

University of Groningen

**A genome-wide association study of radiotherapy induced toxicity in head and neck cancer patients identifies a susceptibility locus associated with mucositis**

Head and Neck Group of the Radiogenomics Consortium; Schack, Line M H; Naderi, Elnaz; Fachal, Laura; Dorling, Leila; Luccarini, Craig; Dunning, Alison M; Ong, Enya H W; Chua, Melvin L K; Langendijk, Johannes A

*Published in:*  
British Journal of Cancer

*DOI:*  
[10.1038/s41416-021-01670-w](https://doi.org/10.1038/s41416-021-01670-w)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2022

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Head and Neck Group of the Radiogenomics Consortium, Schack, L. M. H., Naderi, E., Fachal, L., Dorling, L., Luccarini, C., Dunning, A. M., Ong, E. H. W., Chua, M. L. K., Langendijk, J. A., Alizadeh, B. Z., Overgaard, J., Eriksen, J. G., Andreassen, C. N., & Alsner, J. (2022). A genome-wide association study of radiotherapy induced toxicity in head and neck cancer patients identifies a susceptibility locus associated with mucositis. *British Journal of Cancer*. <https://doi.org/10.1038/s41416-021-01670-w>

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**






If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## ARTICLE OPEN



## Genetics and Genomics

# A genome-wide association study of radiotherapy induced toxicity in head and neck cancer patients identifies a susceptibility locus associated with mucositis

Line M. H. Schack<sup>1,2</sup><sup>✉</sup>, Elnaz Naderi<sup>3,4</sup>, Laura Fachal<sup>5,6</sup>, Leila Dorling<sup>7</sup>, Craig Luccarini<sup>5</sup>, Alison M. Dunning<sup>5</sup>, The Head and Neck Group of the Radiogenomics Consortium\*, The Danish Head and Neck Cancer Group (DAHANCA)\*, Enya H. W. Ong<sup>8</sup>, Melvin L. K. Chua<sup>9,10</sup>, Johannes A. Langendijk<sup>3</sup>, Behrooz Z. Alizadeh<sup>3,4</sup>, Jens Overgaard<sup>1</sup>, Jesper Grau Eriksen<sup>1,2</sup>, Christian Nicolaj Andreassen<sup>1,2</sup> and Jan Alsner<sup>1</sup>

© The Author(s) 2022

**PURPOSE:** A two-stage genome-wide association study was carried out in head and neck cancer (HNC) patients aiming to identify genetic variants associated with either specific radiotherapy-induced (RT) toxicity endpoints or a general proneness to develop toxicity after RT.

**MATERIALS AND METHODS:** The analysis included 1780 HNC patients treated with primary RT for laryngeal or oro/hypopharyngeal cancers. In a non-hypothesis-driven explorative discovery study, associations were tested in 1183 patients treated within The Danish Head and Neck Cancer Group. Significant associations were later tested in an independent Dutch cohort of 597 HNC patients and if replicated, summary data obtained from discovery and replication studies were meta-analysed. Further validation of significantly replicated findings was pursued in an Asian cohort of 235 HNC patients with nasopharynx as the primary tumour site.

**RESULTS:** We found and replicated a significant association between a locus on chromosome 5 and mucositis with a pooled OR for rs1131769\**C* in meta-analysis = 1.95 (95% CI 1.48–2.41;  $p_{\text{pooled}} = 4.34 \times 10^{-16}$ ).

**CONCLUSION:** This first exploratory GWAS in European cohorts of HNC patients identified and replicated a risk locus for mucositis. A larger Meta-GWAS to identify further risk variants for RT-induced toxicity in HNC patients is warranted.

*British Journal of Cancer*; <https://doi.org/10.1038/s41416-021-01670-w>

## INTRODUCTION

Head and neck cancers (HNC) represent a group of patients with increasing loco-regional control and survival rates [1–3]. Curative treatment is usually made up of surgery, radiotherapy (RT) or a combination. Primary RT is often preferred for reasons of cosmesis and preserved organ function. However, RT also comes with the price of RT-induced toxicity. In HNC this toxicity is primarily characterised by mucositis, dysphagia, fibrosis of the skin, atrophy of the epithelial linings and xerostomia. The risk of RT-induced toxicity is influenced by various treatment and patient-related factors. A review of normal tissue complication (NTCP) models from 2010–2017 in standard fractionated RT of HNC identified the following parameters to be associated with RT-induced toxicity endpoints: Mean dose, maximal dose, irradiated volumes, the dose leading to a 50% complication

rate ( $D_{50}$ ), RT technique (Intensity-Modulated Radiation Therapy (IMRT) vs. 3D Conformal RT technique), gender, concomitant chemotherapy and stage of T-site [4]. However, increasing evidence also points to an individual inherent predisposition of developing toxicity [5] and in 2009, the Radiogenomics Consortium [6, 7] was established, facilitating research efforts in genetic variations associated with RT-induced toxicity. In studies of breast-, head and neck and prostate cancer cohorts, several SNPs associated with different morbidities after RT have been identified [8–11]. Two recent GWA studies explored RT-induced toxicity (temporal lobe brain injury and oral mucositis, respectively) in nasopharyngeal HNC patients [12, 13]. And another recent study of 37 patients identified associations between somatic genetic variants in HNC tumour tissue and severity of RT-induced morbidity, which is the first study to link genetic

<sup>1</sup>Department of Experimental Clinical Oncology, Aarhus University Hospital, Aarhus, Denmark. <sup>2</sup>Department of Oncology, Aarhus University Hospital, Aarhus, Denmark. <sup>3</sup>University of Groningen, University Medical Center Groningen, Department of Radiation Oncology, Groningen, The Netherlands. <sup>4</sup>University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, The Netherlands. <sup>5</sup>Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. <sup>6</sup>Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK. <sup>7</sup>Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. <sup>8</sup>Division of Medical Sciences, National Cancer Centre Singapore, Singapore, Singapore. <sup>9</sup>Division of Radiation Oncology, National Cancer Centre Singapore, Singapore, Singapore. <sup>10</sup>Oncology Academic Clinical Programme, Duke-NUS Medical School, Singapore, Singapore. \*Lists of authors and their affiliations appear at the end of the paper. ✉email: [schack@oncology.au.dk](mailto:schack@oncology.au.dk)

Received: 3 December 2020 Revised: 21 November 2021 Accepted: 6 December 2021

Published online: 17 January 2022

variants in the tumour to surrounding normal tissue morbidity [14]. However, GWASs with respect to radiation-induced toxicity in non-nasopharyngeal HNC patients remain to be investigated. The present study aimed to identify genetic variants associated with either specific RT-induced toxicity endpoints or a general proneness to develop toxicity after RT across endpoints. We conducted a two-stage genome-wide study comprising a total of 1780 European HNC patients with larynx and oro/hypopharynx as the primary tumour site. Further validation was pursued in an Asian cohort of HNC patients with nasopharynx as the primary tumour site.

## MATERIALS AND METHODS

We performed a two-stage GWAS including 1780 HNC patients with oropharyngeal/hypopharyngeal or laryngeal cancer from two European cohorts. In a discovery study including 1183 patients, common genetic variants were tested for their association with either specific radiation-induced morbidities or with a general susceptibility to develop toxicity after RT across endpoints. In a replication study including 597 patients, the genome-wide significant variants identified in the discovery study were tested in an independent cohort of comparable HNC patients and if successfully replicated, variants were meta-analysed. A subsequent attempt to replicate significant findings in a cohort of 235 Asian HNC patients with nasopharyngeal cancer was made.

### Discovery study

**DAHANCA cohorts.** Patients in the discovery study were treated within The Danish Head and Neck Cancer Group (DAHANCA) protocols in Denmark for head and neck squamous cell carcinoma in the period of 2000–2011. Eligibility criteria were stringently set to ensure a uniform cohort as follows: Histologically verified pharyngeal or laryngeal squamous cell carcinomas T1-4N0-3M0. Curative intent treatment regimen, consisting of primary RT ± concomitant treatment, >600 days recurrence-free (curative single lymph node extirpation on the neck was accepted) and available biological tissue with germline DNA. Patients were excluded by the death of any cause or switch to palliative treatment regimen during primary RT. Patients were censored by recurrence date in case of recurrence after 600 days. RT doses were 66, 68 (accelerated 6 fx/week regime) or 70 Gy (standard 5 fx/week regime). Fractionation doses were 2 Gy/fx. Concomitant treatment included weekly cisplatin and Nimorazole. Cisplatin was offered to patients with N1–3 disease from 2007, dosage 40 mg/m<sup>2</sup> once weekly during RT. Nimorazole was administered to all patients except patients with glottic laryngeal cancers T1N0M0. RT technique was IMRT from 2006. Before this, 3D-Conformal RT was applied. The 1183 patients included were from the following cohorts:

**DAHANCA10 (N=81):** In a randomised phase III study, the effect of Darbepoetin-alfa in anaemic patients (haemoglobin < 14.5 g/dl for men and < 13.5 g/dl for women) was analysed with loco-regional failure as the primary endpoint. Patients were eligible by the following criteria: Squamous cell carcinoma of the larynx, oropharynx, hypopharynx or oral cavity T1-4N0-3M0 (excluding stage I glottic carcinomas), WHO performance score 0–2, age ≥18 years. The study was interrupted after an interim analysis showing the inferiority of the Darbepoetin-alfa arm [15].

**DAHANCA19 (N=485):** In a randomised, open-label phase III trial in 2007–2012, the effect of zalutumumab, an EGFR-inhibitor, was analysed with loco-regional failure as the primary endpoint. Patients were eligible by histologically verified squamous cell carcinoma of the oral cavity, larynx, oropharynx or hypopharynx, stage T1-4N0-3M0 (excluding stage I + II glottic carcinomas), WHO performance score 0–2, age ≥18 years [16].

**DAHANCA Prospective biobank protocol (N=617):** Consecutive patients with available biobank material, age ≥18 years, treated according to standardised national DAHANCA guidelines [17].

**Endpoints.** Follow-up data were prospectively registered weekly during RT and after completion of RT course, at 2 weeks and at 3, 6, 12, 24, 36, 48 and 60 months. RT-induced toxicity in the present study was scored as the maximal grade at any time during treatment or follow-up.

Acute toxicity endpoints comprised acute dysphagia and mucositis between week 2 after RT start and 14 days after treatment completion. Late endpoints comprised late dysphagia, xerostomia, mucosal atrophy and skin fibrosis between 600 days and 5 years after treatment completion. Under the hypothesis that mucosal atrophy and skin fibrosis are manifestations of similar biological pathway processes, these entered a combined fibrosis/atrophy endpoint. Additionally, tube feeding at 6 months was included as a binary

variable. This endpoint represents a composite multifactorial endpoint dependent on not only RT- and patient-related factors but also on factors such as inter-institutional variation in the availability of specialised knowledge in tube feeding and cultural differences in when to start and when to end tube feeding. It was included as a surrogate marker for severe toxicity after RT.

Acute and late endpoints were registered ordinarily according to standardised DAHANCA forms, which are comparable with LENT-SOMA scales [18]. For the purpose of the present study, endpoints were dichotomised into “moderate-severe” and “severe”. Grading and cut-off points are available in Supplementary Table 1. Severe mucositis (corresponding to ulceration) was not tested independently due to a low number of events.

With the purpose of normalising and combining acute, late and global endpoints, Standardised Total Average Toxicity (STAT) scores were estimated aggregating ordinal endpoints (Supplementary Table 1). This method was previously described elsewhere [19]. STAT<sub>acute</sub> included acute dysphagia and mucositis. STAT<sub>late</sub> included late dysphagia, xerostomia, fibrosis and atrophy. STAT<sub>global</sub> combined STAT<sub>late</sub> and STAT<sub>acute</sub>. Tube feeding was not included in the STAT scores.

**Covariates.** Available covariates were age, sex, total RT dose, concomitant chemotherapy, protocol, and a surrogate for irradiated volume (Supplementary Table 1). As dose-volume histograms were not available, a trichotomised surrogate marker for irradiated volume was estimated from the cancer site and TNM stage. The rationale was as follows: Irradiated volumes depended on site and stage. Patients with T1a-T1bN0M0 glottic laryngeal carcinomas received small box fields to the larynx without an elective volume. The remaining patients with N0 disease received bilateral (or unilateral for tonsillar tumours without the involvement of soft palate or base of tongue) elective volumes by the following criteria: Oropharyngeal and laryngeal cancers without subglottic disease: lymph node levels II and III. Hypopharyngeal cancers: levels II, III and IV. Subglottic laryngeal cancers: levels II, III, IV and VI. In general, patients with N1–3 disease received a larger irradiated volume than patients with N0 disease as the positive nodes were irradiated to the prescribed tumour dose and elective volumes could be expanded beyond those described above. Based on this, the volume surrogate cut-offs were: Volume 1 = T1a-1bN0M0 glottic laryngeal carcinomas, Volume 2 = all other TxN0 sites, Volume 3 = TxN1-3 carcinomas.

**Genotyping, imputation, quality control and bioinformatics.** Germline whole-genome DNA was extracted from buffy coats and SNP genotyping was done using the Infinium OncoArray-500K Bead Chip (Illumina Inc., San Diego, USA). Quality control processes adhered to OncoArray guidelines [20] and are available in Supplementary Document and Supplementary Table 2. Haplotype estimation (phasing) was done using SHAPEITv2(r837). Imputation was done with IMPUTE2 using the last phase of 1000 Genomes, mapped to build GRCh37/hg19 as a reference.

**Statistical analysis.** The cohort was analysed as a pooled dataset of 1183 patients from DAHANCA10, DAHANCA19 and the DAHANCA Prospective biobank protocol as they were eligible by the criteria set. Covariates considered clinically relevant were tested against endpoints in a univariate logistic regression model for binary endpoints and a linear regression model for STAT (continuous) scores (Supplementary Fig. 1). If significant at an  $\alpha = 0.05$  level they entered multivariate analysis and by the method of backwards stepwise selection, final models were created. The total dose was included in the final models for all endpoints. The volume surrogate entered models for all endpoints except severe late dysphagia and tube feeding. Concomitant chemotherapy entered models for acute endpoints. Sex entered models for xerostomia and STAT scores. The protocol entered models for xerostomia, STAT<sub>late</sub> and STAT<sub>global</sub>. Principal components were introduced stepwise in analysis and included if significant ( $\alpha = 0.05$ ).

Associations between genotypes and endpoints were tested using a logistic regression model for binary endpoints and a linear regression model for STAT (continuous) scores. All analyses were adjusted for covariates in a case-control design. A log-additive genetic model was assumed for single endpoints and a linear model was assumed for STAT scores. SNPs were included as a number of effect alleles or imputed genotype dosage resulting in a per-allele odds ratio (OR) or a regression coefficient ( $\beta$ ) for each SNP. The significance level was set to  $p < 5 \times 10^{-8}$  and thus was not adjusted for testing multiple endpoints. This choice was based on including genetic variants suggestively significant in the replication phase.

In a Bayesian approach, we estimated the posterior probability of a false discovery ( $P_{\text{Bayes}}$ ), assuming a prior probability =  $10^{-4}$  and a prior variance =  $0.35^2$  [21, 22]. Top variants with a posterior probability <10% were considered a plausible true positive discovery. Data preparation and

statistical analyses were carried out using Stata13 (StataCorp, TX USA), R Statistical Package v3.3.0 [23] and SNPTTEST [24]. The study is reported in compliance with the STROGAR guidelines for radiogenomic studies [25].

**Power analysis.** Power was estimated using the Genetic Association Study (GAS) Power Calculator [26] and is illustrated for selected endpoint frequencies

in Supplementary Fig. 2. For the discovery phase, in general associations with ORs <1.7 were not likely to be identified with a power of 80%.

### Replication studies

**Dutch cohort.** The study population consisted of 1102 HNC patients eligible by the same criteria as in the discovery study and treated with

**Table 1.** Patient characteristics of discovery and replication cohorts.

|                             | Discovery study       |   |                    |                    | Replication studies   |                           |
|-----------------------------|-----------------------|---|--------------------|--------------------|-----------------------|---------------------------|
|                             | All patients<br>N (%) | DAHANCA Prosp.<br>biobank protocol<br>N (%) | DAHANCA10<br>N (%) | DAHANCA19<br>N (%) | Dutch cohort<br>N (%) | Singapore<br>cohort N (%) |
| Number of patients          | 1183 (100)            | 617 (52)                                    | 81 (7)             | 485 (41)           | 597 (100)             | 235 (100)                 |
| Age in years                |                       |   |                    |                    |                       |                           |
| Median (range)              | 60 (27–90)            | 61 (27–90)                                  | 56 (35–80)         | 59 (32–80)         | 63 (20–89)            | 51 (21–91)                |
| Sex                         |                       |   |                    |                    |                       |                           |
| Male                        | 940 (79)              | 481 (78)                                    | 63 (78)            | 396 (82)           | 399 (67)              | 182 (77)                  |
| Female                      | 243 (21)              | 136 (22)                                    | 18 (22)            | 89 (18)            | 198 (33)              | 53 (23)                   |
| Dose                        |                       |   |                    |                    |                       |                           |
| 66 Gy                       | 765 (65)              | 466 (76)                                    | 56 (69)            | 243 (50)           | 496 (83)              | 0 (0)                     |
| 68 Gy                       | 383 (32)              | 148 (24)                                    | 25 (31)            | 210 (43)           | 35 (6)                | 0 (0)                     |
| 70 Gy                       | 35 (3)                | 3 (0)                                       | 0 (0)              | 32 (7)             | 66 (11)               | 235 (100)                 |
| Volume surrogate            |                       |   |                    |                    |                       |                           |
| T1a N0 M0<br>glottic larynx | 159 (13)              | 159 (26)                                    | 0 (0)              | 0 (0)              | 78 (13)               | 13 (1)                    |
| Tx N0 M0                    | 215 (18)              | 145 (23)                                    | 19 (23)            | 51 (11)            | 221 (37)              | 11 (18)                   |
| Tx N1–3M0                   | 809 (68)              | 13 (51)                                     | 62 (77)            | 434 (89)           | 298 (50)              | 211 (22)                  |
| Chemotherapy                |                       |   |                    |                    |                       |                           |
| No                          | 714 (60)              | 505 (82)                                    | 81 (100)           | 128 (26)           | 445 (75)              | 53 (23)                   |
| Yes                         | 469 (40)              | 112 (18)                                    | 0 (0)              | 357 (74)           | 152 (25)              | 182 (78)                  |
| Endpoint frequency          |                       |   |                    |                    |                       |                           |
| Acute dysphagia             |                       |   |                    |                    |                       |                           |
| Moderate/<br>severe         | 730 (62)              | 304 (49)                                    | 53 (65)            | 373 (77)           | —                     | —                         |
| Severe                      | 410 (35)              | 153 (25)                                    | 29 (36)            | 228 (47)           | —                     | —                         |
| Mucositis                   |                       |   |                    |                    |                       |                           |
| Moderate/<br>severe         | 870 (74)              | 395 (64)                                    | 63 (78)            | 412 (85)           | 170 (28)              | 49 (21)                   |
| Severe                      | 41 (3)                | 19 (3)                                      | 3 (4)              | 19 (4)             | —                     | —                         |
| Late dysphagia              |                       |   |                    |                    |                       |                           |
| Moderate/<br>severe         | 143 (12)              | 54 (9)                                      | 14 (17)            | 75 (15)            | —                     | —                         |
| Severe                      | 60 (5)                | 23 (4)                                      | 4 (5)              | 33 (7)             | —                     | —                         |
| Fibrosis                    |                       |   |                    |                    |                       |                           |
| Moderate/<br>severe         | 628 (53)              | 97 (48)                                     | 54 (67)            | 277 (57)           | —                     | —                         |
| Severe                      | 166 (14)              | 66 (11)                                     | 12 (15)            | 88 (18)            | —                     | —                         |
| Fibrosis/atrophy            |                       |   |                    |                    |                       |                           |
| Moderate/<br>severe         | 822 (69)              | 404 (65)                                    | 67 (83)            | 351 (72)           | —                     | —                         |
| Severe                      | 202 (17)              | 81 (13)                                     | 15 (19)            | 106 (22)           | —                     | —                         |
| Xerostomia                  |                       |   |                    |                    |                       |                           |
| Moderate/<br>severe         | 755 (64)              | 368 (60)                                    | 67 (83)            | 320 (66)           | —                     | —                         |
| Severe                      | 276 (23)              | 149 (24)                                    | 39 (48)            | 88 (18)            | —                     | —                         |
| Tube feeding at<br>6 months | 78 (7)                | 41 (7)                                      | 8 (10)             | 29 (6)             | —                     | —                         |

definitive or postoperative RT or Chemo-RT from 2007 to 2017. All patients were included in the UMCG-HNC prospective data registration program, in which all baseline patient-, tumour- and treatment characteristics, as well as acute and late toxicity, were prospectively scored. Patients were followed up to 5 years after completion of treatment (NCT02435576 at clinicaltrials.gov).

**Endpoints, SNPs and covariates.** A replication study was performed on genetic variants reaching GWAS significance in the discovery phase and including suggestively significant associations with  $p < 10^{-5}$  for the specific endpoint. The replication study included the same covariates associated with the endpoint as in the discovery study. In the discovery cohort, acute toxicity was assessed weekly up to the second week after completion of treatment, whereas in the replication cohort, acute toxicity was assessed weekly during radiation and up to 5 weeks after completion of RT. The included endpoint, mucositis, was defined as moderate or severe on at least one time point between the second week after the start of RT and the fifth week after completion of RT.

**Genotyping, imputation, quality control and bioinformatics.** Out of 1,102 patients, 607 were genotyped on Illumina human hap 550k v.3.0 and 495 on Illumina global screening array. Quality control procedures were performed using standard procedures. First, samples and SNPs with call rates below 98% and MAF  $< 1\%$  were removed. We checked the ethnicity of subjects using multidimensional scaling (MDS) clustering of our samples with HapMap Phase3 individuals using EIGENSTRAT [27]. Samples that deviated more than 3 SD from the mean of their closest clusters were removed. Second, genotypes that passed the QC were imputed on the backbone of Haplotype Reference Consortium reference panel (HRC) version R1.1. Third, we further performed post-imputation QC involved removing SNPs with an imputation quality (info) score of  $R^2 < 0.3$ , with a MAF  $< 0.01$ , SNPs that had a discordant MAF (maximum allowed difference  $< 0.15$ ) compared to the reference panel, and strand ambiguous AT/CG SNPs and multi-allelic SNPs. Two imputed datasets were then merged together using PLINK (version 1.90). As the genotyping array differed between the discovery and the replication cohorts, some variants from the discovery phase were not reproduced and thus not analysed in the replication phase.

**Statistical analysis.** We used logistic regression analysis and included covariates similar to the discovery study. The top ten principal components were included. The effect estimates were reported as OR with their corresponding 95% CI and associated  $p$ -values. The summary results of both discovery and replication studies were meta-analysed using an inverse variance-weighted fixed effects model as implemented in METAL (version 2011-03-25) [28]. A  $p$ -value  $< 0.05$  was considered statistically significant for replication analysis. For the meta-analysis, we maintained a significance threshold of  $p < 5 \times 10^{-8}$ . In all cases, it was a part of the quality control to check that directionality of effect and that MAFs were comparable between discovery and replication phases.

**Singapore cohort.** Genetic variants identified in the discovery phase and successfully replicated were independently tested in an Asian cohort of 235

HNC patients with nasopharyngeal carcinoma treated with curative intent with Radiotherapy (70 Gy/35 fractions)  $\pm$  chemotherapy (clinicaltrials.gov: NCT04340024). Germline DNA was whole-exome sequenced using the Agilent V6 + UTR NovaSeq platform. Genotyping, quality control and bioinformatics was performed using Genome Analysis Tool KIT (GATK v 4.1.8.0) and adhering to the GATK best practices workflow (<https://github.com/gatk-workflows/gatk4-germline-snp-indels>). Germline variants were called for each sample using HaplotypeCaller [29]. This was followed by a joint genotyping step via GATK GenotypeGVCF across the whole cohort. The cohort followed the same protocol as the Dutch cohort in terms of endpoints, covariates included, statistical analysis methods and  $p$ -value significance cut-off.

## RESULTS

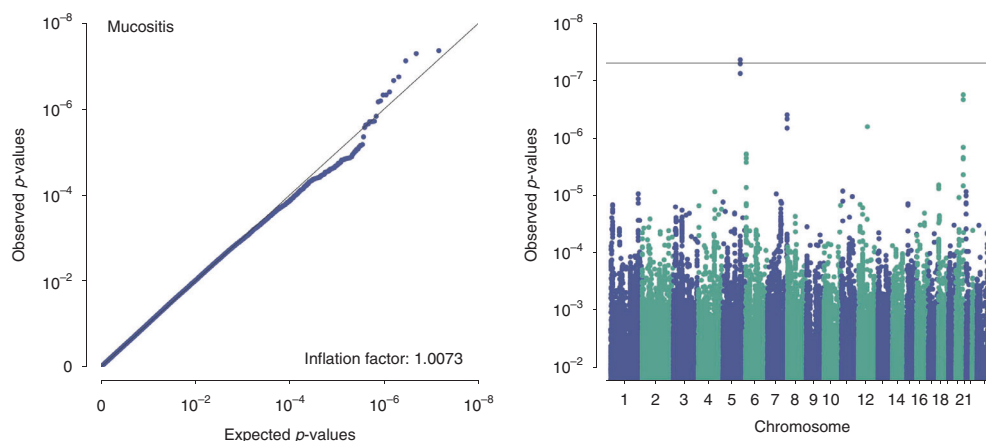
### Discovery study

**Genotyping and patient characteristics.** Genotyping and imputation resulted in a dataset of 14,640,905 common variants in 1237 patients including duplicate controls. After Quality Control processes, 7,178,424 variants (SNPs, deletion/insertion variations (indels) and copy number variations (CNVs)) and 1183 patients remained in analysis. Stepwise outcomes from QC processes are available in Supplementary Table 2. Clinical characteristics are displayed in Table 1.

**GWAS-QQ and Manhattan plots.** The 7,178,424 SNPs were tested for associations with acute dysphagia, mucositis, late dysphagia, xerostomia, skin fibrosis, fibrosis/atrophy, tube feeding at 6 months, STAT<sub>acute</sub>, STAT<sub>late</sub> and STAT<sub>global</sub>. The first 7 endpoints were analysed as binary outcomes, and the first six of these were analysed at two different cut points, 'moderate-to-severe' and 'severe' (Supplementary Table 1). Due to a low number of severe events, mucositis was only evaluated as 'moderate-to-severe'. The analysis identified a significant association for mucositis with QQ and Manhattan plots shown in Fig. 1. QQ and Manhattan plots for all endpoints are shown in Supplementary Fig. 3.

In moderate-to-severe acute dysphagia, an imputed single SNP, rs2024705 on chromosome X, reached genome-wide significance. Imputation quality was rather poor (info = 0.44). This SNP represents an intron variant in LINC00632 (long intergenic non-protein-coding RNA 632) and was considered a potential false-positive finding.

**Mucositis.** Three SNPs (rs28419191, rs10875554 and rs1131769) tagged a locus on chromosome 5 that reached genome-wide significance for association with radiation-induced mucositis (Table 2 and Fig. 2). The strongest association was found for effect allele C (allele frequency 0.87) of rs28419191 with OR = 2.27; 95% CI 1.70–3.05;  $p = 4.4 \cdot 10^{-8}$ ;  $p_{\text{Bayes}} = 0.068$ . The SNPs were well imputed (imputation info score range 0.95–0.96) and in close LD



**Fig. 1** GWAS results for moderate/severe mucositis. QQ plot (left) and Manhattan plot (right).



**Table 2.** Summary results of the association between a locus on chromosome 5 and RT-induced mucositis in discovery, replication and meta-analysis.

| SNP        | Discovery study (N = 1183) |                       |                 |                 | Replication study (Dutch) (N = 597) |  |       |       | Meta-analysis (N = 1780) |           |           |                         |                |
|------------|----------------------------|-----------------------|-----------------|-----------------|-------------------------------------|--|-------|-------|--------------------------|-----------|-----------|-------------------------|----------------|
|            | Effect allele              | MA (MAF) <sup>a</sup> | OR <sup>b</sup> | SE <sup>c</sup> | p-value                             | MA (MAF) <sup>a</sup>                    | OR    | SE    | p-value                  | Pooled OR | Pooled SE | Pooled p-value          | I <sup>2</sup> |
| rs28419191 | C                          | T (0.13)              | 2.272           | 0.344           | 4.4 × 10 <sup>-8</sup>              | T (0.13)                                 | 1.550 | 0.392 | 0.035                    | 1.958     | 0.259     | 3.6 × 10 <sup>-14</sup> | 47.8%          |
| rs1131769  | C                          | T (0.13)              | 2.209           | 0.329           | 7.6 × 10 <sup>-8</sup>              | T (0.12)                                 | 1.650 | 0.349 | 0.016                    | 1.946     | 0.239     | 4.3 × 10 <sup>-16</sup> | 26.4%          |
| rs10875554 | C                          | T (0.13)              | 2.248           | 0.331           | 5.1 × 10 <sup>-8</sup>              | Data not available in replication cohort |       |       |                          |           |           |                         |                |

<sup>a</sup>MA (MAF) Minor allele (frequency).<sup>b</sup>OR odds ratio.<sup>c</sup>SE standard error.

with each other (Supplementary Fig. 4A). The locus spans an area from 4 kb upstream of the ECSCR gene (rs28419191 and rs10875554) to an exon in the *STING1* gene, where rs1131769T > C represents a missense variant changing the codon from H (His) > R (Arg). The locus is in cis-eQTL with *DNAJC18*, *PROB1*, *SLC23A1*, *SPATA24* and in sQTL with *STING1* [30]. *DNAJC18*, *PROB1*, *SLC23A1* and *SPATA24* have a similar expression pattern across tissue types and *ECSCR* and *STING1* share a different expression pattern (Supplementary Fig. 5).

*ECSCR* (Endothelial Cell Surface expressed Chemotaxis and apoptosis Regulator protein, previously known as Endothelial Cell-Specific Molecule 2 (ECSM2)) is expressed mainly in endothelial cells and blood vessels [31] and is involved in the regulation of chemotaxis and apoptosis through modulation of various pathways including EGFR, VEGF and FGFR signaling [32, 33].

*STING1* (STimulator of INterferon response cGAMP interactor 1, previously known as transmembrane protein 173 (TMEM173)) has several splice variants. It is critically involved in the  $\beta$ -interferon synthesis and inflammation/immune system function [34] and non-canonical activation of a STING signalling pathway has been observed following DNA damage [35].

*DNAJC18* is a member (C18) of the DnaJ heat shock protein family (Hsp40) [36]. The transcripts of this family function as chaperone molecules protecting other cellular proteins across pro- and eukaryotic species [37].

*PROB1* (PROline rich Basic protein 1) is a protein-coding gene [36]. Its function in humans is not known, although one study has identified a possible association between alterations in this gene and the eye disorder keratoconus [38].

*SLC23A1* (SoLute Carrier family 23 member 1, formerly SVCT1) is a member of the solute carrier (SLC) group of proteins and is involved in the transport of ascorbic acid (Vitamin C) across the cell membrane [39, 40].

*SPATA24* (spermatogenesis associated 24) also a protein-coding gene, expressed in several tissues, in particular in the testes and prostate. It is preserved in several mammals where it plays a role in Cell differentiation as well as spermatogenesis [39].

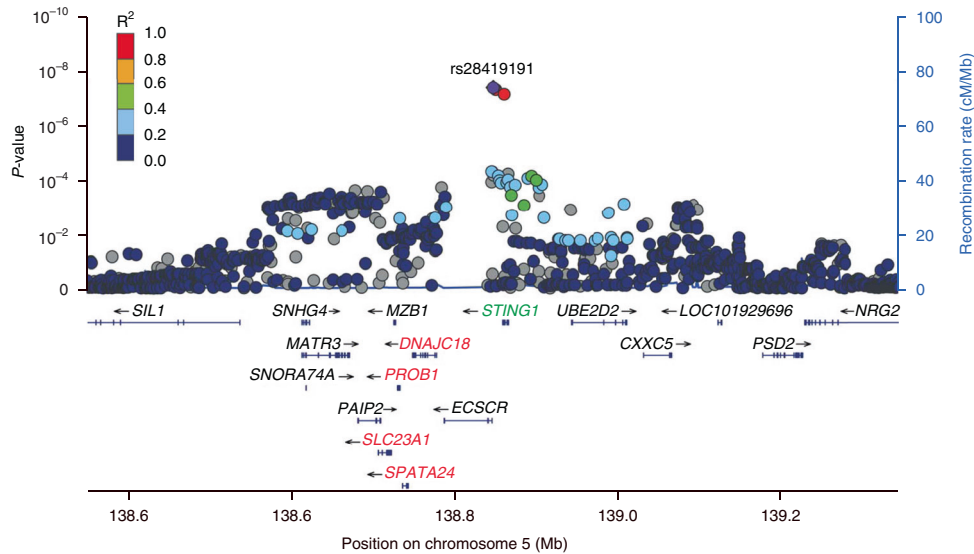
The study did not identify other loci reaching genome-wide significance. A near-significant association was found between a locus on chromosome 6 and STAcute, available in Supplementary Figs. 4B, 6 and in Supplementary Table 3, where a full list of variants reaching  $p$ -values < 10<sup>-5</sup> is available for all endpoints.

## Replication study

### Dutch cohort

Genotyping and patient characteristics: Among 1102 HNC samples, 957 patients and 6,334,277 SNPs passed imputation process and quality controls. The Dutch cohort includes HNC patients treated with both definitive and postoperative RT. Phenotypic quality control yielded 597 patients who were treated with definitive RT and met the remaining eligibility criteria matching the discovery study. Table 1 presents the characteristics of the Dutch replication population.

Replication and meta-analysis: The discovery study identified three variants significantly and 27 variants suggestively associated with mucositis. Two out of 3 genome-wide significant and 21 out of 27 suggestively significant variants were available for analysis in the replication study. We found the rs28419191\*C allele with an OR of 1.55 (95% CI 1.03–2.32;  $p < 0.035$ ) and rs1131769\*C allele with an OR of 1.65 (95% CI 1.10–2.47;  $p < 0.015$ ) significantly associated with increased risk of mucositis in the replication cohort (Table 2). None of the 21 suggestive SNPs from the discovery study was associated with mucositis in the replication cohort (Supplementary Table 3, sheet "Mucositis"). Next, we performed a meta-analysis on summary data for the discovery and replication studies, the combined subjects amounting to 1780 patients. The pooled OR for rs28419191\*C allele was 1.96 (95% CI



**Fig. 2 Locus Zoom plot for a locus on chromosome 5 associated with moderate/severe mucositis.** Variants in orange or red colour are in Linkage Disequilibrium with the specified SNP rs28419191. The locus is in cis-eQTL with genes marked in red and in sQTL with gene marked in green.

1.45–2.46;  $p_{\text{meta-analysis}} = 3.67 \times 10^{-14}$ ) and that for rs1131769\**C* allele was 1.95 (95% CI 1.48–2.41;  $p_{\text{meta-analysis}} = 4.34 \times 10^{-16}$ ) and thus significantly associated to mucositis (Table 2).

**Singapore cohort:** In 235 samples, 49 patients with nasopharyngeal carcinoma experienced moderate-severe mucositis. Patient characteristics are presented in Table 1. In the locus eligible for analysis, rs28419191 and rs10875554 were unavailable for analysis and only rs1131769, was therefore analysed. This variant failed to replicate with an OR = 0.94 (95% CI 0.46–1.92;  $p = 0.86$ ). The genotyping quality was good (0.99). However, the effect allele C had a lower frequency than in the DAHANCA and Dutch cohorts with 0.77 vs 0.87 and 0.87, respectively.

## DISCUSSION

Many studies in radiogenomics have been conducted since the turn of the millennium and several genomic loci have been identified as significantly associated with RT-induced toxicity. Up to now, breast, prostate, gynaecological and lung cancer cohorts are represented in hypothesis-driven studies of associations between SNPs and RT-induced toxicity. In non-hypothesis-driven GWA studies, breast and prostate cancer cohorts have been explored [8–11] and two fairly recent studies explored loci associated with temporal lobe injury [12] and oral mucositis [13] reaching  $p$ -values of  $10^{-7}$  after RT for nasopharyngeal carcinoma.

The present study is the first GWAS study to explore associations with RT-induced toxicity in non-nasopharyngeal HNC cohorts. Previous studies in breast cancer cohorts have indicated that genetic variations may predispose to both an endpoint-specific as well as a more general susceptibility of developing RT-induced toxicity [41, 42] and hence supports the idea that bringing a cancer site with new RT-induced toxicity data to the research field of radiogenomics may contribute both site-specific and general value. The study identified and replicated a locus on chromosome 5 in the *STING1* gene to be significantly associated with RT-induced mucositis and meta-analysis confirmed this association. The expression pattern falls into two groups, rendering it likely that this locus could be involved by either splice variants in the *STING1* pathway with which the locus is in sQTL or the group of genes (*DNAJC18*, *PROB1*, *SLC23A1* and

*SPATA24*) with which the locus is in cis-eQTL (Supplementary Fig. 5).

An effort was done to replicate this finding in an Asian ancestry cohort from Singapore using whole-exome sequencing data. However, it failed to replicate due to most likely a different tumour site (nasopharynx) bringing a different distribution of RT-induced toxicity, a low sample size and the fact that only one variant, rs1131769, was available for analysis and this had a lower minor allele frequency than in the European ancestry cohorts studied here. In accordance with this chain of reasoning, the loci reaching low  $p$ -values in the present study were different from the ones identified in [13] exploring RT-induced oral mucositis in patients with both a different tumour site and ancestry (nasopharyngeal cancers in Chinese patients).

We did not identify SNPs significantly associated with any other acute and late endpoints. This may own to the sample size of 1183 patients in the discovery study, which was a limitation to the study as it rendered the identification of low-penetrance genetic variants less likely.

The significance level in the discovery phase was set to  $p < 5 \times 10^{-8}$  and thus was not adjusted for testing multiple endpoints. This was a pragmatic choice that follows consensus in GWA studies and further was based on the considerations that the replication phase included genetic variants suggestively significant, and as the multiple endpoints tested are not individually independent, a Bonferroni estimate of a corrected  $p$ -value may increase the risk of false-negative discoveries to a larger extent than it will decrease the risk of false-positive findings. However, we do acknowledge this limitation and recommend that the finding in our study should be validated in other independent comparable cohorts before considered conclusive.

It is our hope that this study can serve such future independent studies as summary data for associations reaching  $p$ -values  $< 10^{-5}$  are published in Supplementary Table 3 and may serve as a reference for directionality, effect size and endpoint association. By design, the present study analysed the association between individual SNPs and endpoints. The complexity of various types of epistasis and other causes for modification of gene function (epigenetics, gene-environment interactions) were beyond the scope.

One of the more robust findings in radiogenomics is an association between *ATM* variants and RT-induced morbidity. A meta-analysis across breast and prostate cancer cohorts identified rs1801516 \*A associated with ORs of 1.5 or 1.2 for being in the upper quartile of acute and late toxicity, respectively [9]. We did not identify associations between *ATM* variants and endpoints in this study which may own to the limitation of sample size, not rendering identification of variants in this range of effect size likely.

In conclusion, this is the first study to identify and replicate a locus associated with RT-induced mucositis in HNC patients with a laryngeal or oro/hypopharyngeal tumour site. It serves to extend the field of radiogenomics and adds a piece to the increasing evidence that an individual genetic predisposition is a contributory factor in the development of toxicity after RT. Collaborative efforts of combining cohorts of HNC patients with solid data for RT-induced morbidity would enable identification of genetic variants with lower penetrance, possibilities for validation and in the longer run, to add biology variables to Normal Tissue Complication Models. Such efforts are warranted under the auspices of the Head and Neck Cancer Group in the Radiogenomics Consortium.

### Reporting guidelines

The study was reported complying with the STROGAR guidelines for radiogenomic studies.

### Web resources

dbSNP, <https://www.ncbi.nlm.nih.gov/SNP/>, GAS power calculator, [http://csg.sph.umich.edu/abecasis/cats/gas\\_power\\_calculator/index.html](http://csg.sph.umich.edu/abecasis/cats/gas_power_calculator/index.html), GATK, <https://gatk.broadinstitute.org/hc/en-us>, The Genotype-Tissue Expression (GTEx) Project, [gtexportal.org](http://gtexportal.org), IMPUTE2, [http://mathgen.stats.ox.ac.uk/impute/impute\\_v2.html](http://mathgen.stats.ox.ac.uk/impute/impute_v2.html), LocusZoom, <http://locuszoom.org/>, National Institute of Health LD analysis and plot, <https://analysistools.nci.nih.gov/LDlink/>, PLINK, <https://www.cog-genomics.org/plink2>, R Software, <https://www.r-project.org/>, SHAPEIT, [https://mathgen.stats.ox.ac.uk/genetics\\_software/shapeit/shapeit.html](https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html), SNPedia, <https://www.snpedia.com/index.php/SNPedia>, SNPTEST, [https://mathgen.stats.ox.ac.uk/genetics\\_software/snptest/snptest.html](https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html).

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

Summary data for associations reaching  $p$ -values  $< 10^{-5}$  are available in Supplementary Table 3.

### REFERENCES

- Blanchard P, Baujat B, Holostenco V, Bourredjem A, Baey C, Bourhis J, et al. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): a comprehensive analysis by tumour site. *Radiother Oncol.* 2011;100:33–40.
- Bourhis J, Overgaard J, Audry H, Ang KK, Saunders M, Bernier J, et al. Hyperfractionated or accelerated radiotherapy in head and neck cancer: a meta-analysis. *Lancet.* 2006;368:843–54.
- Overgaard J, Jovanovic A, Godballe C, Grau Eriksen J. The Danish Head and Neck Cancer database. *Clin Epidemiol.* 2016;ume 8:491–6.
- Brodin NP, Kabarriti R, Garg MK, Guha C, Tomé WA. Systematic review of normal tissue complication models relevant to standard fractionation radiation therapy of the Head and Neck Region Published After the QUANTEC Reports. *Int J Radiat Oncol Biol Phys.* 2018;100:391–407.
- Rosenstein BS, West CM, Bentzen SM, Alsner J, Andreassen CN, Azria D, et al. Radiogenomics: radiobiology enters the era of big data and team science. *Int J Radiat Oncol Biol Phys.* 2014;89:709–13.

- West C, Rosenstein BS, Alsner J, Azria D, Barnett G, Begg A, et al. Establishment of a radiogenomics consortium. *Int J Radiat Oncol Biol Phys.* 2010;76:1295–6.
- West C, Rosenstein BS. Establishment of a radiogenomics consortium. *Radiother Oncol.* 2010;94:117–8.
- Herskind C, Talbot CJ, Kerns SL, Veldwijk MR, Rosenstein BS, West CML. Radiogenomics: a systems biology approach to understanding genetic risk factors for radiotherapy toxicity? *Cancer Lett.* 2016;382:95–109.
- Andreassen CN, Rosenstein BS, Kerns SL, Ostrer H, De Ruyscher D, Cesaretti JA, et al. 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. *Radiother Oncol.* 2016;121:431–9.
- Kerns SL, Fachal L, Dorling L, Barnett GC, Baran A, Peterson DR, et al. Radiogenomics consortium genome-wide association study meta-analysis of late toxicity after prostate cancer radiotherapy. *J Natl Cancer Inst.* 2019;112:179–90.
- Lee S, Kerns S, Ostrer H, Rosenstein B, Deasy JO, Oh JH. Machine learning on a genome-wide association study to predict late genitourinary toxicity after prostate radiation therapy. *Int J Radiat Oncol Biol Phys.* 2018;101:128–35.
- Wang T-M, Shen G-P, Chen M-Y, Zhang J-B, Sun Y, He J, et al. Genome-wide association study of susceptibility loci for radiation-induced brain injury. *J Natl Cancer Inst.* 2019;111:620–8.
- Yang D-W, Wang T-M, Zhang J-B, Li X-Z, He Y-Q, Xiao R, et al. Genome-wide association study identifies genetic susceptibility loci and pathways of radiation-induced acute oral mucositis. *J Transl Med.* 2020;18:224.
- Sumner W, Ray X, Sutton L, Rebibo D, Marincola F, Sanghvi P, et al. Gene alterations as predictors of radiation-induced toxicity in head and neck squamous cell carcinoma. *J Transl Med.* 2021;19:212.
- Overgaard J, Hoff CM, Hansen HS, Specht L, Overgaard M, Lassen P, et al. DAHANCA 10—Effect of darbeopetin alfa and radiotherapy in the treatment of squamous cell carcinoma of the head and neck. A multicenter, open-label, randomized, phase 3 trial by the Danish head and neck cancer group. *Radiother Oncol.* 2018;127:12–9.
- Eriksen JG, Maare C, Johansen J, Primdahl H, Evensen J, Kristensen CA, et al. OC-009: Update of the randomised phase III trial DAHANCA 19: -primary C-RT or RT and zalutumumab for squamous cell carcinomas of head and neck. *Radiother Oncol.* 2015;114:10.
- The Danish Head and Neck Cancer Group. 2021. <https://www.dahanca.dk/>.
- LENT SOMA tables. *Radiother Oncol.* 1995;35:17–60.
- Barnett GC, West CML, Coles CE, Pharoah PDP, Talbot CJ, Elliott RM, et al. Standardized Total Average Toxicity score: a scale- and grade-independent measure of late radiotherapy toxicity to facilitate pooling of data from different studies. *Int J Radiat Oncol Biol Phys.* 2012;82:1065–74.
- Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The OncoArray Consortium: A network for understanding the genetic architecture of common cancers. *Cancer Epidemiol Biomark Prev.* 2017;26:126–35.
- Wakefield J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet.* 2007;81:208–27.
- Phelan CM, Kuchenbaecker KB, Tyrer JP, Kar SP, Lawrenson K, Winham SJ, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet.* 2017;49:680–91.
- R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. 2021. <https://www.r-project.org/>.
- Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007;39:906–13.
- Kerns SL, de Ruyscher D, Andreassen CN, Azria D, Barnett GC, Chang-Claude J, et al. STROGAR—STrengthening the Reporting Of Genetic Association studies in Radiogenomics. *Radiother Oncol.* 2014;110:182–8.
- Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet.* 2006;38:209–13.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38:904–9.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics.* 2010;26:2190–1.
- Poplin R, Ruano-Rubio V, DePristo M, Fennell T, Carneiro M, Van der Auwera G, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv.* 2017. <https://doi.org/10.1101/201178>.
- Aguet F, Barbeira AN, Bonazzola R, Brown A, Castel SE, Jo B, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science.* 2020;369:1318–30.
- Ma F, Zhang D, Yang H, Sun H, Wu W, Gan Y, et al. Endothelial cell-specific molecule 2 (ECSM2) modulates actin remodeling and epidermal growth factor receptor signaling. *Genes Cells.* 2009;14:281–93.



32. Verma A, Bhattacharya R, Remadevi I, Li K, Pramanik K, Samant GV, et al. Endothelial cell-specific chemotaxis receptor (ecscr) promotes angioblast migration during vasculogenesis and enhances VEGF receptor sensitivity. *Blood*. 2010;115:4614–22.
33. Shi C, Lu J, Wu W, Ma F, Georges J, Huang H, et al. Endothelial cell-specific molecule 2 (ECM2) localizes to cell-cell junctions and modulates bFGF-directed cell migration via the ERK-FAK pathway. *PLoS ONE*. 2011;6:e21482.
34. Nandakumar R, Windross SJ, Paludan SR. Intercellular communication in the innate immune system through the cGAS-STING pathway. *Methods Enzymol*. 2019;625:1–11.
35. Dunphy G, Flannery SM, Almine JF, Connolly DJ, Paulus C, Jönsson KL, et al. Non-canonical activation of the DNA sensing adaptor STING by ATM and IFI16 mediates NF- $\kappa$ B signaling after nuclear DNA damage. *Mol Cell*. 2018;71:745–760.e5.
36. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Ridwan Amodé M, et al. Ensembl 2021. *Nucleic Acids Res*. 2021;49:D884–91.
37. Qiu XB, Shao YM, Miao S, Wang L. The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cell Mol Life Sci*. 2006;63:2560–70.
38. Karolak JA, Gambin T, Pitarque JA, Molinari A, Jhangiani S, Stankiewicz P, et al. Variants in SKP1, PROB1, and IL17B genes at keratoconus 5q31.1-q35.3 susceptibility locus identified by whole-exome sequencing. *Eur J Hum Genet*. 2017;25:73–8.
39. Tweedie S, Braschi B, Gray K, Jones TEM, Seal RL, Yates B, et al. Genenames.org: The HGNC and VGNc resources in 2021. *Nucleic Acids Res*. 2021;49:D939–46.
40. Tsukaguchi H, Tokui T, Mackenzie B, Berger UV, Chen XZ, Wang Y, et al. A family of mammalian Na<sup>+</sup>-dependent L-ascorbic acid transporters. *Nature*. 1999;399:70–5.
41. Bentzen SM, Overgaard M, Overgaard J. Clinical correlations between late normal tissue endpoints after radiotherapy: Implications for predictive assays of radio-sensitivity. *Eur J Cancer*. 1993;29:1373–6.
42. Andreassen CN, Alsner J, Overgaard J, Herskind C, Haviland J, Owen R, et al. TGFB1 polymorphisms are associated with risk of late normal tissue complications in the breast after radiotherapy for early breast cancer. *Radiother Oncol*. 2005;75:18–21.

## ACKNOWLEDGEMENTS

We are grateful to Don Conroy, Joe Dennis and Michael Lush for excellent technical and bioinformatics assistance and the whole group of people offering support and discussion at The Strangeways Research Laboratory in Cambridge. We also warmly thank Kirsten Hillebrandt and Dorte Grand and for generous technical assistance.

## AUTHOR CONTRIBUTIONS

Study design happened as a full collaboration between all authors. Data collection: DAHANCA, LMHS, MLKC, JAL, JO, JGE, CNA. Data processing and interpretation: LMHS, EN, LF, LD, CL, AMD, EHWO, BZA, JA, CNA. LMHS and EN wrote the material and methods and results sections, LMHS wrote the remaining parts of the manuscript, which was commented on and finally agreed upon by all co-authors.

## FUNDING

This study was made possible by financial support from CIRRO—The Lundbeck Foundation Center for Interventional Research in Radiation Oncology, the Danish Council for Strategic Research, Aarhus University, the Danish Cancer Society (grant no. R17-A776-09-S10, R20-A1040-10-S2 and R90-A6300-14-S2), Radiumstationens

Forskningsfond and Helga and Peter Kornings Fond. Funding partners did not participate in study design, data collection, analysis or preparation of the manuscript.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

The DAHANCA discovery phase and collaboration with international partners was approved by the Scientific Ethics Committee and the Data Protection Agency of Central Denmark Region (j.no. 1–10-72-212-15 and 1-16-02-627-15/2012-58-006). The UMCG prospective data registration program has been reviewed by the local medical ethical committee and is considered as standard care. Written informed consent was obtained from participants for the purpose of genome-wide SNP profiling study as registered in NCT02435576 at clinicaltrials.gov. The Singapore replication study was approved by the local medical ethics committee at National Cancer Centre Singapore and is registered at clinicaltrials.gov (NCT04340024).

## CONSENT FOR PUBLICATION

Not applicable.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41416-021-01670-w>.

**Correspondence** and requests for materials should be addressed to Line M. H. Schack.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022



## THE HEAD AND NECK GROUP OF THE RADIOGENOMICS CONSORTIUM

Behrooz Z. Alizadeh<sup>3,4</sup>, Jan Alsner<sup>1</sup>, Christian Nicolaj Andreassen<sup>1,2</sup>, Gill Barnett<sup>11</sup>, Miguel Elías Aguado Barrera<sup>12</sup>, Neil G. Burnet<sup>13</sup>, Laura M. Calvo<sup>12</sup>, Brenda Diergaard<sup>14</sup>, Tom Dudding<sup>15</sup>, Alison Dunning<sup>16</sup>, Frédéric Duprez<sup>17</sup>, Sarah L. Kerns<sup>18</sup>, Melvin C. L. Kiang<sup>10,19</sup>, Hans Langendijk<sup>3</sup>, Hisham Mehanna<sup>20</sup>, Elnaz Naderi<sup>3,4</sup>, Andy Ness<sup>21</sup>, Jens Overgaard<sup>1</sup>, Line M. H. Schack<sup>1,2</sup>, Adelene Y. L. Sim<sup>22</sup>, An Spiessens<sup>23</sup>, Holly R. Summersgill<sup>24</sup>, Juan F. Tajés<sup>8</sup>, Ana Vega<sup>12</sup>, Ceilidh Welsh<sup>25</sup>, Enya O. H. Wen<sup>23</sup> and Catharine West<sup>26</sup>

<sup>11</sup>Oncology Centre, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. <sup>12</sup>Genetic Unit, Institute of Legal Medicine and Genomic Medicine Group, USC Faculty of Medicine, Madrid, Spain. <sup>13</sup>The Christie, Department of Radiation Oncology, Division of Cancer Sciences, University of Manchester, Manchester, UK. <sup>14</sup>Graduate School of Public Health; UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA. <sup>15</sup>MRC Integrative Epidemiology Unit, Department of Population Health Sciences, Bristol Medical School, University of Bristol, UK AND Bristol Dental School, University of Bristol, Bristol, UK. <sup>16</sup>Centre for Cancer Genetic Epidemiology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK. <sup>17</sup>Department of Radiation Oncology, Ghent University Hospital and Ghent University, Ghent, Belgium. <sup>18</sup>Wilmot Cancer Institute, Department of Radiation Oncology, University of Rochester Medical Center, Rochester, NY, USA. <sup>19</sup>Division of Radiation Oncology, NCCS—National Cancer Centre Singapore, Singapore, Singapore. <sup>20</sup>Institute of Head and Neck Studies and Education, Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK. <sup>21</sup>NIHR Bristol Biomedical Research Centre, University of Bristol NHS Foundation Trust and University of Bristol, Bristol, UK. <sup>22</sup>Division of Medical Sciences, NCCS—National Cancer Centre Singapore, Singapore, Singapore. <sup>23</sup>Department of Radiation Oncology and Experimental Cancer Research, Ghent University Hospital, Ghent, Belgium.

<sup>24</sup>Manchester Cancer Research Centre, Faculty of Life Sciences and The Christie NHS Foundation, Trust, University of Manchester, Manchester, UK. <sup>25</sup>Department of Oncology, University of Cambridge, Cambridge, UK. <sup>26</sup>Institute of Cancer Sciences, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK. The Danish Head and Neck Cancer Group (DAHANCA) is a national multidisciplinary group with contribution from a large number of dedicated health professionals. More information is available at [dahanca.dk](http://dahanca.dk). Members with authorship status in the present manuscript are:

#### **THE DANISH HEAD AND NECK CANCER GROUP (DAHANCA)**

Jan Alsner<sup>1</sup>, Christian Nicolaj Andreassen<sup>1,2</sup>, Jesper Grau Eriksen<sup>1,2</sup>, Jens Overgaard <sup>1</sup> and Line M. H. Schack <sup>1,2</sup> 