Supplemental data

Supplementary Methods

A) Calculation of PRSs

A PRS is defined as

$$PRS = \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k \ldots + \beta_n x_n$$

Where, for the RT-interaction-PRS, β_k is the logarithm of the per-allele IOR for breast cancer for SNP k among women exposed to chest RT in the case-only analysis, and x_k the number of minor alleles for that SNP (0,1, or 2), and n the total number of SNPs in the PRS. If the IOR was infinite, i.e. no first primary breast cancer case carried a copy of the minor allele, it was recalculated for use in the PRS by changing the genotype of the youngest first primary breast cancer case to heterozygous.

For the BC-PRS, we used the 77-SNP PRS developed by Mavaddat *et al*.[30], which was based on all SNPs previously associated with breast cancer risk at the genome-wide level of statistical significance ($P < 5 \times 10^{-8}$) at the time of the study. Although additional breast cancer risk SNPs have been identified since that study, these have not been carefully assessed for use in a PRS nor has a PRS based on all these SNPs been evaluated in multiple breast cancer populations. Before combining the SNPs in the PRS, Mavaddat *et al*. studied all pair-wise interactions between the 77 SNPs and showed that the assumption of a log-additive model holds. In addition, associations of the PRS with breast cancer risk were validated in a study that did not contribute to the discovery phase of the analysis and studies that oversampled breast cancer cases with a positive family history were excluded. In this PRS, β_k is the per-allele log OR for breast cancer associated with the minor allele for SNP k in the general population described by Mavaddat *et al*. (see Supplementary Table 4 in Mavaddat *et al*. for the list of SNPs and corresponding ORs)[30]. SNP rs78540526 in the BC-PRS was not genotyped with the ICOGs array and therefore excluded from the BC-PRS. SNPs rs11552449 and rs75915166 were genotyped under different names on the iCOGs array (rs12022378 and c11_pos69088342, respectively). For SNP rs2363956, we used the log of the inverse per-allele OR in the PRS as the G allele is the minor allele in our study (MAF 0.499), while it is the major allele in the study by Mavaddat *et al*.

B) Definition of chest RT

For the Dutch HL survivors, chest RT was defined as (in)complete mantle field or mediastinal RT, or RT to the lungs or axilla. Subjects with only infradiaphragmatic RT were excluded. For HL survivors from the USA, chest RT was defined as chest or total nodal RT (subjects with only brain, other head, neck, abdomen, spine, pelvis and/or limb RT were excluded). For HL survivors from the UK, chest RT was defined as mantle field, chest, mediastinal, axillary, mini mantle field or partial chest RT (subjects with

only neck, clavicular and/or head or other supradiaphragmatic RT or infradiaphragmatic RT, RT field unknown or chemotherapy only were excluded). Prescribed dosage data were only available for a part of the study population, but, HL survivors had been administrated radiation dosages of approximately 35-45 Gray[6,10].

C) Quality control of genotype data

Genotype calling for both the HL survivors (with and without subsequent breast cancer) and the first primary breast cancer patients was performed by the Breast Cancer Association Consortium (BCAC) using GenCall and with a cluster file based on a plethora of samples typed with the iCOGs array[36]. Subsequently, we excluded low-quality SNPs and subjects with low-quality genotype data. Initial quality control of the SNPs was based on all samples genotyped in the BCAC consortium and resulted in exclusion of 3,577 SNPs violating Hardy-Weinberg equilibrium at P<0.00001 among all controls in BCAC and 5,420 SNPs called in <95% of all subjects in BCAC. We also excluded 214 SNPs that had a call rate of >95% in BCAC but <90% in the subset of HL cases. In addition, we excluded low-frequency SNPs that had a MAF <1% (n=7,838) in the combined set of HL cases and first primary breast cancer cases.
All first primary breast cancer cases from BCAC had already passed quality control, which included assessment of genotypic sex, heterozygosity, call rate, cryptic relatedness, and genetic ancestry.[35] We applied the same exclusion criteria to the HL subjects, which resulted in the exclusion of seven HL subjects with <95% of the SNPs called (two with and five without breast cancer) and 14 cryptic duplicates (six with and six without breast cancer, and two first primary breast cancer cases who appeared to have had HL in the past and were also included as breast cancer after HL cases). Ancestry of HL subjects was determined by computing the genomic kinship with European (Utah residents with Northern and Western European ancestry (CEU)), Asian (Han Chinese individuals from Beijing, China (CHB) and Japanese individuals from Tokyo, Japan (JPT)) and African (Yoruba individuals from Ibadan, Nigeria (YRI)) subjects from HapMap. The identity by state (IBS) matrix with the genomic kinship between all pairs of subjects was then transformed into a distance matrix and classical multidimensional scaling (MDS) was performed. The first two principal components (PCs) from the MDS were used to determine genetic ethnicity. Ten HL subjects with significant Asian or African ancestry were excluded (four with and six without breast cancer). Finally, 327 cases with breast cancer after HL, 4,671 first primary breast cancer cases and 491 HL controls without breast cancer were available for analyses. Quality control was performed using the package GenABEL within the R statistical environment (http://www.r-project.org).
For all SNPs in the RT-interaction-PRS, we assessed the genotype accuracy by comparing the genotypes of 45 CEU HapMap samples that have previously been genotyped with the iCOGs array and have also been sequenced in the 1000 Genomes Project. For all nine SNPs, the concordance rate was 100%. We further validated the genotype calls of the Bonferroni-significant RT-interaction SNPs by TaqMan

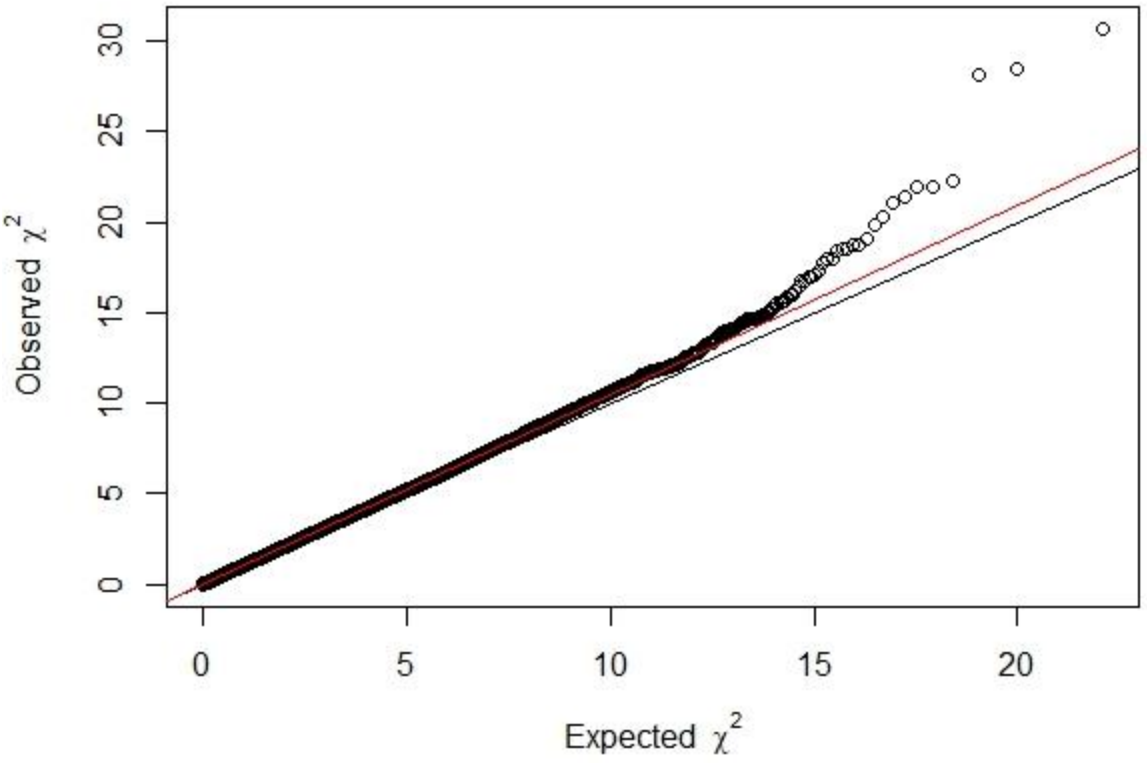
genotyping (Thermo Fisher, USA) in 40 individuals from the Dutch Hodgkin Lymphoma Cohort, including at least 10 carriers and 10 non-carriers per SNP. Primer design failed for rs12086369 and one sample failed in all genotyping assays. The concordance rate for both rs10505506 and rs9461776 was 100% in the remaining 39 samples.

D) Population stratification

When calculating the IOR for every SNP in the case-only analysis, the test statistics may be inflated due to population stratification or cryptic relatedness. We therefore plotted the observed test statistics against the expected null distribution in a Quantile-Quantile (QQ) plot and calculated the genomic inflation factor, lambda. In order to correct for residual population stratification, we estimated the first 10 PCs, describing the 'remaining ethnic genetic differences', among subjects of European origin by computing an IBS matrix and performing MDS for these subjects only. In the case-only analysis, we added PCs as covariates to the logistic regression model until the genomic inflation factor did not further decrease (lambda of 1.05). We only needed to add the first PC, to which we refer as 'ethnicity'. We also added the first PC to the logistic regression models testing the PRSs in the case-control analysis.

E) Asymptotic score test for selection of SNPs

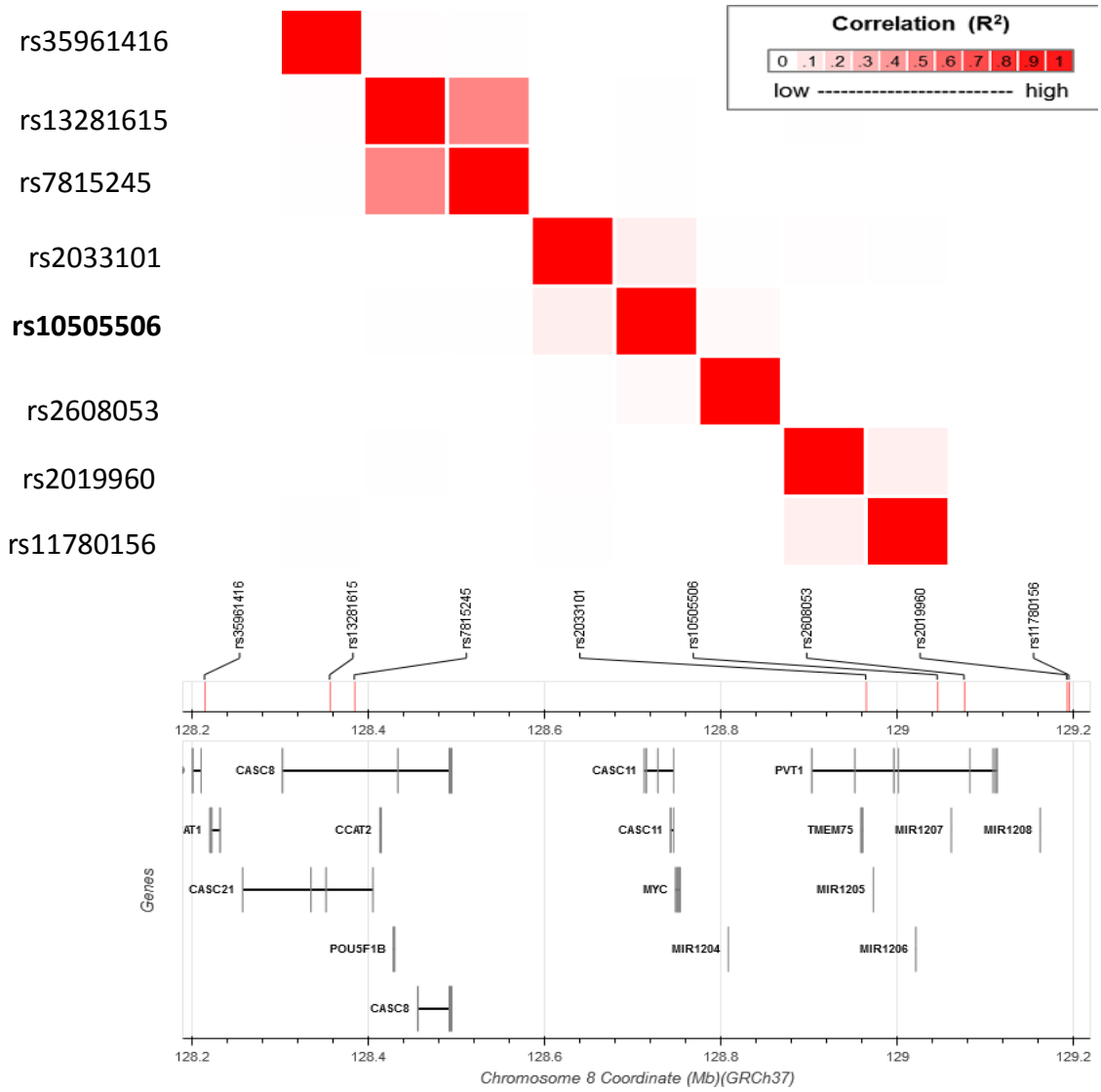
The selection of SNPs for inclusion in the PRS is based on an asymptotic score test for the continuous number of variant alleles in a logistic regression model adjusted for age and year of breast cancer diagnosis (continuous), ethnicity, and country (NL, UK, USA). For SNPs at the lowest frequencies, the number of cases or controls with one or two variant alleles was small. Since calculation of exact P-values in our multivariable model including continuous confounders is not feasible, we performed sensitivity analyses for a set of 50 SNPs (all with P <6.5E-05 (some significant at a 20% FDR, some not), with varying MAF) adjusting only for country. Exact P-values agreed well with asymptotic P-values, although they were sometimes larger and sometimes smaller, mostly by less than a factor of 2. The rank of the SNPs according to their asymptotic and exact P-values was very similar.



Supplementary Figure 1. QQ plot of the observed against the expected distribution of the test statistics in the case-only analysis

Lambda after addition of PC1 for all SNPs with MAF ≥1% amounted to 1.0459.

Supplementary Figure 2. Linkage disequilibrium plot for SNPs at the *PVT1* locus previously associated with breast cancer or HL risk.



We show linkage disequilibrium (r^2 in Europeans from the 1000 Genomes Project[38]) between the Bonferroni-significant *PVT1*-SNP rs10505506 (in bold and underlined) and seven SNPs at 8q24.1, which have previously been associated with breast cancer or Hodgkin lymphoma risk.[53,54] Strongest LD for rs10505506 was observed with rs2033101 (r^2 0.07, D' 0.66). The LD plot was generated using LDlink (Machiela MJ, Chanock SJ. LDlink a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31(21):3555-7).

For reports on LD between the listed breast cancer and HL SNPs and additional SNPs that have been associated with other cancer types within this locus, we refer to Supplementary Table 5 of Enciso-Mora *et al.*[53] and Supplementary Figure 1 of Shi *et al.*[54]. None of the assessed SNPs in these studies are in strong LD with rs10505506 (highest LD was observed for rs10505506-rs2033101: r^2 0.07, D' 0.66).

Supplementary Table 1. Risks of RT-induced breast cancer by RT-interaction-PRS tertiles and stratified analyses

	Breast cancer after Hodgkin lymphoma cases		Hodgkin lymphoma controls without breast cancer		Total	Association with breast cancer after chest RT†		
RT-interaction-PRS‡	N=327		N=491		N=818	OR	95%CI	P
Lowest tertile (≤0.81)	94	28.7%	168	34.2%	262	1.00 (ref)		
Middle tertile (0.81-1.44)	105	32.1%	176	35.8%	281	1.2	0.8 - 1.7	0.348
Highest tertile (≥1.44)	128	39.1%	147	29.9%	275	1.6	1.1 - 2.4	0.007
							P-trend§	0.002
Stratified Analyses								
No gonadotoxic treatment ||	N=158		N=192		N=350			
RT-interaction-PRS								
Lowest tertile (≤0.81)	47	29.7%	69	35.9%	116	1.00 (ref)		
Middle tertile (0.81-1.44)	52	32.9%	65	33.9%	117	1.4	0.8 - 2.4	0.263
Highest tertile (≥1.44)	59	37.3%	58	30.2%	117	1.7	1.0 - 3.0	0.051
							P-trend§	0.023
Gonadotoxic treatment ||	N=152		N=278		N=430			
RT-interaction-PRS								
Lowest tertile (≤0.81)	44	28.9%	90	32.4%	134	1.00 (ref)		
Middle tertile (0.81-1.44)	51	33.6%	103	37.1%	154	1.1	0.7 - 1.9	0.707
Highest tertile (≥1.44)	57	37.5%	85	30.6%	142	1.4	0.8 - 2.4	0.211
							P-trend§	0.204
							P for interaction	0.337
Age at HL treatment ≤20 years	N=175		N=178		N=353			
RT-interaction-PRS								
Lowest tertile (≤0.81)	51	29.1%	63	35.4%	114	1.00 (ref)		
Middle tertile (0.81-1.44)	54	33.9%	57	32.0%	111	1.3	0.7 - 2.2	0.377
Highest tertile (≥1.44)	70	40.0%	58	32.6%	128	1.6	1.0 – 2.8	0.071
							P-trend§	0.027
Age at HL treatment >20 years	N=152		N=313		N=465			
RT-interaction-PRS								
Lowest tertile (≤0.81)	43	28.3%	105	33.5%	148	1.00 (ref)		
Middle tertile (0.81-1.44)	51	33.6%	119	38.0%	170	1.2	0.7 - 2.1	0.413
Highest tertile (≥1.44)	58	38.2%	89	28.4%	147	1.6	1.0 – 2.7	0.070
							P-trend§	0.051
							P for interaction	0.954

CI indicates confidence interval; HL, Hodgkin lymphoma; OR, odds ratio; PRS, polygenic risk score; RT, radiotherapy

† Logistic regression analysis for tertiles of the RT-interaction-PRS, with the lowest tertile as the reference group, to study the association of the RT-interaction-PRS with the risk of RT-induced breast cancer after Hodgkin lymphoma, with adjustment for age at and year of Hodgkin lymphoma diagnosis, country, ethnicity, and the BC-PRS (continuous).

‡ RT-interaction-PRS composed of nine SNPs (MAF≥1%) that reached 20% FDR in the case-only analysis. Tertiles were based on the distribution of the RT-interaction-PRS in cases and controls combined. Seventeen (5.2%) of the breast cancer after Hodgkin lymphoma cases and 34 (6.9%) of the Hodgkin lymphoma controls without breast cancer had a missing genotype for one or more of these nine SNPs. These missing genotypes were imputed with the mode genotype for the specific SNP among Hodgkin lymphoma controls without breast cancer.

§ Computed from a similar model, however where the RT-interaction-PRS was included as a continuous variable instead of a categorical variable.

|| Gonadotoxic treatment was defined as treatment with alkylating chemotherapy and/or pelvic RT. Gonadotoxic treatment was missing for 17 (5.2%) of the breast cancer after Hodgkin lymphoma cases and 21 (4.3%) of the Hodgkin lymphoma controls without breast cancer.

Supplementary Table 2. Association results for the Bonferroni-significant RT-interaction SNPs in the case-control analysis

				Breast cancer after Hodgkin lymphoma cases (n= 327)		Hodgkin lymphoma controls without breast cancer (n=491)		Association with RT-induced breast cancer per SNP†		
SNP	Locus	Chr	Alleles	MAF	N called	MAF	N called	OR‡	95%CI	*P*
rs10505506	*PVT1*	8	G/C	0.407	327	0.339	491	1.3	1.1 - 1.6	0.007
rs12086369	*1p31.1*	1	G/A	0.073	324	0.047	489	1.3	0.9 - 1.9	0.193
rs9461776	*HLA*	6	A/G	0.133	327	0.125	491	1.0	0.8 - 1.4	> 0.5

Chr indicates chromosome; CI, confidence interval; MAF, minor allele frequency; OR, odds ratio; RT, radiotherapy; and SNP, single nucleotide polymorphism

† Logistic regression analysis per SNP (MAF≥1%) to test the log additive effect per allele (per-allele OR) with adjustment for age at and year of Hodgkin lymphoma diagnosis, country, and ethnicity.

Supplementary Table 3. Risks of RT-induced breast cancer by deciles of the BC-PRS

	Breast cancer after Hodgkin lymphoma cases (N=327)		Hodgkin lymphoma controls without breast cancer (N=491)		Total (N=818)	Association with breast cancer after chest-RT†		
BC-PRS, deciles‡	N	%	N	%	N	OR	95% CI	*P*
<10 (≤-0.05)	21	6.4	62	12.6	83	0.6	0.3 - 1.2	0.133
10-20 (>-0.05-0.13)	25	7.6	54	11.0	79	0.8	0.4 - 1.4	0.449
20-30 (>0.13-0.27)	35	10.7	49	10.0	84	1.3	0.8 - 2.3	0.329
30-40 (>0.27-0.39)	31	9.5	52	10.6	83	1.0	0.6 - 1.8	0.917
40-60 (>0.39-0.61)	59	18.0	103	21.0	162	1.0 (ref)		
60-70 (>0.61-0.71)	38	11.6	43	8.8	81	1.5	0.9 - 2.7	0.125
70-80 (>0.71-0.86)	36	11.0	47	9.6	83	1.3	0.8 - 2.3	0.319
80-90 (>0.86-1.04)	32	9.8	47	9.6	79	1.2	0.7 - 2.2	0.463
>90 (>1.04)	50	15.3	34	6.9	84	2.4	1.4 - 4.2	0.002
							P-trend§	9.1E-5

Abbreviations are explained in Supplemental Table 1 and 2; except BC, breast cancer.

† Logistic regression analysis for deciles of the BC-PRS, with the middle quintile (40th to 60th percentile) as the reference group, to study the association of the BC-PRS with the risk of RT-induced breast cancer after Hodgkin lymphoma, with adjustment for age at and year of Hodgkin lymphoma diagnosis, country, ethnicity, and the RT-interaction-PRS in tertiles.

‡ 76 SNPs and corresponding weights for the BC-PRS were extracted from Mavaddat *et al.*, except for one SNP that was excluded as the SNP was not present on the iCOGs array, and deciles of the BC-PRS were based on its distribution in cases and controls combined. Ten (3.1%) of the cases with breast cancer after Hodgkin lymphoma and 9 (1.8%) of the Hodgkin lymphoma controls without breast cancer had a missing genotype for one or more of the SNPs in the BC-PRS. These missing genotypes were imputed with the mode genotype for the specific SNP among Hodgkin lymphoma controls without breast cancer.

§ Computed from a similar model, however where the BC-PRS was included as a continuous variable instead of a categorical variable.