AWARD NUMBER: W81XWH-15-1-0681

TITLE: Genetic Modeling of Radiation Injury in Prostate Cancer Patients Treated with Radiotherapy

PRINCIPAL INVESTIGATOR: Harry Ostrer

CONTRACTING ORGANIZATION: Albert Einstein College of Medicine
Bronx, NY 10461

REPORT DATE: DECEMBER 2019

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
DECEMBER 2019	Final	30 Sep 2015 - 29 SEP 2019

4. TITLE AND SUBTITLE

Genetic Modeling of Radiation Injury in Prostate Cancer Patients Treated with Radiotherapy

5a. CONTRACT NUMBER
W81XWH-15-1-0681

5b. GRANT NUMBER
PC140371

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Barry S. Rosenstein and Harry Ostrer

E-Mail: barry.rosenstein@mssm.edu; harry.ostrer@einstein.yu.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Icahn School of Medicine at Mount Sinai
Albert Einstein College of Medicine

Dept. of Radiation Oncology
Box 1236,One Gustave Levy Place
New York, NY 10029
1300 Morris Park Avenue
Ullmann Building, Room 817
Bronx, NY 10461

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

An expanded meta-analysis was accomplished that identified three highly significant genomic signals (Pmeta<$5x10^{-8}$) with low Bayesian false discovery probability (<2%): single nucleotide polymorphism (SNP) rs17055178 with rectal bleeding (Pmeta=$6.2x10^{-10}$), rs10969913 with decreased urinary stream (Pmeta=$2.9x10^{-10}$) and rs11122573 with hematuria (Pmeta=$1.8x10^{-8}$). Fine scale mapping of these three regions identified a second independent signal (rs147121532) associated with hematuria (Pconditional=$4.69x10^{-6}$). Credible causal variants at these four signals lie in gene-regulatory regions and some modulate expression of nearby genes. Previously identified variants (rs17599026, rs7720298, and rs1801516) showed consistent associations in the new cohorts. In addition, we developed and tested TaqMan quantitative polymerase chain reaction (qPCR) assays for SNPs that were shown to be significant in the GWAS meta analyses.

15. SUBJECT TERMS

Radiogenomics, single nucleotide polymorphisms, prostate cancer, radiation therapy, adverse effects, urinary morbidity, rectal injury

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	20	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

1. INTRODUCTION

As with all forms of treatment for prostate cancer, the goal of radiotherapy is to provide patients with a sustainable cure of their tumor without causing substantial damage to normal tissues and organ function. Clearly, there have been great advances to conform the radiation field to the cancer. However, even with dosimetric improvements, some volume of normal tissue still receives a substantial radiation dose during the course of radiotherapy. This radiation exposure often results in toxicity that compromises organ function and affects the quality of life for the prostate cancer survivor. Therefore, an important goal is to create an assay that could predict which patients are most likely to develop radiation-induced complications. The main approach taken in recent years to achieve this goal has been the identification of genetic markers, primarily single nucleotide polymorphisms (SNPs), which are associated with the development of adverse effects resulting from radiotherapy. The aim of this research is to identify the genetic markers that can serve as the basis for personalized radiotherapy in which cancer management is formulated so that it optimizes the treatment plan for each patient based upon their genetic background. The overall objective of this research project is to create a robust, validated, sensitive and specific SNP-based assay that will be ready for implementation in the clinical setting. This assay will be capable of predicting the risk of developing adverse effects resulting from radiotherapy treatment of prostate cancer -- urinary morbidity and rectal injury. The purpose of the current project is to validate previously identified SNPs and to discover new SNPs in a large, independent cohort and to develop a predictive instrument and companion diagnostic.

2. KEYWORDS:

Radiogenomics, single nucleotide polymorphisms, prostate cancer, radiation therapy, adverse effects, urinary morbidity, rectal injury, sexual dysfunction

3. ACCOMPLISHMENTS:

What were the major goals of the project?

- *Validate previously discovered SNPs and identify additional SNPs via meta-analysis of GWAS using a substantially expanded set of studies in which approximately 7,000 men treated with radiotherapy for prostate cancer have been genotyped using a SNP array that contains a set of genome-wide SNPs as well custom content that contains our previously identified SNPs. (Months 1-18).*

We substantially expanded this meta-analysis during the third year of the project thereby discovering new SNP associations and validating previously identified SNPs as described below.

- *Create polygenic risk models from results of single-SNP analysis and investigate effects of demographic, dosimetric and clinical factors on polygenic risk models. (Months 12-30).*

A substantial effort was devoted during the past year to the development of models using polygenic risk score and machine learning methods, but unfortunately it was not possible to achieve predictive performance, assessed by the area under the receiver operating characteristic curve, that was statistically better than chance. This negative result is likely a consequence of the relatively small number of samples used for these analyses and the modest effect sizes associated with the SNPs constituting these data sets.

- *Develop a low-cost, high-performance genetic assay (Months 1-34)*

TaqMan assays were developed and validated.

- *Export the models developed in Aim 2 to a web-based application that could be used by physicians in practice and/or genetic testing laboratories. (Months 24-36)*

Rather than moving ahead with a web-based application, we have instead disseminated the models to the research community via peer-reviewed publication and presentation at the ASTRO and RGC annual meetings. The next step is to validate these models in an independent clinical study, which is underway through an NIH-funded SBIR Phase II project shared with L2 Diagnostics, LLC.

What was accomplished under these goals?

KEY RESEARCH ACCOMPLISHMENTS:

Expansion of the of the GWAS meta-analysis

During the past year we substantially increased the size of the cohort examined by conducting an individual patient data meta-analysis of six European-ancestry genome-wide association studies (n=3,871) in radiotherapy-treated prostate cancer survivors. Radiotoxicity was graded prospectively in all studies. Cox proportional hazards regression was used to test associations of ~6 million genotyped or imputed variants with urinary and rectal toxicity endpoints (time to first ≥grade 2 event). The meta-analysis of the European cohorts identified three highly significant genomic signals ($P_{meta}<5\times10^{-8}$) with low Bayesian false discovery probability (<2%): single nucleotide polymorphism (SNP) rs17055178 with rectal bleeding ($P_{meta}=6.2\times10^{-10}$), rs10969913 with decreased urinary stream ($P_{meta}=2.9\times10^{-10}$) and rs11122573 with hematuria ($P_{meta}=1.8\times10^{-8}$). Fine scale mapping of these three regions identified a second independent signal (rs147121532) associated with hematuria ($P_{conditional}=4.69\times10^{-6}$). Credible causal variants at these four signals lie in gene-regulatory regions and some modulate expression of nearby genes. Previously identified variants (rs17599026, rs7720298, and rs1801516) showed consistent associations in the new cohorts. This study increases understanding of the architecture of common genetic variants affecting radiotoxicity, points to novel radiobiology mechanisms, and shows further multi-national radiogenomics studies in larger cohorts are worthwhile.

This work included individuals with prostate adenocarcinoma, treated with radiotherapy with curative intent, and followed prospectively for development of urinary and rectal toxicity. All participants gave informed consent, and cohorts were collected following standards indicated by the Declaration of Helsinki. Individuals were excluded if: DNA samples/genotyping failed quality control measures; they had non-European (or non-Japanese) ancestry; and/or data were not available on androgen deprivation therapy, prior prostatectomy, age at treatment, and total biological effective dose (BED).

Participants were assessed prospectively for urinary and rectal toxicity from six months up to 5-years after radiotherapy, with the exception of the UGhent cohort where the maximum follow-up was 3-years. Assessment times were binned into 6-month intervals to enable time-to-event analysis without introducing bias due to variation in exact times of assessment resulting from variation in follow-up clinic scheduling. Four individual toxicity endpoints were analyzed: increased urinary frequency, decreased urinary stream, hematuria, and rectal bleeding. Toxicity grades were harmonized to achieve comparability across cohorts.

Germline DNA from whole blood was genotyped as part of previously completed GWAS. The CCI-EBRT cohort and batch I of the GenePARE cohort (GenePARE-I) were genotyped on the Affymetrix SNP6.0 array (Affymetrix, Inc.; Santa Clara, CA); batch I of RAPPER (RAPPER-I) was genotyped on the Illumina CytoSNP12 array (Illumina, Inc.; San Diego, CA); batch II of GenePARE (GenePARE-II), batch II of RAPPER (RAPPER-II), RADIOGEN, UGhent, and CCI-BT were genotyped on the Illumina OncoArray-500K BeadChip (Illumina, Inc.; San Diego, CA). After filtering, all datasets had genotyping rates >99%. To minimize potential confounding by population structure, individuals with non-European ancestry (European GWAS meta-analysis cohorts) or non-Japanese ancestry (Asian replication cohorts) were excluded based on principal component analysis performed with samples of known ancestry.

Each variant was tested for association with each toxicity endpoint using Cox proportional hazards regression adjusting for covariates selected to reduce heterogeneity in the meta-analysis. In RAPPER and GenePARE, where samples were genotyped in two batches using different arrays, batch (0/1) was included as a binary variable in Cox regression models. Toxicity was defined as time to onset of first occurrence of grade 2 or higher with time binned at six-month intervals. Efron's method was used to break ties. A fixed-effects meta-

analysis using inverse variance weighting was used to combine genetic variant-toxicity association results across studies. A chi-squared test of heterogeneity was performed for each variant. Variants were considered significant if the meta-analysis p-value (P_{meta}) was $<5x10^{-8}$ and the heterogeneity p-value was >0.05. Bayesian false discovery probabilities (BFDP) were calculated as an additional measure of confidence. Clinical variables were tested individually for association with radiotherapy toxicity using Cox proportional hazards regression stratified by cohort. Variables were then combined with genetic variants in multivariable Cox models with a separate model developed for each toxicity outcome. Variables with likelihood ratio p-value <0.05 in the full model were retained. Data management and statistical analyses were conducted using R (version 3.2.2, R Foundation for Statistical Computing, Vienna, Austria), ProbABEL, and Stata (version 14.2, StataCorp LLC, College Station, TX). Pascal was used to compute gene and pathway scores from the genome-wide association results.

Genomic regions for fine scale mapping were defined as the 1Mb interval surrounding each significant independent association. We re-imputed genotypes for the non-directly-genotyped variants at these regions using IMPUTE2 and a reference panel using the standard IMPUTE2 MCMC algorithm for follow-up imputation to improve accuracy at low frequency variants. Variants with imputation info score ≥0.3 in all cohorts and MAF ≥0.02% in at least one cohort were included in the analysis. 4,190 variants across the chr1:230337180-231337180 region; 3,776 at chr5:156903410_157903410 and 3,987 at chr9:30366808-31366808 were evaluated for hematuria, rectal bleeding or decreased urinary stream risk, respectively.

For each cohort, we ran Cox regression independently and meta-analyzed the results, using a fixed-effects meta-analysis (*meta*, https://mathgen.stats.ox.ac.uk/genetics_software/meta/meta.html). Then, the most significant variant (index variant at signal 1) was used to perform conditional analysis in each cohort independently. The conditional results were meta-analyzed and the most significant variant (index variant at signal 2) selected. This loop continued until no variants at p-values of 10^{-4} remained at the region. A preliminary set of credible causal variants (CCVs) was then determined among the variants within two orders of magnitude from the index variant for each signal. The most significant variant (final index variant) within the set was identified by adjusting the effect of each signal by the additional signals. The final credible set was redefined among the variants with p-values within two orders of magnitude smaller than the index variant after being conditioned by the additional index variants at the region.

To define the cumulative posterior probability of the credible set, we estimated the empirical Bayes Factor [21]. For each variant (i) we normalized its effect size ($\hat{\beta}_i$) and variance (σ_i) by its allele frequency (p_i) as follows

$$\beta_{Ni} = \hat{\beta}_i \sqrt{2p_i(1-p_i)}$$
$$\sigma_{Ni}^2 = \sigma_i^2 2p_i(1-p_i)$$

where p_i is the allele frequency for variant i in the OncoArray cohort, and estimated the prior variance (ω) using approach with normalized betas and normalized variance

$$\omega_N = \widehat{\beta_{N_{130}}^2} - \sigma_{Nm}^2$$

We then estimated the cumulative posterior probability of the variants included in the credible set. For regions with more than one independent signal Bayes Factor was estimated using the summary statistics from the conditional analysis, after adjusting for other index variants at the region.

Variants were annotated with Variant Effect Predictor to determine their effect on genes, transcripts, and protein sequence. To evaluated whether our CCVs were located at regulatory regions, we overlapped our CCVs with Encode enhancer-like and promoter-like regions for 73 tissues, primary cells, immortalized cell lines, and *in vitro* differentiated cells with available data for both enhancer- and promoter-like regions. In order to evaluate whether the CCVs could drive the expression of local genes, we accessed the GTEx Portal on 04/19/2018 to retrieve the metasoft results for all tissues in the V7 release. LocusZoom was used to visualize association results for regions containing CCVs. Linkage disequilibrium was estimated using as reference the European ancestry populations from the 1000 Genomes Project.

Table 1 summarizes the characteristics of the 3,871 participants included in the meta-analysis. Q-Q plots (Figure 1) show no genomic inflation, suggesting that population structure was adequately controlled using principal components analysis to restrict the study population to European ancestry individuals. Three independent SNPs reached statistical significance and had a low BFDP (<2%) (Table 2 and Figure 2).

Table 1. Patient characteristics by cohort for the 3,871 men included in the GWAS meta-analysis.

Characteristics	All Cohorts N=3,871	RAPPER, N=2,010	RADIOGEN, N=658	GenePARE, N=492	UGhent, N=311	CCI-BT, N=252	CCI-EBRT, N=148
Age at treatment, median (range)[a]	68 (43, 86)	68 (48, 84)	72 (47, 86)	65 (43, 85)	65 (49, 81)	65 (45, 79)	68 (45, 82)
NCCN risk group, n (%)							
Very low	545 (14.1)	133 (6.6)	100 (15.2)	172 (35.0)	43 (13.8)[b]	89 (35.3)	8 (5.4)
Low	258 (6.7)	82 (4.1)	23 (3.5)	61 (12.4)	21 (6.8)	68 (27.0)	3 (2.0)
Intermediate	2,635 (68.1)	1,566 (77.9)	447 (67.9)	232 (47.2)	173 (55.6)	95 (37.7)	122 (82.4)
High or Very high	410 (10.6)	229 (11.4)	82 (12.5)	27 (5.5)	57 (18.4)	0	15 (10.1)
Not available	23 (0.6)	0	6 (0.9)	0	17 (5.5)	0	0
Stage at diagnosis, n (%)							
T1a-c, T1x	1,443 (37.3)	709 (35.3)	226 (34.3)	249 (50.6)	101 (32.5)	119 (47.2)	38 (25.7)
T2a-c, T2x	2,020 (52.2)	1,084 (53.9)	362 (55.0)	227 (46.1)	126 (40.5)	132 (52.4)	89 (60.1)
T3a-c, T3x	305 (7.9)	182 (9.1)	54 (2.7)	16 (3.3)	37 (11.9)	0	16 (10.8)
T4	14 (0.4)	0	7 (1.1)	0	6 (1.9)	0	1 (0.7)
Not available	89 (2.3)	35 (1.7)	9 (1.4)	0	41 (13.2)	1 (0.4)	4 (2.7)
Gleason at diagnosis, n (%)							
≤6	1,702 (44.0)	605 (30.1)	403 (61.2)	310 (63.0)	142 (45.7)	212 (84.1)	30 (20.3)
7	1,653 (42.7)	1,109 (55.2)	176 (26.8)	124 (25.2)	107 (34.4)	40 (15.9)	97 (65.5)
≥8	265 (6.8)	56 (2.8)	70 (10.6)	58 (11.8)	60 (19.3)	0	21 (14.2)
Not available	251 (6.5)	240 (11.9)	9 (1.4)	0	2 (0.6)	0	0
Pre-treatment PSA, median (range)	8.9 (0, 236.0)	10.1 (0.6, 33.5)	9.7 (0.6, 236.0)	6.2 (0.6, 124.0)	6.6 (0, 150.0)[c]	6.3 (0.5, 16.0)	10.9 (1.4, 80.0)
Radical prostatectomy, n (%)[a]							
Yes	225 (5.8)	0	128 (19.5)	0	97 (31.2)	0	0
No	3,646 (94.2)	2,010 (100)	530 (80.5)	492 (100)	214 (68.8)	252 (100)	148 (100)
Androgen deprivation therapy, n (%)[a]							
Yes	3,047 (78.7)	2,010 (100)	463 (70.4)	248 (50.4)	198 (63.7)	55 (21.8)	73 (49.3)
No	824 (21.3)	0	195 (29.6)	244 (49.6)	113 (36.3)	197 (78.2)	75 (50.7)
Type of radiotherapy, n (%)							
3D-CRT	895 (25.4)	237 (11.8)	658 (100)	0	0	0	0
IMRT	2,239 (57.8)	1,773 (88.2)	0	7 (1.4)	311 (100)	0	148 (100)
Brachytherapy	534 (13.8)	0	0	282 (57.3)	0	252 (100)	0
Brachytherapy + EBRT	203 (5.2)	0	0	203 (41.3)	0	0	0
Total BED[d], median (range)[a]	123 (52, 292)	120 (107, 123)	123 (57, 127)	192 (52, 269)	136 (124, 136)	158 (80, 292)	121 (112, 134)

Abbreviations: NCCN, National Comprehensive Cancer Network; PSA, prostate specific antigen; 3D-CRT, three-dimensional conformal radiotherapy; IMRT, intensity modulated radiotherapy; EBRT, external beam radiotherapy (either 3D-CRT or IMRT); BED, biologic effective dose.

[a] Age at treatment, radical prostatectomy, androgen deprivation therapy, and total BED were included as covariates in the GWAS meta-analysis.

[b] NCCN risk group in the UGhent cohort was defined using pre-radiotherapy PSA rather than PSA at diagnosis.

[c] PSA measurement is pre-radiotherapy but post-prostatectomy in patients who received prior prostatectomy.

[d] Total BED was calculated using an α/β ratio of 3.

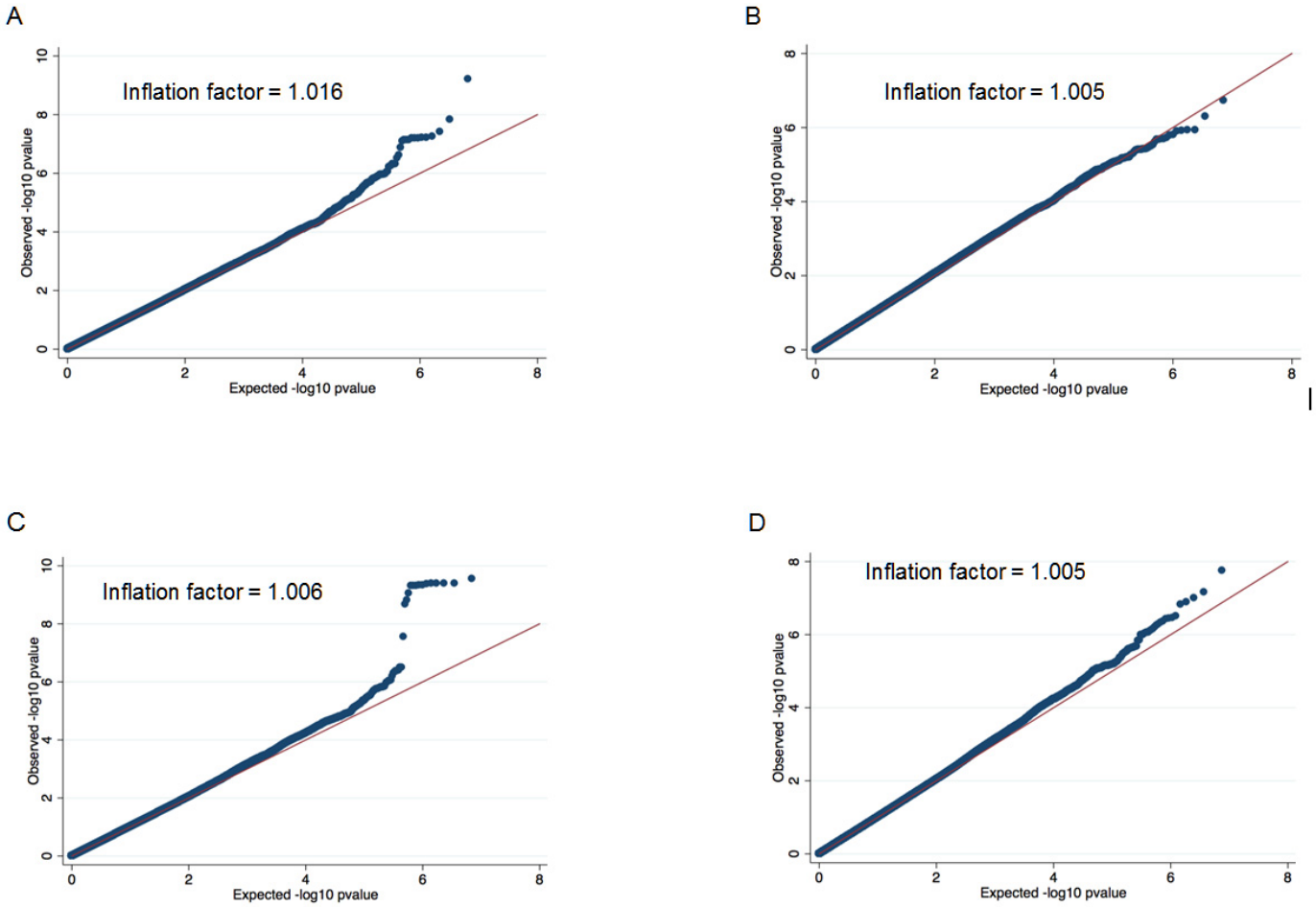

Figure 1. QQ plots showing expected and observed p-values from GWAS meta-analysis of rectal bleeding (A), increased urinary frequency (B), decreased urinary stream (C), and hematuria (D).

Table 2. Study-specific and meta-analysis results for new risk SNPs identified via GWAS meta-analysis. Bold values correspond to meta-analysis results.

Genetic variant	Toxicity outcome	Study	Info[a]	HR (95% CI)[b]	P_{meta}	P_{het}[c]	BFDP[d]
rs17055178	Rectal bleeding	**Meta-analysis**	-	**1.95 (1.58 to 2.40)**	**6.2×10^{-10}**	**0.61**	**0.09%**
chr5:157,403,410[e]		RAPPER	0.81, 0.99	1.78 (1.37 to 2.32)			
Minor allele G		RADIOGEN	0.99	2.58 (1.69 to 3.95)			
MAF[f] 0.09		GenePARE	NA[g]	NA[g]			
		UGhent	0.99	1.38 (0.18 to 10.4)			
		CCI-BT	0.99	2.01 (0.97 to 4.20)			
		CCI-EBRT	0.98	1.27 (0.38 to 4.25)			
rs10969913	Decreased urinary stream	**Meta-analysis**	-	**3.92 (2.57 to 6.00)**	**2.9×10^{-10}**	**0.08**	**1.07%**
chr9:30,866,808[e]		RAPPER	0.61, 0.95	1.86 (0.76 to 4.54)			
Minor allele G		RADIOGEN	0.95	2.03 (0.27 to 15.4)			
MAF[f] 0.05		GenePARE	0.99, 0.95	4.36 (2.55 to 7.46)			
		UGhent	NA[h]	NA[h]			
		CCI-BT	NA[i]	NA[i]			
		CCI-EBRT	0.95	14.3 (3.78 to 54.4)			
rs11122573	Hematuria	**Meta-analysis**	-	**1.92 (1.53 to 2.42)**	**1.8×10^{-8}**	**0.14**	**1.96%**
chr1:230,837,180[e]		RAPPER	0.99	1.42 (0.99 to 2.04)			
Minor allele T		RADIOGEN	0.99	2.40 (1.54 to 3.73)			
MAF[f] 0.06		GenePARE	0.99, 0.99	2.01 (1.25 to 3.22)			
		UGhent	0.99	3.59 (1.72 to 7.49)			
		CCI-BT	NA[j]	NA[j]			
		CCI-EBRT	1.000	0.99 (0.13 to 7.58)			

Abbreviations: SNP, single nucleotide polymorphism; HR, hazard ratio; CI, confidence interval; BFDP, Bayesian false discovery probability; MAF, minor allele frequency; NA, not analyzed.

[a] Imputation info score values in RAPPER are from the cytoSNP12 array and oncoarray respectively; values in Gene-PARE are from the AffySNP6.0 array and oncoarray respectively; values in all other studies are from the oncoarray.

[b] Hazard ratio corresponds to the minor allele with the major allele treated as the reference group.

[c] Heterogeneity p-value.

[d] BFPD estimated assuming a prior variance, W = 0.32^2, and prior probability of a non-null association 0.0001.

[e] Base position according to Genome Reference Consortium Human Build 37 (hg19).

[f] Minor allele frequency from PRACTICAL Oncoarray samples of European ancestry

[g] Rectal bleeding was assigned a single grade in GenePARE using information across all follow up assessments, and so this endpoint was not available for analysis using Cox proportional hazards modeling.

[h] Decreased urinary stream was not assessed in UGhent.

[i] Increased urinary frequency and decreased urinary stream were not analyzed in CCI-BT because pre-radiotherapy assessments were more than one year prior to starting radiotherapy for the majority of participants.

[j] Hematuria was not assessed in CCI-BT.

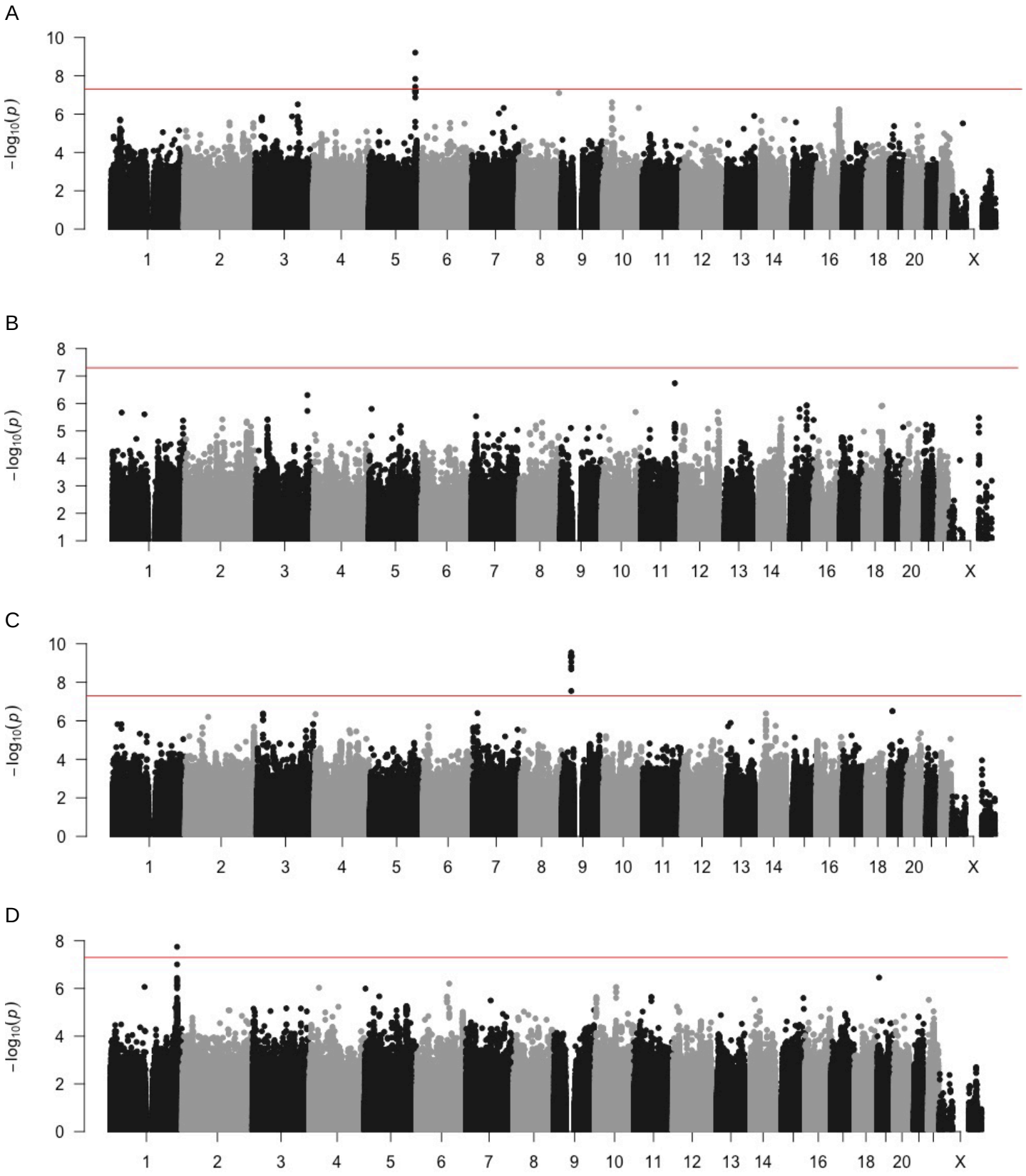

Figure 2. Manhattan plots showing pvalues from GWAS meta-analysis of rectal bleeding (A), increased urinary frequency (B), decreased urinary stream (C), and hematuria (D). Red line denotes p-value = 5x10-8. Numbers on the y-axis denote chromosome number.

Fine-scale mapping defined the CCVs within each of the significantly associated regions (Figure 3). There was evidence for two independent signals at chr1:230337180-231337180 associated with hematuria. The first signal includes 47 CCVs (together explaining 93% of the posterior probability of risk). These CCVs lie in active enhancer-/promoter-like gene-regulatory regions (Figure 4A). Their risk alleles decrease expression of *AGT* (encoding angiotensinogen; ENSG00000135744.7) and *COG2* (encoding conserved oligomeric Golgi complex subunit 2; ENSG00000135775.9) in multiple tissues including vascular arteries (Figure 4B). The second signal with 10 CCVs (explaining 54% of the posterior probability) has risk alleles that decrease expression of *CAPN9* (encoding the intestinal protease, calpain-9; ENSG00000135773.8) and *ARV1* (encoding ARE2 required for viability [ARV1] homolog, fatty acid homeostasis modulator; ENSG00000173409.9). The risk alleles in the second signal were also associated with differential expression of two non-coding RNA genes: decreased expression of ncRNA AL512328.1 (ENSG00000244137.1), which overlaps partially with *AGT* and *CAPN9*; and increased expression of ncRNA LOC101927604 (ENSG00000223393.1). At the chr5:156903410-157903410 region associated with rectal bleeding risk, one signal comprising 15 CCVs accounts for 98% posterior probability at the region (Figure 3B). CCVs in this region overlap active enhancer-like regions in gastrointestinal tissues like large intestine and esophagus (Figure 4A) but none were significantly associated with differential gene expression in GTEx. At the chr9:30366808-31366808 region, associated with decreased urinary stream, a single risk signal including 15 CCVs accounts for 99% of the posterior probability (Figure 3C). None were significantly associated with differential gene expression in the tissues evaluated by GTEX.

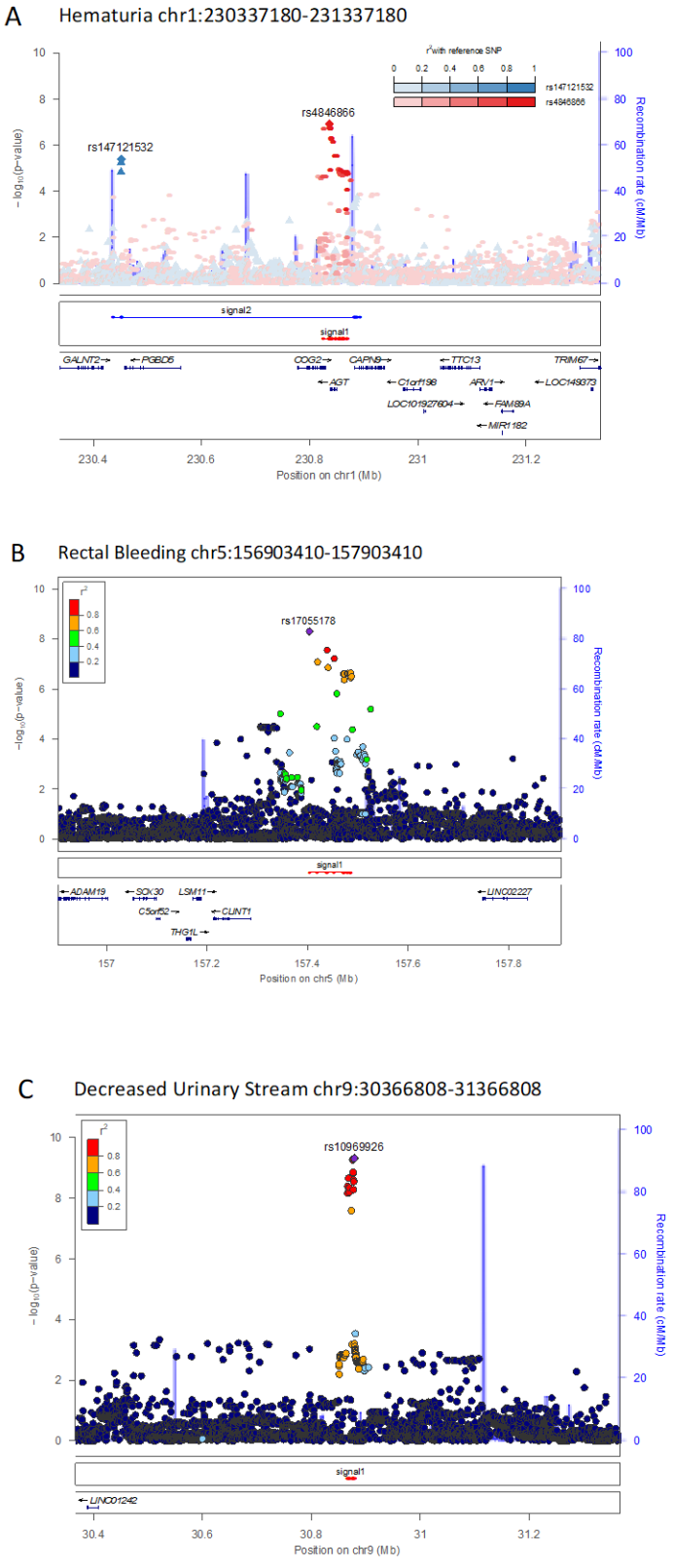

Figure 3. Regional Manhattan plots showing signals defined by fine-mapping of the hematuria risk region chr1:230337180-231337180 (A), rectal bleeding risk region chr5:156903410-157903410 (B), and decreased urinary stream risk region chr9:30366808-31366808 (C).

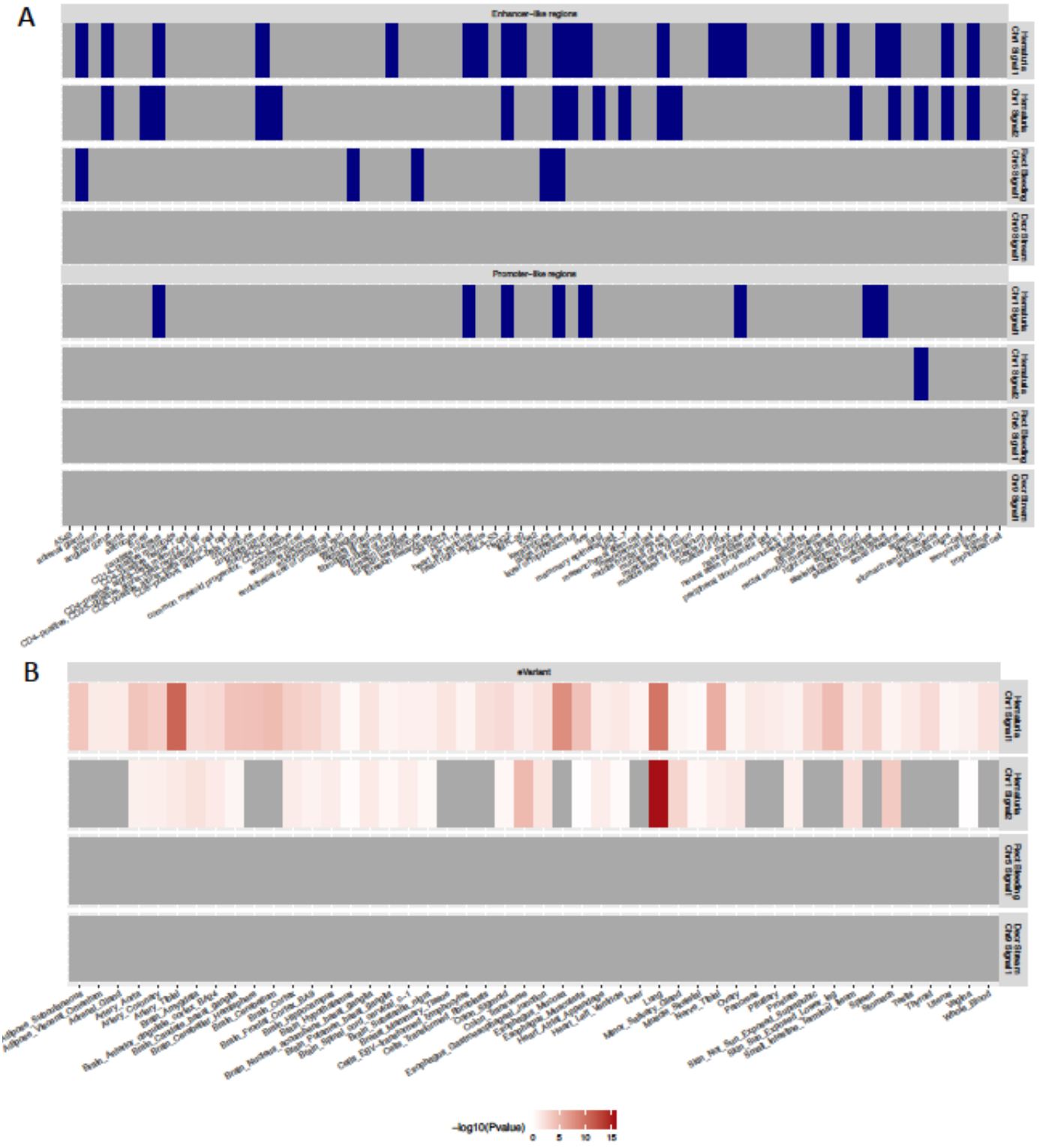

Figure 4. Mapping of credible causal variants (CCVs). In panel A, CCVs overlap with regulatory regions, enhancer- and promoter-like according to ENCODE. X axis: cell lines or tissues. Y axis: independent signals. Top graph shows enhancer-like regions. Bottom graph shows promoter-like regions. Blue: at least one CCV overlap a regulatory region active in the specific cell-line or tissue. Dark grey: any CCV overlap an active regulatory region. Panel B shows co-localization of CCVs with variants driving the expression of a particular transcript according to GTEX. X axis: tissues. Y axis: independent signals. Red, most significant expression p-value out of all CCVs at the signal and all evaluated transcripts for that tissue. Dark grey, no significant variants driving the expression of any transcript in the evaluated tissues.

In an attempt to explore the biological mechanisms underpinning radiotoxicity, we computed gene and pathway scores from the meta-analysis results. Nine pathways were associated ($p<0.05$) with more than one toxicity endpoint, suggesting a common biologic mechanism.

SNPs previously associated with radiotoxicity were evaluated for replication in the new cohorts (Table 3). Three SNPs showed a consistent association signal (rs17599026 with increased urinary frequency, rs7720298 with decreased urinary stream, and rs1801516 with overall toxicity). SNPs in the previously reported *TANC1* gene showed mixed results in the new cohorts.

Table 3. Association results for risk loci identified in prior genetic association studies. Bold values correspond to meta-analysis results.

		Meta-analysis of all studies				Meta-analysis of new studies		
Genetic variant	**Toxicity outcome and results from prior study**	**Study, N**	**Info[a]**	**OR (95% CI)**	**P_{meta}**	**Study, N**	**OR (95% CI)**	**P_{meta}**
rs17599026 chr5:137,763,798[b] *KDM3B* Minor allele T MAF[c] 0.07	Presence of grade 1+ increased urinary frequency at 2 years after radiotherapy OR 3.12 (2.08-4.69) P_{meta} $4.2x10^{-8}$ N = 1,564	**Meta-analysis** RAPPER-I, N=537 RAPPER-II, N=1,255 RADIOGEN, N=597 GenePARE-I, N=185 GenePARE-II, N=161 UGhent, N=281 CCI-BT CCI-EBRT, N=148	- 0.78 0.96 0.96 NA[d] 0.96 0.96 NA[e] NA[d]	**1.55 (1.23 to 1.95)** 1.89 (1.07 to 3.32) 1.27 (0.90 to 1.80) 2.14 (1.24 to 3.67) 3.08 (1.01 to 9.39) 1.10 (0.45 to 2.69) 1.08 (0.44 to 2.64) NA[e] 2.17 (0.71 to 6.62)	**$1.79x10^{-4}$**	**Meta-analysis** RAPPER-II, N=1,255 GenePARE-II, N=161 UGhent, N=281	**1.23 (0.91 to 1.67)** 1.27 (0.90 to 1.80) 1.10 (0.45 to 2.69) 1.08 (0.44 to 2.64)	**0.19**
rs7720298 chr5:13,858,328[b] *DNAH5* Minor allele G MAF[c] 0.30	Presence of grade 1+ decreased urine stream at 2 years after radiotherapy OR 2.71 (1.90-3.86) P_{meta} $3.2x10^{-8}$ N = 1,564	**Meta-analysis** RAPPER-I, N=537 RAPPER-II, N=1,255 RADIOGEN, N=597 GenePARE-I, N=185 GenePARE-II, N=161 UGhent CCI-BT CCI-EBRT	- 0.90 0.98 0.98 0.99 0.98 NA[f] NA[g] NA[h]	**1.43 (1.14 to 1.78)** 1.33 (0.79 to 2.23) 1.27 (0.88 to 1.83) 2.47 (0.63 to 9.75) 1.55 (1.01 to 2.37) 1.61 (0.92 to 2.82) NA[f] NA[g] NA[h]	**$1.62x10^{-3}$**	**Meta-analysis** RAPPER-II, N=1,255 GenePARE-II, N=161	**1.37 (1.01 to 1.86)** 1.27 (0.88 to 1.83) 1.61 (0.92 to 2.82)	**0.05**
rs1801516 chr11:108,175,462[b] *ATM* Minor allele A MAF[c] 0.22	Overall toxicity[i,j] OR 1.21 (0.98, 1.49) p-value not reported N = 2,697	**Meta-analysis** RAPPER-I, N=353 RAPPER-II, N=859 RADIOGEN, N=473 GenePARE-I, N=138 GenePARE-II, N=101 UGhent (N=228) CCI-BT CCI-EBRT (N=84)	- 0.93 NA[d] NA[d] 0.99 NA[d] NA[d] NA[l] 0.99	**1.29 (1.07, 1.55)** 1.61 (1.00, 2.58) 1.36 (1.03, 1.80) 1.32 (0.87, 2.02) 0.41 (0.19, 0.85)[k] 1.45 (0.63, 3.34) 1.41 (0.78, 2.52) NA[l] 1.15 (0.46, 2.86)	**$6.33x10^{-3}$**	**Meta-analysis** RAPPER-II, N=859 GenePARE-II, N=101	**1.37 (1.05, 1.78)** 1.36 (1.03, 1.80) 1.45 (0.63, 3.34)	**0.021**
rs7582141 chr2:159,899,489[b] *TANC1* Minor allele T MAF[c,m] 0.05	Overall toxicity[j] OR 6.17 (2.25, 16.9) P_{meta} $4.2x10^{-10}$ N = 1,742	**Meta-analysis** RAPPER-I, N=590 RAPPER-II, N=1,340 RADIOGEN, N=627 GenePARE-I, N=237 GenePARE-II, N=220 UGhent, N=285 CCI-BT CCI-EBRT, N=148	- 0.76 0.96 0.96 0.98 0.96 0.96 NA[l] 1.00	**1.99 (1.33, 2.98)** 1.91 (0.68, 5.39) 0.56 (0.20, 1.59) 3.98 (1.96, 8.07) 3.18 (1.04, 9.72)[k] 0.85 (0.20, 3.67) 2.16 (0.71, 6.53) NA[l] 0.73 (0.08, 6.38)	**$8.25x10^{-4}$**	**Meta-analysis** RAPPER-II, N=1,340 GenePARE-II, N=220 UGhent, N=285 CCI-EBRT, N=148	**0.98 (0.52, 1.86)** 0.56 (0.20, 1.59) 0.85 (0.20, 3.67)[k] 2.16 (0.71, 6.53) 0.73 (0.08, 6.38)	**0.95**

Table 4 shows multivariable Cox regression models, including the SNPs identified in this study and previously published studies as well as clinical covariates. In each model, SNPs remained independently associated with toxicity after inclusion of clinical covariates.

Table 4. Multivariable models including SNPs and clinical risk factors. All models are stratified by study.

	HR (95% CI)	p-value
Rectal Bleeding		
rs17055178	1.79 (1.46, 2.19)	<0.001
Rectum volume receiving 65Gy[a]	1.43 (1.18, 1.73)	<0.001
Rectum volume receiving 70Gy[b]	4.27 (2.09, 7.51)	<0.001
Sigmoid D_{max} > 31Gy[c]	0.17 (0.02, 0.88)	0.04
Arthritis	1.99 (1.09, 3.36)	0.03
Irritable bowel disease	1.76 (1.05, 2.76)	0.03
Cardiovascular disease	1.45 (1.01, 2.03)	0.04
Urinary Frequency		
rs17599026	1.31 (1.03, 1.63)	0.03
Age at treatment > 75[d]	1.46 (1.13, 1.87)	0.004
Diabetes	1.51 (1.14, 1.97)	0.005
Cardiovascular disease	1.58 (1.04, 2.33)	0.03
Decreased Stream		
rs10969913	2.53 (1.60, 3.82)	<0.001
rs7720298	1.21 (1.02, 1.44)	0.03
Prior TURP	1.62 (1.10, 2.31)	0.02
Prior prostatectomy	0.17 (0.01, 0.78)	0.02
Presence of hemorrhoids	2.04 (1.28, 3.10)	0.004
Hematuria		
rs11122573	1.76 (1.38, 2.21)	<0.001
rs75991123	1.64 (1.25, 2.12)	<0.001
Prior TURP	2.30 (1.68, 3.09)	<0.001
Bladder 74Gy[e]	1.29 (1.09, 1.51)	0.003
Receipt of EBRT[f]	1.92 (1.17, 3.20)	0.01
Age at treatment[g]	2.73 (1.19, 5.76)	0.02

Abbreviations: HR, hazard ratio; TURP, transurethral resection of the prostate

[a] Variable was log transformed and includes a spline at 3.0, the 25th percentile value.
[b] Variable was log transformed and includes a spline at 3.4, the 75th percentile value.
[c] 31Gy is the 25th percentile value.
[d] Reference group are men ≤ 75 at time of treatment.
[e] Variable was log transformed and includes a spline at the median value (0.9Gy).
[f] Reference group is receipt of brachytherapy alone.
[g] Age is treated as a continuous variable if above 75 years.

The work we accomplished during the third year of this project identified three new genomic regions associated with late radiotoxicity. In addition, by carrying out the first radiogenomics fine mapping study, we showed that one region had two independent signals (1q42.2) associated with hematuria. The CCVs driving radiotoxicity were associated with differential expression of local protein coding genes and non-coding RNA genes, which provides some pointers towards possible functional mechanisms. In particular, the signals discovered affect gene regulation, rather than gene coding sequence. An interesting candidate is the *AGT* gene, encoding angiotensinogen, which is converted to the active enzyme angiotensin II through activity of angiotensin converting enzyme (ACE). Prior studies suggest that *AGT* signaling may influence radiation-induced blood vessel wall injury and interstitial fibrosis, and animal and human studies suggest that ACE inhibitors may be radio-protective. However, it is not possible to state with certainty that *AGT* is the target, as the second independent signal in this region did not appear to target the same gene. Where multiple, physically adjacent signals are associated with the same phenotype, as is the case for the two new hematuria risk signals,

Occam's razor predicts that they should act on the same target gene, although their mechanisms of action may differ.

Our analysis represents the largest genome-wide study to date of late radiotoxicity in prostate cancer survivors. As is commonly seen, the initial effect size estimates for the GWAS-identified SNPs were upwardly biased (the so-called "winner's curse") and validation in the present study enabled estimation of effect sizes that more likely reflect their true contribution to risk for developing toxicity. Although unable to replicate the association at 2q24.1 within *TANC1*, it is challenging due to the rarity of minor alleles within European ancestry populations. Ongoing laboratory studies do support a role of *TANC1* in radiation response (personal communication from A.V.), highlighting the importance of functional studies as complementary to association studies.

The multivariable risk models we developed show that genetic variants, treatment variables, and other clinical factors can contribute independent information on risk for developing radiotoxicity. This finding strongly suggests that common variants can improve traditional normal tissue complication probability models, as others have suggested. It is important to continue efforts to identify additional risk SNPs and rare variants, however, these models may be sufficient to move forward into validation studies and potentially prospective studies of clinical utility. Investigators in the field of radiotoxicity prediction are actively developing novel clinical trial approaches for testing the ability of risk models to personalize treatment and improve outcomes.

A strength of our study is the prospective longitudinal assessment of toxicity enabling use of time-to-event analysis in order to maximize information across multiple toxicity assessments. Prior studies by us and ours largely focused on toxicity at a single time-point following radiotherapy; typically two years. Long-term follow-up is clearly important for radiogenomic studies, and future work should aim to use longitudinal analysis. Most GWAS to date focused on populations of European ancestry, as ethnicity can inflate type I error rates and reduce statistical power due to population heterogeneity in allelic effects on a trait. A second strength of our work is the inclusion of multiple independent radiotherapy cohorts from different ancestral backgrounds. It is important to understand how knowledge gained from European ancestry GWAS transfers to other ethnicities, and methods are being developed to detect genetic variants associated with complex traits allowing for population heterogeneity. Trans-ethnic studies suggest susceptibility loci for traits are generally shared between European and East Asians and, because of the larger sample size, cross-population meta-analyses increase statistical power to detect novel loci. Our cohort sizes are still too small to identify heterogeneity in allelic effects between ethnic groups, but we performed the first analysis attempting to explore transferability of SNP-radiotherapy toxicity associations across ethnicities and found evidence of replication of multiple risk loci in one of two Japanese cohorts.

In summary, by performing the largest GWAS meta-analysis and first fine-mapping study in radiogenomics we identified four new regions associated with radiotoxicity following radiotherapy for prostate cancer. We showed that the signals discovered affect gene regulation rather than gene coding sequences, and provide evidence for replication across ethnicities. This study increases understanding of the architecture of common genetic variants affecting radiotoxicity, and shows further multi-national radiogenomics studies in larger cohorts are worthwhile.

Polygenic Scores Results

A substantial effort was devoted during the past year to the development of models using polygenic risk score and machine-learning methods, but unfortunately it was not possible to achieve predictive performance, assessed by the area under the receiver operating characteristic curve, that was statistically better than chance. Even though the number of samples analyzed was substantially increased, this negative result is likely a consequence of modest effect sizes associated with the SNPS identified and validated combined with the lack of the very large sample sizes generally required for the successful use of these approaches. Hence, our study was underpowered to test multi-SNP modeling methods such as polygenic risk scores and machine learning-based methods although these approaches have been used successfully for other polygenic traits and diseases. This approach should therefore be re-evaluated in the context of radiotoxicity as larger cohorts become available.

Develop a low-cost, high-performance genetic assay.

We developed and tested TaqMan quantitative polymerase chain reaction (qPCR) assays for SNPs that were shown to be significant in the GWAS meta analyses (rs1801516, rs264663, rs72735025, rs17599026, rs78394554, rs10969913, rs11122573, rs139882217. For those SNPs where would could not design an assay, we selected the next significant SNP in linkage disequilibrium to develop a test. All of the SNPs converted into TaqMan qPCR assays with 99-100% replication of genotypes (Table 5). The results were well replicated at 30 days, but not 60 days. The reagents were stable when let at room temperature for 24 hours, but not at 30 degrees Celsius. Thus, we have developed a set of assays that could be used for clinical application. We tested these assays on DNA from the 200 patients enrolled through the REQUITE study and observed concordance in the Ostrer research laboratory at Einstein and at the L2 Diagnostics Laboratory, New Haven, CT.

Table 5. Replication of genotypes for TaqMan qPCR assays for SNPs demonstrated to have significant associations

SNP	Initial	30 days	60 days	20 degrees	30 degrees
rs1801516	.99	.99	.49	.98	.71
rs264663	1	.99	.86	.98	.92
rs72735025	.99	.98	.38	.96	.55
rs17599026	1	.98	.50	.90	.85
rs78394554	1	.99	.79	.99	.95
rs10969913	1	.99	.80	.99	.89
rs11122573	1	.93	.63	.94	.80
rs139882217.	1	.98	-	.98	.95

Export the models to a web-based application that could be used by physicians in practice and/or genetic testing laboratories

The models developed under this project are a critical first step towards reducing the incidence and severity of late radiotoxicity in prostate cancer survivors but are not quite ready for clinical use. Rather than moving ahead with a web-based application, we have instead disseminated the models to the research community via peer-reviewed publication and presentation at the ASTRO and RGC annual meetings. The next step is to validate these models in an independent clinical study, which is underway through an NIH-funded SBIR Phase II project shared with L2 Diagnostics, LLC. We are also working with the FDA towards clearance of our assay and accompanying risk models, at which point that will be made available to the clinical community via a web-based or similar application.

What do you plan to do during the next reporting period to accomplish the goals?

N/A

What opportunities for training and professional development has the project provided?

Nothing to Report

How were the results disseminated to communities of interest?

Results of these findings were presented at the annual meeting of the Radiogenomics Consortium (Rochester, NY in June, 2019) and published as open-access in the Journal of the National Cancer Institute (JNCI, 2019 May 16; https://doi.org/10.1093/jnci/djz075).

4. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?
Nothing to Report

What was the impact on other disciplines?
Nothing to Report

What was the impact on technology transfer?
Nothing to Report

What was the impact on society beyond science and technology?
Nothing to Report

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change
Nothing to Report

Actual or anticipated problems or delays and actions or plans to resolve them
Nothing to Report

Changes that had a significant impact on expenditures
Nothing to Report

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Significant changes in use or care of human subjects
Nothing to Report
Significant changes in use or care of vertebrate animals.
Nothing to Report
Significant changes in use of biohazards and/or select agents
Nothing to Report

6. PRODUCTS:

Publications

Kerns SL, Dorling L, Fachal L, Bentzen S, Pharoah PD, Barnes DR, Gómez-Caamaño A, Carballo AM, Dearnaley DP, Peleteiro P, Gulliford SL, Hall E, Michailidou K, Carracedo Á, Sia M, Stock R, Stone NN, Sydes MR, Tyrer JP, Ahmed S, Parliament M, Ostrer H, Rosenstein BS, Vega A, Burnet NG, Dunning AM, Barnett GC, West CM; Radiogenomics Consortium. Meta-analysis of Genome Wide Association Studies Identifies Genetic Markers of Late Toxicity Following Radiotherapy for Prostate Cancer. EBioMedicine. 2016 Aug;10:150-63. doi: 10.1016/j.ebiom.2016.07.022. Epub 2016 Jul 20. PubMed PMID: 27515689; PubMed Central PMCID: PMC5036513.

Kerns SL, Chuang KH, Hall W, Werner Z, Chen Y, Ostrer H, West C, Rosenstein B. Radiation biology and oncology in the genomic era. Br J Radiol. 2018 Nov;91(1091):20170949. doi: 10.1259/bjr.20170949. Epub 2018 Jun 14. Review.

PubMed PMID: 29888979; PubMed Central PMCID: PMC6475928.

Kerns SL, Fachal L, Dorling L, Barnett GC, Baran A, Peterson DR, Hollenberg M, Hao K, Narzo AD, Ahsen ME, Pandey G, Bentzen SM, Janelsins M, Elliott RM, Pharoah PDP, Burnet NG, Dearnaley DP, Gulliford SL, Hall E, Sydes MR, Aguado-Barrera ME, Gómez-Caamaño A, Carballo AM, Peleteiro P, Lobato-Busto R, Stock R, Stone NN, Ostrer H, Usmani N, Singhal S, Tsuji H, Imai T, Saito S, Eeles R, DeRuyck K, Parliament M, Dunning AM, Vega A, Rosenstein BS, West CML. Radiogenomics Consortium Genome-Wide Association Study Meta-analysis of Late Toxicity after Prostate Cancer Radiotherapy. J Natl Cancer Inst. 2019 May 16. pii: djz075. doi: 10.1093/jnci/djz075. [Epub ahead of print] PubMed PMID: 31095341.

Abstract

ASTRO annual meeting in San Antonio, TX; October 21-24, 2018

Sarah L. Kerns, Laura Fachal, Leila Dorling, Gillian C. Barnett, Neil Burnet, Matthew R. Sydes, Emma Hall, David Dearnaley, Alison M. Dunning, Paul D.P. Pharoah, Matthew Parliament, Nawaid Usmani, Kim de Ruyck, Harry Ostrer, Barry S. Rosenstein, Antonio Gómez-Caamaño, Ana Carballo, Paula Peleteiro, Ana Vega, Catharine M.L. West on behalf of the Radiogenomics Consortium

Meta-analysis of genome-wide association studies (GWAS) of late toxicity in 3,874 men treated with radiation for prostate cancer

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name: Harry Ostrer
Project Role: co-PI
Researcher Identifier: 0000-0002-2209-5376
Nearest person month worked: 1
Contribution to Project: Dr. Ostrer oversaw the design and management of this study and worked to develop assays that could be used for risk assessment.
Funding Support: This award

Name: Kinnari Upadhyay
Project Role: Bioinformatician
Researcher Identifier : N/A
Nearest person month worked: 6
Contribution to Project: Ms. Upadhyay developed a database and risk assessment tools for incorporation of genetic data for this project under the supervision of Dr. Ostrer.
Funding Support: This award

Name: Johnny Loke
Project Role: Research associate
Researcher Identifier: N/A
Nearest person month worked: 2
Contribution to Project: Mr. Loke developed qPCR, dPCR, AmpliSeq and hybrid capture sequencing assays for analysis of genetic variants identified in this project under the supervision of Dr. Ostrer.
Funding Support: This award

Name: Ke Hao
Project Role: Co-Investigator
Researcher Identifier : NA
Nearest person month worked: 2
Contribution to Project: Design and implement algorithms in constructing and evaluating polygenic score (PGS) on radiation toxicity traits.
Funding Support: This award

Name: Antonio Di Narzo, PhD
Project Role: Data analyst
Researcher Identifier : NA
Nearest person month worked: 2 months
Contribution to Project: polygenic score data analysis
Funding Support: NA

Name: Gaurav Pandey
Project Role: Co-Investigator
Researcher Identifier : NA
Nearest person month worked: 2
Contribution to Project: Design of machine learning strategies to identify genetic predictors of radiotoxicity
Funding Support: This award

Name: Mehmet Eren Ahsen
Project Role: Data Analyst
Researcher Identifier : NA
Nearest person month worked: 1
Contribution to Project: Implementation of machine learning strategies to identify genetic predictors of radiotoxicity
Funding Support: This award

Name: Barry Rosenstein
Project Role: Principal Investigator
Researcher Identifier : NA
Nearest person month worked: 1
Contribution to Project: Worked with Dr. Kerns to obtain and harmonize dosimetric, clinical and OncoArray genotyping data for all subjects from each cohort comprising this project and to perform statistical analysis for validation of previously discovered SNPs and identification of new SNPs. Worked with Drs. Pandey and Hao to use novel strategies for radiogenomics, sparse learning, polygenic score and ensemble learning, to create polygenic risk models to predict the incidence of radiotherapy toxicity based on the genotype and clinical characteristics.
Funding Support: This award

Name: Sarah Kerns
Project Role: Co-investigator
Researcher Identifier : NA
Nearest person month worked: 5
Contribution to Project: Dr. Kerns performed data management and statistical analyses for the GWAS meta-analysis to identify SNPs associated with radiation toxicity in collaboration with Drs. Rosenstein and Ostrer.
Funding Support: NCI K07 CA187546

Name: Andrea Baran

Project Role: Biostatistician
Researcher Identifier (e.g.ORCID ID): NA
Nearest person month worked: 1
Contribution to Project: Ms. Baran assisted with performing quality checks and data cleaning for the oncoarray SNP datasets analyzed in this project under the supervision of Dr. Kerns.
Funding Support: NCI K07 CA187546 and SBIR HHSN261201500043C

Name: Ashley Amidon Morlang
Project Role: Study Coordinator
Researcher Identifier (e.g.ORCID ID): NA
Nearest person month worked: 1
Contribution to Project: Ms. Morlang assisted with data management related to the clinical and dosimetric data for each cohort included in the GWAS analysis under the supervision of Dr. Kerns. She also coordinated the IRB exemption request/approval required for this project.
Funding Support: This award

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?
Nothing to Report

What other organizations were involved as partners?
Nothing to Report