Although significant shifts in longitudinal frequencies of CAs were observed (**Fig 7A-D)**,we were unable to identify groups of patients with overall lower or higher frequencies of CAs to make inference on radiosensitivity **(Fig 8**, **Supp Fig 1/2**). We attempted to use linear regression for predicting post-therapy average CA frequencies from pre-therapy data, which was unsuccessful overall (**Fig 9**). We also attempted to derive predictive value from our CA data by training XGBoost models to relate pre-therapy counts of CAs per cell with post-therapy average CA frequencies (**Fig 10**, **Supp Table 5**). We speculate that training the XGBoost models with the CA data failed due to either a lack of CA data or also a faulty attempt to relate pre-therapy counts of CAs per cell with post-therapy average CA frequencies (**Supp Table 4**).  Our results suggest that despite the efficiency of dGH in capturing CAs, our number of scored cells per time point/patient (n=30) was inadequate for addressing differential radiosensitivity. Given the relative predominance of inversions in our data, we speculate that when scoring greater numbers of cells, inversions may prove to be a valuable endpoint for predicting patient radiosensitivity. Future studies using dGH to evaluate chromosomal radiosensitivity should consider using more cells per sample (we propose n>60) than used here (n=30).

and pro and made our tools and methodologies available in full.

   degenerative  with degenerative Telomeres are clinical features of interest for radiation late effects, and predicting a patient’s telomeric response to radiation therapy

Developing a means to predict radiosensitivity in patients remains a core objective for radiation therapy as a treatment modality. Accurate determination of a patient’s radiosensitivity status before therapy would enable more personalized treatment planning and potential aversion of radiation late effects. Telomere length, a compelling dual-sided biomarker, to predict patients at increased risks for degenerative radiation late effects or secondary cancers.

Identifying or predicting radiosensitivity in patients remains a core objective for radiation therapy as a treatment modality; especially for pediatric patients. Accurate determination of a patient’s radiosensitivity status before therapy would enable more personalized treatment planning and potential aversion of negative outcomes from radiation therapy, namely radiation late effects. Presently, many assays for predicting radiosensitivity have been developed, albeit with varying degrees of compromise between cost-effectiveness, throughput, and predictive power, which has prevented the adoption of a standard for assessing radiosensitivity. The lack of a single standard for predicting radiosensitivity highlights the need for exploration and testing of new biomarkers and approaches. One approach of particular and rising interest in translational medicine is the use of machine learning models, which can leverage large amounts of patient data and biomarkers to make accurate predictions of patient outcomes.

Telomere length is a particularly compelling biomarker for predicting patient outcomes of radiation therapy, as telomeres can shorten or lengthen in response to IR exposure, with many short telomere pathologies overlapping with radiation late effects (CVD, pulmonary fibrosis, aplastic anemia) and long telomeres being strongly associated with risk for leukemias, a common secondary cancer (and late effect) from radiation therapy. We hypothesized that telomere length could provide a dual-sided biomarker for inferring patients’ risk for types of radiation late effects, and so our core objective was to determine if telomere length measured at pre-therapy, in non-irradiated and *in vitro* irradiated patient samples, could be used to predict changes in post-therapy telomere length for individual patients. When clustering patients by longitudinal changes in mean telomere length and by numbers of short or long telomeres, we observed the same three patients responding to radiation therapy with dramatic and persistent reductions in mean telomere length and increased numbers of short telomeres, suggesting increased risks for degenerative radiation late effects (**Fig 1B/C, Fig 2B, Fig 3B/C**). The remaining patients responded to radiation therapy with increased and persistent increases in mean telomere length and numbers of long telomeres, suggesting increased risks for secondary cancers (**Fig 1B/C, Fig 2C, Fig 3D/E**). These results provided indications that telomere length could be informative, as patients clustered into discrete groups on the basis of telomeric responses to radiation therapy.

Linear regression performed poorly for predicting post-therapy telomeric outcomes using pre-therapy telomeric data (**Fig 4, Supp Table 1A-C**), likely due in part to the nonlinear response of telomere length to radiation, as suggested by the observed differences in patient responses (**Fig 1-3**), but perhaps also because of the low number of observations (n=15 patients). TeloFISH provides not only mean telomere length but also individual telomere length measurements, which in this study totaled 103,040 from the pre-therapy time points. We sought to leverage our large amount of pre-therapy TeloFISH data, and hypothesized that a machine learning model could learn the observed nonlinear relationships between pre-therapy individual telomere length measurements and post-therapy telomeric outcomes.

To the best of our knowledge, we have presented the first implementation of individual telomere length measurements in a machine learning model (**Fig 5**). XGBoost models trained on pre-therapy individual telomere length measurements successfully predicted post-therapy telomeric outcomes with high accuracy when challenged with the test data set (**Fig 6**), strongly suggesting that the models generalized to new data. Our results demonstrate that patients’ telomeric responses to radiation therapy could be predicted with machine learning models trained on pre-therapy individual telomere length measurements, enabling the possibility of using telomere length to inform a patient’s risk for radiation late effects. Our results also raise the possibility of training machine learning models on TeloFISH data from larger patient cohorts for development of tool(s) to aid predictions of types of radiation late effects in a clinical setting.

Chromosome aberrations (CAs) are an attractive biomarker for radiosensitivity as they represent a quantitative endpoint of a cell or tissue’s overall capacity to repair IR-induced DNA damage, which is considered the primary mechanism of action by radiation therapy on cancers. CA assays have had success in predicting patient radiosensitivity, however CA assays are costly in terms of labor and time, with typical assays requiring hundreds or thousands of cells scored. Directional Genomic Hybridization (dGH) is a newer cytogenetic tool for scoring CAs, which has been demonstrated to capture CAs, particularly inversions, at higher frequencies than other assays. Whether dGH could predict radiosensitivity while scoring fewer numbers of cells was unknown. Given the efficiency of dGH in capturing CAs, we hypothesized that it would be possible to score fewer cells (n=30) while retaining sensitivity to differences between chromosomal radiosensitivity in patients, and thus the ability to identify radiosensitive patients.

Although significant shifts in longitudinal frequencies of CAs were observed (**Fig 7A-D)**,we were unable to identify groups of patients with overall lower or higher frequencies of CAs to make inference on radiosensitivity **(Fig 8**, **Supp Fig 1/2**). We attempted to use linear regression for predicting post-therapy average CA frequencies from pre-therapy data, which was unsuccessful overall (**Fig 9**). We also attempted to derive predictive value from our CA data by training XGBoost models to relate pre-therapy counts of CAs per cell with post-therapy average CA frequencies (**Fig 10**, **Supp Table 5**). We speculate that training the XGBoost models with the CA data failed due to either a lack of CA data or also a faulty attempt to relate pre-therapy counts of CAs per cell with post-therapy average CA frequencies (**Supp Table 4**).  Our results suggest that despite the efficiency of dGH in capturing CAs, our number of scored cells per time point/patient (n=30) was inadequate for addressing differential radiosensitivity. Given the relative predominance of inversions in our data, we speculate that when scoring greater numbers of cells, inversions may prove to be a valuable endpoint for predicting patient radiosensitivity. Future studies using dGH to evaluate chromosomal radiosensitivity should consider using more cells per sample (we propose n>60) than used here (n=30).

We have demonstrated the possibility of training machine learning models on pre-therapy individual telomere length measurements for accurate predictions of post-therapy telomeric outcomes, demonstrating feasibility of using telomere length to predict types of radiation late effects. Future studies should examine the relationships between magnitude of change in patients’ telomere length post-radiation therapy with types, severity, and average time of onset of radiation late effects. Our work was performed in the context of prostate cancer patients undergoing IMRT, with telomere data collected by TeloFISH, and XGBoost as the model of choice. Future studies include addressing the potential of predicting post-therapy telomeric outcomes with different assays for measuring individual telomeres, cohorts of male and female radiation therapy patients, different radiation modalities, and different cancer types. We used XGBoost due to its computational efficiency, flexibility, and efficacy in learning nonlinear patterns within data. TeloFISH data would be amenable to other machine learning models and paradigms, including deep, ensemble, and transfer learning, which could enable greater model performance and generalizability at the cost of computational resources.

Maeda, T., Nakamura, K., Atsumi, K., Hirakawa, M., Ueda, Y. and Makino, N. (2013) Radiation-associated changes in the length of telomeres in peripheral leukocytes from inpatients with cancer. *Int J Radiat Biol*, **89**, 106-109.

27. Berardinelli, F., Antoccia, A., Buonsante, R., Gerardi, S., Cherubini, R., De Nadal, V., Tanzarella, C. and Sgura, A. (2013) The role of telomere length modulation in delayed chromosome instability induced by ionizing radiation in human primary fibroblas

Liu, H., Xie, Y., Zhang, Z., Mao, P., Liu, J., Ma, W. and Zhao, Y. (2018) Telomeric Recombination Induced by DNA Damage Results in Telomere Extension and Length Heterogeneity. *Neoplasia*, **20**, 905-916.