Supplementary Materials for:

Variants at 6q21 implicate *PRDM1* in the etiology of therapy-induced second malignancies after Hodgkin lymphoma

Timothy Best[1], Dalin Li[2], Andrew D Skol[3], Tomas Kirchhoff[4], Sarah A Jackson[3], Yutaka Yasui[5], Smita Bhatia[6], Louise C Strong[7], Susan M Domchek[8], Katherine L Nathanson[8], Olufunmilayo I Olopade[3], R Stephanie Huang[3], Thomas M Mack[2,9], David V Conti[2], Kenneth Offit[4], Wendy Cozen[2,9], Leslie L Robison[10], and Kenan Onel[1,11,12]

[1]Committee on Cancer Biology, [3]Department of Medicine, [11]Department of Pediatrics, University of Chicago, Chicago, IL 60637
[2]Department of Preventive Medicine, [9]Department of Pathology, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90089
[4]Department of Medicine, Clinical Genetics Service, Memorial Sloan-Kettering Cancer Center, New York, NY 10021
[5]Department of Public Health Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G3
[6]Department of Population Sciences, City of Hope, Duarte, CA 91010
[7]Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030
[8]Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA 19104
[10]Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, Tennessee, 38105

Table of Contents:

Complete Methods

Discovery Analysis

Our discovery set was comprised of 100 individuals treated for Hodgkin lymphoma (HL) who subsequently developed SMNs and 89 individuals treated for HL who did not develop SMNs. They were individuals of European descent diagnosed with HL as children (aged 10-20) and treated similarly with radiation therapy (25-44 Gy) +/- alykylating chemotherapy. Among the cases, the radiation exposure was to the site at which the subsequent SMN developed. The distribution and frequency of sites exposed to radiation did not differ between cases and controls. All individuals were participants in the Childhood Cancer Survivor Study (CCSS), a multi-institutional retrospective cohort study of survivors of childhood and adolescent cancers[1]. The CCSS includes patients who were diagnosed < 21 years of age with leukemia, central nervous system (CNS) malignancy, HL, non-Hodgkin lymphoma, kidney cancer, neuroblastoma, soft tissue sarcoma, or a bone malignancy at one of 26 participating institutions between 1970 and 1986, and who survived for $\geq$ 5 years after diagnosis. The CCSS collaborating institutions originally identified 20,691 eligible survivors, 17,633 were successfully contacted, and 14,358 (81.4%) completed the baseline survey by self-reported, mailed questionnaire or telephone with a trained interviewer. Initial surveys were conducted primarily between 1995 and 1996, with periodic follow-ups thereafter until the December 2008 data-freeze date for this study.

Controls for the discovery set were selected from among all HL cases without a reported SMN using a modified frequency matching scheme, to ensure comparability of cases and control with regard to length of observation within the cohort. Additionally, controls were followed for a minimum of 27 years following treatment for HL. We selected 27 years as the minimum follow-up time as this represents one standard deviation from the mean time to development of an SMN (latency to SMN) in the case population. On the basis of prevalence rates at this threshold, we estimate that only 2% of controls would subsequently develop SMNs.

The characteristics of the discovery set are described in **Supplementary Table 1A** and **Supplementary Table 2**.

DNA was isolated from EBV-immortalized LCLs (n=30) established from non-malignant peripheral blood lymphocytes using the PureGene DNA extraction kit (Qiagen), from whole blood (n=69) using the PureGene kit (Qiagen), or from saliva (n=90) with the Oragene kit (DNA Genotek). We randomized cases and controls across 96-well plates and used the Affymetrix Genome-Wide Human SNP Array 6.0 to obtain genotypes for all individuals. We used the Birdseed-v2 algorithm (http://www.broadinstitute.org/mpg/birdsuite/birdseed.html) to determine genotypes for the cases and controls in a plate-dependent manner. We excluded 7 controls and 4 cases that had a contrast QC value below the 0.4 threshold required to resolve heterozygote and homozygote distributions. We re-called the genotypes for the plate containing these individuals following their exclusion. The remaining individuals (96 cases and 82 controls) passed all additional QC metrics: > 95% genotyping call rates, inbreeding coefficient < 0.05, and lack of cryptic relatedness.

After exclusion of 227,243 SNPs with minor allele frequency (in the total sample) <0.05, 17,035 SNPs with genotyping call rate < 0.95, and 31 SNPs that deviated strongly from Hardy-Weinberg equilibrium in controls (HWE; $P < 1.0 \times 10^{-5}$), we investigated 665,313 markers for association.

Replication Analysis

We attempted to replicate all three associations surpassing the threshold for genome-wide significance in the discovery set in an independent collection of patients identically treated for childhood HL between the ages of 10-20: 62 cases who subsequently developed SMNs and 71 controls who did not. These patients came from three sources: 40 cases and 52 controls from the CCSS; 12 cases and 3 controls from Memorial Sloan-Kettering Cancer Center (MSKCC)[2]; and 10 cases and 16 controls from the University of Southern California (USC).

Additionally, we attempted to replicate these associations in a separate cohort of 94 patients treated for Hodgkin's lymphoma with high-dose RT (25-44 Gy) as young adults: 57 cases and 37 controls. 11 cases and 1 control were from the CCSS; 8 cases and 1 control were from MSKCC; 8 cases and 35 controls were from USC; 16 cases were from the University of Pennsylvania; and 14 cases were from the University of Chicago.

For both replication sets, cases and controls were self-identified as white, non-Hispanic. Controls were followed for a minimum of 27 years. All cases and controls were treated for HL with high-dose (36-44 Gy) mediastinal radiation. The characteristics of the HL replication sets are described in **Supplementary Table 1B**.

158 samples were genotyped using the Sequenom iPLEX Gold SNP Genotyping platform, according to manufacturer instructions and 69 samples were genotyped using the Illumina Human 610 Beadchip. Concordance among 14 quality-control samples genotyped on all three platforms was 100%.

All participants in both the discovery and replication sets provided written informed consent approved by their local institutional review boards.

Statistical Methods

For the GWAS analysis, we used the PLINK software package[3] to calculate missingness, allele frequency, and rates of heterozygosity, and to perform conditional haplotype tests (http://pngu.mgh.harvard.edu/~purcell/plink/). Associations of single markers with SMNs were assessed using a Chi-square test of homogeneity, with corresponding estimates of ORs and 95% CIs for markers with significant associations. To define a threshold for genome-wide significance, we permuted the case/control status for all samples 10,000 times, while leaving the genotypes unaltered. We recalculated *P* values using the Chi-square test for each SNP in the permuted datasets and stored the *P* value for the most significant SNP from each permutation. From this analysis, we determined that a genome-wide type I error probability of 0.05 was

equivalent to a nominal *P* value of $1x10^{-7}$. Thus, SNPs with *P* values < $1.0x10^{-7}$ were empirically determined to have achieved genome-wide significance. We performed logistic regression, adjusting for age at HL diagnosis, gender, year of HL diagnosis, gonadal radiation (in females), and alkylating chemotherapy exposure to assess the effect of these risk variables on the most significant associations.

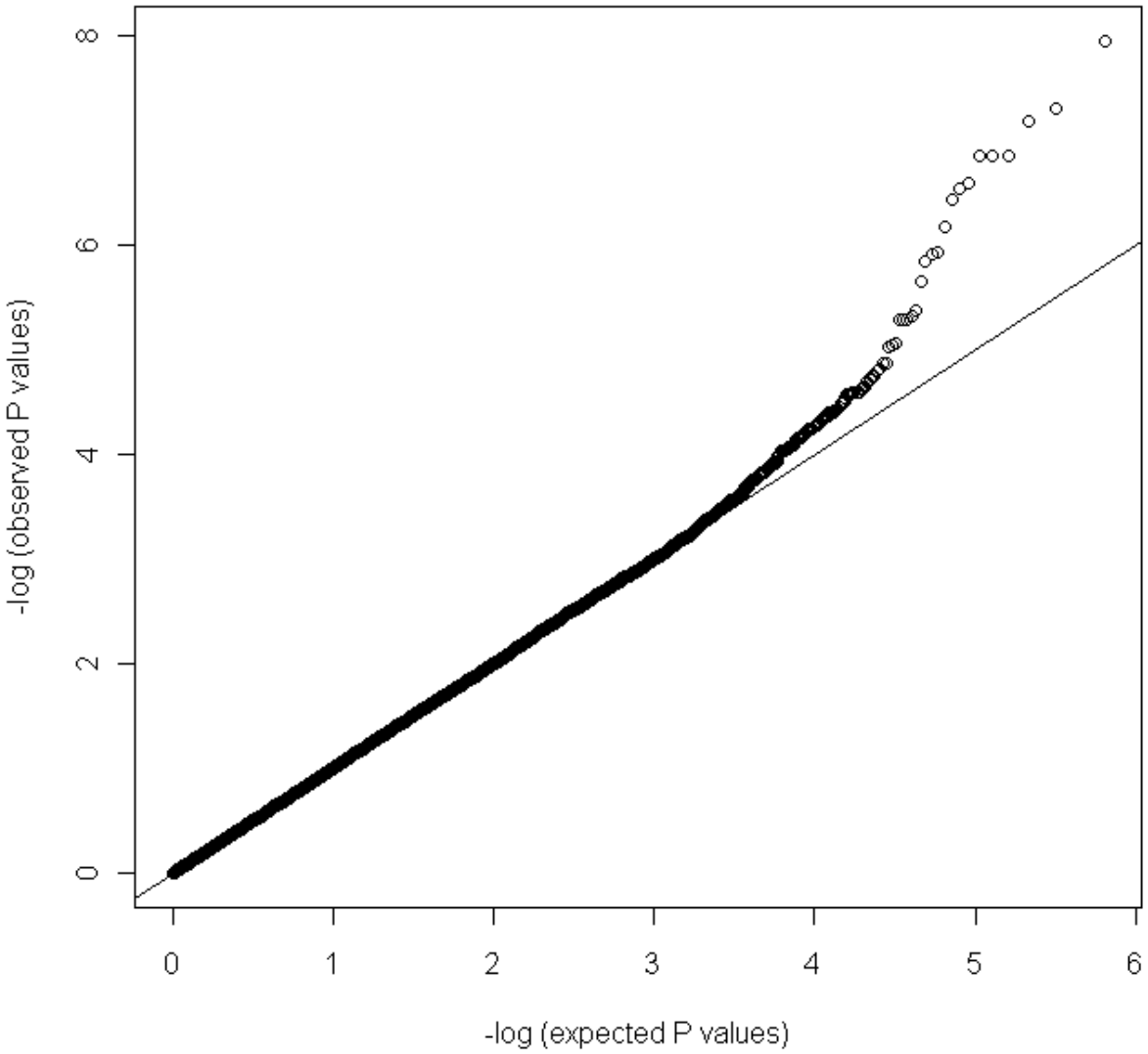
We assessed hidden population structure by constructing a Q-Q plot, using the principal component analysis (PCA)-based approach available in the EIGENSTRAT software package[4], and estimating the genomic control parameter[5]. Genotype imputation was conducted using a hidden Markov model-based algorithm available in the MACH software package and phased CEU HapMap release 22 genotypes as reference for whole genome imputation[6]. Imputation of the 6q21 risk locus was performed using phased 1000 Genomes haplotypes as the reference[7]. We assessed the association of imputed genotypes with SMNs by a Chi-square test of homogeneity; this analysis did not reveal any additional SNPs that surpassed genome-wide significance ($P = 1x10^{-7}$).

Functional Studies of SMN-Associated Genotypes

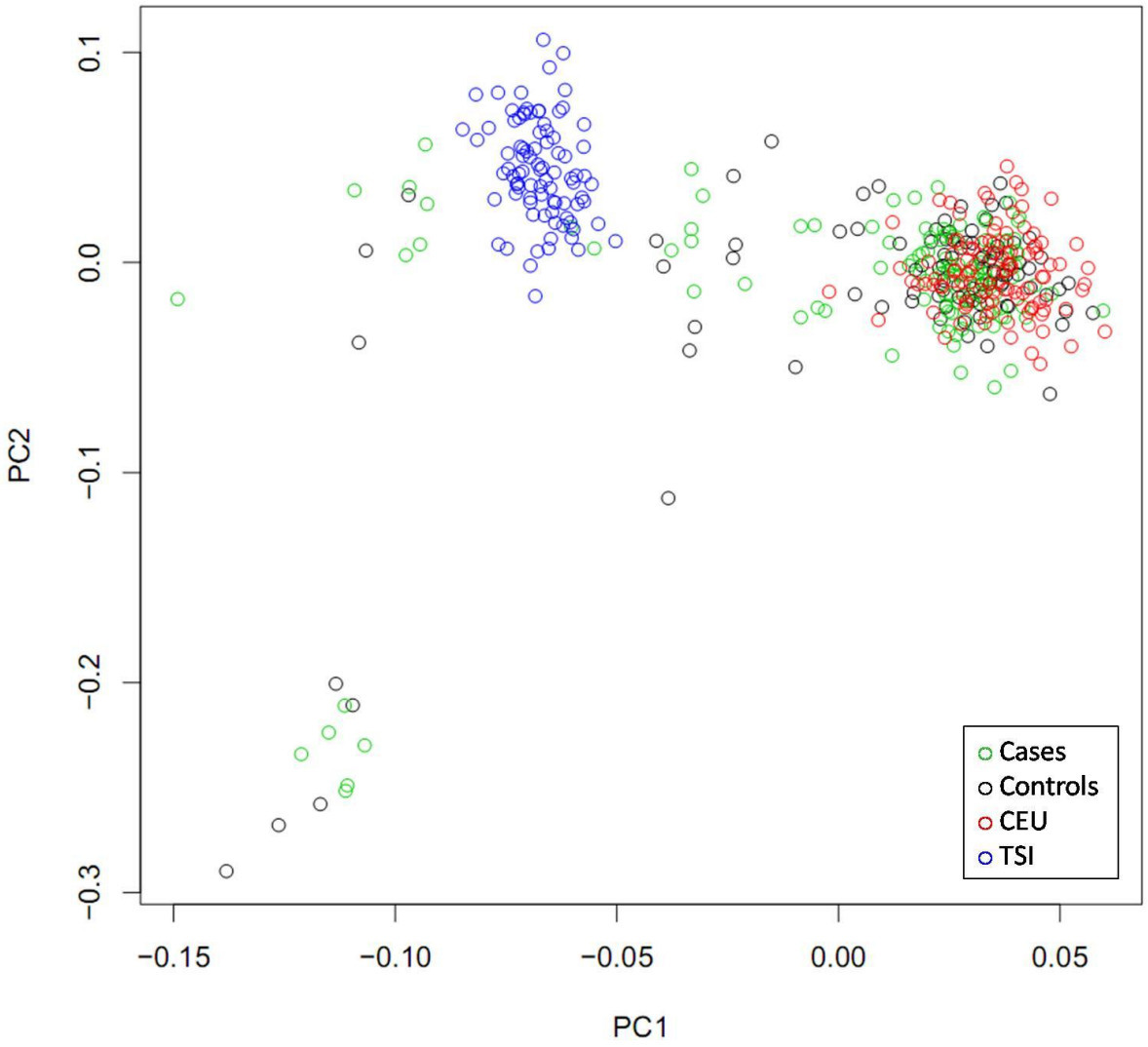
We used lymphoblastoid cell lines (LCLs) from the HapMap CEU population to investigate whether rs4946728 and rs1040411 were expression quantitative trait loci for genes in close physical proximity (cis-eQTLs). Genotypes were downloaded from the International HapMap database[6], and global gene expression was assayed using the Affymetrix GeneChip Human Exon 1.0 ST, as described previously[8]. In Caucasians, rs4946728 and rs1040411 form three common haplotypes of: both major alleles (frequency = 0.49), rs4946278 major allele and rs1040411 minor allele (frequency= 0.21), and both minor alleles (frequency= 0.30). We defined the risk haplotype as containing the major (risk) alleles of both rs4946728 and rs1040411 and the protective haplotype as containing the minor (protective) alleles of both

variants. To determine the association between these haplotypes and gene expression, we used a general linear model.

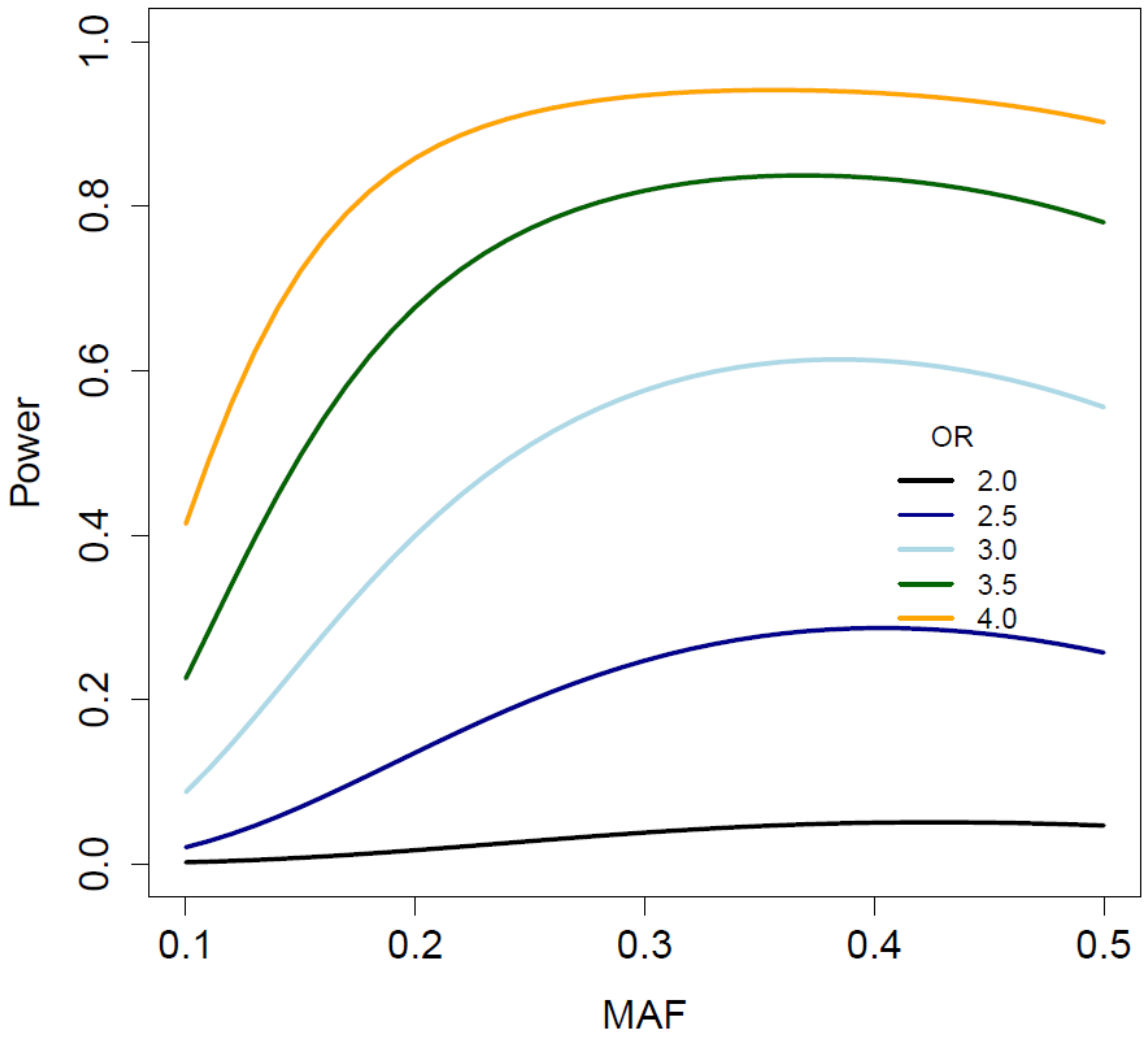
We tested the association between the risk haplotype and protein levels at baseline and following radiation exposure. Four HapMap LCLs homozygous for the risk haplotype and four LCLs homozygous for the protective haplotype were grown in RPMI medium supplemented with 15% FBS and cultured at 37 ^{o}C in a humidified 5% CO2 chamber. When each cell line was > 90% viable, they were adjusted to 3x10^{5} cells/ml and either left untreated or exposed to 10 Gy ionizing radiation. Cell aliquots were removed and protein was isolated by standard methods at 0, 2, and 4 hours following radiation exposure and subjected to immunoblot analysis. ImageJ was used to quantify protein levels using Ran as an internal loading control and t-tests were used to assess significance. Antibodies used include anti-BLIMP1 (PRDM1) (Cell Signaling Technology (CST)), anti-ATG5 (CST), anti-Myc (sc42; Santa Cruz Biotechnology (SCB)), anti-Ran (SCB), anti-rabbit IgG (CST), and anti-mouse IgG (SCB).



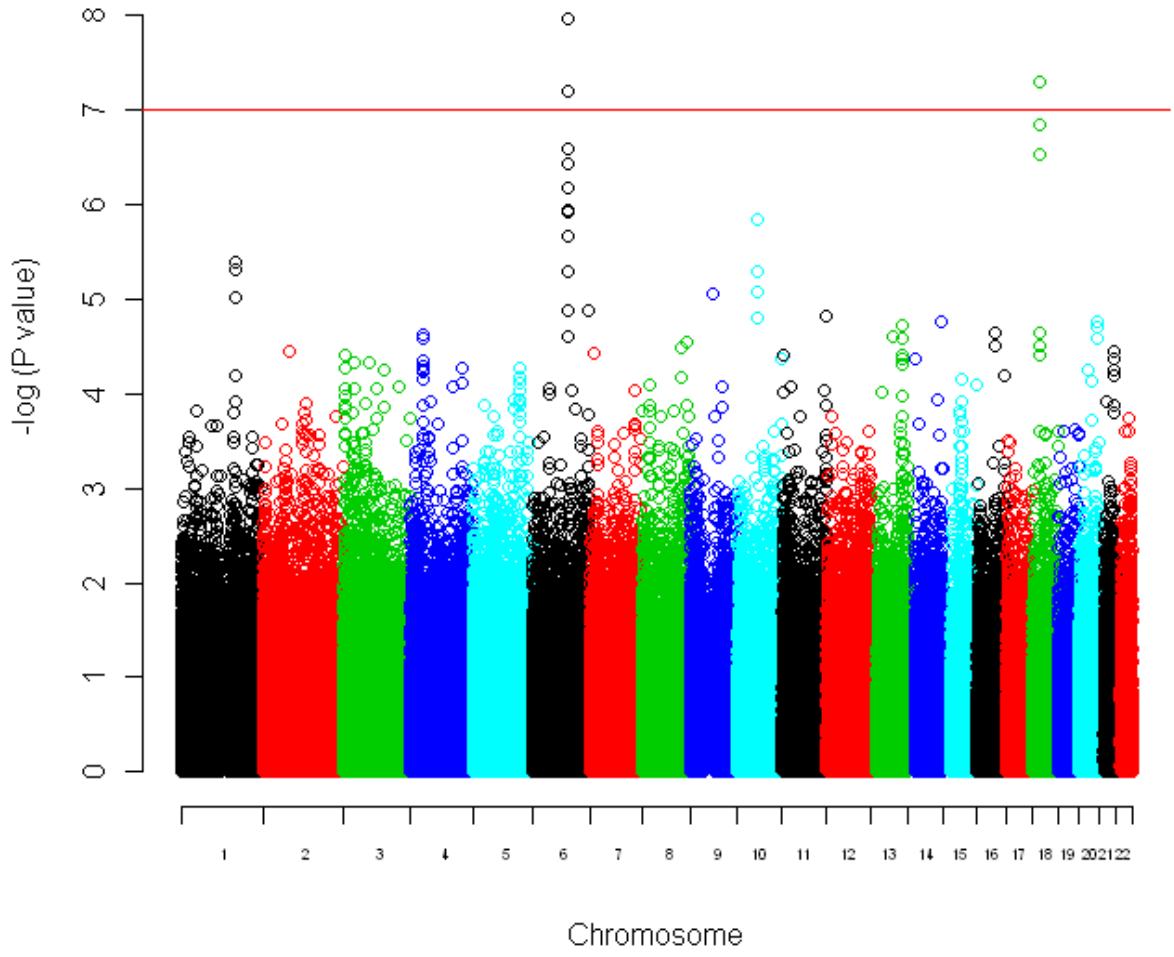
Supplementary Figure 1. Quantile-quantile plot.
Plotted is the observed versus expected distribution of –log (*P* values) for the 665,313 SNPs that passed quality control criteria. The genomic control parameter was 1.007, showing no evidence for inflation.



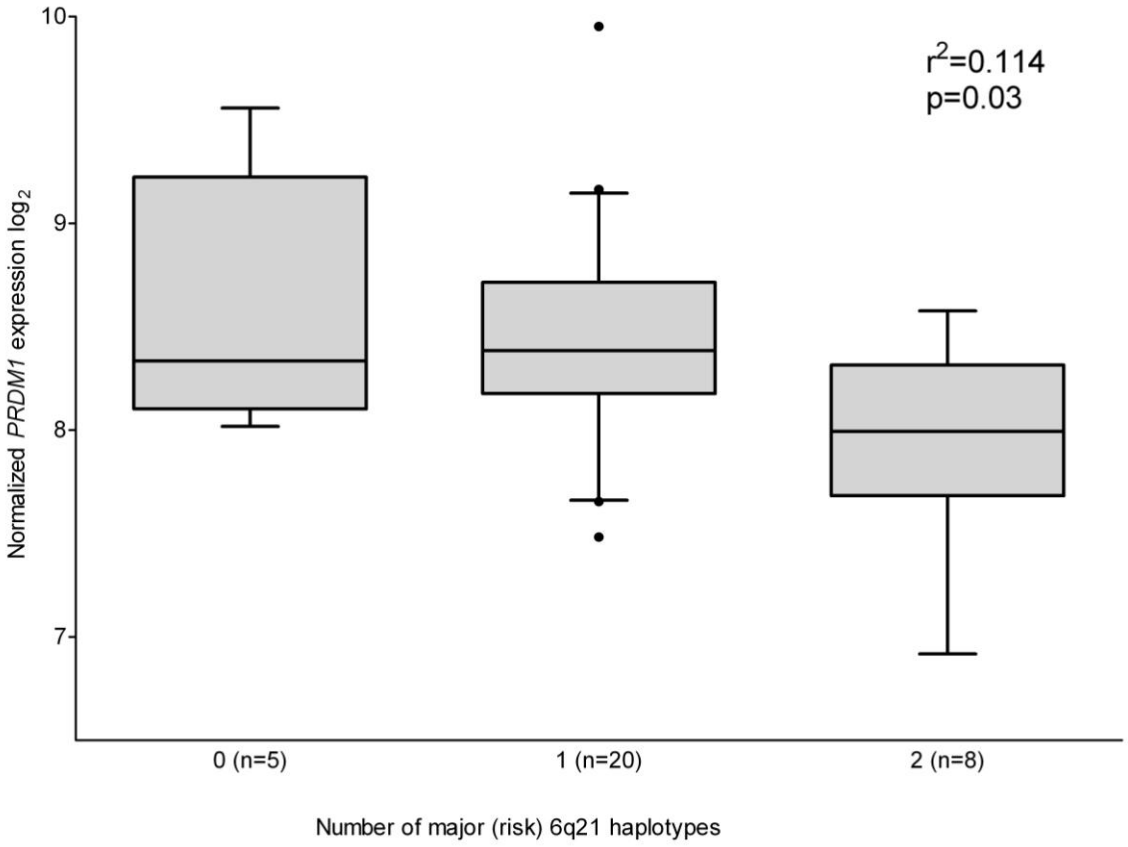
Supplementary Figure 2. Principal component analysis of the genotype data.
Analysis was undertaken using Eigenstrat, and the top two axes of variation are plotted. Cases, controls, northern Europeans from HapMap (CEU) and southern Europeans from HapMap (TSI) are plotted to show intra-European variation. The individuals in the lower left are Ashkenazi. There is no statistically significant difference between case and control populations (P = 0.50).



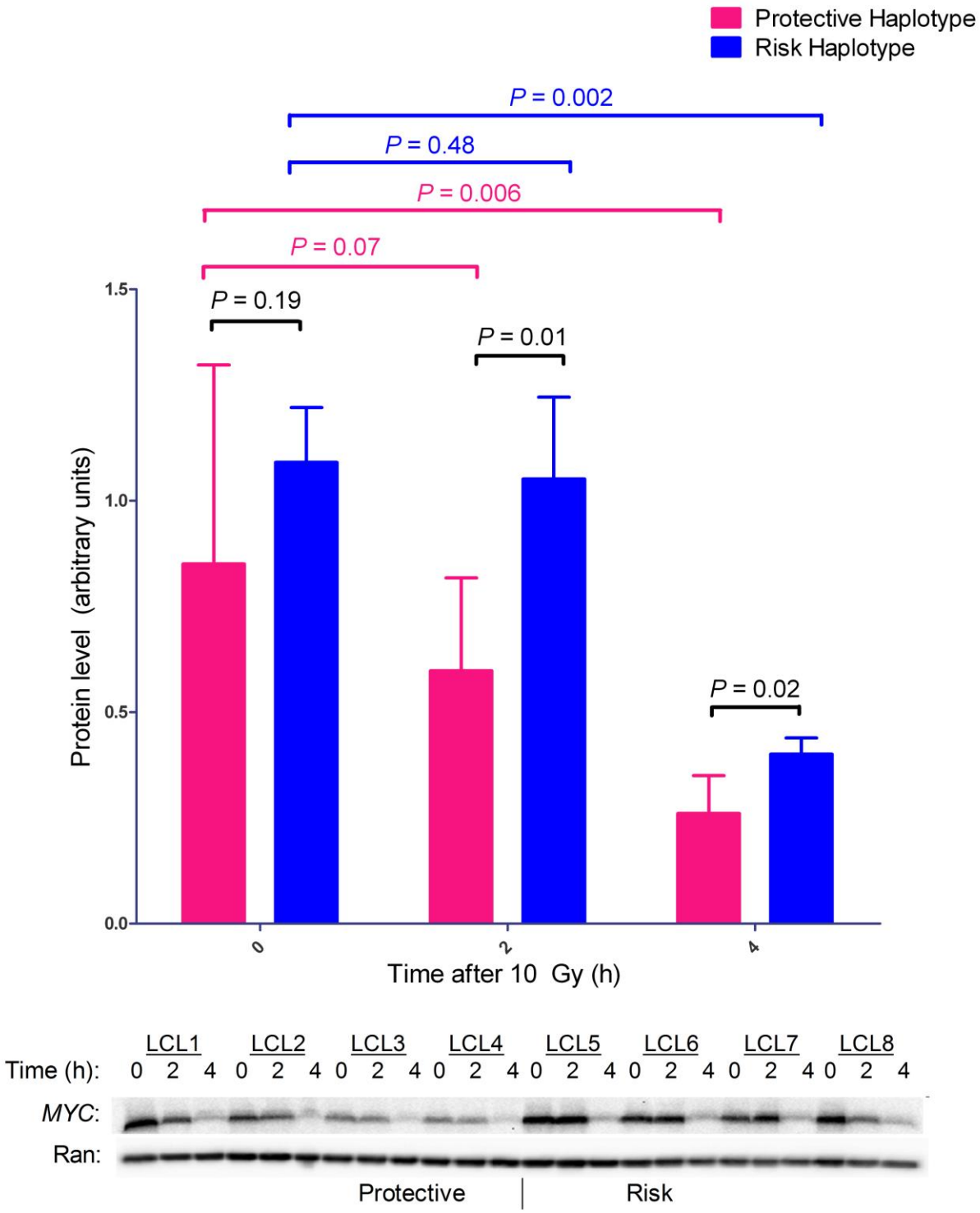
Supplementary Figure 3. Power to detect SMN-associated SNPs.
Shown is the power of the discovery phase to detect associations at a *P* value threshold of $1x10^{-7}$ for different odds ratios using a multiplicative genetic model. At this significance threshold, this study had 80% power to detect common alleles (allele frequency of 35%) with large effect sizes (per-allele odds ratio of 3.5).



<u>Supplementary Figure 4.</u> Variants at 6q21 are associated with SMNs after RT for HL. *P* values were determined for each SNP using a Chi-square test of homogeneity. The threshold for genome-wide significance was determined by permutation of the phenotype data while preserving the genotype data. Three SNPs surpassed the genome-wide significance threshold (*P* value = $1x10^{-7}$): two on chromosome 6q21 (rs4946728 and rs1040411) and one on chromosome 18q11.2 (rs8083533).



<u>Supplementary Figure 5</u>. Relationship between LCL mRNA expression of *PRDM1* carriage of major (risk) 6q21 haplotypes.
A general linear model was used to evaluate the relationship between the expression of *PRDM1* in LCLs and the number of risk haplotypes in each cell line. The risk (major) haplotype was defined as carrying the major alleles for both rs4946728 and rs1040411. The data is shown as mean (inner line), interquartile range (grey box), and overall range in expression.



Supplementary Figure 6. Association of rs4946728/rs1040411 haplotypes with expression of MYC.

Eight lymphoblastoid cell lines, four homozygous for the protective haplotype and four homozygous for the risk haplotype, were treated with 10 Gy gamma irradiation. The risk haplotype was associated with less efficient repression of MYC than was the protective haplotype.

a.

	Discovery	
	Cases ***N*** **= 96**	**Controls** ***N*** **= 82**
Gender		
Male (%)	19 (20)	31 (38)
Female (%)	77 (80)	51 (52)
Age at HL Dx (s.d.)	15.9 (3.0)	15.6 (2.8)
Age at SMN (s.d.)	35.9 (6.6)	-
Latency (s.d.)	20.0 (5.8)	-
Tumor Type (%)		
Breast	59 (61)	-
Thyroid	15 (16)	-
Skin	5 (5)	-
Colorectal	4 (4)	-
Parotid Gland	3 (3)	-
Other[a]	12 (11)	-

b.

	Replication (<20 at treatment)		Replication (>20 at treatment)	
	Cases ***N*** **= 62**	**Controls** ***N*** **= 71**	**Cases** ***N*** **= 57**	**Controls** ***N*** **= 37**
Gender				
Male (%)	6 (10)	14 (20)	(0)	23 (64)
Female (%)	56 (90)	57 (80)	57 (100)	14 (36)
Age at HL Dx (s.d.)	16.4 (2.8)	15.8 (2.8)	25.3 (3.6)	28.5 (6.4)
Age at SMN (s.d.)	35.7 (3.2)	-	44.0 (8.5)	-
Latency (s.d.)	18.8 (3.4)	-	15.9 (6.8)	-
Tumor Type (%)				
Breast	51 (82)	-	55 (96)	-
Thyroid	7 (11)	-	1 (2)	-
Skin	1 (2)	-	0 (0)	-
Parotid Gland	0 (0)	-	0 (0)	-
Other	3 (5)	-	1 (2)	-

<u>Supplementary Table 1</u>. Clinical characteristics of the discovery (a) and replication sets (b) who were treated for HL with radiation +/- chemotherapy.
All cases and controls in both patient sets were treated with high-dose radiation therapy. In sum, 98 SMNs were observed in 96 individuals treated for HL with RT.
[a] Other diagnoses included: osteosarcoma, haemangiosarcoma, large cell diffuse lymphoma, fibromyxoscarcoma, mesothelioma, histiocytoma, ependymoma, and renal cell carcinoma
Note: this table details only individuals whose genotypes passed quality control.

	Discovery	
	Cases ***N* = 96**	**Controls** ***N* = 82**
Treatment Era[a]		
1970-1979 (%)	74 (77)	70 (85)
1980-1986 (%)	24 (25)	12 (15)
Treatment Group		
Radiation only (%)	39 (41)	29 (35)
Chemotherapy + radiation (%)	53 (55)	50 (61)
Missing (%)	4 (4)	3 (4)
RT field by dose[b]		
Supradiaphragmatic <30 Gy (%)	0 (0)	3 (4)
Supradiaphragmatic >=30 Gy (%)	27 (28)	30 (37)
Infradiaphragmatic + supradiaphragmatic, <30 Gy (%)	7 (7)	5 (6)
Infradiaphragmatic + supradiaphragmatic, >=30 Gy (%)	53 (55)	40 (49)
Missing (%)	9 (9)	4 (5)
Alkylating Agent Score[c]		
0 (%)	38 (40)	29 (35)
1 (%)	6 (6)	9 (11)
2 (%)	9 (9)	4 (5)
3 (%)	27 (28)	26 (32)
Missing (%)	18 (19)	14 (17)
SMN and RT field[d]		
In or near RT field (%)	88 (92)	-
Distant to RT field (%)	0 (0)	-
Unable to determine (%)	4 (4)	-

Supplementary Table 2. Treatment characteristics of the discovery set. Cases and controls were treated between 1970 and 1986 with high dose radiation therapy with or without alkylating chemotherapy.

[a] There are two broad eras for Hodgkin treatment in this dataset. Patients in the earlier era (1970-1979) were treated with radiation alone moreso than the later era (1980-1986).

[b] The distribution of sites exposed to radiation and the proportion of patients irradiated at each site did not differ.

[c] The alkylating agent score is a proxy for the dose of alkylating agent the patient received relative to the distribution of doses administered. 0 = no alkylating chemotherapy, 1 = lowest third of the dose distribution, 2 = middle third of the dose distribution, and 3 = highest third of the dose distribution. A high alkylating agent score has been shown to be protective against SMNs in females.

[d] SMNs were determined to appear only within the RT field, although sufficient information regarding the precise location of the SMN relative to the radiation field was not available for four cases.

Note: this table details only individuals whose genotypes passed quality control in the discovery set.

Chr	Position (bp)[a]	SNP	Risk Allele[b]	RAF Cases[c]	RAF Controls	*P* Value[d]	OR [95% CI][e]
6	106697056	rs4946728	C	0.86	0.59	7.51×10^{-7}	4.45 [2.46-8.04]
6	106704716	rs1040411	T	0.68	0.40	2.98×10^{-6}	3.73 [2.15-6.48]
18	22184722	rs8083533	T	0.44	0.18	6.97×10^{-6}	3.23 [1.94-5.39]

Supplementary Table 3. Age at treatment, year of treatment, alkylating chemotherapy exposure, gonadal radiation (in females), and gender do not influence the effect sizes of the associations between rs4946728, rs1040411, or rs8083533 with SMNs following HL.

Logistic regression analysis was undertaken including as covariates: age at HL diagnosis and treatment, year of treatment, alkylating chemotherapy exposure, gonadal radiation (in females), and gender.

[a] Genomic position based on NCBI build-36 coordinates

[b] Risk allele defined as the allele more frequent in the combined cases data set; odds ratios defined with respect to the risk allele

[c] Risk allele frequency

[d] *P* value calculated from logistic regression

[e] Odds ratio [95% confidence interval]

a.

				Breast				Other				
Chr	**Position (bp)[a]**	**SNP**	**Risk Allele[b]**	**RAF[c] Cases**	**RAF Controls**	***P Value*[d]**	**OR [95% CI][e]**	**RAF Cases**	**RAF Controls**	***P* Value**	**OR [95% CI]**	P_{het} [f]
6	106697056	rs4946728	C	0.84	0.64	3.37×10^{-7}	3.01 [1.95-4.65]	0.88	0.64	4.52×10^{-6}	4.17 [2.19-7.97]	0.41
6	106704716	rs1040411	T	0.67	0.44	3.22×10^{-7}	2.55 [1.78-3.67]	0.62	0.44	0.0019	2.07 [1.30-3.29]	0.58
18	22184722	rs8083533	T	0.38	0.26	0.0038	1.75 [1.20-2.55]	0.42	0.26	0.0019	2.10 [1.31-3.37]	0.55

b.

				Females				Males				
Chr	**Position (bp)**	**SNP**	**Risk Allele**	**RAF Cases**	**RAF Controls**	***P Value***	**OR [95% CI]**	**RAF Cases**	**RAF Controls**	***P* Value**	**OR [95% CI]**	P_{het}
6	106697056	rs4946728	C	0.86	0.65	1.77×10^{-7}	3.16 [2.03-4.92]	0.84	0.60	0.0034	3.50 [1.47-8.32]	0.83
6	106704716	rs1040411	T	0.68	0.46	1.25×10^{-6}	2.49 [1.72-3.61]	0.52	0.40	0.18	1.62 [0.81-3.26]	0.29
18	22184722	rs8083533	T	0.38	0.29	0.028	1.54 [1.05-2.27]	0.44	0.19	0.0015	3.37 [1.57-3.37]	0.07

Supplementary Table 4. Subset analysis of SNPs with SMN subtype (a) and gender (*b*).

a. Association of SNPs with breast cancer (110 cases) or other SMN subtypes (48 cases), as compared to 153 controls. The effect size of the association was not significantly different for any SNP tested between breast cancer and other cancer subtypes.

b. Association of SNPs with SMNs in males (25 cases and 45 controls) or females (133 cases and 108 controls). The effect size of the association was not significantly different for any SNP tested between genders.

[a] Genomic position based on NCBI build-36 coordinates

[b] Risk allele defined as the allele more frequent in the combined cases data set; odds ratios defined with respect to the risk allele

[c] Risk allele frequency

[d] Chi-square *P* value

[e] Odds ratio [95% confidence interval]

[f] Breslow-Day *P* value for heterogeneity calculated for the difference between odds ratios

Chr	Position (bp)[a]	SNP	Risk Allele[b]	RAF[c] Cases	RAF Controls	*P* Value[d]	OR [95% CI] [e]
6	106697056	rs4946728	C	0.67	0.75	0.87	0.68 [0.35-1.33]
6	106704716	rs1040411	T	0.53	0.56	0.65	0.89 [0.49-1.61]
18	22184722	rs8083533	T	0.28	0.25	0.69	1.15 [0.58-2.23]

Supplemental Table 5. Association of SNPs with SMNs in patients treated for HL as adults.

[a] Genomic position based on NCBI build-36 coordinates

[b] Risk allele defined as the allele more frequent in the combined cases data set; odds ratios defined with respect to the risk allele

[c] Risk allele frequency

[d] One-sided Chi-squared *P* value

[e] Odds ratio [95% confidence interval]

Position (bp)[a]	SNP	Genotyped/ Imputed[b]	Risk Allele[c]	RAF[d] Cases	RAF Controls	*P* Value[e]	OR [95% CI][f]
106697056	rs4946728	Genotyped	C	0.86	0.59	$1.09x10^{-8}$	4.22 [2.53-7.05]
106704716	rs1040411	Genotyped	A	0.68	0.40	$6.43 x10^{-8}$	3.27 [2.11-5.06]
106695463	rs7759385	Imputed	A	0.87	0.64	$3.69X10^{-7}$	3.75 [2.21-6.36]
106699361	rs6568432	Genotyped	T	0.87	0.64	$3.69X10^{-7}$	3.75 [2.21-6.36]
106703146	rs2065082	Imputed	G	0.87	0.64	$3.69X10^{-7}$	3.75 [2.21-6.36]
106703378	rs9399974	Imputed	A	0.87	0.64	$3.69X10^{-7}$	3.75 [2.21-6.36]
106703606	rs9399975	Genotyped	A	0.87	0.64	$3.69X10^{-7}$	3.75 [2.21-6.36]
106705006	rs1018552	Imputed	C	0.87	0.64	$3.69X10^{-7}$	3.75 [2.21-6.36]
106676749	rs526531	Imputed	G	0.82	0.58	$4.34X10^{-7}$	3.37 [2.08-5.47]
106676768	rs498679	Imputed	C	0.82	0.58	$4.34X10^{-7}$	3.37 [2.08-5.47]
106701412	rs2179175	Genotyped	C	0.87	0.65	$6.70X10^{-7}$	3.65 [2.16-6.2]
106670929	rs533733	Imputed	G	0.81	0.57	$8.55X10^{-7}$	3.23 [2.00-5.20]
106696578	rs11152966	Imputed	C	0.74	0.49	$1.02X10^{-6}$	2.98 [1.91-4.65]
106697111	rs4945743	Imputed	G	0.56	0.30	$1.08X10^{-6}$	2.93 [1.89-4.54]
106705060	rs1885449	Genotyped	A	0.57	0.31	$1.21X10^{-6}$	2.91 [1.88-4.5]
106691788	rs9373835	Imputed	C	0.58	0.32	$1.51X10^{-6}$	2.87 [1.86-4.43]
106692295	rs1008944	Genotyped	T	0.58	0.32	$1.51X10^{-6}$	2.87 [1.86-4.43]
106696189	rs6901662	Genotyped	G	0.57	0.32	$2.18X10^{-6}$	2.83 [1.83-4.37]
106690581	rs12213031	Imputed	T	0.52	0.27	$2.39X10^{-6}$	2.87 [1.84-4.48]
106674727	rs548234	Genotyped	T	0.79	0.57	$5.12X10^{-6}$	2.90 [1.82-4.62]
106675963	rs802791	Imputed	C	0.79	0.57	$5.12X10^{-6}$	2.90 [1.82-4.62]
106700917	rs742109	Genotyped	C	0.71	0.48	$1.31X10^{-5}$	2.61 [1.69-4.04]
106687322	rs6937876	Genotyped	A	0.75	0.54	$2.52X10^{-5}$	2.59 [1.66-4.06]
106688163	rs6939138	Imputed	T	0.75	0.54	$2.52X10^{-5}$	2.59 [1.66-4.06]
106688579	rs9398067	Imputed	C	0.75	0.54	$2.52X10^{-5}$	2.59 [1.66-4.06]
106695307	rs7759216	Imputed	C	0.75	0.54	$2.52X10^{-5}$	2.59 [1.66-4.06]
106695499	rs6568431	Genotyped	C	0.75	0.54	$2.52X10^{-5}$	2.59 [1.66-4.06]
106702780	rs1040893	Imputed	C	0.74	0.53	$4.04X10^{-5}$	2.51 [1.61-3.92]
106704332	rs1012894	Imputed	T	0.74	0.53	$4.04X10^{-5}$	2.51 [1.61-3.92]
106698343	rs11152967	Imputed	C	0.58	0.37	$6.45X10^{-5}$	2.38 [1.55-3.64]
106681487	rs7768653	Imputed	T	0.74	0.54	$6.54X10^{-5}$	2.45 [1.57-3.83]
106683421	rs1883231	Imputed	A	0.74	0.54	$6.54X10^{-5}$	2.45 [1.57-3.83]
106684061	rs4134466	Imputed	G	0.74	0.54	$6.54X10^{-5}$	2.45 [1.57-3.83]
106670247	rs12526490	Imputed	G	0.40	0.21	$8.39X10^{-5}$	2.56 [1.59-4.12]

Supplemental Table 6. Association of imputed SNPs with SMNs in the discovery set.

[a] Genomic position based on NCBI build-36 coordinates

[b] All imputed SNPs had a quality score > 0.9

[c] Risk allele defined as the allele more frequent in the combined cases data set; odds ratios defined with respect to the risk allele

[d] Risk allele frequency

[e] Chi-squared *P* value

[f] Odds ratio [95% confidence interval]

Chr	Position (bp)[a]	SNP	Conditional on:	*P* Value[b]	OR [95% CI] [c]
6	106697056	rs4946728	-	5.33×10^{-9}	3.50 [2.30-5.32]
6	106704716	rs1040411	-	2.67×10^{-7}	2.53 [1.78-3.60]
6	106697056	rs4946728	rs1040411	0.0002	2.61 [1.57-4.35]
6	106704716	rs1040411	rs4946728	0.05	1.55 [1.01-2.40]

Supplemental Table 7. Conditional analysis of genome-wide significant SNPs at 6q21.

[a] Genomic position based on NCBI build-36 coordinates

[b] Logistic *P* value

[c] Odds ratio [95% confidence interval]

SNP A	SNP B	A-B Haplotype	Haplotype	*P* Value[a]
rs4946728	rs4946728	A-C	Risk	$9.88x10^{-10}$
rs4946728	rs1040411	C-C	Mixed	0.9173
rs4946728	rs4946728	C-T	Protective	$6.87x10^{-08}$

Supplemental Table 8. Association of 6q21 haplotypes with SMNs. rs4946728 and rs1040411 form three common haplotypes in Caucasians that represent 99.9% of the haplotypes at this locus.

[a] Asymptotic *P* value comparing the haplotype to all others.

Supplementary References

1. Robison, L.L. et al. Long-term outcomes of adult survivors of childhood cancer. *Cancer* **104**, 2557-64 (2005).
2. Offit, K. et al. Rare variants of ATM and risk for Hodgkin's disease and radiation-associated breast cancers. *Clin Cancer Res* **8**, 3813-9 (2002).
3. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
4. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).
5. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
6. The International HapMap Project. *Nature* **426**, 789-96 (2003).
7. Durbin, R.M. et al. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-73 (2010).
8. Gamazon, E.R. et al. SCAN: SNP and copy number annotation. *Bioinformatics* **26**, 259-62 (2010).