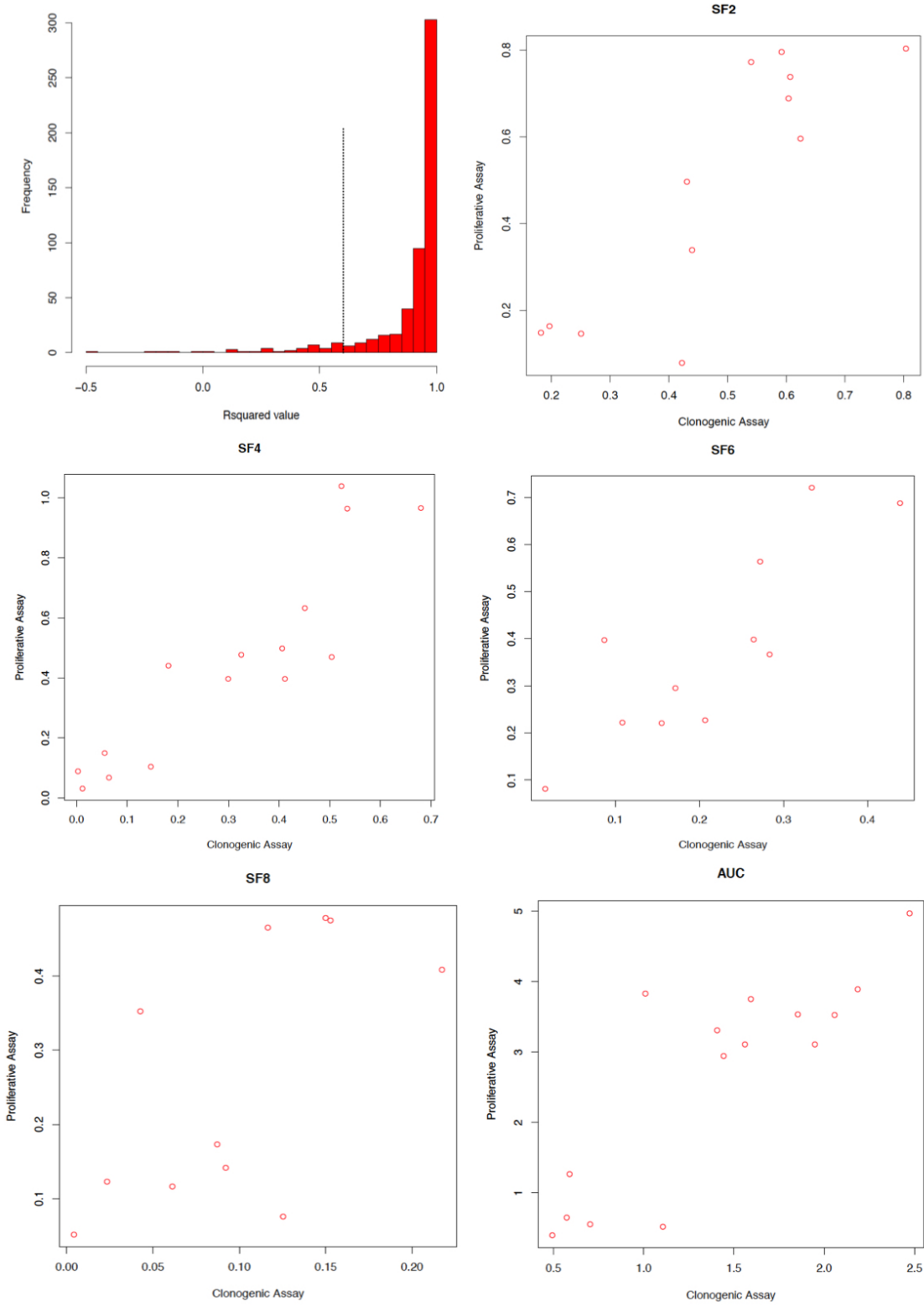
Supplementary Figure 3: Comparison of SF2 and AUC based on tertiles.
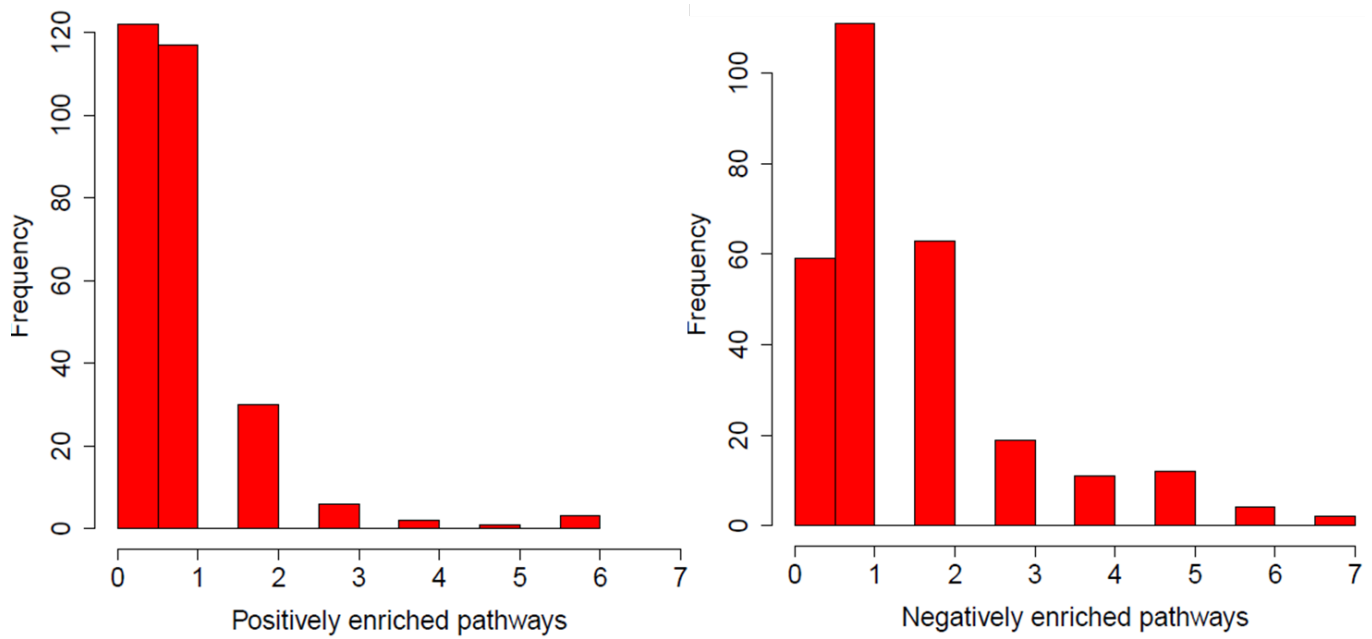
Pathways: SF2 vs. AUC

Supplementary Figure 4: Pathway analysis comparison: SF2 vs. AUC.

Supplementary Figure 5: Univariate correlation between radiation response associated genes under oxic and modeled hypoxic conditions.

Supplementary Figure 6: The tumour types (n=12) represented by a minimum of 15 cell lines were considered for analysis. A total of 281 pathways are enriched for FDR$< 5\%$. Number of positively and negatively enriched pathways in each tissue.
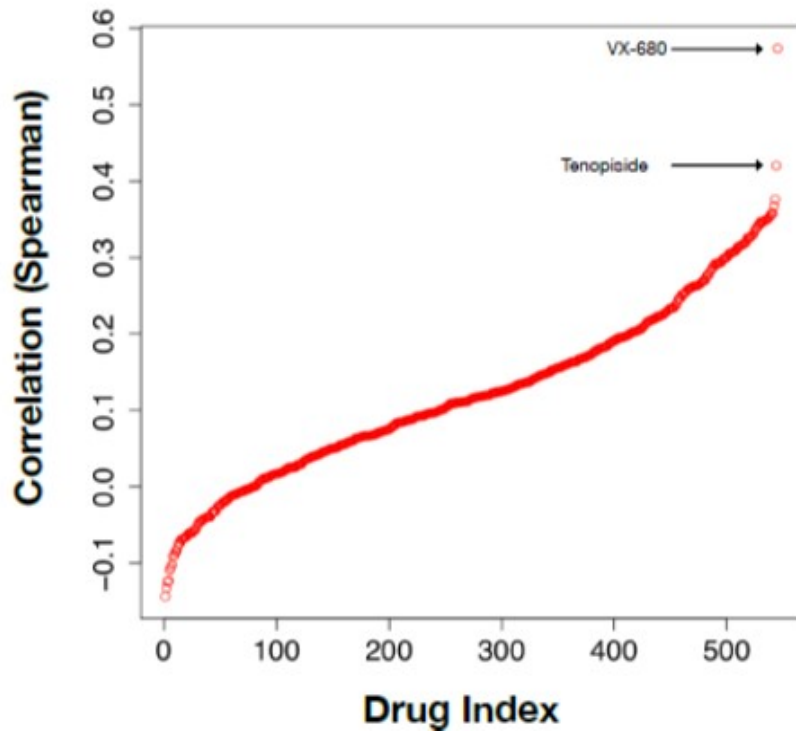
Supplementary Figure 7: Replication of Figure 1 from Yard et al 2016, demonstrating the correlation between drug response and radiation response. Figure produced using RadioGx package.
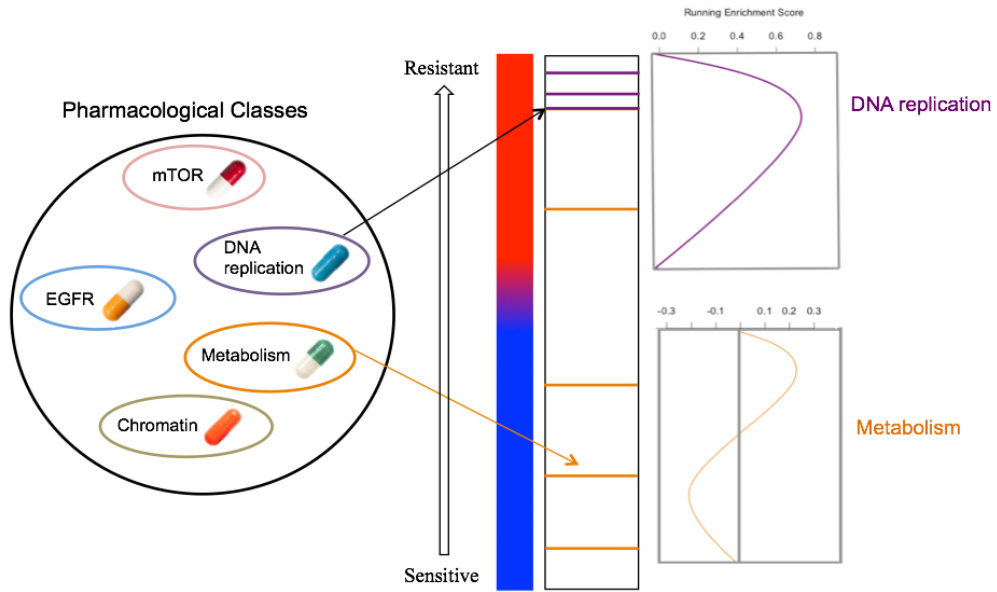
Supplementary Figure 8: Methodology for pharmacological enrichment analysis.