# Large-Scale Meta-GWAS Reveals Common Genetic Factors Linked to Radiation-Induced Acute Toxicities across Cancers

**Running title** Shared Genetic Susceptibility to Radiation toxicity

Elnaz Naderi, PhD[1–3], Miguel E. Aguado-Barrera, PhD[4,5], Line M. H. Schack, PhD[6-8], Leila Dorling, PhD[9], Tim Rattay, PhD FRCS[10], Laura Fachal, PhD[11,12], Holly Summersgill, PhD[13], Laura Martínez-Calvo, PhD[4,5], Ceilidh Welsh, MSc[14], Tom Dudding, PhD[15,16], Yasmin Odding, FRCR[17], Ana Varela-Pazos, PhD[18], Rajesh Jena, PhD[19], David J. Thomson, PhD[20,21], Roel J.H.M. Steenbakkers, PhD[2], Joe Dennis, PhD[12], Ramón Lobato-Busto, PhD[22], Jan Alsner, PhD[6], Andy Ness, PhD[15], Chris Nutting, PhD[23], Antonio Gómez-Caamaño, PhD[5,18], Jesper G. Eriksen, PhD[6], Steve J. Thomas, PhD[15], Amy M. Bates, MSc[14], Adam J. Webb, PhD[24], Ananya Choudhury, FRCR[25], Barry S. Rosenstein, PhD[26], Begona Taboada-Valladares, PhD[5,18], Carsten Herskind, PhD[27], David Azria, MD PhD[28], David P. Dearnaley, MD[29], Dirk de Ruysscher, MD[30], Elena Sperk, MD[27], Emma Hall, PhD[31], Hilary Stobart, MSc[32], Jenny Chang-Claude, PhD[33,34], Kim De Ruyck, PhD[35], Liv Veldeman, PhD[36,37], Manuel Altabas, PhD[38], Maria Carmen De Santis, PhD[39], Marie-Pierre Farcy-Jacquet, MD[40], Marlon R. Veldwijk, PhD[27], Matthew R. Sydes, PhD[41], Matthew Parliament, MD[42], Nawaid Usmani, MD[42], Neil G. Burnet, MD[43], Petra Seibold, PhD[33], R Paul Symonds, PhD[44], Rebecca M. Elliott, MSc[25], Renée Bultijnck, PhD[36,37], Sara Gutiérrez-Enríquez, PhD[45], Meritxell Mollà[46], Sarah L. Gulliford, PhD[47], Sheryl Green, MBBCh[48], Tiziana Rancati, PhD[49], Victoria Reyes, PhD[38], Ana Carballo, PhD[5,18], Paula Peleteiro, PhD[5,18], Paloma Sosa-Fajardo, PhD[5,18], Chris Parker, FRCR[47], Valérie Fonteyne, MD PhD[37], Kerstie Johnson, MBBS MSc[50], Maarten Lambrecht, MD PhD[51], Ben Vanneste, MD PhD[36,37,52], Riccardo Valdagni, MD PhD[39,53], Alexandra Giraldo, PhD[38], Mónica Ramos, PhD[38], Brenda Diergaarde, PhD[54], Geoffrey Liu, MD MSc[55], Suzanne M. Leal, PhD[3,56], Melvin L. K. Chua, PhD FRCR[57,58], Miranda Pring, PhD[15], Jens Overgaard, MD FRCR[6], Luis M. Cascallar-Caneda, MD[18], Fréderic Duprez, PhD[36,37], Christopher J. Talbot, PhD[24], Gillian C. Barnett, MRCP PhD[19], Alison M. Dunning, PhD[12], Ana Vega, PhD[4,5,59],

Christian Nicolaj Andreassen, PhD[6,60], Johannes A. Langendijk, PhD[2], Catharine M.L. West, PhD[61], Behrooz Z. Alizadeh, PhD[1*], Sarah L. Kerns, PhD MPH[62*] On the behalf of the Radiogenomics Consortium

[1] Department of Epidemiology, University Medical Center Groningen, Groningen, the Netherlands.

[2] Department of Radiation Oncology, University Medical Center Groningen, Groningen, the Netherlands.

[3] Center for Statistical Genetics, Gertrude H. Sergievsky Center, and the Department of Neurology, Columbia University Medical Center, New York, NY, USA.

[4] Fundación Pública Galega Medicina Xenómica (FPGMX), Santiago de Compostela, Spain.

[5] Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, Spain.

[6] Department of Experimental Clinical Oncology, Aarhus University Hospital, Aarhus, Denmark.

[7] Department of Oncology, Gødstrup Hospital, Denmark.

[8] NIDO | Centre for Research and Education, Gødstrup Hospital, Denmark.

[9] Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.

[10] Leicester Cancer Research Centre, University of Leicester, Leicester, United Kingdom.

[11] Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK.

[12] Centre for Cancer Genetic Epidemiology, Dept of Oncology, University of Cambridge, Cambridge, CB1 8RN, UK.

[13] Manchester Academic Health Science Centre, The Christie NHS Foundation Trust, Manchester, UK.

[14]Department of Oncology, University of Cambridge, Cambridge, United Kingdom.

[15] Bristol Dental School, University of Bristol, Bristol, UK.

[16] MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.

[17] Bristol Cancer Institute, University Hospitals Bristol NHS Foundation Trust, Bristol, UK.

[18] Department of Radiation Oncology, Complexo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain.

[19] Department of Oncology, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK.

[20] Division of Cancer Sciences, University of Manchester, Manchester, United Kingdom.

[21] The Christie NHS Foundation Trust, Manchester, United Kingdom.

[22] Department of Medical Physics, Complexo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain.

[23] Head and Neck Unit, The Royal Marsden Hospital, London, UK.

[24] Department of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom.

[25] Division of Cancer Sciences, the University of Manchester, Manchester Academic Health Science Centre, Christie Hospital, Manchester, UK.

[26] Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[27] Department of Radiation Oncology, Universitätsmedizin Mannheim, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany.

[28] Fédération Universitaire d'Oncologie Radiothérapie d'Occitanie Méditérranée, Département d'Oncologie Radiothérapie, ICM Montpellier, INSERM U1194 IRCM, University of Montpellier, Montpellier, France.

[29]Division of Radiotherapy and Imaging, The Institute of Cancer Research Department and The Royal Marsden NHS Foundation Trust London, United Kingdom.

[30] MAASTRO Clinic, GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, Netherlands.

[31]Clinical Trials and Statistics Unit, The Institute of Cancer Research, London, UK.

[32] Patient Advocate, Independent Cancer Patients' Voice, UK.

[33] Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[34]University Cancer Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany.

[35] Departments of Basic Medical Sciences and Radiotherapy, Ghent University Hospital, Ghent, Belgium.

[36] Department of Human Structure and Repair, Ghent University, Ghent, Belgium.

[37] Department of Radiation Oncology, Ghent University Hospital, Ghent, Belgium.

[38] Radiation Oncology Department, Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain.

[39] Radiation Oncology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy.

[40] Fédération Universitaire d'Oncologie Radiothérapie d'Occitanie Méditérranée, Département d'Oncologie Radiothérapie, CHU Carémeau, Nîmes, France.

[41] MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, University College London, London, UK.

[42] Division of Radiation Oncology, Department of Oncology, Cross Cancer Institute, University of Alberta, Edmonton, Canada.

[43] The Christie NHS Foundation Trust, Manchester, and University of Manchester, M20 4BX, UK.

[44] Cancer Research Centre, University of Leicester, Leicester, United Kingdom.

[45] Hereditary Cancer Genetics Group, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron Hospital Campus, Barcelona, Spain.

[46]Radiation Oncology Department, Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain

[47]Department of Uro-oncology, The Royal Marsden NHS Foundation Trust and Institute of Cancer Research, London, United Kingdom.

[48]Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, USA.

[49] Data Science Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy.

[50] Leicester Cancer Research Centre, University of Leicester, United Kingdom.

[51] Radiation Oncology, KU Leuven, Leuven, Belgium.

[52] Department of Radiation Oncology (Maastro Clinic), GROW School for Oncology and Developmental Biology, Maastricht University Medical Center, Maastricht, the Netherlands.

[53] Prostate Cancer Program, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy.

[54] Department of Human Genetics, School of Public Health, University of Pittsburgh, and UPMC Hillman Cancer Center, Pittsburgh, PA, USA.

[55] Princess Margaret Cancer Centre, Temerty Faculty of Medicine, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada.

[56] Taub Institute for Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, NY, USA.

[57] Division of Radiation Oncology, National Cancer Centre Singapore, Singapore.

[58]Oncology Academic Clinical Programme, Duke-NUS Medical School, Singapore.

[59]Grupo de Medicina Xenómica, Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Universidade de Santiago de Compostela, Santiago de Compostela, Spain.

[60] Department of Oncology, Aarhus University Hospital, Aarhus, Denmark.

[61] Translational Radiobiology Group, Division of Cancer Sciences, University of Manchester, Manchester Academic Health Science Centre, Christie NHS Foundation Trust Hospital, Manchester, UK.

[62] Department of Radiation Oncology, The Medical College of Wisconsin, Milwaukee, WI, USA.

* Senior authors contributed equally

Corresponding Author:

Sarah L. Kerns, PhD MPH

Medical College of Wisconsin

8701 W Watertown Plank Rd

Milwaukee, WI 53226, USA

Phone: 414-955-4680

Email: skerns@mcw.edu

**ABSTRACT**

**Background:** This study was designed to identify common genetic susceptibility and shared genetic variants associated with acute radiation-induced toxicity (RIT) across four cancer types (prostate, head and neck, breast, and lung).

**Methods:** A GWAS meta-analysis was performed using 19 cohorts including 12,042 patients. Acute standardized total average toxicity ($rSTAT_{acute}$) was modelled using a generalized linear regression model for additive effect of genetic variants adjusted for demographic and clinical covariates. LD score regression estimated shared SNP-based heritability of $rSTAT_{acute}$ in all patients and for each cancer type.

**Results:** Shared SNP-based heritability of $STAT_{acute}$ among all cancer types was estimated at 10% (se=0.02), and was higher for prostate (17%, se=0.07), head and neck (27%, se=0.09), and breast (16%, se=0.09) cancers. We identified 130 suggestive associated SNPs with $rSTAT_{acute}$ ($5.0x10^{-8}$<P-value<$1.0x10^{-5}$) across 25 genomic regions. rs142667902 showed the strongest association (effect allele A; effect size -0.17; P-value=$1.7x10^{-7}$), which is located near *DPPA4*, encoding a protein involved in pluripotency in stem cells, which are essential for repair of radiation-induced tissue injury. Gene-set enrichment analysis identified 'RNA splicing via endonucleolytic cleavage and ligation' (P=5.1 $x10^{-6}$, $P_{corrected}$ =0.079) as the top gene set associated with $rSTAT_{acute}$ among all patients. *In-silico* gene expression analysis showed the genes associated with $rSTAT_{acute}$ were statistically significantly up-regulated in skin (not sun exposed $P_{corrected}$=0.004; sun exposed $P_{corrected}$=0.026).

**Conclusions:** There is shared SNP-based heritability for acute RIT across and within individual cancer sites. Future meta-GWAS among large radiotherapy patient cohorts are worthwhile to identify the common causal variants for acute radiotoxicity across cancer types.

**Keywords**

Radiogenomics, GWAS, acute radiation-induced toxicity, SNP-based heritability.

## INTRODUCTION

Approximately 50% of cancer patients receive radiotherapy[1]. Some suffer with radiation-induced toxicities (RITs) in nearby normal tissues that can occur during or shortly after radiotherapy (i.e., acute RITs)[2], or months to years later (i.e., late RITs). While historical understanding of radiobiology separated tissues into classes defined as early and late responding, the contemporary view is that most, if not all, tissues have both an acute and late phase, but the biologic mechanism(s) underlying early and late injury may differ. Next to radiation dose distributions, individual variability in RIT[3] occurs in part due to host factors including comorbidities, body habitus, and underlying genetic susceptibility[4]. Rare pathogenic variants in DNA damage genes, e.g. *ATM*, result in monogenic syndromes with very high risk of RITs, but these do not explain the variation among most patients. Genome-wide association studies (GWAS) have identified common genetic susceptibility variants (e.g., single nucleotide polymorphisms, SNPs) for late RITs, but few studies have focused on acute RITs[5–14].

While some SNPs affect toxicity risk in a tissue-specific manner, others may be common across multiple tissues with relative contributions differing for acute and late RIT[15]. Prior GWAS of late RITs suggest genetic susceptibility depends on molecular mechanisms specific to a given tissue[5], however, given that acute reactions across tissues depend on DNA damage and cell death, we hypothesized that there is shared susceptibility to acute RIT across tissues. Therefore, we aimed to carry out a GWAS meta-analysis of RITs across tissues among 19 cohorts and 12,042 patients representing four cancers (prostate, head and neck, breast, and lung) from the Radiogenomic Consortium (https://epi.grants.cancer.gov/radiogenomics/). The large study population also enabled us to estimate for the first time SNP-based heritability, i.e., the proportion of variation in acute RIT explained by common genetic variants.

## METHODS

### Study Participants

The study design was a retrospective analysis of prospectively collected cohorts including 12,042 patients who presented with cancers of the prostate, head and neck, breast or lung. Patients received radiotherapy (alone or as part of a combination regimen) with curative intent and were followed for the development of acute toxicity (Table 1 & Supplementary Table 1). Each study obtained ethical approval from local review boards, and all participants provided written informed consent. Supplementary Tables 2-5 summarizes the characteristics of the cohorts, and a description is in the Supplementary Methods.

**Assessment of acute RIT**

The primary endpoint was acute RIT, adjusted for demographic and clinical covariates. The grading schema and assessment schedules are given in Supplementary Table 1. To achieve a composite score describing overall acute RIT, we used the standardized total average toxicity (STAT) method described previously[16] and in the Supplementary Methods. STAT is a scale- and grade-independent measure of RIT developed to facilitate meta-analysis and data pooling.

$STAT_{acute}$ was calculated using toxicity assessments collected within 90 days from start of radiotherapy, which is a commonly used definition for acute RIT in cooperative group trials (e.g., RTOG). When more than one assessment was available within this timeframe, the worst score was used. The individual toxicity endpoints used for calculating $STAT_{acute}$ within each cancer type (Supplementary Tables 6-9) reflect the different organs at risk within the treatment field. Patients with high baseline toxicity such that the grade could not worsen were excluded (Supplementary Table 10).

**Genotyping, quality control and imputation**

Whole blood or buffy coat was stored at -80°C, DNA was isolated using standard procedures (i.e., silica membrane spin columns), and genotyped using genome-wide SNP arrays as part of prior GWAS according to Supplementary Table 11. Germline DNA sequences are (near) constant across tissues; thus the SNPs present in blood cells will be identical to those in all cells of the normal tissues susceptible to acute RIT. Standard pre-imputation quality control (QC) filters removed: SNPs

with low call rate, low MAF, and not meeting Hardy-Weinberg; samples with discordant sex, higher than expected pairwise identity by descent, and non-European ancestry determined by principal components analysis (PCA) or multidimensional scaling. SNPs were imputed on the Michigan Imputation Server using the Haplotype Reference Consortium (HRC) r1.1 2016 reference panel or using IMPUTE2 software and the 1000 Genome Phase 3 European reference panel. Post-imputation filters removed SNPs with low MAF (<1%) and/or imputation quality (INFO<0.3). Supplementary Table 11 summarizes the QC steps and number of SNPs imputed for each cohort.

**Covariables**

Demographic and clinical factors are listed in Supplementary Tables 12-15. Backward stepwise selection procedure with a conservative p-value of 0.2 and $STAT_{acute}$ as the dependent variable was used to identify the most influential covariates within the breast, prostate, and lung cancer cohorts. In HNC cohorts, a pre-defined list of covariates (Supplementary Table 13) was used. Residuals from the final multivariable linear model defined $rSTAT_{acute}$, our primary endpoint, in each cohort. $STAT_{acute}$ unadjusted for demographic and clinical covariates was our secondary endpoint.

**GWAS analysis**

SNP associations with $rSTAT_{acute}$ and $STAT_{acute}$ were independently analyzed in each cohort by linear regression assuming an additive model including the first 10 principal components to control for population stratification. SNPs were represented by the number of effect alleles or imputed genotype dosage. Statistical analyses were carried out using PLINK/1.90b3.44[17] and SNPTEST[18], and GWAS results were checked using the GWASinspector[19] package in R.

**Meta-analysis**

GWAS results were meta-analyzed across all cohorts and separately per cancer type by two independent centers using the inverse variance weighted, fixed-effects method as implemented in PLINK/1.90b3.44[17] and METAL[20] (version 2011-03-25). The Cochran's Q test for heterogeneity was performed. SNPs were considered as genome-wide significant if they had a $P<5.0x10^{-8}$ for

association with outcome, heterogeneity P>0.05 and were available in at least 50% of samples. All tests of statistical significance are 2-sided.

**LD score regression (LDSC), SNP-based heritability and genetic correlation**

LDSC[21] used summary statistics (~1.1 million SNPs; $X^2$ statistics from the GWAS meta-analysis) on the LD scores across the genome. An LDSC intercept close to 1 suggests no confounding bias, whereas an inflated intercept (>1) may indicate population stratification, confounding, or model misspecifications. We filtered the included variants to the subset included in HapMap3, and excluded variants with: duplicated rs-numbers, ambiguity, MAF<0.01. We used the default European LD score file based on the European 1000 genome reference panel. Cross-trait LDSC estimated genetic correlation[22] for acute RITs between the four cancer types in one-by-one comparisons. The slope of the regression estimated the genetic covariance between two RIT endpoints.

**Gene set and *in-silico* tissue expression analysis**

MAGMA[23] gene set association analysis was implemented in FUMA[24]. The gene-wide p-value was calculated by combining the p-value of all SNPs inside genes after accounting for LD and outliers. We allowed for a window of 10 kb up and downstream of each gene to capture SNPs in nearby regulatory regions. Next, we calculated competitive gene set p-value on the gene-wide p-value after accounting for gene size, gene set density and LD between genes. We defined a gene set as statistically significant if its joint p-value was below the threshold corresponding to a FDR < 0.05.

*In-silico* tissue expression analysis was based on the MAGMA gene property in FUMA. The normalized gene expressions (reads per kilo base per million, RPKM) of 53 normal tissue types were obtained from Genotype-Tissue Expression (GTEx).V8[25]. To obtain differentially expressed gene sets (DEG) for 53 tissue types, we used the normalized expression (zero mean of log2 (RPKM+1)). Two-sided Student's t-tests were performed per gene per tissue compared to all other tissues. Genes with Bonferroni $P_{adjusted}$<0.05 and absolute log fold change ≥0.58 were defined as a DEG set in a given tissue.

## RESULTS

### Patient characteristics

Table 1 and Supplementary Tables 2-5 summarize the patient and clinical characteristics of the cohorts and the treatments received. Figure 1 shows the combined distribution of $STAT_{acute}$ and $rSTAT_{acute}$ scores for the four cancer types and Supplementary Figure 1 shows the distributions for each participating cohort. Supplementary Tables 12-15 list the covariates used in statistical analyses to derive $rSTAT_{acute}$. Table 1 and supplementary Figure 1 describe the distribution of $STAT_{acute}$ and $rSTAT_{acute}$ per cohort.

### Meta-GWAS of acute RIT

The additive effect of >6 million SNPs on $rSTAT_{acute}$ (n=10,398) and $STAT_{acute}$ (n=11,115) was estimated. The Q-Q plots showed no genomic inflation, suggesting population structure was adequately controlled using 10 PCs and only including individuals of European-ancestry (Figure 2). No SNP reached a genome-wide significant. However 130 SNPs with a $5.0x10^{-8}$<p-value<$1.0x10^{-5}$ spanning 25 genomic regions had a suggestive association with $rSTAT_{acute}$. The strongest association with an effect size of -0.174 ($p=1.7x10^{-7}$) per copy of the A allele was for rs142667902. The nearest gene to this SNP is *DPPA4* (Figure 3), which encodes a protein involved in the maintenance of pluripotency in stem cells. From association analysis with $STAT_{acute}$, the number of suggestive SNPs decreased to 66 across 27 genomic regions with rs113548225 displaying the strongest association with an effect size of 0.157 ($p=2.2x10^{-7}$) per copy of the A allele. Supplementary Tables 16-17 contain the suggestively associated SNPs and Supplementary Figure 2 displays Manhattan and Q-Q plots.

We found no genome-wide significant SNPs associated with $rSTAT_{acute}$ or $STAT_{acute}$ for the individual cancer sites. The suggestive findings are summarized in Supplementary Tables 18-25 and Supplementary Figures 2-3.

### SNP-based heritability and genetic correlation

The LDSC intercept close to one for all regression models (Table 2) confirmed negligible inflation attributable to relatedness, and that observed associations were due to the polygenic architecture of RITs. SNP-based heritability (±se) estimates for $rSTAT_{acute}$ were 12±0.07% (prostate), 16±0.09% (breast) and 15±0.09% (HNC). The joint estimated SNP-based heritability for $rSTAT_{acute}$ was 7±0.09%. SNP-based heritability for $STAT_{acute}$ was estimated as 17±0.07% (prostate), 27±0.09% (HNC) and 16±0.09% (breast). The joint SNP-based heritability for $STAT_{acute}$ was 10±0.02%. SNP-based heritability estimates for $STAT_{acute}$ and $rSTAT_{acute}$ in lung cancer were imprecise due to small sample size (se up to 0.40) precluding statistical inference. A one-to-one cross cancer type joint analysis of both $rSTAT_{acute}$ and $STAT_{acute}$ showed no statistically significant genetic correlations between pairs of cancer types (Supplementary Table 26).

**Gene set analysis**

The gene set p-value was computed using the gene-based p-value for 4728 curated gene sets (including canonical pathways) and 6166 GO terms obtained from MsigDB v5.2. We used Ensembl gene models for 19,079 genes, and used Bonferroni-corrected p-value threshold of $2.6x10^{-6}$. MAGMA identified 'protein glycosylation in Golgi' as statistically significantly associated with $rSTAT_{acute}$ in HNC ($P=2.4\ x10^{-6}$, $P_{corrected}=0.037$). The next top-ranking pathways was 'RNA splicing via endonucleolytic cleavage and ligation' ($P=5.1\ x10^{-6}$, $P_{corrected}=0.079$, overall $rSTAT_{acute}$). Detailed results of the top 10 gene-sets per cancer type are shown in Supplementary table 27-31.

***In-silico* tissue expression analysis**

The genes related to overall $rSTAT_{acute}$ reached statistically significant up-regulated expression in skin not-sun exposed ($P=7.2x10^{-5}$, $P_{corrected}=0.004$) and skin sun exposed ($P=4.8x10^{-4}$, $P_{corrected}=0.026$) tissues (Figure 4). No tissue reached a significant p-value in the individual cancer types, however the genes associated with acute toxicity in breast and lung patients had maximum expression in their corresponding tissues (breast mammary and lung tissues), and for HNC, esophagus mucosa ranked as the second most expressed tissue (Supplementary Figure 4).

## DISCUSSION

We identified 130 suggestive SNPs underlying shared genetic susceptibility to acute RIT and showed acute RIT is likely to have a moderate SNP-based heritability of 10%. Higher heritability estimates within cancer types confirmed that the genetic susceptibility of acute RIT is partially tissue type-specific. Gene set analysis identified pathways not previously associated with acute RITs that should be explored functionally and as potential targets for interventions to reduce radiation injury.

### Suggestive genome-wide associations with acute RIT

The top SNP associated with $rSTAT_{acute}$ was rs142667902, nearest *DPPA4*, which encodes a nuclear factor involved in the maintenance of pluripotency in stem cells[26]. The pathogenesis of acute RIT involves the turnover and transit time for pluripotent stem cells to repopulate damaged tissue, thus a role of DPPA4 in RIT is plausible. Gene set analysis identified "RNA splicing via endonucleolytic cleavage and ligation" associated with acute RIT. Exposure to ionizing radiation can disrupt the coupling of RNA splicing with gene transcription involved in DNA repair, cell-cycle control, and apoptosis. This emerging trend sheds light on the complex cellular response to DNA damage[27]. Interestingly, gene expression analysis estimated statistically significantly up-regulated expression in skin not-sun and sun-exposed tissues for genes related to overall $rSTAT_{acute}$. A simple interpretation is that skin is the shared organ at risk for all cancer types affected acutely by radiotherapy. In line with Fessé et al[28], our results suggest that skin and damage to the skin due to sun exposure (non-ionizing radiation) might be interesting to explore further for the understanding the mechanism involved in the response of tissues to DNA damage.

We found 95 suggestive SNPs located in 21 genomic regions associated with acute toxicity in prostate cancer patients. The top SNP (rs72954279) was near to a pseudogene *OACYLP*. A trans-ancestry meta-GWAS identified rs35283980 mapped to *OACYLP* associated with susceptibility to prostate cancer[29], though no studies have been published investigating a role in normal tissue RIT. The top gene set in prostate cancer was "adrenergic receptor activity". Adrenergic receptors are

found throughout the body in many cell types. The bladder is particularly rich in the receptors, which are functionally important regulators of the activities of muscles. Pharmacomechanical and molecular approaches have revealed roles for the beta(3)-adrenoceptor in the urinary bladder and gastrointestinal tract smooth muscle, both organs susceptible to acute RIT during prostate cancer treatment[30]. Pullikuth et al. showed that adrenergic receptor signaling regulates tumor response to ionizing radiation[31] and our finding suggests it would be worthwhile to investigate a role in normal tissue responses. Given activity of the receptor would impact multiple tissues, it is a good candidate for further study.

There were 83 suggestive SNPs in 21 genomic regions associated with acute RIT in HNC patients. The top SNP, rs137992872, mapped to *TCF20* encoding a widely expressed transcriptional co-regulator. Our analyses suggested that "protein glycosylation in Golgi" is a potential mechanism involved in susceptibility to acute RIT in HNC. Approximately half of all proteins undergo glycosylation, and this modification has a substantial impact on diverse cellular processes in all tissues. Published studies linked up-regulation of glycosylation genes with radioresistance[33,34]. Inhibition of glycosylation has also been shown to enhance sensitivity to cisplatin (a DNA damaging agent) in HNC cells[32]. Toth et al[33] found that plasma protein glycosylation changes in response to partial body irradiation and the effects last during follow-up.

Of 26 SNPs in 14 genomic regions possibly associated with acute RIT among breast cancer cohorts, the top SNP was rs16882722, mapped near the tumor suppressor *UNC5D,* a netrin receptor involved in apoptosis[34]. Moelans et al. observed an association between *DUSP26* and *UNC5D* loss and chemo-radiotherapy resistance which predicted for worse survival in breast cancer patients[35]. The top gene set associated with RIT in breast cancer was "natural killer cell lectin-like receptor binding". Natural Killer cells are innate immune cells that can respond to inflammatory signals such as interferons and interleukins present at the site of normal tissue injury and can potentiate vascular [damage36,37], and our findings suggest it would be worthwhile to investigate their role in the pathogenesis of RIT in breast cancer.

We found 30 SNPs in 10 genomic regions suggestive of an association with acute RIT in lung cancer patients. The nearest gene to the top SNP (rs1471101) was *MLLT3*. Ayako et al. found a joint effect of *MLLT1* and *MLLT3* genes on the ATM-signaling pathway and a role in repressing genotoxic stresses due to DNA double strand breaks (DSBs) and maintaining genome integrity[38]. Furthermore, our analysis showed the highest expression of genes in lung tissue among all 53 tissues tested and that the top gene set in RIT in lung cancer was "Debiasi apoptosis by reovirus infection dn". Our findings suggest comparing the mechanisms of reovirus-induced apoptosis with radiation-induced apoptosis might identify similarities in tissue damage pathogenesis.

Our observations highlight the complexity of RIT and suggest new avenues to increase understanding of the pathogenesis of acute RIT in a tissue-specific manner. The bioinformatic analyses can point to potential mechanisms but should be used for hypothesis generating and must be followed up with subsequent functional validation studies. Validation studies and subsequent functional characterization of RIT-associated SNPs in cell lines and animal models will be important next steps to understand the molecular mechanisms involved and, potentially, identify targetable pathways for intervention.

**Heritability of acute RIT**

Our first estimate of shared SNP-based heritability of acute RIT across four cancer types was 10%. The estimates were higher for prostate (17%), breast (16%) and HN (27%) cancers. These SNP-based heritability estimates are comparable with those for complex traits such as coronary artery disease (5%)[39], autism spectrum disorder (12%)[40] and schizophrenia (26%)[41]. SNP-based heritability estimates depend on study size, thus our estimates will improve with larger studies[42]. Also, narrow-sense heritability used here misses heritability due to rare variants with large effects that are not tagged by common SNPs, and due to non-additive genetic variation and/or epigenetic factors[43]. Therefore, it is likely that the level of heritability of acute RIT will be higher than that reported in our study.

**Limitations**

No SNP achieved the stringent threshold for genome-wide significance , which is a challenge in GWAS[44]. The rigorous correction for many statistical tests reduces false positives, but it may mask real associations. A second limitation is the lack of ancestral diversity in our cohorts due to limited statistical power to perform a multi-ethnic GWAS, and this limits the generalizability of our findings to non-European and admixed populations. Future studies should be conducted on extended sample size, with particular effort devoted to building cohorts in non-European patient populations and more precisely defining phenotypes for RITs. While we examined common SNPs with a MAF greater than 1%, investigating the rare variants would provide significant insights.

**Conclusion**

There are many common variants potentially associated with acute RITs across tumor sites, and it is worthwhile to carry out larger studies that have the statistical power to identify the causal variants. Our large meta-GWAS provides the first evidence for the heritability of common genetic variants associated with acute RIT, which is higher within than across tissues. Further investigation to verify and expand our findings is merited to identify multiple genome-wide significant loci with pooled clinically relevant effect sizes that can be used in clinical practice.

## DATA AVAILABILITY

This study was done using cohorts involved in the Radiogenomics Consortium. Summary statistics for GWAS results will be available to download from the GWAS Catalog.



## FUNDING

EN was supported by a scholarship for a Ph.D. from the University of Groningen, Groningen, the Netherlands. TD is funded as an Academic Clinical Fellow by the National Institute for Health Research, UK. DJT is supported by a grant from The Taylor Family Foundation and Cancer Research UK [C19941/A30286]. MLKC is supported by the National Medical Research Council Singapore Clinician Scientist Award (NMRC/CSA-INV/0027/2018), National Research Foundation Proton Competitive Research Program (NRF-CRP17-2017-05), Ministry of Education Tier 3 Academic Research Fund (MOE2016-T3-1-004), the Duke-NUS Oncology Academic Program Goh Foundation Proton Research Programme, NCCS Cancer Fund, and the Kua Hong Pak Head and Neck Cancer Research Programme. GCB is supported by Cancer research UK RadNet Cambridge [C17918/A28870]. RADIOGEN research was supported by Spanish Instituto de Salud Carlos III (ISCIII) funding, an initiative of the Spanish Ministry of Economy and Innovation partially supported by European Regional Development FEDER Funds (INT20/00071, INT15/00070, INT17/00133, INT16/00154; PI19/01424; PI16/00046; PI13/02030; PI10/00164); by AECC grant PRYES211091VEGA and through the Autonomous Government of Galicia (Consolidation and structuring program: IN607B). CNA and LMHS received funding from the Danish cancer Society (Grant R231-A14074-B2537). TR was funded by a National Institute of Health Research (NIHR) Clinical Lectureship (CL 2017-11-002) and is supported by the NIHR Leicester Biomedical Research Centre. This publication presents independent research funded by the NIHR. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. REQUITE received funding from the European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreement no. 601826. SGE

is supported by the Government of Catalonia 2021SGR01112. LD was supported by the European Union Horizon 2020 research and innovation programs BRIDGES (grant number, 634935).

**CONFLICTS OF INTEREST**

The authors declare no conflicts of interest. No one was paid to write this article by a pharmaceutical company or agency.

**ACKNOWLEDGMENTS**

The study sponsor(s) were not involved in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

We thank all patients who participated in the study and the participating clinic staff for their contribution to data collection.

This publication presents data from the Head and Neck 5000 study. The study was a component of independent research funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0707-10034). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. Core funding was also provided through awards from Above and Beyond, University Hospitals Bristol and Weston Research Capability Funding and the NIHR Senior Investigator award to Professor Andy Ness. Genotyping was funded by World Cancer Research Fund Pilot Grant (grant number: 2018/1792), Above and Beyond, Wellcome Trust Research Training Fellowship (201237/Z/16/Z) and Cancer Research UK Cancer Research UK Programme Grant, the Integrative Cancer Epidemiology Programme (grant number: C18281/A19169). The VHIO authors acknowledge the Cellex Foundation for providing research equipment and facilities and thank CERCA Program/Generalitat de Catalunya for institutional support.

**REFERENCES**


1. Baskar R, Lee KA, Yeo R, Yeoh KW. Cancer and radiation therapy: Current advances and future directions. *Int J Med Sci*. 2012. doi:10.7150/ijms.3635

2. Dragun A.E. *Encyclopedia of Radiation Oncology*.; 2013. doi:10.1007/978-3-540-85516-3

3. Bentzen SM, Overgaard J. Patient-to-patient variability in the expression of radiation-induced normal tissue injury. *Semin Radiat Oncol*. 1994. doi:10.1016/S1053-4296(05)80034-7

4. Andreassen CN, Alsner J, Overgaard J. Does variability in normal tissue reactions after radiotherapy have a genetic basis - Where and how to look for it? *Radiother Oncol*. 2002. doi:10.1016/S0167-8140(02)00154-8

5. Barnett GC, Thompson D, Fachal L, et al. A genome wide association study (GWAS) providing evidence of an association between common genetic variants and late radiotherapy toxicity. *Radiother Oncol*. 2014;111(2):178-185. doi:10.1016/j.radonc.2014.02.012

6. Naderi E, Petra A, Crijns G, et al. A two‐stage genome‐wide association study of radiation‐induced acute toxicity in head and neck cancer. *J Transl Med*. 2021;1:1-10. doi:10.1186/s12967-021-03145-1

7. Fachal L, Gómez-Caamaño A, Barnett GC, et al. A three-stage genome-wide association study identifies a susceptibility locus for late radiotherapy toxicity at 2q24.1. *Nat Genet*. 2014;46(8):891-894. doi:10.1038/ng.3020

8. Kerns SL, Fachal L, Dorling L, et al. Radiogenomics Consortium Genome-Wide Association Study Meta-Analysis of Late Toxicity After Prostate Cancer Radiotherapy. *J Natl Cancer Inst*. 2020;112(2):179-190. doi:10.1093/jnci/djz075

9. Deichaite I, Hopper A, Krockenberger L, et al. Germline genetic biomarkers to stratify patients for personalized radiation treatment. *J Transl Med*. 2022;20(1):360. doi:10.1186/s12967-022-03561-x

10. Kerns SL, Dorling L, Fachal L, et al. Meta-analysis of Genome Wide Association Studies Identifies Genetic Markers of Late Toxicity Following Radiotherapy for Prostate Cancer. *EBioMedicine*. 2016;10:150-163. doi:10.1016/j.ebiom.2016.07.022

11. De Ruyck K, Duprez F, Werbrouck J, et al. A predictive model for dysphagia following IMRT for head and neck cancer: introduction of the EMLasso technique. *Radiother Oncol*. 2013;107(3):295-299. doi:10.1016/j.radonc.2013.03.021

12. Werbrouck J, De Ruyck K, Duprez F, et al. Acute normal tissue reactions in head-and-neck cancer patients treated with IMRT: influence of dose and association with genetic polymorphisms in DNA DSB repair genes. *Int J Radiat Oncol Biol Phys*. 2009;73(4):1187-1195. doi:10.1016/j.ijrobp.2008.08.073

13. Naderi E, Schack LMH, Welsh C, et al. Meta-GWAS identifies the heritability of acute radiation-induced toxicities in head and neck cancer. *Radiother Oncol*. 2022;176:138-148. doi:10.1016/j.radonc.2022.09.016

14. Schack LMH, Naderi E, Fachal L, et al. A genome-wide association study of radiotherapy induced toxicity in head and neck cancer patients identifies a susceptibility locus associated with mucositis. *Br J Cancer*. 2022;126(7):1082-1090. doi:10.1038/s41416-021-01670-w

15. Guo Z, Shu Y, Zhou H, Zhang W, Wang H. Radiogenomics helps to achieve personalized therapy by evaluating patient responses to radiation treatment. *Carcinogenesis*. 2014;36(3):307-317. doi:10.1093/carcin/bgv007

16. Barnett GC, West CML, Coles CE, 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. doi:10.1016/j.ijrobp.2011.03.015

17. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007. doi:10.1086/519795

18. 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(7):906-913. doi:10.1038/ng2088

19. Ani A, van der Most PJ, Snieder H, Vaez A, Nolte IM. GWASinspector: comprehensive quality control of genome-wide association study results. *Bioinformatics*. 2021;37(1):129-130. doi:10.1093/bioinformatics/btaa1084

20. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010. doi:10.1093/bioinformatics/btq340

21. Bulik-Sullivan B, Loh PR, Finucane HK, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015. doi:10.1038/ng.3211

22. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015. doi:10.1038/ng.3406

23. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol*. 2015. doi:10.1371/journal.pcbi.1004219

24. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017. doi:10.1038/s41467-017-01261-5

25. Manuscript A. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585. doi:10.1038/ng.2653.The

26. Maldonado-Saldivia J, van den Bergen J, Krouskos M, et al. Dppa2 and Dppa4 Are Closely Linked SAP Motif Genes Restricted to Pluripotent Cells and the Germ Line . *Stem Cells*. 2007. doi:10.1634/stemcells.2006-0269

27. Shkreta L, Chabot B. The RNA splicing response to DNA damage. *Biomolecules*. 2015;5(4):2935-2977. doi:10.3390/biom5042935

28. Fessé P, Qvarnström F, Nyman J, Hermansson I, Ahlgren J, Turesson I. UV-Radiation Response Proteins Reveal Undifferentiated Cutaneous Interfollicular Melanocytes with Hyperradiosensitivity to Differentiation at 0.05 Gy Radiotherapy Dose Fractions. *Radiat Res*. 2019;191(1):93-106. doi:10.1667/RR15078.1

29. Conti D V., Darst BF, Moss LC, et al. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat Genet*. 2021. doi:10.1038/s41588-020-00748-0

30. Tanaka Y, Horinouchi T, Koike K. New insights into beta-adrenoceptors in smooth muscle:

distribution of receptor subtypes and molecular mechanisms triggering muscle relaxation. *Clin Exp Pharmacol Physiol*. 2005;32(7):503-514. doi:10.1111/j.1440-1681.2005.04222.x

31. Hassan S, Pullikuth A, Nelson KC, et al. β2-adrenoreceptor Signaling Increases Therapy Resistance in Prostate Cancer by Upregulating MCL1. *Mol Cancer Res*. 2020;18(12):1839-1848. doi:10.1158/1541-7786.MCR-19-1037

32. Noda I, Fujieda S, Seki M, et al. Inhibition of N-linked glycosylation by tunicamycin enhances sensitivity to cisplatin in human head-and-neck carcinoma cells. *Int J cancer*. 1999;80(2):279-284. doi:10.1002/(sici)1097-0215(19990118)80:2<279::aid-ijc18>3.0.co;2-n

33. Tóth E, Vékey K, Ozohanics O, et al. Changes of protein glycosylation in the course of radiotherapy. *J Pharm Biomed Anal*. 2016;118:380-386. doi:10.1016/j.jpba.2015.11.010

34. Zhu Y, Yu M, Chen Y, et al. Down-regulation of UNC5D in bladder cancer: UNC5D as a possible mediator of cisplatin induced apoptosis in bladder cancer cells. *J Urol*. 2014. doi:10.1016/j.juro.2014.01.108

35. Moelans CB, van Maldegem CMG, van der Wall E, van Diest PJ. Copy number changes at 8p11-12 predict adverse clinical outcome and chemo- and radiotherapy response in breast cancer. *Oncotarget*. 2018. doi:10.18632/oncotarget.24904

36. Vivier E, Tomasello E, Baratin M, Walzer T, Ugolini S. Functions of natural killer cells. *Nat Immunol*. 2008;9(5):503-510. doi:10.1038/ni1582

37. Münz C. Natural killer cells and autoimmunity. *Nat Kill Cells*. 2010;5:461-467. doi:10.1016/B978-0-12-370454-2.00034-X

38. Ui A, Yasui A. Collaboration of MLLT1/ENL, Polycomb and ATM for transcription and genome integrity. *Nucleus*. 2016. doi:10.1080/19491034.2016.1177681

39. Chen H, Wang T, Yang J, Huang S, Zeng P. Improved Detection of Potentially Pleiotropic Genes in Coronary Artery Disease and Chronic Kidney Disease Using GWAS Summary Statistics. *Front Genet*. 2020. doi:10.3389/fgene.2020.592461

40. Grove J, Ripke S, Als TD, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019. doi:10.1038/s41588-019-0344-8

41. Ripke S, Neale BM, Corvin A, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014. doi:10.1038/nature13595

42. Evans LM, Tahmasbi R, Vrieze SI, et al. Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat Genet*. 2018. doi:10.1038/s41588-018-0108-x

43. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. *Nat Genet*. 2017. doi:10.1038/ng.3941

44. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*. 2017;101(1):5-22. doi:10.1016/j.ajhg.2017.06.005

**Table 1.** Characteristics of study cohorts.

| | Cohort | N | Age, median (range) | Female (%) | $STAT_{acute}$ mean (sd) |
|---|---|---|---|---|---|
| Prostate N= 4,213 | RAPPER-CHHiP | 1,487 | 68 (50-84) | na | 0.000 (0.782) |
| | RAPPER-RT01 | 219 | 66 (50-79) | na | 0.000 (0.779) |
| | RADIOGEN-PrCa | 647 | 72 (47-86) | na | 0.000 (0.448) |
| | GenePARE | 225 | 65 (47-82) | na | -0.007 (0.681) |
| | CCI-EBRT | 151 | 68 (45-82) | na | 0.000 (0.301) |
| | REQUITE-PrCa | 1,348 | 70 (46-88) | na | 0.004 (0.494) |
| | UGhent-PrCa | 136 | 65 (49-79) | na | 0.001 (0.419) |
| HNC N= 4,042 | UMCG-HANS | 1,279 | 65 (19-93) | 32.7 | 0.005 (0.781) |
| | DAHANCA | 1,183 | 60(27-90) | 21 | -0.003 (0.802) |
| | Ghent-HNC | 273 | 60(31-87) | 16.5 | -0.014 (0.821) |
| | RAPPER-HNC | 187 | 61 (39-85) | 7.6 | 0.078 (0.779) |
| | NIMRAD | 270 | 73 (44-87) | 21 | -0.008 (0.719) |
| | Head and Neck 5000 | 672 | 60 (25-94) | 22 | 0.098 (0.881) |
| | RADIOGEN-HNC | 178 | 63 (35-92) | 11.8 | 0.000 (0.588) |
| Breast N= 2,966 | RAPPER-breast | 907 | 59 (26-83) | 100 | 0.006 (0.993) |
| | REQUITE-breast | 2,059 | 58 (23-90) | 100 | 0.007 (0.786) |
| Lung N= 821 | RADIOGEN-Lung | 154 | 65 (41-89) | 13.6 | -0.006 (0.546) |
| | REQUITE-Lung | 467 | 70 (39-91) | 31.0 | 0.002 (0.545) |
| | CONVERT | 200 | 64 (29-82) | 48.0 | -0.015 (0.524) |

Abbreviations: N number of patients, STAT: standardized total average toxicity, sd: standard deviation, HNC: Head and Neck cancer, na: not applicable.

**Table 2.** SNP-based heritability of $STAT_{acute}$ and $rSTAT_{acute}$ overall and per cancer type.

| | | Minimum N | Number of SNPs | Lambda GC | Mean Chi^2 | Intercept (se) | h2 (se) |
|---|---|---|---|---|---|---|---|
| $STAT_{ACUTE}$ | Overall | 7,410 | 1,087,229 | 1.022 | 1.022 | 1.013 (0.006) | 0.102 (0.026) |
| | Prostate cancer | 2,681 | 1,152,958 | 1.007 | 1.013 | 1.020 (0.006) | 0.171 (0.069) |
| | Breast cancer | 1,890 | 1,067,908 | 1.004 | 1.009 | 1.007 (0.006) | 0.161 (0.096) |
| | HNC | 2,292 | 1,148,120 | 1.019 | 1.018 | 1.005 (0.005) | 0.268 (0.088) |
| | Lung cancer | 547 | 709,968 | 1.011 | 1.013 | 1.019 (0.007) | 0.831 (0.399) |
| $rSTAT_{ACUTE}$ | Overall | 6,739 | 1,079,330 | 1.007 | 1.014 | 1.012 (0.006) | 0.070 (0.028) |
| | Prostate cancer | 2,389 | 1,152,836 | 1.005 | 1.009 | 1.016 (0.006) | 0.125 (0.076) |
| | Breast cancer | 1,786 | 1,067,908 | 1.007 | 1.009 | 1.012 (0.006) | 0.158 (0.097) |
| | HNC | 2,256 | 1,140,063 | 1.013 | 1.009 | 1.005 (0.006) | 0.152 (0.091) |
| | Lung cancer | 500 | 709,969 | 1.005 | 1.008 | 1.023 (0.007) | 0.526 (0.424) |

Abbreviations; N: number of analyzed subjects, SNP: Single nucleotide polymorphism, GC: Genomic control, h2: SNP-based heritability, se: standard error, intercept: protects from bias from population stratification and cryptic relatedness.

## Figure legends

**Figure 1.** Histograms of $STAT_{acute}$ and $rSTAT_{acute}$ distribution per cancer type and curve of normal-log distribution.

**Figure 2**. Manhattan (right) and QQ (left) plots of overall meta-analysis for $STAT_{acute}$ and $rSTAT_{acute}$. Mirror

Manhattan Plot: The GWAS for $STAT_{acute}$ and $rSTAT_{acute}$ are displayed in the top and bottom panels respectively. The X-axis represents genomic locations, while the Y-axis indicates -log10 P-values for SNP associations with the outcome. Each SNP is a dot.

QQ Plot: Observed -log10 P-values are on the Y-axis, and expected -log10 P-values are on the X-axis. Every SNP is represented as a dot, with the red line signifying the null hypothesis of no genuine association. Notable deviations from the expected P-value distribution appear primarily at the tail, complemented by the lambda coefficients, indicating effective control of population stratification.

**Figure 3.** Locus Zoom plot for top locus associated with $rSTAT_{acute}$: Purple diamond shows the top SNP and variants in red color are in Linkage Disequilibrium with the top SNP. The *Y* axis shows observed $-\log_{10}$ P-values, and the *X* axis shows the position across the genome with genes mapped there.

**Figure 4.** Tissue expression analysis in 53 tissue types for genes related to overall $rSTAT_{acute}$**.** Tissue expression analysis testing the positive relationship between all annotated genes using the full distribution of SNPs p-values and the average expression of genes per tissue type based on GTEx RNA-seq data.

Figure 1



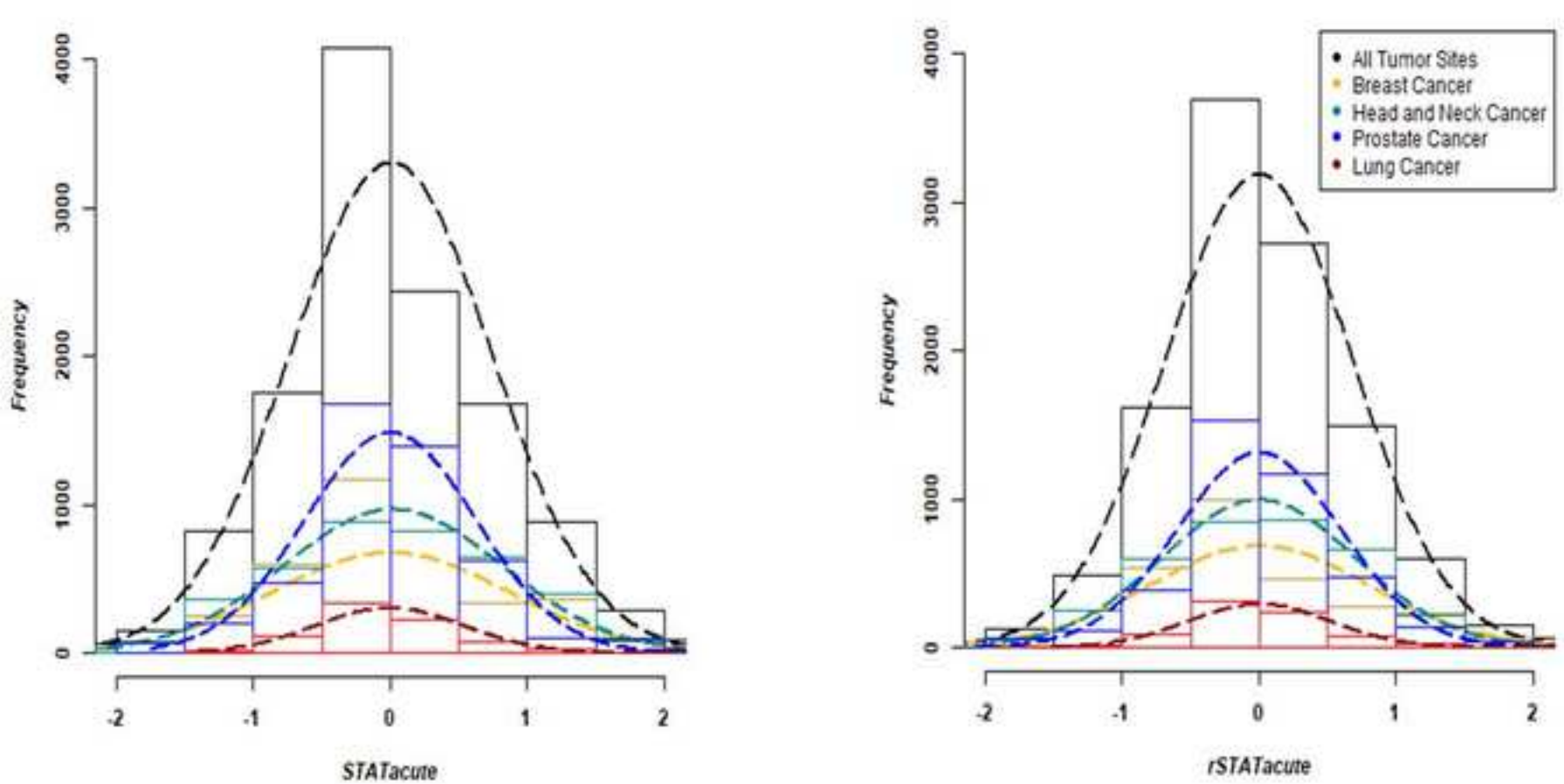
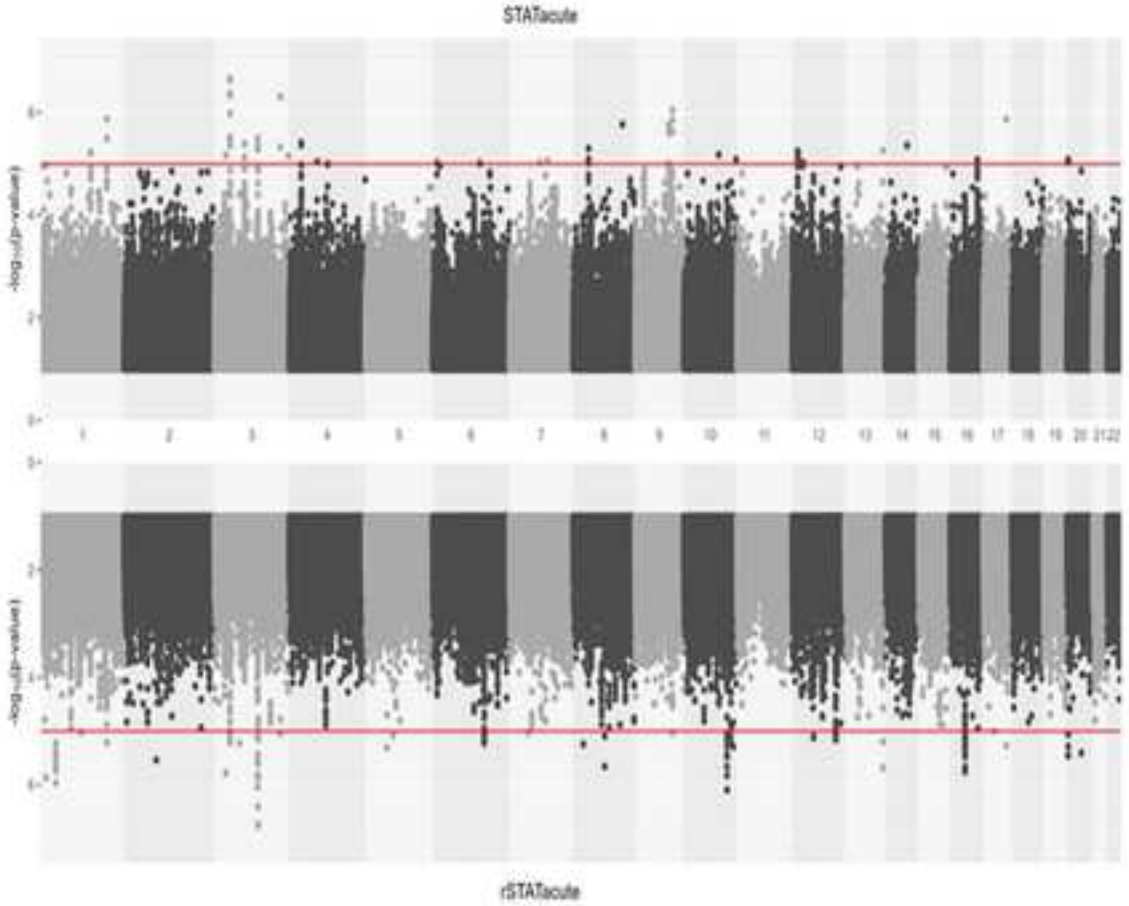


Figure 2





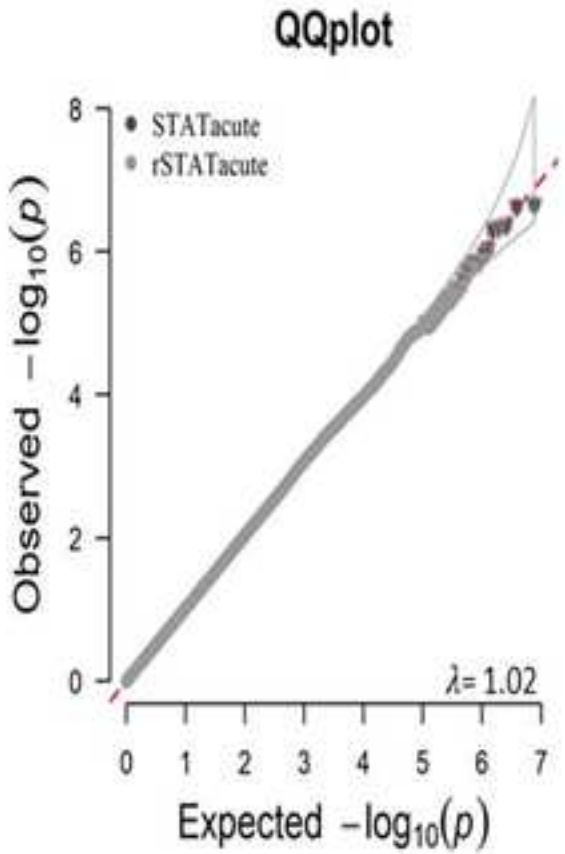



Figure 3



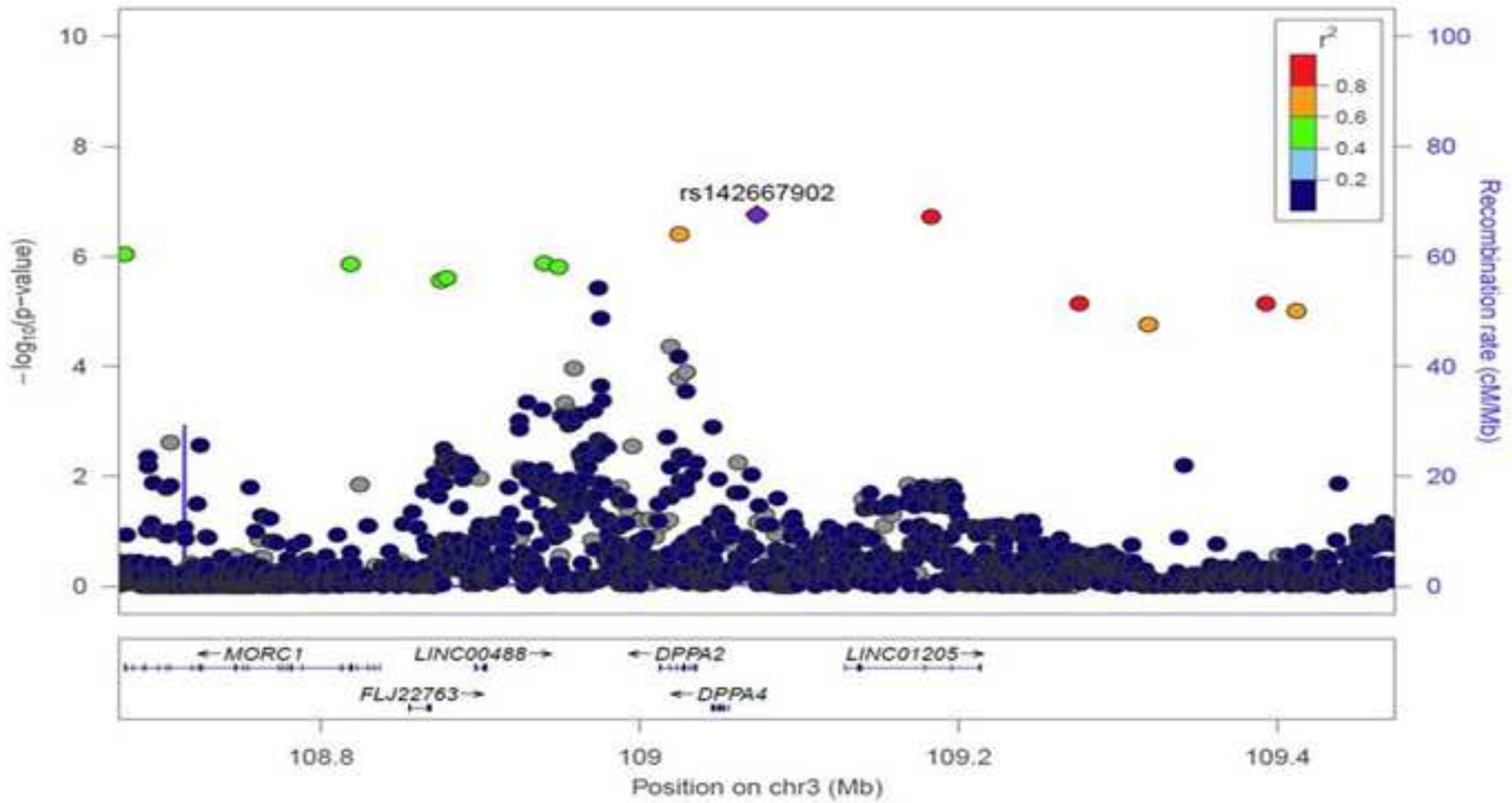


Figure 4