nature genetics

Sequence variants at the *TERT-CLPTM1L* locus associate with many cancer types

Thorunn Rafnar[*,1,44], Patrick Sulem[1,44], Simon N Stacey[1], Frank Geller[1], Julius Gudmundsson[1], Asgeir Sigurdsson[1], Margret Jakobsdottir[1], Hafdis Helgadottir[1], Steinunn Thorlacius[1], Katja K H Aben[2,3], Thorarinn Blöndal[1], Thorgeir E Thorgeirsson[1], Gudmar Thorleifsson[1], Kristleifur Kristjansson[1], Kristin Thorisdottir[4], Rafn Ragnarsson[5], Bardur Sigurgeirsson[4], Halla Skuladottir[6], Tomas Gudbjartsson[7,8], Helgi J Isaksson[9], Gudmundur V Einarsson[10], Kristrun R Benediktsdottir[8,9], Bjarni A Agnarsson[8,9], Karl Olafsson[11], Anna Salvarsdottir[11], Hjordis Bjarnason[1], Margret Asgeirsdottir[1], Kari T Kristinsson[1], Sigurborg Matthiasdottir[1], Steinunn G Sveinsdottir[12], Silvia Polidoro[13,14], Veronica Höiom[15], Rafael Botella-Estrada[16], Kari Hemminki[17], Peter Rudnai[18], D Timothy Bishop[19], Marcello Campagna[20], Eliane Kellen[21,22], Maurice P Zeegers[23,24], Petra de Verdier[25], Ana Ferrer[26], Dolores Isla[26], Maria Jesus Vidal[26], Raquel Andres[26], Berta Saez[27], Pablo Juberias[28], Javier Banzo[29], Sebastian Navarrete[30], Alejandro Tres[26,31], Donghui Kan[32], Annika Lindblom[25], Eugene Gurzau[33], Kvetoslava Koppova[34], Femmie de Vegt[3], Jack A Schalken[35], Henricus F M van der Heijden[36], Hans J Smit[37], René A Termeer[38], Egbert Oosterwijk[35], Onno van Hooij[35], Eduardo Nagore[16], Stefano Porru[20], Gunnar Steineck[15,39], Johan Hansson[15], Frank Buntinx[21,40], William J Catalona[32], Giuseppe Matullo[13,14], Paolo Vineis[13,41], Anne E Kiltie[42], José I Mayordomo[26,31], Rajiv Kumar[17], Lambertus A Kiemeney[2,3,35], Michael L Frigge[1], Thorvaldur Jonsson[7,8], Hafsteinn Saemundsson[11], Rosa B Barkardottir[9], Eirikur Jonsson[10], Steinn Jonsson[8,43], Jon H Olafsson[4,8], Jeffrey R Gulcher[1], Gisli Masson[1], Daniel F Gudbjartsson[1], Augustine Kong[1], Unnur Thorsteinsdottir[1,8] & Kari Stefansson[1,8]

The common sequence variants that have recently been associated with cancer risk are particular to a single cancer type or at most two. Following up on our genome-wide scan of basal cell carcinoma[1], we found that rs401681[C] on chromosome 5p15.33 satisfied our threshold for genome-wide significance (OR = 1.25, $P = 3.7 \times 10^{-12}$). We tested rs401681 for association with 16 additional cancer types in over 30,000 cancer cases and 45,000 controls and found association with lung cancer (OR = 1.15, $P = 7.2 \times 10^{-8}$) and urinary bladder, prostate and cervix cancer (ORs = 1.07–1.31, all $P < 4 \times 10^{-4}$). However, rs401681[C] seems to confer protection against cutaneous melanoma (OR = 0.88, $P = 8.0 \times 10^{-4}$). Notably, most of these cancer types have a strong environmental component to their risk. Investigation of the region led us to rs2736098[A], which showed stronger association with some cancer types. However, neither variant could fully account for the association of the other. rs2736098 corresponds to A305A in the telomerase reverse transcriptase (TERT) protein and rs401681 is in an intron of the *CLPTM1L* gene.

Cancer is caused by a complex interplay between genetic and environmental factors. Highly penetrant mutations explain only a small fraction of cancer cases and most genetic cancer risk is thought to be due to the contribution of many common sequence variants of low penetrance. Recently, genome-wide association (GWA) studies have yielded common sequence variants that associate with cancer risk of the prostate, breast, colon and rectum, lung, urinary bladder and skin[1–14]. Notably, in most cases the variants reported seem to be specific to the particular cancer type under study. This tissue specificity holds true even in the region on chromosome 8q24 where several independent variants have been found that associate with risk of cancer of the prostate, breast and bladder[2,4,6,14–16]. Only one of the prostate cancer variants has been shown also to associate with risk of another cancer, colorectal cancer[7].

We previously conducted a GWA study of basal cell carcinoma (BCC) of the skin and found two signals that reached genome-wide significance in a combined analysis of Icelandic and European sample sets[1]. Here, we followed up the initial GWA scan, increasing the effective sample size by genotyping additional Icelandic BCC cases (total of 1,025 cases) using the Illumina HumanCNV370-duo chip.

[*]A full list of author affiliations appears at the end of the paper.

Received 28 October 2008; accepted 5 November 2008; published online 18 January 2009; doi:10.1038/ng.296

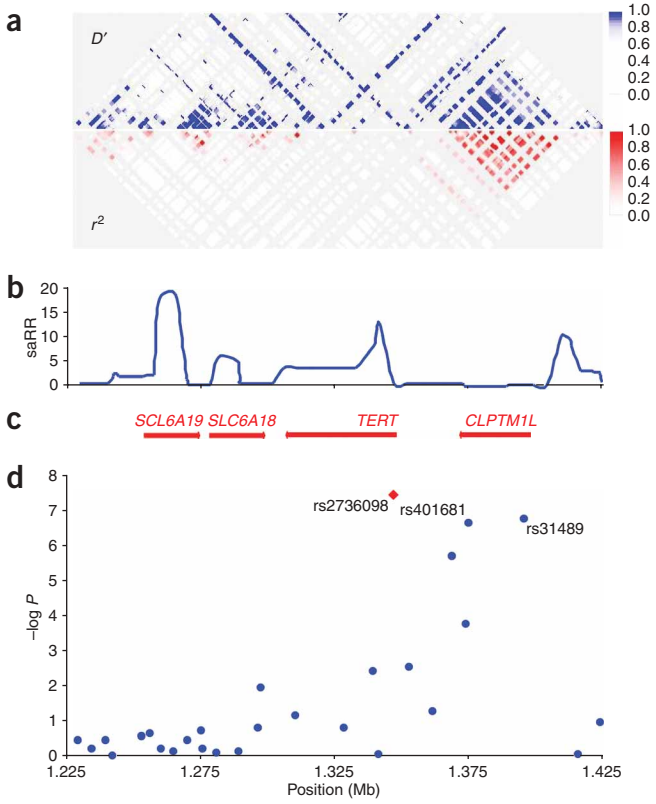

Figure 1 A schematic view of the association results and LD structure in a region on chromosome 5p15.33. (**a**) The pairwise correlation structure in a 200-kb interval (1.225–1.425 Mb, NCBI B36) on chromosome 5. The upper plot shows pairwise D' for 100 common SNPs (with MAF > 5%) from the HapMap (v22) CEU dataset. The lower plot shows the corresponding r^2 values. (**b**) Estimated recombination rates (saRR) in cM/Mb from the HapMap Phase II data. (**c**) Location of known genes in the region. (**d**) Schematic view of the association with BCC in the Icelandic sample set consisting of cases genotyped by chip or *in silico* (blue dots). Red diamond shows the location of rs2736098 and corresponding significance of association to BCC, testing for the HapMap CEU markers absent on the chip for individuals directly genotyped on chip.

Furthermore, we used a method where known genotypes of relatives are used to provide information on BCC cases not genotyped (*in silico* genotyping) to add genotypes that are equivalent to an additional 480 BCC cases[17]. Analysis of this larger dataset confirmed that the two previously reported loci contained the strongest signals (**Supplementary Fig. 1** online). The third strongest signal was on chromosome 5p15.33 in an area of high linkage disequilibrium (LD) and was represented by two correlated SNPs, rs401681 and rs31489 ($D' = 1$ and $r^2 = 0.87$ in data from HapMap CEU and the Icelandic controls; **Fig. 1**). The allele C of rs401681 had an OR of 1.25 (95% CI = 1.15–1.36; $P = 2.3 \times 10^{-7}$) and allele C of rs31489 had an OR of 1.25 (95% CI = 1.15–1.36; $P = 1.9 \times 10^{-7}$) (**Supplementary Table 1** online). We selected rs401681 for follow-up genotyping in an additional 744 BCC cases from Iceland, including some who were used in the initial analysis through *in silico* genotyping, and 525 BCC cases and 515 controls from Eastern Europe (for a summary of cases and controls used in this paper, see **Table 1**). In the combined analysis of the Icelandic and Eastern European samples, the association of rs401681[C] with BCC reached genome-wide significance (OR = 1.25; $P = 3.7 \times 10^{-12}$; **Table 2**). We did not observe heterogeneity of the ORs between the Icelandic and East European groups ($P = 0.35$). Furthermore, judging from the results from both groups, the association between rs401681[C] and BCC did not show significant deviation from the multiplicative risk model ($P > 0.05$).

rs401681 resides in an LD block that contains the *CLPTM1L* (cisplatin resistance related protein CRR9p) gene and the 5′ end of the *TERT* (human telomerase reverse transcriptase) gene. CLPTM1L is a predicted transmembrane protein that is expressed in a range of normal and malignant tissues including skin, lung, breast, ovary and cervix. Expression of CLPTM1L has been shown to sensitize ovarian cancer cells to cisplatin-induced apoptosis[18]. The *TERT* gene encodes the catalytic subunit of the telomerase ribonucleoprotein complex (telomerase). The major role of telomerase is to catalyze the *de novo* addition of telomeric repeat sequences onto chromosome ends and thereby counterbalance telomere-dependent replicative aging[19]. Several studies have reported an association between short telomeres and increased risk of cancer at several sites, including BCC, cancers of the lung, head and neck, bladder, kidney, esophagus and breast, as well as lymphoma[20–24]. Furthermore, this region is frequently amplified in many types of cancer such as lung and cervical cancer[25,26].

Given the relevance of this genomic region to cancer biology, we assessed the association of rs401681[C] with 16 additional cancer types in individuals of European ancestry (**Table 1**). Altogether, we directly genotyped rs401681 on approximately 20,500 cases and 36,000 controls. Using *in silico* genotyping, we also added information corresponding to genotypes of 4,265 Icelandic cancer cases (**Table 1** and **Supplementary Table 2** online). The results from the samples genotyped directly and *in silico* were combined with summary-level data from publicly available GWA datasets, specifically the dataset on colorectal cancer from the UK Institute of Cancer Research (ICR)[27], the dataset on lung cancer from the International Agency for Research on Cancer (IARC)[11] and the datasets on prostate and breast cancer from the Cancer Genetics Markers of Susceptibility (CGEMS) study group[4]. In total, we assessed the association of rs401681[C] with 17 individual cancer sites using more than 33,800 cancer cases and 45,800 controls.

Table 1 Cancer cases and controls used in the study

Cancer site	No. sample sets	No. cases: Directly genotyped	No. cases: Genotyped *in silico*[a]	No. controls
Basal cell carcinoma (skin)	2	2,294	271	29,405
Lung	4	3,613	652	34,666
Bladder	9	3,945	202	34,988
Prostate	4	8,951	522	37,901
Cervix	1	276	93	28,890
Breast	2	3,089	556	30,030
Colon and rectum	2	1,966	529	29,817
Melanoma	3	2,381	62	30,839
Endometrium	1	387	83	28,890
Kidney	2	784	203	30,722
Lymphoma	1	178	70	28,890
Multiple myeloma	1	64	62	28,890
Ovary	1	363	134	28,890
Pancreas	1	75	226	28,890
Squamous cell carcinoma (skin)	1	547	ND	28,890
Stomach	1	277	485	28,890
Thyroid	1	413	115	28,890
Total		**29,603**	**4,265**	**45,846**

[a]Effective sample size, see Methods. ND, not done.

Of the 16 cancer sites tested in addition to BCC, 5 sites showed nominally significant association ($P < 0.05$) with rs401681[C] with the same direction of the effect (**Table 2** and **Supplementary Table 3** online). Of those, 4 cancer sites showed significant association after we accounted for the 16 additional cancer sites tested ($P < 0.05/16 = 0.003$; **Table 2**). These cancer sites are lung, urinary bladder, prostate and cervix of the uterus. The strongest association of rs401681[C], following BCC, was with lung cancer, with an OR of 1.15 ($P = 7.2 \times 10^{-8}$) in the combined analysis of samples from Iceland, The Netherlands and Spain, in addition to the lung cancer dataset from the IARC. We observed an association between rs401681[C] and bladder cancer with a combined OR of 1.12 ($P = 5.7 \times 10^{-5}$) for the nine European case-control groups tested. For prostate cancer, we analyzed data from five groups (over 9,000 cases) and demonstrated a significant effect that is consistent among the groups tested with combined OR = 1.07 ($P = 3.6 \times 10^{-4}$). For cervical cancer, where we only had samples from Iceland, we detected a significant association with rs401681[C] (OR = 1.31, $P = 2.6 \times 10^{-4}$). No signs of heterogeneity or deviation from the multiplicative model were observed.

We did not detect an association between rs401681[C] and breast cancer (OR = 0.98; $P = 0.340$), even though a large sample set was used (3,645 cases and 30,030 controls) (**Supplementary Table 3**). Endometrial cancer showed a trend that did not reach significance after adjustment for the number of tests (OR = 1.21, $P = 5.5 \times 10^{-3}$). Notably, we observed a significant association between rs401681[C] and protection against cutaneous melanoma (OR = 0.88, $P = 8.0 \times 10^{-4}$) in a sample set consisting of 2,443 melanoma cases and 30,839 controls from Iceland, Sweden and Spain. We note that a recently

Table 2 Association of rs401681[C] on 5p15.33 with basal cell carcinoma and cancers of the lung, bladder, prostate and cervix

	Number		Frequency				
Study population	Cases	Controls	Cases	Controls	OR	95% CI	*P* value
Basal cell carcinoma							
Iceland all[a]	2,040	28,890	0.604	0.545	1.27	1.19–1.36	9.5×10^{-12}
Iceland genotyped[b]	1,769	28,890	0.602	0.545	1.26	1.17–1.35	4.3×10^{-10}
Eastern Europe	525	515	0.616	0.575	1.16	0.97–1.39	0.098
All combined[c]	**2,565**	**29,405**	**0.610**	**0.560**	**1.25**	**1.18–1.34**	$\mathbf{3.7 \times 10^{-12}}$
Lung cancer							
Iceland all[a]	1,449	28,890	0.575	0.545	1.13	1.04–1.23	3.6×10^{-3}
Iceland genotyped[b]	797	28,890	0.584	0.545	1.18	1.06–1.32	2.8×10^{-3}
The Netherlands	529	1,832	0.610	0.570	1.18	1.02–1.35	0.021
Spain	367	1,427	0.582	0.538	1.19	1.01–1.41	0.034
IARC	1,920	2,517	0.617	0.586	1.16	1.06–1.27	8×10^{-4}
All combined[c]	**4,265**	**34,666**	**0.596**	**0.560**	**1.15**	**1.10–1.22**	$\mathbf{7.2 \times 10^{-8}}$
Bladder cancer							
Iceland all[a]	780	28,890	0.583	0.545	1.16	1.05–1.29	4.5×10^{-3}
Iceland genotyped[b]	578	28,890	0.583	0.545	1.17	1.03–1.32	0.012
The Netherlands	1,277	1,832	0.584	0.570	1.06	0.96–1.17	0.27
UK	707	506	0.564	0.514	1.23	1.04–1.44	0.014
Italy-Torino	329	379	0.550	0.545	1.02	0.84–1.24	0.84
Italy-Brescia	122	156	0.574	0.564	1.04	0.74–1.46	0.82
Belgium	199	378	0.603	0.554	1.22	0.95–1.56	0.11
Eastern Europe	214	515	0.619	0.575	1.20	0.96–1.51	0.12
Sweden	346	905	0.545	0.521	1.10	0.92–1.31	0.30
Spain	173	1,427	0.546	0.538	1.03	0.83–1.29	0.78
All combined[c]	**4,147**	**34,988**	**0.578**	**0.535**	**1.12**	**1.06–1.18**	$\mathbf{5.7 \times 10^{-5}}$
Prostate cancer							
Iceland all[a]	2,276	28,890	0.569	0.545	1.10	1.03–1.17	3.75×10^{-3}
Iceland genotyped[b]	1,754	28,890	0.564	0.545	1.08	1.00–1.16	0.042
The Netherlands	994	1,832	0.576	0.570	1.02	0.92–1.14	0.67
Chicago, US	635	693	0.581	0.568	1.06	0.90–1.23	0.49
Spain	459	1,427	0.559	0.538	1.09	0.94–1.26	0.27
CGEMS	5,109	5,059	0.558	0.543	1.06	1.00–1.11	0.036
All combined[c]	**9,473**	**37,901**	**0.569**	**0.553**	**1.07**	**1.03–1.11**	$\mathbf{3.6 \times 10^{-4}}$
Cervical cancer							
Iceland all[a]	369	28,890	0.611	0.545	1.31	1.13–1.51	2.6×10^{-4}
Iceland genotyped[b]	**276**	**28,890**	**0.611**	**0.545**	**1.31**	**1.03–1.32**	$\mathbf{1.9 \times 10^{-3}}$

Shown are the numbers of cases and controls (*N*), allelic frequencies, the allelic ORs with *P* values based on the multiplicative model.
[a]Results obtained from combining data from individuals genotyped directly or *in silico*. See supplementary material. [b]Results from directly genotyped individuals only. [c]For the combined study populations, the reported control frequency was the average, unweighted control frequency of the individual populations, and the OR and the *P* value were estimated using the Mantel-Haenszel model.

Table 3 Association of rs2736098[A] on 5p15.33 with basal cell carcinoma and cancers of the lung, bladder, prostate and cervix

	Number		Frequency				
Study population	Cases	Controls	Cases	Controls	OR	95% CI	*P* value
Basal cell carcinoma							
Iceland	1,600	3,667	0.327	0.272	1.30	1.18–1.43	1.4×10^{-7}
Eastern Europe	496	491	0.249	0.264	0.93	0.77–1.12	0.45
All combined[a]	**2,096**	**4,158**	**0.288**	**0.268**	**1.21**	**1.11–1.32**	$\mathbf{1.3 \times 10^{-5}}$
Lung cancer							
Iceland	687	3,667	0.306	0.272	1.18	1.03–1.35	0.014
The Netherlands	525	1,740	0.326	0.286	1.21	1.04–1.41	0.014
Spain	365	1,384	0.271	0.229	1.25	1.04–1.51	0.019
All combined[a]	**1,577**	**6,791**	**0.301**	**0.262**	**1.20**	**1.10–1.31**	$\mathbf{3.2 \times 10^{-5}}$
Bladder cancer							
Iceland	460	3,667	0.283	0.272	1.05	0.90–1.22	0.53
The Netherlands	1,212	1,740	0.308	0.286	1.11	0.90–1.24	0.066
UK	677	486	0.313	0.270	1.24	1.03–1.49	0.023
Italy-Torino	322	375	0.278	0.249	1.16	0.91–1.48	0.23
Italy-Brescia	99	132	0.272	0.227	1.28	0.83–1.97	0.26
Belgium	188	365	0.293	0.271	1.11	0.84–1.46	0.46
Eastern Europe	206	491	0.323	0.264	1.33	1.03–1.71	0.026
Sweden	332	436	0.294	0.227	1.41	1.12–1.77	0.0031
Spain	173	1,384	0.249	0.229	1.11	0.86–1.43	0.42
All combined[a]	**3,669**	**9,076**	**0.290**	**0.255**	**1.16**	**1.08–1.23**	$\mathbf{1.3 \times 10^{-4}}$
Prostate cancer							
Iceland	1,640	3,667	0.290	0.272	1.09	0.99–1.20	0.076
The Netherlands	983	1,740	0.319	0.286	1.17	1.04–1.32	0.0096
Chicago, US	627	679	0.305	0.268	1.20	1.01–1.43	0.039
Spain	449	1,384	0.252	0.229	1.13	0.95–1.34	0.17
All combined[a]	**3,699**	**7,470**	**0.291**	**0.264**	**1.13**	**1.06–1.21**	$\mathbf{1.3 \times 10^{-4}}$
Cervical cancer							
Iceland	249	3,667	0.295	0.272	1.12	0.91–1.37	0.28

Shown are the corresponding numbers of cases and controls (*N*), allelic frequencies of variants, the allelic odds-ratio (OR) with *P* values based on the multiplicative model.
[a]For the combined study populations, the reported control frequency was the average, unweighted control frequency of the individual populations, and the OR and the *P* value were estimated using the Mantel-Haenszel model.

published study of telomere length in individuals with skin cancers has shown that although short telomeres are associated with increased risk of BCC, long telomeres are associated with increased risk of melanoma[24]. The rs401681[C] variant was also marginally associated with protection against colorectal cancer (OR = 0.95, $P = 8.4 \times 10^{-3}$), although this was not significant after we took into account the number of cancer sites tested. We did not observe any association with cancers of the kidney, stomach, thyroid, ovary, pancreas, lymphoma, multiple myeloma or SCC of the skin. However, the moderate sizes of the sample sets tested do not allow us to draw definitive conclusion about the lack of association and further assessment will be needed in larger sample sets.

To explore the potential contributions from other variants in the region, we indirectly tested SNPs that are in the HapMap CEU database, but not on the HumanHap300 or HumanCNV370-duo chips (**Supplementary Methods** online). Using the chip-genotyped Icelandic samples, we selected rs2736100 and rs4975616 to tag rs2736098 for an indirect test of its association with BCC, giving a *P* of 3.9×10^{-8}. rs2736098 is a synonymous coding SNP (A305A) in the second exon of the *TERT* gene (**Fig. 1**). We proceeded to directly genotype rs2736098 in all available cases of the five cancer types showing association with rs401681, in a subset of all the other cancer cases (total of 14,389 cancer cases), in all available non-Icelandic controls and in a subset of the Icelandic controls (total of 9,703 controls). With the exception of cervical cancer, rs2736098[A] was significantly associated with each of the other four cancers that associated significantly with rs401681[C] (**Table 3**). Allele A of rs2736098 is positively correlated with rs401681[C] and the D' between rs2736098 and rs401681 is very high (0.94), but the value of r^2 (0.39), although substantial, is more moderate. For three of the cancers that associated significantly with rs2736098[A]—lung, bladder and prostate—the estimated OR for rs2736098[A] was higher than that for rs401681[C]. In the case of BCC, the OR for rs2736098[A] was, however, lower than that for rs401681[C], mainly because the latter did not show association with the disease in the Eastern European samples (heterogeneity $P = 0.0035$).

We examined the joint effect of rs401681[C] and rs2736098[A], for each of the five cancers, using only samples typed for both SNPs (**Table 4**). After adjustment for rs2736098[A], the association of rs401681[C] remained significant in all except prostate cancer. After adjustment for rs401681[C], rs2736098[A] remained significant for three cancers, lung, bladder and prostate. Overall, these results indicate

Table 4 Joint analysis of rs401681[C] and rs2736098[A] for BCC and cancers of the lung, bladder, prostate and cervix

		rs401681[C] adjusted for rs2736098[A]			rs2736098[A] adjusted for rs401681[C]		
Cancer type	No. populations	OR	95% CI	*P* value	OR	95% CI	*P* value
Basal cell carcinoma	2	1.20	1.10–1.31	7.8×10^{-5}	1.09	0.99 –1.21	0.091
Lung cancer	3	1.11	1.01–1.21	0.024	1.14	1.03–1.25	0.010
Bladder cancer	9	1.07	1.00–1.16	0.036	1.12	1.04–1.20	0.0034
Prostate cancer	4	1.01	0.95–1.08	0.68	1.13	1.05–1.21	0.0015
Cervical cancer	1	1.27	1.03–1.55	0.022	0.97	0.77–1.22	0.80

that neither rs401681[C] nor rs2736098[A] can, by themselves, fully account for the association observed between sequence variants in this region and the five cancer types. This suggests that a unique variant capturing the effect of both rs401681[C] and rs2736098[A] remains to be discovered or, alternatively, that the region contains more than one variant that predisposes to cancers at the same or different sites, analogous to the region on 8q24 where independent variants have been found that associate with different cancer types. We analyzed the association between the 27 SNPs depicted in **Figure 1** and the 17 cancer types studied using the Icelandic sample sets and found that 15 sites showed an association with one or more of these SNPs at the $P < 0.05$ level (**Supplementary Table 4** online).

We assessed known missense variants in *CLPTM1L* and *TERT*, the potentially functional *TERT* promoter variant (–1327T/C), as well as the variable number of tandem repeats (VNTRs) in introns 2, 6 and 12 of *TERT* and found that none of these variants associated significantly with any of the five cancer sites or could account for the association observed with rs401681 or rs2736098 (**Supplementary Note** and **Supplementary Tables 5** and **6** online). Furthermore, no association was observed between rs401681[C] or rs2736098[A] and the RNA expression of *TERT* or *CLPTM1L* in whole blood ($N = 991$) or adipose tissue ($N = 662$) (**Supplementary Note** and **Supplementary Table 7** online).

We postulated that the cancer-associated sequence variants in the *TERT* gene might be associated with shorter telomeres. In order to test this hypothesis, we examined the association between rs401681 and rs2736098 and telomere length in DNA from whole blood, using a quantitative PCR assay. To limit variability, we took into account several factors that have been reported to affect telomere length, including age, sex and smoking status[28,29], and selected from our database 276 females born between 1925 and 1935 who reported to have never smoked and who had not been diagnosed with cancer. To maximize the contrast, only women homozygous for allele C or allele T at rs401681 were included in the test. In these subjects, rs401681[C] and rs2736098[A] were associated with shorter telomeres with nominal significance ($P = 0.017$ and 0.027, respectively; **Supplementary Fig. 2** and **Supplementary Table 8** online). However, when we tested telomere length in a group of 260 younger women (selected by the same criteria regarding smoking and cancer, but born between 1940 and 1950), there was no association between telomere length and the risk alleles. Indeed, the effect estimates, although not significant ($P = 0.08$ and 0.28 for rs401681 and rs2736098, respectively), were in the opposite direction (**Supplementary Fig. 2** and **Supplementary Table 8**). These results suggest that the variants may lead to an increase in the gradual shortening of telomeres over time, the effect only becoming apparent after a certain age. Further testing of these hypotheses is warranted.

We assessed the association of rs401681[C] and rs2736098[A] with the major histological types of lung cancer (**Supplementary Table 9** online). For all histological types except carcinoids, the frequency of the risk variants was higher than in controls, with the highest frequencies found in squamous cell carcinomas. Among the cervix cancer cases, squamous cell carcinomas showed a trend toward stronger association with the risk variants than adenocarcinomas (not significant). In prostate cancer, the variants were not associated with age at diagnosis or disease aggressiveness, as defined by Gleason score ≥ 7 or stage T3 or higher, node positive or metastatic disease, or a combination of these criteria. We found no association between rs401681[C] and either smoking status or nicotine addiction or any of several pigment phenotypes assessed (hair and eye color, freckling or Fitzpatrick skin-type score) (**Supplementary Note**). Furthermore, neither rs401681[C] nor rs2736098[A] was more strongly associated with bladder cancer risk in smokers (current or former, $n = 4,346$) than in nonsmokers ($n = 556$). Finally, rs401681[C] did not associate with longevity ($P = 0.50$) (**Supplementary Note**).

Of note, four of the five cancers associated with the risk variants are cancer types that have strong environmental contribution to risk—smoking and occupational exposures for lung and bladder cancer, UV irradiation for BCC and infection with human papillomavirus for cervical cancer. Most cancers in these organs arise in the epithelial layer that is in closest contact with the environment. Although no strong environmental risk factors are currently known for prostate cancer, several external factors such as diet, physical activity and inflammation may have an effect on disease risk. Although telomere length is partly inherited[30], various environmental factors such as smoking and radiation also affect telomere length[28].

In conclusion, we have discovered sequence variants in the region of the *TERT* and *CLPTM1L* genes that associate with risk of many types of cancer. The biology of the *TERT* gene makes it a compelling candidate for a gene that predisposes to many cancers. Further investigations of the potential effects of genetic variants at the 5p15.33 locus on the functions and expression of *TERT* and *CLPTM1L* are warranted.

METHODS

Genotyping. Detailed information on all case-control sample sets can be found in the **Supplementary Methods**. Whole-genome association studies have been done on the following cancers in the Icelandic population: prostate cancer, breast cancer, lung cancer, BCC, melanoma, urinary bladder cancer and colorectal cancer[1–3,9,14]. All cases and controls were assayed using genotyping systems and specialized software from Illumina (Human Hap300 and HumanCNV370-duo Bead Arrays). Furthermore, all Dutch bladder cancer cases and controls have been genotyped with the HumanCNV370-duo Bead Arrays[14]. These chips provide about 75% genomic coverage in the Utah CEPH (CEU) HapMap samples for common SNPs at $r^2 > 0.8$. We discarded SNP data if the minor allele frequency in the combined case and control was <0.001 or had less than 95% yield or showed a very significant distortion from Hardy-Weinberg equilibrium in the controls ($P < 1 \times 10^{-10}$). Any chips with a call rate below 98% of the SNPs were excluded from the genome-wide association analysis.

All single SNP genotyping was done using the Centaurus (Nanogen) platform. We evaluated the quality of each Centaurus SNP assay by genotyping each assay in the CEU HapMap samples and comparing the results with the HapMap publicly released data. Assays with >1.5% mismatch rate were not used and a linkage disequilibrium (LD) test was used for markers known to be in LD. Approximately 10% of the Icelandic case samples that were genotyped on the Illumina platform were also genotyped using the Centaurus assays and the observed mismatch rate was lower than 0.5%. All genotyping was carried out at deCODE Genetics.

Assessment of telomere length. We selected whole blood as the tissue for analyzing telomere length because of its accessibility, but studies have shown that the length of telomeres is very similar within different tissues of the same individual but varies considerably between individuals. Telomeres were measured using a quantitative Taqman PCR assay. We used the RNAseP endogenous control assay (cat.no. 4316844, Applied Biosystems) to correct for DNA input. This quantitative PCR method has been shown to give as consistent results as Southern blot and FISH-based telomere measurements. All reactions were run on ABI7900TH real-time PCR system (Applied Biosystems). All assays were done in duplicate and repeated in an independent experiment. Primers and probes used are listed in **Supplementary Table 10** online. The use of RNAseP is a standard procedure in gene dosage measurements with real-time qRT-PCR. The main limitation of the method is that it measures relative telomere length rather than actual telomere length. However, in our study the relative telomere length is sufficient to determine whether there is a difference in telomere length between individuals depending on their genotype.

Regression analysis of telomere length data. We analyzed a total of 528 females in two batches, with each batch consisting of three DNA plates: batch 1 included 268 women with a mean age at blood sampling of 72.8 y (s.d. 5.0), batch 2 included 260 women with a mean age at blood sampling of 57.8 y (s.d. 4.6). The relationship between the SNPs showing association and telomere length was analyzed by multiple regression. The logarithm of the ratio between telomere and RNAseP was taken as dependent variable, and the covariates age at blood sampling and DNA plate were included in the models. We analyzed SNPs showing association using multiple linear regression. The experiments were carried out at two different points in time and were analyzed separately.

hTERT minisatellite genotyping. We genotyped the five VNTR polymorphisms in the intronic regions of the hTERT gene by size fractionation of PCR products by gel electrophoresis. Two polymerase chain reaction systems were used: Recombinant Taq (Fermentas) and Extensor-High Fidelity Master Mix (ABgene). PCR conditions were according to the manufacturer's instructions. DNA input material was 30 ng, and primers used are listed in **Supplementary Table 10**.

Analysis of RNA expression. Samples of RNA from human adipose and whole blood were hybridized to Agilent Technologies Human 25K microarrays. Expression changes between two samples were quantified as the mean logarithm ($\log_{10}$) expression ratio (MLR) compared to a reference pool RNA sample. The array probes for *CLPTM1L* and *TERT* genes were in the 3′ untranslated regions of the genes.

Association analysis. A likelihood procedure implemented in the NEMO software was used for the association analyses. An attempt was made to genotype all individuals and all SNPs reported, and for each of the SNPs, the yield was higher than 95% in every study group. We tested the association of an allele with cancer using a standard likelihood ratio statistic that, if the subjects were unrelated, would have asymptotically a χ^2distribution with one degree of freedom under the null hypothesis. Allelic frequencies rather than carrier frequencies are presented for the markers in the main text. Allele-specific ORs and associated *P* values were calculated assuming a multiplicative model for the two chromosomes of an individual. Results from multiple case-control groups were combined using a Mantel-Haenszel model in which the groups were allowed to have different population frequencies for alleles, haplotypes and genotypes but were assumed to have common relative risks. All *P* values are reported as two-sided.

For the analysis of the Icelandic samples, the same set of cancer-free controls used in the BCC discovery analysis was used for all other cancer types, introducing a potential bias. However, because of the lack of association with common cancers like breast and colorectal cancer and also because of the modest effect sizes for the cancers associating with rs401681[C], the frequency of the variant is not substantially different in the Icelandic cancer-free controls (0.545) compared to the whole group of Icelanders ($N = 36{,}139$) genotyped with the BeadChips (0.547), which includes all cancer cases. Therefore, the potential bias introduced into the estimation of the association of the 16 cancers with rs401681[C] is small. Furthermore, this effect is confined to the Icelandic part of our study.

Test of ungenotyped HapMap markers. To test for SNPs that are in the CEU section of the HapMap database, but that are absent on the Illumina chip, we use a method based on haplotypes of two markers on the chip. We computed associations with a linear combination of the different haplotypes chosen to act as surrogates to HapMap markers in the regions. In the 5p13.33 region displayed in **Figure 1** (corresponding to a 200-kb interval), we tested with this method 95 markers in addition to the ones on the chip. These calculations were based on 1,025 BCC cases and 28,890 controls genotyped on the chip. Of those markers, rs2736098 had the most significant association with BCC.

Genomic control and inflation factors. To adjust for possible population stratification and the relatedness among individuals, we divided the χ^2 statistics from the initial scan of basal cell carcinoma in Iceland using the method of genomic control; that is, the 304,000 test statistics were divided by their mean, which was 1.22. In the cases where the method of genomic control was not directly applicable (if the genome-wide association results are not available for the same groups), we used the genealogy to estimate the inflation factor. As some of the Icelandic cases and controls are related to each other, both within and between groups, the χ^2 statistics have a mean >1. We estimated the inflation factor by simulating genotypes through the Icelandic genealogy, as described previously, and corrected the χ^2 statistics for Icelandic ORs accordingly. The estimated inflation factor for different analyses is presented in **Supplementary Table 11** online.

***In silico* genotyping of ungenotyped individuals.** We extended the classical SNP case-control association study design by including ungenotyped cases with genotyped relatives, using information from genotyped individuals in the Icelandic population and the genealogy of all Icelanders. For a detailed description of the calculations of the probability of genotypes, see **Supplementary Methods**.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the individuals that participated in the study and whose contribution made this work possible. We also thank the nurses at deCODE's participant recruitment center and the personnel at deCODE's core facilities. We acknowledge the Icelandic Cancer Registry for assistance in the ascertainment of the Icelandic subjects with cancer. The project was funded in part by the European Commission (POLYGENE: LSHC-CT-2005-018827 and GENADDICT: LSHM-CT-2004-005166), the National Institutes of Health (R01-DA017932) and a research investment grant of the Radboud University Nijmegen Medical Centre. The Leeds Bladder Cancer Study was funded by Cancer Research UK and Yorkshire Cancer Research. Torino Bladder Cancer Case Control Study was supported by a grant to ECNIS (Environmental Cancer Risk, Nutrition and Individual Susceptibility), a network of excellence operating within the European Union 6th Framework Program, Priority 5: "Food Quality and Safety" (Contract No 513943); and by a grant of the compagnia di San Paolo, the Italian Association for Cancer Research, Italy and the Piedmont Region Progetti di Ricerca Sanitaria Finalizzata. J.H. is supported by the Swedish Cancer Society, The Radiumhemmet Research Funds, The Swedish Research Council and the Karolinska Institutet Research Funds.

AUTHOR CONTRIBUTIONS

The study was designed and results were interpreted by T.R., P.S., S.N.S., F.G., J.G., J.R.G., D.F.G., A.K., S.T., U.T. and K.S. Statistical analysis was carried out by P.S., F.G., D.F.G., M.L.F., G.T. and A.K. Subject ascertainment, recruitment, biological material collection and collection of clinical and lifestyle information

was organized and carried out by T.E.T., K. Kristjansson, S.G.S., R.R., B. Sigurgeirsson, K.T., J.H.O., S.J., H.H., T.G., H.J.I., E.J., T.J., G.V.E., R.B.B., K.R.B., B.A.A., H. Skuladottir, K.O., A. Salvarsdottir, H. Saemundsson, J.H., V.H., E.N., S. Polidoro, S. Porru, R.B.-E., R.K., K.H., P.R., K. Koppova, E.G., G.S., D.T.B., A.E.K., M.C., E.K., M.P.Z., P.V., P.d.V., G. Matullo, A.F., D.I., M.J.V., R.A., B. Saez, P.J., J.B., S.N., A.T., D.K., A.L., F.d.V., F.B., W.J.C., J.A.S., H.F.M.v.d.H., H.J.S., R.A.T., E.O., O.v.H., K.K.H.A., J.I.M., L.A.K. Principal collaborators for the non-Icelandic populations were L.A.K. (The Netherlands), J.I.M. (Zaragoza, Spain), A.E.K. (UK), G. Matullo and P.V. (Torino), S.P. (Brescia), M.P.Z. and F.B. (Belgium), R.K. (Eastern Europe), G.S. (bladder cancer, Sweden), J.H. (melanoma, Sweden), E.N. (Valencia, Spain) and W.J.C. (Chicago, USA). Genotyping and laboratory experiments were carried out by A. Sigurdsson, T.B., M.J., H.H., H.B., M.A., K.T.K. and S.M. Bioinformatic analysis was carried out by P.S., T.R., A. Sigurdsson, T.E.T., G. Masson, T.B. and G.T. Authors T.R., P.S., D.F.G., A.K., U.T. and K.S. drafted the manuscript. All authors contributed to the final version of the paper.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Stacey, S.N. *et al.* Common variants on 1p36 and 1q42 are associated with cutaneous basal cell carcinoma but not with melanoma or pigmentation traits. *Nat. Genet.* **40**, 1313–1318 (2008).
2. Gudmundsson, J. *et al.* Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat. Genet.* **39**, 631–637 (2007).
3. Stacey, S.N. *et al.* Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.* **39**, 865–869 (2007).
4. Yeager, M. *et al.* Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* **39**, 645–649 (2007).
5. Haiman, C.A. *et al.* Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat. Genet.* **39**, 638–644 (2007).
6. Easton, D.F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087–1093 (2007).
7. Tomlinson, I. *et al.* A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat. Genet.* **39**, 984–988 (2007).
8. Gudbjartsson, D.F. *et al.* ASIP and TYR pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nat. Genet.* **40**, 886–891 (2008).
9. Thorgeirsson, T.E. *et al.* A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* **452**, 638–642 (2008).
10. Eeles, R.A. *et al.* Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.* **40**, 316–321 (2008).
11. Hung, R.J. *et al.* A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* **452**, 633–637 (2008).
12. Amos, C.I. *et al.* Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat. Genet.* **40**, 616–622 (2008).
13. Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* **40**, 310–315 (2008).
14. Kiemeney, L.A. *et al.* Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat. Genet.* **40**, 1307–1312 (2008).
15. Amundadottir, L.T. *et al.* A common variant associated with prostate cancer in European and African populations. *Nat. Genet.* **38**, 652–658 (2006).
16. Haiman, C.A. *et al.* A common genetic risk factor for colorectal and prostate cancer. *Nat. Genet.* **39**, 954–956 (2007).
17. Gudbjartsson, D.F. *et al.* Many sequence variants affecting diversity of adult human height. *Nat. Genet.* **40**, 609–615 (2008).
18. Yamamoto, K., Okamoto, A., Isonishi, S., Ochiai, K. & Ohtake, Y. A novel gene, CRR9, which was up-regulated in CDDP-resistant ovarian tumor cell line, was associated with apoptosis. *Biochem. Biophys. Res. Commun.* **280**, 1148–1154 (2001).
19. Blackburn, E.H. Switching and signaling at the telomere. *Cell* **106**, 661–673 (2001).
20. Wu, X. *et al.* Telomere dysfunction: a potential cancer predisposition factor. *J. Natl. Cancer Inst.* **95**, 1211–1218 (2003).
21. Shen, J. *et al.* Short telomere length and breast cancer risk: a study in sister sets. *Cancer Res.* **67**, 5538–5544 (2007).
22. Widmann, T.A., Herrmann, M., Taha, N., Konig, J. & Pfreundschuh, M. Short telomeres in aggressive non-Hodgkin's lymphoma as a risk factor in lymphomagenesis. *Exp. Hematol.* **35**, 939–946 (2007).
23. Risques, R.A. *et al.* Leukocyte telomere length predicts cancer risk in Barrett's esophagus. *Cancer Epidemiol. Biomarkers Prev.* **16**, 2649–2655 (2007).
24. Han, J. *et al.* A prospective study of telomere length and the risk of skin cancer. *J. Invest. Dermatol.* advance online publication, doi: 10.1038/jid.2008.238 (31 July 2008).
25. Zhang, A. *et al.* Genetic alterations in cervical carcinomas: frequent low-level amplifications of oncogenes are associated with human papillomavirus infection. *Int. J. Cancer* **101**, 427–433 (2002).
26. Kang, J.U., Koo, S.H., Kwon, K.C., Park, J.W. & Kim, J.M. Gain at chromosomal region 5p15.33, containing TERT, is the most frequent genetic event in early stages of non-small cell lung cancer. *Cancer Genet. Cytogenet.* **182**, 1–11 (2008).
27. Tomlinson, I.P. *et al.* A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.* **40**, 623–630 (2008).
28. Valdes, A.M. *et al.* Obesity, cigarette smoking, and telomere length in women. *Lancet* **366**, 662–664 (2005).
29. Frenck, R.W. Jr., Blackburn, E.H. & Shannon, K.M. The rate of telomere sequence loss in human leukocytes varies with age. *Proc. Natl. Acad. Sci. USA* **95**, 5607–5610 (1998).
30. Slagboom, P.E., Droog, S. & Boomsma, D.I. Genetic determination of telomere size in humans: a twin study of three age groups. *Am. J. Hum. Genet.* **55**, 876–882 (1994).

[1]deCODE Genetics, Sturlugata 8, 101 Reykjavik, Iceland. [2]Comprehensive Cancer Center East, 6501 BG Nijmegen, The Netherlands. [3]Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands. [4]Department of Dermatology, University of Iceland, 101 Reykjavik, Iceland. [5]Departments of Plastic Surgery, [6]Medical Oncology and [7]Surgery, Landspitali-University Hospital, 101 Reykjavik, Iceland. [8]Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland. [9]Departments of Pathology, [10]Urology and [11]Obstetrics and Gynecology, Landspitali-University Hospital, 101 Reykjavik, Iceland. [12]Clinical Research Centre, Krokhals 5D, 110 Reykjavik, Iceland. [13]ISI Foundation (Institute for Scientific Interchange), 10133 Torino, Italy. [14]Department of Genetics, Biology and Biochemistry, University of Torino, 10126 Torino, Italy. [15]Department of Oncology-Pathology, Cancer Centre Karolinska, Karolinska Institutet, Karolinska University Hospital Solna Stockholm, S-171 76, Sweden. [16]Department of Dermatology, Instituto Valenciano de Oncologia, 46009, Valencia, Spain. [17]Division of Molecular Genetic Epidemiology, German Cancer Research Centre, Heidelberg, D-69120, Germany. [18]National Institute of Environmental Health, H-1450 Budapest, Hungary. [19]Section of Epidemiology & Biostatistics, Leeds Institute of Molecular Medicine, St. James's University Hospital, LS9 7TF Leeds, UK. [20]Department of Experimental and Applied Medicine - Section of Occupational Medicine and Industrial Hygiene, University of Brescia, 25125 Brescia, Italy. [21]Department of General Practice, Catholic University of Leuven, 3000 Leuven, Belgium. [22]Leuven University Centre for Cancer Prevention (LUCK), 3000 Leuven, Belgium. [23]Unit of Genetic Epidemiology, Department of Public Health and Epidemiology, University of Birmingham, B15 2TT Birmingham, UK. [24]Department of Complex Genetics, Cluster of Genetics and Cell Biology, Nutrition and Toxicology Research Institute, Maastricht University, 6200 MD Maastricht, The Netherlands. [25]Department of Molecular Medicine and Surgery, Karolinska Institutet, S171 76 Stockholm, Sweden. [26]Division of Medical Oncology, University Hospital, 50009 Zaragoza, Spain. [27]Health Science Institute, 50009 Zaragoza, Spain. [28]Division of Dermatology, University Hospital, 50009 Zaragoza, Spain. [29]Divisions of Nuclear Medicine and [30]Radiation Oncology, University Hospital, 50009 Zaragoza, Spain. [31]Health Science Institute, Nanotechnology Institute of Aragon, 50009, Zaragoza, Spain. [32]Department of Urology, Northwestern University Feinberg School of Medicine, Chicago, Illinois 60611, USA. [33]Environmental Health Center, 400166 Cluj Napoca, Romania. [34]State Health Institute, SK-975 56 Banska Bystrica, Slovakia. [35]Departments of Urology and [36]Pulmonary Diseases, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB, Nijmegen, The Netherlands. [37]Department of Pulmonary Medicine, Rijnstate Hospital, Arnhem, The Netherlands. [38]Department of Pulmonary Diseases, Canisius-Wilhelmina Hospital, Weg door Jonkerbos 100, 6532 SZ, Nijmegen, The Netherlands. [39]Department of Oncology, Sahlgrenska University Hospital, S-413 45 Goteborg, Sweden. [40]Department of General Practice, Maastricht University, 6200 MD Maastricht, The Netherlands. [41]Department of Epidemiology and Public Health, Imperial College, W2 1PG London, UK. [42]Section of Experimental Oncology, Leeds Institute of Molecular Medicine, St. James's University Hospital, LS9 7TF Leeds, UK. [43]Department of Medicine, Landspitali-University Hospital, 101 Reykjavik, Iceland. [44]These authors contributed equally to this work. Correspondence should be addressed to K.S. (kstefans@decode.is) or T.R. (thorunn.rafnar@decode.is).