

Cross-platform validation of a mouse blood gene signature for quantitative reconstruction of radiation dose

Shanaz A. Gandhi (✉ sg2423@cumc.columbia.edu)

Columbia University Irving Medical Center

Igor Shuryak

Columbia University Irving Medical Center

Brian Ponnaiya

Columbia University Irving Medical Center

Xuefeng Wu

Columbia University Irving Medical Center

Guy Garty

Columbia University Irving Medical Center

Shad R. Morton

Columbia University Irving Medical Center

Salan P. Kaur

Columbia University Irving Medical Center

Sally A. Amundson

Columbia University Irving Medical Center

Research Article

Keywords: Mouse, radiation, biomarker, gene expression

Posted Date: February 4th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1325734/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

In the search for biological markers after a large-scale exposure of the human population to radiation, gene expression is a sensitive endpoint easily translatable to in-field high throughput applications. Primarily, the *ex-vivo* irradiated healthy human blood model has been used to generate available gene expression datasets. This model has limitations i.e., lack of signaling from other irradiated tissues and deterioration of blood cells cultures over time. *In vivo* models are needed; therefore, we present our novel approach to define a gene signature in mouse blood cells that quantitatively correlates with radiation dose (at 1 Gy/min). Starting with available microarray datasets, we selected 30 radiation-responsive genes and performed cross-validation/training-testing data splits to downselect 16 radiation-responsive genes. We then tested these genes in an independent cohort of irradiated adult C57BL/6 mice (50:50 both sexes) and measured mRNA by quantitative RT-PCR in whole blood at 24h. Dose reconstruction using net Signal (difference between geometric means of top 3 positively correlated and top 4 negatively correlated genes with dose), was highly improved over the microarrays, with a root mean square error of ± 1.04 Gy in male and female mice combined. There were no significant sex-specific differences in mRNA or cell counts after irradiation.

Introduction

In the wake of a radiation event that results in the exposure of many in the human population to radiation, there will be a need for rapid dosimetry for medical management and triage for treatment¹⁻⁵. Biological dosimetry has been shown to be very useful in this regard and the best-characterized methodology is the gold standard cytogenetic measurement of double strand breaks using the dicentric assay^{4,6}. Other assays are being developed, tested, and validated with the potential of shortening the time-to-result. Molecular assays, such as gene expression measurements, are among the most promising approaches to bring this critical time down from days to a few hours after sampling⁴. They also reduce the large-scale equipment and space needed for high throughput cytogenetic biodosimetry^{7,8}. Many laboratories have tested gene expression in human blood to develop panels of gene targets in mRNA, microRNA, and non-coding RNA towards this goal of reconstructing exposure⁹⁻¹³. Most of these studies, however, have been performed on human blood irradiated in cell cultures¹⁴⁻²⁰ which is a powerful method mimicking much of the *in vivo* human blood response in an experimental set up^{16,21}. However, for investigating longer term *in vivo* responses, the human blood culture approach has its limitations, most importantly, degeneration of the health of blood cells over several days in culture. Thus, to study longer times after irradiation, we need *in vivo* model systems. Large and small animal models have been proposed for this purpose and used extensively following the Animal rule^{22,23} to study responses at longer times after irradiation, and enable studies of protracted exposures, internal emitters and more complicated scenarios involving partial body exposures²⁴⁻²⁷.

Here, we will describe our approaches using a small animal model – the house mouse *Mus musculus* (C57BL/6N, Charles River Laboratory), for developing and validating gene expression signatures,

specifically at 24 h after acute radiation exposures. Many laboratories, including ours, have published extensively on gene expression in mouse blood after irradiation (as recently reviewed ^{9,21}), but most of these studies focus on classifying samples as exposed vs. unexposed, or above/below a threshold 2 Gy dose. We present an approach wherein we identify a set of genes that can quantitatively reconstruct doses between 0 and 8 Gy in a continuous rather than discrete manner.

We started with meta-analysis of transcriptomic microarray data in mouse blood at 24 h after irradiation, training and testing (cross validation using random separation of the data) to select about 30 genes that correlated strongly with dose. From this gene set we selected a smaller target set of genes to test transition to the quantitative real-time PCR platform for validation studies. Here, we used an independent cohort of animals and used a similar approach to that used to develop our human radiation-responsive gene expression signature ²⁸. In this approach, quantitative dose reconstruction was performed based on the difference in median signals of sets of genes that were positively or negatively correlated with radiation dose. This small-scale systems biology approach was used for dimension reduction to refine and then validate our dose reconstruction signature that consists of 7 transcripts.

Results

Meta-analysis of transcriptomic data and gene selection for dose reconstruction

For this analysis, we compiled datasets from microarray analyses published by our group and others as listed in Table 1. The maximum dose was 10 Gy and all studies used a similar acute dose rate for the irradiations, around 1 Gy/min, as described in the respective publications (see Table 1). All studies used the wild-type C57BL/6 strain. For these analyses we compiled normalized data using BRB ArrayTools ²⁹, and generated a table with genes as rows and samples as columns with the signal intensities of corresponding genes. We identified genes that showed a significant dose response, either positively or negatively correlated with dose. This approach resulted in a set of genes that were ranked by correlation to dose and to each other. We selected a set of 20 up regulated genes (Figure 1A) and 10 down regulated genes (Figure 1B) with strong dose correlations at 24 h after acute irradiation for further analyses (net_signal vs. dose are plotted in Figure 2).

Using these genes for dose reconstruction on microarray data a robust dose response curve was generated (Figure 2), and then used for the model for dose reconstruction (Figure 3). Biological analysis using gene ontology and pathway analyses suggested that most of these genes are known to be involved in the DNA damage response to stress and radiation in the blood, and therefore are good candidates to measure and reconstruct dose in mice.

Next, we assessed the stability of the dose reconstruction model by 1000 random splits of the data into training/testing halves. The results of this approach gave an R^2 value of 0.695 (standard deviation, SD 0.0388). When the data were randomly split into a training-testing set, the root mean squared error (RMSE) for the training set was 1.86 Gy (SD 0.122) and RMSE for the test set was similar at 1.9 Gy (SD

0.126). This shows that the model fits were relatively stable over random splits, but the dose reconstruction quality on these microarray data was not greatly improved ($R^2 < 0.7$, RMSE ~ 1.9 Gy). Especially the day 1 samples were reconstructed with a larger error ($R^2 < 0.6$, RMSE ~ 2.2 Gy). However, from these calculations we obtained a list of dose-responsive genes (up and down) that were relatively stable over time (Supplementary data table 1 lists all the genes with annotations). A subset of these genes was selected for qPCR analyses using an independently irradiated cohort of adult mice.

Blood counts and Quantitative RT-PCR analysis in an independent mouse cohort

At necropsy, an aliquot of blood was used for immunophenotyping, measuring levels of leucocytes, T and B cells and neutrophils (Figure 4). Mouse blood cell-count data were \log_n -transformed and analyzed by linear regression, fitted using a generalized least squares method (to allow variance to depend on a power of the fitted values) using the *gls* function in *R*. The goal was to characterize the dose responses for B and T cells, and to assess potential differences in dose response parameters between males and females. Sex was included as a binary variable (Sex, males=1, females=0) and as an interaction term (Sex×Dose). There were no significant differences between males and females by this analysis; the Sex and Sex×Dose terms did not achieve statistical significance for either B-cells or T-cells. Supplementary Figure 1A shows the fitted parameters, along with the dose response of T cells (Supplementary Figure 1B), and B cells (Supplementary Figure 1C).

Isolated RNA was processed to cDNA and further to measure mRNA changes for genes that were selected based on signal intensity in the microarrays. For both up and down regulated groups of genes, to ensure that we were inclusive across a range of sensitivity, from low to high copy number RNA, we chose genes based on signal groupings as follows: high copy number transcripts signal intensity $>10,000$; low copy number transcript signal intensity <2000 ; and an intermediate group $2000 < \text{signal} < 10,000$. These genes were: *Ccng1* (cyclin G1), *Aen* (apoptosis enhancing nuclease1), *Sgta* (small glutamine-rich tetratricopeptide repeat (TPR)-containing, alpha), *Grn* (granulin), *Ccnd1* (cyclin D1), *Phlda3* (pleckstrin homology-like domain, family A, member 3), *Rhoc* (ras homolog gene family, member C), *Xdh* (xanthine dehydrogenase), *Bax* (BCL2-associated X protein), *Cd5l* (CD5 antigen-like), *Tcn2* (transcobalamin 2) and *Lrg1* (leucine-rich alpha-2-glycoprotein 1) in the up regulated gene group. *Cd19* (CD19 antigen), *Cxcr5* (chemokine (C-X-C motif) receptor 5), *Ly6D* (lymphocyte antigen 6 complex, locus D) and *Ccr7* (chemokine (C-C motif) receptor 7) were in the down regulated group. Geometric mean of *Actb* (actin B) and *Gapdh* (glyceraldehyde-3-phosphate dehydrogenase) were used for all analyses.

Further DCt analyses showed that *Phlda3* [correlation coefficients of -0.6(M), -0.7(F)], *Rhoc* [correlation coefficients of -0.5(M), -0.7(F)], and *Lrg1* [correlation coefficients of -0.6(M), -0.6(F)] were best correlated with dose among the up regulated genes tested. *Cxcr5* [correlation coefficients of 0.9(M), 0.8(F)], *Cd19* [correlation coefficients of 0.9(M), 0.9(F)], *Ly6D* [correlation coefficients of 0.9(M), 0.9(F)] and *Ccr7* [correlation coefficients of 1.0(M), 0.9(F)] gave the best dose correlations among the down regulated genes. Results of our qRT-PCR analyses for these genes with the best fold change to dose correlations combining male and female animals are shown in Figure 5. Corresponding best-fit parameter

values for the multiple regression model are shown in Table 2 (relative expression of these 7 genes compared with sham control is shown separately for males and females in Supplementary Figure 2).

Our assessments of potential sex effects (Table 2) suggested that all terms containing Sex were not statistically significant, and they were dropped from the model (dose reconstruction for males and females is shown separately in Supplementary Figure 3). The retained simpler regression model (containing only N and N²) is described in Table 3. The R² for dose reconstruction from this model was 0.84 and RMSE was 1.04 Gy (Figure 6). These performance metric values suggest an improvement with qRT-PCR measurements compared to the “training” microarray data analysis where RMSE values were ~1.9 Gy, as described above.

Discussion

Previous work from our group and others has emphasized the use of the mouse model as a tractable system for discovery and validation of radiation biodosimetry endpoints that cannot be studied directly using healthy humans^{9,21,30-33}. With the caveat that there are species specific³⁴ and even strain specific differences^{35,36} in radiation response, we trained an algorithm using the transcriptomics datasets available in the NCBI Gene Expression Omnibus database for quantitative radiation responses at 24 h. From this analysis we selected the top dose-correlated genes (positive and negative) and trained and tested our method using the net signal (N) as a correlate of dose. Typically, gene signatures that have potential for field applications are viewed to consist of a few genes to tens of genes³⁷⁻³⁹. Therefore, we started with 30 top ranked genes for quantitative RT-PCR testing, then further reduced the number of genes to a lower threshold, beyond which addition of more genes would not dramatically change the average error measured by RMSE.

Gene ontology using the PANTHER database for the 30 top genes from microarray data analyses suggested enrichment of biological processes related to the intrinsic apoptotic signaling pathway in response to DNA damage by p53 (Bonferroni corrected p-value <0.012, genes *Phlda3*, *Aen*, *Bax* and *Rps27l*) and leukocyte activation (Bonferroni corrected p-value <0.003, genes *Ly6D*, *Cxcr5*, *Ccr7*, *Cd19* and *Gm*). Further detailed analysis of the gene ontology tree indicated leukocyte activation processes specifically related to B-cell activation and differentiation (Bonferroni corrected p-value <0.013) were enriched. Pathway analysis using Ingenuity (IPA) also suggested that p53 signaling (p-value <0.0013), IL-7 signaling (p-value <0.0025) and interferon signaling (p-value <0.0009) were top signaling pathways affected by these genes. These results of biological functions represented in our signature are congruent with the major processes represented among human gene expression changes after irradiation in blood^{14,16} based on *ex vivo* studies. These responses, as well as p53 and DNA damage response at the mRNA level, are well established in the field of radiation biology. The interesting point here is that some of these genes are correlated with cell development and may provide an indication not only of dose but also increased damage to specific blood cell subpopulations after radiation. Additionally, the observation that immune cell changes, such as activation and differentiation are enriched biological functions among the

top genes, indicates that the *in vivo* model captures additional information about the radiation response, which may correlate with cell depletion in the blood after increasing doses. B-cell markers *Cd19* and *Ly6D* are highly down regulated after irradiation in blood and may be indicative of cell-number changes in this cell subpopulation, shown to be radio-sensitive to acute and protracted irradiation^{34,40,41}.

Next, we processed an independent cohort of male and female C57BL/6 adult mice to measure these genes using the real time qRT-PCR platform. The range of doses between 0 and 8 Gy, was similar to that in the microarray studies and whole-body irradiations were delivered at a conventional dose rate of 1 Gy/min. We measured cell counts and mRNA changes in blood at 24h after the acute dose irradiations and compared the results and dose reconstruction across the two platforms. Both T and B cell counts in the blood dropped significantly at all irradiation doses compared with controls, however, no differences were observed by sex, even at the highest 8 Gy dose (Mann-Whitney p-values comparing T and B cell count vs. dose for males and females was ~0.04 at only one dose). The 7-gene signature used in this study were down selected from the results of qRT-PCR experiments and are *Phlda3* (pleckstrin homology-like domain, family A, member 3), *Rhoc* (ras homolog gene family, member C) and *Lrg1* (leucine-rich alpha-2-glycoprotein 1) in the up regulated gene group. *Cd19* (CD19 antigen), *Cxcr5* (chemokine (C-X-C motif) receptor 5), *Ly6D* (lymphocyte antigen 6 complex, locus D) and *Ccr7* (chemokine (C-C motif) receptor 7) were in the down regulated group. Among these genes *Phlda3* is a well-known radiation response gene in both normal human and cancer cells⁴²⁻⁴⁴ and a known target of p53⁴⁵. *Lrg1* protein in blood has been detected as an early survival biomarker after radiotherapy for squamous cell carcinoma⁴⁶⁻⁴⁸ and as a serum biomarker for radiation exposure⁴⁹. *Cd19* is a commonly used biomarker for detection of B cell development and diagnosis of leukemia⁵⁰. *Ly6D* mRNA expression is induced by ATM and p53 after radiation in MCF10A cells⁵¹ and is also a surface protein involved in B cell lineage development⁵². *Cxcr5* mRNA is up regulated in T cells after antigen exposure driven by radiation-resistant cells⁵³ but a role in response to radiation has not been established. *Ccr7* mRNA is down regulated in response to radiation in dendritic cells and may influence cancer cell migration^{54,55}. Together, these mRNAs make a blood-based signature in response to radiation that is highly reproducible for dose reconstruction of acute radiation exposure with the low error of 1 Gy with the same accuracy in both sexes. Also, we established that although -omics approaches are best at discovery, testing-training, and dimension reduction; moving to a more streamlined and targeted platform such as the well-established and widely used PCR assays gave a significant improvement of both measurement of individual genes and application of the dose reconstruction algorithm (compared in Supplementary data Figure 4). We also compared Pearson's correlation R^2 values measured by the microarray and qRT-PCR platforms for each gene, and for all down-regulated genes there was a significant difference (adjusted p-value < 0.05) in qRT-PCR measurements, but no significant differences for up regulated genes.

Our approach was to start with microarray transcriptomics data, casting a wide net to capture the best radiation response genes for continuous dose reconstruction, then move to a qRT-PCR platform for independent validation, and finally narrow these down to <10 genes that may be translatable to the field. This signature can form the basis for further experiments using the mouse model, simulating more

complicated exposure scenarios with realistic parameters such as dose rate, presence of neutrons, partial body exposure and internal emitters. Such studies may provide indications for enhancement of the core 7-gene signature. Eventually, the combined and expanded signature must be applied to the human condition and homologous genes identified that can be used in simulated experiments.

In summary, we present here our data mining approach for identifying and independently validating a set of genes that can be used to quantitatively reconstruct dose in mouse blood cells 24 h after whole body irradiation at the dose rate of 1 Gy/min. We took advantage of the -omic data available in the literature, collated them together and statistically determined the best candidates for quantitative dose reconstruction. We tested genes on a different platform, the gold standard real time qRT-PCR, using available Taqman assays and improved the dose reconstruction by reducing error, going from an RMSE of 1.9 Gy with microarray data to 1.0 Gy with qRT-PCR. We also showed that a few genes were sufficient for dose reconstruction, 3 up- and 4 down- regulated mRNA used in combination. This core 7-mRNA signature can be further expanded to test other parameters of radiation exposure in this small animal model system.

Methods

Statistical meta-analysis of microarray data

Gene expression datasets were selected based on a literature search for the terms “radiation”, “biodosimetry” and “mouse blood” on the NCBI gene expression omnibus website (<https://www.ncbi.nlm.nih.gov/geo/>) and are listed in Table 1 with the GEO accession numbers for uploaded datasets and publication PMIDs. We started by pooling the mouse microarray data sets using normalized and averaged signal intensities, dropping missing values wherever possible. Doses were as indicated in Table 1 (details in Supplementary data table 1). For some datasets only one non-zero irradiation dose was available, and in other cases a dose range. After merging all datasets, we retained $n = 90$ of 0 Gy samples; $n = 14$ of 1 Gy samples; $n = 50$ of 2 Gy samples; $n = 59$ of 3 Gy samples; $n = 11$ of 4 Gy samples; $n = 60$ of 6 Gy samples; $n = 15$ of 7 Gy samples; $n = 6$ of 8 Gy samples and $n = 48$ of 10 Gy samples.

Using the compiled data sets, we wrote customized code in the *R* programming language to look for genes positively or negatively correlated with radiation dose (based on Spearman’s correlation coefficients). We focused on genes that were stable over time (1-7 days, wherever the data was available) after irradiation by calculating the correlation coefficients with dose for all time points combined for each gene. We retained for further analysis only those genes whose correlation coefficients with dose were statistically significant (p -values < 0.05 after Bonferroni correction).

Since the retained genes were often strongly correlated with each other and tended to have similar dose response shapes, we did not treat each gene as a separate predictor of dose, but instead sought to combine gene signals to reduce the data dimensionality and increase the robustness of resulting dose reconstructions. We selected the top 20 genes with positive correlation with dose, and the top 10 genes

negatively correlated with dose and calculated median signals in each of these two groups. The difference between group medians was treated as the net signal (N), which was then used for dose reconstruction. Sensitivity analysis showed that choosing somewhat different numbers of genes in each group (e.g., 10 and 10, 30 and 30) did not substantially change the correlation of the net signal with dose.

From these calculations we obtained a list of dose-responsive genes (up and down) that were relatively stable over time (Supplementary data table 1 lists all the genes with annotations). A subset of these genes was then selected based on signal intensity range/relative copy number, for qRT-PCR analyses using an independently irradiated cohort of adult mice. For dose reconstruction, we fitted the following simple function by robust regression (using the *nprob* function in the R programming language), where D is dose (Gy), N is the net gene signal (difference between median signals of the gene groups with positive and negative correlations with radiation dose, which varied from 0 to 10 Gy), T is time (days) after irradiation (which varied from 0.25 to 7 days), and k_1 - k_3 are model parameters (see Figure 3 legend for parameter values used):

$$D = k_1 N + k_2 N^2 + k_3 \sqrt{T} \quad \text{equation (1)}$$

The structure of equation (1) was not mechanistically motivated, but empirically established based on examination of the data patterns. We evaluated alternative structures, where T was raised to the first or second powers instead of a power of $\frac{1}{2}$, and where the T terms acted multiplicatively rather than additively. We also evaluated fitting these models by an ordinary least squares' procedure (*nls* function in R) instead of by the robust procedure. This empirical exploration suggested that the structure of equation (1) fitted by a robust algorithm generated the best performance metrics during random training/testing splits of the data, which are discussed in the Results section.

Animal irradiations and sampling

All animal husbandry and experimental procedures were conducted in accordance with applicable federal and state guidelines and approved by the Animal Care and Use Committees of Columbia University (Assurance Number: A3007-01). C57BL/6NCrl mice were purchased from Charles River Laboratories (Wilmington, MA). For validation of gene expression using the real-time qPCR method, we irradiated 5 male and 5 female young adult (~12 week) C57BL/6 mice to 0, 1, 2, 3, 4, 6 and 8 Gy of x rays using an XRAD-320 Biological Irradiator (Precision X-ray, North Branford, CT) at a dose rate of 1 Gy/min at the Center for Radiological Research. The irradiator is equipped with a custom-made Thoraeus filter (1.25 mm Sn, 0.25 mm Cu, 1.5 mm Al, HVL 4mm Cu) and dose rate from the X-Rad-360 is calibrated periodically using a factory-calibrated Accudose 10x6-6 Ionization Chamber. 24 h after irradiation, we collected blood using cardiac puncture following euthanasia with CO₂ and split the blood for blood counts and RNA isolation. The majority volume of blood was added to 4X volume of PAX solution (Becton Dickinson, NJ) and inverted 10X before freezing at -80°C. After overnight freezing, the samples were thawed at room temperature and left for >2 h before processing for total RNA isolation using the

PAXgene® Blood RNA kit (Qiagen, Valencia, CA) according to the manufacturer's protocol. Isolated RNA was quantified using the Nanodrop One spectrophotometer (ThermoFisher) and $A_{260/280}$ ratios recorded (average yields of RNA and metrics are shown in Supplementary data table 2).

A 20 mL aliquot of whole blood was processed for immunophenotyping using a CytoFLEX flow cytometer (Beckman Coulter Inc., Brea, CA). Cell counts were quantified by standard flow cytometry methods, using antibodies specific to mouse blood cell surface antigens as follows: Neutrophils (Biolegend, catalog# 127627, Brilliant Violet 421™ anti-mouse Ly-6G), WBC (Biolegend, catalog#103115, APC/Cy7 anti-mouse CD45), B cells (Biolegend catalog#115508, PE anti-mouse CD19) and T cells (Biolegend catalog#100312 APC anti-mouse CD3ε). Flow cytometry data were analyzed with CytExpert 2.3 (Beckman Coulter Inc., Brea, CA).

qRT-PCR analysis of dose reconstruction genes in mouse blood

We prepared complimentary DNA (cDNA) from 1 microgram total mRNA using the High-Capacity® cDNA Kit (Life Technologies, Foster City, CA). Quantitative real-time RT-PCR (qRT-PCR) was performed for the selected genes using Taqman® assays (Life Technologies) using the most 3' assays as follows: *Ccng1* (cyclin G, Mm00438084_m1), *Aen* (apoptosis enhancing nuclease1, Mm00471554_m1), *Sgta* (small glutamine-rich tetratricopeptide repeat (TPR)-containing, alpha, Mm00458535_m1), *Grn* (granulin, Mm00433848_m1), *Ccnd1* (cyclin D1, Mm00432359_m1), *Phlda3* (pleckstrin homology-like domain, family A, member 3, Mm00449846_m1), *Rhoc* (ras homolog gene family, member C, Mm00455906_m1), *Xdh* (xanthine dehydrogenase, Mm00442110_m1), *Bax* (BCL2-associated X protein, Mm00432051_m1), *Cd5l* (CD5 antigen-like, Mm00437567_m1), *Tcn2* (transcobalamin 2, Mm00443660_m1) and *Lrg1* (leucine-rich alpha-2-glycoprotein 1, Mm01278767_m1) in the up regulated gene group. *Cd19* (CD19 antigen, Mm00515420_m1), *Cxcr5* (chemokine (C-X-C motif receptor 5, Mm00432086_m1), *Ly6D* (lymphocyte antigen 6 complex, locus D, Mm00521959_m1) and *Ccr7* (chemokine (C-C motif) receptor 7, Mm00432608_m1) were in the down regulated group. These genes were selected based on range of signal intensity/relative transcript copy number from the arrays. Geometric mean of *Actb* (actin B) and *Gapdh* (GAPDH glyceraldehyde-3-phosphate dehydrogenase) were used for all analyses. All PCRs were performed in duplicate using 15 ng cDNA as input for standard PCR conditions. Relative fold-induction was calculated by the $2^{-\Delta\Delta CT}$ method using Expression Suite software (ThermoFisher) and Microsoft Excel software⁵⁶.

We calculated the net signal (N; difference between average signals for genes with positive and negative correlations with dose) for the selected qPCR genes and relative quantification was calculated based on delta Ct measurements (using geometric mean of two stable housekeeping genes, *Actb* and *Gapdh*). In the N calculations we used geometric mean instead of median for the up and down-regulated gene groups because this marginally improved the correlation with dose. We then performed a multiple linear regression to reconstruct dose using the following predictor variables: N, N^2 , Sex (female=0, male=1), Sex \times N, and Sex \times N^2 .

GO and pathway analyses

We used PANTHER database gene expression tools to analyze the gene lists generated by our meta-analysis of the microarray datasets⁵⁷. Top biological processes and functions enriched among these genes were identified as those with Bonferroni corrected p-value <0.05. We also uploaded the lists of mouse dose response genes to Ingenuity Pathway Analysis® Software (IPA from Ingenuity®: <http://www.ingenuity.com>) and performed core analysis for top pathways and biological functions. We also compared gene lists using the Venny tool⁵⁸.

References

1. Bushberg, J. T. *et al.* Nuclear/radiological terrorism: emergency department management of radiation casualties. *The Journal of emergency medicine* **32**, 71–85, doi:10.1016/j.jemermed.2006.05.034 (2007).
2. Flood, A. B. *et al.* Advances in a framework to compare bio-dosimetry methods for triage in large-scale radiation events. *Radiat Prot Dosimetry* **159**, 77–86, doi:10.1093/rpd/ncu120 (2014).
3. Flood, A. B. *et al.* A Framework for Comparative Evaluation of Dosimetric Methods to Triage a Large Population Following a Radiological Event. *Radiat Meas* **46**, 916–922, doi:10.1016/j.radmeas.2011.02.019 (2011).
4. Sullivan, J. M. *et al.* Assessment of biodosimetry methods for a mass-casualty radiological incident: medical response and management considerations. *Health Phys* **105**, 540–554, doi:10.1097/HP.0b013e31829cf221 (2013).
5. Grace, M. B. *et al.* Rapid radiation dose assessment for radiological public health emergencies: roles of NIAID and BARDA. *Health Phys* **98**, 172–178, doi:10.1097/01.HP.0000348001.60905.c0 (2010).
6. Steinhauser, G., Brandl, A. & Johnson, T. E. Comparison of the Chernobyl and Fukushima nuclear accidents: a review of the environmental impacts. *Sci Total Environ* **470-471**, 800–817, doi:10.1016/j.scitotenv.2013.10.029 (2014).
7. Repin, M., Pampou, S., Garty, G. & Brenner, D. J. RABiT-II: A Fully-Automated Micronucleus Assay System with Shortened Time to Result. *Radiat Res* **191**, 232–236, doi:10.1667/rr15215.1 (2019).
8. Garty, G. *et al.* The RABiT: a rapid automated biodosimetry tool for radiological triage. II. Technological developments. *Int J Radiat Biol* **87**, 776–790, doi:10.3109/09553002.2011.573612 (2011).
9. Amundson, S. A. Transcriptomics for radiation biodosimetry: progress and challenges. *Int J Radiat Biol*, 1–9, doi:10.1080/09553002.2021.1928784 (2021).
10. Templin, T., Amundson, S. A., Brenner, D. J. & Smilenov, L. B. Whole mouse blood microRNA as biomarkers for exposure to gamma-rays and (56)Fe ion. *Int J Radiat Biol* **87**, 653–662, doi:10.3109/09553002.2010.549537 (2011).
11. Templin, T. *et al.* Radiation-induced micro-RNA expression changes in peripheral blood cells of radiotherapy patients. *Int J Radiat Oncol Biol Phys* **80**, 549–557, doi:10.1016/j.ijrobp.2010.12.061

- (2011).
12. Aryankalayil, M. J. *et al.* Radiation-Induced Long Noncoding RNAs in a Mouse Model after Whole-Body Irradiation. *Radiat Res* **189**, 251–263, doi:10.1667/rr14891.1 (2018).
 13. Aryankalayil, M. J. *et al.* Microarray analysis of miRNA expression profiles following whole body irradiation in a mouse model. *Biomarkers* **23**, 689–703, doi:10.1080/1354750x.2018.1479771 (2018).
 14. Ghandhi, S. A., Smilenov, L. B., Elliston, C. D., Chowdhury, M. & Amundson, S. A. Radiation dose-rate effects on gene expression for human biodosimetry. *BMC Med Genomics* **8**, 22, doi:10.1186/s12920-015-0097-x (2015).
 15. Park, J. G. *et al.* Developing Human Radiation Biodosimetry Models: Testing Cross-Species Conversion Approaches Using an Ex Vivo Model System. *Radiat Res*, doi:10.1667/rr14655.1 (2017).
 16. Paul, S. & Amundson, S. A. Development of gene expression signatures for practical radiation biodosimetry. *Int J Radiat Oncol Biol Phys* **71**, 1236–1244, doi:10.1016/j.ijrobp.2008.03.043 (2008).
 17. Beer, L. *et al.* High dose ionizing radiation regulates micro RNA and gene expression changes in human peripheral blood mononuclear cells. *BMC Genomics* **15**, 814, doi:10.1186/1471-2164-15-814 (2014).
 18. Dressman, H. K. *et al.* Gene expression signatures that predict radiation exposure in mice and humans. *PLoS Med* **4**, e106, doi:10.1371/journal.pmed.0040106 (2007).
 19. El-Saghire, H. *et al.* Gene set enrichment analysis highlights different gene expression profiles in whole blood samples X-irradiated with low and high doses. *Int J Radiat Biol* **89**, 628–638, doi:10.3109/09553002.2013.782448 (2013).
 20. Fachin, A. L. *et al.* Gene expression profiles in human lymphocytes irradiated in vitro with low doses of gamma rays. *Radiat Res* **168**, 650–665, doi:10.1667/rr0487.1 (2007).
 21. Lacombe, J., Sima, C., Amundson, S. A. & Zenhausern, F. Candidate gene biodosimetry markers of exposure to external ionizing radiation in human blood: A systematic review. *PLoS One* **13**, e0198851, doi:10.1371/journal.pone.0198851 (2018).
 22. Snoy, P. J. Establishing efficacy of human products using animals: the US food and drug administration's "animal rule". *Vet Pathol* **47**, 774–778, doi:10.1177/0300985810372506 (2010).
 23. Williams, J. P. *et al.* Animal models for medical countermeasures to radiation exposure. *Radiat Res* **173**, 557–578, doi:10.1667/RR1880.1 (2010).
 24. MacVittie, T. J. The MCART Consortium animal models series. *Health Phys* **103**, 340–342, doi:10.1097/HP.0b013e318261175a (2012).
 25. MacVittie, T. J., Farese, A. M. & Jackson, W., 3rd. The Hematopoietic Syndrome of the Acute Radiation Syndrome in Rhesus Macaques: A Systematic Review of the Lethal Dose Response Relationship. *Health Phys* **109**, 342–366, doi:10.1097/HP.0000000000000352 (2015).
 26. Singh, V. K., Newman, V. L., Berg, A. N. & MacVittie, T. J. Animal models for acute radiation syndrome drug discovery. *Expert opinion on drug discovery* **10**, 497–517,

- doi:10.1517/17460441.2015.1023290 (2015).
27. Gandhi, S. A. *et al.* Whole thorax irradiation of non-human primates induces persistent nuclear damage and gene expression changes in peripheral blood cells. *PloS one* **13**, e0191402 (2018).
 28. Gandhi, S. A., Shuryak, I., Morton, S. R., Amundson, S. A. & Brenner, D. J. New Approaches for Quantitative Reconstruction of Radiation Dose in Human Blood Cells. *Sci Rep* **9**, 18441, doi:10.1038/s41598-019-54967-5 (2019).
 29. Simon, R. Analysis of DNA microarray expression data. *Best Pract Res Clin Haematol* **22**, 271–282, doi:10.1016/j.beha.2009.07.001 (2009).
 30. Lucas, J. *et al.* A translatable predictor of human radiation exposure. *PLoS One* **9**, e107897, doi:10.1371/journal.pone.0107897 (2014).
 31. Paul, S. *et al.* Gene expression response of mice after a single dose of ¹³⁷Cs as an internal emitter. *Radiat Res* **182**, 380–389, doi:10.1667/rr13466.1 (2014).
 32. Gandhi, S. A. *et al.* Effect of ⁹⁰Sr internal emitter on gene expression in mouse blood. *BMC Genomics* **16**, 586, doi:10.1186/s12864-015-1774-z (2015).
 33. Paul, S., Kleiman, N. J. & Amundson, S. A. Transcriptomic responses in mouse blood during the first week after in vivo gamma irradiation. *Sci Rep* **9**, 18364, doi:10.1038/s41598-019-54780-0 (2019).
 34. Gandhi, S. A., Smilenov, L., Shuryak, I., Pujol-Canadell, M. & Amundson, S. A. Discordant gene responses to radiation in humans and mice and the role of hematopoietically humanized mice in the search for radiation biomarkers. *Sci Rep* **9**, 19434, doi:10.1038/s41598-019-55982-2 (2019).
 35. Haston, C. K. Mouse genetic approaches applied to the normal tissue radiation response. *Front Oncol* **2**, 94, doi:10.3389/fonc.2012.00094 (2012).
 36. Sproull, M., Shankavaram, U. & Camphausen, K. Comparison of Proteomic Biodosimetry Biomarkers Across Five Different Murine Strains. *Radiat Res* **192**, 640–648, doi:10.1667/RR15442.1 (2019).
 37. Huang, T. C. *et al.* Gene expression analysis with an integrated CMOS microarray by time-resolved fluorescence detection. *Biosens Bioelectron* **26**, 2660–2665, doi:10.1016/j.bios.2010.03.001 (2011).
 38. Gu, J. *et al.* Development of an integrated fingerstick blood self-collection device for radiation countermeasures. *PLoS One* **14**, e0222951, doi:10.1371/journal.pone.0222951 (2019).
 39. Ostheim, P. *et al.* Gene expression for biodosimetry and effect prediction purposes: promises, pitfalls and future directions - key session ConRad 2021. *Int J Radiat Biol*, 1–12, doi:10.1080/09553002.2021.1987571 (2021).
 40. Gandhi, S. A. *et al.* Dose and Dose-Rate Effects in a Mouse Model of Internal Exposure to ¹³⁷Cs. Part 1: Global Transcriptomic Responses in Blood. *Radiat Res*, doi:10.1667/RADE-20-00041 (2020).
 41. Heylmann, D., Rodel, F., Kindler, T. & Kaina, B. Radiation sensitivity of human and murine peripheral blood lymphocytes, stem and progenitor cells. *Biochim Biophys Acta* **1846**, 121–129, doi:10.1016/j.bbcan.2014.04.009 (2014).
 42. Amundson, S. A. *et al.* Stress-specific signatures: expression profiling of p53 wild-type and -null human cells. *Oncogene* **24**, 4572–4579, doi:10.1038/sj.onc.1208653 (2005).

43. Mukherjee, S., Laiakis, E. C., Fornace, A. J., Jr. & Amundson, S. A. Impact of inflammatory signaling on radiation biodosimetry: mouse model of inflammatory bowel disease. *BMC Genomics* **20**, 329, doi:10.1186/s12864-019-5689-y (2019).
44. Suresh Kumar, M. A. *et al.* Gene Expression in Parp1 Deficient Mice Exposed to a Median Lethal Dose of Gamma Rays. *Radiat Res* **190**, 53–62, doi:10.1667/rr14990.1 (2018).
45. Kawase, T. *et al.* PH domain-only protein PHLDA3 is a p53-regulated repressor of Akt. *Cell* **136**, 535–550, doi:10.1016/j.cell.2008.12.002 (2009).
46. Walker, M. J. *et al.* Discovery and Validation of Predictive Biomarkers of Survival for Non-small Cell Lung Cancer Patients Undergoing Radical Radiotherapy: Two Proteins With Predictive Value. *EBioMedicine* **2**, 841-850, doi:10.1016/j.ebiom.2015.06.013 (2015).
47. Wang, Y. *et al.* The Clinical Prognostic Value of LRG1 in Esophageal Squamous Cell Carcinoma. *Curr Cancer Drug Targets* **19**, 756–763, doi:10.2174/1568009619666190204095942 (2019).
48. Ouerhani, A., Chiappetta, G., Souiai, O., Mahjoubi, H. & Vinh, J. Investigation of serum proteome homeostasis during radiation therapy by a quantitative proteomics approach. *Biosci Rep* **39**, doi:10.1042/bsr20182319 (2019).
49. Huang, J. *et al.* Proteomic Profiling for Serum Biomarkers in Mice Exposed to Ionizing Radiation. *Dose Response* **17**, 1559325819894794, doi:10.1177/1559325819894794 (2019).
50. Otero, D. C., Anzelon, A. N. & Rickert, R. C. CD19 function in early and late B cell development: I. Maintenance of follicular and marginal zone B cells requires CD19-dependent survival signals. *J Immunol* **170**, 73–83, doi:10.4049/jimmunol.170.1.73 (2003).
51. Kurosawa, M. *et al.* Expression of LY6D is induced at the surface of MCF10A cells by X-ray irradiation. *Febs j* **279**, 4479–4491, doi:10.1111/febs.12034 (2012).
52. Inlay, M. A. *et al.* Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development. *Genes Dev* **23**, 2376–2381, doi:10.1101/gad.1836009 (2009).
53. Chen, X., Ma, W., Zhang, T., Wu, L. & Qi, H. Phenotypic Tfh development promoted by CXCR5-controlled re-localization and IL-6 from radiation-resistant cells. *Protein Cell* **6**, 825–832, doi:10.1007/s13238-015-0210-0 (2015).
54. Liu, C. *et al.* Gamma-ray irradiation impairs dendritic cell migration to CCL19 by down-regulation of CCR7 and induction of cell apoptosis. *Int J Biol Sci* **7**, 168–179, doi:10.7150/ijbs.7.168 (2011).
55. Cummings, R. J. *et al.* Exposure to ionizing radiation induces the migration of cutaneous dendritic cells by a CCR7-dependent mechanism. *J Immunol* **189**, 4247–4257, doi:10.4049/jimmunol.1201371 (2012).
56. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408, doi:10.1006/meth.2001.1262 (2001).
57. Mi, H., Muruganujan, A. & Thomas, P. D. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* **41**, D377-386, doi:10.1093/nar/gks1118 (2013).

58. Oliveros, J. C. An interactive tool for comparing lists with Venn Diagrams. (2007).
59. Paul, S., Smilenov, L. B., Elliston, C. D. & Amundson, S. A. Radiation Dose-Rate Effects on Gene Expression in a Mouse Biodosimetry Model. *Radiat Res* **184**, 24–32, doi:10.1667/RR14044.1 (2015).
60. Broustas, C. G., Xu, Y., Harken, A. D., Garty, G. & Amundson, S. A. Comparison of gene expression response to neutron and x-ray irradiation using mouse blood. *BMC Genomics* **18**, 2, doi:10.1186/s12864-016-3436-1 (2017).
61. Rudqvist, N. *et al.* Global Gene Expression Response in Mouse Models of DNA Repair Deficiency after Gamma Irradiation. *Radiat Res* **189**, 337–344, doi:10.1667/rr14862.1 (2018).

Declarations

Acknowledgements

We would like to acknowledge Mr. Hai Tang for assistance with mouse handling and irradiations. This study was funded by the NIAID grant U19A1067773 for the Center for High-Throughput Minimally Invasive Radiation Biodosimetry.

Author contributions

SAG, IS and SAA designed the study, performed the analyses and drafted the manuscript. BP and XW were responsible for animal work and cell counts, GG performed irradiations and dosimetry, SRM and SPK performed RNA isolations, PCR and data collection and analysis.

Competing interest statement

All authors declare that they have no competing interests.

Ethics and approval and consent to participate

All animal husbandry and experimental procedures were conducted in accordance with applicable federal and state guidelines and approved by the Animal Care and Use Committees of Columbia University (Assurance Number: A3007-01).

Data availability statement

All except 2 microarray datasets used for meta-analyses in this study are publicly available in the NCBI Gene Expression Omnibus database (<https://www.ncbi.nlm.nih.gov/geo/>) with the processed microarray data in Supplementary data table 1. The remaining two are indicated in Table 1 as accession pending approval and PMID/citation pending. The microarray datasets can be made available upon request.

Tables

Table 1. Microarray data used for meta-analyses with GEO accession numbers and citations

NCBI-GEO dataset GSE#	Dose range	PMID/Citation
124612	0 -10 Gy	31797975 / Paul et al ³³
Submitted, accession pending	0, 7 Gy	NA
62623	0 - 4 Gy	26114327 / Paul et al ⁵⁹
85323	0 - 4 Gy	28049433 / Broustas et al ⁶⁰
99176	0, 8 Gy	29351057 / Rudqvist et al ⁶¹
114142	0, 7 Gy	31046668 / Mukherjee et al ⁴³
184361	0, 7 Gy	NA
52403	0 - 6 Gy	25255453 / Lucas et al ³⁰

Table 2 Assessment of sex effects on dose reconstruction using mouse qRT-PCR data. Sex was included in the multiple regression model as a binary variable: male = 1, female = 0.

Parameter	Best-fit value	Standard error	p-value
Intercept	2.341	0.232	2.73×10^{-15}
N	-0.880	0.073	$< 2 \times 10^{-16}$
N ²	0.057	0.025	0.0259
Sex×N	-0.003	0.103	0.970
Sex×N ²	0.044	0.041	0.287
Sex	-0.289	0.370	0.438

Table 3 Best-fit parameter values for the preferred dose reconstruction model using mouse qRT-PCR data.

Parameter	Best-fit value	Standard error	p-value
Intercept	2.244	0.178	$< 2 \times 10^{-16}$
N	-0.883	0.051	$< 2 \times 10^{-16}$
N ²	0.074	0.020	0.00034

Figures

Figure 1

Correlation matrices of genes from the microarray meta-analysis that correlate with dose. **A.** Genes that are positively correlated with dose, and **B.** Genes that are negatively correlated with dose. Shown here in the matrix are pairwise correlations with each other. The size of the oval blue area indicates the correlation level, a broad oval indicates lower correlation, and the narrower oval indicates higher correlation (as indicated in the color gradient key). The table on the side shows the genes as listed and the correlation R^2 values with dose from the microarray analyses (ranging from 0.6 to 0.7 for positive correlation and -0.3 to -0.7 for negative correlation). All genes in each group are highly correlated with each other indicating that the patterns are highly similar in both the up and down regulated gene groups.

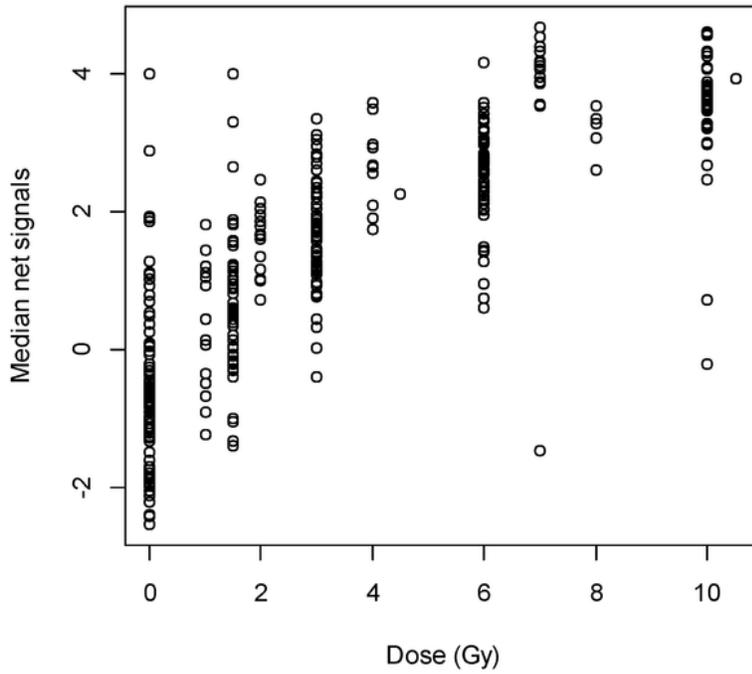


Figure 2

Normalized N vs. Dose (Gy) plots for all genes included in the microarray signature. Net signal (N) is the difference between the median of the DCt of up regulated genes and down regulated genes. This value is a better indicator of dose response than either group alone, or any one gene alone, and was used for developing the model.

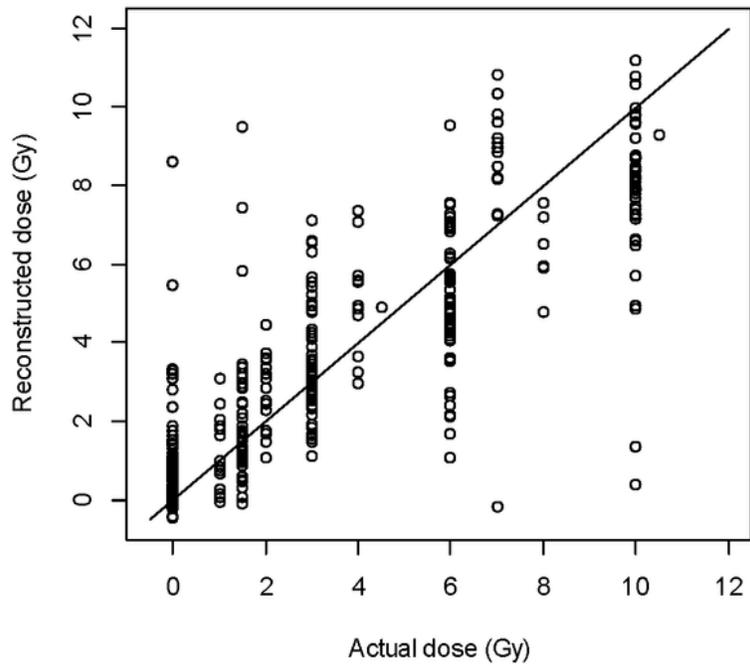


Figure 3

Dose reconstruction based on mouse microarray data analyzed using a nonlinear model equation (1). The model was based on the net gene signal (difference between median signals of the gene groups with positive and negative correlations with radiation dose), and time (days) after irradiation. The best-fit

parameter values for this dose reconstruction model were: $k_1 = 0.906$ (standard error SE = 0.086, p-value = $<2 \times 10^{-16}$) Gy^{-1} , $k_2 = 0.274$ (SE = 0.028, p-value = $<2 \times 10^{-16}$) Gy^{-2} , $k_3 = 0.549$ (SE = 0.069, p-value = 1.5×10^{-14}) $\text{days}^{-1/2}$.

Figure 4

Blood cell counts from flow cytometry analysis in an independent mouse cohort. **A.** Leucocytes (CD45+ cells)/ mL of blood, **B.** B cells (CD19+)/ mL of blood, **C.** T cells (CD3e+)/mL of blood; and **D.** Myeloid cells/ mL of blood are plotted with dose as the independent variable. All doses of irradiation produce a significant change of T and B cell numbers compared to control levels; however, B cells are more sensitive. 5 males and 5 females were used for this analysis and data is combined, as there were no differences by sex. (p-values are indicated in the figure as *p<0.05)

Figure 5

Quantitative RT-PCR results in an independent mouse cohort. Responses of male (n=5) and female mice (n=5) are plotted individually as a function of radiation dose for the 7-gene set measured by $2^{-\text{DDCt}}$ method. The response of these 7 genes was the same in both sexes and is plotted separately for male and female mice in Supplementary Figure 2.

Figure 6

Dose reconstruction in the independent mouse cohort using qRT-PCR. **A.** Plot of the best-fitting regression equation **B.** plot of reconstructed dose vs. actual dose, which was used to determine RMSE.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarydatatable1.xlsx](#)
- [Supplementarydatatable2.xlsx](#)
- [Supplementaryfigures.pdf](#)