

OPEN

New Approaches for Quantitative Reconstruction of Radiation Dose in Human Blood Cells

Shanaz A. Ghandhi^{1,2*}, Igor Shuryak^{1,2}, Shad R. Morton¹, Sally A. Amundson¹ & David J. Brenner¹

In the event of a nuclear attack or large-scale radiation event, there would be an urgent need for assessing the dose to which hundreds or thousands of individuals were exposed. Biodosimetry approaches are being developed to address this need, including transcriptomics. Studies have identified many genes with potential for biodosimetry, but, to date most have focused on classification of samples by exposure levels, rather than dose reconstruction. We report here a proof-of-principle study applying new methods to select radiation-responsive genes to generate quantitative, rather than categorical, radiation dose reconstructions based on a blood sample. We used a new normalization method to reduce effects of variability of signal intensity in unirradiated samples across studies; developed a quantitative dose-reconstruction method that is generally under-utilized compared to categorical methods; and combined these to determine a gene set as a reconstructor. Our dose-reconstruction biomarker was trained using two data sets and tested on two independent ones. It was able to reconstruct dose up to 4.5 Gy with root mean squared error (RMSE) of ± 0.35 Gy on a test dataset using the same platform, and up to 6.0 Gy with RMSE of ± 1.74 Gy on a test set using a different platform.

In the event of a nuclear attack or large-scale radiation event, there would be an urgent need for assessing the dose to which hundreds or thousands of individuals were exposed^{1–4}. Many approaches are currently being tested for radiation biodosimetry, including well-known methods based on chromosomal damage, that may take days to complete in a high-throughput manner⁴; to methods such as gene expression that can have a time-to-result of several hours. Gene expression can be easily quantified from very small blood samples. The assays can be subjected to multiplexing and modified for high-throughput or developed for integrated point-of-care approaches^{5–8}. As an added advantage gene expression has the potential to be used as an indicator or predictor of specific injury on an individualized basis.

Our group has conducted multiple transcriptomic studies of different radiation exposure conditions, leading to the identification of many radiation responsive genes that have the potential to be useful for dose reconstruction and exposure characterization. Many other studies have also used transcriptomics to identify candidates for biodosimetry. These are summarized in the review by Lacombe *et al.*⁹, in which a systematic review of available datasets identified genes that could be used for classifying samples by doses above or below 2 Gy. Other studies that have used a comparative approach to identify genes that can reconstruct dose are by Lu *et al.*¹⁰ and Macaeva *et al.*¹¹.

Most human radiation biodosimetry studies have used *ex-vivo* irradiation of blood and have focused on the simplest type of human exposure; a total body photon (x-ray or gamma-ray) exposure at an acute dose rate (~ 1 Gy/min), with mRNA levels measured at 24 h after the exposure. Once a basic dose-reconstruction signature is established to estimate dose in a quantitative manner, increasingly complex scenarios can also be considered. Gene expression holds promise for distinguishing other relevant characteristics of exposure; including dose rate^{12,13} such as from exposure to fallout¹⁴; partial shielding; and the presence of neutrons^{15,16}, which would be relevant to the blast from an improvised nuclear device¹⁷. Genes for discrimination of such dose modifying factors could be tested and added to future biodosimetric signatures.

The goal of the study presented here was to apply a rigorous statistical approach to existing transcriptomic data to develop a signature for dose reconstruction of total body photon exposures. To our knowledge, this is the first report of a continuous (non-discrete) dose reconstruction gene signature with stringent testing on

¹Columbia University Irving Medical Center, 630, W 168th street, VC11-237, New York, NY, 10032, USA. ²These authors contributed equally: Shanaz A. Ghandhi and Igor Shuryak. *email: sg2423@cumc.columbia.edu

Datasets for gene signature	NCBI-GEO dataset GSE#	Dose range	PMID/Citation
1 st Training set	8917	0.5 to 8.0 Gy	18572087 / Paul <i>et al.</i> ¹⁸
2 nd training set (normalizer genes)	90909	0.1 to 4.0 Gy	28140791 / Broustas <i>et al.</i> ¹⁵
1 st test set	65292	0.56 to 4.45 Gy	25963628 / Ghandhi <i>et al.</i> ¹²
2 nd test set	58613	1.5 to 6.0 Gy	25255453 / Lucas <i>et al.</i> ¹⁹

Table 1. *Ex vivo* irradiated Human blood gene expression datasets used in this analysis.

independent datasets for radiation biodosimetry. Our approach has several novel aspects. First, we developed a new customized normalization procedure to reduce the effects of variability of gene intensities in unirradiated control samples across different studies. Second, we performed quantitative dose reconstructions rather than the more commonly used categorical approach. Finally, by combining these two design methods and applying them to independent datasets we obtained a robust gene expression signature that can reconstruct radiation dose, and which is more accurate than other gene expression tests that have been reported in the literature.

Materials and Methods

Microarray data pre-processing and meta-analysis. In this study, we focused our analyses on human blood irradiated *ex vivo* to generate a human gene signature for dose reconstruction. We used datasets that had been generated in our group^{12,15,18} in independent studies performed at different times and using similar acute dose-rate exposures to gamma rays or x-rays as well as a study from Lucas *et al.*¹⁹, which used a similar gamma-ray dose range. First, we downloaded the datasets from the NCBI GEO database²⁰, details as in Table 1. We collated each Agilent dataset using BRB-ArrayTools²¹ and our standard import parameters described previously^{12,18}.

For the Lucas *et al.* dataset (which used an Affymetrix platform, [HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array); the default importer in BRB-ArrayTools was used to collate the data from .cel files. BRB-ArrayTools was then used to normalize the data by median array across common platforms and filter the data to remove genes with more than 20% missing values. The resulting gene sets were then exported from BRB-ArrayTools for further analysis.

Computational methods to determine and test the continuous dose reconstruction signature. *Data analysis outline.* Two independent *ex vivo* photon-irradiated human blood data sets were used for radiation-responsive signature generation (training), and two independent data sets (obtained at different times and from different donors) were used for signature testing/validation, as shown in Table 1. These data sets, which contain log₂ transformed gene signal intensities, were imported into R 3.5.1 software²², where all data analysis steps were performed. These steps were as follows:

- Using the first training data set, identify genes strongly correlated with radiation dose, which we called “signature” genes.
- Using the second training data set, identify “normalizer” genes that can reduce the effects of variability in signature gene signal intensities in unirradiated samples across different data sets.
- Test the combined signature and normalizer gene set on each testing data set by generating continuous dose reconstructions and comparing them with true dose values.

Each step of this approach is described in more detail below.

Identification of radiation-responsive signature genes. Using the first training data set (Paul *et al.* GSE8917¹⁸, dose range 0 to 8 Gy); we generated a list of genes with positive Spearman’s correlations with radiation dose, with p-values ≤ 0.05 (with Bonferroni correction) for the Spearman’s correlation coefficient. This analysis was performed using the *cor* and *cor.test* commands in R on each gene in the data set²². As an additional test for robustness to the Bonferroni correction of p-values, synthetic noise variables were added to each data set to serve as benchmarks of reconstructor performance^{23,24}. We added 40,000 synthetic noise variables per data set that were drawn from the normal distribution, and 20,000 per data set from the uniform distribution, using the same mean and SD as all real genes combined. The ratio of noise variables to real genes was approximately 3:1. The rationale for this noise injection into the data set is that only those reconstructors (genes in this case) that outperform all noise variables can be regarded as the strongest ones. Therefore, we retained only those genes for further analysis that: (a) had Bonferroni corrected p-values ≤ 0.05 for the Spearman’s correlation coefficient with dose, and (2) had Spearman’s correlation coefficient with dose values larger than those for any of the synthetic noise variables.

Because samples exposed to different doses came from the same blood donor, we fitted linear mixed-effects models for each of these significantly radiation-responsive “signature” genes to account for correlations of gene signal intensities by blood donor. The model structure contained a common dose response slope for all donors (fixed effect), but intercepts were allowed to vary by donor (random effect). In other words, the dose response was assumed to be in common for a given gene in all blood donors, but the baseline gene value in unirradiated samples was allowed to vary by donor. More complicated models with random components; for both intercepts and slopes, did

Training dataset 1: GSE8917

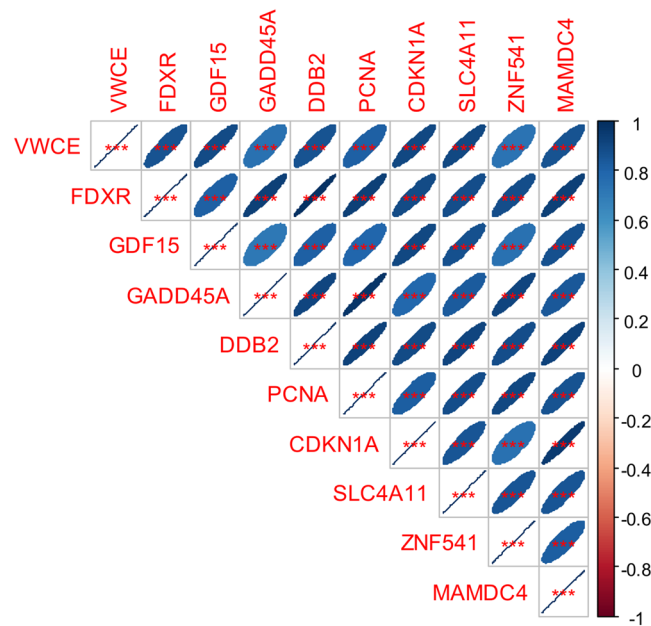


Figure 1. Matrix of Spearman's correlation coefficients (pairwise, without correction for multiple testing) between 10 selected genes with the strongest positive Spearman's correlations with dose in training data set 1 (GSE8917). A color-coded correlation scale is provided on the right of the plot. Based on the scale, blue ellipses represent positive correlations of a given gene pair, and red ones represent negative correlations. Crossed out boxes represent meaningless correlations of a given variable with itself. Darker color tones and narrower ellipses represent larger correlation coefficient magnitudes. Red star symbols indicate statistical significance levels: ***indicates $p < 0.001$, **indicates $p < 0.01$, *indicates $p < 0.05$, no stars indicates $p > 0.05$. These p-values here are intended only for visualization: due to multiple comparisons, only 3-star significance levels are likely to indicate strong associations. Among these 10 genes, which were all strongly positively correlated with dose and strongly positively correlated with each other, there were no negative correlations among gene pairs, so only blue ellipses with 3 star symbols are shown.

not converge in many cases on this data set, so they were not used. We retained only those genes with p-values ≤ 0.05 (with Bonferroni correction) for the dose-response slope parameter.

To further validate the robustness of the identified radiation-responsive gene signature list, we performed repeated k-means clustering on the same data set using the *kmeans* function in R²². There were 50 repeats, with different initial random number seeds. The goal was to identify which genes most frequently appear in the top-scoring cluster (the one with the largest Spearman's correlation coefficient with radiation dose) out of 50 clustering repeats with different initial random number seeds, and to compare these frequencies with those for synthetic noise variables. We varied the average cluster size from 30 genes per cluster to other values (e.g. 10, 50) to assess the sensitivity of the results to this parameter.

We combined the k-means clustering analysis with: (1) randomization of the outcome variable (radiation dose) by random permutation of the sample labels (doses), and (2) a scenario replacing the dose values with a linear function of a manually selected reconstruction gene- signature (a synthetic noise variable). The goal of these procedures was to assess the false positive rate (i.e. when a gene or set of genes is found to be significantly associated with the randomized dose values in scenario 1) and the sensitivity of the analysis (i.e. what effect size is required to detect the true reconstructor in scenario 2 above).

To assess the correlations between the identified radiation-responsive genes, we calculated the Spearman's correlation matrix of the genes with each other using the commands *cor* and *e* in R²². The matrix showed that, although the magnitudes of gene signal intensities varied, very strong Spearman's correlations (≥ 0.73 for any gene pair) were found, suggesting that the dose response shapes for all of these genes are very similar (Fig. 1). Consequently, treating each gene as a separate reconstructor of dose would result in severe multi-collinearity. To avoid this problem and group the genes together into one robust reconstructor, we calculated median signal values for all of these signature genes. In other words, the reconstructor value in each sample (i.e. at each dose for each blood donor) was the median value of all the radiation-responsive signature genes identified in the analyses described above (Table 2 and details in Supplementary File 1, signature genes).

We repeated all the same procedures of gene identification and robustness testing on the same training data set, but looking for genes with negative Spearman's correlations with radiation dose. However, the correlation coefficient values and subsequent dose reconstruction results for downregulated genes turned out to be much weaker than those for upregulated genes were; so genes with negative Spearman's correlations with radiation dose were not included in the final analysis (Supplementary File 1, negatively correlated genes).

Gene symbol		
ANKRA2	DRAM1	RPS27L
ANXA4	GADD45A	SESNI
ARHGEF3	GDF15	SLC4A11
ASCC3	IL21R	SLC7A6
ASTN2	LIG1	TNFRSF10B
BBC3	MAMDC4	TRIAP1
HIST1H2BD	MAP4K4	UROD
CDKN1A	PCNA	VWCE
DDB2	PHPT1	WIG1
EI24	PLK3	XPC
FBXO22	PPM1D	ZNF337
FDXR	PTP4A1	ZNF541
	REV3L	

Table 2. Signature genes for continuous dose reconstruction.

Identification of normalizer genes. In our analyses involving independent datasets, performed at different times and from different donors, we have found that baseline expression of genes in unirradiated sample is an issue that can confound the method for dose reconstruction. Although we did not observe any systematic variability in baseline levels of genes, random variability appeared to be considerable (2–4 fold between signals). To reduce the effect of these fluctuations on subsequent dose reconstructions, we developed a method that allowed us to compare gene expression between two independent datasets or platforms using normalizer genes (Supplementary File 1, normalizer genes).

In our analyses, the definition of normalizer gene is one that satisfies the following criteria:

- (1) The sum of squared differences between a potential normalizer gene's signal and median signal of the signature gene group has to be as small as possible in unirradiated control samples across ≥ 2 training data sets.
- (2) The Spearman's correlation coefficient for a potential normalizer gene's signal values with radiation dose across ≥ 2 training data sets had to be as close to zero as possible.

These criteria were set to identify genes that co-vary with radiation responsive signature genes in unirradiated samples, but do not have radiation responses themselves. We imported the second training data set (Broustas *et al.* GSE90909¹⁵, 0 to 4 Gy dose range) into R and searched for normalizer genes using both training data sets combined. Two different normalizer gene groups were found, separately for signature genes with positive and negative correlations with radiation dose. We performed normalization across both training data sets by subtracting median signal values for all normalizer genes from median signal values for all signature genes. The retained number of normalizer genes in each group (positive and negative correlations with dose) was selected to be close to the number of signature genes (Supplementary File 1, normalizer genes).

Testing the combination of signature + normalizer genes on independent data sets. To create a “standard curve” for relating normalized median gene signals to radiation dose, we fitted a polynomial model (or a robust version that down-weights outliers using the *rlm* function in R²² on data from the two training data sets combined, using dose as the dependent variable and normalized median gene signals as reconstructors (the independent variables). These models had the following structure, where k_0 , k_1 and k_2 are adjustable parameters, and S is the normalized median gene signal in each sample:

$$Dose = k_0 + k_1 \times S + k_2 \times S^4 \quad (1)$$

This structure was selected from several alternatives (e.g. different powers of S and different numbers of adjustable parameters) using the Akaike information criterion with sample size correction (AICc).

We tested the ability of these fits to reconstruct radiation dose on two independent testing data sets generated using either the same microarray platform (Ghandhi *et al.*¹² with dose range 0 to 4.45 Gy, using acute dose samples only) or a different platform (Lucas *et al.*¹⁹, with dose range 0 to 6 Gy). This was done in each data set by calculating the normalized median gene signal values (S) for the signature and normalizer gene groups for each sample. These values were then used as input for the polynomial or robust polynomial models (with original parameter values k_0 , k_1 and k_2 derived from fitting training data) to produce dose reconstructions from the testing data. Model performance was evaluated by calculating the coefficient of determination (R^2) and root mean squared error (RMSE), comparing true and estimated doses.

Pathway and network analysis and comparisons. We performed gene ontology analysis using DAVID Functional Annotation Tool, ver 6.8²⁵. We uploaded the human signature gene list to the program and performed functional annotation using biological processes 5, which are ontology-tree child terms. We also uploaded the lists of human signature genes to Ingenuity Pathway Analysis[®] Software (IPA from Ingenuity[®]: <http://www.ingenuity.com>)

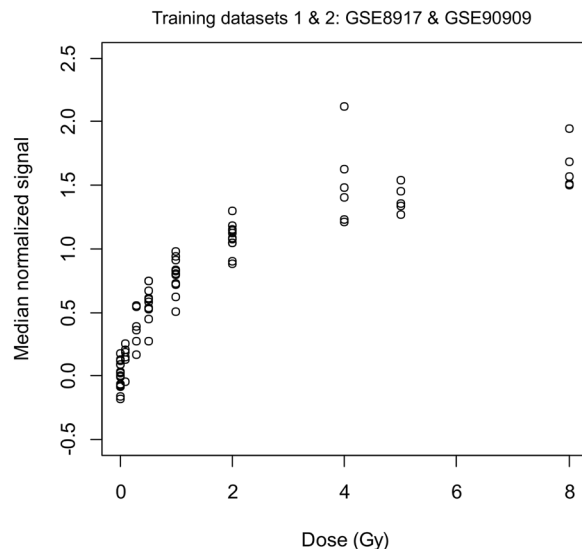


Figure 2. Microarray median normalized gene expression values of the human gene signature in both training data sets combined. Up-regulated genes that were identified using a biostatistics correlation approach in the training data sets (GSE8917 AND GSE90909) are shown in this plot, with corresponding normalizer gene values subtracted. Median log₂-transformed signal values are plotted against dose (Gy). Each dot in the graph is a single sample. Spearman's correlation coefficient with dose was 0.969.

and performed prediction analysis for upstream regulators. This method identifies potential upstream regulators and ranks them by statistical significance and provides information about the direction of activation of that regulatory protein based on the downstream gene targets from the gene list. The program provides a z score derived from the number of target genes, their relative expression, and the type of relationship between the regulator and target genes (either activation or inhibition) from the published literature. We compared gene lists using Venny²⁶.

Ethics, approval, and consent to participate. No animal or human experiments were performed as part of this study.

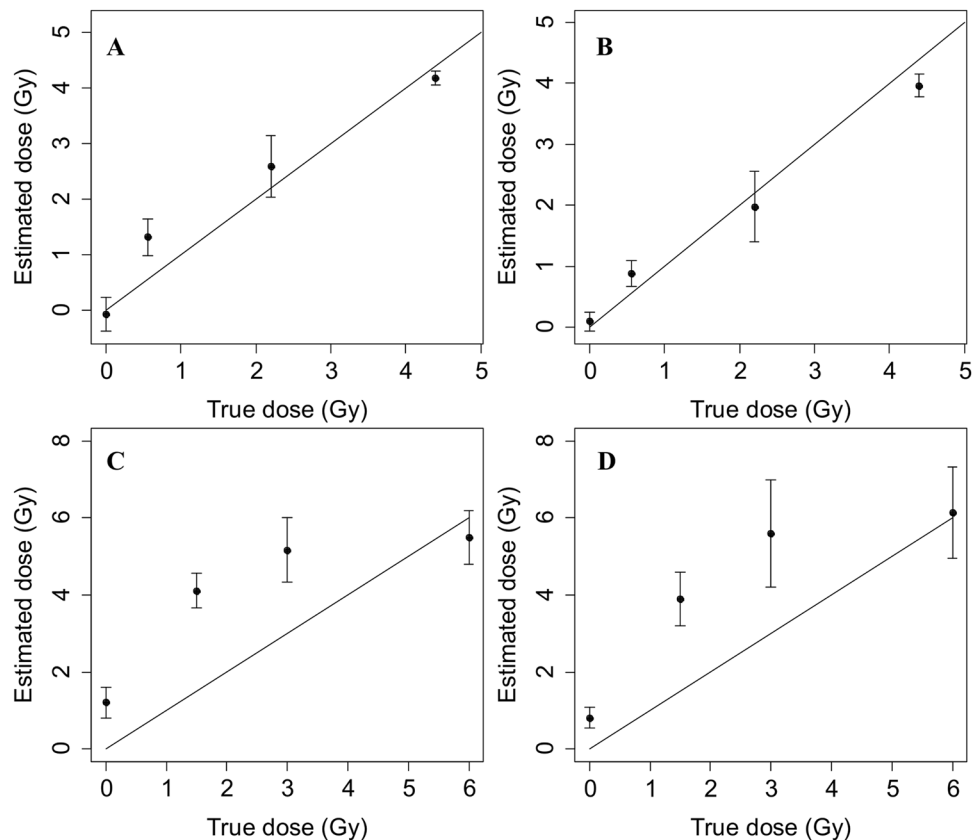
Results

Gene expression results. Published datasets of whole genome gene expression 24 h after *ex vivo* exposure of peripheral blood from healthy human donors to x-rays or γ -rays (Table 1) were used for this analysis. Genes with more than 20% missing values within a dataset were filtered out and excluded from further analysis. Replicate probes and gene symbols were averaged, which reduced the number of genes in the datasets by ~10% and yielded an average of ~20,000 genes for further analysis from each dataset.

Radiation responsive gene signature identification and normalization. In the first training data set, we identified 37 genes (Supplementary File 1, signature genes) with strong positive Spearman's correlations with dose, which withstood our tests for significance and robustness, described in Materials and Methods. Three genes with strong negative correlations with dose were also identified (Supplementary File 1, negatively correlated genes). The genes in each of these two groups were strongly correlated with each other (Fig. 1, showing positively correlated genes and Supplementary File 1, negatively correlated genes). In other words, although their signal levels varied, the dose response shapes were very similar for all genes within each group: a nonlinear (concave) increase with dose for the first group, and a nonlinear decrease with dose in the second. Very similar gene lists were generated by the repeated k-means clustering procedure (described in Materials and Methods). For example, the well-known radiation responsive genes *FDXR* and *DDB2* were found in the top-scoring cluster in 40 out of 50 repeats, whereas several other genes, such as *GADD45A* and *PCNA* were found in 9–10 out of 50 repeats. By comparison, synthetic noise variables were found in the top-scoring cluster in only ≤ 7 out of 50 repeats.

Randomization of the outcome (permutation of dose labels) did not produce any false positives: the highest score for any variable was 11 out of 50, and synthetic noise variables were intermingled with real genes in terms of scores. The alternative test of introducing an artificial dependence of the outcome (dose) on a selected noise variable also performed well in identifying the true predictor in 35 out of 50 repeats. These results demonstrate the ability of the proposed methods to separate strong predictors from weak ones and to validate radiation-responsive biomarker signatures generated by previous analyses.

To reduce the effects of variations in baseline levels of signature genes on dose reconstruction, we implemented a search for normalizer genes that co-varied with signature genes in unirradiated samples but did not have a radiation dose response. This was done by pooling two training data sets. When median signal intensities for normalizer genes were subtracted from the median signal intensities for signature genes, with positive correlations with dose, the results showed a very strong Spearman's correlation with dose across both training data sets ($R^2 = 0.969$, Fig. 2). In contrast, the same procedure applied to genes with negative correlations with dose



Panels A & B: Test dataset 1: GSE65292

Panels C & D: Test dataset 2: GSE58613

Figure 3. Comparisons of true and estimated radiation doses. Panel A Regular polynomial model on testing data set 1, $R^2=0.914$, RMSE = 0.43 Gy. Panel B Robust polynomial model on testing data set 1, $R^2=0.952$, RMSE = 0.35 Gy. (Panels A & B: GSE65292) Panel C Regular polynomial model on testing data set 2, $R^2=0.660$, RMSE = 1.74 Gy. Panel D Robust polynomial model on testing data set 2, $R^2=0.677$, RMSE = 1.74 Gy. (Panels C & D: GSE58613). Error bars represent standard deviations, and the solid line represents a theoretically perfect 1:1 correlation of true and estimated doses.

Testing data set	True dose (Gy)	Reconstructed dose (Gy) using polynomial model		Reconstructed dose (Gy) using robust polynomial model	
		mean	SD	mean	SD
Ghandhi <i>et al.</i>	0.00	-0.07	0.30	0.09	0.16
	0.56	1.32	0.32	0.88	0.21
	2.20	2.59	0.55	1.98	0.58
	4.40	4.18	0.13	3.96	0.19
Lucas <i>et al.</i>	0.00	1.20	0.40	0.81	0.27
	1.50	4.11	0.45	3.89	0.69
	3.00	5.16	0.84	5.60	1.40
	6.00	5.49	0.70	6.14	1.18

Table 3. Comparison of true doses with reconstructed dose values. SD are standard deviations.

produced much weaker results (Spearman's correlation with dose was -0.715), and we excluded this gene group from further analysis.

We fitted the polynomial model (Eq. 1 above), using the data for median-normalized signal intensities for signature genes with positive correlations with dose (S) as independent variables, and dose as the dependent variable. Best-fit parameters for the polynomial model were $k_0 = -0.43$ (standard error = 0.23), $k_1 = 2.63$ (0.36), $k_2 = 0.16$ (0.06). For the robust polynomial model, the best-fit parameter values were $k_0 = -0.10$ (0.08), $k_1 = 1.36$ (0.13), $k_2 = 0.45$ (0.02).

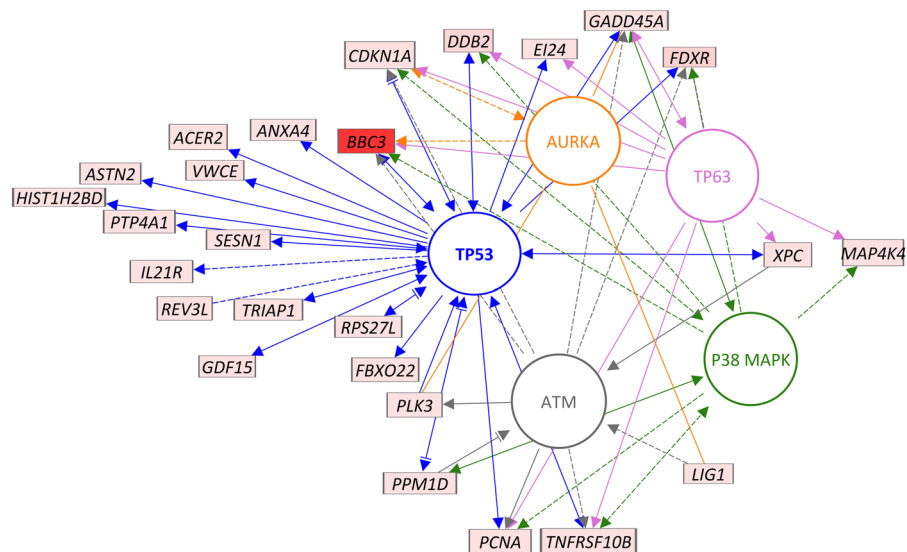


Figure 4. Network analysis showing interactions between the dose-reconstruction signature genes and potential upstream regulators. The network shown here was generated in the Path Designer tool of Ingenuity Pathway Analysis (IPA) program. A subset of signature genes are shown here with connections (indirect, dashed lines; and direct, solid lines/arrows) to the top regulators reconstructed using the IPA algorithm. Nodes/entities are either proteins (circles) or mRNA/genes (rectangles). ATM, AURKA, p38MAPK, TP53 and TP63 were the top regulators (different colors to differentiate between them, with arrows/lines of the same colors connecting the regulator and gene) with high z scores for activation. All mRNA for signature genes (all up regulated by radiation) shown overlaid with signal/expression values (low to high expression, light to dark red) from training dataset 1 (GSE8917).

Testing the reconstructor gene signature. Using normalized median gene signal values in the first testing data set (Ghandhi *et al.*¹²) as inputs (S values) for the polynomial and robust polynomial models, we generated dose reconstructions. The same procedure was performed using second testing data set (Lucas *et al.*¹⁹). Comparisons of true and reconstructed dose values are shown in Table 3 and Fig. 3. The dose reconstruction results were very close to the true doses for the first testing data set, which was generated on the same microarray platform as the two training data sets. The results on the second testing data set, which came from a different platform, showed larger error magnitudes. However, even in this case the dose reconstructions were relatively accurate at the highest tested doses.

Biological functions of signature genes (networks and pathways, regulators). We examined the biological significance of the genes in the signature using gene ontology (GO) and pathway analyses. GO analysis using the DAVID database²⁵ suggested enrichment of biological processes in DNA damage response and mitotic cell cycle (Supplementary File 2, worksheet DAVID GO). Apoptosis and cell cycle arrest as well as UV activation of cells were also significantly over-represented among these genes (Benjamini p-value < 0.05) and also in the IPA network core analysis and biological functions results (Supplementary File 2, worksheet IPA functions). Selecting the pathway category in the ontology and pathway tools, indicated similar processes being implicated, such as cell death, double strand break repair, mismatch repair and damage of lymphoid cells. We then used IPA to build networks and determine if there were common upstream regulators for this set of genes (Fig. 4). Top projected upstream regulators for these genes were AURKA (z score + 9.8), ATM (z score 8.7), p38MAPK (z score + 6.8), p53 (z score + 14.2) and p63 (z score + 9.2) (Supplementary File 2, worksheet IPA upstream regulators).

Comparison of gene response with other studies. We further visually compared our radiation dose signature genes and their expression across the datasets used in this study in the training/testing analysis. We also compared gene expression side-by-side with another dataset (GSE102971, dose range from 0 to 7 Gy) generated in our group, which was performed independently to test similarity of human gene expression responses to *ex vivo* irradiation with non-human primates²⁷. The heat map in Fig. 5, displays the scale and changes of signals for 27 of the genes from the reconstructor gene signature (some genes were trimmed for this visualization because of missing values in one or more dataset). GSE102971 data were from different human donors, but most of the genes were induced similarly as in GSE8917, which was a training set used here. Some genes such as *CDKN1A* and *ASTN2* were also induced but the mean signal intensities were lower than in the training set. Some genes such as *DDB2* and *TNFRSF10B* showed very similar levels of gene induction and baseline gene expression. We also included neutron data from GSE90909¹⁵, also from our group, as we have found similar responses after neutron doses in *ex vivo* irradiated human blood (last dataset in Fig. 5).

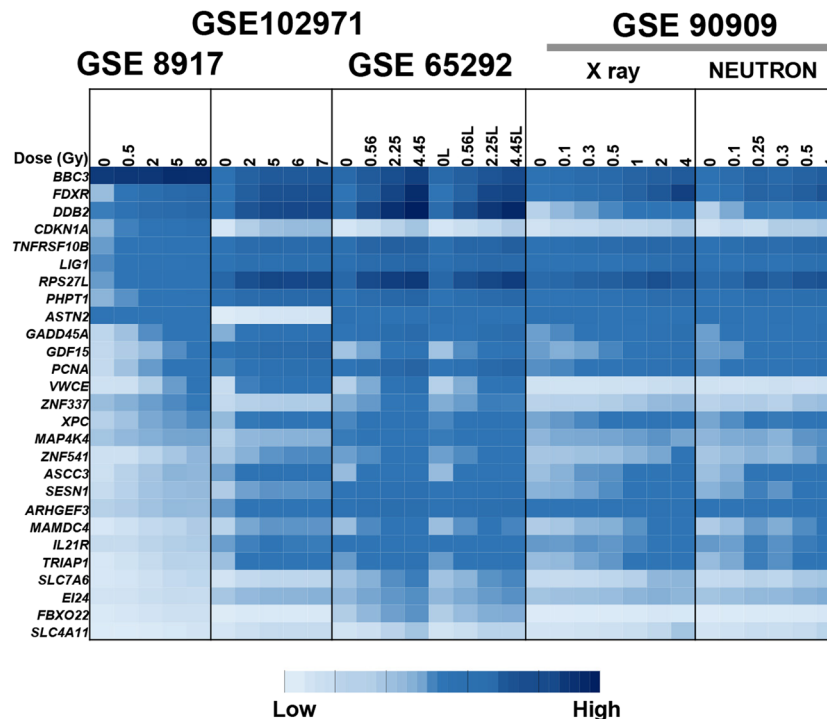


Figure 5. Heat map of gene expression of genes from human blood *ex vivo* irradiated studies, including an independent study for comparison. Shown here are the normalized expression values for a subset of signature genes (except those that had missing values in some of the datasets) in the sham-irradiated samples (0 Gy) and irradiated samples from different datasets. Starting from the left, first, the training study GSE8917; second, an independent human blood *ex vivo* irradiation data set GSE102971; third, the first test dataset GSE65292, in which the same donor blood samples were split to study dose-rate effects (suffix L, low dose-rate); and finally, the second training dataset GSE90909 in which the same donor blood samples were split to study LET effects using pure neutrons. Supplementary File 3, shows the data of the gene expression values.

Discussion

We have developed a robust gene expression signature, which can work effectively to estimate radiation dose in the simplest context. With very stringent biostatistical approaches and by asking a focused question, we have generated a list of target genes that can accurately reconstruct the dose to which a sample was exposed. These results showed that the use of radiation-responsive signature genes positively correlated with dose is robust against several statistical testing approaches: correlation analysis, mixed effects modeling, synthetic noise. The accuracy of our human gene signature was in a very good range, especially on a dataset from the same platform as the training data (Table 3, Fig. 3). Compared with other studies using gene expression results where error for dose reconstruction was within the range ± 2.2 Gy²⁸ and with reported microarray-based mean absolute differences of 1.5 to 2.4 Gy for reconstruction of a 4 Gy dose²⁹, our estimation of dose was preferable. Lacombe *et al.*⁹ performed a systematic review of 24 independent studies to identify 30 dose reconstruction genes at any time from 2 to 48 h after exposure. They demonstrated the ability of genes from this set to discriminate between doses above or below 2 Gy, but did not test actual dose reconstruction on independent datasets. In contrast, our study focuses on the response at 24 h after exposure, which is the earliest time after a large-scale event when it is thought that first responders may reasonably be able to start assessing the affected population. There were 15 genes in common between the gene-sets reported in our study (Human sig (37 genes)) and Lacombe (31 genes), shown in Fig. 6. We also confirmed that most of the genes in the signature identified here were part of the consensus gene set (Paul (64) in Fig. 6) used to classify samples by dose in the initial analysis of the study that provided our first training dataset¹⁸. The genes common to all three signatures were *ASCC3*, *BBC3*, *CDKN1A*, *DDB2*, *E124*, *FBXO22*, *FDXR*, *GADD45A*, *PCNA*, *PHPT1*, *RPS27L*, *SESN1*, *TNFRSF10B*, *TRIAP1* and *XPC* all of which are known radiation-response genes^{30–35}. There were an additional 16 genes (*ANKRA2*, *ANXA4*, *ARHGEF3*, *ASTN2*, *GDF15*, *IL21R*, *LIG1*, *MAMDC4*, *PLK3*, *PPM1D*, *PTP4A1*, *SLC4A11*, *SLC7A6*, *UROD*, *VWCE* and *ZNF337*) in common between our robust dose reconstructor and the Paul consensus gene set (64 genes). Interestingly, some genes (*HIST1H2BD*, *DRAM1*, *MAP4K4*, *REV3L*, *WIG1* and *ZNF541*) were only included in the signature identified in this study; most of these are involved in the p53 and/or radiation response^{36–41}.

The ultimate goal of radiation biodosimetry is to reconstruct the dose to people who were exposed *in vivo*. Several studies have reported that some, but not all, of the *in vivo* gene expression response to radiation measured in blood cells is recapitulated when the blood is irradiated *ex vivo*, and that informed selection of genes allows signatures based on *ex vivo* exposures to reconstruct dose levels of *in vivo* exposures^{19,42–44}. Many of the genes commonly suggested for biodosimetry have also been shown to respond to the stress of blood cells being cultured

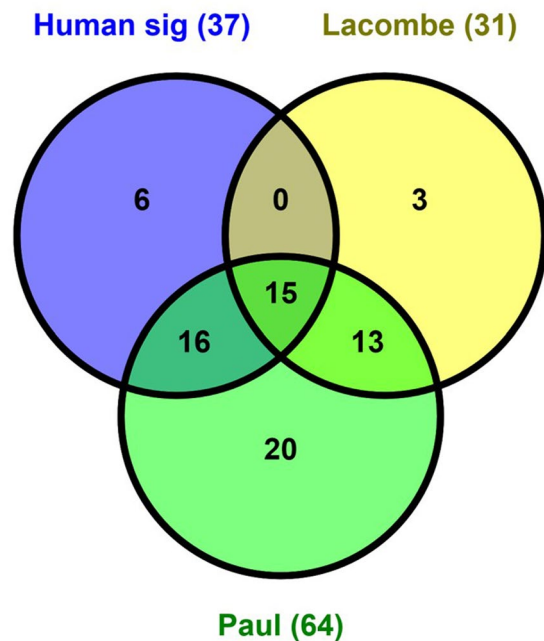


Figure 6. Venn diagram showing overlap between our signature genes and others. We compared overlap between gene signatures published in two other studies. The Human sig (signature) genes were the 37 genes from the biostatistics method described here, the Lacombe study also used a meta-analysis approach from which 31 genes were identified from published studies at different times post-irradiation, and the Paul *et al.*, 64 consensus-gene set that has been used for classification of radiation dose levels.

outside the body⁴³. Future implementation of a dose-reconstruction gene signature may require modification to account for the *in vivo* context. The goal of the analysis described here was to determine a core gene signature to reconstruct the gamma-equivalent dose in irradiated human blood in an experimental scenario that simulates the potential real-life exposures that may occur after a radiation accident or bomb. The set of genes selected by these analyses are mostly well-known radiation response genes that correlate very well with dose. We also showed as a proof-of-principle that our strategy of using normalizer genes to correct for variability in the signature genes in unirradiated control samples across datasets, could work to generate continuous dose reconstructions in a training/testing framework.

Conclusions

We identified radiation-responsive “signature” genes with continuous dose responses and strong positive correlations with dose that are consistent across several data sets. The gene group with negative correlations with dose was much smaller/weaker in these *ex vivo* blood data sets. The identified radiation-responsive “signature” genes are biologically relevant to the stress response, overlap with previous findings, and stood up to various tests: addition of synthetic noise, mixed effects modeling of donor effects, different forms of clustering. We have tested the dose reconstruction capacity of these genes across independent datasets and across more than one platform. The performance of dose reconstruction was best in the dataset that used the same measurement platform, having an acceptable level of error of ± 0.35 Gy. This gene set shows great promise for reconstruction of individual radiation dose and may be developed further to be informative for the more complex exposures likely to be encountered in a realistic radiological or nuclear event.

Data availability

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/>) under accession numbers GSE8917, GSE90909, GSE65292, GSE102971 and GSE58613.

Received: 20 June 2019; Accepted: 19 November 2019;

Published online: 05 December 2019

References

1. Bushberg, J. T. *et al.* Nuclear/radiological terrorism: emergency department management of radiation casualties. *The Journal of emergency medicine* **32**, 71–85, <https://doi.org/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. *Radiation protection dosimetry* **159**, 77–86, <https://doi.org/10.1093/rpd/ncu120> (2014).
3. Rump, A. *et al.* Medical management of victims contaminated with radionuclides after a “dirty bomb” attack. *Military Medical Research* **5**, 27, <https://doi.org/10.1186/s40779-018-0174-5> (2018).

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, <https://doi.org/10.1097/HP.0b013e31829cf221> (2013).
5. Huang, T. C. *et al.* Gene expression analysis with an integrated CMOS microarray by time-resolved fluorescence detection. *Biosens Bioelectron* **26**, 2660–2665, <https://doi.org/10.1016/j.bios.2010.03.001> (2011).
6. Brengues, M. *et al.* Biodosimetry on small blood volume using gene expression assay. *Health physics* **98**, 179–185, <https://doi.org/10.1097/01.HP.0000346706.44253.5c> (2010).
7. Lacombe, J., Phillips, S. L. & Zenhausern, F. Microfluidics as a new tool in radiation biology. *Cancer Lett* **371**, 292–300, <https://doi.org/10.1016/j.canlet.2015.11.033> (2016).
8. Hurth, C. *et al.* A miniature quantitative PCR device for directly monitoring a sample processing on a microfluidic rapid DNA system. *Biomed Microdevices* **16**, 905–914, <https://doi.org/10.1007/s10544-014-9895-8> (2014).
9. 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, <https://doi.org/10.1371/journal.pone.0198851> (2018).
10. Lu, T. P., Hsu, Y. Y., Lai, L. C., Tsai, M. H. & Chuang, E. Y. Identification of gene expression biomarkers for predicting radiation exposure. *Scientific reports* **4**, 6293, <https://doi.org/10.1038/srep06293> (2014).
11. Macaeva, E. *et al.* Radiation-induced alternative transcription and splicing events and their applicability to practical biodosimetry. *Scientific reports* **6**, 19251, <https://doi.org/10.1038/srep19251> (2016).
12. 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 medical genomics* **8**, 22, <https://doi.org/10.1186/s12920-015-0097-x> (2015).
13. 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, <https://doi.org/10.1667/RR14044.1> (2015).
14. Buddemeier, B. R. Nuclear detonation fallout: Key considerations for internal exposure and population monitoring. LLNL-TR-754319. Lawrence Livermore National Lab (LLNL), Livermore, CA. (2018).
15. Broustas, C. G. *et al.* Impact of Neutron Exposure on Global Gene Expression in a Human Peripheral Blood Model. *Radiat Res*, <https://doi.org/10.1667/rr0005.1> (2017).
16. Broustas, C. G., Xu, Y., Harken, A. D., Taylor, G. & Amundson, S. A. Comparison of gene expression response to neutron and x-ray irradiation using mouse blood. *BMC genomics* **18**, 2, <https://doi.org/10.1186/s12864-016-3436-1> (2017).
17. Kramer, K., Li, A., Madrigal, J., Sanchez, B. & Millage, K. Report DTRA-TR-13-045 (R1): Monte Carlo Modeling of the Initial Radiation Emitted by a Nuclear Device in the National Capitol Region. Defense Threat Reduction Agency, Fort Belvoir, VA. (2016).
18. Paul, S. & Amundson, S. A. Development of gene expression signatures for practical radiation biodosimetry. *Int J Radiat Oncol Biol Phys* **71**, 1236–1244, <https://doi.org/10.1016/j.ijrobp.2008.03.043> (2008).
19. Lucas, J. *et al.* A translatable predictor of human radiation exposure. *PLoS one* **9**, e107897, <https://doi.org/10.1371/journal.pone.0107897> (2014).
20. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic acids research* **41**, D991–995, <https://doi.org/10.1093/nar/gks1193> (2013).
21. Simon, R. *et al.* Analysis of gene expression data using BRB-ArrayTools. *Cancer informatics* **3**, 11–17 (2007).
22. Team, R. R. A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria (2013).
23. Shuryak, I. Advantages of Synthetic Noise and Machine Learning for Analyzing Radioecological Data Sets. *PLoS One* **12**, e0170007, <https://doi.org/10.1371/journal.pone.0170007> (2017).
24. Shuryak, I. Modeling species richness and abundance of phytoplankton and zooplankton in radioactively contaminated water bodies. *Journal of environmental radioactivity* **192**, 14–25, <https://doi.org/10.1016/j.jenvrad.2018.05.016> (2018).
25. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44–57, <https://doi.org/10.1038/nprot.2008.211> (2009).
26. Oliveros, J. C. An interactive tool for comparing lists with Venn Diagrams. (2007).
27. Park, J. G. *et al.* Developing Human Radiation Biodosimetry Models: Testing Cross-Species Conversion Approaches Using an *Ex Vivo* Model System. *Radiat Res*, <https://doi.org/10.1667/rr14655.1> (2017).
28. Pogosova-Agadjanyan, E. L. *et al.* Identification of radiation-induced expression changes in nonimmortalized human T cells. *Radiat Res* **175**, 172–184 (2011).
29. Abend, M. *et al.* Examining Radiation-Induced *In Vivo* and *In Vitro* Gene Expression Changes of the Peripheral Blood in Different Laboratories for Biodosimetry Purposes: First RENEB Gene Expression Study. *Radiat Res* **185**, 109–123, <https://doi.org/10.1667/rr14221.1> (2016).
30. Hwang, P. M. *et al.* Ferredoxin reductase affects p53-dependent, 5-fluorouracil-induced apoptosis in colorectal cancer cells. *Nat Med* **7**, 1111–1117, <https://doi.org/10.1038/nm1001-1111> (2001).
31. Xu, J. & Morris, G. F. p53-mediated regulation of proliferating cell nuclear antigen expression in cells exposed to ionizing radiation. *Mol Cell Biol* **19**, 12–20, <https://doi.org/10.1128/mcb.19.1.12> (1999).
32. Hollander, M. C. *et al.* Analysis of the mammalian gadd45 gene and its response to DNA damage. *J Biol Chem* **268**, 24385–24393 (1993).
33. Velasco-Miguel, S. *et al.* PA26, a novel target of the p53 tumor suppressor and member of the GADD family of DNA damage and growth arrest inducible genes. *Oncogene* **18**, 127–137, <https://doi.org/10.1038/sj.onc.1202274> (1999).
34. He, H. & Sun, Y. Ribosomal protein S27L is a direct p53 target that regulates apoptosis. *Oncogene* **26**, 2707–2716, <https://doi.org/10.1038/sj.onc.1210073> (2007).
35. Finnberg, N. *et al.* DR5 knockout mice are compromised in radiation-induced apoptosis. *Mol Cell Biol* **25**, 2000–2013, <https://doi.org/10.1128/mcb.25.5.2000-2013.2005> (2005).
36. Bakshi, M. V. *et al.* In-Utero Low-Dose Irradiation Leads to Persistent Alterations in the Mouse Heart Proteome. *PLoS One* **11**, e0156952, <https://doi.org/10.1371/journal.pone.0156952> (2016).
37. Chauhan, V., Howland, M., Greene, H. B. & Wilkins, R. C. Transcriptional and secretomic profiling of epidermal cells exposed to alpha particle radiation. *The open biochemistry journal* **6**, 103–115, <https://doi.org/10.2174/1874091x01206010103> (2012).
38. Cui, L. *et al.* Radiation induces autophagic cell death via the p53/DRAM signaling pathway in breast cancer cells. *Oncol Rep* **35**, 3639–3647, <https://doi.org/10.3892/or.2016.4752> (2016).
39. Meng, C. *et al.* MicroRNA-26b suppresses autophagy in breast cancer cells by targeting DRAM1 mRNA, and is downregulated by irradiation. *Oncology letters* **15**, 1435–1440, <https://doi.org/10.3892/ol.2017.7452> (2018).
40. Varmeh-Ziaie, S. *et al.* Wig-1, a new p53-induced gene encoding a zinc finger protein. *Oncogene* **15**, 2699–2704, <https://doi.org/10.1038/sj.onc.1201454> (1997).
41. Wang, W. *et al.* REV3L modulates cisplatin sensitivity of non-small cell lung cancer H1299 cells. *Oncol Rep* **34**, 1460–1468, <https://doi.org/10.3892/or.2015.4121> (2015).
42. Amundson, S. A. *et al.* Human *in vivo* radiation-induced biomarkers: gene expression changes in radiotherapy patients. *Cancer Res* **64**, 6368–6371, <https://doi.org/10.1158/0008-5472.CAN-04-1883> (2004).
43. Paul, S. *et al.* Prediction of *in vivo* radiation dose status in radiotherapy patients using *ex vivo* and *in vivo* gene expression signatures. *Radiat Res* **175**, 257–265, <https://doi.org/10.1667/RR2420.1> (2011).
44. Port, M. *et al.* Validating Baboon *Ex Vivo* and *In Vivo* Radiation-Related Gene Expression with Corresponding Human Data. *Radiat Res* **189**, 389–398, <https://doi.org/10.1667/rr14958.1> (2018).

Acknowledgements

This study was funded by the NIAID grant U19A1067773.

Author contributions

S.A.G. and S.A.A. conceived the study, performed the gene ontology and network analyses and drafted the manuscript. I.S. performed all biostatistics analyses and drafted the manuscript. D.J.B. reviewed the biostatistics analysis and helped draft the manuscript and S.R.M. assisted with the biostatistical analyses and helped prepare the manuscript and figures.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-54967-5>.

Correspondence and requests for materials should be addressed to S.A.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019