Radiogenomics: Identification of Genomic Predictors for Radiation Toxicity

Barry S. Rosenstein, PhD*†

The overall goal of radiogenomics is the identification of genomic markers that are predictive for the development of adverse effects resulting from cancer treatment with radiation. The principal rationale for a focus on toxicity in radiogenomics is that for many patients treated with radiation, especially individuals diagnosed with early-stage cancers, the survival rates are high, and therefore a substantial number of people will live for a significant period of time beyond treatment. However, many of these patients could suffer from debilitating complications resulting from radiotherapy. Work in radiogenomics has greatly benefited from creation of the Radiogenomics Consortium (RGC) that includes investigators at multiple institutions located in a variety of countries. The common goal of the RGC membership is to share biospecimens and data so as to achieve large-scale studies with increased statistical power to enable identification of relevant genomic markers. A major aim of research in radiogenomics is the development of a predictive instrument to enable identification of people who are at greatest risk for adverse effects resulting from cancer treatment using radiation. It is anticipated that creation of a predictive assay characterized by a high level of sensitivity and specificity will improve precision radiotherapy and assist patients and their physicians to select the optimal treatment for each individual.

Semin Radiat Oncol 27:300-309 © 2017 Elsevier Inc. All rights reserved.

Introduction

The goals of research in radiogenomics fall into two general areas. The first main objective being pursued by investigators in this field is identification of genomic markers, primarily single nucleotide polymorphisms (SNPs) that could serve as the basis of an assay to predict the relative susceptibility for patients with newly diagnosed cancer to develop adverse effects if they were to be treated with radiation. SNPs represent a major source of genetic variation between individuals as approximately once every 1000 nucleotides, more than 5% of people have an alternate base pair at a particular nucleotide, although the frequency of specific SNPs depends on ethnic, racial, and geographic location. In addition, as the costs for whole exome and whole genome sequencing continue to decrease, it is likely that information will increasingly become available for many subjects in radiogenomic studies as to the presence of rare variants, which may be associated with various outcomes resulting from radiotherapy. It has come to be recognized that patient-related characteristics, including genomic factors, could represent an important basis influencing susceptibility for development of radiation-related toxicities. It should be noted that adverse effects resulting from radiotherapy are relatively common, with approximately 2%-5% of patients developing some form of grade 3 complication and 10%-20% experiencing moderate grade 2 toxicity. Although great strides have been made to localize the dose of radiation to the cancer, normal tissues and organs still often are subjected to a substantial dose of radiation as part of treatment, which can result in significant complications. While radiotherapy is often curative, the adverse effects resulting from treatment can place a major financial burden on both individuals as well as the health care system.

*Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY.
†Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Conflicts of interest: none.

This research was supported by Grants and contracts from the United States National Institutes of Health (1R01CA134444 and HHSN261201500043C), the American Cancer Society (RSGT-05-200-01-CCE), and the United States Department of Defense (PC074201 and PC140371).

*Address reprint requests to Barry S. Rosenstein, PhD, Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029. E-mail: barry.rosenstein@mssm.edu
Even though the emphasis of research in radiogenomics has been on the identification of SNPs associated with outcomes, it is likely that epigenetic and other “panomic” factors are also of importance and likely to be eventually incorporated into any predictive instrument that is developed as it evolves and improves in sensitivity and specificity. Nevertheless, the development of a SNP-based test would enhance precision radiotherapy as it will enable selection of patients who might benefit from a strictly surgical or drug treatment or use of a more conformal form of radiotherapy that spares normal tissues. Use of such a genetic or genomic predictive assay could enhance the therapeutic index through a decrease in the rate of complications. In addition, it may be feasible to dose escalate and possibly improve the cure rate for patients predicted to be at lower risk for radiation-induced injuries.

The second main aim of radiogenomics, which represents a more far-reaching goal, is the use of information gained through radiogenomic research to assist with the development of agents that could prevent or mitigate normal tissue or organ toxicities that may result from treatment with high doses of radiation. As genes are identified whose encoded products are affected by SNPs that reside either within or near these genes, it will then be possible to conduct mechanistic and functional studies to enhance an understanding as to the potential role that these gene products play in the development of adverse outcomes resulting from exposure to radiation. Thus, it is anticipated that a greater understanding of the molecular pathways that play a role in the development of radiation injuries could lead to the development of pharmacologic agents with a capability to either prevent or mitigate these toxicities. However, progress toward this overall goal is dependent upon the validation of SNPs in multiple cohorts that have been discovered as associated with normal tissue toxicities resulting from cancer radiotherapy.

Factors That Facilitate Research and Challenges in Radiogenomics

The following are among the positive factors that facilitate a radiogenomics approach:

(1) The outcome of interest occurs in response to a specific exposure, radiation—this is in contrast to many candidate gene and genome-wide association studies (GWAS) that are performed to identify genetic variants associated with an increased probability to develop a certain disease for which there may be numerous environmental and lifestyle factors that could influence the probability of an individual developing a particular disease in addition to genetic influences.

(2) Adverse effects resulting from cancer radiotherapy are relatively common—this factor enables studies of a more modest size to be performed compared with research focused on relatively rare phenotypes.

(3) Findings of radiogenomic studies are actionable—for example, men diagnosed with prostate cancer often are provided with limited information in terms of basing a treatment decision as to whether radiotherapy, surgery, or active surveillance represents the best course of action. However, a man receiving the results of a predictive assay suggestive of radiosensitivity may opt for surgery or active surveillance if his prostate-specific antigen (PSA) and Gleason scores are consistent with this recommendation. Conversely, men predicted to be at low risk for the development of toxicities following prostate cancer radiotherapy may consider a more aggressive treatment. Similarly, a treatment decision for women diagnosed with early-stage breast cancer may not be clear as to a choice between mastectomy or lumpectomy followed by radiotherapy. For women predicted to be at high risk for fibrotic responses in the breast and other adverse effects, mastectomy rather than radiotherapy may be advisable, especially as adverse radiation effects could compromise the ability for breast reconstruction. However, women at low risk may feel greater confidence to proceed with limited surgery followed by radiotherapy. Young people diagnosed with Hodgkin’s disease, and their families, are also often faced with a difficult decision as to whether radiation should be used to treat their cancer in addition to chemotherapy. The cure rates for these forms of cancer, particularly when diagnosed at an early stage, are relatively high. Thus, many of these individuals will live for a substantial length of time following treatment. However, for those patients receiving radiotherapy, their risk for development of a serious complication or secondary malignancy caused by radiation exposure could be increased. Thus, development of a robust predictive assay would allow patients diagnosed with these and other forms of cancer to reach a more informed treatment decision.

(4) A predictive assay can guide radiotherapy treatment planning—for patients predicted to be at a substantially greater risk for the development of adverse effects resulting from exposure to radiation, more stringent steps could be taken to limit the exposure of normal tissues and organs to radiation. One approach could be the use of protons or carbon ions as patients at high risk might specifically benefit from the more conformal treatment that these alternate, but generally more costly forms of radiotherapy, may enable.

(5) Biological insight can lead to the development of preventative or mitigating agents—although development of a predictive instrument to help guide treatment is the initial priority of research in radiogenomics, once genetic variants associated with radiation toxicities have been identified and validated in multiple cohorts, it would be beneficial to embark upon mechanistic and functional studies to help elucidate the molecular basis as to the pathways through which an alteration in the products of these genes results in an increased susceptibility to dysfunctionality to particular organs or
tissues. Such information could lead to the development of agents to help prevent or mitigate adverse outcomes following radiotherapy.

**Challenges in Radiogenomics**

The following are among the challenges of studies focused on radiogenomics:

(1) **Dosimetry matters**—it is important to obtain detailed treatment and dosimetric data for multivariable modeling. Unfortunately, this is not routinely accomplished for many studies and is a particular problem when attempting to combine data from multiple studies. This is a critical aspect of the REQUITE project\(^{16}\), in which a series of dosimetric parameters, including the full dose volume histograms (DVHs) and Digital Imaging and Communications in Medicine (DICOM) images have been obtained for the roughly 5000 patients treated with radiotherapy who have been enrolled into this study.

(2) **Need baseline (preradiotherapy) symptom assessment**—to determine the effect of radiation for symptoms that appear following treatment, it is essential to obtain baseline information for patients before the initiation of therapy. For example, men diagnosed with prostate cancer often already suffer from different forms of urinary or sexual function problems before being exposed to high doses of radiation as part of their treatment. It is, therefore, essential to know the change experienced by each subject from a baseline score.

(3) **Long-term follow-up (≥2 years)**—the development of symptoms varies with time and grade of effect. As opposed to studies whose aim is the identification of genetic variants associated with susceptibility for a particular disease, in which the outcome is essentially dichotomous, the development of adverse effects resulting from radiotherapy varies with both time and grade of complication. The question often arises as to whether a patient should be evaluated at a specific time point or grade level of toxicity or both, or if the data should be analyzed as a continuous variable.

(4) **Variability in tools used to measure adverse effects**—when combining multiple cohorts, as is often done in radiogenomic studies, different measures and scales to quantify toxicities are commonly employed across the centers contributing biospecimens or data or both to a particular project. It can, therefore, be difficult to harmonize outcomes. Although this heterogeneity has likely limited the discovery of many genetic variants, clearly it has not completely prevented their identification as there is often adequate similarity and overlap between the evaluation instruments used by different groups of investigators to enable successful association studies.

(5) **"Outcomes" are multiple and incompletely understood**—multiple measures of toxicity for a particular outcome are routinely employed. For example, urinary toxicity in men treated for prostate cancer can be measured using hematuria, nocturia, straining, urgency, and other end points. In addition, many factors influence toxicity including dose, volume irradiated, time, comorbid conditions, and interaction with other modalities. Therefore, multivariable analyses of radiogenic studies are essential as multiple factors will influence the development of different forms of toxicity. An important effort to combine different forms of toxicity into one score has been attempted in what is termed a standardized total average toxicity (STAT) score.\(^{17}\)

(6) **Requirement for large sample sizes and multiple cohorts**—previous GWAS that have been performed demonstrate the need for large sample sizes.\(^{18-20}\) This is of importance, as the relative risk associated with any particular SNP is generally relatively small. In addition, to identify SNPs with a low minor allele frequency, large sample sizes are needed. Beyond that, multiple cohorts are essential to validate SNPs that were discovered through an initial GWAS. An important factor that has helped to address the need for multiple cohorts of large sample size was creation of the Radiogenomics Consortium (RGC) that was established in 2009\(^{21}\) and is a National Cancer Institute/NIH-supported Cancer Epidemiology Consortium (https://epi.grants.cancer.gov/radiogenomics/). The RGC currently consists of 217 investigators at 123 member institutions in 30 countries. The shared goals of the RGC members are to bring together collaborators for potential projects to share data and biospecimens so as to increase the statistical power of radiogenomic studies. The RGC has also facilitated the performance of cross-center validation studies, which are indispensable for the development of a predictive assay to gain widespread clinical acceptance.\(^{22}\)

Nevertheless, despite these many challenges, an increasing number of well-performed radiogenomic candidate gene studies (CGS) and GWAS have been successfully performed, which are now producing an increasing number of genetic markers that have been discovered and validated in multiple cohorts as outlined below.

**SNPs That Have Been Identified and Validated**

The initial research performed in radiogenomics involved CGS that focused on genes encoding proteins with known associations to pathways involved in responses to radiation, such as DNA repair processes and cell cycle checkpoint control. Although a number of positive associations were reported, these studies often did not adequately correct for multiple-hypothesis testing and generally were not validated in subsequent studies,\(^ {23}\) with several exceptions. The main advance in radiogenomics research has been achieved through the use of SNP microarrays and the performance of GWAS in which large numbers of SNPs across the genome have been evaluated.
Using this approach, several large GWAS and CGS have been accomplished involving a rigorous analysis for association between particular SNPs and specific outcomes. The results of these studies are summarized in the Table.

**Prostate Cancer**

A series of studies examining common SNPs in candidate genes were initially performed but little evidence was obtained to validate any of the SNPs examined. However, once the cost substantially diminished for genotyping using DNA microarrays, the focus of research in radiogenomics shifted toward the performance of GWAS. It should be noted that owing to the necessity to employ a correction for multiple-hypothesis testing, genome-wide significance for a GWAS is generally thought to be met only for those SNP associations with a $P < 5 \times 10^{-8}$.

The first GWAS performed in radiogenomics was to identify SNPs associated with erectile dysfunction in African American men treated with radiotherapy for prostate cancer. Through this study, a SNP (rs2268363) in the FSHR gene, which encodes follicle stimulating hormone, was identified ($P = 5.46 \times 10^{-8}$). This hormone is expressed in sertoli cells located in the testis and is involved in the development and function of this organ. Disturbance of the FSHR signaling pathway can result in small testis size, abnormal spermatogenesis, and infertility.

A 3-stage GWAS was conducted using discovery and replication cohorts that included the use of STAT score as the measure of combined adverse effects following prostate cancer radiotherapy. A locus encompassing the TANC1 gene was associated with STAT score for overall late toxicity with an odds ratio (OR) of approximately 6 (combined $P = 6.5 \times 10^{-7}$). This haplotype block is located within the IFNK gene whose product is a type 1 interferon that regulates cytokine release, which could influence the development of urinary complications following treatment of prostate cancer with radiation as these factors could play a role in the inflammatory response resulting from damage to tissues exposed to high doses of radiation. Also, SNP rs13035033 that is located in MYO3B, which encodes actin-based motor protein myosin IIIB and is highly expressed in the kidney, was associated with urinary straining ($P = 5.0 \times 10^{-9}$).

A replicated study involving men who received prostate cancer radiotherapy identified a SNP (rs2788612) located in the KCND3 gene that was strongly associated with late rectal incontinence ($RR = 9.91, P = 1.05 \times 10^{-12}$). KCND3 encodes a member of the potassium channel, voltage-gated, shal-related subfamily, which is expressed in smooth muscle and thus may play a role in sphincter function.

A fixed-effect meta-analysis was performed using data from four cohorts consisting of 1564 men treated for prostate cancer for which a GWAS was performed in which toxicity was measured at a 2-year time point. Two SNPs were identified in this study that met genome-wide significance. One was rs17599026, which resides on chromosome 5q31.2 and associated with urinary frequency and characterized by an OR of 3.1 (95% CI: 2.1-4.7, $P = 4.2 \times 10^{-4}$). This SNP is located in an intronic region of KDM3B, 23 bp downstream of exon 20. This gene is highly expressed in bladder tissue, which is consistent with a potential role for the encoded protein in normal bladder function. Thus, its alteration may increase the likelihood for a urinary complication upon exposure to a high dose of radiation.

rs7720298, which resides on chromosome 5p15.2, was associated with decreased urine stream with an OR of 2.7 (95% CI: 1.9-3.9, $P = 3.2 \times 10^{-5}$). This SNP is located in an intronic region downstream of DNAH5 exon 30. This gene encodes the dynem, axonemal, and heavy chain 5 protein that is part of a protein complex that is associated with microtubule formation. Mutations in DNAH5 can result in primary ciliary dyskinesia resulting from abnormal cellular cilia and flagella. This gene is expressed in both the kidney and bladder, consistent with a possible role in the function of these organs. Therefore, a variant in DNAH5 may enhance the probability for an adverse urinary effect following radiotherapy. In addition to identification of a strong association between these two SNPs with urinary morbidity following radiotherapy, an important aspect of this study is the demonstration that meta-analysis of a multicohort consisting of subjects who were evaluated using variable toxicity instruments is able nevertheless to yield results identifying genetic markers of importance for outcomes following cancer treatment with radiation.

**Breast Cancer**

An increasing focus for radiogenomics investigators is the identification of SNPs associated with the development of adverse normal tissue outcomes resulting from radiotherapy of...
<table>
<thead>
<tr>
<th>SNP Number</th>
<th>Gene</th>
<th>Gene Product</th>
<th>Function</th>
<th>Adverse Effect</th>
<th>GWAS or CGS</th>
<th>P Value</th>
<th>Odds Ratio (95% CI)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2268363</td>
<td><em>FSHR</em></td>
<td>Follicle stimulating hormone</td>
<td>Expressed in sertoli cells located in the testis; involved in the development and function of the testis</td>
<td>Erectile dysfunction</td>
<td>GWAS</td>
<td>$5.4 \times 10^{-8}$</td>
<td>7.0 (3.4-14.7)</td>
<td>24</td>
</tr>
<tr>
<td>rs264663</td>
<td>TANC1</td>
<td>Tetrameric peptide repeat, ankyrin repeat, and coiled-coil domain-containing protein 1</td>
<td>Recruitment of fusion-competent myoblasts during myotube formation</td>
<td>Overall late GU and GI toxicity</td>
<td>GWAS</td>
<td>$4.6 \times 10^{-11}$</td>
<td>6.6 (2.2-19.6)</td>
<td>27</td>
</tr>
<tr>
<td>rs7120482</td>
<td>Upstream of SLC36A4</td>
<td>Solute carrier family 36 member</td>
<td>High-affinity/low-capacity nonproton-coupled amino acid transporter; modulates action of the mTOR complex 1 signaling pathway</td>
<td>Rectal bleeding</td>
<td>GWAS</td>
<td>$5.4 \times 10^{-8}$</td>
<td>3.1 (1.7-5.6)</td>
<td>30</td>
</tr>
<tr>
<td>rs71630638</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$6.9 \times 10^{-7}$</td>
<td>2.9 (1.6-5.2)</td>
<td></td>
</tr>
<tr>
<td>rs11648233</td>
<td><em>HSD17B2</em></td>
<td>Hydroxysteroid 17β-dehydrogenase 2</td>
<td>Oxidative metabolism of androgens and estrogens</td>
<td>Erectile dysfunction</td>
<td>GWAS</td>
<td>$9.1 \times 10^{-5}$</td>
<td>1.8 (1.2-2.8)</td>
<td>29</td>
</tr>
<tr>
<td>rs17779457</td>
<td>IFNκ</td>
<td>Type I interferon</td>
<td>Regulates cytokine release</td>
<td>AUA symptom score</td>
<td>GWAS</td>
<td>$6.5 \times 10^{-7}$</td>
<td>2.4 (1.1-3.6)</td>
<td>(BC)</td>
</tr>
<tr>
<td>rs13035033</td>
<td>MYO3B</td>
<td>Myosin IIIb</td>
<td>Actin-based motor protein; highly expressed in the kidney</td>
<td>Urinary straining</td>
<td>GWAS</td>
<td>$5.0 \times 10^{-9}$</td>
<td>0.9 (0.6-1.2)</td>
<td>(BC)</td>
</tr>
<tr>
<td>rs2788612</td>
<td>KCNQ3</td>
<td>Potassium voltage-gated channel, shal-related subfamily member 3</td>
<td>Regulates epithelial electrolyte transport, smooth muscle contraction, and cell volume</td>
<td>Late rectal incontinence</td>
<td>GWAS</td>
<td>$1.1 \times 10^{-12}$</td>
<td>9.9 (RR)</td>
<td>39</td>
</tr>
<tr>
<td>rs17599026</td>
<td>KDM3B</td>
<td>Lysine demethylase 3B</td>
<td>Histone demethylase; specifically demethylates Lys-9 of histone H3</td>
<td>Urinary frequency</td>
<td>GWAS</td>
<td>$4.2 \times 10^{-8}$</td>
<td>3.1 (2.1-4.7)</td>
<td>41</td>
</tr>
<tr>
<td>rs7720298</td>
<td>DNAH5</td>
<td>Dynein, axonemal, and heavy chain 5</td>
<td>Part of a protein complex associated with microtubule formation</td>
<td>Decreased urine stream</td>
<td></td>
<td>$3.2 \times 10^{-8}$</td>
<td>2.7 (1.9-3.9)</td>
<td></td>
</tr>
</tbody>
</table>

**Prostate**

**Breast**

<table>
<thead>
<tr>
<th>SNP Number</th>
<th>Gene</th>
<th>Gene Product</th>
<th>Function</th>
<th>Adverse Effect</th>
<th>GWAS or CGS</th>
<th>P Value</th>
<th>Odds Ratio (95% CI)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1800629</td>
<td>TNFα</td>
<td>Tumor necrosis factor alpha</td>
<td>Multifunctional proinflammatory cytokine</td>
<td>Telangiectasia</td>
<td>CGS</td>
<td>0.0028</td>
<td>NP</td>
<td>44</td>
</tr>
<tr>
<td>rs2857595</td>
<td></td>
<td>Intergenic region between NCR3 and AIF1</td>
<td>NCR3-longevity protein in tumor cells</td>
<td>Overall late toxicity</td>
<td>CGS</td>
<td>0.01</td>
<td>2.0 (1.0-3.9)</td>
<td></td>
</tr>
<tr>
<td>rs1139793</td>
<td>TXNRD2</td>
<td>Thioredoxin reductase 2</td>
<td>Selenocysteine-containing flavoenzyme; maintains thioredoxins in a reduced state, thereby playing a key role in regulating the cellular redox environment</td>
<td>Subcutaneous fibrosis</td>
<td>CGS</td>
<td>0.012</td>
<td>3.2 (1.3-8.2)</td>
<td>45</td>
</tr>
<tr>
<td>SNP</td>
<td>Gene</td>
<td>Function</td>
<td>Tumor Type</td>
<td>Study</td>
<td>Effect Size</td>
<td>CI</td>
<td>p-value</td>
<td>Cohort</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------</td>
<td>-------------------------------------------------------------------------</td>
<td>------------</td>
<td>-------</td>
<td>-------------</td>
<td>---------------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td>rs2682585</td>
<td>XRCC1</td>
<td>X-ray repair cross complementing 1</td>
<td>Skin</td>
<td>CGS</td>
<td>0.02 (CGS)</td>
<td>0.8 (0.6-1.0)</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>rs1801516</td>
<td>ATM</td>
<td>ATM telangiectasia mutated</td>
<td>Overall</td>
<td>CGS</td>
<td>NP</td>
<td>1.2 (1.0-1.4)</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Serine/threonine protein kinase that is recruited and activated by DNA double-strand breaks; phosphorylates several key proteins that initiate activation of the DNA damage checkpoint, leading to cell cycle arrest, DNA repair, or apoptosis</td>
<td>Overall acute toxicity</td>
<td>CGS</td>
<td>1.5 (1.2-1.9)</td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Acute skin toxicity</td>
<td>CGS</td>
<td>1.7 (1.1-2.7)</td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Telangiectasia</td>
<td>CGS</td>
<td>1.3 (1.1-1.7)</td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Fibrosis</td>
<td>CGS</td>
<td>1.3 (1.0-1.6)</td>
<td>51</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Lung**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Function</th>
<th>Tumor Type</th>
<th>Study</th>
<th>Effect Size</th>
<th>CI</th>
<th>p-value</th>
<th>Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2868371</td>
<td>HSPB1</td>
<td>Heat shock protein family B (small) member 1</td>
<td>Pneumonitis</td>
<td>CGS</td>
<td>0.02 (TC)</td>
<td>0.29 (0.09-0.97)</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Esophagitis</td>
<td>CGS</td>
<td>0.045 (TC)</td>
<td>0.25 (0.07-0.88)</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.031 (VC)</td>
<td>0.25 (0.07-0.88)</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>rs1800469</td>
<td>TGFβ1</td>
<td>Transforming growth factor beta 1</td>
<td>Esophagitis</td>
<td>CGS</td>
<td>0.045 (TC)</td>
<td>2.5 (1.0-6.0)</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.023 (VC)</td>
<td>2.5 (1.1-5.6)</td>
<td>54</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: AUA, American Urological Association; BC, beta coefficient; CGS, candidate gene study; CI, confidence interval; GI, gastrointestinal; GU, genitourinary; GWAS, genome-wide association study; NP, not provided; RR, relative risk; HR, hazard ratio; TC, training cohort; VC, validation cohort.
breast cancer. One example was a study in which more than 2000 patients with breast cancer from four cohorts treated with radiotherapy were genotyped for SNPs related to the TGFB pathway and associations reported for several outcomes, including breast induration, telangiectasia, and overall toxicity. Significant and replicated associations with adverse outcomes following breast radiotherapy were reported for the TNFa SNP rs1800629 and rs2857595, which is located 25.7 kb from rs1800629 and resides in the intergenic region between NCR3 and AIF1.

Another validated study of patients with breast cancer identified SNP rs1139793 in TXNRD2 associated with subcutaneous fibrosis following radiotherapy. TXNRD2 encodes the mitochondrial selenoprotein thioredoxin reductase 2, which plays a central role in preventing oxidative damage. Thus, alteration of the protein encoded by this gene could affect upon reactive oxygen species produced through irradiation, and therefore influence the risk for fibrosis development following radiotherapy.

A separate study used a 2-stage design to investigate associations between SNPs in genes whose products are involved with responses to oxidative stress with toxicities following radiation treatment of approximately 2600 women diagnosed with breast cancer. The rs2682585 SNP in XRCC1 was found to be associated with risk for skin toxicities (OR = 0.77; 95% CI: 0.61-0.96; P = 0.02) and STAT score (−0.08; 95% CI: −0.15 to −0.02; P = 0.016). The protein encoded by XRCC1 plays a role in base excision repair of oxidative damage produced by radiation.

A GWAS was performed in which more than 1500 patients who received radiotherapy for breast cancer were examined for SNPs associated with adverse effects resulting from treatment. The quantile-quantile (Q-Q) plots from this study revealed a larger number of associations at the P < 5 × 10−7 level than would be expected by chance. This result provides evidence that common genetic variants are associated with risk for development of adverse effects following radiotherapy.

A study of more than 5000 patients treated for either breast or prostate cancer with radiotherapy reported an association between the rs1801516 in the ATM gene with ORs of 1.5 for acute and 1.2 for late toxicity.

**Lung Cancer**

It was reported in studies of patients treated with radiotherapy for non–small cell lung cancer (NSCLC) that the HSPB1 rs2868371 SNP was associated with grade 3 or greater radiation pneumonitis (P = 0.02), and that this SNP was also associated with the development of grade 3 or greater radiation-induced esophagitis in both training (P = 0.045) and validation cohorts (P = 0.031). HSP27 is a heat shock protein whose plasma concentrations are under genetic control of HSBP1. HSP27 increases cellular resistance to heat shock, oxidative stress, and inflammatory mediators. In addition, HSP27 additionally increases the antioxidant activity in cells and limits the toxicity of oxidized proteins through its chaperone activity. Thus, it is plausible that modulation of HSP27 levels through genetic alterations in HSPB1 could affect the sensitivity of patients with NSCLC for the development of lung pneumonitis and esophagitis following radiotherapy. In addition, it was reported that the TGFB1 rs1800469 SNP was associated with a higher risk of radiation esophagitis in both the training (P = 0.045) and validation (0.023) sets of patients with NSCLC. Radiation can activate TGFB1 from its latent form, which plays an important role in the etiology of radiation-induced inflammatory processes.

**Model Development**

An important factor in development of a radiogenomic predictive instrument is the creation of a suitable model. One approach is to build upon a normal tissue complication (NTCP) model, which has its basis dosimetric parameters, with the addition of genetic information and other patient-specific factors. Several predictive models have been created using the EMLasso technique, which represents a statistical approach for model building. This methodology helps to avoid overfitting or underfitting, includes cross-validation, employs the smallest number of parameters, and is appropriately designed for datasets with missing values, a situation common in radiogenomic research. This technique has been used for several studies, including predictive models for dysphagia resulting from head and neck cancer treatment, genitourinary toxicities following treatment of prostate cancer with radiotherapy, and esophagitis following chemoradiation treatment. In addition, decision analytic methods, such as decision curve analysis and net benefit, can assist with the quantification of clinical usefulness. Several other approaches have also been suggested to evaluate radiation-induced normal tissue effects based on dosimetric and clinical factors, including logit-equivalent uniform dose and relative seriality, nearest-neighbor prediction, and the Lyman-Kutcher-Berman model.

Another key factor to consider for the development of models is that normal tissue radiosensitivity for any particular tissue or organ is a complex trait dependent upon the expression of multiple genes whose variation is a reflection of the collective effect of numerous sequence variants. Hence, susceptibility for the appearance of adverse effects in a specific organ is likely the manifestation of several molecular pathways, which can be affected by the presence of SNPs in multiple genes. Therefore, any predictive instrument to be developed will need to incorporate a multi-SNP component.

An important aspect associated with model building is that this type of analysis provides insight as to the number of SNPs that will be necessary to create a clinically useful predictive instrument. One approach to address this issue has been through the employment of simulation data, which is informative as to the robustness of predictive models for discrimination between individuals at high risk for development of complications following radiotherapy and those at low risk. Several conclusions were obtained from these simulation experiments, including the following: (1) inclusion of SNPs present in the genome with a high-risk allele frequency and larger effect size enhances the accuracy of the model, (2)
increasing the number of SNPs included in a risk model improves the discrimination accuracy as quantitated through the use of the area under the curve for a receiver-operating characteristic curve, and (3) high area under the curve values can be achieved through the use of 50-100 common risk SNPs with effect sizes of 1.05-1.5. Substantial progress in radiogenomics research is being made toward discovery and validation of this number of SNPs. The results of these simulations are reassuring as they indicate that a relatively modest number of SNPs could form the basis of a clinically useful instrument capable of predicting risk for development of a particular form of toxicity resulting from cancer radiotherapy. It is, therefore, anticipated that a predictive test should be available for clinical use in the near future. Such an instrument should substantially improve upon and assist the clinical decision-making process.

**Design of Clinical Trials**

Now that substantial progress has been achieved in radiogenomics to identify biomarkers associated with development of adverse effects resulting from radiotherapy and advances have been realized toward model building, efforts are being focused on optimal design and patient selection for interventional trials using radiogenomic biomarkers. One important point to consider in the design of clinical trials is that unlike disease susceptibility, the risk for radiation-induced toxicities is continuous for the development of toxicities with an increasing incidence of complications associated with larger radiation doses or volumes irradiated. In addition, a SNP-predictive assay will likely require one or multiple thresholds for classifying risk into discrete categories. Thus, a classical biomarker trial design may not be appropriate, while an approach using a risk factor stratification methodology could be more suitable.

**Current Research and Future Directions**

In total, three large studies are currently in progress whose main goal is to discover new SNPs and validate previously identified genetic biomarkers predictive of susceptibility for the development of adverse effects resulting from radiotherapy. One such study involves roughly 6000 men treated for prostate cancer, which encompasses multiple cohorts created by RGC investigators. DNA samples from all of these men have been genotyped using a GWAS chip, and detailed clinical data are available with a minimum of 2 years follow-up. The following are the goals of this project: (1) discover new SNP associations and validate previously identified SNPs linked with the development of adverse outcomes resulting from radiotherapy, (2) build clinically useful multi-SNP models that incorporate dosimetric and clinical factors to predict susceptibility for the development of toxicities following radiotherapy, and (3) develop a low-cost, high-performance assay, and companion risk assessment tool to predict risk for development of complications resulting from treatment with radiation. Related to this aim, research is being conducted that is supported by the NIH Small Business Innovation program to help rapidly translate the findings from this project into an assay ready for implementation in the clinic and routine medical care.

A second large project is REQUITE (validating predictive models and biomarkers of radiotherapy toxicity to reduce side-effects and improve quality-of-life in cancer survivors). This is a multicenter study involving member investigators of the RGC. An important aspect of this project is that it addresses the problem of data harmonization, which is a significant challenge in most radiogenomic studies that involve multiple cohorts in which the subjects were followed using a variety of evaluative instruments. For REQUITE, identical categories of clinical and dosimetric information were obtained for all subjects, and the same health professional and patient-reported outcome forms were used at all enrolling centers. The objectives of REQUITE are to (1) perform a multicenter, observational cohort study in which epidemiologic, treatment, longitudinal toxicity, and quality-of-life data are collected from approximately 5000 patients treated with radiotherapy for breast, prostate, or lung cancer; (2) produce a centralized biobank in which DNA is isolated from patients enrolled in the observational study and create a centralized data management system for secure collection, integration, mining, sharing, and archiving of all project data; (3) validate published SNP biomarkers of radiosensitivity and discover new variants associated with specific forms of adverse effects following radiotherapy; (4) validate clinical and dosimetric predictors of radiotherapy toxicity and incorporate biomarker data; (5) design interventional trials to reduce long-term adverse cancer treatment effects; and (6) deliver interventional trial protocols using validated models incorporating biomarkers to identify patient subpopulations likely to benefit from interventions and to serve as a resource exploitable for future studies exploring relationships between adverse effects resulting from radiotherapy and the genetics of radiosensitivity using developing technologies such as next-generation sequencing.

A third project involves roughly 4500 women treated for breast cancer with radiotherapy in which blood samples and detailed clinical and follow-up information are available. These come from three cohorts in which blood samples, treatment, and dosimetric and follow-up data have been obtained for approximately (1) 1000 women treated as part of RTOG 1005, which is a trial examining the use of a hypofractionated protocol with a concurrent boost for treatment of breast cancer; (2) 2000 women enrolled into the REQUITE study, and (3) 1500 subjects from a series of clinical protocols performed at New York University School of Medicine.

It is anticipated that the results of these 3 large projects in radiogenomics will result in the discovery of new SNPs, and the validation of previously identified genetic markers, which will form the basis of an assay to predict outcomes from radiotherapy that will be ready for application in routine cancer care.

**Conclusion**

Substantial progress in radiogenomic research has been achieved toward the creation of a test predictive of the
susceptibility for individual cancer patients as to the development of adverse effects resulting from radiotherapy, which often have a deleterious effect upon the quality-of-life for these individuals. It is also likely that identification of SNPs and genes whose encoded products play a role in the molecular etiology of the development of radiation-induced toxicities will advance our understanding as to the molecular basis through which these adverse outcomes arise. It is expected that this knowledge should assist in the development of agents to prevent or mitigate these radiation-induced injuries. Thus, it is anticipated that clinical implementation of a predictive instrument characterized by a high level of sensitivity and specificity will substantially enhance the ability to select an optimal treatment for people diagnosed with cancer and thereby improve outcomes and advance precision radiotherapy.

References
32. Heublein S, Kazi S, Ogumudspotto SP: Proton-assisted amino-acid transporters are conserved regulators of proliferation and amino-acid-dependent mTORC1 activation. Oncogene 29:4068-4079, 2010