Radiotherapy and Oncology 111 (2014) 178–185

Contents lists available at ScienceDirect
Radiotherapy and Oncology
journal homepage: www.thegreenjournal.com

Radiogenomics

A genome wide association study (GWAS) providing evidence of an association between common genetic variants and late radiotherapy toxicity

Gillian C. Barnett a,b,⇑, Deborah Thompson a, Laura Fachal c, Sarah Kerns d, Chris Talbot e, Rebecca M. Elliott f, Leila Dorling g, Charlotte E. Coles g, David P. Dearnaley h, Barry S. Rosenstein d, Ana Vega c, Paul Symonds i, John Yarnold h, Caroline Baynes a, Kyriaki Michailidou a, Joe Dennis a, Jonathan P. Tyrer j, Jennifer S. Wilkinson g, Antonio Gómez-Caamaño i, George A. Tanteles g, Radka Platte a, Rebecca Mayes a, Don Conroy a, Mel Maranian a, Craig Luccarini a, Sarah L. Gulliford j, Matthew R. Sydes i, Emma Hall m, Joanne Haviland m, Vivek Misra n, Jennifer Titley m, Søren M. Bentzen o, Paul D.P. Pharoah a, Neil G. Burnet b,1, Alison M. Dunning a,1, Catharine M.L. West a,⇑

a,1 Joint last authors.

a Centre for Cancer Genetic Epidemiology, University of Cambridge, Strangeways Research Laboratory; b University of Cambridge, Department of Oncology, Oncology Centre, Cambridge University Hospitals NHS Foundation Trust, UK; c Fundación Pública Galega de Medicina Xénómica-SERGBAS, Grupo de Medicina Xénómica, CIBERER, IDIS, Santiago de Compostela, Spain; d Department of Radiation Oncology, Icahn Mount Sinai School of Medicine, NY, USA; e Department of Genetics, University of Leicester; f Institute of Cancer Sciences, University of Manchester, Manchester Academic Health Science Centre, Christie Hospital; g Oncology Centre, Cambridge University Hospitals NHS Foundation Trust; h Institute of Cancer Research and Royal Marsden NHS Foundation Trust, Sutton; i Department of Cancer Studies and Molecular Medicine, University Hospitals of Leicester, UK; j Department of Radiation Oncology, Complejo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain; k The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; l Cancer and Other Non-Infectious Diseases, MRC Clinical Trials Unit, London; m Institute of Cancer Research-Clinical Trials and Statistics Unit, Sutton; n Department of Clinical Oncology, Christie Hospital, Manchester, UK; o Division of Biostatistics and Bioinformatics, Greenebaum Cancer Center; Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, USA

A R T I C L E   I N F O

Article history:
Received 27 November 2013
Received in revised form 30 January 2014
Accepted 17 February 2014
Available online 28 April 2014

Keywords:
Radiotherapy
Late toxicity
Adverse effects
Genetic
Genome-wide association scan

A B S T R A C T

Background and purpose: This study was designed to identify common single nucleotide polymorphisms (SNPs) associated with toxicity 2 years after radiotherapy.

Materials and methods: A genome wide association study was performed in 1850 patients from the RAPPER study: 1217 received adjuvant breast radiotherapy and 633 had radical prostate radiotherapy. Genotype associations with both overall and individual endpoints of toxicity were tested via univariable and multivariable regression. Replication of potentially associated SNPs was carried out in three independent patient cohorts who had radiotherapy for prostate (516 RADIOGEN and 862 Gene-PARE) or breast (355 LeND) cancer.

Results: Quantile–quantile plots show more associations at the P < 5 × 10 −7 level than expected by chance (164 vs. 9 for the prostate cases and 29 vs. 4 for breast cases), providing evidence that common genetic variants are associated with risk of toxicity. Strongest associations were for individual endpoints rather than an overall measure of toxicity in all patients. However, in general, significant associations were not validated at a nominal 0.05 level in the replication cohorts.

Conclusions: This largest GWAS to date provides evidence of true association between common genetic variants and toxicity. Associations with toxicity appeared to be tumour site-specific. Future GWAS require higher statistical power, in particular in the validation stage, to test clinically relevant effect sizes of SNP associations with individual endpoints, but the required sample sizes are achievable.

© 2014 Published by Elsevier Ireland Ltd. Radiotherapy and Oncology 111 (2014) 178–185

Late toxicity from radiotherapy, which may continue to accumulate years after completion of treatment, is generally irreversible and often decreases health-related quality of life. Examples include bowel or urinary incontinence after radiotherapy for prostate cancer, or breast shrinkage after post-operative radiotherapy for breast cancer. Known causes of variation in incidence or
severity of toxicity include radiotherapy dose, dose distribution, co-morbidities such as diabetes, and concurrent chemotherapy [1]. However, large patient-to-patient variability in response remains, after allowing for known risk factors, and there is evidence that these are intrinsic to the patient. Studies on the heritability of susceptibility to radiotherapy toxicity are limited, but estimates of heritability of in vitro cellular radiosensitivity range from 60% to 80% [1]. With a few exceptions [2–5], radiogenomic studies published to-date used a candidate gene approach, in which polymorphisms in or near genes thought to be important in the pathogenesis of late toxicity are investigated. However, positive associations proved difficult to replicate [6–8]. This study aimed to identify common genetic variants associated with late radiotherapy toxicity using a phased genome-wide association study (GWAS) design. The initial phase used samples and data from the UK RAPPER (Radiogenomics: Assessment of Polymorphisms for Predicting the Effects of Radiotherapy) study [9,10], with a replication phase in three independent cohorts.

In response to the need for improving the quality of research in Radiogenomics and increasing the transparency and completeness of reporting, the Radiogenomics Consortium recently published the STROGAR guidelines. This study adheres to these important guidelines [11].

Materials and methods

RAPPER GWAS patients

RAPPER (UKCRN1471) recruits patients from clinical trials and observational studies, which prospectively collect toxicity data [9]. RAPPER is approved by the East of England Cambridge South Research Ethics Committee (05/Q0108/365) and informed consent is obtained from all patients. This study involved 1850 patients with two-year radiotherapy toxicity data from six cohorts (Supplementary Fig. 1 is the CONSORT diagram). All patients have undergone potentially curative treatment with radiotherapy as a major component.

There were 1217 breast cancer patients who received adjuvant radiotherapy following conservative surgery: 935 from the Cambridge Breast IMRT Trial (ISRCTN21474421) [12], 56 from a prospective study at the Christie Hospital, Manchester [6], 55 from the IMPORT (Intensity Modulated and Partial Organ RadioTherapy) LOW (ISRCTN12852634) trial of partial breast radiotherapy [13], and 171 from the RACE (Radiation Complications and Epidemiology) study [14]. There were 633 patients who received radical prostate radiotherapy following neoadjuvant androgen suppression: 223 from the MRC RT01 trial (ISRCTN47772397) [15], and 410 from stages 1 and 2 of the Conventional or Hypofractionated High Dose Intensity Modulated Radiotherapy for Prostate Cancer (CHHiP) trial (ISRCTN97182923) [16].

Blood was taken prior to radiotherapy (Cambridge IMRT, Manchester study, RACE), at a minimum of 6 months following end of treatment (MRC RT01 and CHHiP trials) or at any point in the trial (IMPORT LOW).

Replication cohorts

Replication of the most significant associations identified in RAPPER was carried out in cohorts from the Radiogenomics Consortium (RGC) with late toxicity data available. This included 1378 prostate cancer patients with 2-year toxicity data: 516 patients treated with conformal radical or post-prostatectomy radiotherapy at the Clinical University Hospital of Santiago de Compostela, Spain, in the RADIOGEN trial [17,18], and 862 treated with brachytherapy with or without additional external beam radiotherapy at Mount Sinai Hospital, in the Gene-PARE study [2,3]. Replication was also carried out in 355 breast cancer patients from the LeND study [8,19]. All patients gave written informed consent for use of their samples in genetic research. The RADIOGEN protocol was approved by the ethics review board of the Galician Ethics Committee for Clinical Research. Gene-PARE was approved by the Mount Sinai Medical Center Institutional Review Board. LeND received both local and national ethics approval.

Covariates

Data were available on prescribed radiotherapy dose, age, ethnicity, smoking history and diabetes mellitus for patients recruited to the Manchester prospective study, the Cambridge IMRT, CHHiP and RT01 trials. In breast cancer patients, use of tamoxifen and chemotherapy, radiotherapy breast boost, breast volume, co-morbidity cardiovascular disease and cosmesis after surgery were also recorded. In RACE, breast volume was estimated by three independent observers from baseline clinical photographs; in the other studies breast volume was calculated from the radiotherapy plans. In prostate patients data on hypertension, previous surgery, clinical stage, risk of seminal vesicle involvement, concomitant hormone therapy and baseline symptoms were also documented. Data on acute toxicity were available in the Cambridge IMRT, Manchester prospective breast study, RT01, CHHiP and Spanish patients. Dose–volume metrics were available from dose-volume histograms in CHHiP and RADIOGEN patients.

Assessment of toxicity

Toxicity was assessed at two years following radiotherapy using standardised scoring systems (Supplementary Table 1). To overcome the use of different toxicity scoring systems in different tissues and studies, we used the scale-independent STAT (Standardised Total Average Toxicity) score as the main outcome measure. STAT scores provide a standardised, scale-independent measure of toxicity and were derived using individual endpoints, as described previously [20]. The individual endpoints studied in breast patients, namely pain, telangiectasia, and breast shrinkage, have been shown to be measures of radiotherapy toxicity unrelated to previous breast surgery [12]; endpoints in the prostate patients, namely urinary incontinence, decreased stream, urine frequency, nocturnal frequency, proctitis, rectal bleeding, and rectal incontinence, show a radiation dose–volume response [21]. Changes in scores from baseline (pre-hormone treatment in prostate patients) to those recorded at two years were calculated [22]. Erectile dysfunction was not analysed, as few men had adequate, self-reported, erectile function at baseline.

Genotyping, quality control and imputation

Samples were genotyped using the Illumina CytoSNP12 array. After standard quality control exclusions, genotypes were available for 249,679 SNPs in 1773 patients with estimated European ancestry. Imputed genotype dosages were based on the HapMap2 CEU reference panel (Supplementary Methods).

Statistical analysis

Analysis of late toxicity with genotype was performed using a standard approach. Univariable analysis (UVA) was performed by linear regression of mean toxicity scores against the number of minor alleles (0, 1 or 2) or the imputed genotype dosage, using a 1–degree of freedom (df) trend test. Multivariable analyses (MVA) were performed of overall toxicity (STAT) and of individual endpoints against all covariates identified from UVA of patient- and treatment-related factors with probability of association of
GWAS of late radiation toxicity

For each endpoint, the resulting residuals estimate the risk of toxicity not explained by available patient- and treatment-related factors [20]. The means of these residuals for overall toxicity (rSTAT) and for individual end-points were correlated with genotype using linear regression, as for the UVA. Differences due to study populations were accounted for by adjusting for trial.

Using P-values obtained from the test of association, the prevalence of the corresponding toxicity endpoint and the minor allele frequencies (MAF), the effect size of SNPs associated with toxicity on MVA were expressed as a relative risk (RR) of toxicity in carriers versus non-carriers (Supplementary Methods). The prevalence of moderate/severe overall late toxicity in unselected patients is assumed to be 20%.

The nominal level at which an individual association is considered significant in genome-wide studies is \( P = 5 \times 10^{-8} \). SNPs with low MAF generate false positive associations more often. For a less common variant, a result declared significant at a certain P-value is more likely to be a false positive than for a more common variant. To relax the assumptions made in the distribution of the endpoint, non-parametric analysis was also performed, using Spearman’s rank correlation coefficient, for all SNPs associated with toxicity with linear regression \( P < 10^{-7} \) and MAF \(< 0.05\). For toxicity end-points scored on an ordinal scale, polychotomous logistic regression was also performed for SNPs significant on linear regression. All analyses used Stata version 10.1 and the GenABEL package implemented in the R statistical package [23].

Using the number of genotyped or imputed SNPs with MAF > 0.05 (2,168,129), the number of associated SNPs expected by chance was estimated for breast (4 endpoints) and prostate (8 endpoints) patients.

### Power calculations

The RAPPER GWAS was powered to detect significant associations between common SNPs and late radiotherapy toxicity that were tumour-site independent (i.e. displayed effects in both breast and prostate patients) (Supplementary Table 2). Assuming a 20% incidence of toxicity in the population, the power to detect an association at \( P < 1 \times 10^{-7} \) would be 99.6% for a SNP with MAF = 0.15 and RR = 3, and only 1% for a SNP with MAF = 0.05 and RR = 2.

### Results

A total of 2,417,493 genotyped or imputed SNPs were analysed for association with overall, tumour-site independent toxicity (STAT) in 1773 patients. The SNPs were also analysed for association with 10 individual toxicity endpoints, seven in 579 prostate patients and three in 1194 breast patients.

Table 1 shows the number of observed SNPs with P-values below a given significance threshold compared to the number expected by chance alone. Marked excesses of significant associations were evident, particularly for prostate cancer endpoints. QQ and Manhattan plots for STAT and for four of the ten site-specific endpoints are presented in Figs. 1 and 2. The Q–Q plots display deviation from the null distribution at the tail, strongly suggesting that common SNPs are associated with risk of radiotherapy toxicity. The multivariable Q–Q plots for overall toxicity, rectal bleeding and nocturia show more associations with \( P \)-values of \(< 10^{-7} \) than would be expected by chance. For telangiectasia and rectal incontinence the line deviates earlier and there are more \( P \)-values of \(< 10^{-2} \) than would be expected by chance, suggesting hundreds of SNPs may be associated with these endpoints. The Q–Q plots show scant evidence for population stratification on either UVA or MVA (Supplementary Methods).

This study had greatest power to detect associations with STAT. Each individual endpoint had fewer subjects and consequently reduced power, but despite this, nine individual toxicity endpoints showed more significant associations than expected by chance (Fig. 1 and Supplementary Fig. 2). The endpoints showing strongest evidence for SNP associations all relate to prostate toxicity (rectal bleeding, nocturnal frequency and rectal incontinence). QQ plots for all other endpoints are included in Supplementary Fig. 2, only the plot for breast shrinkage assessed by photographs showed no deviation from the null distribution.

### Selection of SNPs for replication

On the basis of the GWAS results from the first phase (including additional genotyping in the same samples of 103 SNPs for which the initial significant results had been based on imputed genotype dosages), 177 SNPs were selected for genotyping in a rapid replication phase. SNPs were selected for replication if UVA or MVA \( P < 10^{-6} \) and MAF \( \geq 0.05 \), or MAF \(< 0.05 \) and \( P < 10^{-4} \) by non-parametric test, or if located in or near possible candidate genes. Where multiple correlated SNPs were associated with toxicity \((R^2 > 0.9)\) the most strongly associated SNP was selected. The SNPs included in this replication stage included 23 SNPs displaying evidence for association with STAT in all patients, 63 SNPs associated with toxicity in breast and 91 SNPs associated with toxicity in prostate patients.

### Rapid replication of potential associations with STAT

None of the 23 potential STAT associations from RAPPER became more statistically significant on inclusion of data from the independent breast and prostate replication cohorts (Table 2, Supplementary Table 3). One of the strongest multivariable associations with STAT was with SNP rs13116075 at 4q28.3, close to CCRN4L \((P = 5.80 \times 10^{-5})\). The minor allele was associated with increased toxicity in all sets of patients tested. Each additional minor allele was associated with an increase in STAT with RR of 1.52.

### Rapid replication of potential associations with prostate endpoints

Table 3 and Supplementary Table 4 show SNPs potentially associated with late toxicity on MVA in the joint analysis of the

### Table 1

The number of observed SNPs with P-values below a given significance threshold compared to the number expected by chance alone.

<table>
<thead>
<tr>
<th>Significance level</th>
<th>Expected number of SNPs with MAF &gt; 0.05 below significance level</th>
<th>Observed number of SNPs with MAF &gt; 0.05 below significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prostate</td>
<td>Breast</td>
</tr>
<tr>
<td>( P &lt; 5 \times 10^{-8} )</td>
<td>0.9</td>
<td>0.4</td>
</tr>
<tr>
<td>( P &lt; 5 \times 10^{-7} )</td>
<td>8.7</td>
<td>4.3</td>
</tr>
<tr>
<td>( P &lt; 5 \times 10^{-6} )</td>
<td>86.7</td>
<td>43.4</td>
</tr>
<tr>
<td>( P &lt; 5 \times 10^{-5} )</td>
<td>867.3</td>
<td>433.6</td>
</tr>
</tbody>
</table>

* Expected number of SNPs at each P-value threshold were calculated by multiplying \( P \), \( N \) and \( M \), where \( P \) = P-value threshold, \( N \) = number of SNPs analysed with MAF > 0.05 and \( M \) = number of endpoints analysed, \( M_{prostate} = 8 \) and \( M_{breast} = 4 \).
Fig. 1. Q–Q plots of the observed chi$^2$ statistics obtained from linear regression of mean toxicity scores against the number of minor alleles (0, 1 or 2) or the imputed genotype dosage, using a 1-degree of freedom (df) trend test, versus the chi$^2$ statistics expected under the null hypothesis of no association. QQ plots are demonstrated for (a) overall (cancer-site-independent) toxicity in all patients, (b) telangiectasia in breast patients, (c) rectal bleeding, (d) nocturnal frequency and (e) rectal incontinence in prostate patients and genotype at imputed SNPs with MAF > 0.05 in univariable analysis (UVA) and multivariable analysis (MVA). Shaded regions are the 95% concentration bands that are formed by calculating the 2.5th and 97.5th centiles of the distribution of the order statistic under random sampling and the null hypothesis. The QQ plots display deviation from the null distribution at the tail (top 10%), suggesting that common SNPs are associated with risk of radiotherapy toxicity.

Fig. 2. Manhattan plots of observed log$_{10}$P-values vs. SNP position from univariable and multivariable analysis of (a) overall toxicity in all patients, (b) telangiectasia in breast patients, (c) rectal bleeding, (d) nocturnal frequency, (e) rectal incontinence in prostate patients. Several chromosome regions contain groups of SNPs which show evidence of association as shown by points representing small P-values ($P < 10^{-8}$) aligning almost vertically.
GWAS of late radiation toxicity

**Table 2** Results of SNPs potentially associated with overall toxicity on multivariable analysis in the RAPPER, LeND, RADILOGEN and Gene-PARE studies.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Gene</th>
<th>Ch Position</th>
<th>MAF</th>
<th>Beta RAP</th>
<th>SE RAP</th>
<th>P RAP</th>
<th>RR RAP</th>
<th>Beta ph2</th>
<th>SE ph2</th>
<th>P ph2</th>
<th>n comb</th>
<th>Beta comb</th>
<th>SE comb</th>
<th>P comb</th>
<th>RR comb</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs13116075</td>
<td>CCRN4L</td>
<td>4</td>
<td>0.15</td>
<td>0.28</td>
<td>0.06</td>
<td>1.21</td>
<td>10^(-6)</td>
<td>1.71</td>
<td>0.05</td>
<td>0.06</td>
<td>2219</td>
<td>0.17</td>
<td>0.04</td>
<td>5.80</td>
<td>10^(-5)</td>
</tr>
<tr>
<td>rs12243039</td>
<td>C10orf113</td>
<td>10</td>
<td>0.01</td>
<td>1.01</td>
<td>0.20</td>
<td>2.30</td>
<td>10^(-6)</td>
<td>3.31</td>
<td>0.37</td>
<td>0.34</td>
<td>0.28</td>
<td>1518</td>
<td>0.81</td>
<td>0.18</td>
<td>6.05</td>
</tr>
<tr>
<td>rs2185268</td>
<td>Near</td>
<td>18</td>
<td>0.33</td>
<td>0.18</td>
<td>0.04</td>
<td>4.99</td>
<td>10^(-5)</td>
<td>1.47</td>
<td>0.01</td>
<td>0.09</td>
<td>0.87</td>
<td>1515</td>
<td>0.15</td>
<td>0.04</td>
<td>2.14</td>
</tr>
<tr>
<td>rs718304</td>
<td>GABRB3</td>
<td>15</td>
<td>0.04</td>
<td>0.44</td>
<td>0.12</td>
<td>1.59</td>
<td>10^(-4)</td>
<td>1.95</td>
<td>0.42</td>
<td>0.23</td>
<td>0.07</td>
<td>1520</td>
<td>0.44</td>
<td>0.1</td>
<td>2.86</td>
</tr>
<tr>
<td>rs596917</td>
<td>NCR2</td>
<td>6</td>
<td>0.6</td>
<td>0.17</td>
<td>0.04</td>
<td>1.63</td>
<td>10^(-4)</td>
<td>1.44</td>
<td>0.03</td>
<td>0.05</td>
<td>0.58</td>
<td>2189</td>
<td>0.1</td>
<td>0.03</td>
<td>0.0016</td>
</tr>
<tr>
<td>rs2881208</td>
<td>SATB2</td>
<td>2</td>
<td>0.36</td>
<td>0.36</td>
<td>0.04</td>
<td>1.82</td>
<td>10^(-4)</td>
<td>1.43</td>
<td>0.05</td>
<td>0.09</td>
<td>0.55</td>
<td>1520</td>
<td>0.14</td>
<td>0.04</td>
<td>3.22</td>
</tr>
<tr>
<td>rs4496520</td>
<td>CCRN4L</td>
<td>3</td>
<td>0.17</td>
<td>0.62</td>
<td>0.06</td>
<td>4.37</td>
<td>10^(-4)</td>
<td>1.45</td>
<td>0.18</td>
<td>0.11</td>
<td>0.12</td>
<td>1520</td>
<td>0.19</td>
<td>0.05</td>
<td>1.13</td>
</tr>
<tr>
<td>rs4234649</td>
<td>C10orf113</td>
<td>3</td>
<td>0.19</td>
<td>0.06</td>
<td>0.47</td>
<td>10^(-4)</td>
<td>1.48</td>
<td>0.21</td>
<td>0.11</td>
<td>0.07</td>
<td>1518</td>
<td>0.42</td>
<td>0.02</td>
<td>1.51</td>
<td>10^(-4)</td>
</tr>
<tr>
<td>rs17798101</td>
<td>Near</td>
<td>18</td>
<td>0.34</td>
<td>0.62</td>
<td>0.06</td>
<td>8.16</td>
<td>10^(-4)</td>
<td>1.49</td>
<td>0.02</td>
<td>0.01</td>
<td>0.81</td>
<td>1816</td>
<td>0.15</td>
<td>0.05</td>
<td>0.0036</td>
</tr>
<tr>
<td>rs4849101</td>
<td>LINC00470</td>
<td>2</td>
<td>0.43</td>
<td>0.13</td>
<td>0.04</td>
<td>8.25</td>
<td>10^(-4)</td>
<td>1.37</td>
<td>-0.004</td>
<td>0.09</td>
<td>0.96</td>
<td>1520</td>
<td>0.11</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>rs10060885</td>
<td>ANKH</td>
<td>5</td>
<td>0.03</td>
<td>0.55</td>
<td>0.15</td>
<td>0.0037</td>
<td>2.50</td>
<td>0.13</td>
<td>0.2</td>
<td>0.52</td>
<td>1933</td>
<td>0.39</td>
<td>0.12</td>
<td>0.0015</td>
<td></td>
</tr>
</tbody>
</table>

Ch = chromosome, MAF = minor allele frequency, Beta = beta coefficient of regression, SE = standard error of the beta coefficient, P = P value, RAP = phase 1 RAPPER set, ph2 = LeND, RADILOGEN and Gene-PARE studies datasets, comb = combined result of phase 1 and phase 2; n comb = total number of patients with genotype and toxicity data.

RAPPER, RADILOGEN and Gene-PARE studies. Comparison of regression coefficients revealed seven SNPs where the minor allele was associated with increased toxicity in all three cohorts on MVA, giving final P-values between 6.79 × 10^-4 and 3.57 × 10^-7. Of note, each additional minor allele of SNP rs12353488 (P = 3.57 × 10^-7) was associated with RR of 1.52 of increased urinary frequency. Furthermore, SNP rs2788612, in KCND3 was associated with increased risk of rectal incontinence in RAPPER patients with P < 10^-10 and this decreased to P = 1.05 × 10^-12 on inclusion of RADILOGEN data. However, only one Spanish patient experienced rectal incontinence, so this result should be interpreted as preliminary.

**Rapid replication of potential associations with breast endpoints**

Table 4 and Supplementary Table 5 show SNPs potentially associated with late toxicity on MVA of RAPPER and LeND. Three SNPs were associated with telangietasia (rs16958536) or STAT1 associated with late toxicity on MVA of RAPPER and LeND. Three SNPs were considered to detect SNPs with tumour-site-independent effects, but in which the toxicity-associated SNPs could exert a clinical effect, within individual patients[20,24].

**Re-estimating optimal study power**

Sample sizes required to detect these potential associations at nominal genome-wide significance (z = 5 × 10^-8) were calculated. For tumour site-independent toxicity, a phase 1 study of 4750 mixed cancer patients would have 80% power to detect the observed difference of 0.17 of the SD of the residual of overall toxicity for a SNP with the same MAF as rs13116075 (0.15). Similarly, 3250 prostate cancer patients would give 80% power to detect a difference of 0.21 of the SD of the residual of urinary frequency for a SNP with the same MAF as rs12353488 (0.11). For overall breast toxicity, a study of 3000 patients would have 80% power to detect an effect size equal to that of rs2840044 (MAF = 0.33).

**Discussion**

The Q–Q and Manhattan plots presented here provide good evidence that common genetic variants are associated with a cancer patient’s risk of developing late radiotherapy toxicity. Originally it was hypothesised that genetic variation would be linked with toxicity generalised to all tissues. This study was therefore powered to detect SNPs with tumour-site-independent effects, but in fact found stronger associations with tumour-site-specific toxicity. Although the most significant associations in breast cancer patients were with overall toxicity rather than the individual endpoints, those SNPs were not associated with overall toxicity in prostate cancer patients. In prostate cancer patients, potential associations with individual endpoints were more numerous, despite the reduced power. These observations contrast with known rare genetic variants in DNA damage response genes, which have large effects on radiosensitivity and risk of toxicity, irrespective of site irradiated [11]. Conversely, it is consistent with clinical studies of the (lack of) association between multiple endpoints within individual patients [20,24].

None of the SNPs with potential associations to radiotherapy toxicity in this study have previously been reported as associated with prostate or breast cancer susceptibility [25–27], nor are they in genes previously considered to be candidates for radiotherapy toxicity [6]. There are, however, biologically plausible mechanisms by which the toxicity-associated SNPs could exert a clinical effect, although these are yet to be investigated. For example, rs2788612, associated with increased risk of late rectal incontinence (P = 1.05 × 10^-12), is located in KCND3 (potassium voltage-gated channel, Shal-related subfamily, member 3), expressed in smooth muscle, and might therefore be involved in sphincter function.

SNPs rs575018 and rs505994, close to FAM174A and ST8SIA4, were potentially associated with overall toxicity in breast cancer patients in UVA. Other SNPs in these genes have been reported to be potentially associated with body mass index (BMI) [28]. Breast volume is highly correlated with BMI and both are clearly associated with radiotherapy toxicity in the breast [20,29]. The significance of the associations with overall toxicity in the Cambridge breast IMRT patients, for whom BMI data were available, dropped to 10^-3 when adjusted for either BMI or breast volume, suggesting the apparent associations could be attributable to increased breast volume and/or BMI.

Our study also highlights the need for improved toxicity data collection for radiogenomics studies. Differences between treatment regimens and toxicity scales used in the UK, Spanish and USA cohorts may have reduced the power of the replication stage despite the use of STAT scores in the analysis. Harmonisation of toxicity data collected from different studies is challenging [30], particularly when data are collected at different time points and using different scales. For example, Gene-PARE collected data on urinary toxicity using the IPSS scale, whereas the LENT-SOMA and CTCAE scales were used in the RAPPER and RADILOGEN.

A limitation of this study was the sample size, which was restricted by the small numbers of subjects and studies that cur-
### Table 3
Results of SNPs potentially associated with toxicity on multivariable analysis in the RAPPER, RADIOGEN and Gene-PARE prostate cancer patients.

<table>
<thead>
<tr>
<th>Late toxicity</th>
<th>Gene</th>
<th>Ch</th>
<th>SNP</th>
<th>Position</th>
<th>MAF Beta RAP</th>
<th>SE RAP</th>
<th>P RAP</th>
<th>RR RAP</th>
<th>Beta ph2 RAP</th>
<th>SE ph2</th>
<th>P ph2</th>
<th>n comb</th>
<th>Beta comb RAP</th>
<th>SE comb</th>
<th>P comb</th>
<th>RR comb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decreased stream</td>
<td></td>
<td>6</td>
<td>rs1527708</td>
<td>72255359</td>
<td>0.06 0.77</td>
<td>0.12 1.27 \times 10^{-9}</td>
<td>6.46 -0.01</td>
<td>0.15 0.93</td>
<td>1800 0.36</td>
<td>0.1 2.87 \times 10^{-4}</td>
<td>1.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decreased stream</td>
<td></td>
<td>6</td>
<td>rs35663023</td>
<td>72200225</td>
<td>0.06 0.75</td>
<td>0.12 1.77 \times 10^{-9}</td>
<td>6.39 -0.008</td>
<td>0.15 0.95</td>
<td>1801 0.35</td>
<td>0.1 3.00 \times 10^{-4}</td>
<td>1.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectal incontinence</td>
<td>KCND3</td>
<td>1</td>
<td>rs2788612</td>
<td>112128405</td>
<td>0.05 0.89</td>
<td>0.15 2.91 \times 10^{-9}</td>
<td>7.98 -0.0009</td>
<td>0.0005 0.85</td>
<td>1124 0.66</td>
<td>0.09 1.05 \times 10^{-12}</td>
<td>9.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urine frequency</td>
<td></td>
<td>9</td>
<td>rs12353488</td>
<td>83705231</td>
<td>0.11 0.54</td>
<td>0.09 1.98 \times 10^{-8}</td>
<td>3.36 0.08</td>
<td>0.06 0.19</td>
<td>1800 0.21</td>
<td>0.05 3.57 \times 10^{-5}</td>
<td>1.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decreased stream</td>
<td>CLVS1</td>
<td>8</td>
<td>rs11785638</td>
<td>62622368</td>
<td>0.09 0.58</td>
<td>0.11 6.93 \times 10^{-8}</td>
<td>5.17 0.06</td>
<td>0.1 0.55</td>
<td>1410 0.25</td>
<td>0.07 6.79 \times 10^{-4}</td>
<td>1.60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urine incontinence</td>
<td>LINC00478</td>
<td>21</td>
<td>rs2832779</td>
<td>16690395</td>
<td>0.09 0.54</td>
<td>0.10 9.76 \times 10^{-8}</td>
<td>5.47 -0.07</td>
<td>0.13 0.59</td>
<td>1123 0.31</td>
<td>0.08 1.11 \times 10^{-4}</td>
<td>3.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urine frequency</td>
<td>Near UBR4</td>
<td>1</td>
<td>rs7527580</td>
<td>19251095</td>
<td>0.05 0.67</td>
<td>0.13 4.68 \times 10^{-7}</td>
<td>4.26 -0.01</td>
<td>0.17 0.93</td>
<td>1797 0.41</td>
<td>0.1 7.67 \times 10^{-5}</td>
<td>1.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectal incontinence</td>
<td>ERLIN1</td>
<td>10</td>
<td>rs11595238</td>
<td>101920399</td>
<td>0.04 0.84</td>
<td>0.17 4.84 \times 10^{-7}</td>
<td>7.06 -0.0009</td>
<td>0.0003 0.74</td>
<td>1123 0.39</td>
<td>0.08 1.93 \times 10^{-5}</td>
<td>6.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urine frequency</td>
<td>KIF13A</td>
<td>6</td>
<td>rs13198614</td>
<td>18046764</td>
<td>0.06 0.6</td>
<td>0.12 1.75 \times 10^{-6}</td>
<td>3.41 0.11</td>
<td>0.09 0.24</td>
<td>1788 0.28</td>
<td>0.07 1.38 \times 10^{-4}</td>
<td>1.62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urine frequency</td>
<td>Near VAMP4</td>
<td>1</td>
<td>rs11800109</td>
<td>168370756</td>
<td>0.07 0.5</td>
<td>0.11 7.17 \times 10^{-6}</td>
<td>3.11 0.11</td>
<td>0.06 0.068</td>
<td>1796 0.2</td>
<td>0.05 1.46 \times 10^{-4}</td>
<td>1.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urine frequency</td>
<td>Near AP15</td>
<td>11</td>
<td>rs12576830</td>
<td>42939051</td>
<td>0.07 0.48</td>
<td>0.11 3.21 \times 10^{-5}</td>
<td>4.46 0.16</td>
<td>0.12 0.19</td>
<td>1124 0.32</td>
<td>0.08 8.70 \times 10^{-5}</td>
<td>3.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectal incontinence</td>
<td>GNAS3</td>
<td>3</td>
<td>rs9832625</td>
<td>29304024</td>
<td>0.08 0.51</td>
<td>0.12 3.78 \times 10^{-5}</td>
<td>4.75 -0.0001</td>
<td>0.0002 0.64</td>
<td>1124 0.24</td>
<td>0.06 7.78 \times 10^{-5}</td>
<td>4.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectal incontinence</td>
<td>KCND3</td>
<td>1</td>
<td>rs12025303</td>
<td>112276452</td>
<td>0.11 0.41</td>
<td>0.1 4.44 \times 10^{-5}</td>
<td>4.41 -0.0001</td>
<td>0.0002 0.58</td>
<td>1124 0.2</td>
<td>0.05 7.40 \times 10^{-5}</td>
<td>4.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decreased stream</td>
<td>TSGA10</td>
<td>2</td>
<td>rs4850895</td>
<td>99102990</td>
<td>0.06 0.5</td>
<td>0.13 1.15 \times 10^{-4}</td>
<td>4.10 0.3</td>
<td>0.19 0.11</td>
<td>1802 0.42</td>
<td>0.11 1.83 \times 10^{-4}</td>
<td>1.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proctitis</td>
<td>PCDH9</td>
<td>13</td>
<td>rs17579023</td>
<td>65786896</td>
<td>0.08 0.40</td>
<td>0.11 3.08 \times 10^{-4}</td>
<td>2.48 0.13</td>
<td>0.07 0.07</td>
<td>1789 0.21</td>
<td>0.06 5.01 \times 10^{-4}</td>
<td>1.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectal incontinence</td>
<td>Between SAMD12 &amp; TNFRSF11B</td>
<td>8</td>
<td>rs952403</td>
<td>119770333</td>
<td>0.08 0.44</td>
<td>0.12 3.74 \times 10^{-4}</td>
<td>4.12 -0.0001</td>
<td>0.0002 0.66</td>
<td>1123 0.22</td>
<td>0.06 4.24 \times 10^{-4}</td>
<td>4.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectal incontinence</td>
<td>Between PLCZ1 &amp; PLEKHA5</td>
<td>12</td>
<td>rs1532054</td>
<td>18904175</td>
<td>0.1 0.38</td>
<td>0.11 3.92 \times 10^{-4}</td>
<td>3.90 -0.0009</td>
<td>0.0002 0.69</td>
<td>1124 0.21</td>
<td>0.06 2.10 \times 10^{-4}</td>
<td>4.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectal incontinence</td>
<td>ZFHX4</td>
<td>8</td>
<td>rs10957819</td>
<td>77900179</td>
<td>0.03 0.72</td>
<td>0.21 5.18 \times 10^{-4}</td>
<td>5.37 -0.0007</td>
<td>0.0004 0.84</td>
<td>1124 0.36</td>
<td>0.10 5.72 \times 10^{-4}</td>
<td>5.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rectal incontinence</td>
<td>FANCC</td>
<td>9</td>
<td>rs4647355</td>
<td>95158698</td>
<td>0.02 0.76</td>
<td>0.22 7.37 \times 10^{-4}</td>
<td>6.04 -0.0009</td>
<td>0.0005 0.86</td>
<td>1124 0.45</td>
<td>0.12 2.97 \times 10^{-4}</td>
<td>6.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ch = chromosome, MAF = minor allele frequency, Beta = beta coefficient of regression, SE = standard error of the beta coefficient, P = P value, RAP = phase 1 RAPPER set, ph2 = RADIOGEN + Gene-PARE datasets combined, comb = combined result of phase 1 and phase 2; n comb = total number of patients with genotype and toxicity data.

Proctitis was the only rectal toxicity endpoint measured in the USA Gene-PARE dataset.

RR = per allele relative risk of late toxicity endpoint for carriers of each SNP.
rently have sufficient follow-up. In addition the number of SNPs included in the rapid replication phase was limited by available funds. Post-hoc power calculations using the validation study sample size estimate that the study had essentially 100% power to validate the top association at a nominal significance level of 0.001 with prostate toxicity endpoints where the relative risk was 6.46 with a MAF of 0.06 corresponding to a 97% risk of toxicity in the 6% carrying the minor allele vs. 15% risk of toxicity in non-carriers. This suggests that the effect size estimate is biased in the training set, as would be expected. Therefore, to estimate the true underlying effect size with sufficient accuracy, a larger independent validation study is required.

The power calculations we have made will inform the design of further radiogenomic studies and based on the assumptions we have used, we can calculate that sample sizes of at least 3000 are required to reliably detect similar tissue-specific effects at nominal genome-wide significance. Given our limited sample size, it is likely that associations detected here, if real, represent some of the most strongly toxicity-related loci, although they may well overestimate the true effect sizes (the so-called “winner’s curse”). Our Q-Q plots indicate the likely existence of many more SNP associations with progressively smaller effects, which will require correspondingly larger sample sizes for confirmation. Such study sizes are achievable through collaboration within the RGC [7,3]. Furthermore, it would be interesting to extend the study to different tumour sites, as additional data become available.

In conclusion, this study provides good evidence that common genetic variants are associated with a cancer patient’s risk of developing radiotherapy toxicity. Thus the provision of personalised radiotherapy based on the patient’s genetic risk of toxicity is a real future possibility and so further investment in this field is merited to identify loci with clinically relevant effect sizes [6].

**Conflict of interest**

The authors disclose no potential conflicts of interest.

**Acknowledgements**

Dr. Gillian Barnett is funded the National Institute of Health Research (NIHR) and previously by a fellowship from Cancer Research UK and the Royal College of Radiologists [C26990/ A8740]. She also received funding from Addenbrooke’s Charitable Trust. Dr. Charlotte Coles and Prof. Neil Burnet are supported by the Cambridge National Institute for Health Research (NIHR) Bio-medical Research Centre. Dr. Alison Dunning is funded by Cancer Research UK [C8197/A10865] and the Joseph Mitchell Trust. Miss Jennifer Wilkinson, breast research radiographer, is funded by the Breast Cancer Campaign. Prof Catharine West is supported by Cancer Research UK and Experimental Cancer Medicine Centre (ECMC) funding. The collaborative group (RAPPER) is funded by Cancer Research UK (UK C1094/A11728), Laboratory infrastructure was funded by Cancer Research UK [C8197/A10123]. Prof. Søren Bentzen acknowledges support from the National Cancer Institute Grant no. P30 CA 134274-04.

The RACE study was funded by Cancer Research UK, and Prof. John Yarnold is supported by the Royal Marsden and ICR NHS Bio-medical Research Centre. The IMPORT LOW (CRUK/06/003) study was funded by Cancer Research UK [C1491/A6035]. The Institute of Cancer Research-Clinical Trials and Statistics Unit (ICR-CTSU) receives core funding from Cancer Research UK.

Prof. John Yarnold, Prof. David Dearnaley and Dr. Sarah Guillford are supported by the Royal Marsden NHS Foundation Trust and Institute of Cancer Research NIHR Biomedical Research Centre for Cancer. Dr. Guillford is supported by CRUK Programme Grant to the Section of Radiotherapy, Royal Marsden NHS Foundation Trust and Institute of Cancer Research, Grant no. C46/A2131. RT01 was supported by the UK Medical Research Council. CHHIP was supported by the Department of Health and Cancer Research UK; trial recruitment was facilitated within centres by the NIHR-funded National Cancer Research Network.

**Funding**

Funding was received from Cancer Research UK, the Royal College of Radiologists, Breast Cancer Campaign, the National Institute of Health Research, UK Medical Research Council, National Cancer Institute, the American Cancer Society, the Experimental Cancer Medicine Centre, the Instituto de Salud Carlos and Fondo Europeo de Desarrollo Regional and the Isabel Barreto programme from Xunta de Galicia and Fondo Social Europeo.
and Fondo Social Europeo, and was granted by an EMBO short term fellowship (ASTF 418–2011). RADIogen would like to acknowledge CESGA (Supercomputing Centre of Galicia, Santiago de Compostela, Spain; https://www.cesga.es/en) for providing supercomputing infrastructures.

Prof. Barry Rosenstein, Dr. Sarah Kerns and the Gene-PARE Study are funded by grants RSGT-05-200-01-CCE from the American Cancer Society, PC074201 from the Department of Defense and 1R01CA134444 from the National Institutes of Health.

The LeND cohort was funded by a grant from the Breast Cancer Campaign to Prof. R. Paul Symonds and Dr. Chris Talbot.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.radonc.2014.02.012.

References