

Fine-mapping identifies multiple prostate cancer risk loci at 5p15, one of which associates with *TERT* expression

Zsofia Kote-Jarai^{1,*}, Edward J. Saunders^{1,†}, Daniel A. Leongamornlert^{1,†}, Malgorzata Tymrakiewicz^{1,†}, Tokhir Dadaev^{1,†}, Sarah Jugurnauth-Little^{1,†}, Helen Ross-Adams², Ali Amin Al Olama³, Sara Benlloch³, Silvia Halim², Roslin Russel², Alison M. Dunning³, Craig Luccarini³, Joe Dennis³, David E. Neal^{4,5,‡}, Freddie C. Hamdy^{6,7,‡}, Jenny L. Donovan^{8,‡}, Ken Muir^{9,‡}, Graham G. Giles^{10,11,‡}, Gianluca Severi^{10,11,‡}, Fredrik Wiklund^{12,‡}, Henrik Gronberg^{12,‡}, Christopher A. Haiman^{13,‡}, Fredrick Schumacher^{13,‡}, Brian E. Henderson^{13,‡}, Loic Le Marchand^{14,‡}, Sara Lindstrom^{15,‡}, Peter Kraft^{15,‡}, David J. Hunter^{15,‡}, Susan Gapstur^{16,‡}, Stephen Chanock^{17,‡}, Sonja I. Berndt^{17,‡}, Demetrius Albanes^{18,‡}, Gerald Andriole^{19,‡}, Johanna Schleutker^{20,21,‡}, Maren Weischer^{22,‡}, Federico Canzian^{23,‡}, Elio Riboli^{24,‡}, Tim J. Key^{25,‡}, Ruth C. Travis^{26,27,‡}, Daniele Campa^{20,21,‡}, Sue A. Ingles^{13,‡}, Esther M. John^{26,27,‡}, Richard B. Hayes^{28,‡}, Paul Pharoah^{3,‡}, Kay-Tee Khaw^{29,‡}, Janet L. Stanford^{30,31,‡}, Elaine A. Ostrander^{32,‡}, Lisa B. Signorello^{33,34,‡}, Stephen N. Thibodeau^{35,‡}, Dan Schaid^{35,‡}, Christiane Maier^{36,‡}, Walther Vogel^{36,‡}, Adam S. Kibel^{37,‡}, Cezary Cybulski^{38,‡}, Jan Lubinski^{38,‡}, Lisa Cannon-Albright^{39,40,‡}, Hermann Brenner^{41,‡}, Jong Y. Park^{42,‡}, Radka Kaneva^{43,‡}, Jyotsna Batra^{44,45,‡}, Amanda Spurdle^{46,‡}, Judith A. Clements^{44,45,‡}, Manuel R. Teixeira^{47,48,‡}, Koveela Govindasami¹, Michelle Guy¹, Rosemary A. Wilkinson¹, Emma J. Sawyer¹, Angela Morgan¹, Ed Dicks³, Caroline Baynes³, Don Conroy³, Stig E. Bojesen²², Rudolf Kaaks²³, Daniel Vincent⁴⁹, François Bacot⁴⁹, Daniel C. Tessier⁴⁹, COGS-CRUK GWAS-ELLIPSE (Part of GAME-ON) Initiative[¶], The UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology[§], The UK ProtecT Study Collaborators[§], The PRACTICAL Consortium[§], Douglas F. Easton³ and Rosalind A. Eeles¹

¹The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK, ²Cancer Research UK, Li Ka Shing Centre, Cambridge Research Institute, Robinson Way, Cambridge CB2 0RE, UK, ³Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge—Strangeways Laboratory, Worts Causeway, Cambridge CB1 8RN, UK, ⁴Surgical Oncology (Uro-Oncology: S4), University of Cambridge—Addenbrooke's Hospital, Hills Road, Cambridge, UK, ⁵Cancer Research UK, Li Ka Shing Centre, Cambridge Research Institute, Cambridge CB2 2QQ, UK, ⁶Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK, ⁷Faculty of Medical Science, University of Oxford—John Radcliffe Hospital, Oxford OX3 9DU, UK, ⁸School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK, ⁹Warwick Medical School, University of Warwick, Coventry, UK, ¹⁰Cancer Epidemiology Centre, The Cancer Council

*To whom correspondence should be addressed at: Oncogenetics Team, The Institute of Cancer Research, 15 Cotswold Road, Sutton, Surrey SM2 5NG, UK. Tel: +44 2087224027; Fax +44 2087224110; Email: zsofia.kote-jarai@icr.ac.uk

[†]These authors are joint second authors.

[‡]These authors contributed equally to this work.

[¶]See Supplementary Note.

[§]Full list in Supplementary Notes.

© The Author 2013. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permission@oup.com

Victoria, 1 Rathdowne Street, Carlton Victoria, Australia, ¹¹Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, Victoria, Australia, ¹²Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden, ¹³Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA, USA, ¹⁴University of Hawaii Cancer Center, Honolulu, HI, USA, ¹⁵Program in Molecular and Genetic Epidemiology, Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA, ¹⁶Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA, ¹⁷Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD 20892, USA, ¹⁸Nutritional Epidemiology Branch, National Cancer Institute, NIH, EPS-3044, Bethesda, MD 20892, USA, ¹⁹Division of Urologic Surgery, Washington University School of Medicine, St. Louis, MO, USA, ²⁰Department of Medic Biochemistry and Genetics, Institute of Biomedical Technology, University of Turku, Turku, Finland, ²¹BioMediTech, University of Tampere and FimLab Laboratories, Tampere, Finland, ²²Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev Ringvej 75, Herlev DK-230, Denmark, ²³Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany, ²⁴Department of Epidemiology & Biostatistics, School of Public Health, Imperial College London, UK, ²⁵Cancer Epidemiology Unit, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK, ²⁶Cancer Prevention Institute of California, Fremont, CA, USA, ²⁷Stanford University School of Medicine, Stanford, CA, USA, ²⁸Division of Epidemiology, Department of Environmental Medicine, NYU Langone Medical Center, NYU Cancer Institute, New York, NY 1004, USA, ²⁹Clinical Gerontology Unit, University of Cambridge, Cambridge, UK, ³⁰Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, ³¹Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, USA, ³²National Human Genome Research Institute–National Institute of Health, 50 South Drive, Rm. 5351, Bethesda, MD, USA, ³³International Epidemiology Institute, 1555 Research Blvd., Suite 550, Rockville, MD, USA ³⁴Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA, ³⁵Mayo Clinic, Rochester, MN, USA, ³⁶Department of Urology and Institute of Human Genetics, University Hospital Ulm, Ulm, Germany, ³⁷Division of Urologic Surgery, Dana-Farber Cancer Institute, Brigham and Women's Hospital, 45 Francis Street- ASB II-3, Boston, MA 02245, USA, ³⁸International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland, ³⁹Division of Genetic Epidemiology, Department of Medicine, University of Utah School of Medicine, Salt Lake City, UT 84132, USA, ⁴⁰George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, UT 84148, USA, ⁴¹Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany, ⁴²Division of Cancer Prevention and Control, H. Lee Moffitt Cancer Center, 15902 Magnolia Dr., Tampa, FL, USA, ⁴³Molecular Medicine Center and Department of Medical Chemistry and Biochemistry, Medical University–Sofia, 2 Zdrave St, Sofia 1531, Bulgaria, ⁴⁴Australian Prostate Cancer Research Centre–Qld, Institute of Health and Biomedical Innovation, ⁴⁵School of Biomedical Science, Queensland University of Technology, Brisbane, Australia, ⁴⁶Molecular Cancer Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Australia, ⁴⁷Department of Genetics, Portuguese Oncology Institute, Porto, Portugal, ⁴⁸Portugal and Biomedical Sciences Institute (ICBAS), Porto University, Porto, Portugal and ⁴⁹Centre d'innovation Génome Québec et Université McGill, 740 avenue Dr-Penfield, Montréal, QC H3A 0G1, Canada,

Received June 1, 2012; Revised and Accepted February 18, 2013

Associations between single nucleotide polymorphisms (SNPs) at 5p15 and multiple cancer types have been reported. We have previously shown evidence for a strong association between prostate cancer (PrCa) risk and rs2242652 at 5p15, intronic in the telomerase reverse transcriptase (*TERT*) gene that encodes TERT. To comprehensively evaluate the association between genetic variation across this region and PrCa, we performed a fine-mapping analysis by genotyping 134 SNPs using a custom Illumina iSelect array or Sequenom MassArray iPLEX, followed by imputation of 1094 SNPs in 22 301 PrCa cases and 22 320 controls in The PRACTICAL consortium. Multiple stepwise logistic regression analysis identified four signals in the promoter or intronic regions of *TERT* that independently associated with PrCa risk. Gene expression analysis of normal prostate tissue showed evidence that SNPs within one of these regions also associated with *TERT* expression, providing a potential mechanism for predisposition to disease.

INTRODUCTION

We have previously reported an association between prostate cancer (PrCa) risk and rs2242652 on 5p15 (1). rs2242652 lies in intron 4 of telomerase reverse transcriptase (*TERT*) that encodes TERT, the catalytic subunit of the telomerase ribonucleoprotein complex (2). Telomerase catalyzes the *de novo* addition of telomere repeat sequences on to chromosome ends and, thereby, counterbalances telomere-dependent replicative senescence. Several studies have reported an association between shorter telomeres in lymphocytes and increased risk of various cancer types (3–5), although evidence from prospective studies is ambiguous. Associations between single nucleotide polymorphisms (SNPs) in the *TERT* region and multiple cancer types have been reported, and these have been comprehensively reviewed recently (6–8); however, no consistent correlation has been observed thus far between the cancer-associated SNPs in *TERT* and either gene expression or telomere length (TL).

Initial evidence for association with PrCa risk at 5p15 was reported by Rafnar *et al.* (7) for rs401681 and rs2736098 ($P = 3.6 \times 10^{-4}$ and $P = 1.3 \times 10^{-4}$). Subsequently, we found much stronger evidence of association for rs2242652, a SNP only weakly correlated with rs401681 and rs2736098 ($r^2 = 0.19$ and $r^2 = 0.10$, respectively, in Hapmap CEU) (1). We, therefore, concluded that rs2242652 is more strongly associated with variant(s) causally related to PrCa risk. rs2242652 is strongly correlated with rs10069690 ($r^2 = 0.80$) that is associated with oestrogen receptor negative (ER –ve) breast cancer (9), but is not correlated with SNPs previously associated with other cancer types. The *TERT* locus is characterized by low linkage disequilibrium (LD), raising the possibility that additional SNPs could be independently related to PrCa risk and that these could also differ from those predisposing to other cancers.

RESULTS

To further elucidate the association of the 5p15 *TERT* locus with PrCa risk, we have performed a high-resolution fine-mapping of SNPs across the region through a combination of direct genotyping and imputation. Using a custom Illumina iSelect genotyping array (iCOGS) designed for the Collaborative Oncology Gene-Environment Study (http://ec.europa.eu/research/health/medical-research/cancer/fp7-projects/cogs_en.html), we initially genotyped 114 SNPs spanning 135 kb of the *SLC6A18–TERT–CLPTMIL* region in lymphocyte extracted DNA from 22 301 PrCa cases and 22 320 matched controls. These data enabled us to select a narrower 20 kb interval (Chr5:1278590–1299850, GRCh37/hg19) within which variants exhibited substantially stronger associations with PrCa. An additional 25 SNPs within this interval were genotyped in a subset of 2831 PrCa cases and 2440 controls by Sequenom MassArray iPLEX. We imputed all 44 621 samples genotyped in the iCOGS PRACTICAL (<http://ccge.medschl.cam.ac.uk/consortia/practical>) sample set for variants in the 1000 Genome Phase 1 integrated variant set (March 2012) for the interval Chr5:1227693–1361669 using IMPUTE v2.2.2. Concordance between imputed and genotyped SNPs for the 20 SNPs in the Sequenom panel that passed quality control

(QC) was >90% (Materials and Methods and Supplementary Material, Fig. S1). Associations between PrCa risk and the imputed dataset of 1094 SNPs were assessed using a 1 df trend test adjusted for study and six principal components to correct for inflation (10). Samples used in the analysis were predominantly of European single ancestry, and individuals with >15% minority ancestries were excluded (see Materials and Methods and summary data of imputation in Supplementary Material, Table S1). This analysis identified 44 SNPs associated with PrCa risk at $P < 10^{-5}$ (Supplementary Material, Figs S1 and S2 and Supplementary Material, Table S2). To determine independently associated variants in this region, we performed forward and backward stepwise logistic regression (LR); SNPs were included in the model, if they were significant at $P < 10^{-4}$ after adjustment for other SNPs (Table 1 and Supplementary Material, Table S2). Both regression models identified multiple independent associations, reflecting the complexity of this region. Across both models, six SNPs were ascertained to be independent. To further validate their independence, we performed an additional LR analysis using only these SNPs. This retained four SNPs independently significant at $P < 0.05$ (the same SNPs as were selected by the backwards model, Table 1). These SNPs highlight clusters of highly or moderately correlated variants, with only modest LD between these groups of variants, suggesting the presence of four separate regions containing PrCa risk variants (Fig. 1, Supplementary Material, Fig. S2).

Region 1 begins within intron 2 and stretches into intron 4 of *TERT* and contains our previously reported association rs2242652. This variant remained the most strongly associated PrCa risk SNP after univariate analysis ($P = 1.0 \times 10^{-23}$) and remained significant in the forward LR model, whereas the backward LR model identified a different significant SNP, rs7725218, that is only modestly correlated with rs2242652 ($r^2 = 0.40$). However, after the multiple regression analysis, only rs7725218 remained independently significant (Table 1). Several SNPs in this cluster are correlated with these variants at $r^2 > 0.5$, including rs10069690 that was previously reported to be associated with ER –ve breast cancer (9), suggesting that the prostate and breast cancer risks may be driven by the same variant(s).

Region 2 is entirely situated within intron 2 of *TERT* and also contains a portion of the *TERT* promoter CpG island. In the single SNP analysis, the most significant SNP is c5–1291331 ($P = 3.8 \times 10^{-23}$); however, this is no longer significant at $P < 10^{-4}$ after adjustment for other SNPs in the region. Instead, in the backwards LR model, another SNP is identified, rs2853676, and this SNP remained independently significant in the final regression model. This SNP has been reported to be associated with risk of glioma (11). The most studied polymorphism of the *TERT* region, rs2736100, that was reported to be strongly associated with lung cancer and testicular cancer and is in a putative regulatory element (6) is located within this region, but this SNP is only weakly correlated with the PrCa risk association ($r^2 = 0.2$) and was not significant at $P < 10^{-4}$ after adjustment for other SNPs.

Region 3 spans from exon 2 into the near promoter of *TERT*. The most strongly associated SNPs in the single SNP analysis were rs7712562 ($P = 3.8 \times 10^{-23}$) and rs6554754 ($P = 1.1 \times 10^{-18}$). In the conditional analysis, however, the

Table 1. Results of LR analysis

Marker	Region	Position (Hg19/ GRCh37)	Univariate LR OR (95% C.I.)/ <i>P</i> -value	Forward LR OR (95% C.I.)/ <i>P</i> -value	Backward LR OR (95% C.I.)/ <i>P</i> -value	LR of SNPs significant in forward or backward model
rs2242652	1	1 280 028	0.84 (0.81–0.87)/ 1.0×10^{-23}	0.92 (0.88–0.97)/0.002		0.96 (0.91–1.01)/0.134
rs7725218 ^a	1	1 282 414	0.88 (0.86–0.91)/ 3.5×10^{-17}	0.92 (0.89–0.96)/ 3.6×10^{-5}	0.90 (0.87–0.93)/ 2.6×10^{-12}	0.92 (0.88–0.95) / 9.7×10^{-6}
rs2853676 ^a	2	1 288 547	1.09 (1.05–1.12)/ 1.4×10^{-7}		1.08 (1.05–1.12)/ 4.9×10^{-6}	1.06 (1.03–1.10) / 7.7×10^{-4}
rs2853669	3	1 295 349	1.12 (1.08–1.15)/ 2.7×10^{-13}	1.11 (1.08–1.15)/ 7.0×10^{-11}		1.05 (0.98–1.13)/0.131
rs2736107	3	1 297 854	1.12 (1.08–1.15)/ 1.4×10^{-11}		1.16 (1.13–1.20)/ 3.0×10^{-19}	1.09 (1.01–1.17) /0.028
rs13190087 ^a	4	1 298 733	1.20 (1.12–1.29)/ 1.2×10^{-7}	1.18 (1.10–1.26)/ 2.2×10^{-6}	1.19 (1.11–1.27)/ 1.1×10^{-6}	1.18 (1.10–1.27) / 2.4×10^{-6}

The table shows SNPs that remained significant after forward or backward stepwise LR (Forward LR, Backward LR) analyses of 44 imputed or genotyped SNPs in the *TERT* region associated at $P < 10^{-5}$ with PrCa risk in single SNP analysis (Univariate LR). Additional LR analysis of these six SNPs showed that four SNPs (bold) remained independently significant at $P < 0.05$, representing four independent regions.

^aGenotyped SNPs.

evidence for association was defined by two different SNPs, rs2853669 (forward model $P = 1.11 \times 10^{-11}$) and rs2736107 (backward model $P = 1.16 \times 10^{-19}$). rs2853669 has been reported previously to be associated with breast cancer risk (12). Two other SNPs in this region, rs2736108 and rs2736109, which are strongly correlated ($r^2 = 0.94$), have been reported to be associated with breast and ovarian cancer risk (13); these two SNPs are highly correlated with rs2736107 ($r^2 = 0.95$) that remained as an independent signal after multiple LR, whereas rs2853669 did not (Table 1). Although this region extends into the coding sequence, the SNPs that best define it according to the models are all located immediately in the 5' promoter region, suggesting that modulation of *TERT* transcription appears to be the most likely mechanism underlying the risk association at this region.

The fourth association signal, rs13190087, lies 3.5 kb 5' to *TERT*. This SNP is independently significant in both the forward and backward stepwise models and in the final regression analysis. Furthermore, it is not correlated with any of the other association signals (Table 1 and Supplementary Material, Table S2).

To explore the existence of specific risk haplotypes within the association signals, we selected SNPs correlated at $r^2 > 0.2$ with the four 'top' SNPs that had remained significant after multiple regression. Haplotypes containing the top SNP and with a *P*-value smaller than that of any single marker included in the haplotype analysis are shown in Supplementary Material, Table S3. In region 1, the A/A haplotype of rs2242652/rs7725218 (both minor alleles) is more significantly associated with risk than rs7725218 alone (Supplementary Material, Table S3b). This suggests that rs2242652 and rs7725218 (or markers strongly correlated with them) are both related to risk, but combine in a non-multiplicative manner, or that there is a single, as yet untested, causal variant in region 1 partially correlated with both markers. In region 3, the most significant two-marker haplotype (rs2736107/rs2735940) is more significant than rs2736107 alone, again supporting the existing of either two independent

signals or a partially correlated untested causal variant. The haplotype analysis also suggests a possible combined effect of SNPs in regions 2 and 3; the T/T haplotype of rs2853676/ rs7449190 is more significant than single marker effect of rs2853676.

To investigate whether SNPs in any of these regions were associated with *TERT* gene expression, we performed quantitative PCR (qPCR) assays on RNA isolated from 195 histologically benign prostate tissue samples using the Fluidigm Biomark™ HD system. These samples were identified and selected from core biopsy specimens from fresh frozen radical prostatectomy from men with elevated prostate specific antigen (PSA) level (median age 61 years). mRNA samples were analysed for *TERT* and *CLPTMIL* and normalized to housekeeping genes β -actin and 18S RNA. We found evidence that the protective alleles of rs10054203, rs10069690, rs2242652, rs7725218 and rs7713218 (all in region 1) were significantly associated with increased *TERT* expression ($P = 0.01–0.0009$), but no association was observed for *CLPTMIL* (Fig. 2, Supplementary Material, Table S4). We found no evidence for association between any of the SNPs significant in the univariate analysis in regions 2–4 and *TERT* expression. This provides further evidence that the functional basis of the region 1 risk signal differs from that of the other regions.

DISCUSSION

Within the *TERT* locus at 5p15, we have identified four association signals that are independently associated with PrCa risk after multiple LR analysis (Table 1). Haplotype analyses also confirm the existence of four association signals, but identify stronger risk haplotypes in three of the four regions, suggesting either the presence of untyped causal variants in these regions or non-multiplicative interactions between two or more variants. Three of these risk signals are represented by SNPs in localized clusters of moderate LD, whereas the fourth appears to be more tightly defined. These association regions select variants that are intronic or closely upstream for all known transcripts of *TERT*. Whereas the four SNPs

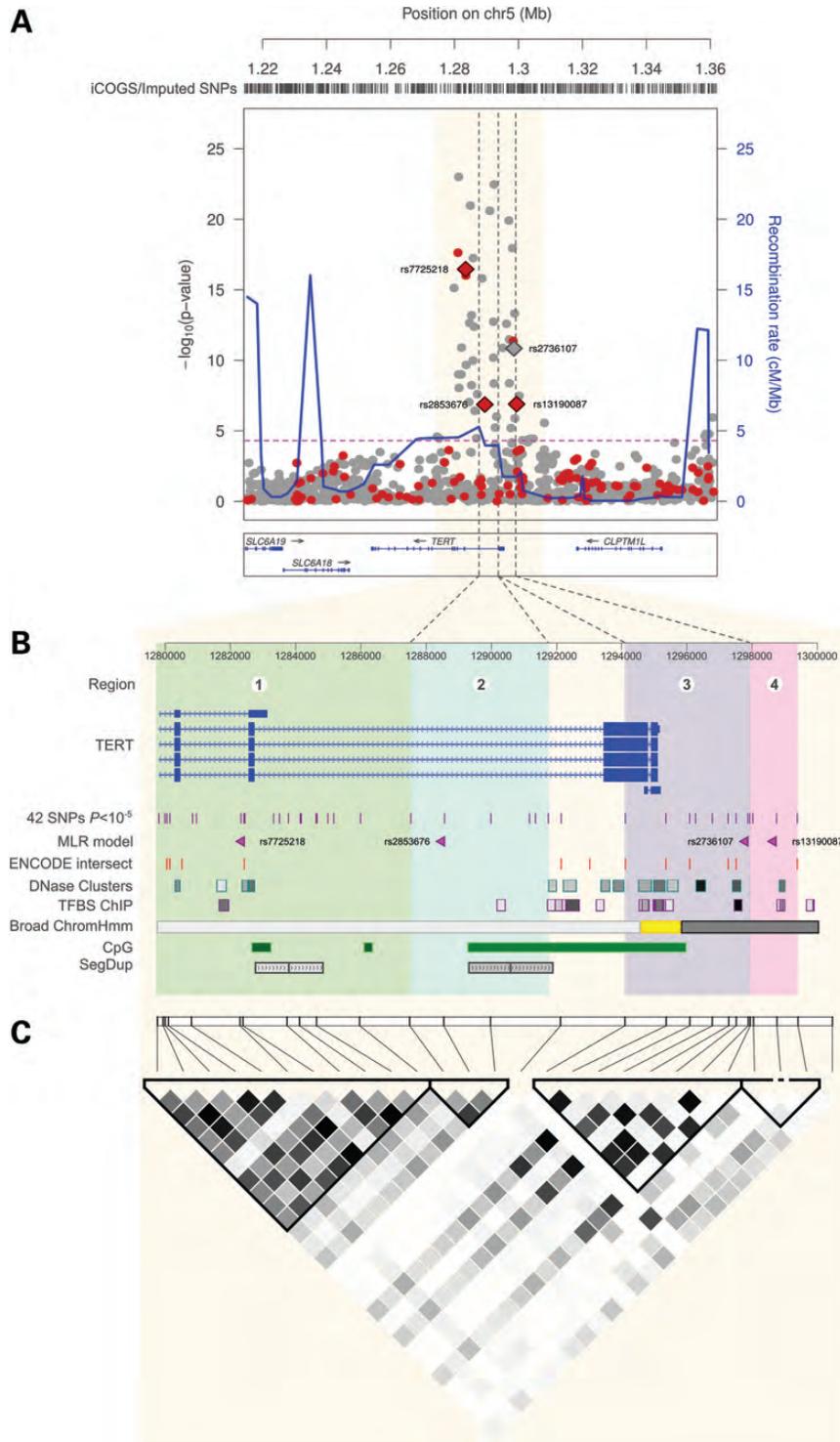


Figure 1. Results of *TERT* fine-mapping analysis. (A) Regional association plot of the imputed iCOGS genotype data. Typed SNPs are indicated in red and imputed SNPs in grey. Diamonds denote SNPs significantly associated with PrCa after multiple LR analyses. The 20 kb interval is denoted by the shaded region that is expanded below. Forty-four SNPs were associated with PrCa after multiple LR analyses at $P < 10^{-5}$ (indicated by the red line). (B) Expanded detail for the 20 kb interval. The positions of 42 SNPs located within this window significant at $P < 10^{-5}$ are marked (42 SNPs $P < 10^{-5}$), as are the 4 SNPs independently significant after multiple LR model and SNPs that overlap with ENCODE annotations (ENCODE intersect), including DNase I hypersensitivity (DNase Clusters) and TFBS CHIP signals. The positions of *TERT* gene transcripts from Ensembl 65 (*TERT*), CpG island regions (CpG), segmental duplications (SegDup) and ENCODE chromatin state (Broad ChromHmm) are also indicated. The light grey rectangle (Broad ChromHmm) denotes a region of heterochromatin, the yellow rectangle a weak enhancer and the dark grey rectangle a polycomb-repressed region. All tracks were generated using the Hg19 build of the UCSC genome browser. The locations of regions 1–4 are indicated as coloured rectangles and numbered. (C) LD plot for the 20 kb interval. r^2 values are derived from imputed data for the UKGPCS subset of iCOGS samples. Triangles indicate the boundaries of regions 1–4. MLR, multiple LR.

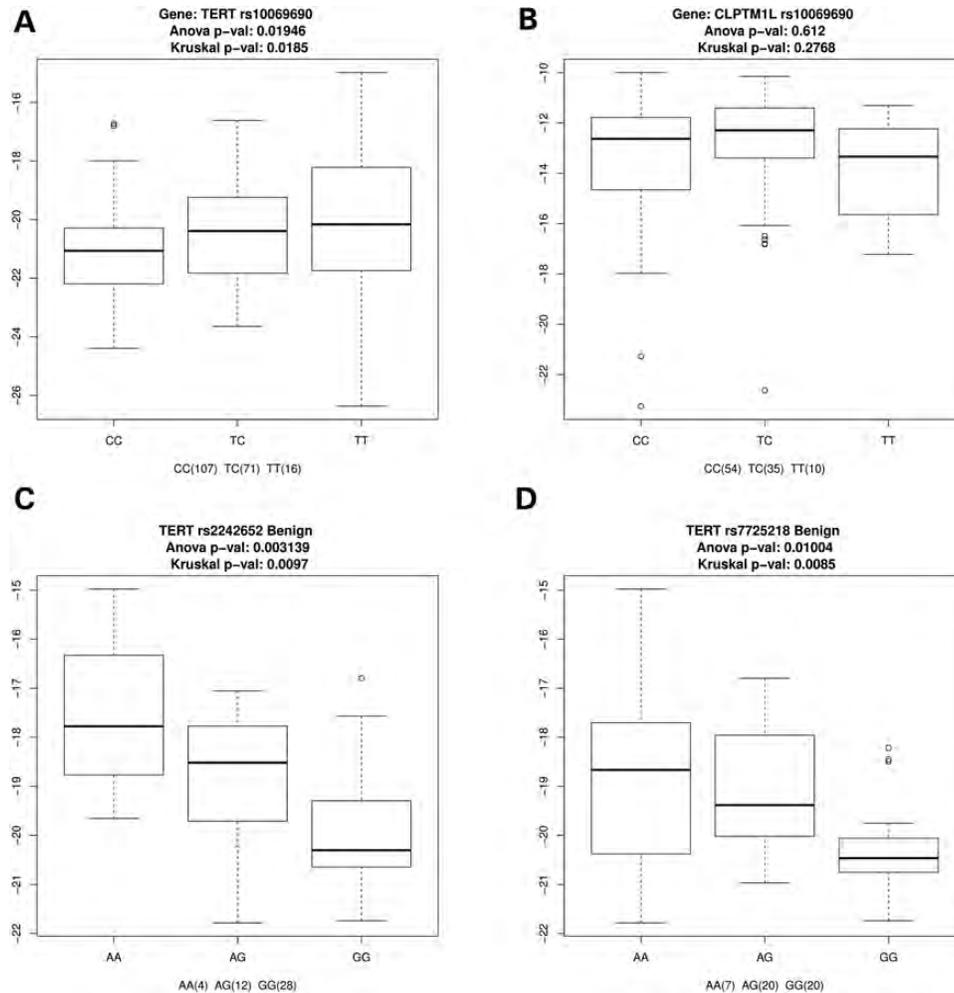


Figure 2. mRNA expression levels in benign prostate tissue for three SNPs in region 1 of the *TERT* locus. A significant increase in *TERT* expression was observed for the minor (protective) alleles of (A) rs10069690, (C) rs2242652 and (D) rs7725218. No effect on expression of the *CLPTM1L* gene was observed, data are shown only for (B) rs10069690.

representing the independently associated signals in the regression models could be candidate causative variants for further analyses, any variants that are correlated with these SNPs could potentially confer the functional effects that modify disease risk.

The regulation of *TERT* has been studied in much detail. There are transcription factor binding sites (TFBS) in the *TERT* promoter for several genes that are known to influence PrCa development and progression while chromatin remodeling via acetylation and methylation also appears to play a critical role (14,15). This implies that the variants we have identified could manifest their effect through modification of these elements. We have shown that SNPs in region 1 are associated with *TERT* expression in benign prostate tissue (Fig. 2, Supplementary Material, Table S4) providing evidence that variants in this region may affect PrCa risk through regulation of gene expression.

Our analysis identified four independent association signals at the *TERT* locus; however, the precise functional variants that are responsible for altering the risk of PrCa remain to be established and could arise through any variants in LD

with the SNPs we have identified. Comparing our findings with functional data from the Encyclopedia of DNA Elements (ENCODE) Project (16) [obtained through HaploReg (17) and the UCSC genome browser (18)] can help to predict the most likely functional SNPs (Fig. 1, Supplementary Material, Table S5). In region 1, rs7725218, the SNP that remained significant in the final analysis, is situated within a DNase I hypersensitivity region and predicted to alter an Mrg TFBS. In addition, rs2242652, which is in moderate LD with rs7725218 ($r^2 = 0.4$), is also situated in a DNase I hypersensitivity region and predicted to disrupt HEN1, Zfx and E2A TFBS consensus sequences. The minor, lower risk alleles of both these variants are associated with increased *TERT* expression (Fig. 2) that would be consistent with these SNPs modifying functional regulatory elements. In addition, another SNP rs7734992 also overlaps a DNase I hypersensitivity region and is predicted to alter an Mtf1 TFBS. Region 3 encompasses the near promoter region of the *TERT* gene and as expected contains several variants with potential functional effects. rs2853669, which was significant in the forward analysis only, is located immediately 5' to the *TERT* transcription

start site, within a DNase I hypersensitivity region. ChIP-seq data indicate that this SNP is situated within an RNA polymerase II binding site, whereas histone modification data suggest that it lies inside a weak enhancer element. This SNP is also predicted to disrupt an RBP-Jkappa TFBS and has previously been demonstrated to modify telomerase activity in lung cancer cells (19), providing further support for a direct functional effect arising from this SNP. Another SNP in region 3 that ENCODE data suggests may exert a functional effect is rs2736108. This SNP lies within a DNase I hypersensitivity site, and ChIP-seq data indicate that it is within an EBF1 TFBS. This SNP did not itself remain significant after LR analysis; however, it is very highly correlated with rs2736107 ($r^2 = 0.95$), the SNP in region 3 that remained significant after the multiple regression analysis. Lastly, rs2736098, which is also correlated with rs2736107 ($r^2 = 0.8$), is located within a DNase I hypersensitivity region and is predicted to alter TFBS for NRSF and LRF. The SNP that defines region 4 according to all statistical models, rs13190087, has no obvious functional effect itself and, however, is correlated with one other variant, rs71595003 ($r^2 = 0.67$). This SNP overlaps a DNase I hypersensitivity site, and ChIP-seq data also indicate that it overlaps TFBS for TCF12 and MAFK, although it is also predicted to disrupt an aryl hydrocarbon receptor binding motif.

In addition to the biological insights provided by the ENCODE project, (20) showed that rs7705526 in region 1 and SNPs in region 3, including rs2736108, are strongly associated with mean TL in lymphocytes. Whereas the correlation between the region 1 TL SNP and our PrCa risk SNPs is weak, the variants associated with PrCa and TL are strongly correlated in region 3 ($r^2 = 0.94$); therefore, it remains possible that this region could influence PrCa risk through a TL-dependent mechanism.

Overall, our results demonstrate that four sets of variants within a narrow interval at 5p15 are independently associated with PrCa risk and that one of these regions significantly affects *TERT* expression. It has been reported previously that elevated *TERT* expression improves PrCa survival (21), and we have demonstrated that the lower risk alleles of variants in region 1 are associated with elevated *TERT* expression, thereby suggesting a plausible mechanism by which these variants could affect disease. Deep re-sequencing of this region may provide further insight by helping to uncover additional associated variants, further refine these loci and facilitate selection of prospective causal variants for functional validation studies. The phenomenon whereby multiple loci are subsequently identified to explain an initial GWAS association signal has also been observed for other PrCa regions such as 11q13 and 8q24 and highlights the value of fine-scale mapping of risk associations to fully define their contribution to cancer susceptibility.

MATERIALS AND METHODS

Samples

Samples for the iCOGS replication were drawn from 25 studies participating in the PRACTICAL Consortium. The majority of studies were population-based or hospital-based case-control studies, or nested case-control studies; some studies

selected samples by age or oversampled for cases with a family history of PrCa. In total, genotype data for 22 301 PrCa cases and 22 320 matched controls were available after QC (10). A subset of 2831 cases and 2440 controls from the UKGPCS study were selected for genotyping by Sequenom iPLEX MassARRAY technology.

Genotyping of 5p15 SNPs on the iCOGS chip

All known SNPs from the March 2010 (Build 36) release of the 1000 Genomes Project with minor allele frequency >0.02 in Europeans in a 135 kb interval (Chr5:1227693-1361669) encompassing the *SLC6A18*, *TERT* and *CLPTMIL* genes were identified. All SNPs correlated at $r^2 > 0.1$ with a published cancer association, plus an additional tagging set to cover the remaining known SNPs, were included on the array. This generated a panel of 114 SNPs that were genotyped using a custom Illumina Infinium array (iCOGS).

Selection and genotyping of further SNPs

Based on iCOGS data, the SNPs associated with PrCa clustered within an ~ 20 kb interval (Chr5:1278590-1299850), with no SNPs outside of this region showing evidence for association (Fig. 1). Data from the 1000 genomes project (1000 Genomes August 2010 dataset called by Broad in Nov 2011 across 283 European samples) indicated that the PrCa interval contained 104 putative SNPs, of which 52 had minor allele frequency (MAF) $>2\%$. To fine-map the PrCa susceptibility region at high depth, we used the Tagger feature of Haploview to design a panel to capture all MAF $>2\%$ variants at $r^2 > 0.9$. These criteria required genotyping of 45 SNPs, 17 of which had previously been genotyped on the iCOGS array (6 were significant at $P < 10^{-6}$, a further 3 at $P < 10^{-4}$ and the remainder showed no evidence of association). Additionally, a proxy search using the 1000 Genomes Pilot 1 CEU panel was performed to identify any further SNPs correlated at $r^2 > 0.4$ with rs2242652 or any of the iCOGS $P < 10^{-4}$ SNPs. This added further 6 SNPs to the fine-mapping panel, bringing the number of SNPs to be genotyped in addition to the iCOGS array to 34.

Genotyping assays were designed using the Sequenom MassARRAY Assay Designer 4.0 software. During the assay design process, nine SNPs in RepeatMasked or segmentally duplicated regions were unable to be designed and were excluded. The remaining 25 SNPs were genotyped using the Sequenom MassARRAY iPLEX Platform (Sequenom, San Diego, CA, USA), of which 20 passed QC: SNPs were excluded, if more than 15% of samples failed.

All assays were performed in 384-well plates, including a mix of cases and controls, with 4 blank samples and 8 random duplicates for QC. Duplicate samples were 99.6% concordant.

Imputation

Imputation was performed on 22 301 cases and 22 320 control samples across 114 iCOGS SNPs from the *TERT* interval that passed pre-imputation QC metrics: missing genotypes $\leq 3\%$, MAF >0.01 and Hardy-Weinberg Equilibrium among controls $P < 10^{-6}$ (10). IMPUTE v2.2.2 (22) was

used to impute the interval Chr5:1227693-1361669 (GRCh37/hg19) using a 1000 Genomes Phase 1 integrated variant set (SNPs and indels) from 5 March 2012, settings in Supplementary Material, Figure S1.

This generated an iCOGS imputed dataset of 1094 SNPs. Concordance was checked by two methods; firstly, 5271 samples were analysed for concordance across the 20 SNPs genotyped by Sequenom, but not on the iCOGS chip, with concordance of >90%. Secondly, IMPUTE v2.2.2 'leave one out' internal concordance check gave 86.3% concordance at SNPs $r^2 \geq 0.3$ and 90.1% concordance at SNPs $r^2 \geq 0.9$ with the 114 SNPs on the iCOGS chip across all 44 621 samples (for a full breakdown by r^2 , see Supplementary Material, Table S6). Given the high concordance across both methods, we performed imputation using a 1000 Genomes variant set alone, without implementing a two panel imputation.

Statistical analysis

Association tests were performed on genotypes in the MaCH dosage format (0–2) converted from the IMPUTE genotype posterior probabilities using GenABEL (23), and haplotype analyses were performed on 'best guess' genotypes converted using GenGen; calls are generated only, if the posterior probability is higher than 0.9, unless otherwise stated.

Associations between each SNP and PrCa risk were analysed using a per-allele trend test, adjusted for study and six principal components (10). Odds ratios (ORs) and 95% confidence limits were estimated using unconditional LR. Tests of homogeneity of the ORs across strata were assessed using likelihood ratio test. SNPs significant at $P < 10^{-5}$ were considered for further analysis. To determine independently associated SNPs, we used forward and backward stepwise LR; SNPs were included in the model, if they were significant at $P < 10^{-4}$ after adjustment for other SNPs. To further assess the independence of these associations, an additional LR analysis was performed using the SNPs retained in these models.

Haplotype analyses (Chi-squared test) were performed using Unphased 3.16 (24) using all marker combinations and a window size of two. Haplotypes were filtered to select only haplotypes containing the top SNP and with a P -value smaller than that of any single marker. These haplotypes were then rerun in PLINK (25) (LR), to correct for the same covariates used in the original association analyses.

Gene expression analysis

Tissue sections were obtained from biopsies taken from fresh frozen radical prostatectomy samples of 195 European men (mean age 61.5 years). Ten to 14 cores from each biopsy sample were excised, and the pathology of each core was determined based on the H&E staining of the two adjacent sections. All patients who underwent surgery had elevated (>3 ng/ml) PSA levels (mean PSA 9.52 ng/ml, range 3.4–40 ng/ml). qPCR assays were performed using the Fluidigm Biomark™ HD system with 48 × 48 and 96 × 96 dynamic array plates according to the manufacturer's instructions. TaqMan assays for *TERT* Hs00972656_m1 and Hs00972649_m1 were tested, but only assay Hs00972656_m1 worked reliably, so all data generated were based on this. Each assay was performed in

triplicate on each plate, and at least two replicate plates were run for each assay. Other TaqMan assays included Hs00363947_m1 (*CLPTMIL*), 4319413E (18S RNA) and 4326315E (*β-actin*). Data for all repeats were normalized to housekeeping genes *β-actin* and 18S RNA. Multiple 'no template' control samples were included in each reaction plate. Data were also normalized across reaction plates through the inclusion of three commercially sourced 'control' RNA samples across all reaction plates. Clontech qPCR human reference total RNA (Clontech, Mountain View, CA, USA, Cat No. 636 690), Ambion FirstChoice human brain RNA reference (Life Technologies Corporation, Carlsbad, CA, USA, Cat No. AM6050) and Applied Biosystems' TaqMan control total RNA (human) (Life Technologies Corporation, Carlsbad, CA, USA, Cat No. 4 307 281) also acted as positive controls for target gene expression. In addition, 1000 permutation tests were performed on the available data. Hits with Kruskal–Wallis $P < 0.05$ were considered significant.

URLS

http://ec.europa.eu/research/health/medical-research/cancer/fp7-projects/cogs_en.html
<http://ccge.medschl.cam.ac.uk/consortia/practical>
<http://pngu.mgh.harvard.edu/purcell/plink/>
http://mathgen.stats.ox.ac.uk/impute/impute_v2.html
<http://www.openbioinformatics.org/genegen/index.html>
<http://genome.ucsc.edu/>
<http://www.broadinstitute.org/mammals/haploreg/haploreg.php>

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank all the patients and control men who took part in this study. Further acknowledgements for individual studies and individual investigators are listed in the Supplementary Material.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by European Commission's Seventh Framework Programme grant agreement No. 223175 (HEALTH-F2-2009-223175), Cancer Research UK Grants C5047/A7357, C1287/A10118, C5047/A3354, C5047/A10692, C16913/A6135 and The National Institute of Health (NIH) Cancer Post-Cancer GWAS initiative grant: No. 1 U19 CA 148537-01 (the GAME-ON initiative). We would like to thank the following for funding support: The Institute of Cancer Research and The Everyman Campaign, The Prostate Cancer Research Foundation (now Prostate Action), Prostate Research Campaign UK (now Prostate Action), The Orchid Cancer Appeal, The National Cancer Research Network, UK and The National Cancer Research Institute (NCRI), UK. We are grateful for support of NIHR funding

to the NIHR Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. Funding to pay the Open Access publication charges for this article was provided by European Commission's Seventh Framework Programme grant agreement No. 223175 (HEALTH-F2-2009-223175).

REFERENCES

- Kote-Jarai, Z., Al Olama, A.A., Giles, G.G., Severi, G., Schleutker, J., Weischer, M., Campa, D., Riboli, E., Key, T., Gronberg, H. *et al.* (2011) Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nat. Genet.*, **43**, 785–791.
- Wyatt, H.D., West, S.C. and Beattie, T.L. (2010) InTERTpreting telomerase structure and function. *Nucleic Acids Res.*, **38**, 5609–5622.
- Jang, J.S., Choi, Y.Y., Lee, W.K., Choi, J.E., Cha, S.I., Kim, Y.J., Kim, C.H., Kam, S., Jung, T.H. and Park, J.Y. (2008) Telomere length and the risk of lung cancer. *Cancer Sci.*, **99**, 1385–1389.
- Martinez-Delgado, B., Yanowsky, K., Inglada-Perez, L., de la, H.M., Caldes, T., Vega, A., Blanco, A., Martin, T., Gonzalez-Sarmiento, R., Blasco, M. *et al.* (2012) Shorter telomere length is associated with increased ovarian cancer risk in both familial and sporadic cases. *J. Med. Genet.*, **49**, 341–344.
- Pooley, K.A., Sandhu, M.S., Tyrer, J., Shah, M., Driver, K.E., Luben, R.N., Bingham, S.A., Ponder, B.A., Pharoah, P.D., Khaw, K.T. *et al.* (2010) Telomere length in prospective and retrospective cancer case-control studies. *Cancer Res.*, **70**, 3170–3176.
- Mocellin, S., Verdi, D., Pooley, K.A., Landi, M.T., Egan, K.M., Baird, D.M., Prescott, J., De Vivo, I. and Nitti, D. (2012) Telomerase reverse transcriptase locus polymorphisms and cancer risk: a field synopsis and meta-analysis. *J. Natl. Cancer Inst.*, **104**, 840–854.
- Rafnar, T., Sulem, P., Stacey, S.N., Geller, F., Gudmundsson, J., Sigurdsson, A., Jakobsdottir, M., Helgadóttir, H., Thorlacius, S., Aben, K.K. *et al.* (2009) Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. *Nat. Genet.*, **41**, 221–227.
- Turnbull, C., Rapley, E.A., Seal, S., Pernet, D., Renwick, A., Hughes, D., Ricketts, M., Linger, R., Nsengimana, J., Deloukas, P. *et al.* (2010) Variants near DMRT1, TERT and ATF7IP are associated with testicular germ cell cancer. *Nat. Genet.*, **42**, 604–607.
- Haiman, C.A., Chen, G.K., Vachon, C.M., Canzian, F., Dunning, A., Millikan, R.C., Wang, X., Ademuyiwa, F., Ahmed, S., Ambrosone, C.B. *et al.* (2011) A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat. Genet.*, **43**, 1210–1214.
- Eeles, R.A., Al Olama, A.A., Benlloch, B., Saunders, E.J., Leongamornlert, D.A., Tymrakiewicz, M., Ghousaini, M., Luccarini, C., Dennis, J., Jugurnauth-Little, S. *et al.* (2013) Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat. Genet.*, in press.
- Shete, S., Lau, C.C., Houlston, R.S., Claus, E.B., Barnholtz-Sloan, J., Lai, R., Il'yasova, D., Schildkraut, J., Sadetzki, S., Johansen, C. *et al.* (2011) Genome-wide high-density SNP linkage search for glioma susceptibility loci: results from the Gliogene Consortium. *Cancer Res.*, **71**, 7568–7575.
- Shen, J., Gammon, M.D., Wu, H.C., Terry, M.B., Wang, Q., Bradshaw, P.T., Teitelbaum, S.L., Neugut, A.I. and Santella, R.M. (2010) Multiple genetic variants in telomere pathway genes and breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **19**, 219–228.
- Beesley, J., Pickett, H.A., Johnatty, S.E., Dunning, A.M., Chen, X., Li, J., Michailidou, K., Lu, Y., Rider, D.N., Palmieri, R.T. *et al.* (2011) Functional polymorphisms in the TERT promoter are associated with risk of serous epithelial ovarian and breast cancers. *PLoS One*, **6**, e24987.
- Kyo, S., Takakura, M., Taira, T., Kanaya, T., Itoh, H., Yutsudo, M., Ariga, H. and Inoue, M. (2000) Sp1 cooperates with c-Myc to activate transcription of the human telomerase reverse transcriptase gene (hTERT). *Nucleic Acids Res.*, **28**, 669–677.
- Kyo, S., Takakura, M., Fujiwara, T. and Inoue, M. (2008) Understanding and exploiting hTERT promoter regulation for diagnosis and treatment of human cancers. *Cancer Sci.*, **99**, 1528–1538.
- The ENCODE Project Consortium. (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.*, **9**, e1001046.
- Ward, L.D. and Kellis, M. (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.*, **40**, D930–D934.
- Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M. and Haussler, D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
- Hsu, C.P., Hsu, N.Y., Lee, L.W. and Ko, J.L. (2006) Ets2 binding site single nucleotide polymorphism at the hTERT gene promoter—effect on telomerase expression and telomere length maintenance in non-small cell lung cancer. *Eur. J. Cancer*, **42**, 1466–1474.
- Bojesen, S.E., Pooley, K.A., Johnatty, S.A., Beesley, J., Michailidou, K., Tyrer, J.P., Edwards, S.L., Pickett, H.A., Shen, H.C., Smart, C.H. *et al.* (2013) Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat. Genet.*, in press.
- Taylor, B.S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B., Arora, V.K., Kaushik, P., Cerami, E., Reva, B. *et al.* (2010) Integrative genomic profiling of human prostate cancer. *Cancer Cell*, **18**, 11–22.
- Howie, B.N., Donnelly, P. and Marchini, J. (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.
- Aulchenko, Y.S., Ripke, S., Isaacs, A. and van Duijn, C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, **23**, 1294–1296.
- Dudbridge, F. (2008) Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum. Hered.*, **66**, 87–98.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. and Sham, P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.