

Original Article

Radiation Response Prediction Model based on Integrated Clinical and Genomic Data Analysis

Bum-Sup Jang¹, Ji-Hyun Chang², Seung Hyuck Jeon³, Myung Geun Song⁴, Kyung-Hun Lee⁵, Seock-Ah Im⁵, Jong-Il Kim⁶, Tae-You Kim⁵, Eui Kyu Chie^{2,7}

¹Department of Radiation Oncology, Seoul National University Bundang Hospital, Seongnam,

²Department of Radiation Oncology, Seoul National University College of Medicine, Seoul,

³Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and

Technology, Daejeon, ⁴Biomedical Research Institute, Seoul National University Hospital, Seoul,

⁵Department of Internal Medicine, Seoul National University Hospital, Seoul, ⁶Department of

Biomedical Sciences, Seoul National University College of Medicine, Seoul, ⁷Institute of

Radiation Medicine, Medical Research Center, Seoul National University, Seoul, Korea

Running Title: Radiation response prediction using genomic and clinical data

Correspondence: Eui Kyu Chie

Department of Radiation Oncology, Seoul National University College of Medicine, 101

Daehak-ro, Jongno-gu, Seoul 03080, Korea

Tel: 82-2-2072-3705 Fax: 82-2-765-3317 E-mail: ekchie93@snu.ac.kr

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article'; doi:10.4143/crt.2021.759

Abstract

Purpose

The value of the genomic profiling by targeted gene-sequencing on radiation therapy response prediction was evaluated through integrated analysis including clinical information. Radiation response prediction model was constructed based on the analyzed findings.

Materials and Methods

Patients who had the tumor sequenced using institutional cancer panel after informed consent and received radiotherapy for the measurable disease served as the target cohort. Patients with irradiated tumor locally controlled for more than 6 months after radiotherapy were defined as the durable local control (DLC) group, otherwise, non-durable local control (NDLC) group. Significant genomic factors and domain knowledge were used to develop the Bayesian Network model to predict radiotherapy response.

Results

Altogether, 88 patients were collected for analysis. Of those, 41 (43.6%) and 47 (54.4%) patients were classified as the NDLC and DLC group, respectively. Somatic mutations of NOTCH2 and BCL were enriched in the NDLC group, whereas, mutations of CHEK2, MSH2, and NOTCH1 were more frequently found in the DLC group. Altered DNA repair pathway was associated with better local failure-free survival (HR 0.40, 95%CI 0.19-0.86, $p=0.014$). Smoking somatic signature was found more frequently in the DLC group. AUC of the Bayesian Network model predicting probability of 6-month local control was 0.83.

Conclusion

Durable radiation response was associated with alterations of DNA repair pathway and smoking somatic signature. Bayesian network model could provide helpful insights for high precision radiotherapy. However, these findings should be verified in prospective cohort for further individualization.

Keyword Radiation therapy, Response, Targeted-gene, Sequencing, Bayesian Network

Introduction

Genome sequencing is often used for tumor samples to help find the genomic alterations suitable for targeted therapy [1,2]. With increased use of institutional panel sequencing, there are challenges to interpret clinical and/or potential value of reported mutation. For druggable targets, utility-based ranking system such as the European Society for Medical Oncology scale for clinical actionability of molecular targets [3] or the American College of Medical Genetics guidelines [4] are introduced to prioritize alterations.

Radiation therapy (RT) is another crucial component for cancer treatment. Several syndromes developed from germline variations have been reported to be related with radiation sensitivity. However, these variations have not been linked to the radiation response of the tumor [5,6]. To our knowledge, only limited number of studies have investigated the association between the genomic profile of the patient-derived tumor and the clinical RT response. Genomic profile as well as clinical factors such as diagnosis, RT dose or fractions are well acknowledged factors related with RT response [7]. The clinical value of the genomic profiling needs to be evaluated accompanying clinical information. Thus, the model integrating clinical and genomic factors is unmet need in predicting RT response.

All exons and selected introns from tumor were sequenced with an institutional targeted next generation sequencing (NGS) panel, named FiRST cancer panel, to identify somatic mutations, copy number alterations, and structural variations of tumor samples obtained from informed-consent patients. Of those, we evaluated RT response in patients with measurable disease. Clinical data and genomic alterations were integrated by using machine learning framework to develop radiation response prediction model.

Materials and Methods

1. Study population and Clinical Data Collection

Among patients whose tumors were sequenced using the FiRST cancer gene panel version 3.1, patients who received radiotherapy for a measurable lesion were analyzed in the current study. Measurable tumor was defined according to the new response evaluation criteria in solid tumor, the revised RECIST guideline version 1.1 [8]. Baseline characteristics including age, gender, and diagnosis are collected. The site and type (metastatic vs. primary) of the sequenced specimen and match of specimen collection site against RT site was also collected. As for the radiotherapy, intent of treatment, dose/fractions, biologically effective dose (BED), treatment site, modality, and the completeness of the prescribed treatment were reviewed.

To evaluate the RT response, we tracked the changes of the RT-treated tumor in follow-up medical images as well as patient symptoms. Response to RT was defined as radiologic response or symptom relief greater than partial response of the treated target. For example, for metastatic bony lesions, revised RECIST guideline was used for bony lesions with measurable extraosseous extension. For lesions not deemed measurable, symptomatic response was employed for response assessment. Those not meeting the criteria were considered no response. The duration of local control was calculated from the first day of RT to tumor progression. The durable local control (DLC) group was defined as group of patients whose irradiated tumor was locally controlled for more than 6-months. The remaining patients were defined as the non-DLC (NDLC) group.

2. Genomic Data from Cancer Gene Panel

The FiRST cancer panel version 3.1, employed in current study, was developed to identify exon of 183 genes, introns of 23 fusion genes, the *TERT* promoter region, 8

microsatellite-instability markers, and 45 drug-target lesions. Detailed list of target genes is found in previous studies [9,10]

Formalin-fixed paraffin-embedded tumor tissue was used to extract DNA, which was sequenced by NextSeq 550Dx (Illumina Inc., USA). After quality assurance protocol, the FASTQ file was generated and aligned to the hg19 reference genome with the Burrows-Wheeler Aligner-men (v0.7.17) [11] and the Genomic Analysis Toolkit Best Practice [12]. For SNV and InDel detection, GATK UnifiedGenotyper (v4.0.6.0), SNVer (v0.5.3), and LoFreq (v2.1.0) were employed. The Delly (v0.7.8) and Manta (v1.4.0) were used to identify structural variation (SV), and the THetA2 (v0.7) and CNVKit (v0.9.3) were used to reveal copy number variation (CNV). Filtering parameters are as follows: variant allele frequency $\geq 5\%$, reads supporting for alternative allele ≥ 10 , and reads supporting each strand for alternative allele > 5 for SNV and InDel; copy number ≥ 6 and copy number ≤ 1 for CNV; split reads supporting for alternative allele ≥ 10 for translocation. By applying the threshold ($> 1\%$) of frequency in variants found in multiple databases, the germline variants were filtered out. Afterwards, SNP/Indel variants were annotated by the Catalogue of Somatic Mutations in Cancer (COSMIC) database version 91, and those variants not retrieved from the COSMIC were further filtered out.

To derive somatic signatures, we adopted 'deconstructSigs' version 1.8.0 [13] using the COSMIC somatic signatures as reference. As per previous study [2], we derived eight signatures from COSMIC somatic signatures: aging, APOBEC, smoking, BRCA1/2, MMR, UV, POLE, and TMZ.

To visualize alterations, 'maftools' [14] and 'circlize' [15] based on 'R' statistical software version 4.0.3 were used.

3. Prediction Model regarding Radiotherapy Response

We employed the Bayesian Network model to develop a RT response prediction model. The node indicates RT parameters, SV, CNV, and diagnosis. The edge indicates causal or effect relationship. The conditional dependencies of variables are represented with a directed acyclic graph (DAG) and joint probabilities. Bayesian network specifies the full joint probability distribution as follows:

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P\left(\frac{X_i}{X_{j(i)}}\right)$$

where $X_{j(i)}$ denotes the parents of X_i . Each node indicates one of a set of n dimensional variables $\{x_1, x_2, \dots, x_n\}$. P denotes this conditional probability distribution for variable X_i . The Bayesian network structure is constructed to predict the probability of the DLC at 6-month via augmented Naive Bayes method. Three Bayesian network models— clinical, genomic, and clinico-genomic integrated model were generated. Mean area under the receiver operating characteristic curve (AUC) values and mean area under the precision-recall curve (AUPRC) were computed through 5-fold cross validation method.

In the clinico-genomic integrated model, target optimization tree algorithm was used to find the best evidence to maximize the DLC rate at 6-month. The target dynamic profile and embedded target optimization functions were utilized to serve the purpose. These functions are intended to search the optimum combination of variables that have a nonlinear relationship with the target and correlations between them.

Further, we performed direct effect contribution analysis to identify which type of data (clinical vs. genomic) serves as dominating factor on the local control prediction. While varying the x -value, a mean value analysis infers the corresponding mean values of the target node. Then, direct effect is the derivatives of their respective effect curves, which were computed at

their a priori mean values as follows:

$$De_x = \frac{\delta_y}{\delta_x}$$

The standardized direct effect normalizes the direct effect by taking into account the ratio between the standard deviation of the variables (x) and the Target Node (y).

$$SDe_x = \frac{\delta_y}{\delta_x} \times \frac{\sigma_x}{\sigma_y}$$

Finally, contribution is computed as follows:

$$C_x = \frac{|SDe_x|}{\sum_{xi \in X} |SDe_x|}$$

All Bayesian Network development, evaluation, optimization analysis, and estimation of direct effect contribution were performed using the BayesiaLab 8 (Bayesia S.A.S. France).

4. Statistical Analysis

Logistic regression was used for enrichment analysis. Chi-square tests were used to compare somatic signatures between the DLC and non-DLC groups. Univariate and multivariate Cox proportional hazard models were established to investigate factors associated with local failure-free survival that was defined as the duration of local control as mentioned above. Kaplan-Meier survival analyses and log-rank tests were also performed using STATA 16 (StataCorp LLC, USA). To compare AUC among clinical, genomic, and clinico-genomic models, the DeLong's pairwise comparison method was used [16]. To compare AUPRC, we adopted bootstrapping method with iterations of 1000 and estimated confidence interval. The

PRISM 8 (GraphPad Software, USA) was used to perform t-test and depict bar-graphs.

Results

1. Patient and treatment characteristics

Among 178 patients enrolled in the protocol for tumor sequencing, there were 88 patients (49.44%) who underwent RT. Tumor samples were collected and sequenced between March 2013 to March 2019. Among 88 patients, breast cancer was the leading diagnosis in the study population (n=28, 32%), followed by sarcoma (n=16, 18%), melanoma (n=9, 10%), and colorectal cancer (n=5, 6%) (Fig. 1A). As for the local control end-point, 43.6% (n=41) were grouped to the NDLC group, whereas 54.4% (n=47) were assigned to the DLC group, respectively. The most common RT-treated site was bone (Fig. 1B), and the leading objective of RT was palliation (Fig. 1C). Distribution of various factors were balanced between two groups except for the performance status (Table 1). Patients with deteriorated performance measured in higher Eastern Cooperative Oncology Group (ECOG) grade were more frequent in the NDLC group compared to the DLC group (p=0.003).

2. Landscape of somatically altered genes and survival analysis

Distribution of somatically altered genes found in the COSMIC database is summarized in oncoplot (Fig. 2), regarding several oncogenic pathways: TP53, NOTCH, RTK-RAS, DNA repair, PI3K, NRF2, cell cycle, and androgen receptor (AR). Alterations within *NOTCH2* (25/41, p=0.017) and *BCL2* (7/41, p=0.018) gene were enriched in the NDLC group. Meanwhile, alterations in *TSHR* (9/47, p=0.014), *MSH2* (11/47, p=0.014), *NOTCH1* (16/47, p=0.031), and *CHEK2* (15/47, p=0.049) were enriched in the DLC group.

In addition to the oncogenic pathways, we further investigated several COSMIC

signatures: Aging, APOBEC, smoking, BRCA1/2, MMR, UV, POLE, and TMZ. Distribution of altered somatic signatures found in patients in terms of oncogenic pathways and the COSMIC signature are summarized in S1 Table. Alteration of smoking signature and altered DNA repair pathway were more frequent in the DLC group compared to the NDLC group (46.8% vs. 24.4%, $p=0.045$, Fig. 3B and 76.6% vs. 53.7%, $p=0.027$, Fig. 3C, respectively). Univariate and multivariate survival analysis in terms of local failure free survival was conducted (S2 Table). Multivariate Cox proportional regression model revealed that the completion of RT (HR 0.21, 95% CI=0.06-0.74, $p=0.014$) and alteration of DNA repair pathway were significant factors for better local failure-free survival (HR 0.40, 95% CI=0.19-0.86, $p=0.018$, Fig. 2D). Although altered smoking signature showed numerically better local failure-free survival, the difference was not statistically significant (HR 0.65, 95% CI=0.32-1.31, $p=0.232$). Logistic regression analysis was conducted to investigate related factors in the DLC subgroup (S3 Table). Altered smoking signature (OR=2.73, 95% CI=1.09-6.81, $p=0.031$), performance status (OR=0.37, 95% CI=0.19-0.74, $P=0.005$), and altered DNA pathway (OR=2.83, 95%CI=1.13-7.04, $p=0.026$) were statistically significant factors.

3. Pattern of structural variation

There was no clear pattern of distribution of SV's including insertion, inversion, duplication, deletion, translocation, and total number of SV counts (Fig. 4A). When distribution of various SV's were compared between the groups, duplication event was more frequent in the DLC group ($p=0.040$, Fig. 4B). Among duplicated genes, *PUS1*, *POLE*, *FANCD2*, *IGFRI*, *ZNF726* were duplicated for the DLC group, meanwhile, *FANCD2* and *NRG1* were duplicated for the NDLC group. In the circos plots, patterns of SV and CNV for the NDLC group (Fig. 4C) and the DLC group (Fig. 4D) were depicted, and differential findings in each group were

also described. For example, *NOTCH3* deletion was observed in the NDLC group, meanwhile, *BRCA*, *RAD21*, *ATR* deletion was found in the DLC group. Copy number amplification of *NOTCH2*, *ERBB2*, *CDK12*, *TERT* was found in the NDLC group, whereas, *MYC*, *NOTCH3* in the DLC group. Furthermore, we investigated the fusion event per diagnosis in both groups. *PDGFRB* gene fusion in colorectal cancer was only found in the NDLC group, whereas melanoma with *ALK* gene fusion was only found in the DLC group. In particular, *NTRK3* gene fusion was found in sarcoma and inflammatory myofibroblastic tumor, only in the DLC group. Overall, compared to the NDLC group, more fusion events were observed in the DLC group (Fig. 4E and 4F).

4. The Bayesian network model and its application

To integrate these results with clinical factors, we sought a model that can consider both genomic and clinical information to predict local control after RT. Finally, we constructed the Bayesian network model with significant somatic signatures, SV, CNV, diagnosis, and RT-related factors (Fig. 5A). Regarding somatic signatures, the alteration of DNA repair pathway and smoking signature was adopted. Regarding domain knowledge in radiation oncology, the model integrated RT parameters including RT site, RT dose (BED_{10}), the intent of RT, and the completion of RT, all of which are considered to be important for local control. This model used both clinical and genomic information to predict the probability of local control at 6-month. Additionally, we built a model using clinical information only (S4A Fig.) and genomic information only (S4B Fig.), respectively. Model were compared in terms of AUC and AUPRC calculated from 10-fold validation. There was no statistically significant differences between clinico-genomic and clinical models with ROC curve comparison ($p=0.530$, Fig. 5B), between genomic and clinico-genomic model ($p=0.385$), and clinical and genomic model ($p=0.325$),

respectively. Regarding the precision-recall curve, clinico-genomic model showed the highest AUPRC of 0.879, followed by clinical model of 0.861, and genomic model of 0.834 (Fig. 5C). AUPRC difference between the employed models was statistically significant for clinico-genomic model and genomic model (AUPRC difference=0.044, 95% bootstrap CI=0.010-0.091), whereas not significant between clinico-genomic model and clinical model (AUPRC difference=0.017, 95% bootstrap CI=-0.009-0.053) and clinical model and genomic model (AUPRC difference=0.027, 95% bootstrap CI=-0.005-0.059).

5. Clinical application clinical scenarios

There was a difference of the priority between clinical information and genomic information in each cancer for predicting the probability of local control (Fig. 5D). For example, to predict local control after RT, impact of clinical information was dominant in adenoid cystic carcinoma, meanwhile genomic information was more important in the hypopharyngeal cancer and choroid plexus carcinoma. However, application of derived model in clinical practice should be done with caution as it was developed from very limited number of patients.

The other usage of the model could be target optimization. This model can seek optimal combinations of clinical and genomic parameters to achieve high local control after RT. For example, when a patient with pancreatic cancer undergoes full course RT with dose of $BED_{10} > 60$ Gy and tumor demonstrates no DNA repair pathway mutation, then maximized local control could be expected if tumor carries smoking signature and *CDKN2B* amplification (S4D Fig.).

With the Bayesian network model, we can estimate the probability of local control at 6-month in various genomic and clinical parameter settings. For example, when palliative RT was given to painful bone metastases in patients with renal cell carcinoma, we can estimate local control probability at 6-month per prescribed RT dose in BED_{10} . It would be very unlikely

to be controlled for RT dose of $BED_{10} < 40$ Gy (S5A Fig.), approximately 39.6% control could be expected for RT dose of $BED_{10} \leq 60$ Gy (S5B Fig.), and over 99.8% probability for RT dose of $BED_{10} > 60$ Gy (S5C Fig.), respectively. Low probability of local control by low dose reflects the real world situation and the clinical domain knowledge.

Discussion

We developed a computational model to predict RT response of a measurable lesion using clinical and genomic information. Prediction was based on an individual characteristic including diagnosis, treatment parameters, somatic mutation, and structural/copy number alterations. Although quite crude in current stage, recommendation to undergo RT could be personalized based on clinical and genomic information. This in turn could collectively optimize the treatment for an individual patient, and serve as patient selection tool for clinical trials.

In clinical setting, tumor histology is an important factor for predicting radiation response [17]. Breast and prostate primaries are considered as radiosensitive tumors. In contrast, non-small lung cancer, melanoma, renal cell carcinoma, and sarcoma are considered as radioresistant tumors. In terms of local control, radiation dose, the completion of RT, and the intent of RT are important for radiation oncologists. Given that many factors are involved in RT response, our model has a clear advantage of integrating both genomic and clinical factors.

The Bayesian network is the probabilistic graphical model [18] that can be developed from multi-dimensional data and human domain knowledge. This statistical framework can provide causal inference and optimization. Further, the Bayesian network model can integrate complex data, optimize the target variable, and perform contribution analysis. Specific domain knowledge elicited by expertise can also be incorporated in the model. Luo et al. proposed hierarchical relationships based on the Bayesian network to predict tumor local control in 68

non-small cell lung cancer patients before and during RT [19]. Furthermore, Luo et al. developed more advanced Bayesian Network model to predict radiation pneumonitis after RT [20]. Various types of datasets including single nucleotide polymorphisms, micro RNAs, cytokines, radiation dosimetric data were used in the model. These factors were traditionally reported to be associated with radiation pneumonitis. In line with the aforementioned studies, current study also showed that clinical domain knowledge and heterogenous dataset could be integrated in the Bayesian Network machine learning framework.

Clinico-genomic model may have additional implication over clinical model alone. First, we benchmarked the performance of derived clinical, genomic, and clinico-genomic models. We found that clinico-genomic model is superior to other models regarding AUPRC, which is considered more reliable evaluation metrics in an unbalanced dataset [21]. In a similar study by Oh et al., the Bayesian Network model based on clinical, dosimetric, and various blood borne information demonstrated a slightly higher performance, compared to the model based on clinical and dosimetric information [22]. Thus, systemic approach for individual would be feasible using the clinico-genomic model. Second, the model output is the probability that can be easily interpreted by clinicians. The Cox proportional hazard model, which is commonly used in clinical research, cannot estimate the survival probability at any particular time. Meanwhile, the Bayesian Network model can provide clinicians with the probability in real time. This makes it possible to cope with many challenging clinical scenarios. Third, we further investigated the impact of given information on specific cancer type. We found the diversity of clinical and genomic information that contributed to the probability of local control. These results could be helpful in decision making process. Similarly, Penson et. al. used targeted panel DNA sequencing data and clinical information including age and gender to predict cancer histology with machine learning algorithm [23]. By using the random forest classifier, authors

CANCER RESEARCH AND TREATMENT (CRT)

showed that most informative individual features for predicting 22 tumor types with random forest classifier. Thus, a selective use of machine learning can provide clinical implications to the oncologists.

There are several limitations. The number of study population was quite limited and harbored cancer type was various. In particular, the interpretation of contribution analysis should be done with utmost caution, as the number of patients allocated to each cancer type was extremely small resulting in very limited generalizability. Many patients in this study were heavily treated with various systemic treatments including chemotherapy and hormonal therapy. As detail of the systemic treatment was not introduced in the model, it might serve as underlying confounding factor. However, we prioritized the domain knowledge in radiation oncology as clinical information. Because, first, response to prior systemic treatment was evaluated. More importantly, local control of irradiated lesion is more likely to be influenced by the local treatment given, namely RT in current clinical setting. Although there is a possibility of collinearity between clinical variables and diagnosis or genomic profile, the current Bayesian network model was constructed based on Augmented Naive Bayes algorithm assuming that a variable is dependent on its target class and other variables. It is our understanding that the Bayesian network does not require strict independence of random variables. In conclusion, durable radiation response was associated with alterations of DNA repair pathway and smoking somatic signature. Bayesian network model enriched with genomic information could provide helpful insights for high precision radiotherapy. However, these findings should be verified comprehensively in more homogeneous and prospective cohort for further individualization.

Ethical Statement

This study was approved by the institutional review board (IRB No. H-2105-142-1220).

Author Contributions

Conceived and designed the analysis: Jang BS, Kim TY, Chie EK.

Collected the data: Chang JH, Jeon SH, Song MG, Lee KH, Im SA, Kim JI, Kim TY, Chie EK.

Contributed data or analysis tools: Chang JH, Jeon SH, Song MG, Lee KH, Im SA, Kim JI, Kim TY, Chie EK.

Performed the analysis: Jang BS, Kim TY.

Wrote the paper: Jang BS.

Conflicts of Interest

Conflict of interest relevant to this article was not reported.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Ministry of Science and ICT, Republic of Korea (NRF-2020R1F1A1073616). The biospecimens and data used in this study were provided by the Biobank of Seoul National University Hospital, a member of Korea Biobank Network.

The SNUH FiRST Cancer Panel v3.1 used in this study was supported by the grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea [grant number: HI14C1277].

References

1. Hyman DM, Taylor BS, Baselga J. Implementing Genome-Driven Oncology. *Cell*. 2017;168:584–99.
2. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med*. 2017;23:703–13.
3. Mateo J, Chakravarty D, Dienstmann R, Jezdic S, Gonzalez-Perez A, Lopez-Bigas N, et al. A framework to rank genomic alterations as targets for cancer precision medicine: the ESMO Scale for Clinical Actionability of molecular Targets (ESCAT). *Ann Oncol*. 2018;29:1895–902.
4. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer. *J Mol Diagnostics*. 2017;19:4–23.
5. Pollard JM, Gatti RA. Clinical Radiation Sensitivity With DNA Repair Disorders: An Overview. *Int J Radiat Oncol Biol Phys*. 2009;74:1323–31.
6. Bergom C, West CM, Higginson DS, Abazeed ME, Arun B, Bentzen SM, et al. The Implications of Genetic Testing on Radiation Therapy Decisions: A Guide for Radiation Oncologists. *Int J Radiat Oncol Biol Phys*. 2019;105:698–712.
7. Forker LJ, Choudhury A, Kiltie AE. Biomarkers of Tumour Radiosensitivity and Predicting Benefit from Radiotherapy. *Clin Oncol*. 2015;27:561–9.
8. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45:228–47.
9. Suh J, Jeong CW, Choi S, Ku JH, Kim HH, Kim K, et al. Sharing the initial experience

- of pan-cancer panel analysis in high-risk renal cell carcinoma in the Korean population. *BMC Urol.* 2020;20:125.
10. Park C, Kim M, Kim MJ, Kim H, Ock C-Y, Keam B, et al. Clinical application of next-generation sequencing-based panel to BRAF wild-type advanced melanoma identifies key oncogenic alterations and therapeutic strategies. *Mol Cancer Ther.* 2019;19:molcanther.0457.2019.
 11. Li H. Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. *Bioinformatics.* 2012;28:1838–44.
 12. Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr Protoc Bioinforma.* 2013;43:11.10.1-11.10.33.
 13. Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* 2016;17:1–11.
 14. Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018;28:1747–56.
 15. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular visualization in R. *Bioinformatics.* 2014;30:2811–2.
 16. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics.* 1988;44:837.
 17. Laufer I, Rubin DG, Lis E, Cox BW, Stubblefield MD, Yamada Y, et al. The NOMS Framework: Approach to the Treatment of Spinal Metastatic Tumors. *Oncologist.* 2013;18:744–51.

18. Husmeier D, Dybowski R, Roberts S, editors. Probabilistic Modeling in Bioinformatics and Medical Informatics. London: Springer London; 2005. doi:10.1007/b138794.
19. Luo Y, McShan D, Ray D, Matuszak M, Jolly S, Lawrence T, et al. Development of a Fully Cross-Validated Bayesian Network Approach for Local Control Prediction in Lung Cancer. *IEEE Trans Radiat Plasma Med Sci.* 2019;3:232–41.
20. Luo Y, El Naqa I, McShan DL, Ray D, Lohse I, Matuszak MM, et al. Unraveling biophysical interactions of radiation pneumonitis in non-small-cell lung cancer via Bayesian network analysis. *Radiother Oncol.* 2017;123:85–92.
21. Ozenne B, Subtil F, Maucort-Boulch D. The precision-recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *J Clin Epidemiol.* 2015;68:855–9.
22. Oh JH, Craft J, Al Lozi R, Vaidya M, Meng Y, Deasy JO, et al. A Bayesian network approach for modeling local failure in lung cancer. *Phys Med Biol.* 2011;56:1635–51.
23. Penson A, Camacho N, Zheng Y, Varghese AM, Al-Ahmadie H, Razavi P, et al. Development of Genome-Derived Tumor Type Prediction to Inform Clinical Cancer Care. *JAMA Oncol.* 2020;6:84.

Table 1. Patient Characteristics

Variables	NDLC (n=41, 43.6%)	DLC (n=47, 54.4%)	p-value
Age at RT (year)	53 (18-76)	54 (29-80)	0.801
Sex			1.000
Female	25 (61.0%)	28 (59.6%)	
Male	16 (39.0%)	19 (40.4%)	
Specimen Type			0.082
Metastatic	21 (51.2%)	33 (70.2%)	
Primary	20 (48.8%)	14 (29.8%)	
RT Site			0.588
Abdomen	3 (7.3%)	4 (8.5%)	
Bone	17 (41.5%)	22 (46.8%)	
Brain	7 (17.1%)	4 (8.5%)	
Breast	1 (2.4%)	0 (0.0%)	
Chest wall	1 (2.4%)	1 (2.1%)	
Cutaneous	1 (2.4%)	0 (0.0%)	
Head and neck	3 (7.3%)	6 (12.8%)	
Lung	0 (0.0%)	2 (4.3%)	
Lymph node	3 (7.3%)	1 (2.1%)	
Mediastinum	1 (2.4%)	4 (8.5%)	
Pelvis	3 (7.3%)	3 (6.4%)	
Soft tissue	1 (2.4%)	0 (0.0%)	
RT Intent			0.213
Definitive	2 (4.9%)	7 (14.9%)	
Palliative	37 (90.2%)	35 (74.5%)	
Pre-op	0 (0.0%)	2 (4.3%)	
Salvage	2 (4.9%)	3 (6.4%)	
RT Modalities			0.529
2-dimensional radiotherapy	6 (14.6%)	5 (10.6%)	
3-dimensional conformal radiotherapy	8 (19.5%)	15 (31.9%)	
Brachytherapy	0 (0.0%)	1 (2.1%)	
Intensity modulated radiotherapy	22 (53.7%)	23 (48.9%)	
Stereotactic body radiotherapy	5 (12.2%)	3 (6.4%)	
ECOG performance status at RT (grade)			0.003
0	2 (4.9%)	14 (29.8%)	
1	25 (61.0%)	27 (57.4%)	
2	12 (29.3%)	4 (8.5%)	
3	2 (4.9%)	2 (4.3%)	
Completion of RT			0.093
No	5 (12.2%)	1 (2.1%)	
Yes	36 (87.8%)	46 (97.9%)	
Extracranial disease at RT			0.080
Absence	6 (14.6%)	15 (31.9%)	

CANCER RESEARCH AND TREATMENT (CRT)

Presence	35 (85.4%)	32 (68.1%)	
RT Dose (Gy)	30 (15-68)	30 (14-68)	0.604
RT Fractions	10 (1-30)	10 (1-38)	0.744
BED ₁₀ (Gy)			0.438
≤50	7 (17.1%)	12 (25.5%)	
>50	34 (82.9%)	35 (74.5%)	

RT, radiation therapy; BED₁₀, biologically effective dose with a/b=10; DLC, durable local control group; NDLC, non-durable local control group; ECOG, Eastern Cooperative Oncology Group. P-value was computed by Chi-square test

Accepted Article

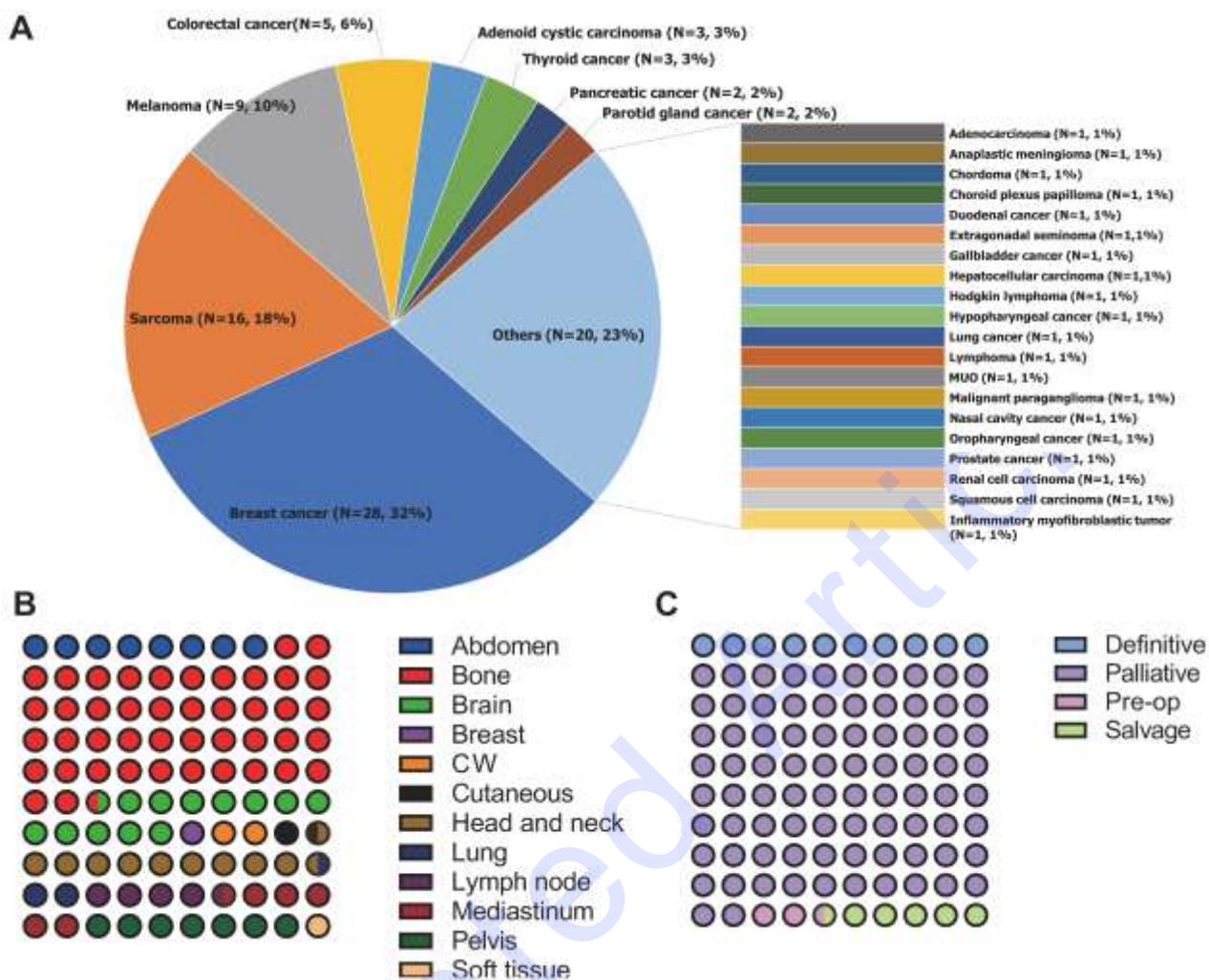


Fig. 1. (A) Distribution of primary cancer in the study cohort. (B) Dot plot representing RT site. (C) Dot plot showing the intent of RT. A dot represent one percent. RT, radiation therapy; CW, chest wall.

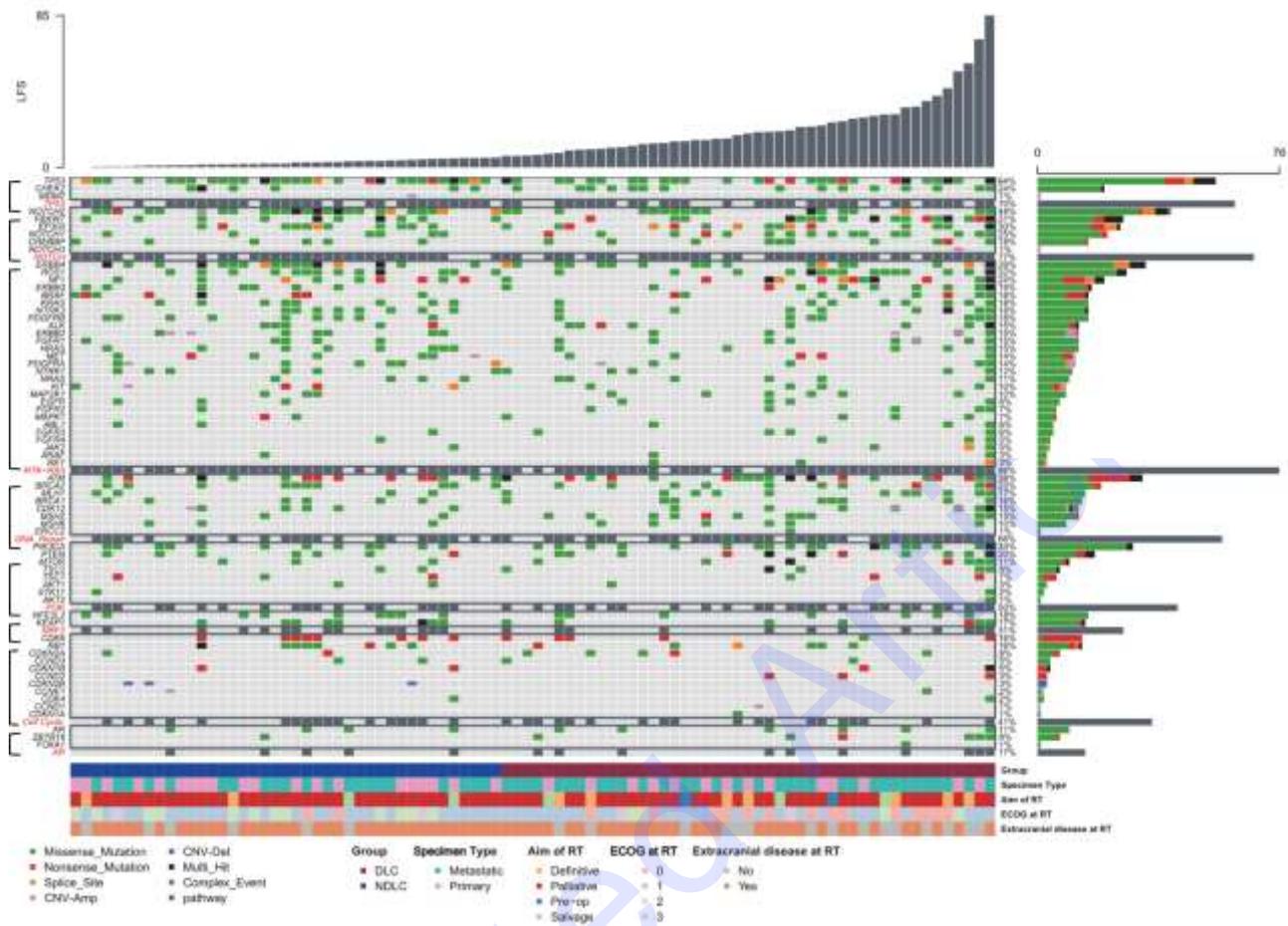


Fig. 2. Oncoplot showing the gene alterations found in Catalogue of Somatic Mutations in Cancer (COSMIC) database version 91. RT, radiation therapy; DLC, durable local control group; NDLC, non-durable local control group; ECOG, Eastern Cooperative Oncology Group performance status; CNV, copy number variation; Amp, amplification; Del, deletion; LFS, local-failure free survival.

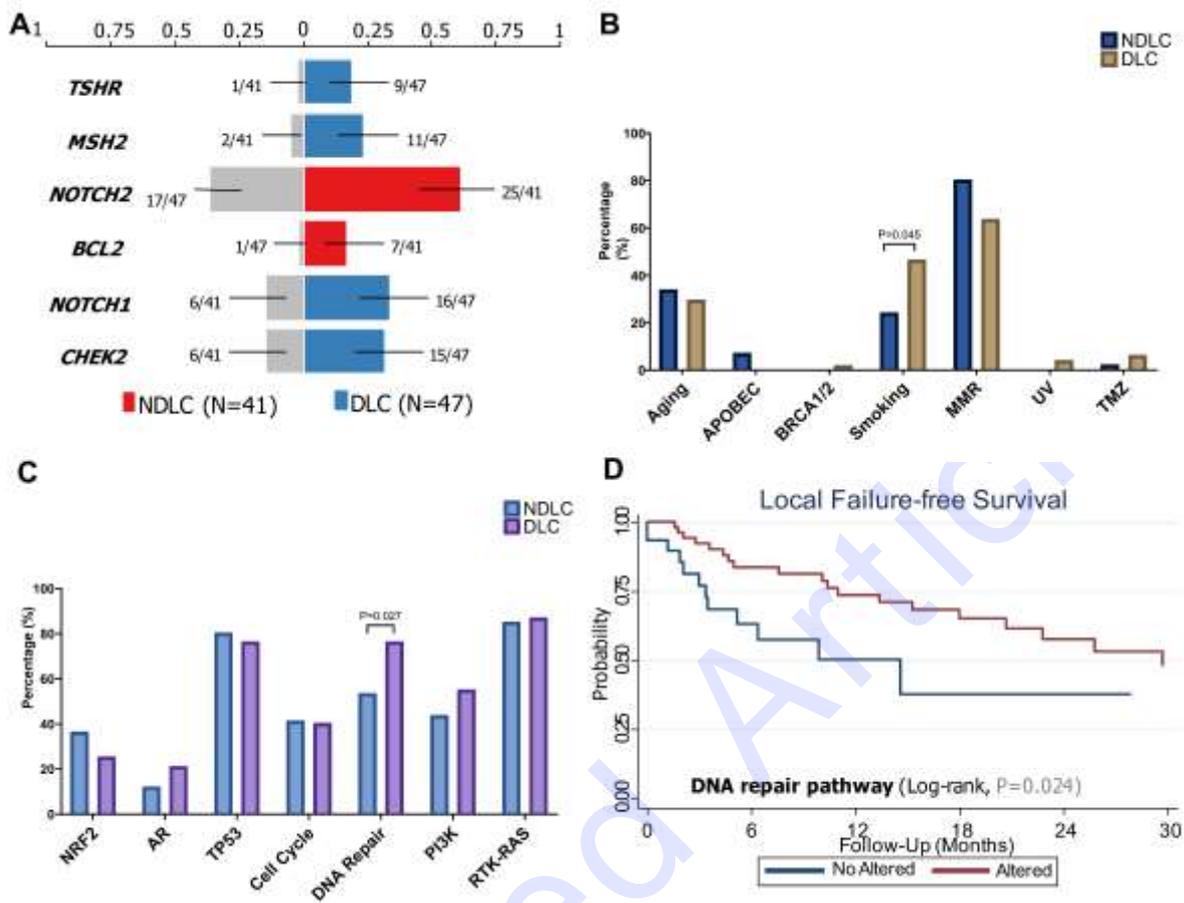


Fig. 3. (A) Bar plot showing the enrichment of gene mutation. The number of altered patient / a total number of patients are presented within bar plot. Bar plots representing percentage of patients having altered somatic signature (B) and pathway (C). P-value was estimated by Fisher's exact test. (D) Kaplan-Meier curve depicting local failure-free survival between the altered and the non-altered DNA repair pathway. P-value was computed by log-rank test. DLC, durable local control group; NDLC, non-durable local control group; MMR, mismatch repair; UV, ultraviolet; TMZ, temozolomide; AR, androgen-receptor.

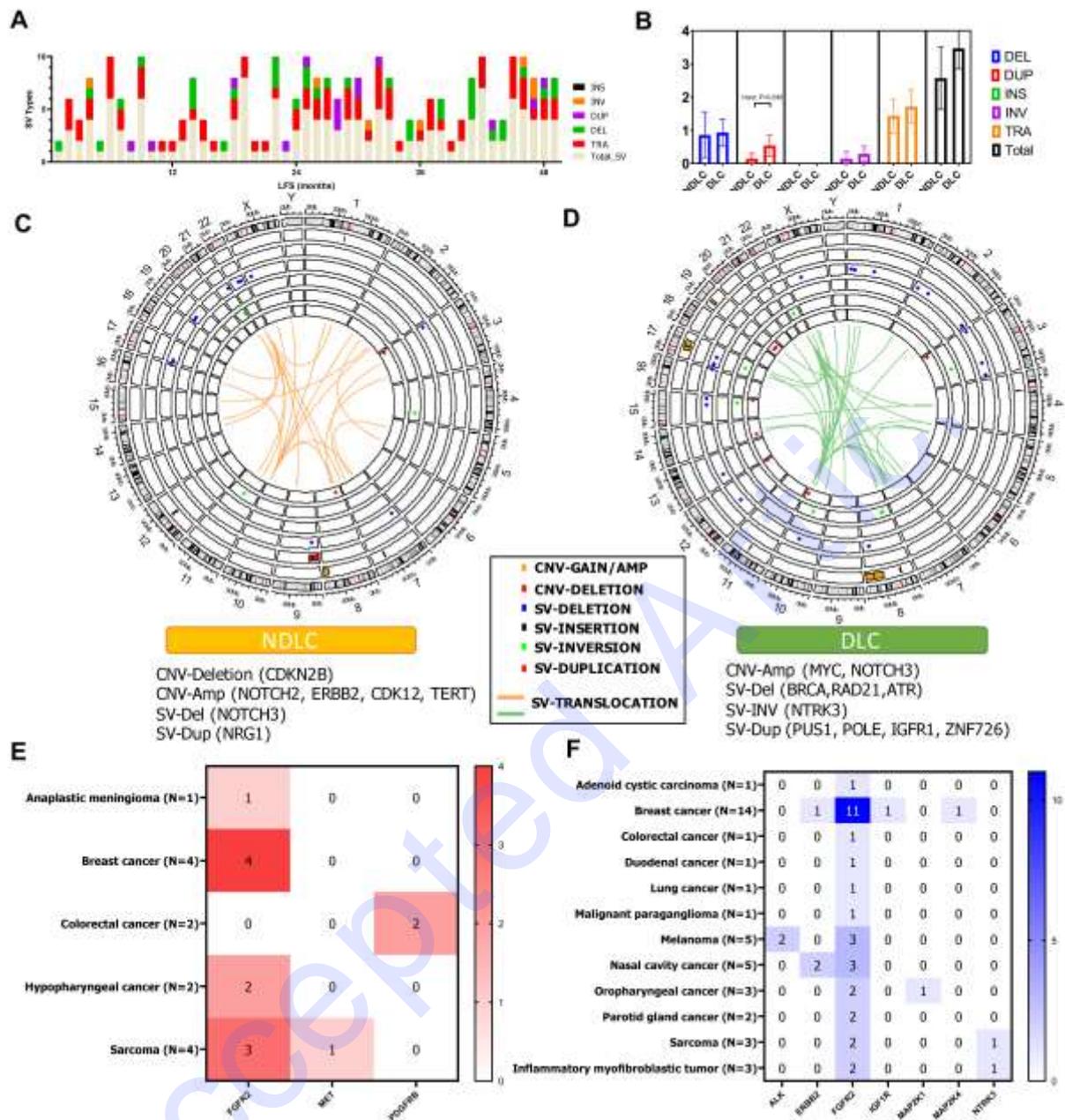


Fig. 4. (A) Cumulative bar plots showing the structural variation. Patients are sorted in ascending order in x-axis. (B) Bar plot comparing the number of structural variation events between the NDLC and the DLC groups. Circos plots depicting structural variation and copy number variations in the NDLC (C) and the DLC (D) groups. Heat map showing the number of patients having kinase fusion event in the NDLC (E) and the DLC (F) groups. DLC, durable local control group; NDLC, non-durable local control group; SV, structural variation; Del or DEL, deletion; Dup or DUP, duplication; Amp, amplification; INV, inversion; INS, insertion; TRA; translocation.

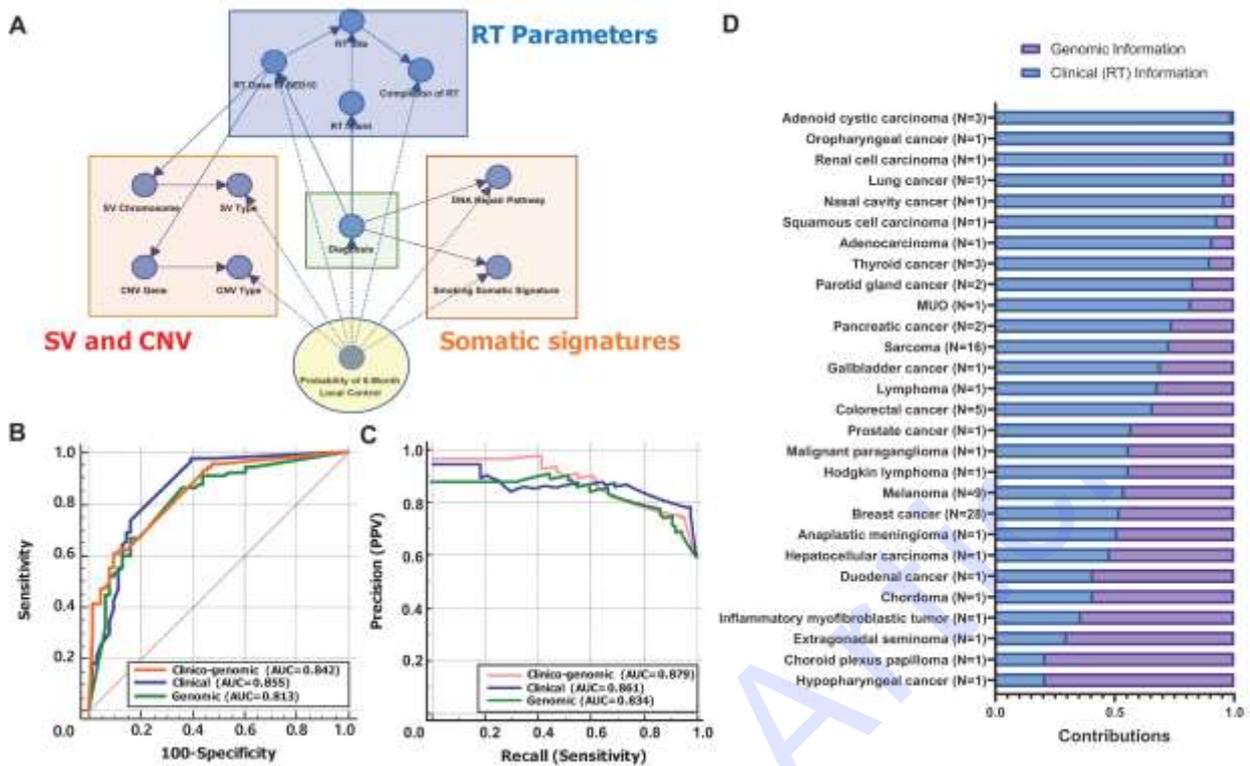


Fig. 5. (A) The Bayesian network model integrating genomic information and clinical domain knowledge. The final prediction is the probability of local control at 6 months after local RT (yellow circle). A receiver operating characteristic curve (B) and precision-recall curve (C) comparing clinico-genomic, clinical, and genomic Bayesian network models. Area under the curve values are also presented in the plots. (D) Regarding probability prediction, contribution of genomic and clinical information in each cancer are represented. RT, radiation therapy; SV, structural variation; CNV, copy number variation; PPV, positive predictive value.

S1 Table. Distribution of alteration of somatic signatures

Variables	NDLC (n=41, 43.6%)	DLC (n=47, 54.4%)	p-value
Aging			
No altered	27 (65.9%)	33 (70.2%)	0.819
altered	14 (34.1%)	14 (29.8%)	
APOBEC			
No altered	38 (92.7%)	47 (100.0%)	0.097
altered	3 (7.3%)	0 (0.0%)	
Smoking			
No altered	31 (75.6%)	25 (53.2%)	0.045
altered	10 (24.4%)	22 (46.8%)	
BRCA1/2			
No altered	41 (100.0%)	46 (97.9%)	1.000
altered	0 (0.0%)	1 (2.1%)	
MMR			
No altered	8 (19.5%)	17 (36.2%)	0.101
altered	33 (80.5%)	30 (63.8%)	
UV			
No altered	41 (100.0%)	45 (95.7%)	0.497
altered	0 (0.0%)	2 (4.3%)	
POLE			
No altered	38 (92.7%)	45 (95.7%)	0.661
altered	3 (7.3%)	2 (4.3%)	
TMZ			
No altered	40 (97.6%)	44 (93.6%)	0.620
altered	1 (2.4%)	3 (6.4%)	
NRF			
No altered	26 (63.4%)	35 (74.5%)	0.354
altered	15 (36.6%)	12 (25.5%)	
DNA Repair			
No altered	19 (46.3%)	11 (23.4%)	0.027
altered	22 (53.7%)	36 (76.6%)	
AR			
No altered	36 (87.8%)	37 (78.7%)	0.395
altered	5 (12.2%)	10 (21.3%)	
PI3K			
No altered	23 (56.1%)	21 (44.7%)	0.393
altered	18 (43.9%)	26 (55.3%)	
RTK_RAS			
No altered	6 (14.6%)	6 (12.8%)	1.000
altered	35 (85.4%)	41 (87.2%)	
TP53			

CANCER RESEARCH AND TREATMENT (CRT)

No altered	8 (19.5%)	11 (23.4%)	0.796
altered	33 (80.5%)	36 (76.6%)	
Cell Cycle			
No altered	24 (58.5%)	28 (59.6%)	1.000
altered	17 (41.5%)	19 (40.4%)	
NOTCH			
No altered	10 (24.4%)	10 (21.3%)	0.802
altered	31 (75.6%)	37 (78.7%)	

DLC, durable local control group; NDLC, non-durable local control group. p-value was estimate by Fisher's exact test.

Accepted Article

S2 Table. Cox proportional hazard model for local-failure free survival

Variables	Univariate			Multivariate		
	HR	95%CI	p-value	HR	95%CI	p-value
Age at RT (year), incremental	0.98	0.95-1.01	0.142			
Sex						
Female	1.00		0.090			
Male	1.81	0.91-3.58				
ECOG at RT, incremental	1.42	0.93-2.16	0.104			
BED ₁₀ (Gy)			0.223			
≤50	1.00					
>50	0.64	0.31-1.31				
Completion of RT			0.027			0.014
No	1.00			1.00		
Yes	0.25	0.08-0.85		0.21	0.06-0.74	
Extracranial disease at RT			0.526			
No	1.00					
Yes	1.28	0.60-2.73				
Aging signature			0.707			
No altered	1.00					
Altered	1.15	0.56-2.32				
APOBEC signature			0.284			
No altered	1.00					
Altered	3.03	0.40-22.96				
Smoking signature			0.232			
No altered	1.00					
Altered	0.65	0.32-1.32				
BRCA1/2 signature			0.972			
No altered	1.00					
Altered	0.00	∞				
MMR signature			0.464			
No altered	1.00					
Altered	1.32	0.63-2.80				
UV signature			0.239			
No altered	1.00					
Altered	2.39	0.56-10.17				
POLE signature			0.250			
No altered	1.00					
Altered	1.87	0.64-5.48				
TMZ signature			0.181			
No altered	1.00					
Altered	3.95	0.53-29.62				
NRF Pathway			0.884			
No altered	1.00					
Altered	0.95	0.45-2.01				
DNA Repair Pathway			0.028			0.018
No altered	1.00			1.00		

CANCER RESEARCH AND TREATMENT (CRT)

Altered	0.43	0.20-0.91		0.40	0.19-0.86
AR Pathway			0.286		
No altered	1.00				
Altered	0.59	0.23-1.55			
RTK-Ras Pathway			0.789		
No altered	1.00				
Altered	1.18	0.35-3.92			
TP53 Pathway			0.479		
No altered	1.00				
Altered	0.75	0.34-1.67			
Cell Cycle			0.328		
No altered	1.00				
Altered	0.69	0.33-1.45			
NOTCH Pathway			0.580		
No altered	1.00				
Altered	1.31	0.50-3.42			

HR, hazard ratio; CI, confidence interval; RT, radiation therapy; BED10, biologically effective dose with $a/b=10$; ECOG, Eastern Cooperative Oncology Group performance status; MMR, mismatch repair; UV, ultraviolet; TMZ, temozolomide; AR, androgen-receptor.

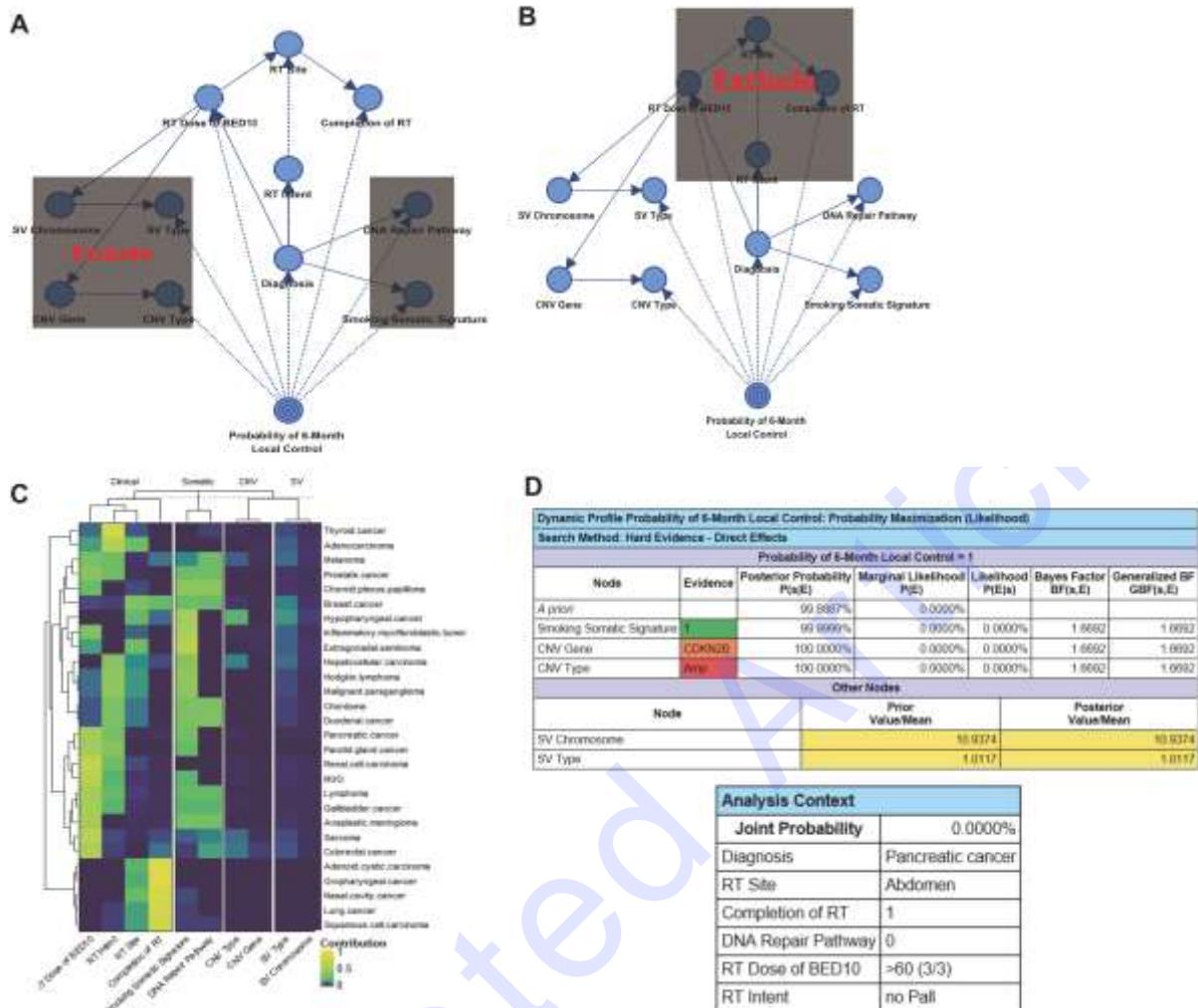
S3 Table. Logistic regression analysis for the DLC

	OR	95% CI	p-value
Age at RT (year), incremental	1.01	0.98-1.04	0.552
Sex			
Female	1.00		
Male	1.06	0.45-2.50	0.893
ECOG at RT, incremental	0.37	0.19-0.74	0.005
BED ₁₀ (Gy)			
≤50	1.00		
>50	1.46	0.61-3.52	0.397
Completion of RT			
No	1.00		
Yes	6.39	0.71-57.14	0.097
Extracranial disease at RT			
No	1.00		
Yes	0.37	0.13-1.06	0.063
Aging signature			
No altered	1.00		
Altered	0.82	0.33-2.01	0.662
APOBEC signature			
No altered	1.00		
Altered	1.00	N/A	
Smoking signature			
No altered	1.00		
Altered	2.73	1.09-6.81	0.031
BRCA1/2 signature			
No altered	1.00		
Altered	1.00	N/A	
MMR signature			
No altered	1.00		
Altered	0.43	0.16-1.13	0.086
UV signature			
No altered	1.00		
Altered	1.00	N/A	
POLE signature			
No altered	1.00		
Altered	0.56	0.09-3.55	0.541
TMZ signature			
No altered	1.00		

CANCER RESEARCH AND TREATMENT (CRT)

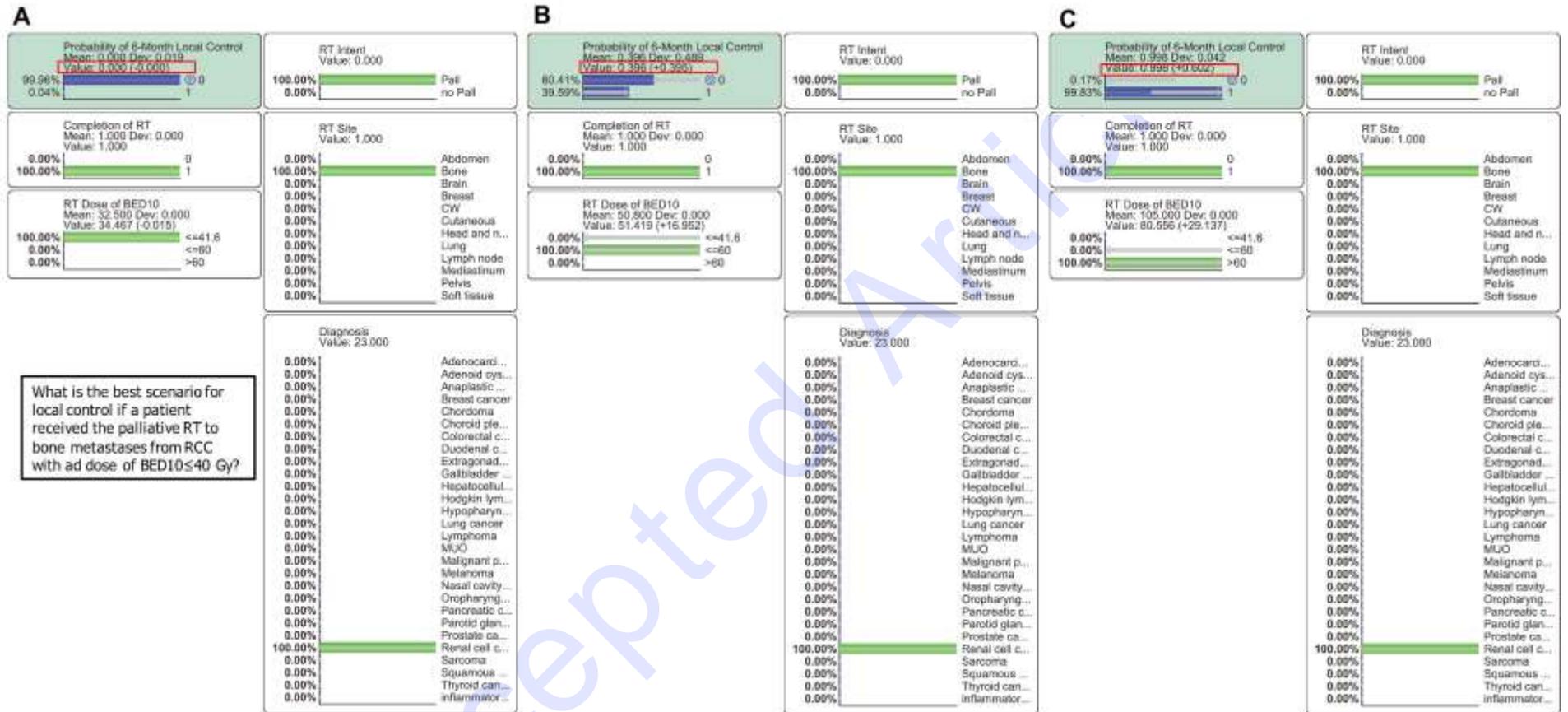
Altered	2.73	0.27-27.29	0.393
NRF Pathway			
No altered	1.00		
Altered	0.59	0.24-1.48	0.206
DNA Repair Pathway			
No altered	1.00		
Altered	2.83	1.13-7.04	0.026
AR Pathway			
No altered	1.00		
Altered	1.95	0.61-6.25	0.236
RTK-Ras Pathway			
No altered	1.00		
Altered	1.17	0.35-3.96	0.799
TP53 Pathway			
No altered	1.00		
Altered	0.79	0.28-2.21	0.658
Cell Cycle			
No altered	1.00		
Altered	0.96	0.41-2.25	0.921
NOTCH Pathway			
No altered	1.00		
Altered	1.19	0.44-3.24	0.728

OR, odds ratio; CI, confidence interval; RT, radiation therapy; BED10, biologically effective dose with $a/b=10$; ECOG, Eastern Cooperative Oncology Group performance status; MMR, mismatch repair; UV, ultraviolet; TMZ, temozolomide; AR, androgen-receptor.



S4 Fig. Bayesian network models only using clinical information (A) and using genomic profile (B). (C) Heatmap showing the detail contribution in clinical and genomic information according to primary cancer. (D) Example of target optimization using Bayesian network model for maximizing local control at 6 months after RT. RT, radiation therapy; SV, structural variation; CNV, copy number variation; BED, biologically effective dose; Amp, amplification.

CANCER RESEARCH AND TREATMENT (CRT)



S5 Fig. The predicted probability of local control at 6-month following RT with BED₁₀ < 40 Gy (A), BED₁₀ ≤ 60 Gy (B), and > 60 Gy (C) for a given clinical scenario (text box).