Published in final edited form as:

Phys Med Biol. ; 62(16): R179–R206. doi:10.1088/1361-6560/aa7c55.

Radiogenomics and Radiotherapy Response Modeling

Issam El Naqa¹, Sarah L. Kerns², James Coates³, Yi Luo¹, Corey Speers¹, Catharine M.L. West⁴, Barry S. Rosenstein⁵, and Randall K. Ten Haken¹

¹Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, USA

²Department of Radiation Oncology, University of Rochester, Rochester, NY, USA

³Department of Oncology, University of Oxford, Oxford, UK

⁴Division of Cancer Sciences, Manchester Academic Health Science Centre, Christie Hospital NHS Trust, University of Manchester, Manchester, UK

⁵Departments of Radiation Oncology and Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Abstract

Advances in patient-specific information and biotechnology have contributed to a new era of computational medicine. Radiogenomics has emerged as a new field that investigates the role of genetics in treatment response to radiation therapy. Radiation oncology is currently attempting to embrace these recent advances and add to its rich history by maintaining its prominent role as a quantitative leader in oncologic response modeling. Here, we provide an overview of radiogenomics starting with genotyping, data aggregation, and application of different modeling approaches based on modifying traditional radiobiological methods or application of advanced machine learning techniques. We highlight the current status and potential for this new field to reshape the landscape of outcome modeling in radiotherapy and drive future advances in computational oncology.

Keywords

Radiogenomics; outcome modeling; tumor control; normal tissue toxicity; machine learning

Introduction

It is currently widely accepted that multiple heterogeneous factors shape a patient's response to radiotherapy (Lambin *et al.*, 2013b). Categorization of these factors enhances our ability to characterize the way ionizing radiation interacts with living tissues. Radiotherapy is a clinical treatment that applies a physical tool based on high energy particles to modify the biological environment of targeted cells. Hence, these factors can be broadly divided into clinical, physical and biological factors (El Naqa, 2014a).

Efforts to develop quantitative methods to better understand radiation effects have been ongoing since the discovery of x-rays at the end of the 19th century. It is therefore not surprising that applications of physical principles and mathematical relationships have played prominent roles in characterization of radiotherapy outcomes. Given the limited data

and computational power, early investigations took a *reductionist* approach with a focus on applying power-laws to better understand the effects of dose, volume, and time on radiotherapy response, three main defining elements which remain important today in the application of radiotherapy. For instance, Kingery, an instructor in dermatology at the University of Michigan, developed one of the earliest mathematical models of radiation response termed the saturation method using repetitive applications of single exponential decay (Kingery, 1920).

Treatment outcomes in radiotherapy are usually characterized by tumor control probability (TCP) and the surrounding normal tissues complication probability (NTCP) (Steel, 2002; Webb, 2001). Traditionally, these outcomes are modeled using information about the dose distribution and the fractionation scheme (Moissenko *et al.*, 2005). However, it is recognized that radiation response may also be affected by multiple clinical prognostic factors (Marks, 2002) and more recently, inherited common genetic variations have been suggested as playing an important role in radiosensitivity (Alsner *et al.*, 2008; West *et al.*, 2007).

The widely used and often controversial linear-quadratic (LQ) radiation cell kill model traces its conception to the 1940s (Lea and Catcheside, 1942) and remains the cornerstone of radiobiology modeling. Recent advances in biotechnology and the ability to generate massive amounts of data accompanied with improved computational power has led to the emergence of ‘Omics’ technologies as *holistic* approaches to analyze whole genomes (“genomics” - coined by T.H. Roderick, a geneticist, over a beer in Bethesda, MD (Kuska, 1998)). This notion has expanded to include the analysis of all mRNAs (transcriptomics), proteins (proteomics) and metabolites (metabolomics) of a specific biological sample in a non-targeted and non-biased manner using data-driven rather than traditional hypothesis-driven approaches (Horgan and Kenny, 2011). In this context, it seemed that advances in computational biology outpaced progression in radiobiological modeling, at least, in the arena of data-driven approaches. However, in recent years, radiotherapy started to catch up to this data-centric era. Notably, in 2010, authors from the QUANTEC (Quantitative Analyses of Normal Tissue Effects in the Clinic) effort supported a data-pooling culture of retrospective and prospective data (Deasy *et al.*, 2010). Moreover, the QUANTEC introductory papers emphasized that, in addition to traditional dose-volume metrics, endogenous biological markers (biomarkers) (Bentzen *et al.*, 2010b) and computer extracted imaging features (radiomics) could act as predictors of radiation side effects and surrogate endpoints for clinical outcomes (Jeraj *et al.*, 2010). This is equally true in the case of tumor response where physical and also biological factors have a role in determining treatment success (Begg *et al.*, 2011). In addition, a significant effort has been launched to incorporate concepts of “Big Data” to enhance and accelerate both clinical and basic science research in radiotherapy (Benedict *et al.*, 2016; Rosenstein *et al.*, 2016). As part of this effort, radiogenomics (Rosenstein *et al.*, 2014; Abazeed *et al.*, 2013; Andreassen *et al.*, 2012; Coates *et al.*, 2015; Kerns *et al.*, 2014b) and radiomics (Lambin *et al.*, 2012; Aerts *et al.*, 2014; El Naqa, 2014b) have emerged as promising spheres for active investigation.

Radiogenomics was originally used to describe the study of genetic variation associated with response to radiation therapy to distinguish it from radiomics, which involves the analysis of large amounts of quantitative features from medical images using data-characterization

algorithms associated with prognosis and treatment response. However, the latter also could fall under radiogenomics when attempting to correlate imaging and genetic markers. More recent efforts have been directed towards integrative systems-based approaches which would encompass both data types as part of a radiotherapy ‘pan-omics’ framework (El Naqa, 2014a) vis-a-vis a better decision support system (Lambin *et al.*, 2013a).

The focus of this review will be on the biological aspects of radiogenomics, which has witnessed tremendous growth recently as highlighted by the national and international efforts underpinned by the establishment of the international Radiogenomics Consortium to coordinate and lead efforts in this area (West and Rosenstein, 2010). The following sections provide an overview of radiogenomics starting with genotyping, data aggregation, and application of different modeling approaches based on modifying traditional radiobiological methods or using advanced data-driven approaches such as machine learning techniques.

1.1 Biomarkers and the world of –omics

Molecular biology has witnessed tremendous progress in recent years due to the extraordinary advances in biotechnology and the success of the Human Genome Project and its offshoots. These technologies provide tools to screen a large number of biological molecules and to identify biomarkers that characterize human disease or response to treatment. These biomarkers follow from the central dogma of biology (Crick, 1970), in which biological information is expressed via the sequential transcription of the deoxyribonucleic acid (DNA) genetic code into ribonucleic acid (RNA) and subsequent translation of the RNA into proteins, which further interact with intermediaries of metabolic reactions (metabolites) to determine cell function. Epigenetic modifications of DNA and associated histone molecules contribute an additional layer of regulatory influence on this process.

A biomarker is formally defined as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathological processes, or pharmacological responses to a therapeutic intervention”(Group, 2001). Biomarkers can be divided into *prognostic* and *predictive* biomarkers. A prognostic biomarker informs about a patient’s likely cancer outcome (e.g., disease recurrence, disease progression, death) independent of the treatment received, e.g., Prostate-specific antigen (PSA) serum levels at the time of a prostate cancer diagnosis (Ballman, 2015). A predictive biomarker informs about the likely benefit from a therapeutic intervention, e.g., mutations in the epidermal growth factor receptor (EGFR) gene predicting benefit from gefitinib (Oldenhuis *et al.*, 2008). In the context of radiogenomics, biomarkers lend themselves primarily to the predictive category. Moreover, biomarkers could be categorized based on the biochemical source of the marker into *exogenous* or *endogenous*.

Exogenous biomarkers are based on introducing a foreign substance into the patient’s body such as those used in molecular imaging (e.g., positron emission tomography with fluorodeoxyglucose (FDG-PET), a metabolic analog). Conversely, endogenous biomarkers can further be classified as (1) ‘expression biomarkers,’ measuring baseline or changes in gene expression, protein or metabolite levels or (2) ‘genomics biomarkers,’ based on structural variations of tumors or normal tissues, in the underlying DNA genetic code. The

different categories of predictive biomarkers that are used for radiation therapy response and the multiple types of specimens involved such as tissue or peripheral blood samples are shown in Figure 1 as part of the general radiogenomics platform.

1.2 Radiotherapy and –omics

In radiation oncology, biomarkers have the potential to be used to classify patients based on predicted responses or to stratify patients for optimization of treatment plans via dose reduction or dose escalation. Data-intensive models that can capture underlying trends in toxicity and/or factors relevant for tumor control, without under- or over-fitting, can provide valuable patient-specific data to improve outcomes by personalizing treatment. Modern, high-throughput sequencing technology has facilitated the discovery of novel biomarkers from larger numbers of samples; however, the highly non-linear mechanisms of biological response to radiation therapy have meant that the functional relationships governing clinical effects remain largely unknown or, at the least, poorly understood.

The goal of radiogenomics as noted earlier is to identify the genetic markers that are associated with and may predict of the development of adverse effects and likelihood of tumor response resulting from cancer radiotherapy. This information will assist patients and their physicians to select the optimal treatment for each patient individually based upon genetic factors and other patient characteristics to maximize therapeutic benefit. For instance, recent developments in next-generation sequencing have made it possible to identify prognostic and predictive signatures of tumor response based on genomic profiles for personalized decision making in prostate cancer treatment (Fraser *et al.*, 2015).

To date, large collaborative studies in radiogenomics focused on late side-effects, since for many people diagnosed with cancer, especially early stage disease, the long-term survival rates are high. However, many of these long-term survivors will suffer from radiotherapy complications, which may impact upon their quality of life. Patients predicted to be at high risk for radiation-induced complications based upon genomic factors may opt for a non-radiation treatment, if available, or highly conformal treatments such as proton therapy. In contrast, patients predicted to be at low risk may undergo radiotherapy, potentially with dose escalation, which could improve their chance for tumor control. An equally important consideration is the predicted (and actual) tumor response to radiotherapy. If a tumor, regardless of the normal-tissue toxicity risk, was known to be inherently resistant to radiation treatment based on biological factors, then alternative treatments could be pursued, if feasible. Conversely, tumors predicted to be exquisitely sensitive to radiation treatment may be effectively treated and cured with lower doses of radiation, thus further reducing the risk of normal tissue toxicity while not compromising tumor control.

2 Radiogenomics Overview

2.1 From radiosensitivity assays to radiogenomics

Radiation-induced cell death is caused principally by an accumulation of DNA damage (Painter *et al.*, 1980). Much of the damage induced can be readily repaired. In some cells the damage triggers programmed cell death (apoptosis) to avoid propagation of damaged genetic

material (Elmore, 2007). The predominant type of cell death following irradiation, however, occurs as cells attempt to divide during mitosis (mitotic catastrophe) (Eriksson and Stigbrand, 2010). Biological contributors to the intrinsic radiosensitivity of cells include DNA damage response capacity, factors controlling cell cycle progression, and potential to undergo apoptosis. Physical quantities such as dose-rate and radiation quality can also play important roles in determining radiation response. Extrinsic factors such as oxygen tension or physical quality of radiation are also well known to affect radiosensitivity (Brahimi-Horn *et al.*, 2007). Such competing biophysical effects contributing towards survivability make predicting outcomes in the clinic extremely challenging.

Early pioneering efforts in developing assays for measuring human tumor and normal tissue radiosensitivity involved clonogenic assays, for example to measure surviving fraction at 2 Gy (2SF2) (Buffa *et al.*, 2001; Dunne *et al.*, 2003). Experimental observations of *in vitro* work led to the development of mechanistic (or analytical) models to predict both tumor control probability (TCP) and normal tissue complication probability (NTCP). Such models have, however, fallen short of their mark in that their usefulness in a prospective clinical setting has proven to be limited due to the inherent heterogeneity of patients' radiogenomic profiles, which is inadequately accounted for by these models (Adamus-Gorka *et al.*, 2011).

Radiogenomic models hold promise in improving predictive performance over canonical TCP/NTCP models as they can be designed around any number or type of variables that include biological, physical, and other factors. Moreover, advanced informatics approaches that can identify, integrate and use prospectively-selected "pan-omic" information have started to pave the path for biologically-driven models to improve outcomes (El Naqa, 2014a).

A substantial focus of research efforts in radiation biology has been directed towards the development of assays to predict response to radiotherapy. The sections below highlight radiogenomics efforts for developing predictive radiosensitivity assays for tumor and normal tissues response.

2.1.1 Tumor response—Interest in developing radiosensitivity assays to predict tumor response stems from research carried out in the 1980s (Fertil and Malaise, 1981; Deacon *et al.*, 1984; Fertil and Malaise, 1985). The work highlighted a relationship between the initial slope of radiation survival curves for cell lines derived from different tumor types and their associated clinical radioresponsiveness. That is, cell lines from radiocurable tumors such as lymphomas were more radiosensitive than those from cancers that were refractory to radiotherapy such as melanoma. Studies were subsequently established to measure the radiosensitivity of human tumors with the initial focus on clonogenic assays. The largest study was carried out in 156 cervix cancer patients undergoing radiotherapy and demonstrated that measurements of primary tumor SF2 was an independent prognostic factor for survival outcomes (West *et al.*, 1993, 1997; West, 2007). However, these studies were generally of modest size yielding equivocal results. Thus, interest moved towards investigating alternative assays that were more rapid, such as those involving DNA damage, chromosome alterations or apoptosis. For example, there is good evidence that high protein expression in pre-treatment tumor samples of the DNA damage response gene *MRE11*

predicts benefit for radiotherapy rather than cystectomy in patients with muscle-invasive bladder cancer (Choudhury *et al.*, 2010; Laurberg *et al.*, 2012).

With the advent of high-throughput profiling approaches, studies have attempted to derive gene expression signatures associated with tumor radiosensitivity. Different approaches can and have been used, e.g., supervised classification based on outcome following radiotherapy (Pramana *et al.*, 2007) or measurements of radiosensitivity in cell lines (Amundson *et al.*, 2008). The most developed work comes from Torres-Roca and colleagues who developed a signature using a database of 48 human tumor cell lines with radiosensitivity (SF2) and gene expression data (Torres-Roca, 2012). Signaling pathway network analysis identified 10 genes related to SF2 that were used to build a rank-based linear regression algorithm to predict radiosensitivity (a radiosensitivity index, RSI). The RSI has been shown to be prognostic in multiple cohorts involving several tumor types (Eschrich *et al.*, 2009; Ahmed *et al.*, 2015) and to predict benefits from radiotherapy in breast cancer (Eschrich *et al.*, 2012). The RSI was combined with the linear quadratic model to derive a genomic-adjusted radiation dose, which was prognostic in multiple cohorts and was proposed as a framework to design genomically-guided radiotherapy trials (Scott *et al.*, 2017).

2.1.2 Normal tissue response—The interest in developing a test to predict risk of toxicity following radiotherapy coincided with the pioneering work of the 1980s in studying tumor response (Fertil and Malaise, 1981; Deacon *et al.*, 1984; Fertil and Malaise, 1985). It was known that skin fibroblasts cultured from individuals with the cancer pre-disposing syndrome ataxia telangiectasia (associated with hyper-sensitivity to radiotherapy) were roughly 3-fold more sensitive to radiation *in vitro* than cells cultured from non-syndromic individuals (Taylor *et al.*, 1975). There was, however, little evidence that differences in radiosensitivity could be measured *in vitro* for most individuals. With the introduction of the linear quadratic model, it was shown that differences in the radiosensitivity of fibroblasts could be measured using parameters reflecting the initial slope of radiation survival curves (Malaise *et al.*, 1987). It became accepted that intrinsic radiosensitivity varies among skin fibroblasts cultured from non-syndromic individuals due to ‘unidentified genetic factors’ (Little *et al.*, 1988). This work led to studies attempting to measure radiosensitivity to predict for risk of toxicity based on skin fibroblast, lymphocyte, and gene expression assays as discussed below.

2.1.2.1 Skin fibroblast radiosensitivity assays: As a manifestation of an inherent genetic susceptibility to radiation toxicity, studies were performed in which the *in vitro* radiosensitivity of skin fibroblasts was measured. It should be noted that although it could be expected that cell killing by radiation may play a central role in the etiology of early effects, late radiation effects in the skin are more likely a manifestation of a cytokine cascade induced by radiation resulting in an inflammatory response leading to a fibrotic reaction (Bentzen, 2006). Several initial studies reported an association between dermal fibroblast radiation sensitivity with the severity for both early and late effects (Burnet *et al.*, 1992; Loeffler *et al.*, 1990; Oppitz *et al.*, 2001). However, replication studies generally were not able to validate these initial findings as there was a lack of correlation between *in vitro*

fibroblast radiosensitivity with late effects and only a weak association with early skin responses (Begg *et al.*, 1993; Peacock *et al.*, 2000).

2.1.2.2 Lymphocyte assays: Early studies using blood lymphocytes showed a correlation between *in vitro* radiosensitivity with the development of toxicities exhibited by patients following radiotherapy employing a variety of assays, including clonogenic (West *et al.*, 1995), chromosome damage (Jones *et al.*, 1995) and DNA damage (Alapetite *et al.*, 1999). A relatively large prospective study reported that lymphocyte radiosensitivity measured using a clonogenic assay (SF2) was an independent predictor for risk of toxicity (West and Barnett, 2011). However, like the experience using fibroblast assays, findings tended to be equivocal with too many small and badly designed studies. Researchers continue to explore cell based assays with a particular interest in gamma H2AX formation (Pouliliou *et al.*, 2015). Experimental variations and lymphocyte cell type heterogeneity were highlighted as limitations of such work (Crompton and Ozsahin, 1997).

Probably the most promising work involving lymphocytes employs an apoptosis assay that considers the cell-type specific radiosensitivities. Using this approach, it has been reported that the response of CD4 and CD8 T-lymphocytes to irradiation correlates with radiation-induced late effects following radiotherapy (Mirjolet *et al.*, 2016; Azria *et al.*, 2015). Most notably, an inverse correlation has been reported between radiation-induced T-lymphocyte apoptosis, particularly for CD8 cells, with the development of late effects in patients from whom the lymphocytes were derived.

2.1.2.3 Gene expression profiling: The development of gene expression microarrays and next-generation RNA sequencing methods provided the ability to measure the expression of a large number of genes following irradiation. It is now possible to identify genes which are expressed following irradiation (i.e., change in gene expression after irradiation compared to before irradiation) and correlate with the development of radiation-induced injuries resulting from radiotherapy. Expression was studied in fibroblast cell lines derived from breast cancer patients exhibiting a range of subcutaneous fibrotic reactions following post-mastectomy radiotherapy (Rodningen *et al.*, 2008). This work identified a set of 18 genes that could differentiate patients who were at low risk for fibrosis compared with patients at high risk based upon the differential expression of these genes in the two populations. Using quantitative real time PCR, it was found that the relative magnitude of the increase in gene expression of irradiated compared with unirradiated fibroblasts provided even greater discrimination between patients who were either sensitive or resistant to the development of subcutaneous fibrosis following radiotherapy. The results of this study indicated that differential gene expression of specific genes could distinguish between patients at low risk from those at high risk for a fibrotic response. Further work has identified a set of 67 radiation-induced gene expressions capable of differentiating between severe radiosensitive and normal reacting patients (Mayer *et al.*, 2011).

2.2 Radiogenomics categories

2.2.1 Structural variations (SNPs/CNVs)—There are several classes of genomic variants, all of which may potentially impact normal tissue radiosensitivity and susceptibility for development of toxicity. The following will focus on the two most common ones.

2.2.1.1 Single nucleotide polymorphisms (SNPs): Single nucleotide polymorphisms (SNPs) and small insertions and deletions (indels) are the most widely studied genomic variants to date in the context of radiosensitivity. Common SNPs are single base pair variable sites that are generally present in at least 1% of the population. SNPs can affect protein coding sequences, non-coding introns, and intergenic regions that may include cis-acting and/or trans-acting regulatory sites. There are approximately 10 million SNPs in the average human genome, occurring approximately every 300 nucleotides. In contrast to common SNPs, rare variants are present in less than 1% of the population. Rare variants can also occur in protein coding sequences, non-coding introns, and intergenic regions. Because of their rarity, very large sample sizes are needed to detect associations with disease, but rare variants may be more causally related to disease than SNPs. SNPs generally tag a genomic locus, or region, that contains one or more causal variants.

Progress in research aimed at identifying specific genetic variants associated with radiotherapy toxicity largely reflects progress in development of genomic technology and resources. Early studies used the so-called candidate gene approach, in which a small number of SNPs in or near genes known to play a role in cellular radiation response were tested for association with clinical toxicity. These studies were limited in the number of SNPs investigated and the genes of interest, and few positive associations were replicated in additional studies (Barnett *et al.*, 2012; Kerns *et al.*, 2014b). One challenge with these early studies was our incomplete understanding of the genetic architecture of complex diseases such as cancer. Advances in genotyping and sequencing technology have led to development of large reference population datasets and enabled performance of large genome-wide association studies of numerous complex diseases and traits. We now know that most common SNPs have very modest effects on increasing or decreasing risk for developing disease, and large sample sizes are needed to detect replicable SNP associations. Early radiogenomics studies were largely under-powered to detect such modest effects.

More recent large candidate gene studies in radiogenomics have been successful. For example, SNPs rs1800629 and rs2857595, which tag a locus upstream of *TNF*, showed a replicated association with late toxicity in over 2,000 breast cancer patients treated with radiotherapy (Talbot *et al.*, 2012). In another study, rs2682585 in *XRCC1* showed a replicated association with late skin toxicity and overall toxicity in over 2,500 breast cancer radiotherapy patients (Seibold *et al.*, 2015). SNP rs2868371 in *HSPB1* was associated with multiple late complications in lung cancer radiotherapy patients (Pang *et al.*, 2013; Guerra *et al.*, 2012) including pneumonitis and esophagitis. In a large individual patient data meta-analysis of over 5,000 breast and prostate cancer patients, rs1801516, a protein coding SNP in *ATM*, was associated with overall acute and late toxicity (Andreassen *et al.*, 2016).

Advances in technology, and decreases in costs, for large-scale genotyping and sequencing projects have enabled the application of genome-wide approaches to the study of

radiosensitivity. In contrast to candidate-gene studies, genome-wide association studies (GWAS) of late radiotherapy toxicity have led to the identification of SNPs that lie in or near genes not previously known to play a role in radiation response. For example, SNPs in *TANC1* were found to be significantly associated with overall late urinary and rectal toxicity in prostate cancer patients in a three-stage GWAS (Fachal *et al.*, 2014). *TANC1* plays a role in regeneration of damaged muscle tissue, representing a pathway not previously implicated in radiotherapy toxicity. This study also found that the top selected SNP, rs264663, was associated with expression of *TANC1* mRNA, and it showed a statistical correlation with total biologically effective dose of radiation. A meta-analysis of four GWAS, also in prostate cancer patients, identified two additional loci: SNP rs17599026, which lies in *KDM3B*, was associated with increased urinary frequency; and rs7720298, which lies in *DNAH5*, was associated with decreased urine stream (Kerns *et al.*, 2016). Neither of these genes was previously implicated in cellular radiation response, and they appear to be novel radiosensitivity genes.

While radiogenomics GWAS are proving to be important for uncovering novel radiation biologic pathways, the results of such studies can also be useful for development of predictive models. Several published studies have shown a role for genetic variants, mainly SNPs, in existing predictive models of radiotherapy response. For example, Tucker *et al.* showed that inclusion of five SNPs located in or near several candidate genes (*TGFB1*, *VEGF*, *TNF*, *XRCC1* and *APEX1*) significantly improved performance of the Lyman NTCP model of radiation pneumonitis in non-small cell lung cancer patients (Tucker *et al.*, 2013). Another study found that inclusion of the top selected 12 SNPs from a GWAS along with clinical factors (age, hormone therapy, and radiation modality) improved performance of a logistic regression model of erectile dysfunction among prostate cancer patients (Kerns *et al.*, 2013). While these initial results are promising, they likely suffer from overfitting and require testing in independent studies. Evidence from polygenic models of other complex diseases suggests that many SNPs will be required to develop predictive models with high sensitivity and specificity. Indeed, results from a simulation study of radiogenomics outcomes suggests that tens to hundreds of SNPs will be required for good performance models, depending on the allele frequency and effect size of each SNP (Kerns *et al.*, 2015). Results of published GWAS suggest that many more radiosensitivity SNPs remain to be discovered (Barnett *et al.*, 2014), and collaborative efforts will result in increased sample sizes for studies that are adequately powered for further SNP discovery.

2.2.1.2 Copy number variations (CNVs): Gene dosage, or copy number variations (CNVs), are a second genetic variable that may be relevant to radiotherapy toxicities (Mei *et al.*, 2015; Shlien and Malkin, 2009). CNVs capture a much larger part of the genome, spanning thousands to millions of base pairs, whereas SNPs each describe changes specific to only one nucleotide. Thus, CNVs are indicative of more macro-scale changes in the genome that may play more significant roles when using a datamining approach. The objective of CNV analysis is to identify the chromosomal regions at which the number of copies of a gene deviates from two. These could be gains (CNV>2) or losses (CNV<2) (Yavas *et al.*, 2009). CNVs have been identified in several complex diseases such Crohn's disease (McCarroll *et al.*, 2008), psoriasis (de Cid *et al.*, 2009), autism (Qiao *et al.*, 2013;

Pinto *et al.*, 2010), and susceptibility to cancer (Kuiper *et al.*, 2010; Liu *et al.*, 2009). CNVs are less numerous than SNPs but they affect up to 10% of the genome, disrupting coding sequences or interrupting long range gene regulation, therefore, accounting for more differences among individuals (Zhang *et al.*, 2009). Moreover, as noted earlier, the association probability of a SNP with radiation-induced toxicities might be small and inefficient because of the large number of SNPs. Thus individual SNPs are expected to have limited impact in contrast with a particular CNV, which can affect up to several Mbs of DNA per variant leading to a more substantial impact on gene expression. In addition, there is evidence to suggest that there is limited overlap between SNPs and CNVs indicating potential complementary effect (Stranger *et al.*, 2007). The *XRCC1* CNV has been shown as a potential risk factor of radiation-induced toxicities in prostate cancer (Coates *et al.*, 2015). Large-scale profiling of 533 genetically annotated human tumor cell lines identified several genetic features (somatic copy number alterations, gene mutations and basal mRNA expression) that regulate survival after exposure to radiation (Yard *et al.*, 2016). These included functionally important genes in cell cycle regulation (*E2F1*), chromosome maintenance (*KIF3B*), glutathione synthesis (*GSS*) and apoptosis (*BCL2L1*).

2.2.2 Gene expression (mRNA, miRNA, lncRNA)—Gene expression profiling entails the use of techniques to investigate the relative rate of expression of many mRNA transcripts at a single time with respect to some baseline genes (called housekeeping genes) or across different time points (e.g., gene expression changes between before and after radiation therapy), which is more typical of radiogenomics applications. Microarray techniques using fluorophores and unbiased sequencing approaches (RNA-seq) are the most commonly used methods for determining such expression levels.

The advent of high-throughput expression profiling techniques has meant that mRNA levels can be readily quantified and integrated into outcome models. For example, several mRNA signatures have been developed to predict the efficacy of neoadjuvant chemotherapy in breast cancer patients. Several signatures have been derived that reflect tumor hypoxia and predict benefit from combining hypoxia-modifying treatment with radiotherapy (Eustace *et al.*, 2013; Yang *et al.*, 2017). The most developed tumor radiosensitivity mRNA signature is the one developed by Torres-Roca and colleagues (see Section 2.1.1). Alternatively, rather than searching for signatures, which don't explain underlying radiobiology, mRNA panels can be used to further explore and elucidate the mechanisms by which certain cancers exhibit radiosensitivity or resistance – examples of such cases include the well-known National Cancer Institute (NCI) mRNA expression panel (Kim *et al.*, 2012). The study by Yard *et al.* revealed gene expression sets that correlated with radiation sensitivity including DNA damage response, cell cycle, chromatin organization and RNA metabolism while genes that correlated with radiation resistance included cellular signaling, lipid metabolism and transport, stem-cell state, cellular stress and inflammation (Yard *et al.*, 2016).

Micro RNAs (miRNA) have also become of interest in oncology in recent times. They have been identified as playing oncogenic as well as tumor suppressor roles and may play a role in disease progression (Wang and Luo, 2015). In terms of radiosensitivity, miRNAs have been reported to modulate the *PI3K/AKT* pathway, influence or be influenced by hypoxia, *MAPK* signal transduction and *TGF-β1* expression while miRNA implication in

radioresistance has also been identified (Buffa *et al.*, 2011; Huang *et al.*, 2010; Cai *et al.*, 2013).

Another class of RNA transcripts that are gaining interest are long non-coding RNAs (lncRNAs), which are abundant in the genomes of higher organisms and seem to be part of the genome regulatory machinery. They are defined as transcripts of 200 up to 100,000 bp lacking an open reading frame (Bertone *et al.*, 2004). Recent studies suggest that lncRNA plays a role in intercellular communication (gene-silencing) in response to irradiation (O'Leary *et al.*, 2015; Zhang *et al.*, 2016).

Challenges to dealing with the quantification of RNA expression levels include the inability to quantify transcripts whose expression levels may be intrinsically low but of significant biological effect while for microarrays the complementary DNA (cDNA) sequences must already be known. Although RNA expression analysis techniques can readily generate large amounts of data, to be useful such variables must be understood in terms of their biology – simply identifying statistically significant variables from large pools of data does not, by itself, give much insight into the underlying biological interactions taking place.

2.2.3 Epigenetics changes—Epigenetics refers to heritable changes in the genome or chromatin that do not involve direct changes in the DNA sequence. Examples include DNA methylation, histone modification (methylation, acetylation), and non-coding RNAs, among others (Rakyan *et al.*, 2011). Unlike DNA variants such as SNPs and CNVs, epigenetic biomarkers are dynamic over time, among different tissue and cell types, and in response to environmental exposures. These features make them particularly challenging to study, but also potentially of great importance in radiation response, in which there is a large environmental exposure acting on the tissues at risk. Because epigenetic markers change over time, it can be difficult to determine whether a particular variant is a cause or consequence of the disease or outcome of interest. This is of particular importance in studies of radiosensitivity. Whereas in SNP/CNV studies, genomic DNA can be isolated from a blood sample collected at any time before or after radiation, epigenetic studies would require that blood (or other tissue) be collected prior to radiotherapy if the aim is to identify predictive markers. In this way, epigenetic studies are similar to studies of mRNA expression. Similarly, while germline DNA variants are the same across all normal cells or tissue types, epigenetic variants can vary from one tissue to another, and even among cell types comprising a single tissue or organ. Thus, for studies of normal tissue toxicity following radiotherapy, it may be important to study the particular tissue or cell type affected. This can be challenging given the complex interplay between cell types within an organ at risk as well as immune cells responding systemically to local radiation damage. Effects of irradiation on DNA methylation, a key epigenetic mechanism regulating the expression of genetic information could be used to predict response (Miousse *et al.*, 2017). Several studies have highlighted specific epigenetic variants and their impact on late toxicity (Weigel *et al.*, 2016; Weigel *et al.*, 2015). For instance, *KDM3B* has been shown to have a potential role on histone demethylation and radiosensitivity (Kerns *et al.*, 2016).

2.3 Resources, consortiums and nucleic acid databases

Research in radiogenomics has made substantial progress in recent years to identify the genetic/genomic factors associated with the development of normal tissue toxicities following radiotherapy. One of the major factors responsible for the progress that has been made in this field of research is due to formation of the Radiogenomics Consortium (RGC) in 2009 (West *et al.*, 2010). It was recognized that in order to conduct definitive studies, it would be necessary to both markedly increase the size of cohorts being examined and to include multiple cohorts for meta-analyses and replication studies. To achieve this aim, the RGC was established and became a National Cancer Institute/NIH-supported Cancer Epidemiology Consortium (<http://epi.grants.cancer.gov/Consortia/single/rgc.html>). The RGC currently consists of 217 investigators at 123 institutions in 30 countries. The common aims of the RGC investigators are to identify SNPs associated with radiotherapy adverse effects and to develop predictive assays ready for clinical implementation. The goal of the RGC is to bring together collaborators to pool samples and data for increased statistical power of radiogenomic studies. An important function of the RGC is that it has facilitated cross-center validation studies, which are essential for a predictive instrument to achieve widespread clinical implementation (Lambin *et al.*, 2013b). Initiatives for validation are equally important. The REQUITE Project sponsored by the European Union aims to develop a centralized database for validation of biomarker data (West *et al.*, 2014).

Generally speaking, modern radioresponse modeling schemes in radiotherapy could be divided into *top-down* or *bottom-up* approaches. *Top-down* approaches follow from generalizing the notion of systems biology into a systems radiobiology approach, where intra-radiotherapy changes and post-radiotherapy treatment outcomes could be optimized through using complex system analyses (e.g., machine learning methods) (El Naqa, 2013; Oh *et al.*, 2012). On the other hand, *bottom-up* approaches are based on first principles of radiation physics and biology to model cellular damage temporally and spatially (e.g., multi-scale modeling) (El Naqa *et al.*, 2012; Prokopiou *et al.*, 2015; Torres-Roca, 2012; Oh *et al.*, 2012; Pater *et al.*, 2014; Pater *et al.*, 2016; Stewart *et al.*, 2015). Typically, bottom-up approaches would apply advanced numerical methods such as Monte-Carlo (MC) to estimate the molecular spectrum of damage in clustered and non-clustered DNA lesions ($\text{Gbp}^{-1} \text{Gy}^{-1}$) (Nikjoo *et al.*, 2006). The temporal and spatial evolution of the effects from ionizing radiation can be divided into three phases: physical, chemical, and biological. Different available MC codes aim to emulate these phases along the molecular and cellular scales to varying extents such as KURBUC (Nikjoo *et al.*, 2016), the Geant4-DNA project (Incerti *et al.*, 2016) and PARTRAC (Friedland *et al.*, 2011). A detailed review of many current MC particle track codes and their potential use for radiobiological outcome modelling is given in (El Naqa *et al.*, 2012). In the context of radiogenomics, the focus in the literature has been mainly on top-down approaches (e.g., analytical or data-driven), which will be the subject of this review and they generically follow the schematic depicted in Figure 2 (El Naqa, 2013).

2.3.1 Single versus multiple biomarkers—The addition of relevant biomarkers to clinical or dosimetric-based outcomes models is likely to improve prediction performance by capturing more of the underlying manifestation of the pathophysiology in question. Thus,

more than one biological variable or type of biological variable may be included in a model depending upon the complexity of the outcome being studied. Indeed, it has been demonstrated that combinations of SNPs and CNVs can improve prediction performance when used alongside dosimetric and clinical data to predict late radiation toxicity in prostate cancer patients treated with hypofractionated radiotherapy (Coates *et al.*, 2015). Furthermore, SNPs or CNVs alone did not improve performance as significantly as when used together.

Data-driven modeling does not require specific variables to stratify one-another thereby allowing for dosimetric, clinical, and biological variables to be treated equally but weighted as per cross-validated performance enhancement (so called “mixed-type” models). It has been demonstrated on a limited patient population that data-driven models are more robust and have higher classification power when considering late radiotherapy outcomes and radiogenomic model-types (Coates *et al.*, 2015). In the context of analytical models, biomarkers can be added through dose modifying factors (DMFs) as discussed in detail below (Coates *et al.*, 2016).

2.3.2 Analytical models—Analytical techniques, also known as mechanistic or phenomenological modeling approaches, attempt to predict outcomes or effects based on reductionist simplifications or radiobiological effects. They include basic mechanisms of action that agree with experimental results and are thus thought to be partially theory-based. Analytical models can be either tumor control probability (TCP) or normal tissue complication probability (NTCP) approaches; however, analytical TCP models that describe effects at the cellular level are most common based on the effectiveness of radiation to induce cell death in tumors.

Undoubtedly, the most well-known analytical model used for TCP calculations in the linear-quadratic (LQ) model. The model is relatively straightforward in terms of mathematics and closely follows *in vitro* survival curves at fraction sizes < 5–6 Gy, after which point significant deviations can occur, especially with high linear energy transfer (LET) charged particle therapy. Modifications to the classical LQ model and further limitations have been discussed. (Kirkpatrick *et al.*, 2008; Coates *et al.*, 2016).

The most commonly employed NTCP model used in literature today is the Lyman-Kutcher-Berman (LKB) model. Variants of the classical LKB model have been proposed in literature, to include clinical or biological risk factors (Defraene *et al.*, 2012; Tucker *et al.*, 2013; Gulliford *et al.*, 2012). Different formulations of the LKB model can be found in literature incorporating specific risk factors via use of DMF. DMFs can stratify relevant variables, such as the tolerance dose for a 50% complication for (TD50) of the LKB model.

2.3.2.1 Statistical approaches: Data-driven modeling, also known as statistical modeling techniques, relies not only on mechanistic principles relating to disease manifestation but also on empirical combinations of relevant variables. In this context, the observed treatment outcome is considered as the result of functional mapping of several input variables through a statistical learning process that aims to estimate dependencies from data (El Naqa *et al.*, 2006a).

Among the most commonly used statistical methods for radiogenomics modeling are functions with sigmoidal shapes (S-curved) such as logistic regression, which also provide numerical stability. The results of this type of approach are expressed in the model parameters, which are chosen in a stepwise fashion to define the abscissa of the regression model. However, it is the user's responsibility to determine whether interaction terms or higher order variables should be added. Penalty techniques based on ridge (L2-norm) or Lasso (L1-norm) methods could aid in the process by eliminating least relevant variables and imposing sparsity conditions (Hastie *et al.*, 2015). An alternative solution to ameliorate this problem is offered by applying machine learning methods.

2.3.2.2 Machine learning: Machine learning techniques are a class of artificial intelligence (e.g., neural networks, support vector machines, deep neural networks, decision trees, random forests, etc.), which are able to emulate human intelligence by learning the surrounding environment from the given input data and can detect nonlinear complex patterns in such data. Based on the human-machine interaction, there are two common types of learning: supervised and unsupervised. Supervised learning is used when the endpoints of the treatments such as tumor control or toxicity grades are known; these endpoints are provided by experienced oncologists following institutional or National Cancer Institute (NCI) criteria and it is the most commonly used learning method in outcomes modeling. Nevertheless, unsupervised methods such as clustering methods or the use of principal component analysis (PCA) provide means to reduce the learning problem curse of dimensionality, feature extraction, and to aid in the visualization of multivariable data and the selection of the optimal learning method parameters for supervised learning methods (El Naqa *et al.*, 2015; Kang *et al.*, 2015).

In radiotherapy, neural networks were extensively investigated to model post-radiation treatment outcomes for cases of lung injury (Munley *et al.*, 1999; Su *et al.*, 2005) and biochemical failure and rectal bleeding in prostate cancer (Gulliford *et al.*, 2004; Tomatis *et al.*, 2012). A rather more robust approach of machine learning methods is support vector machines (SVMs), which are universal constructive learning procedures based on the statistical learning theory (Vapnik, 1998). For discrimination between patients who are at low risk versus high risk of treatment toxicity, the main idea of SVM would be to separate these two classes with 'hyper-planes' that maximize the margin between them in the nonlinear feature space defined by an implicit kernel mapping (El Naqa *et al.*, 2009; El Naqa *et al.*, 2010; El Naqa, 2012). Recently, neural networks have witnessed a recent incarnation, with the advent of deep learning methods such as convolutional neural networks with their ability to abstract data at different levels and result in improved predictions (Goodfellow *et al.*, 2016). However, these methods have been stigmatized in medical applications as black boxes, hindering their implementation in practical clinical contexts (Foster *et al.*, 2014).

In an effort to overcome the black box stigma of generic machine learning algorithms, approaches incorporating more engineering systems-like methods based on graphical techniques (e.g., decision trees and Bayesian networks [BNs]) have witnessed increased used in outcome modeling of cancer (Oh *et al.*, 2011a; Lee *et al.*, 2015; Jayasurya *et al.*, 2010; Luo *et al.*, 2017). In particular, a BN provides a probabilistic graphical representation of the relationships between the variables represented as nodes in a directed acyclic graph

(DAG), which encodes the presence and direction of relationship influence among the variables themselves and the clinical endpoint of interest. The relationship between parent and child nodes is modeled by conditional probabilities using Bayes chain rule. These methods are also robust for variable uncertainties and missing data, which would make them candidates for clinical applications (Koller and Friedman, 2009; Sinoquet and Mourad, 2014). Example application of these methods in radiogenomics modeling will be discussed below.

3 Examples of radiogenomics modeling

3.1 Prostate cancer

Building appropriate models is a critical step in the development of a robust predictive instrument. Simulation data can help provide a sense as to the number of genetic variants (SNPs/CNVs) that would be necessary to improve upon the performance of a NTCP model (Janssens *et al.*, 2006). For example, a simulation analysis provided evidence as to how well a predictive model might be capable of discriminating between those men at risk for a particular complication following radiotherapy from those unlikely to develop that form of radiation-induced morbidity for a range of assumptions about the genetic architecture of normal tissue toxicity (Kerns *et al.*, 2015). The results of simulation experiments suggest that: (1) increasing numbers of SNPs included in the risk model improves discrimination accuracy as measured by the area under the receiver-operating characteristic curve (AUC); (2) inclusion of SNPs with larger effect sizes and/or higher risk allele frequency improves the accuracy of the model; and (3) relatively high AUC values can be achieved with roughly 50–100 common risk SNPs with effect sizes in the range of 1.05–1.5. The results of these simulation experiments are encouraging since they suggest that a relatively small number of SNPs can serve as the basis of an assay that would markedly improve the ability to predict the risk of a particular patient to develop an adverse effect resulting from radiotherapy. An example of a logistic regression model for rectal bleeding using Lasso is shown in Figure 3 (Kerns *et al.*, 2015). Recently, high-throughput gene expression and clinical data were used to develop and validate a 24-gene expression signature that predicts response to post-prostatectomy radiotherapy (PORTOS) in matched training and validation cohorts of patients with prostate cancer. The study showed that patients with high PORTOS had a lower incidence of distant metastasis than that in patients with low scores (Zhao *et al.*, 2016).

3.2 Breast cancer

Many genetic studies of radiosensitivity were initiated in breast cancer. As noted earlier, an 18-gene panel was developed for subcutaneous fibrotic reactions following post-mastectomy radiotherapy (Rodningen *et al.*, 2008). Another set of 67 genes differentiating severe radiosensitive and normal reacting patients was identified (Mayer *et al.*, 2011). A panel of 147 gene signature as was more recently correlated with radiosensitivity using random forest machine learning. The signature was further refined to 51 genes that were enriched for concepts involving cell-cycle arrest and DNA damage response. It was validated in an independent dataset and shown to be the most significant factor in predicting local recurrence on multivariate analysis outperforming all clinically used clinicopathologic features as shown in Figure 4 (Speers *et al.*, 2015).

3.3 Lung cancer

It has been noted in the literature that pathway modeling of microscopic molecular interactions can be represented using systems biology approaches, e.g., developing graphs of network connections as in electrical power lines grids (Alon, 2007). In the case of radiation therapy, the radiogenomics interaction can be represented as a graph (network) where the nodes represent genes or proteins and the edges may represent probabilistic similarities or interactions between these nodes using approaches such as Bayesian networks (BNs), which have been applied successfully for radiogenomics modeling. In one example, data derived from dosimetric radiation metrics (physical variables) and blood-based cytokines (biomarkers) were used to predict local control in non-small cell lung cancer (NSCLC) (Oh *et al.*, 2011b). The model demonstrated better prediction of local control by combining physical and biological variables compared with using either category alone. Moreover, by employing this network type approach with other bioinformatics tools, robust biomarkers for radiation-induced lung inflammation, radiation pneumonitis (RP), in patients with lung cancer could be identified from large-scale proteomic studies. These findings were further validated in an independent dataset using an enzyme-linked immunosorbent assay (ELISA) (Oh *et al.*, 2011c). Biomarkers identified from proteomics analysis along with other candidate cytokines and dosimetric variables (i.e., mean lung dose, mean heart dose, V20, etc.) obtained from published reports were used for constructing a BN for RP, which provided better prediction than known individual biomarkers or a combination of variables using conventional regression modeling counterparts (Lee *et al.*, 2015). In a recent study of NSCLC patients following radiotherapy, BN models of RP were generated that used genetic markers including SNP and miRNA in addition to clinical, dosimetric, and cytokine variables before and during the course of the radiotherapy (Luo *et al.*, 2017) as shown in Figure 5. It was noted that the performance of the network improved by incorporating during treatment cytokine changes achieving an area under the receiver -operating characteristic (ROC) curve of 0.87.

4 Challenges, Issues, Controversies

4.1 Pan- vs. p-OMICs

Due to advances in biotechnology, and complex biologic analysis the amount of potential radiotherapy data available for association with outcomes has grown exponentially exponential in the past decade. At the same time cancer incidences have generally plateaued. The fact that the number of patient-specific (clinical, physical and biological) variables (p) $\gg$ the number of patient samples (n) are a challenge for class inference methods of statistical learning. This p -omics phenomenon (too many variables to few samples) can result in statistical modeling panic and may yield undesirable spurious correlations, reversal paradoxes, or misleading ghost analytics. There are several methods to mitigate the effects of such problems, from basic data reduction or feature selection methods (Kantardzic, 2011) to more advanced approaches based on incorporating prior knowledge, utilization of information theory or machine learning ensembles (El Naqa, 2016).

4.2 Correlative versus functional studies

Biomarker studies have been for years the subject of intense scrutiny, primarily due to inconsistent conclusions and lack of clarity in population characteristics and applied methodologies among different studies. Therefore, guidelines on REporting recommendations for tumor MARKer prognostic studies (REMARK) have been proposed in a joint effort between National Cancer Institute and the European Organization for Research and Treatment of Cancer (NCI–EORTC) to enhance transparency and reproducibility with a checklist of about 20 items that include hypothesis statement, description of patients, specimens, assays, statistics, data and analysis presentation to discussion of limitations and clinical implications (McShane *et al.*, 2005). This list was later updated with examples of good reporting (Altman *et al.*, 2012). A further validation would involve testing using functional biological assays of the presented hypothesis of the identified biomarker (Alinezhad *et al.*, 2016). For genetic risks, a checklist of 25 items recommended for strengthening the reporting of Genetic RIsK Prediction Studies (GRIPS) was further developed particularly to improve the knowledge synthesis and application of information from multiple genetic studies that might differ in design, conduct or analysis (Janssens *et al.*, 2011). Following suite, the Radiogenomics Consortium produced an 18-item checklist for STrengthening the Reporting Of Genetic Association studies in Radiogenomics (STROGAR), which aims at helping researchers improve their design and reporting of new radiogenomics studies, interpret published research, and facilitate the discovery of SNPs that are genuinely associated with radiotherapy toxicity (Kerns *et al.*, 2014a).

4.3 Patient characteristics and confounding factors

Typically overlooked are confounding effects that exist and should be considered in radiogenomics investigations. Examples of these include ancestry and population sub-structure as a common confounding factor in GWAS, or clinical factors and radiation dose/volume parameters as potential effect modifiers or the role for gene-by-environment association studies. Some of these issues have been a major subject of the REMARK/GRIPS/STROGAR guidelines discussed above.

4.3.1 Clinical factors—Clinical data in radiotherapy typically refers to cancer diagnostic information (site, histology, stage, grade, etc.) and patient-related characteristics (age, gender, co-morbidities, etc.). In some instances, other treatment modalities information (surgery, chemotherapy, hormonal treatment, etc.) would be also classified under this category. For instance, when chemotherapy and radiotherapy are combined, they can act synergistically to improve tumor cell and patient survival (Belani *et al.*, 2005; De Ruyscher *et al.*, 2012) or when more recently combining radiotherapy with immunotherapeutic agents to limit distant failures (Zeng *et al.*, 2014; Rekers *et al.*, 2014; Demaria *et al.*, 2015). However, little progress has been achieved when combining molecularly targeted drugs with radiotherapy (Sharma *et al.*, 2016) and radiogenomics may facilitate such needed advancement of drug–radiotherapy combinations. The mining of such clinical data could be challenging if the data is embedded in physician notes (i.e., unstructured), however, there are good opportunities for natural language processing (NLP) techniques to assist in the organization of such data for the purposes of radiogenomics analytics (Shivade *et al.*, 2013).

4.3.2 Dose-volume metrics—This type of data is related to the treatment planning process in radiotherapy, which involves radiation dose simulations using computed tomography imaging; specifically dose-volume metrics derived from dose-volume histograms (DVHs) graphs (Ten Haken and Kessler, 2001). Dose-volume metrics have been extensively studied in the radiation oncology literature for outcomes modelling (Blanco *et al.*, 2005; Bradley *et al.*, 2004; Hope *et al.*, 2006; Hope *et al.*, 2005; Levegrun *et al.*, 2001; Marks, 2002; Ten Haken *et al.*, 1993; Bentzen *et al.*, 2010a). These metrics are extracted from the DVH such as volume receiving certain dose (V_x), minimum dose to x% volume (D_x), mean, maximum and minimum dose, etc. (Deasy and El Naqa, 2008). Moreover, a dedicated software tool called ‘DREES’ was developed for deriving these DVH metrics and modeling of radiotherapy response (El Naqa *et al.*, 2006b) that was applied to radiogenomics modeling (Coates *et al.*, 2015).

4.4 Fitting issues and validation

The objective of any prospective outcomes classifier or model is to accurately make predictions based off previously acquired data. Radiogenomics modeling has transitioned from simple TCP/NTCP models with few parameters into machine learning models with massive number of parameters and equally high risk of overfitting the data and not generating to out-of-sample data. For such models to be of use they must be robust enough to account for heterogeneity across large patient cohorts. From a practical point of view, they must also be able to be trained on smaller subsets of patients. The gold standard for validation remains independent evaluation of model performance on an unseen cohort; however, independent evaluation is not always practical to rely on in many cases due to the lack of availability of such data until close to clinical implementation and is thus of little use in the early stages of model building.

Once a model is prepared for validation, it is imperative that the data being used to test the performance is indicative of the factors used to construct the model. Cases whereby a model may be expected to perform differently include changes in treatment modality (e.g., 3D-Conformal Radiotherapy [3D-CRT] versus Volumetric Arc Therapy [VMAT]), and should thus be considered prior to testing in addition to other confounding factors that may differ between the two study populations. as mentioned in Section 4.3.

Two primary challenges exist when it comes to constructing robust models and attempting to apply them to a wider population: *over-fitting*, whereby a model is not robust enough to apply to a wider population, and *under-fitting*, whereby the true signal is mistaken for noise and is being fit instead. In addition to traditional multiple comparison corrections such as Bonferroni adjustments or permutation testing (Johnson *et al.*, 2010), more advanced statistical techniques can be used to determine if the model over-fit or under-fit a given signal; methods to accomplish such tasks include statistical resampling (cross-validation or bootstrapping) or information theory approaches (Coates *et al.*, 2016).

The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) recommendation developed a set of 22 items checklist for the reporting of studies developing, validating, or updating a prediction model such as radiogenomics models (Collins *et al.*, 2015). The checklist includes items similar to

REMARK/GRIPS/STROGAR, in addition to issues related to validation and performance reporting. For validation, the TRIPOD recommendations emphasize the importance of internal validation as a necessary part of model development. For model evaluation, it suggests reporting performance measures (with confidence intervals) of the proposed prediction model.

4.5 Integration into clinical practice

The incorporation of radiogenomics into radiotherapy trials would allow for the identification of patients who are at risk of toxicity and escalation of dose or combined drug-radiation therapy in patients that are likely to fail. It has been long thought that strategically planned biomarker evaluations in phase II studies will allow for better screening of therapeutics and improved phase III trial results (McShane *et al.*, 2009; Freidlin *et al.*, 2012). To date, few studies have employed radiogenomics as part of their decision-making strategy. Mehta *et al.* designed a window-of-opportunity study for early assessment of antiangiogenic therapy in breast cancer using a combination of magnetic resonance imaging scans and core biopsies for exon array gene expression analysis (Mehta *et al.*, 2011). A new emerging concept to accelerate biomarker application in clinical trials is called *Basket trials*, which is predicated on the hypothesis that the presence of a biomarker that can predict response to a targeted therapy (or combined drug-radiation) independent of tumor histology (Redig and Jänne, 2015). This approach was applied in a lung cancer study with five targeted therapies in patients grouped by molecular markers along with their tumor histology (Lopez-Chavez *et al.*, 2015). Alternatively, one can envision a systematic decision-tree or a Bayesian network approach (cf. Figure 5) that would combine these radiogenomics biomarkers with patient characteristics to tailor drug-radiation combination to individual patients based on their predicted benefit/risk tradeoffs.

5 Recommendations

As noted earlier, transparency and reproducibility are essential for the success of radiogenomics as a predictive tool and potentially as a guide for designing personalized or adaptive clinical trials in radiotherapy. The REMARK/GRIPS/STROGAR/TRIPOD guidelines collectively contribute towards achieving these goals. In addition, the confounding effects of dosimetric and clinical variables need to be included in the dataset and accounted for in the analysis to avoid pitfalls of heterogeneous populations. It is also important to notice that successful application of advanced radiogenomics methodologies such as complex system analysis or machine-learning methods needs to account for the intricacies of existing radiobiology knowledge to be successful and not attempt to re-invent the wheel. Radiogenomics modeling is not about building a black box but rather a tool to help improve our understanding of radiobiological mechanisms, generate new hypotheses and support better personalized clinical design and decision-making. Another aspect that is necessary to achieve success in this area relates to the sharing of data whether through centralized or federated databases, which is a necessity to successfully identify and validate trends from smaller institutional cohorts. Data-pooling from multiple institutions may lead to privacy concerns that are unique at the local and international levels particularly when dealing with genetic data. Such concerns can greatly hinder the development of prospective

models and the evaluation thereof unless accounted for using approaches such as rapid (distributed) learning (Lambin *et al.*, 2013a). If structured data-pooling protocols can be agreed upon then radiogenomics-based models are likely to flourish more due to larger databases and easier independent evaluation.

6 Conclusions

The heterogeneity of cancer is well reflected in the diverse array of diagnostic and therapeutic options employed in oncology today. Because the complexity of the underlying interactions between biological tissue and ionizing radiation, it is likely that more advanced modeling techniques will become necessary for radiotherapy response predictions to better understand the underlying trends across populations. Novel treatment-related analytics coupled with breakthroughs in radiogenomics modeling have the capacity to inform physicians and patients on the benefits/risks associated with individual treatments, thereby contributing directly to improved outcomes. However, this will require a multidisciplinary approach involving all stake holders.

Acknowledgments

RTH/EN/YL would like to acknowledge support by the National Institutes of Health grant number P01 CA059827. Dr. Kerns is supported by K07CA187546. Grants and contracts to BSR 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). CMLW is supported by Cancer Research UK. The authors would like to acknowledge Mr. Steven Kronenberg for help generating Figure 1.

8 Bibliography

- Abazeed ME, Adams DJ, Hurov KE, Tamayo P, Creighton CJ, Sonkin D, Giacomelli AO, Du C, Fries DF, Wong KK, Mesirov JP, Loeffler JS, Schreiber SL, Hammerman PS, Meyerson M. Integrative radiogenomic profiling of squamous cell lung cancer. *Cancer research*. 2013; 73:6289–98. [PubMed: 23980093]
- Adamus-Gorka M, Mavroidis P, Lind BK, Brahme A. Comparison of dose response models for predicting normal tissue complications from cancer radiotherapy: application in rat spinal cord. *Cancers (Basel)*. 2011; 3:2421–43. [PubMed: 24212817]
- Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Cavalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature communications*. 2014; 5:4006.
- Ahmed KA, Chinnaiyan P, Fulp WJ, Eschrich S, Torres-Roca JF, Caudell JJ. The radiosensitivity index predicts for overall survival in glioblastoma. *Oncotarget*. 2015; 6:34414–22. [PubMed: 26451615]
- Alapetite C, Thirion P, de la Rochefordiere A, Cosset JM, Moustacchi E. Analysis by alkaline comet assay of cancer patients with severe reactions to radiotherapy: defective rejoining of radioinduced DNA strand breaks in lymphocytes of breast cancer patients. *Int J Cancer*. 1999; 83:83–90. [PubMed: 10449613]
- Alinezhad S, Väänänen R-M, Mattsson J, Li Y, Tallgrén T, Tong Ochoa N, Bjartell A, Åkerfelt M, Taimen P, Boström PJ, Pettersson K, Nees M. Validation of Novel Biomarkers for Prostate Cancer Progression by the Combination of Bioinformatics, Clinical and Functional Studies. *PloS one*. 2016; 11:e0155901. [PubMed: 27196083]
- Alon, U. An introduction to systems biology : design principles of biological circuits. Boca Raton, FL: Chapman & Hall/CRC; 2007.
- Alsner J, Andreassen CN, Overgaard J. Genetic markers for prediction of normal tissue toxicity after radiotherapy. *Semin Radiat Oncol*. 2008; 18:126–35. [PubMed: 18314067]

- Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *BMC Medicine*. 2012; 10:51. [PubMed: 22642691]
- Amundson SA, Do KT, Vinikoor LC, Lee RA, Koch-Paiz CA, Ahn J, Reimers M, Chen Y, Scudiero DA, Weinstein JN, Trent JM, Bittner ML, Meltzer PS, Fornace AJ Jr. Integrating global gene expression and radiation survival parameters across the 60 cell lines of the National Cancer Institute Anticancer Drug Screen. *Cancer research*. 2008; 68:415–24. [PubMed: 18199535]
- Andreassen CN, Barnett GC, Langendijk JA, Alsner J, De Ruyscher D, Krause M, Bentzen SM, Haviland JS, Griffin C, Poortmans P, Yarnold JR. Conducting radiogenomic research--do not forget careful consideration of the clinical data. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*. 2012; 105:337–40. [PubMed: 23245646]
- Andreassen CN, Rosenstein BS, Kerns SL, Ostrer H, De Ruyscher D, Cesaretti JA, Barnett GC, Dunning AM, Dorling L, West CM, Burnet NG, Elliott R, Coles C, Hall E, Fachal L, Vega A, Gomez-Caamano A, Talbot CJ, Symonds RP, De Ruyck K, Thierens H, Ost P, Chang-Claude J, Seibold P, Popanda O, Overgaard M, Dearnaley D, Sydes MR, Azria D, Koch CA, Parliament M, Blackshaw M, Sia M, Fuentes-Raspall MJ, Ramon YCT, Barnadas A, Vesprini D, Gutierrez-Enriquez S, Molla M, Diez O, Yarnold JR, Overgaard J, Bentzen SM, Alsner J. International Radiogenomics C. Individual patient data meta-analysis shows a significant association between the ATM rs1801516 SNP and toxicity after radiotherapy in 5456 breast and prostate cancer patients. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*. 2016; 121:431–9. [PubMed: 27443449]
- Azria D, Riou O, Castan F, Nguyen TD, Peignaux K, Lemanski C, Lagrange JL, Kirova Y, Lartigau E, Belkacemi Y, Bourgier C, Rivera S, Noel G, Clippe S, Mornex F, Hennequin C, Kramar A, Gourgou S, Pelegri A, Fenoglietto P, Ozsahin EM. Radiation-induced CD8 T-lymphocyte Apoptosis as a Predictor of Breast Fibrosis After Radiotherapy: Results of the Prospective Multicenter French Trial. *EBioMedicine*. 2015; 2:1965–73. [PubMed: 26844275]
- Ballman KV. Biomarker: Predictive or Prognostic? *Journal of Clinical Oncology*. 2015; 33:3968–71. [PubMed: 26392104]
- Barnett GC, Coles CE, Elliott RM, Baynes C, Luccarini C, Conroy D, Wilkinson JS, Tyrer J, Misra V, Platte R, Gulliford SL, Sydes MR, Hall E, Bentzen SM, Dearnaley DP, Burnet NG, Pharoah PD, Dunning AM, West CM. Independent validation of genes and polymorphisms reported to be associated with radiation toxicity: a prospective analysis study. *Lancet Oncol*. 2012; 13:65–77. [PubMed: 22169268]
- Barnett GC, Thompson D, Fachal L, Kerns S, Talbot C, Elliott RM, Dorling L, Coles CE, Dearnaley DP, Rosenstein BS, Vega A, Symonds P, Yarnold J, Baynes C, Michailidou K, Dennis J, Tyrer JP, Wilkinson JS, Gomez-Caamano A, Tanteles GA, Platte R, Mayes R, Conroy D, Maranian M, Luccarini C, Gulliford SL, Sydes MR, Hall E, Haviland J, Misra V, Tittley J, Bentzen SM, Pharoah PD, Burnet NG, Dunning AM, West CM. A genome wide association study (GWAS) providing evidence of an association between common genetic variants and late radiotherapy toxicity. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*. 2014; 111:178–85. [PubMed: 24785509]
- Begg AC, Russell NS, Knaken H, Lebesque JV. Lack of Correlation of Human Fibroblast Radiosensitivity in Vitro with Early Skin Reactions in Patients Undergoing Radiotherapy. *International Journal of Radiation Biology*. 1993; 64:393–405. [PubMed: 7901301]
- Begg AC, Stewart FA, Vens C. Strategies to improve radiotherapy with targeted drugs. *Nat Rev Cancer*. 2011; 11:239–53. [PubMed: 21430696]
- Belani CP, Choy H, Bonomi P, Scott C, Travis P, Haluschak J, WJC. Combined Chemoradiotherapy Regimens of Paclitaxel and Carboplatin for Locally Advanced Non-Small-Cell Lung Cancer: A Randomized Phase II Locally Advanced Multi-Modality Protocol. *Journal of Clinical Oncology*. 2005; 23:5883–91. [PubMed: 16087941]
- Benedict SH, El Naqa I, Klein EE. Introduction to Big Data in Radiation Oncology: Exploring Opportunities for Research, Quality Assessment, and Clinical Care. *International journal of radiation oncology, biology, physics*. 2016; 95:871–2.

- Bentzen SM. Preventing or reducing late side effects of radiation therapy: radiobiology meets molecular pathology. *Nat Rev Cancer*. 2006; 6:702–13. [PubMed: 16929324]
- Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, Haken RKT, Yorke ED. Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): An Introduction to the Scientific Issues. *International journal of radiation oncology, biology, physics*. 2010a; 76:S3–S9.
- Bentzen SM, Parliament M, Deasy JO, Dicker A, Curran WJ, Williams JP, Rosenstein BS. QUANTEC: Vision Paper. *International journal of radiation oncology, biology, physics*. 2010b; 76:S145–S50.
- Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, Gerstein M, Snyder M. Global Identification of Human Transcribed Sequences with Genome Tiling Arrays. *Science*. 2004; 306:2242–6. [PubMed: 15539566]
- Blanco AI, Chao KS, El Naqa I, Franklin GE, Zakarian K, Vicic M, Deasy JO. Dose-volume modeling of salivary function in patients with head-and-neck cancer receiving radiotherapy. *Int J Radiat Oncol Biol Phys*. 2005; 62:1055–69. [PubMed: 15990009]
- Bradley J, Deasy JO, Bentzen S, El-Naqa I. Dosimetric correlates for acute esophagitis in patients treated with radiotherapy for lung carcinoma. *Int J Radiat Oncol Biol Phys*. 2004; 58:1106–13. [PubMed: 15001251]
- Brahimi-Horn MC, Chiche J, Pouyssegur J. Hypoxia and cancer. *J Mol Med (Berl)*. 2007; 85:1301–7. [PubMed: 18026916]
- Buffa FM, Camps C, Winchester L, Snell CE, Gee HE, Sheldon H, Taylor M, Harris AL, Ragoussis J. microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. *Cancer research*. 2011; 71:5635–45. [PubMed: 21737487]
- Buffa FM, Davidson SE, Hunter RD, Nahum AE, West CML. Incorporating biologic measurements (SF2, CFE) into a tumor control probability model increases their prognostic significance: a study in cervical carcinoma treated with radiation therapy. *International Journal of Radiation Oncology*Biology*Physics*. 2001; 50:1113–22.
- Burnet NG, Nyman J, Turesson I, Wurm R, Yarnold JR, Peacock JH. Prediction of normal-tissue tolerance to radiotherapy from in-vitro cellular radiation sensitivity. *Lancet*. 1992; 339:1570–1. [PubMed: 1351551]
- Cai J, Guan H, Fang L, Yang Y, Zhu X, Yuan J, Wu J, Li M. MicroRNA-374a activates Wnt/beta-catenin signaling to promote breast cancer metastasis. *J Clin Invest*. 2013; 123:566–79. [PubMed: 23321667]
- Choudhury A, Nelson LD, Teo MT, Chilka S, Bhattarai S, Johnston CF, Elliott F, Lowery J, Taylor CF, Churchman M, Bentley J, Knowles MA, Harnden P, Bristow RG, Bishop DT, Kiltie AE. MRE11 expression is predictive of cause-specific survival following radical radiotherapy for muscle-invasive bladder cancer. *Cancer research*. 2010; 70:7017–26. [PubMed: 20843819]
- Coates J, Jeyaseelan AK, Ybarra N, David M, Faria S, Souhami L, Cury F, Duclos M, El Naqa I. Contrasting analytical and data-driven frameworks for radiogenomic modeling of normal tissue toxicities in prostate cancer. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*. 2015
- Coates J, Souhami L, El Naqa I. Big Data Analytics for Prostate Radiotherapy. *Frontiers in Oncology*. 2016; 6:149. [PubMed: 27379211]
- Collins GS, Reitsma JB, Altman DG, Moons KM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): The tripod statement. *Annals of Internal Medicine*. 2015; 162:55–63. [PubMed: 25560714]
- Crick F. Central dogma of molecular biology. *Nature*. 1970; 227:561–3. [PubMed: 4913914]
- Crompton NEA, Ozsahin M. A Versatile and Rapid Assay of Radiosensitivity of Peripheral Blood Leukocytes Based on DNA and Surface-Marker Assessment of Cytotoxicity. *Radiation Research*. 1997; 147:55–60. [PubMed: 8989370]
- de Cid R, Riveira-Munoz E, Zeeuwen PLJM, Robarge J, Liao W, Dannhauser EN, Giardina E, Stuart PE, Nair R, Helms C, Escaramis G, Ballana E, Martin-Ezquerro G, Heijer Md, Kamsteeg M, Joosten I, Eichler EE, Lazaro C, Pujol RM, Armengol L, Abecasis G, Elder JT, Novelli G, Armour JAL, Kwok P-Y, Bowcock A, Schalkwijk J, Estivill X. Deletion of the late cornified envelope

LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat Genet.* 2009; 41:211–5. [PubMed: 19169253]

- De Ruyscher D, van Baardwijk A, Steevens J, Botterweck A, Bosmans G, Reymen B, Wanders R, Borger J, Dingemans AM, Bootsma G, Pitz C, Lunde R, Geraedts W, Oellers M, Dekker A, Lambin P. Individualised isotoxic accelerated radiotherapy and chemotherapy are associated with improved long-term survival of patients with stage III NSCLC: a prospective population-based study. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2012; 102:228–33. [PubMed: 22100659]
- Deacon J, Peckham MJ, Steel GG. The radioresponsiveness of human tumours and the initial slope of the cell survival curve. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 1984; 2:317–23. [PubMed: 6395213]
- Deasy JO, Bentzen SM, Jackson A, Ten Haken RK, Yorke ED, Constine LS, Sharma A, Marks LB. Improving Normal Tissue Complication Probability Models: The Need to Adopt a “Data-Pooling” Culture. *International Journal of Radiation Oncology • Biology • Physics.* 2010; 76:S151–S4.
- Deasy JO, El Naqa I. Image-based modeling of normal tissue complication probability for radiation therapy. *Cancer Treat Res.* 2008; 139:215–56. [PubMed: 18236719]
- Defraene G, Van den Bergh L, Al-Mamgani A, Haustermans K, Heemsbergen W, Van den Heuvel F, Lebesque JV. The Benefits of Including Clinical Factors in Rectal Normal Tissue Complication Probability Modeling After Radiotherapy for Prostate Cancer. *International Journal of Radiation Oncology*Biography*Physics.* 2012; 82:1233–42.
- Demaria S, Golden EB, Formenti SC. Role of Local Radiation Therapy in Cancer Immunotherapy. *JAMA Oncol.* 2015; 1:1325–32. [PubMed: 26270858]
- Dunne AL, Price ME, Mothersill C, McKeown SR, Robson T, Hirst DG. Relationship between clonogenic radiosensitivity, radiation-induced apoptosis and DNA damage/repair in human colon cancer cells. *Br J Cancer.* 2003; 89:2277–83. [PubMed: 14676806]
- El Naqa I. Machine learning methods for predicting tumor response in lung cancer. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.* 2012; 2:173–81.
- El Naqa, I. Informatics in Radiation Oncology. Starkschall, G., Siochi, C., editors. Boca Raton, FL: CRC Press, Taylor and Francis; 2013. p. 257-75.
- El Naqa I. Biomedical informatics and panomics for evidence-based radiation therapy. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.* 2014a; 4:327–40.
- El Naqa I. The role of quantitative PET in predicting cancer treatment outcomes. *Clin Transl Imaging.* 2014b; 2:305–20.
- El Naqa I. Perspectives on making big data analytics work for oncology. *Methods.* 2016; 111:32–44. [PubMed: 27586524]
- El Naqa I, Bradley JD, Lindsay PE, Blanco AI, Vicic M, Hope AJ, Deasy JO. Multi-variable modeling of radiotherapy outcomes including dose-volume and clinical factors. *Int J Radiat Oncol Biol Phys.* 2006a; 64:1275–86. [PubMed: 16504765]
- El Naqa I, Bradley JD, Lindsay PE, Hope AJ, Deasy JO. Predicting radiotherapy outcomes using statistical learning techniques. *Phys Med Biol.* 2009; 54:S9–S30. [PubMed: 19687564]
- El Naqa I, Deasy JO, Mu Y, Huang E, Hope AJ, Lindsay PE, Apte A, Alaly J, Bradley JD. Datamining approaches for modeling tumor control probability. *Acta Oncol.* 2010; 49:1363–73. [PubMed: 20192878]
- El Naqa, I, Li, R., Murphy, MJ., editors. *Machine Learning in Radiation Oncology: Theory and Application.* Switzerland: Springer International Publishing; 2015.
- El Naqa I, Pater P, Seuntjens J. Monte Carlo role in radiobiological modelling of radiotherapy outcomes. *Phys Med Biol.* 2012; 57:R75–97.
- El Naqa I, Suneja G, Lindsay PE, Hope AJ, Alaly JR, Vicic M, Bradley JD, Apte A, Deasy JO. Dose response explorer: an integrated open-source tool for exploring and modelling radiotherapy dose-volume outcome relationships. *Phys Med Biol.* 2006b; 51:5719–35. [PubMed: 17068361]
- Elmore S. Apoptosis: a review of programmed cell death. *Toxicol Pathol.* 2007; 35:495–516. [PubMed: 17562483]
- Eriksson D, Stigbrand T. Radiation-induced cell death mechanisms. *Tumour Biol.* 2010; 31:363–72. [PubMed: 20490962]

- Eschrich SA, Fulp WJ, Pawitan Y, Foekens JA, Smid M, Martens JW, Echevarria M, Kamath V, Lee JH, Harris EE, Bergh J, Torres-Roca JF. Validation of a radiosensitivity molecular signature in breast cancer. *Clin Cancer Res.* 2012; 18:5134–43. [PubMed: 22832933]
- Eschrich SA, Pramana J, Zhang H, Zhao H, Boulware D, Lee JH, Bloom G, Rocha-Lima C, Kelley S, Calvin DP, Yeatman TJ, Begg AC, Torres-Roca JF. A gene expression model of intrinsic tumor radiosensitivity: prediction of response and prognosis after chemoradiation. *International journal of radiation oncology, biology, physics.* 2009; 75:489–96.
- Eustace A, Mani N, Span PN, Irlam JJ, Taylor J, Betts GN, Denley H, Miller CJ, Homer JJ, Rojas AM, Hoskin PJ, Buffa FM, Harris AL, Kaanders JH, West CM. A 26-gene hypoxia signature predicts benefit from hypoxia-modifying therapy in laryngeal cancer but not bladder cancer. *Clin Cancer Res.* 2013; 19:4879–88. [PubMed: 23820108]
- Fachal L, Gomez-Caamano A, Barnett GC, Peleteiro P, Carballo AM, Calvo-Crespo P, Kerns SL, Sanchez-Garcia M, Lobato-Busto R, Dorling L, Elliott RM, Dearnaley DP, Sydes MR, Hall E, Burnet NG, Carracedo A, Rosenstein BS, West CM, Dunning AM, Vega A. A three-stage genome-wide association study identifies a susceptibility locus for late radiotherapy toxicity at 2q24.1. *Nat Genet.* 2014; 46:891–4. [PubMed: 24974847]
- Fertil B, Malaise EP. Inherent cellular radiosensitivity as a basic concept for human tumor radiotherapy. *International journal of radiation oncology, biology, physics.* 1981; 7:621–9.
- Fertil B, Malaise EP. Intrinsic radiosensitivity of human cell lines is correlated with radioresponsiveness of human tumors: analysis of 101 published survival curves. *International journal of radiation oncology, biology, physics.* 1985; 11:1699–707.
- Foster KR, Koprowski R, Skufca JD. Machine learning, medical diagnosis, and biomedical engineering research - commentary. *BioMedical Engineering OnLine.* 2014; 13:94. [PubMed: 24998888]
- Fraser M, Berlin A, Bristow RG, van der Kwast T. Genomic, pathological, and clinical heterogeneity as drivers of personalized medicine in prostate cancer. *Urol Oncol.* 2015; 33:85–94. [PubMed: 24768356]
- Freidlin B, McShane LM, Polley MY, Korn EL. Randomized phase II trial designs with biomarkers. *J Clin Oncol.* 2012; 30:3304–9. [PubMed: 22869885]
- Friedland W, Dingfelder M, Kundrat P, Jacob P. Track structures, DNA targets and radiation effects in the biophysical Monte Carlo simulation code PARTRAC. *Mutation research.* 2011; 711:28–40. [PubMed: 21281649]
- Goodfellow, I., Bengio, Y., Courville, A. *Deep learning.* Cambridge, Massachusetts: The MIT Press; 2016.
- Group BDW. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther.* 2001; 69:89–95. [PubMed: 11240971]
- Guerra JL, Gomez D, Wei Q, Liu Z, Wang LE, Yuan X, Zhuang Y, Komaki R, Liao Z. Association between single nucleotide polymorphisms of the transforming growth factor beta1 gene and the risk of severe radiation esophagitis in patients with lung cancer. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2012; 105:299–304. [PubMed: 23022171]
- Gulliford SL, Partridge M, Sydes MR, Webb S, Evans PM, Dearnaley DP. Parameters for the Lyman Kutcher Burman (LKB) model of Normal Tissue Complication Probability (NTCP) for specific rectal complications observed in clinical practise. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology.* 2012; 102:347–51. [PubMed: 22119373]
- Gulliford SL, Webb S, Rowbottom CG, Corne DW, Dearnaley DP. Use of artificial neural networks to predict biological outcomes for patients receiving radical radiotherapy of the prostate. *Radiotherapy and Oncology.* 2004; 71:3–12. [PubMed: 15066290]
- Hastie, T., Tibshirani, R., Wainwright, M. *Statistical learning with sparsity : the lasso and generalizations.* Boca Raton: CRC Press, Taylor & Francis Group; 2015.
- Hope AJ, Lindsay PE, El Naqa I, Alaly JR, Vicic M, Bradley JD, Deasy JO. Modeling radiation pneumonitis risk with clinical, dosimetric, and spatial parameters. *Int J Radiat Oncol Biol Phys.* 2006; 65:112–24. [PubMed: 16618575]

- Hope, AJ., Lindsay, PE., El Naqa, I., Bradley, JD., Vicic, M., Deasy, JO. ASTRO 47TH ANNUAL MEETING; Denver, CO. 2005. p. S231vol. Series 63)
- Horgan RP, Kenny LC. 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist*. 2011; 13:189–95.
- Huang X, Le QT, Giaccia AJ. MiR-210--micromanager of the hypoxia pathway. *Trends Mol Med*. 2010; 16:230–7. [PubMed: 20434954]
- Incerti S, Douglass M, Penfold S, Guatelli S, Bezak E. Review of Geant4-DNA applications for micro and nanoscale simulations. *Physica medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics*. 2016; 32:1187–200.
- Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM. Predictive testing for complex diseases using multiple genes: fact or fiction? *Genet Med*. 2006; 8:395–400. [PubMed: 16845271]
- Janssens ACJW, Ioannidis JPA, van Duijn CM, Little J, Khoury MJ. for the GG. Strengthening the reporting of genetic risk prediction studies: the GRIPS statement. *European Journal of Clinical Investigation*. 2011; 41:1004–9. [PubMed: 21434891]
- Jayasurya K, Fung G, Yu S, Dehing-Oberije C, De Ruyscher D, Hope A, De Neve W, Lievens Y, Lambin P, Dekker AL. Comparison of Bayesian network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy. *Medical physics*. 2010; 37:1401–7. [PubMed: 20443461]
- Jeraj R, Cao Y, Ten Haken RK, Hahn C, Marks L. Imaging for Assessment of Radiation-Induced Normal Tissue Effects. *International Journal of Radiation Oncology*Biophysics*. 2010; 76:S140–S4.
- Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA, O'Brien SJ. Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics*. 2010; 11:724. [PubMed: 21176216]
- Jones LA, Scott D, Cowan R, Roberts SA. Abnormal radiosensitivity of lymphocytes from breast cancer patients with excessive normal tissue damage after radiotherapy: chromosome aberrations after low dose-rate irradiation. *Int J Radiat Biol*. 1995; 67:519–28. [PubMed: 7775827]
- Kang J, Schwartz R, Flickinger J, Beriwal S. Machine Learning Approaches for Predicting Radiation Therapy Outcomes: A Clinician's Perspective. *International Journal of Radiation Oncology*Biophysics*. 2015; 93:1127–35.
- Kantardzic, M. *Data Mining*. John Wiley & Sons, Inc; 2011. p. 53-86.
- Kerns SL, de Ruyscher D, Andreassen CN, Azria D, Barnett GC, Chang-Claude J, Davidson S, Deasy JO, Dunning AM, Ostrer H, Rosenstein BS, West CM, Bentzen SM. STROGAR - STrengthening the Reporting Of Genetic Association studies in Radiogenomics. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*. 2014a; 110:182–8. [PubMed: 23993398]
- Kerns SL, Dorling L, Fachal L, Bentzen S, Pharoah PD, Barnes DR, Gomez-Caamano A, Carballo AM, Dearnaley DP, Peleteiro P, Gulliford SL, Hall E, Michailidou K, Carracedo A, 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 C. Meta-analysis of Genome Wide Association Studies Identifies Genetic Markers of Late Toxicity Following Radiotherapy for Prostate Cancer. *EBioMedicine*. 2016; 10:150–63. [PubMed: 27515689]
- Kerns SL, Kundu S, Oh JH, Singhal SK, Janelins M, Travis LB, Deasy JO, Janssens AC, Ostrer H, Parliament M, Usmani N, Rosenstein BS. The Prediction of Radiotherapy Toxicity Using Single Nucleotide Polymorphism-Based Models: A Step Toward Prevention. *Semin Radiat Oncol*. 2015; 25:281–91. [PubMed: 26384276]
- Kerns SL, Stock R, Stone N, Buckstein M, Shao Y, Campbell C, Rath L, DeRuyscher D, Lammering G, Hixson R, Cesaretti J, Terk M, Ostrer H, Rosenstein BS. A Two-Stage Genome-Wide Association Study to Identify Single Nucleotide Polymorphisms Associated with Development of Erectile Dysfunction Following Radiotherapy for Prostate Cancer. *International journal of radiation oncology, biology, physics*. 2013; 85:e21–e8.

- Kerns SL, West CM, Andreassen CN, Barnett GC, Bentzen SM, Burnet NG, Dekker A, De Ruyscher D, Dunning A, Parliament M, Talbot C, Vega A, Rosenstein BS. Radiogenomics: the search for genetic predictors of radiotherapy response. *Future Oncol.* 2014b; 10:2391–406. [PubMed: 25525847]
- Kim HS, Kim SC, Kim SJ, Park CH, Jeung H-C, Kim YB, Ahn JB, Chung HC, Rha SY. Identification of a radiosensitivity signature using integrative metaanalysis of published microarray data for NCI-60 cancer cells. *BMC Genomics.* 2012; 13:348. [PubMed: 22846430]
- Kingery LB. Saturation in roentgen therapy: Its estimation and maintenance: preliminary report. *Archives of Dermatology and Syphilology.* 1920; 1:423–33.
- Kirkpatrick JP, Meyer JJ, Marks LB. The linear-quadratic model is inappropriate to model high dose per fraction effects in radiosurgery. *Seminars in radiation oncology.* 2008; 18:240–3. [PubMed: 18725110]
- Koller, D., Friedman, N. Probabilistic graphical models : principles and techniques. Cambridge, MA: MIT Press; 2009.
- Kuiper RP, Ligtenberg MJL, Hoogerbrugge N, Geurts van Kessel A. Germline copy number variation and cancer risk. *Current Opinion in Genetics & Development.* 2010; 20:282–9. [PubMed: 20381334]
- Kuska B. Beer, Bethesda, and Biology: How “Genomics” Came Into Being. *Journal of the National Cancer Institute.* 1998; 90:93. [PubMed: 9450566]
- Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, Zegers CML, Gillies R, Boellard R, Dekker A, Aerts HJWL. Radiomics: Extracting more information from medical images using advanced feature analysis. *European Journal of Cancer.* 2012; 48:441–6. [PubMed: 22257792]
- Lambin P, Roelofs E, Reymen B, Velazquez ER, Buijsen J, Zegers CML, Carvalho S, Leijenaar RTH, Nalbantov G, Oberije C, Scott Marshall M, Hoebbers F, Troost EGC, van Stiphout RGPM, van Elmpt W, van der Weijden T, Boersma L, Valentini V, Dekker A. ‘Rapid Learning health care in oncology’ – An approach towards decision support systems enabling customised radiotherapy. *Radiotherapy and Oncology.* 2013a; 109:159–64. [PubMed: 23993399]
- Lambin P, van Stiphout RGPM, Starmans MHW, Rios-Velazquez E, Nalbantov G, Aerts HJWL, Roelofs E, van Elmpt W, Boutros PC, Granone P, Valentini V, Begg AC, De Ruyscher D, Dekker A. Predicting outcomes in radiation oncology[multifactorial decision support systems. *Nat Rev Clin Oncol.* 2013b; 10:27–40. [PubMed: 23165123]
- Laurberg JR, Brems-Eskildsen AS, Nordentoft I, Fristrup N, Schepeler T, Ulhoi BP, Agerbaek M, Hartmann A, Bertz S, Wittlinger M, Fietkau R, Rodel C, Borre M, Jensen JB, Orntoft T, Dyrskjot L. Expression of TIP60 (tat-interactive protein) and MRE11 (meiotic recombination 11 homolog) predict treatment-specific outcome of localised invasive bladder cancer. *BJU Int.* 2012; 110:E1228–36. [PubMed: 23046361]
- Lea D, Catchside D. The mechanism of the induction by radiation of chromosome aberrations in *Tradescantia*. *J Genet.* 1942:216–45.
- Lee S, Ybarra N, Jeyaseelan K, Faria S, Kopek N, Brisebois P, Bradley JD, Robinson C, Seuntjens J, El Naqa I. Bayesian network ensemble as a multivariate strategy to predict radiation pneumonitis risk. *Med Phys.* 2015; 42:2421–30. [PubMed: 25979036]
- Levegrun S, Jackson A, Zelefsky MJ, Skwarchuk MW, Venkatraman ES, Schlegel W, Fuks Z, Leibel SA, Ling CC. Fitting tumor control probability models to biopsy outcome after three-dimensional conformal radiation therapy of prostate cancer: Pitfalls in deducing radiobiologic parameters for tumors from clinical data. *International Journal of Radiation Oncology Biology Physics.* 2001; 51:1064–80.
- Little JB, Nove J, Strong LC, Nichols WW. Survival of human diploid skin fibroblasts from normal individuals after X-irradiation. *Int J Radiat Biol.* 1988; 54:899–910. [PubMed: 2903888]
- Liu W, Sun J, Li G, Zhu Y, Zhang S, Kim ST, Wiklund F, Wiley K, Isaacs SD, Stattin P, Xu J, Duggan D, Carpten JD, Isaacs WB, Gronberg H, Zheng SL, Chang BL. Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer. *Cancer Res.* 2009; 69:2176–9. [PubMed: 19258504]

- Loeffler JS, Harris JR, Dahlberg WK, Little JB. In Vitro Radiosensitivity of Human Diploid Fibroblasts Derived from Women with Unusually Sensitive Clinical Responses to Definitive Radiation Therapy for Breast Cancer. *Radiation Research*. 1990; 121:227–31. [PubMed: 2305041]
- Lopez-Chavez A, Thomas A, Rajan A, Raffeld M, Morrow B, Kelly R, Carter CA, Guha U, Killian K, Lau CC, Abdullaev Z, Xi L, Pack S, Meltzer PS, Corless CL, Sandler A, Beadling C, Warrick A, Liewehr DJ, Steinberg SM, Berman A, Doyle A, Szabo E, Wang Y, Giaccone G. Molecular Profiling and Targeted Therapy for Advanced Thoracic Malignancies: A Biomarker-Derived, Multiarm, Multihistology Phase II Basket Trial. *Journal of Clinical Oncology*. 2015; 33:1000–7. [PubMed: 25667274]
- Luo Y, El Naqa I, McShan DL, Ray D, Lohse I, Matuszak MM, Owen D, Jolly S, Lawrence TS, Kong FS, Ten Haken RK. Unraveling biophysical interactions of radiation pneumonitis in non-small-cell lung cancer via Bayesian network analysis. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*. 2017
- Malaise EP, Fertil B, Deschavanne PJ, Chavaudra N, Brock WA. Initial slope of radiation survival curves is characteristic of the origin of primary and established cultures of human tumor cells and fibroblasts. *Radiat Res*. 1987; 111:319–33. [PubMed: 3628719]
- Marks LB. Dosimetric predictors of radiation-induced lung injury. *Int J Radiat Oncol*. 2002; 54:313–6.
- Mayer C, Popanda O, Greve B, Fritz E, Illig T, Eckardt-Schupp F, Gomolka M, Benner A, Schmezer P. A radiation-induced gene expression signature as a tool to predict acute radiotherapy-induced adverse side effects. *Cancer Lett*. 2011; 302:20–8. [PubMed: 21236564]
- McCarroll SA, Huett A, Kuballa P, Chlewicki SD, Landry A, Goyette P, Zody MC, Hall JL, Brant SR, Cho JH, Duerr RH, Silverberg MS, Taylor KD, Rioux JD, Altshuler D, Daly MJ, Xavier RJ. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat Genet*. 2008; 40:1107–12. [PubMed: 19165925]
- McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Statistics Subcommittee of the NCIEW Go CD. Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst*. 2005; 97:1180–4. [PubMed: 16106022]
- McShane LM, Hunsberger S, Adjei AA. Effective incorporation of biomarkers into phase II trials. *Clin Cancer Res*. 2009; 15:1898–905. [PubMed: 19276274]
- Mehta S, Hughes NP, Buffa FM, Li SP, Adams RF, Adwani A, Taylor NJ, Levitt NC, Padhani AR, Makris A, Harris AL. Assessing early therapeutic response to bevacizumab in primary breast cancer using magnetic resonance imaging and gene expression profiles. *J Natl Cancer Inst Monogr*. 2011; 2011:71–4. [PubMed: 22043045]
- Mei H, Sun S, Bai Y, Chen Y, Chai R, Li H. Reduced mtDNA copy number increases the sensitivity of tumor cells to chemotherapeutic drugs. *Cell Death Dis*. 2015; 6:e1710. [PubMed: 25837486]
- Miousse IR, Kutanzi KR, Koturbash I. Effects of ionizing radiation on DNA methylation: from experimental biology to clinical applications. *Int J Radiat Biol*. 2017:1–13.
- Mirjolet C, Merlin JL, Dalban C, Maingon P, Azria D. Correlation between radio-induced lymphocyte apoptosis measurements obtained from two French centres. *Cancer Radiother*. 2016; 20:391–4. [PubMed: 27372557]
- Moissenko, V., Deasy, JO., Van Dyk, J. The Modern Technology of Radiation Oncology: A Compendium for Medical Physicists and Radiation Oncologists. Van Dyk, J., editor. Madison, WI: Medical Physics Publishing; 2005. p. 185-220.
- Munley MT, Lo JY, Sibley GS, Bentel GC, Anscher MS, Marks LB. A neural network to predict symptomatic lung injury. *Physics in Medicine and Biology*. 1999; 44:2241–9. [PubMed: 10495118]
- Nikjoo H, Emfietzoglou D, Liamsuwan T, Taleei R, Liljequist D, Uehara S. Radiation track, DNA damage and response-a review. *Reports on progress in physics Physical Society*. 2016; 79:116601.
- Nikjoo H, Uehara S, Emfietzoglou D, Cucinotta FA. Track-structure codes in radiation research. *Radiat Meas*. 2006; 41:1052–74.
- O'Leary Valerie B, Ovsepian Saak V, Carrascosa Laura G, Buske Fabian A, Radulovic V, Niyazi M, Moertl S, Trau M, Atkinson Michael J, Anastasov N. PARTICLE, a Triplex-Forming

Long ncRNA, Regulates Locus-Specific Methylation in Response to Low-Dose Irradiation. *Cell Reports*. 2015; 11:474–85. [PubMed: 25900080]

Oh JH, Craft J, Al Lozi R, Vaidya M, Meng Y, Deasy JO, Bradley JD, El Naqa I. A Bayesian network approach for modeling local failure in lung cancer. *Phys Med Biol*. 2011a; 56:1635–51. [PubMed: 21335651]

Oh JH, Craft J, Al Lozi R, Vaidya M, Meng Y, Deasy JO, Bradley JD, Naqa IE. A Bayesian network approach for modeling local failure in lung cancer. *Phys Med Biol*. 2011b; 56:1635–51. [PubMed: 21335651]

Oh JH, Craft JM, Townsend R, Deasy JO, Bradley JD, El Naqa I. A bioinformatics approach for biomarker identification in radiation-induced lung inflammation from limited proteomics data. *J Proteome Res*. 2011c; 10:1406–15. [PubMed: 21226504]

Oh JH, Wong HP, Wang X, Deasy JO. A bioinformatics filtering strategy for identifying radiation response biomarker candidates. *PloS one*. 2012; 7:e38870. [PubMed: 22768051]

Oldenhuis CN, Oosting SF, Gietema JA, de Vries EG. Prognostic versus predictive value of biomarkers in oncology. *Eur J Cancer*. 2008; 44:946–53. [PubMed: 18396036]

Oppitz U, Baier K, Wulf J, Schakowski R, Flentje M. The in vitro colony assay: a predictor of clinical outcome. *Int J Radiat Biol*. 2001; 77:105–10. [PubMed: 11213342]

Painter R, Meyn R, Withers H. The role of DNA damage and repair in cell killing induced by ionizing radiation. *Radiat Biol Cancer Res*. 1980:59–68.

Pang Q, Wei Q, Xu T, Yuan X, Lopez Guerra JL, Levy LB, Liu Z, Gomez DR, Zhuang Y, Wang LE, Mohan R, Komaki R, Liao Z. Functional promoter variant rs2868371 of HSPB1 is associated with risk of radiation pneumonitis after chemoradiation for non-small cell lung cancer. *International journal of radiation oncology, biology, physics*. 2013; 85:1332–9.

Pater P, Backstom G, Villegas F, Ahnesjo A, Enger SA, Seuntjens J, El Naqa I. Proton and light ion RBE for the induction of direct DNA double strand breaks. *Med Phys*. 2016; 43:2131. [PubMed: 27147325]

Pater P, Seuntjens J, El Naqa I, Bernal MA. On the consistency of Monte Carlo track structure DNA damage simulations. *Med Phys*. 2014; 41:121708. [PubMed: 25471955]

Peacock J, Ashton A, Bliss J, Bush C, Eady J, Jackson C, Owen R, Regan J, Yarnold J. Cellular radiosensitivity and complication risk after curative radiotherapy. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*. 2000; 55:173–8. [PubMed: 10799729]

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Almeida J, Bacchelli E, Bader GD, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bolte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Bryson SE, Carson AR, Casallo G, Casey J, Chung BH, Cochrane L, Corsello C, Crawford EL, Crossett A, Cytrynbaum C, Dawson G, de Jonge M, Delorme R, Drmic I, Duketis E, Duque F, Estes A, Farrar P, Fernandez BA, Folstein SE, Fombonne E, Freitag CM, Gilbert J, Gillberg C, Glessner JT, Goldberg J, Green A, Green J, Guter SJ, Hakonarson H, Heron EA, Hill M, Holt R, Howe JL, Hughes G, Hus V, Igliozzi R, Kim C, Klauck SM, Klevzon A, Korvatska O, Kustanovich V, Lajonchere CM, Lamb JA, Laskawiec M, Leboyer M, Le Couteur A, Leventhal BL, Lionel AC, Liu XQ, Lord C, Lotspeich L, Lund SC, Maestrini E, Mahoney W, Mantoulan C, Marshall CR, McConachie H, McDougale CJ, McGrath J, McMahon WM, Merikangas A, Migita O, Minshew NJ, Mirza GK, Munson J, Nelson SF, Noakes C, Noor A, Nygren G, Oliveira G, Papanikolaou K, Parr JR, Parrini B, Paton T, Pickles A, Pilorge M, Piven J, Ponting CP, Posey DJ, Poustka A, Poustka F, Prasad A, Ragoussis J, Renshaw K, Rickaby J, Roberts W, Roeder K, Roge B, Rutter ML, Bierut LJ, Rice JP, Salt J, Sansom K, Sato D, Segurado R, Sequeira AF, Senman L, Shah N, Sheffield VC, Soorya L, Sousa I, Stein O, Sykes N, Stoppioni V, Strawbridge C, Tancredi R, Tansey K, Thiruvahindrapuram B, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H, Vincent JB, Volkmar F, Wallace S, Wang K, Wang Z, Wassink TH, Webber C, Weksberg R, Wing K, Wittemeyer K, Wood S, Wu J, Yaspan BL, Zurawiecki D, Zwaigenbaum L, Buxbaum JD, Cantor RM, Cook EH, Coon H, Cuccaro ML, Devlin B, Ennis S, Gallagher L, Geschwind DH, Gill M, Haines JL, Hallmayer J, Miller J, Monaco AP, Nurnberger JI Jr, Paterson AD, Pericak-Vance MA, Schellenberg GD, Szatmari P, Vicente AM, Vieland VJ, Wijsman EM,

- Scherer SW, Sutcliffe JS, Betancur C. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*. 2010; 466:368–72. [PubMed: 20531469]
- Pouliliou SE, Lialiaris TS, Dimitriou T, Giatromanolaki A, Papazoglou D, Pappa A, Pistevo K, Kalamida D, Koukourakis MI. Survival Fraction at 2 Gy and gammaH2AX Expression Kinetics in Peripheral Blood Lymphocytes From Cancer Patients: Relationship With Acute Radiation-Induced Toxicities. *International journal of radiation oncology, biology, physics*. 2015; 92:667–74.
- Pramana J, Van den Brekel MW, van Velthuysen ML, Wessels LF, Nuyten DS, Hofland I, Atsma D, Pimentel N, Hoebbers FJ, Rasch CR, Begg AC. Gene expression profiling to predict outcome after chemoradiation in head and neck cancer. *International journal of radiation oncology, biology, physics*. 2007; 69:1544–52.
- Prokopiou S, Moros EG, Poleszczuk J, Caudell J, Torres-Roca JF, Latifi K, Lee JK, Myerson R, Harrison LB, Enderling H. A proliferation saturation index to predict radiation response and personalize radiotherapy fractionation. *Radiat Oncol*. 2015; 10:159. [PubMed: 26227259]
- Qiao Y, Tyson C, Hrynchak M, Lopez-Rangel E, Hildebrand J, Martell S, Fawcett C, Kasmara L, Calli K, Harvard C, Liu X, Holden JJ, Lewis SM, Rajcan-Separovic E. Clinical application of 2.7M Cytogenetics array for CNV detection in subjects with idiopathic autism and/or intellectual disability. *Clin Genet*. 2013; 83:145–54. [PubMed: 22369279]
- Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet*. 2011; 12:529–41. [PubMed: 21747404]
- Redig AJ, Jänne PA. Basket Trials and the Evolution of Clinical Trial Design in an Era of Genomic Medicine. *Journal of Clinical Oncology*. 2015; 33:975–7. [PubMed: 25667288]
- Rekers NH, Troost EG, Zegers CM, Germeraad WT, Dubois LJ, Lambin P. Stereotactic ablative body radiotherapy combined with immunotherapy: present status and future perspectives. *Cancer radiotherapie : journal de la Societe francaise de radiotherapie oncologique*. 2014; 18:391–5. [PubMed: 25179250]
- Rodningen OK, Borresen-Dale AL, Alsner J, Hastie T, Overgaard J. Radiation-induced gene expression in human subcutaneous fibroblasts is predictive of radiation-induced fibrosis. *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*. 2008; 86:314–20. [PubMed: 17963910]
- Rosenstein BS, Capala J, Efstathiou JA, Hammerbacher J, Kerns SL, Kong FM, Ostrer H, Prior FW, Vikram B, Wong J, Xiao Y. How Will Big Data Improve Clinical and Basic Research in Radiation Therapy? *International journal of radiation oncology, biology, physics*. 2016; 95:895–904.
- Rosenstein BS, West CM, Bentzen SM, Alsner J, Andreassen CN, Azria D, Barnett GC, Baumann M, Burnet N, Chang-Claude J, Chuang EY, Coles CE, Dekker A, De Ruyck K, De Ruyscher D, Drumea K, Dunning AM, Easton D, Eeles R, Fachal L, Gutierrez-Enriquez S, Haustermans K, Henriquez-Hernandez LA, Imai T, Jones GD, Kerns SL, Liao Z, Onel K, Ostrer H, Parliament M, Pharoah PD, Rebbeck TR, Talbot CJ, Thierens H, Vega A, Witte JS, Wong P, Zenhausern F, Radiogenomics C. Radiogenomics: radiobiology enters the era of big data and team science. *International journal of radiation oncology, biology, physics*. 2014; 89:709–13.
- Scott JG, Berglund A, Schell MJ, Mihaylov I, Fulp WJ, Yue B, Welsh E, Caudell JJ, Ahmed K, Strom TS, Mellon E, Venkat P, Johnstone P, Foekens J, Lee J, Moros E, Dalton WS, Eschrich SA, McLeod H, Harrison LB, Torres-Roca JF. A genome-based model for adjusting radiotherapy dose (GARD): a retrospective, cohort-based study. *Lancet Oncol*. 2017; 18:202–11. [PubMed: 27993569]
- Seibold P, Behrens S, Schmezer P, Helmbold I, Barnett G, Coles C, Yarnold J, Talbot CJ, Imai T, Azria D, Koch CA, Dunning AM, Burnet N, Bliss JM, Symonds RP, Rattay T, Suga T, Kerns SL, Bourcier C, Vallis KA, Sautter-Bihl ML, Classen J, Debus J, Schnabel T, Rosenstein BS, Wenz F, West CM, Popanda O, Chang-Claude J. XRCC1 Polymorphism Associated With Late Toxicity After Radiation Therapy in Breast Cancer Patients. *International journal of radiation oncology, biology, physics*. 2015; 92:1084–92.
- Sharma RA, Plummer R, Stock JK, Greenhalgh TA, Ataman O, Kelly S, Clay R, Adams RA, Baird RD, Billingham L, Brown SR, Buckland S, Bulbeck H, Chalmers AJ, Clack G, Cranston AN, Damstrup L, Ferraldeschi R, Forster MD, Golec J, Hagan RM, Hall E, Hanauske A-R,

- Harrington KJ, Haswell T, Hawkins MA, Illidge T, Jones H, Kennedy AS, McDonald F, Melcher T, O'Connor JPB, Pollard JR, Saunders MP, Sebag-Montefiore D, Smitt M, Staffurth J, Stratford IJ, Wedge SR. on behalf of the NCA-PJWG. Clinical development of new drug-radiotherapy combinations. *Nat Rev Clin Oncol*. 2016; 13:627–42. [PubMed: 27245279]
- Shivade C, Raghavan P, Fosler-Lussier E, Embi PJ, Elhadad N, Johnson SB, Lai AM. A review of approaches to identifying patient phenotype cohorts using electronic health records. *Journal of the American Medical Informatics Association*. 2013
- Shlien A, Malkin D. Copy number variations and cancer. *Genome Med*. 2009; 1:62. [PubMed: 19566914]
- Sinoquet, C., Mourad, RI. Probabilistic graphical models for genetics, genomics, and postgenomics. Oxford: Oxford University Press; 2014.
- Speers C, Zhao S, Liu M, Bartelink H, Pierce LJ, Feng FY. Development and Validation of a Novel Radiosensitivity Signature in Human Breast Cancer. *Clin Cancer Res*. 2015; 21:3667–77. [PubMed: 25904749]
- Steel, GG. Basic clinical radiobiology. London New York: Arnold; Oxford University Press; 2002.
- Stewart RD, Streitmatter SW, Argento DC, Kirkby C, Goorley JT, Moffitt G, Jevremovic T, Sandison GA. Rapid MCNP simulation of DNA double strand break (DSB) relative biological effectiveness (RBE) for photons, neutrons, and light ions. *Phys Med Biol*. 2015; 60:8249–74. [PubMed: 26449929]
- Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavaré S, Deloukas P, Hurles ME, Dermitzakis ET. Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. *Science*. 2007; 315:848–53. [PubMed: 17289997]
- Su M, Miftena M, Whiddon C, Sun X, Light K, Marks L. An artificial neural network for predicting the incidence of radiation pneumonitis. *Medical Physics*. 2005; 32:318–25. [PubMed: 15789575]
- Talbot CJ, Tanteles GA, Barnett GC, Burnet NG, Chang-Claude J, Coles CE, Davidson S, Dunning AM, Mills J, Murray RJ, Popanda O, Seibold P, West CM, Yarnold JR, Symonds RP. A replicated association between polymorphisms near TNF α and risk for adverse reactions to radiotherapy. *Br J Cancer*. 2012; 107:748–53. [PubMed: 22767148]
- Taylor AM, Harnden DG, Arlett CF, Harcourt SA, Lehmann AR, Stevens S, Bridges BA. Ataxia telangiectasia: a human mutation with abnormal radiation sensitivity. *Nature*. 1975; 258:427–9. [PubMed: 1196376]
- Ten Haken, R., Kessler, M. 3-D Conformal and Intensity Modulated Radiation Therapy. Purdy, J., et al., editors. Madison, WI: Advanced Medical Publishing; 2001. p. 43–59.
- Ten Haken RK, Martel MK, Kessler ML, Hazuka MB, Lawrence TS, Robertson JM, Turrisi AT, Lichter AS. Use of Veff and iso-NTCP in the implementation of dose escalation protocols. *International journal of radiation oncology, biology, physics*. 1993; 27:689–95.
- Tomatis S, Rancati T, Fiorino C, Vavassori V, Fellin G, Cagna E, Mauro FA, Girelli G, Monti A, Baccolini M, Naldi G, Bianchi C, Menegotti L, Pasquino M, Stasi M, Valdagni R. Late rectal bleeding after 3D-CRT for prostate cancer: development of a neural-network-based predictive model. *Physics in Medicine and Biology*. 2012; 57:1399. [PubMed: 22349550]
- Torres-Roca JF. A molecular assay of tumor radiosensitivity: a roadmap towards biology-based personalized radiation therapy. *Personalized medicine*. 2012; 9:547–57. [PubMed: 23105945]
- Tucker SL, Li M, Xu T, Gomez D, Yuan X, Yu J, Liu Z, Yin M, Guan X, Wang LE, Wei Q, Mohan R, Vinogradskiy Y, Martel M, Liao Z. Incorporating single-nucleotide polymorphisms into the Lyman model to improve prediction of radiation pneumonitis. *International journal of radiation oncology, biology, physics*. 2013; 85:251–7.
- Vapnik, V. Statistical Learning Theory. New York: Wiley; 1998.
- Wang W, Luo Y-p. MicroRNAs in breast cancer: oncogene and tumor suppressors with clinical potential. *Journal of Zhejiang University. Science. B*. 2015; 16:18–31. [PubMed: 25559952]
- Webb, S. The physics of three-dimensional radiation therapy : conformal radiotherapy, radiosurgery, and treatment planning. Bristol, UK ; Philadelphia: Institute of Physics Pub; 2001.
- Weigel C, Schmezer P, Plass C, Popanda O. Epigenetics in radiation-induced fibrosis. *Oncogene*. 2015; 34:2145–55. [PubMed: 24909163]

- Weigel C, Veldwijk MR, Oakes CC, Seibold P, Slynko A, Liesenfeld DB, Rabionet M, Hanke SA, Wenz F, Sperk E, Benner A, Rosli C, Sandhoff R, Assenov Y, Plass C, Herskind C, Chang-Claude J, Schmezer P, Popanda O. Epigenetic regulation of diacylglycerol kinase alpha promotes radiation-induced fibrosis. *Nature communications*. 2016; 7:10893.
- West C, Azria D, Chang-Claude J, Davidson S, Lambin P, Rosenstein B, De Ruyscher D, Talbot C, Thierens H, Valdagni R, Vega A, Yuille M. The REQUITE Project: Validating Predictive Models and Biomarkers of Radiotherapy Toxicity to Reduce Side-effects and Improve Quality of Life in Cancer Survivors. *Clinical Oncology*. 2014; 26:739–42. [PubMed: 25267305]
- West C, Rosenstein BS. Establishment of a Radiogenomics Consortium. *International Journal of Radiation Oncology*Biophysics*. 2010; 76:1295–6.
- West C, Rosenstein BS, Alsner J, Azria D, Barnett G, Begg A, Bentzen S, Burnet N, Chang-Claude J, Chuang E, Coles C, De Ruyck K, De Ruyscher D, Dunning A, Elliott R, Fachal L, Hall J, Haustermans K, Herskind C, Hoelscher T, Imai T, Iwakawa M, Jones D, Kulich C, Langendijk JH, O'Neils P, Ozsahin M, Parliament M, Polanski A. Establishment of a Radiogenomics Consortium. *International journal of radiation oncology, biology, physics*. 2010:76.
- West CM. Radiobiological Modelling in Radiation Oncology. 2007:35–50.
- West CM, Barnett GC. Genetics and genomics of radiotherapy toxicity: towards prediction. *Genome Medicine*. 2011; 3:52. [PubMed: 21861849]
- West CM, Davidson SE, Roberts SA, Hunter RD. Intrinsic radiosensitivity and prediction of patient response to radiotherapy for carcinoma of the cervix. *Br J Cancer*. 1993; 68:819–23. [PubMed: 8398714]
- West CM, Davidson SE, Roberts SA, Hunter RD. The independence of intrinsic radiosensitivity as a prognostic factor for patient response to radiotherapy of carcinoma of the cervix. *Br J Cancer*. 1997; 76:1184–90. [PubMed: 9365167]
- West CM, Elyan SA, Berry P, Cowan R, Scott D. A comparison of the radiosensitivity of lymphocytes from normal donors, cancer patients, individuals with ataxia-telangiectasia (A-T) and A-T heterozygotes. *Int J Radiat Biol*. 1995; 68:197–203. [PubMed: 7658145]
- West CML, Elliott RM, Burnet NG. The Genomics Revolution and Radiotherapy. *Clinical Oncology*. 2007; 19:470–80. [PubMed: 17419040]
- Yang L, Taylor J, Eustace A, Irlam J, Denley H, Hoskin PJ, Alsner J, Buffa FM, Harris AL, Choudhury A, West CM. A Gene Signature for Selecting Benefit from Hypoxia Modification of Radiotherapy for High Risk Bladder Cancer Patients. *Clin Cancer Res*. 2017
- Yard BD, Adams DJ, Chie EK, Tamayo P, Battaglia JS, Gopal P, Rogacki K, Pearson BE, Phillips J, Raymond DP, Pennell NA, Almeida F, Cheah JH, Clemons PA, Shamji A, Peacock CD, Schreiber SL, Hammerman PS, Abazeed ME. A genetic basis for the variation in the vulnerability of cancer to DNA damage. *Nature communications*. 2016; 7:11428.
- Yavas G, Koyuturk M, Ozsoyoglu M, Gould MP, LaFramboise T. An optimization framework for unsupervised identification of rare copy number variation from SNP array data. *Genome Biol*. 2009; 10:R119. [PubMed: 19849861]
- Zeng J, Baik C, Bhatia S, Mayr N, Rengan R. Combination of stereotactic ablative body radiation with targeted therapies. *The Lancet. Oncology*. 2014; 15:e426–34. [PubMed: 25186046]
- Zhang F, Gu W, Hurles ME, Lupski JR. Copy Number Variation in Human Health, Disease, and Evolution. *Annual Review of Genomics and Human Genetics*. 2009; 10:451–81.
- Zhang Y, He Q, Hu Z, Feng Y, Fan L, Tang Z, Yuan J, Shan W, Li C, Hu X, Tanyi JL, Fan Y, Huang Q, Montone K, Dang CV, Zhang L. Long noncoding RNA LINP1 regulates repair of DNA double-strand breaks in triple-negative breast cancer. *Nat Struct Mol Biol*. 2016; 23:522–30. [PubMed: 27111890]
- Zhao SG, Chang SL, Spratt DE, Erho N, Yu M, Ashab HA-D, Alshalalfa M, Speers C, Tomlins SA, Davicioni E, Dicker AP, Carroll PR, Cooperberg MR, Freedland SJ, Karnes RJ, Ross AE, Schaeffer EM, Den RB, Nguyen PL, Feng FY. Development and validation of a 24-gene predictor of response to postoperative radiotherapy in prostate cancer: a matched, retrospective analysis. *The Lancet Oncology*. 2016; 17:1612–20. [PubMed: 27743920]

Figure 1.

The human body is a valuable resource for varying solid and fluid types of specimens, which can yield different –omic (genomics, transcriptomics, proteomics, metabolomics, radiomics) predictive biomarkers, in addition to dosimetric and clinical factors used in radiotherapy that would undergo major processes of annotation, curation, and preparation before being applied into radiogenomics modeling of treatment outcomes (e.g., tumor response, toxicity).

Figure 2.

The informatics understanding of heterogeneous variable interactions as a feedback into the treatment planning system to improve patient's outcomes. The noise reflects the uncertainties in the measured (clinical, physical, and biological) factors). The computer represents the modeling process, which could be embedded in a treatment planning system. The gear crankshaft is part of the continuous feedback during the course of treatment or a clinical trial to refine or take new updated measurements (El Naqa, 2013).

Figure 3.

Comparison between the predicted incidence of grade 2+ rectal bleeding and the actual incidence of grade 2+ rectal bleeding. The predicted outcomes were produced after applying the logistic regression to outputs of the LASSO model using 2 principal components on the validation data set with 484 SNPs that entered the LASSO. Based on the sorted predicted outcomes, the patients were binned into 6 groups, with the first being the lowest toxicity group and the sixth being the highest. The ratio above each group represents the observed number of patients who experienced grade 2+ rectal bleeding and the total number of patients in the group (Kerns *et al.*, 2015).

Figure 4. Ten-year receiver operating characteristic (ROC) curves (A) and Kaplan–Meier survival estimate (B) analysis in validation dataset of radiosensitivity in breast cancer using random forest machine learning (Speers *et al.*, 2015).

Figure 5.

A radiogenomic model using systems biology techniques based on Bayesian networks is used for modeling radiation pneumonitis (RP) in lung cancer. First row shows pre-treatment BN modeling of RP. (a) Markov blanket, (b) BN structure, and (c) ROC analysis on cross-validation. The second row shows during-treatment BN modeling of RP. (d) Markov blanket, (e) BN structure, and (f) ROC analysis on cross-validation (Luo *et al.*, 2017).