Published in final edited form as:

Radiat Res. 2017 June ; 187(6): 708–721. doi:10.1667/RR14655.1.

Developing Human Radiation Biodosimetry Models: Testing Cross-Species Conversion Approaches Using an *Ex Vivo* Model System

Jin G. Park^{a,1}, Sunirmal Paul^{d,1,2}, Natalia Briones^{a,3}, Jia Zeng^{a,b}, Kristin Gillis^a, Garrick Wallstrom^{a,b}, Joshua LaBaer^{a,c,4}, and Sally A. Amundson^{d,4}

^aBiodesign Center for Personalized Diagnostic, Biodesign Institute, Arizona State University, Arizona

^bDepartment of Biomedical Informatics, Arizona State University, Arizona

^cSchool of Molecular Sciences, Arizona State University, Arizona

^dCenter for Radiological Research, Columbia University Medical Center, New York

Abstract

In the event of a large-scale radiation exposure, accurate and quick assessment of radiation dose received would be critical for triage and medical treatment of large numbers of potentially exposed individuals. Current methods of biodosimetry, such as the dicentric chromosome assay, are time consuming and require sophisticated equipment and highly trained personnel. Therefore, scalable biodosimetry approaches, including gene expression profiles in peripheral blood cells, are being investigated. Due to the limited availability of appropriate human samples, biodosimetry development has relied heavily on mouse models, which are not directly applicable to human response. Therefore, to explore the feasibility of using non-human primate (NHP) models to build and test a biodosimetry algorithm for use in humans, we irradiated *ex vivo* peripheral blood samples from both humans and rhesus macaques with doses of 0, 2, 5, 6 and 7 Gy, and compared the gene expression profiles 24 h later using Agilent human microarrays. Among the dose-responsive genes in human and using non-human primate, 52 genes showed highly correlated expression patterns between the species, and were enriched in p53/DNA damage response, apoptosis and cell cycle-related genes. When these interspecies-correlated genes were used to build biodosimetry models with using NHP data, the mean prediction accuracy on non-human primate samples was about 90% within 1 Gy of delivered dose in leave-one-out cross-validation. However, tests on human samples suggested that human gene expression values may need to be adjusted prior to application of the NHP model. A “multi-gene” approach utilizing all gene values for cross-species conversion and applying the converted values on the NHP biodosimetry models, gave a leave-one-out cross-validation prediction accuracy for human samples highly comparable (up to 94%) to that for non-human primates. Overall, this study demonstrates that a robust NHP

⁴Addresses for correspondence: Biodesign Center for Personalized Diagnostic and School of Molecular Sciences, Arizona State University, Arizona; Joshua.Labaer@asu.edu; and Center for Radiological Research, Columbia University Medical Center, New York; saa2108@cumc.columbia.edu.

¹These authors contributed equally to this work.

²Currently at Department of Radiology, Rutgers New Jersey Medical School, New Jersey.

³Currently at Integrated Cancer Genomics Division, Translational Genomics Research Institute, AZ.

biodosimetry model can be built using interspecies-correlated genes, and that, by using multiple regression-based cross-species conversion of expression values, absorbed dose in human samples can be accurately predicted by the NHP model.

INTRODUCTION

Many recent studies have focused on developing and improving methods for biodosimetry as a necessary component of triage in the wake of possible large-scale radiological or nuclear exposure incident. Biological response-based estimates of absorbed radiation dose can be used to help reassure the “worried well”, to inform medical management of casualties, to help target treatment most appropriately in resource-limited situations, and for long-term risk assessment. The dicentric assay is considered to be the gold standard for radiation biodosimetry. Although new approaches to increase throughput in dicentric assay are being developed (1–3), it remains time consuming and requires sophisticated equipment and technical expertise, supporting interest in alternative approaches. These approaches include automation of DNA repair and cytogenetic assays (4), protein biomarkers (5–7) and metabolomic methods in urine (8–10) or serum (11, 12).

Gene expression changes measured in easily accessible peripheral blood (PB) samples also show promise for radiation biodosimetry, and have been developed by several groups using various models that include human blood samples irradiated *ex vivo* (13–18), blood samples from patients that were total-body irradiated (19–24) and blood samples from mice (20, 21, 25–27). None of these models alone are sufficient for use of biodosimetry in humans. The most predictive model would rely on samples from normal human populations exposed to radiation *in vivo*, but there are many limitations to obtaining these samples. The majority of humans receiving 1 Gy or more of radiation are cancer patients, raising concerns about the possible influence on the transcriptional profile of both the underlying disease and additional treatments that are almost always given prior to or concurrently with radiation. The doses and post-exposure sampling times that can be studied are restricted to the fixed treatment plans, the radiation is often targeted to small areas of the body, and may involve delivering the radiation dose into smaller multiple fractions, none of which represent a realistic scenario for a mass-exposure event.

Mice provide a convenient *in vivo* model that allows detailed dose-response testing and choice of post-exposure assay time, but they are phylogenetically removed from humans and not the best model for development of a biodosimetric test intended for humans. The *ex vivo* model, in which blood from healthy donors is irradiated outside the body and held under culture conditions to allow development of the transcriptional response, is the only human model that allows detailed acute dose-response studies. It recapitulates many of the responses seen in patients exposed *in vivo*, but cannot capture the full response of a complete organism (19, 22).

Due to their close phylogenetic relationship to humans, non-human primates (NHPs) are often considered the preferred model for translational development. However, there have been few reports using NHP models for radiation biodosimetry development. Levels of C-reactive protein and circulating interleukins (28) were assessed in NHP blood samples at

various times after irradiation and were found to correlate well with lymphocyte and neutrophil counts. Measurements of four proteins expressed in NHP plasma were also found to be informative of radiation dose at up to 7 days after exposure (29). The γ -H2AX residual-focus assay has been demonstrated to be dose dependent in NHPs, both for total-body (30) and partial-body (31) irradiations. Small molecule metabolite changes in the urine of irradiated NHPs have been compared with those seen in rodents and human patients, and showed both similarities and differences (32), underlining the need for cross-species comparisons. To our knowledge, there have been no reports of gene expression studies for radiation biodosimetry using NHPs.

As a first test of the potential usefulness of NHPs as a model for the development of gene expression based radiation biodosimetry for humans, we used the irradiated *ex vivo* blood model to compare the gene expression response in NHP and human blood exposed to a broad range of radiation doses. Expression levels of a large number of genes could predict radiation doses in both human and NHP 24 h postirradiation. In fact, 52 dose-responsive genes were found to overlap between human and NHP datasets with strong interspecies correlations. We found that even restricting the models to genes whose expression levels correlated with radiation dose in both species, NHP data models still did not provide satisfactory dose prediction of human samples when raw expression values were applied, indicating that cross-species conversion was required. By identifying the interspecies-correlated genes, the resulting simplified dataset was ideal for testing various modeling approaches of cross-species value conversion between human gene expression values and corresponding NHP values. We found these converted values could then directly support human dose prediction using NHP biodosimetry models developed across a broad dose range.

MATERIALS AND METHODS

Experimental Setup, Sample Preparation and Microarray Data Acquisition

At 24 h after irradiation *ex vivo*, using doses ranging from 2–7 Gy, RNA from nonirradiated control and irradiated human and NHP peripheral blood from both sexes was analyzed using the same microarray platform. PB samples were drawn from 5 healthy male and 10 female Chinese rhesus macaques under a protocol approved by the University of Illinois at Chicago IACUC and shipped at $20 \pm 2^\circ\text{C}$ for arrival at Columbia University in New York within 24 h of blood draw. In parallel, de-identified human PB samples were collected, under a Columbia University IRB protocol and with written informed consent, from 10 male and 10 female volunteers, into sodium citrate Vacutainer tubes (Becton Dickinson and Company, Franklin Lakes, NJ) as used for NHP samples and maintained at $20 \pm 2^\circ\text{C}$ overnight to simulate the NHP sample shipping. PB samples from both NHP and humans were divided into 2 ml aliquots and diluted 1:1 with RPMI 1640 medium (Mediatech, Herndon, VA) supplemented with 10% heat-inactivated fetal bovine serum (HyClone, Logan, UT). These aliquots were exposed to 0, 2, 5, 6 and 7 Gy Cesium-137 gamma rays at 24 h after blood draw using a Gammacell-40 ^{137}Cs irradiator (AECL, Ottawa, Ontario, Canada) with a dose rate of 0.8 Gy/min, and then placed in a 37°C humidified incubator with 5% CO_2 for 24 h.

White blood cell and differential cell counts were obtained from the cultured blood samples before irradiation and again immediately before RNA extraction at 24 h postirradiation (48 h after blood collection). The RNA extraction was done using the 5-Prime PerfectPure™ blood RNA purification kit (5 Prime, Inc., Gaithersburg, MD) as per the manufacturer's recommendations. Globin mRNA reduction was performed using human GLOBINclear™ (Ambion Inc., Austin, TX) on both human and NHP samples to prepare them for microarray analysis.

Cyanine-3 (Cy3)-labeled (cRNA) was prepared from 200 ng input RNA using Agilent's One-Color Quick Amp labeling kit (Agilent Technologies) according to manufacturer's instructions. After purification of cRNA by RNAeasy column (QIAGEN, Valencia, CA), the cRNA yield and Cy3 dye incorporation was measured with a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific). The cRNA (1.65 µg) was fragmented and hybridized to Agilent's whole genome microarrays (G4845A, Agilent Technologies) using the Agilent Gene Expression Hybridization Kit. Human samples were hybridized at 65°C according to Agilent's standard protocol; however, NHP samples were incubated at 63°C to optimize hybridization of NHP probes on the human whole genome array platform. The hybridization was performed for 17 h with recommended rotation, followed by washing according to Agilent's instructions. Microarray slides were immediately scanned with the Agilent Scanner (G2404B, Agilent Technologies) using the recommended settings. The scanned images were analyzed with Agilent Feature Extraction (v10.7) using default parameters for background correction and flagging of nonuniform features.

Data Processing, Gene Set Selection, Pathway Analysis, p53 Target Identification

The human and NHP Agilent microarray data were quantilenormalized with the `qnorm` function in *R*, which transformed the quantiles of dataset so that individual data sets had the same distribution (60). As summarized in Supplementary Fig. S1 (<http://dx.doi.org/10.1667/RR1466.1.S1>), we built and tested quadratic and isotonic linear models with the `scikit-learn`, a Python-based machine-learning software (61), for each feature in human and NHP datasets, and selected ones with dose prediction accuracy within 1.0 Gy equal to or greater than 50% to identify dose-responsive gene sets. In case of obtaining multiple features for a gene, one with the best accuracy was selected. Then, the interspecies-correlated features intersecting human and NHP dose-responsive feature sets were selected. For the heat map display, expression levels were standardized to the Z scores, and we utilized the `Scikit-learn` package to perform the principal component analysis. All plots were generated with the `matplotlib` Python package. Using the dose-responsive genes in each species and the interspecies-correlated genes, we performed pathway analyses with the DAVID tools (33), and individual terms were manually assigned to four major groups of p53/DNA damage, apoptosis, chromatin/cell cycle, and immune response.

Four publically available p53 ChIP-Seq data sets on lncRNA tumor suppressor (62), motif and track discovery on functional regulatory interactions (63), cistrome of p53 mutants (64) and enhancer regulation by p53 in response to DNA damage (65) were collected. Direct p53 targets were compiled for determining the presence of p53 targets in the selected genes.

Building Human and NHP Biodosimetry Models

Dose-predictive models were built and tested by using the scikit-learn package (61). Four linear regression [Bayesian Ridge (BR), Elastic Net (EN), Ridge (RI) and Least Squares (LS)] and four ensemble (Ada Boost, Extra Trees, Gradient Boost and Random Forests) were built with selected gene sets and tested by leave-one-out cross-validation (LOOCV). When multiple features were found for a gene, one with the highest correlation with dose (for dose-responsive genes) and between species (for interspecies-correlated genes) was used for model training. All ensemble models were built with 500 individual sub-models. To minimize overfitting and select the most informative features, optimal gene sets were selected by the backward stepwise feature selection process with cross-validation by using the *RFECV* function in scikit-learn. The prediction accuracy was measured by the percentage of samples with dose prediction within 1.0 or 0.5 Gy of true dose.

Building Human-to-NHP Cross-Species Conversion Models

For the conventional approach of single-gene linear regression, human-to-NHP regression was performed on either: 1. average values across doses in human and NHP data sets or 2. individual human values and average values across doses in NHP, where only the gene expression values for that gene were utilized to build the conversion models. We tested four linear regression (Bayesian Ridge, Elastic Net, Least Squares and Ridge) and four ensemble regression (Ada Boost, Extra Trees, Gradient Boost and Random Forest) approaches. The cross-species converted values by each model were tested with selected linear and ensemble NHP biodosimetry models, and the percent accuracies within 0.5 and 1.0 Gy of true dose were reported for both fitting and cross-validations. For comparison, human expression values were directly applied to NHP biodosimetry models without conversion.

To develop human-to-NHP conversion models with a multi-gene approach, we also evaluated four linear regression and four ensemble regression models, as mentioned above, where expression values of all genes in the biodosimetry models were used to train cross-species conversion models by multiple regressions. The cross-species converted values by each model were tested with selected linear and ensemble NHP biodosimetry models, and percent accuracies within 0.5 and 1.0 Gy of actual doses were calculated.

RESULTS

Dose Response of Irradiated Ex Vivo Peripheral Blood Cells in Humans and NHP

Cultured peripheral blood samples from human ($n = 20$ per dose) and NHP ($n = 25$ per dose) were either untreated (0 Gy) or irradiated at 2, 5, 6 and 7 Gy. mRNA expression levels were measured by microarrays at 24 h postirradiation. To identify a pool of candidate genes that could be used to build biodosimetry models, we tested individual genes for their dose predictability by building regression models and calculating the prediction accuracy within 1 Gy. Since we planned to test a variety of modeling approaches including linear and ensemble models, we selected genes with linear trends across all doses as well as within only certain dose ranges that could be utilized by decision tree-based models. Therefore, we performed both isotonic linear modeling and quadratic modeling on each feature (i.e., spots on the microarrays) and selected genes with fitting accuracy of $\geq 50\%$ in either model for the

initial pool of biomarker candidates (outlined in Supplementary Fig. S1; <http://dx.doi.org/10.1667/RR1466.1.S1>). In cases where there were multiple microarray features for a single gene, we selected the feature with the best accuracy. This resulted in the identification of 210 human and 800 NHP dose-responsive genes, among which *PCNA* (84% accuracy) and *RASGRP2* (73% accuracy) were the best performing individual human and NHP genes, respectively.

To further refine the gene set to be used for testing cross-species conversion approaches in this study, we calculated interspecies Pearson correlation coefficients (R) of the means across doses for the 53 intersecting genes between the human and NHP dose-predictive gene sets. For this gene set, 52 genes with $R > 0.9$ were selected as the interspecies-correlated genes. (Supplementary Table S1; <http://dx.doi.org/10.1667/RR1466.1.S2>). With these 52 genes (36 up-regulated and 16 down-regulated after increasing radiation dose; Fig. 1A and B for human and NHP, respectively), we then performed principal component analysis (PCA) to examine the sample groupings by this gene set, which showed that the 5, 6 and 7 Gy samples were grouped together, but separated well from 0 and 2 Gy samples in both human (Fig. 1C) and NHP (Fig. 1D). Three apparent outliers (one each for 0, 2 and 7 Gy irradiated samples, 3 total out of 125) in the NHP dataset were removed from the figures and from the data sets used for testing various development approaches, so as to not skew the results for determining the most accurate methods to incorporate into the cross-species biodosimeter.

Notably, when compared to the genes selected based on prediction accuracy, different sets of genes were identified, often with small fold changes (e.g. <2 folds) and P values over 0.05 in t tests on 0 vs. 7 Gy samples (volcano plots, Supplementary Fig. S1; <http://dx.doi.org/10.1667/RR1466.1.S1>). Therefore, if a conventional method of selecting differentially expressed genes based on fold change and P values derived from statistical tests (e.g., t test) were used, this subset of genes, which proved to be informative in predicting dose, would have been missed.

Biological Functions of Dose-Responsive Genes

We performed gene ontology/pathway enrichment analyses of the dose-responsive genes in each species and the interspecies-correlated genes using the DAVID tool (33) with GO biological processes plus PANTHER (34), KEGG (35) and REACTOME (36) pathway terms (Supplementary Table S2; <http://dx.doi.org/10.1667/RR1466.1.S2>). Dose-responsive human and NHP genes were commonly enriched with pathway terms related to three major functions: p53/DNA damage response, apoptosis and cell cycle (Fig. 2). This underscores the strong concordance of biological responses to radiation between humans and NHP. There were also differences between the pathways that predominated between the two species. Most notably, immune responses were exclusively enriched in NHP. However, when we inspected the pathway annotation of individual genes, the interspecies-correlated genes encompassed all four major functions including immune responses. More specifically, four genes (*BAX*, *EOMES*, *TNFSF4* and *EDA2R*) among the 52 interspecies-correlated genes were associated with immune responses, although the enrichment as a gene set was not statistically significant (adjusted $P = 0.23$, Supplementary Table S3; <http://dx.doi.org/10.1667/RR1466.1.S2>). Since p53 is a key driver of radiation responses, we collected

publically available ChIP-Seq datasets on p53 binding in various cell lines and identified 25 direct transcriptional targets of p53 among the 52 interspecies-correlated genes. Interestingly, all four immune-related genes (*BAX*, *EOMES*, *TNFSF4* and *EDA2R*) were among the p53 targets (Supplementary Table S3A), indicating potentially a broader role of p53 in mediating cellular response in peripheral blood cells upon radiation through direct regulation of immune-related genes. Overall, these interspecies-correlated genes spanning multiple radiation-related pathways represent an optimal gene set from which to build *ex vivo* biodosimetry models and test cross-species conversion approaches.

Building Human and NHP Ex Vivo Biodosimetry Models with Species-Specific and Interspecies-Correlated Dose-Responsive Genes

For development of a robust and widely applicable approach of utilizing NHP biodosimetry models to predict radiation dose in humans, we tested various types of regression models using our *ex vivo* data as a model system. Throughout this study, in search of optimal modeling approaches, we tested eight widely used machine-learning methods, including four linear (Bayesian Ridge, Elastic Net, Ridge and Least Squares) and four ensemble (Ada Boost, Extra Trees, Gradient Boost and Random Forests) regressions. To improve robustness of the conventional Least Square model (Ordinary Least Square), Ridge (Tikhonov regularization) and Bayesian Ridge (37, 38) regressions apply constraints on the size of coefficients, in which the latter utilizes Bayesian inference. Elastic Net (39) combines Ridge with the Lasso (40) model that is used for sparse learning. Unlike linear models that make predictions based on a single model, ensemble models generate multiple sub-models and make the final prediction by combining multiple predictions from sub-models, which provides greater robustness by reducing overfitting. Extra Trees (41) and Random Forests (42) generate a collection of decision trees based on random subsets of samples and features (i.e., genes in our models) and average the predicted values of individual decision trees (also known as bootstrap aggregating or bagging). Extra Trees, an extension of Random Forests, involves an additional randomization step in choosing the optimal decision splits. Two boosting models, Ada Boost (43) (AB or Adaptive Boosting) and Gradient Boost (44) (GB or Gradient Tree Boosting), iteratively add improved decision tree models by assigning differential weights on individual samples towards better performance on incorrectly predicted samples in previous trees (AB) or minimization of overall prediction errors of ensemble of decision trees (GB).

First, we constructed eight dosimetry models based on the entire set of dose-responsive genes in each species and tested for dose-prediction accuracy within 1.0 Gy or 0.5 Gy of delivered radiation dose by a stringent LOOCV test. By utilizing all 210 dose-responsive human genes, we observed a broad range of accuracies (51–98% within 1 Gy for EN and RI models, Supplementary Table S4A; <http://dx.doi.org/10.1667/RR1466.1.S2>), with linear regression models (except EN) generally outperforming ensemble models. To identify an optimal set of genes for individual models, we performed a backward stepwise feature selection with embedded cross-validations, which eliminated iteratively the least informative gene until the maximal performance and robustness was achieved. With these feature-selected gene sets, ranging from 4 to 113 genes, four linear and four ensemble models were re-trained and tested by LOOCV (Fig. 3A and Supplementary Table S4A). The BR and LS

models showed the best overall performance with 100% accuracies within both 1.0 Gy and 0.5 Gy in LOOCV testing. When models were built and tested by LOOCV with 800 dose-responsive genes for the NHP data set, we observed slightly lower accuracies for both the full sets of dose-predictive genes (mean accuracy across 8 models within 1 Gy of 76% in NHP vs. 85% in human, Supplementary Table S4) and the feature-selected sets (87% in NHP vs. 89% in human, Fig. 3A and B and Supplementary Table 4). Overall, the LS and AB models showed the highest accuracies among linear and ensemble models, respectively, for both human and NHP.

Although these species-specific models represent optimal biodosimetry models for each species, to develop robust cross-species conversion approaches, we then built NHP biodosimetry models with 52 interspecies-correlated genes using the same four linear and four ensemble regression methods, on which various cross-species conversion approaches could be subsequently tested. Using the entire gene set, the LOOCV accuracies ranged from 42% (for EN) to 91% (for BR and RI) within 1 Gy of true doses (Fig. 3D and Supplementary Table S4D; <http://dx.doi.org/10.1667/RR1466.1.S2>). After recursive feature selection, the BR (24 genes) and RI (23 genes) models had the highest accuracy of 96% within 1 Gy among all models, while AB (15 genes, 87% accuracy within 1Gy) and ET (11 genes, 85% accuracy within 1Gy) outperformed the other ensemble models. In comparison, when we analyzed human datasets with the same set of 52 interspecies-correlated genes, we observed slightly better performance (mean accuracy of 86% across 8 models within 1 Gy) than NHP (mean accuracy of 82%) (Fig. 3C and Supplementary Table S4C).

Despite exclusion of a large fraction (93.5% or 748 among 800 genes) of NHP-specific dose-responsive genes during the selection for interspecies-correlation to build optimal cross-species biodosimetry models, the dose-prediction accuracy decreased by only an average of 5% (87% to 82% with 800 or 52 genes, respectively) after feature selection. This implied that the selected set of 52 interspecies-correlated genes enriched in radiation-related pathways of p53/DNA damage, apoptosis, cell cycle and immune response still included genes that were highly informative in predicting dose. As shown in Supplementary Table S3 (<http://dx.doi.org/10.1667/RR1466.1.S2>), genes that were most frequently selected by the feature selection were not preferentially linked to particular pathways among the four major pathway groups. To explore methods to utilize NHP biodosimetry models to predict absorbed radiation dose in humans, as a proof-of-concept, we selected four of the best performing NHP biodosimetry models including the two linear models (BR and LS) and two ensemble models (AB and ET), and utilized them to test various cross-species conversion approaches. The performance of the BR and AB dosimetry models, representing the linear and ensemble models, respectively, are shown in Figs 4, 5 and 6. Performance of the other models tested are shown in Supplementary Figs. S3, S4, and S5 (<http://dx.doi.org/10.1667/RR1466.1.S1>).

Building Human-to-NHP Conversion Models: Conventional Approaches

Gene expression trends across doses of dose/interspecies-correlated genes were very similar for human and NHP responses; whereas, the magnitude and shape of dose-response curves were mostly distinct between the species, as well as across genes (Supplementary Fig. S2;

<http://dx.doi.org/10.1667/RR1466.1.S1>). These data suggest that directly using human expression values with NHP biodosimetry models will result in sub-optimal performance. Indeed, when applying unconverted human values to the four selected NHP biodosimetry models, we observed prediction accuracies within 1.0 Gy from 39% (for the ET model) to only up to 54% (for the BR model). These accuracies were much lower than the benchmark accuracies on NHP values with an average of 90% accuracy (Fig. 3D), ranging from 87% to 96% within 1.0 Gy in LOOCV across 4 models (referred to as “Human Unconverted” Fig. 4; Supplementary Materials, Fig. S3 and Table S5; <http://dx.doi.org/10.1667/RR1466.1.S2>). These results clearly indicate that expression values of human genes should be converted to corresponding NHP values for accurate dose prediction, even for this directly comparable *ex vivo* model system.

One simple approach to cross-species conversion is a linear transformation of each gene value [referred as single-gene conversion models hereafter (Fig. 5)], such as multiplying human values of each gene with a single-gene-specific conversion factor (e.g., a ratio of mean expression values). Although this average-based approach resulted in substantial increases in accuracies for all four biodosimetry models [ranging from 67% to 75% within 1 Gy (referred to as “Ave” in Fig. 4); Supplementary Fig. S3 and Table S5], all models still showed much lower performance than the original NHP biodosimetry models with accuracies around 90% in LOOCV. In addition, when we performed cross-species conversion of all human expression values against the mean NHP values for the corresponding dose by applying conventional regression methods (i.e., LS), we also observed lower LOOCV accuracies (43–64% within 1 Gy, (referred to as “Regr” in Fig. 4; Supplementary Fig. S3; <http://dx.doi.org/10.1667/RR1466.1.S1> and Supplementary Table S5; <http://dx.doi.org/10.1667/RR1466.1.S2>) than for NHP models. Overall, cross-species conversion by these conventional conversion approaches resulted in a substantial reduction of dose-prediction accuracy from those of original NHP biodosimetry models based on dose and interspecies correlated genes.

Building Human-to-NHP Conversion Models: A Multi-Gene-Based Regression Approach

Since conventional approaches for converting gene expression values by linear regressions did not provide robust cross-species conversion for the biodosimetry models, as an alternative, we developed and tested a multi-gene conversion approach, where expression values of all genes (versus one gene of interest) in NHP biodosimetry models were used in combination to convert expression values of each gene (Fig. 5). This concept of a multiple regression (or linear combination of genes) approach has been applied successfully to impute missing expression values in large transcriptomics data sets (45, 46). In the context of irradiation *ex vivo*, the feature-selected genes in biodosimetry models function within closely related biological pathways in response to radiation. Therefore, we hypothesized that there were certain degrees of correlation among expression levels of the genes, and that expression values of other genes might be informative in cross-species conversion and could potentially improve robustness of conversion models.

To test if this approach improved cross-species conversion of human expression values, we constructed four linear and four ensemble conversion models by utilizing all feature-selected

genes in each of four selected NHP biodosimetry models and first compared the accuracy of converting expression values themselves with those of the conventional single gene-based conversion models. While the conventional approaches of conversion (referred to as “Ave” and “Regr” in Fig. 6A and Supplementary Fig. S4; <http://dx.doi.org/10.1667/RR1466.1.S1>) showed 78% and 86% mean accuracies in LOOCV of human-to-NHP conversion within log2-transformed expression value of 0.5 from the NHP values, multi-gene conversion models showed >99% accuracies, with the exception of EN (Fig. 6A and Supplementary Fig. S4).

One potential drawback of multi-gene-based models is that conversion accuracies can be affected by any fluctuation in gene expression values of highly informative genes in predicting the expression values of a given gene, which can be caused by both biological and technical variability. Thus, to test robustness of conversion, we iteratively introduced artificial random errors (± 1.0 of log2-transformed expression values) for randomly selected genes, and tested how the conversion accuracies were affected in single and multi-gene-based conversion models. Interestingly, as shown in Fig. 6A and Supplementary Fig. S4 (<http://dx.doi.org/10.1667/RR1466.1.S1>), we observed that multi-gene-based approaches were more robust (a 2.6% point decrease in accuracy from LOOCV) even in the presence of random errors, when compared to single-gene-based models (a 9% point decrease).

Next, we applied human values converted by various cross-species conversion models onto four selected NHP biodosimetry models and tested for dose-prediction accuracy by LOOCV. In general, when compared to single gene-based models, multi-gene conversion models markedly increased the prediction accuracies within 1.0 Gy by 17–30% across all tested models (Fig. 7 and Supplementary Fig. S5; <http://dx.doi.org/10.1667/RR1466.1.S1>), excluding the EN model, where the combination of the BR biodosimetry model with the AB cross-species conversion models showed the highest LOOCV accuracy of 94% within 1.0 Gy (Fig. 7 and Supplementary Table S5; <http://dx.doi.org/10.1667/RR1466.1.S2>). Among multi-gene conversion models, we observed that ensemble conversion models generally outperformed linear conversion models across all four biodosimetry models. As previously discussed, since any noise or variation in expression measurements of heavily weighted genes could affect conversion of all genes, we generated 1,000 randomized datasets, each with artificial noise (± 1.0 in log2-transformed data) in expression values of a randomly selected gene, and then examined the effect on dose prediction. Even with the introduction of randomized errors, all dosimetry models suffered only marginal decreases (an average of 8.7% points) in accuracy (Supplementary Table S5). These results indicate that utilizing expression values of other functionally related genes during cross-species conversion provided greater robustness in the presence of potential measurement errors.

DISCUSSION

Gene ontology and pathway enrichment analysis showed that four major categories of cellular functions (p53/DNA damage pathway, apoptosis, chromatin assembly/cell cycle and immune response) were enriched among genes that were regulated in a dose-dependent fashion in response to ionizing radiation irradiation *ex vivo* in human and NHP models (Fig. 2 and Supplementary Tables S2 and S3; <http://dx.doi.org/10.1667/RR1466.1.S2>). Many of

the same biological processes and pathways, notably functions related to DNA damage response and p53 pathways, were commonly enriched in human and NHP dose-responsive genes and interspecies-correlated genes. For example, the response to DNA damage stimulus (GO, 0006974) term was highly enriched in human (adjusted $P = 8.96\text{E-}7$) and NHP (adjusted $P = 5.05\text{E-}8$) dose-responsive genes, as well as in interspecies-correlated genes (adjusted $P = 1.05\text{E-}8$). Similarly, the p53 signaling pathway (KEGG, hsa04115) was enriched in human (adjusted $P = 1.11\text{E-}9$) and NHP (adjusted $P = 1.99\text{E-}5$) dose-responsive genes and interspecies-correlated genes (adjusted $P = 6.5\text{E-}13$). The vast majority of the DNA damage-response genes showing dose responses in both human and NHP blood cells (*CDKN1A*, *MDM2*, *PCNA*, *PPM1D*, *ZMAT3*, *DDB2*, *CCNG1*, *GAD-D45A*, *BBC3*, *AEN*, *RPS27L* and *POLH*) are known to be transcriptionally regulated by p53 in response to stresses such as ionizing radiation (Supplementary Table S3). Such genes have been reported by multiple groups as top candidates for use in biodosimetry (23, 47–54). However, because the p53 response is not unique to ionizing radiation exposure, a robust dose-prediction model should also include genes from a broad spectrum of pathways. Supporting this, the most informative and frequently utilized genes in feature-selected biodosimetry models based on interspecies-correlated genes consisted of genes from all the 4 major pathway groups (Supplementary Table S3; <http://dx.doi.org/10.1667/RR1466.1.S2>).

In contrast to DNA damage-response genes, there was more interspecies variation in the specific genes with apoptosis, chromatin/cell cycle and immune-response-related functions. Apoptosis-related genes were more enriched in human dose-responsive genes, while chromatin-assembly and immune-response genes were enriched only in NHP dose-responsive genes. Importantly, when we selected interspecies-correlated genes to build cross-species conversion models, genes in all four major pathways were included. The chromatin-packaging related genes, revealed by the gene ontology analysis, may be attractive for radiation biodosimetry in this regard. About half the dose-responsive genes that were associated with these functions were conserved between humans and NHP, providing a strong basis for gene selection. Histone gene expression has previously been reported to be down-regulated by stresses including chemical inhibition of DNA synthesis (55) and ionizing radiation (56–58). Although some of the response of histones to ionizing radiation appears to be p53 dependent (57), a p53-independent down-regulation of histone genes also occurs in response to radiation in a variety of cell lines, including those of lymphoid origin (58). Coordinate down-regulation of a set of genes involved in processes of DNA packaging and mitosis was also described as the major p53-independent transcriptional response to ionizing radiation across a diverse set of 60 cancer cell lines (59). Although there has been less emphasis on the potential of histone and other DNA packaging genes for use in radiation biodosimetry, their robust and well-conserved down-regulation supports their inclusion in dose-responsive models.

In addition, the results indicated that biodosimetry models based on the interspecies-correlated genes (i.e., a subset of dose-responsive genes) had comparable dose-prediction accuracies to those based on the full set of dose-responsive genes each of both species (Fig. 3 and Supplementary Table S4; <http://dx.doi.org/10.1667/RR1466.1.S2>). Therefore, the interspecies-correlated gene sets were utilized to build biodosimetry models and to test cross-species conversion approaches in this study. Overall, excluding the EN model (42%

accuracy within 1 Gy in LOOCV), NHP biodosimetry models built with feature-selected interspecies-correlated genes by both linear and ensemble regression approaches showed comparable performance (average accuracy of 93% and 83% within 1 Gy, respectively) in LOOCV testing (Fig. 3 and Supplementary Table S4). The weak performance of EN model might be due to the inclusion of a “lasso”-like sparse learning term. Indeed, when we built biodosimetry models using a lasso regression, they showed no predictive power (data not shown).

Our results clearly indicate that expression values need to be converted for accurate dose prediction in human utilizing NHP biodosimetry models. Gene expression trends across doses of interspecies-correlated genes were very similar for human and NHP response, while the magnitude and shape of dose-response curves were mostly distinct between the species as well as across genes (Supplementary Fig. S2; <http://dx.doi.org/10.1667/RR1466.1.S1>). Directly applying human expression values to NHP biodosimetry models resulted in sub-optimal performances with prediction accuracies within 1.0 Gy from 0% to only up to 27% for AB model (Fig. 7 and Supplementary Table S5; <http://dx.doi.org/10.1667/RR1466.1.S2>). On the other hand, multi-gene conversion models increased the accuracies for within 1.0 Gy by 29–45% across all tested biodosimetry models when compared to single gene-based conversion models. This indicates that expression values of other genes with radiation-related functions are informative in cross-species conversion and can improve the robustness of conversion models.

Based on these results, the same principles seem promising for developing other cross-species dataset conversion models for potential irradiated *in vivo* sample modeling. We plan to further explore this approach in future studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project has been funded in whole or in part with federal funds from the Biomedical Advanced Research and Development Authority, office of the Assistant Secretary for Preparedness and Response, Office of the Secretary, Department of Health and Human Services under Contract No. HHS01201000008C

References

1. Flegal FN, Devantier Y, McNamee JP, Wilkins RC. Quickscan dicentric chromosome analysis for radiation biodosimetry. *Health Phys.* 2010; 98(2):276–81. [PubMed: 20065694]
2. Gruel G, Gregoire E, Lecas S, Martin C, Roch-Lefevre S, Vaurijoux A, et al. Biological dosimetry by automated dicentric scoring in a simulated emergency. *Radiat Res.* 2013; 179(5):557–69. [PubMed: 23560627]
3. Vaurijoux AGG, Pouzoulet F, Gregoire E, Martin C, Roch-Lefevre S, Voisin P, Voisin P, Roy L. Strategy for population triage based on dicentric analysis. *Radiat Res.* 2009; 171(5):541–8. [PubMed: 19580489]
4. Garty G, Chen Y, Salerno A, Turner H, Zhang J, Lyulko O, et al. The RABIT: a rapid automated biodosimetry tool for radiological triage. *Health Phys.* 2010; 98(2):209–17. [PubMed: 20065685]

5. Ivey RG, Subramanian O, Lorentzen TD, Paulovich AG. Antibody-based screen for ionizing radiation-dependent changes in the Mammalian proteome for use in biodosimetry. *Radiat Res.* 2009; 171(5):549–61. [PubMed: 19580490]
6. Kim D, Marchetti F, Chen Z, Zaric S, Wilson RJ, Hall DA, et al. Nanosensor dosimetry of mouse blood proteins after exposure to ionizing radiation. *Sci Rep.* 2013; 3:2234. [PubMed: 23868657]
7. Sharma M, Moulder JE. The urine proteome as a radiation biodosimeter. *Adv Exp Med Biol.* 2013; 990:87–100. [PubMed: 23378004]
8. Chen C, Brenner DJ, Brown TR. Identification of urinary biomarkers from X-irradiated mice using NMR spectroscopy. *Radiat Res.* 2011; 175(5):622–30. [PubMed: 21338244]
9. Laiakis EC, Mak TD, Anizan S, Amundson SA, Barker CA, Wolden SL, et al. Development of a metabolomic radiation signature in urine from patients undergoing total body irradiation. *Radiat Res.* 2014; 181(4):350–61. [PubMed: 24673254]
10. Tyburski JBPA, Krausz KW, Slavík J, Fornace AJ Jr, Gonzalez FJ, Idle JR. Radiation metabolomics. 1. Identification of minimally invasive urine biomarkers for gamma-radiation exposure in mice. *Radiat Res.* 2008; 170(1):1–14. [PubMed: 18582157]
11. Goudarzi M, Weber WM, Mak TD, Chung J, Doyle-Eisele M, Melo DR, et al. Metabolomic and lipidomic analysis of serum from mice exposed to an internal emitter, cesium-137, using a shotgun LC-MS(E) approach. *J Proteome Res.* 2015; 14(1):374–84. [PubMed: 25333951]
12. Laiakis EC, Strassburg K, Bogumil R, Lai S, Vreeken RJ, Hankemeier T, et al. Metabolic phenotyping reveals a lipid mediator response to ionizing radiation. *J Proteome Res.* 2014; 13(9):4143–54. [PubMed: 25126707]
13. Gruel G, Voisin P, Vaurijoux A, Roch-Lefevre S, Gregoire E, Maltere P, et al. Broad modulation of gene expression in CD4+ lymphocyte subpopulations in response to low doses of ionizing radiation. *Radiat Res.* 2008; 170(3):335–44. [PubMed: 18763857]
14. Mori M, Benotmane MA, Tirone I, Hooghe-Peters EL, Desaintes C. Transcriptional response to ionizing radiation in lymphocyte subsets. *Cell Mol Life Sci.* 2005; 62(13):1489–501. [PubMed: 15971001]
15. Paul S, Amundson SA. Gene expression signatures of radiation exposure in peripheral white blood cells of smokers and non-smokers. *Int J Radiat Biol.* 2011; 87(8):791–801. [PubMed: 21801107]
16. Paul S, Ghandhi SA, Weber W, Doyle-Eisele M, Melo D, Guilmette R, et al. Gene expression response of mice after a single dose of 137CS as an internal emitter. *Radiat Res.* 2014; 182(4):380–9. [PubMed: 25162453]
17. Paul SAS. Development of gene expression signatures for practical radiation biodosimetry. *Int J Radiat Oncol Biol Phys.* 2008; 71(4):1236–44. [PubMed: 18572087]
18. Amundson SA, Do KT, Shahab S, Bittner M, Meltzer P, Trent J, et al. Identification of potential mRNA biomarkers in peripheral blood lymphocytes for human exposure to ionizing radiation. *Radiat Res.* 2000; 154(3):342–6. [PubMed: 11012342]
19. Amundson SAGM, McLeland CB, Epperly MW, Yeager A, Zhan Q, Greenberger JS, Fornace AJ Jr. Human in vivo radiation-induced biomarkers: gene expression changes in radiotherapy patients. *Cancer Res.* 2004; 64(18):6368–71. [PubMed: 15374940]
20. Dressman HKMG, Chao NJ, Meadows S, Marshall D, Ginsburg GS, Nevins JR, Chute JP. Gene expression signatures that predict radiation exposure in mice and humans. *PLoS Med.* 2007; 4(4):e106. [PubMed: 17407386]
21. Lucas J, Dressman HK, Suchindran S, Nakamura M, Chao NJ, Himburg H, et al. A translatable predictor of human radiation exposure. *PLoS One.* 2014; 9(9):e107897. [PubMed: 25255453]
22. Meadows SKDH, Muramoto GG, Himburg H, Salter A, Wei Z, Ginsburg GS, Chao NJ, Nevins JR, Chute JP. Gene expression signatures of radiation response are specific, durable and accurate in mice and humans. *PLoS One.* 2008; 3(5):e1912. [PubMed: 18382685]
23. Paul SBC, Turner HC, McLane A, Wolden SL, Amundson SA. Prediction of in vivo radiation dose status in radiotherapy patients using ex vivo and in vivo gene expression signatures. *Radiat Res.* 2011; 175(3):257–65. [PubMed: 21388269]
24. Templin T, Paul S, Amundson SA, Young EF, Barker CA, Wolden SL, et al. Radiation-induced micro-RNA expression changes in peripheral blood cells of radiotherapy patients. *Int J Radiat Oncol Biol Phys.* 2011; 80(2):549–57. [PubMed: 21420249]

25. Meadows SK, Dressman HK, Daher P, Himburg H, Russell JL, Doan P, et al. Diagnosis of partial body radiation exposure in mice using peripheral blood gene expression profiles. *PLoS One*. 2010; 5(7):e11535. [PubMed: 20634956]
26. Ossetrova NI, Blakely WF. Multiple blood-proteins approach for early-response exposure assessment using an in vivo murine radiation model. *Int J Radiat Biol*. 2009; 85(10):837–50. [PubMed: 19863200]
27. Tucker JD, Grever WE, Joiner MC, Konski AA, Thomas RA, Smolinski JM, et al. Gene expression-based detection of radiation exposure in mice after treatment with granulocyte colony-stimulating factor and lipopolysaccharide. *Radiat Res*. 2012; 177(2):209–19. [PubMed: 22128785]
28. Ha CT, Li XH, Fu D, Moroni M, Fisher C, Arnott R, et al. Circulating interleukin-18 as a biomarker of total-body radiation exposure in mice, minipigs, and nonhuman primates (NHP). *PLoS One*. 2014; 9(10):e109249. [PubMed: 25290447]
29. Ossetrova NI, Sandgren DJ, Blakely WF. Protein biomarkers for enhancement of radiation dose and injury assessment in nonhuman primate total-body irradiation model. *Radiat Prot Dosimetry*. 2014; 159(1–4):61–76. [PubMed: 24925901]
30. Redon CE, Nakamura AJ, Gouliava K, Rahman A, Blakely WF, Bonner WM. The use of gamma-H2AX as a biodosimeter for total-body radiation exposure in non-human primates. *PLoS One*. 2010; 5(11):e15544. [PubMed: 21124906]
31. Redon CE, Nakamura AJ, Gouliava K, Rahman A, Blakely WF, Bonner WM. Q(gamma-H2AX), an analysis method for partial-body radiation exposure using gamma-H2AX in nonhuman primate lymphocytes. *Radiat Meas*. 2011; 46(9):877–81. [PubMed: 21949480]
32. Johnson CHPA, Krausz KW, Kalinich JF, Tyburski JB, Kang DW, Luecke H, Gonzalez FJ, Blakely WF, Idle J. Radiation metabolomics. 5. Identification of urinary biomarkers of ionizing radiation exposure in nonhuman primates by mass spectrometry-based metabolomics. *Radiat Res*. 2012; 178(4):328–40. [PubMed: 22954391]
33. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4(1):44–57. [PubMed: 19131956]
34. Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res*. 2010; 38(Database issue):D204–10. [PubMed: 20015972]
35. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*. 2014; 42(Database issue):D199–205. [PubMed: 24214961]
36. Croft D, Mundo AF, Haw R, Milacic M, Weiser J, Wu G, et al. The Reactome pathway knowledgebase. *Nucleic Acids Res*. 2014; 42(Database issue):D472–7. [PubMed: 24243840]
37. Hsiang TC. Bayesian View on Ridge Regression. *Statistician*. 1975; 24(4):267–8.
38. Landram FG. Bayesian Approach to Ridge Regression. *Oper Res*. 1975; 23 B393–B.
39. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J Roy Stat Soc B*. 2005; 67:301–20.
40. Tibshirani R. Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B Met*. 1996; 58(1): 267–88.
41. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. 2006; 63(1):3–42.
42. Breiman L. Random forests. *Mach Learn*. 2001; 45(1):5–32.
43. Freund Y, Schapire RE. A decision-theoretic generalization of online learning and an application to boosting. *J Comput Syst Sci*. 1997; 55(1):119–39.
44. Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann Stat*. 2001; 29(5):1189–232.
45. Kim H, Golub GH, Park H. Missing value estimation for DNA microarray gene expression data: local least squares imputation. *Bioinformatics*. 2005; 21(2):187–98. [PubMed: 15333461]
46. Aittokallio T. Dealing with missing values in large-scale studies: microarray data imputation and beyond. *Brief Bioinform*. 2010; 11(2):253–64. [PubMed: 19965979]
47. Paul S, Amundson SA. Development of gene expression signatures for practical radiation biodosimetry. *Int J Radiat Oncol Biol Phys*. 2008; 71(4):1236–44. [PubMed: 18572087]

48. Turtoi A, Brown I, Oskamp D, Schneeweiss FH. Early gene expression in human lymphocytes after gamma-irradiation-a genetic pattern with potential for biodosimetry. *Int J Radiat Biol.* 2008; 84(5):375–87. [PubMed: 18464067]
49. Filiano AN, Fathallah-Shaykh HM, Fiveash J, Gage J, Cantor A, Kharbanda S, et al. Gene expression analysis in radiotherapy patients and C57BL/6 mice as a measure of exposure to ionizing radiation. *Radiat Res.* 2011; 176(1):49–61. [PubMed: 21361780]
50. Pogosova-Agadjanyan EL, Fan W, Georges GE, Schwartz JL, Kepler CM, Lee H, et al. Identification of radiation-induced expression changes in nonimmortalized human T cells. *Radiat Res.* 2011; 175(2):172–84. [PubMed: 21268710]
51. Brzoska K, Kruszewski M. Toward the development of transcriptional biodosimetry for the identification of irradiated individuals and assessment of absorbed radiation dose. *Radiat Environ Biophys.* 2015; 54(3):353–63. [PubMed: 25972268]
52. Li MJ, Wang WW, Chen SW, Shen Q, Min R. Radiation dose effect of DNA repair-related gene expression in mouse white blood cells. *Med Sci Monit.* 2011; 17(10):BR290–7. [PubMed: 21959603]
53. Riecke A, Rufa CG, Cordes M, Hartmann J, Meineke V, Abend M. Gene expression comparisons performed for biodosimetry purposes on in vitro peripheral blood cellular subsets and irradiated individuals. *Radiat Res.* 2012; 178(3):234–43. [PubMed: 22769026]
54. Tucker JD, Joiner MC, Thomas RA, Grever WE, Bakhmutsky MV, Chinkhota CN, et al. Accurate gene expression-based biodosimetry using a minimal set of human gene transcripts. *Int J Radiat Oncol Biol Phys.* 2014; 88(4):933–9. [PubMed: 24444760]
55. Heintz N, Sive HL, Roeder RG. Regulation of human histone gene expression: kinetics of accumulation and changes in the rate of synthesis and in the half-lives of individual histone mRNAs during the HeLa cell cycle. *Mol Cell Biol.* 1983; 3(4):539–50. [PubMed: 6406835]
56. Datta R, Weichselbaum R, Kufe DW. Ionizing radiation downregulates histone H1 gene expression by transcriptional and post-transcriptional mechanisms. *Radiat Res.* 1993; 133(2):176–81. [PubMed: 8438058]
57. Su C, Gao G, Schneider S, Helt C, Weiss C, O'Reilly MA, et al. DNA damage induces downregulation of histone gene expression through the G1 checkpoint pathway. *EMBO J.* 2004; 23(5):1133–43. [PubMed: 14976556]
58. Meador JA, Ghandhi SA, Amundson SA. p53-independent downregulation of histone gene expression in human cell lines by high- and low-let radiation. *Radiat Res.* 2011; 175(6):689–99. [PubMed: 21520998]
59. Amundson SA, Do KT, Vinikoor LC, Lee RA, Koch-Paiz CA, Ahn J, et al. Integrating global gene expression and radiation survival parameters across the 60 cell lines of the National Cancer Institute Anticancer Drug Screen. *Cancer Res.* 2008; 68(2):415–24. [PubMed: 18199535]
60. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics.* 2003; 19(2):185–93. [PubMed: 12538238]
61. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res.* 2011; 12:2825–30.
62. Sanchez Y, Segura V, Marin-Bejar O, Athie A, Marchese FP, Gonzalez J, et al. Genome-wide analysis of the human p53 transcriptional network unveils a lncRNA tumour suppressor signature. *Nat Commun.* 2014; 5:5812. [PubMed: 25524025]
63. Janky R, Verfaillie A, Imrichova H, Van de Sande B, Standaert L, Christiaens V, et al. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput Biol.* 2014; 10(7):e1003731. [PubMed: 25058159]
64. Schlereth K, Heyl C, Krampitz AM, Mernberger M, Finkernagel F, Scharfe M, et al. Characterization of the p53 cistrome–DNA binding cooperativity dissects p53's tumor suppressor functions. *PLoS Genet.* 2013; 9(8):e1003726. [PubMed: 23966881]
65. Younger ST, Kenzelmann-Broz D, Jung H, Attardi LD, Rinn JL. Integrative genomic analysis reveals widespread enhancer regulation by p53 in response to DNA damage. *Nucleic Acids Res.* 2015; 43(9):4447–62. [PubMed: 25883152]

FIG. 1.

Heat map and PCA plots of interspecies-correlated genes in irradiated *ex vivo* samples.

Expression patterns of 52 interspecies-correlated genes in human (panel A) and NHP (panel B) datasets. PCA plots based on the 52 genes are shown for human (panel C) and NHP (panel D). Three NHP outliers were excluded in all plots.

FIG. 3.

LOOCV performance of feature-selected NHP and human *ex vivo* biodosimetry models.

Panel A: Human dose-responsive gene-based models were tested on human data. Panel B: NHP dose-responsive gene-based models were tested on NHP data. Panel C: Interspecies-correlated gene-based models were tested on the human data set. Panel D: Interspecies-correlated gene-based models were tested on the NHP data set. BR= Bayesian Ridge, EN = Elastic Net, LS: Least Squares, RI = Ridge, AB: Ada Boost, ET = Extra Tress, GB: Gradient Boost, RF = Random Forests.

FIG. 4.

Performance of NHP *ex vivo* dosimetry models with unconverted and single-gene converted human expression values. Unconverted and converted values by various conversion models applied to four selected NHP dosimetry models, and percentage accuracy dose prediction measured by LOOCV. Panel A: Bayesian Ridge. Panel B: Ada Boost. For mean-converted values (Ave) and regression-converted values (Regr) for single-gene-based conversion, fitting values are shown. BR = Bayesian Ridge, AB = Ada Boost.

FIG. 5.
Schematic overview of single-gene- and multi-gene-based cross-species conversion approaches.

FIG. 6.

Cross-species conversion accuracies by single- and multi-gene conversion models in LOOCV and randomization tests. Panel A: Conversion accuracies within 0.5 of NHP mean values are shown. Panel B: Dose-response curves before and after conversion of representative single-gene and multi-gene conversion models. Ave = mean-based conversion for each gene, Regr = linear regression of each gene, BR = Bayesian Ridge, AB = Ada Boost.

A**BR Biodosimetry Model** **B****AB Biodosimetry Model** **FIG. 7.**

Performance of NHP *ex vivo* dosimetry models with unconverted and converted human expression values. Unconverted and converted values by various conversion models applied to four selected NHP dosimetry models, and percentage accuracy dose prediction measured by LOOCV. Panel A: Bayesian Ridge. Panel B: Ada Boost. For mean-converted values (Ave) and regression-converted values (Regr) for single-gene-based conversion, fitting values are shown. BR = Bayesian Ridge, EN = Elastic Net, LS = Least Squares, RI = Ridge, AB = Ada Boost, ET = Extra Tress, GB = Gradient Boost, RF = Random Forests.