RESEARCH ARTICLE

Germ-line variation at a functional p53 binding site increases susceptibility to breast cancer development

Jianjun Liu · Kartiki Vasant Desai · Yuqing Li · Shakeela Banu · Yew Kok Lee ·
Dianbo Qu · Tuomas Heikkinen · Kirsimari Aaltonen · Taru A. Muranen ·
Tasneem Shabbir Kajiji · Carine Bonnard · Kristiina Aittomäki · Karl von Smitten ·
Carl Blomqvist · John L. Hopper · Melissa C. Southey · Hiltrud Brauch ·
The GENICA Consortium · Georgia Chenevix-Trench · Jonathan Beesley ·
Amanda B. Spurdle · Xiaoqing Chen · Kathleen Cuningham Foundation Consortium
for Research into Familial Breast Cancer · Australian Ovarian Cancer Study Group ·
Kamila Czene · Per Hall · Heli Nevanlinna · Edison T. Liu

Abstract Multiple lines of evidence suggest regulatory variation to play an important role in phenotypic evolution and disease development, but few regulatory polymorphisms have been characterized genetically and molecularly. Recent technological advances have made it possible to identify bona fide regulatory sequences experimentally on a genome-wide scale and opened the window for the biological interrogation of germ-line polymorphisms within these sequences. In this study, through a forward genetic analysis of bona fide p53 binding sites identified by a genome-wide chromatin immunoprecipitation and sequence analysis, we discovered a SNP (rs1860746) within the motif sequence of a p53 binding site where p53 can function as a regulator of transcription. We found that the minor allele (T) binds p53 poorly and has low transcriptional regulation activity as compared to the major allele (G). Significantly, the homozygosity of the minor allele was found to be associated with an increased risk of ER negative breast cancer (OR = 1.47, $P = 0.038$) from the analysis of five independent breast cancer samples of European origin consisting of 6,127 breast cancer patients and 5,197 controls. rs1860746 resides in the third intron of the PRKAG2 gene that encodes the γ subunit of the AMPK protein, a major sensor of metabolic stress and a modulator of p53 action. However, this gene does not appear to be regulated by p53 in lymphoblastoid cell lines nor in a cancer cell line. These results suggest that either the rs1860746 locus regulates another gene through distant interactions, or that this locus is in linkage disequilibrium

Gene Environment Interaction and Breast Cancer in Germany:Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ U. Hamann), Heidelberg, Germany; Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University of Tübingen (H Brauch, C Justenhoven), Tübingen, Germany; Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn (YD Ko, C Baisch), Germany; BGFA - Research Institute of Occupational Medicine of the German Social Accident Insurance, Ruhr University Bochum (T Bruening, B Pesch, V Harth, S Rabstein), Germany.

J. Liu (✉) · Y. Li · D. Qu · T. S. Kajiji · C. Bonnard
Human Genetics, Genome Institute of Singapore,
Singapore 138672, Singapore
e-mail: liuj3@gis.a-star.edu.sg

K. V. Desai · S. Banu · Y. K. Lee · E. T. Liu (✉)
Cancer Biology, Genome Institute of Singapore,
Singapore 138672, Singapore
e-mail: liue@gis.a-star.edu.sg

T. Heikkinen · K. Aaltonen · T. A. Muranen · H. Nevanlinna
Department of Obstetrics and Gynecology, Helsinki University
Central Hospital, Helsinki, Finland

K. Aaltonen · C. Blomqvist
Department of Oncology, Helsinki University Central Hospital,
Helsinki, Finland

K. Aittomäki
Department of Clinical Genetics, Helsinki University Central
Hospital, Helsinki, Finland

K. von Smitten
Department of Surgery, Helsinki University Central Hospital,
Helsinki, Finland

with a second causal mutation. This study shows the feasibility of using genomic scale molecular data to uncover disease associated SNPs, but underscores the complexity of determining the function of regulatory variants in human populations.

Introduction

There is a great interest in the role of regulatory variation in phenotypic evaluation and disease development. Early evolutionary biologists suggested that genetic variation within regulatory sequences is the main driving force behind phenotypic evolution (Wray 2007). Based largely on the observation that coding sequences usually show limited divergence between closely related species such as human and chimpanzees (King and Wilson 1975), it was argued that such moderate divergence of coding sequences cannot account for the profound phenotypic difference between species and it was proposed that regulatory mutations within non-coding sequences (regulatory variation) constitutes the main genetic basis of phenotypic evolution. This interest on regulatory variation is also motivated by the discovery of several regulatory variants that underline complex disease traits (Knight 2005) and further augmented by the recent findings that gene expression shows substantial variation in human and

many model organisms, and that such expression variation is, at least partially, due to germ-line genetic variation (Sladek and Hudson 2006).

Despite this great interest, only few regulatory polymorphisms have been characterized molecularly and linked to disease development. For example, a T/G polymorphism within the intronic promoter of MDM2, a strong negative regulator of p53 protein activity, was shown to influence the binding affinity of transcription activator Sp1 and thus the expression level of MDM2, which in turn resulted in decreased levels of p53 protein, leading to accelerated tumor formation (Bond et al. 2004). Recently, the meta-analysis of all the 21 subsequent association studies of the polymorphism showed a convincing association of the homozygous genotype of the minor allele with an increased risk for cancer development, especially in lung cancer and smoking-related cancers (Hu et al. 2007). The limited progress made on identifying disease-related regulatory variants is largely due to a difficulty in delineating regulatory sequences and thus their germ-line polymorphisms. The identification of regulatory sequences has been pursued by using a comparative genomics approach where the conservation analysis of non-coding sequences between closely related species is used as a primary tool for inferring regulatory sequences. Such an approach is prone to high false discoveries and is confounded by the fact that many functional transcription factor binding sites reside on species specific repetitive sequences (Bourque et al. 2008). Using eQTL strategies, recent efforts attempted to map regulatory variation to particular genomic regions (Pastinen and Hudson 2004), but the extensive linkage disequilibrium in human and model organism genomes limits the mapping resolution of such genetical genomic analyses. Recent advances on technologies like chromatin immunoprecipitation (ChIP) followed by hybridization to an array chip (ChIP-Chip) (Cawley et al. 2004) or by shotgun sequencing of ChIP pull-down DNA fragments (ChIP-seq or ChIP-PET) (Cawley et al. 2004; Wei et al. 2006) have made it possible to identify bona fide regulatory sequences on a genome-wide scale. These binding sites can then be subjected to functional interrogation of germ-line polymorphisms within the binding sequences.

In this study, we performed a forward genetic study of a group of the p53 binding sites that were identified through a genome-wide ChIP-PET analysis (Wei et al. 2006). By the identification of known single nucleotide polymorphisms (SNPs) within p53 binding motif sequences and subsequently the molecular and genetic characterization of such polymorphisms, we found a common SNP within an intronic p53 binding motif that could influence p53 protein binding, transcription regulation and breast cancer development.

K. Czene · P. Hall
Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

C. Blomqvist
Department of Oncology, Radiology and Clinical Immunology, Uppsala University Hospital, Uppsala, Sweden

J. L. Hopper · M. C. Southey
Australian Breast Cancer Family Study, Departments of Public Health and Pathology, The University of Melbourne, Melbourne, Australia

H. Brauch
Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology Stuttgart, University of Tübingen, Tübingen, Germany

G. Chenevix-Trench · J. Beesley · A. B. Spurdle · X. Chen · Australian Ovarian Cancer Study Group
Queensland Institute of Medical Research, Brisbane, Australia

Kathleen Cuningham Foundation Consortium
for Research into Familial Breast Cancer · Australian Ovarian Cancer Study Group
Peter MacCallum Cancer Center, Melbourne, VIC, Australia

Methods

Samples

The current study included the clinical samples from five breast cancer studies of European women (Table 1). The discovery sample set consisted of 3,512 cases and 2,739 controls from the SASBAC (Sweden) and HEBCS (Finland) studies, and the validation sample set included 2,615 cases and 2,458 controls from the GENICA (Germany), ABCFS (Australia) and kConFab (Australia and New Zeland).

The SASBAC study consisted of 1,596 breast cancer patients that were randomly selected from a population-based Swedish cohort that included all Swedish-born breast cancer patients between 50 and 74 years of age and resident in Sweden between October 1993 and March 1995. A total of 1,730 age-matched controls were randomly selected from the Swedish Registry of Total Population. A total of 1,290 cases and 1,483 controls of the SASBAC study provided DNA and were successfully genotyped in the current study. Finnish breast cancer cases consisted of two series of unselected breast cancer patients and additional familial cases ascertained at the Helsinki University Central

Hospital. The first series of 884 patients was collected in 1997–1998 and 2000 and covers 79% of all consecutive, newly diagnosed cases during the collection periods. And the second series, containing 986 consecutive newly diagnosed patients, was collected in 2001–2004 and covers 87% of all such patients treated at the hospital during the collection period. An additional 538 familial breast cancer cases were collected at the same hospital. A total of 1,287 anonymous, healthy female population controls were collected from the same geographical regions in Southern Finland as the cases. A total of 2,222 cases and 1,256 controls of the HEBCS study were successfully genotyped in the current study. The ABCFS study is a population based case–control-family study. Briefly, 1,610 cases were composed of three age-groups of patients from two metropolitan areas selected in 1992–1999 in Melbourne and 1993–1998 in Sydney, Australia. A total of 1,077 controls were identified from the electoral rolls from the same geography areas with 5-year age frequency matched. A total of 1,117 cases and 601 controls of the ABCFS study provided DNA and were successfully genotyped in the current study. The GENICA study consisted of 1,021 incident breast cancer cases and 1,015 age matched controls enrolled between 2000 and 2004 from the Greater Bonn area, Germany. The controls were

Table 1 Summary of the samples from the five breast cancer studies of European population

Study	Controls	Cases							
		All	Menopause		Family history		ER status	
			Pre-	Post-	Yes	No	ER−	ER+
HEBCS	1,256	2,222	502	1,171	899	1,297	408	1,716
SASBAC	1,483	1,290	14	1,276	199	1,062	158	694
ABCFS	601	1,117	786	294	137	980		
GENICA	1,002	1,015	247	754	160	855	214	751
KCONFAB	855	483	129	331	340	61	73	176
Total	5,197	6,127	1,678	3,826	1,735	4,255	853	3,337

Table 2 Analysis of the association of SNP with cancer susceptibility under a recessive model of inheritance in the combined Swedish and Finnish samples

Sample	Size	Genotype Freq (%)		OR (95% CI)	P value
		GG or GT	TT		
Controls	2,739	97.08	2.92		
All cases	3,512	96.07	3.93	1.34 (1.01, 1.77)	0.04
ER+ cases	2,492	96.27	3.73	1.26 (0.92, 1.72)	0.15
ER− cases	588	95.58	4.42	1.48 (0.93, 2.35)	0.10
Postmenopausal cases	2,447	96.24	3.76	1.30 (0.96, 1.76)	0.10
Premenopausal cases	502	94.82	5.18	1.66 (1.00, 2.75)	0.05
Sporadic cases	2,359	96.31	3.69	1.25 (0.92, 1.70)	0.16
Familial cases	1,098	95.72	4.28	1.48 (1.01, 2.17)	0.04

randomly selected from population registries for 31 communities in the greater Bonn area who matched to cases in 5-year age classes. Both cases and controls were of Caucasian ethnicity and below 80 years of age. A total of 1,015 cases and 1,002 controls of the GENICA study were successfully genotyped in the current study. The sample of the kConFab study consisted of 640 cases from multiple-case breast and breast-ovarian families recruited though family cancer clinics across Australia and New Zealand from 1998 to the present. A total of 1,009 controls were ascertained by the Australian Ovarian Cancer Study identified from the electoral rolls from all over Australia from 2002 to 2006. From these studies, 483 patients with no family history of mutations in BRCA1 and BRCA2, or who were the index for the family, and the youngest breast cancer affected family member, and 855 controls provided DNA and were successfully genotyped in the current study. All the samples of the five studies were used in several previous genetic association studies (Easton et al. 2007; Ahmed et al. 2009; Dunning et al. 2009). In total, 6,127 breast cancer cases and 5,197 controls were analyzed in the current study.

All the samples were recruited with informed consent, and this study was approved by local Institutional Review Boards.

SNP genotyping

Genotyping analysis of SNPs was performed by using the MALDI-TOF mass spectrometry-based MassARRAY™ system from the Sequenom (San Diego, CA, US) (the samples of SASBAC, ABCFS, GENICA, and KONFAB) as well as the TaqMan assays from the AppliedBiosystesm (ABI) (Foster City, CA, US) (HSBCA). All genotyping plates included positive and negative controls, DNA samples were randomly assigned to the plates, and all genotyping results were generated and checked by laboratory staff unaware of case–control status. The genotype frequency was in Hardy–Weinberg equilibrium in each of the five samples.

Lymphoblatoid cell lines and culture

Lymphoblastoid cell lines (LCLs) used in this study were obtained from the Coriell depository (http://www.coriell.org/). Cells were cultured in RPMI medium supplemented with 20% fetal bovine serum. For ChIP, real-time qPCR and western blot analyses, cells were treated with 5FU at the concentration of 375 μM for various numbers of hours. All the drug treatments were done during the log phase of cell growth (about 1–1.5 millions of cells per ml). Cells were harvested after culture with or without drug treatment(s) and stored at −80 degrees. 5FU was obtained from the Sigma.

ChIP analysis

ChIP assays were performed in LCLs using the protocol described previously (Weinmann and Farnham 2002; Wells and Farnham 2002). The DO1 monoclonal antibody for p53 (Santa Crux Biotechnology, Santa Cruz, CA) was used for immunoprecipitation, and real-time quantitative PCR analyses were performed in triplicate using the PRISM 7,900 Sequence Detection System and the SYBR protocol as described (Wei et al. 2006). The real-time PCR analysis was performed using the following primers: *CCATCCT GCCTGAGCATGTCTGAAC* (forward) and *CCGGCTTTG CCAGACAATTGG* (reverse) (For *PRKAG2*); *CAGGCTG TGGCTCTGATTGGCTTTC* (forward) and *GCTGGCAGAT CACATACCCTGTTCAGAGTA* (reverse) (For CDKN1A); *ACCCACACTGTGCCCATCTACGAG* (forward) and *TCT CCTTAATGTCACGCACGATTTCC* (reverse) (For *Actin*). Relative occupancy was calculated by determining the immunoprecipitation efficiency (ratios of the amount of immunoprecipitated DNA over that of the input sample) and normalized to the level observed at a control region, which was defined as 1.0. The control region was a distal site around the binding site for Actin and not enriched by the immunoprecipitation.

Allele enrichment analysis of ChIP pull-down DNAs by real-time PCR

The allele enrichment analysis of the ChIP input and pull-down DNAs from heterozygous cell lines was performed in triplicate by real-time quantitative PCR using a made-to-order TaqMan SNP assay for rs1804674 from the ABI. The quality of the TaqMan SNP assay was first verified by genotyping 30 CEPH DNA samples, and all the genotype results are consistent with the ones from the HAPMAP project (data not shown). For real-time PCR analysis, the Ct value difference (ΔCt) between G and T alleles of a ChIP pull-down DNA was normalized by the ΔCt value of the corresponding input DNA (reflecting the equal numbers of G and T alleles in normal genomic DNAs from the heterozygous cell lines). The normalized ΔCt value ($\Delta\Delta$Ct) was then used to calculate the enrichments (Fold Change using the formula of $2^{\Delta\Delta Ct}$) of the wild-type G allele over the mutant T allele in the ChIP pull-down DNA.

Expression analysis by real-time PCR

We studied the induction of p53 in LCLs using 50, 100, 150, and 375 μM 5FU. We achieved maximal response at 100 μM 5FU with minimal cell death over 48 h of treatment (data not shown). To determine gene expressions changes, cells were plated at 0.3 million cells per ml and treated with 100 μM 5FU or DMSO (vehicle) for 4, 8, and

24 h. For each time point, cells were harvested and total RNA was extracted using QIAGEN RNeasy Kit. The SuperScript III First-Strand Synthesis System (Invitrogen, CA, USA) was used to reverse transcribe 2 μg of total RNA to 20 μl cDNA. cDNA was diluted to 80 and 2 μl was used as template for real-time PCR.

Real-time PCR analysis was done in the ABI Prism 7700 sequence detection system using SYBR Green from ABI. Primers were designed using the online Primer 3 program. Primer sequences are as follows: PRKAG2: *CCC TATCAGTGGGAATGCAC* (forward), *GCTCATCCAGGTT CTGCTTC* (reverse); CDKN1A: *TTAGCAGCGGAACAAG GAGT* (forward), *CAACTACTCCCAGCCCCATA* (reverse); Beta-ACTIN: *TCCCTGGAGAAGAGCTACGA* (forward), *AGGAAGGAAGGCTGGAAGAG* (reverse). Ct values obtained for PRKAG2 and CDKN1A were normalized to Beta-ACTIN Ct values. The normalized Ct (ΔCt) values were then used to calculate the difference ($\Delta\Delta$Ct) between 5FU and DMSO treated samples for each timepoint. Fold change of PRKAG2 and CDKN1A at each timepoint was then calculated as $2^{-\Delta\Delta Ct}$.

Primers for PRKAG2 isoforms

Since we did not observe change in PRKAG2 expression using primers that detect all isoforms, we hypothesized that p53 may differentially regulate them. Primers were designed using the online NCBI Primer-BLAST program, ensuring specificity of primer pairs to the intended template. Primer sequences are as follows: PRKAG2 isoform 'a' (NM_0162 03.3): *CGCGTGCACATTCCGGACCT* (forward), *GCCGAA GGGGCTGTCCACCT* (reverse) PRKAG2 isoform 'b' (NM_024429.1): *CCTCCTCCTCCCCCTCAGGC* (forward), *TGAGTCTTCTACTGCTTCGTCCTCG* (reverse) PRKAG2 isoform 'c' (NM_001040633.1): *CCAGGACCCCTTGGGG CTGA* (forward), *CCGAAGGGGCTGTCCACCTTT* (reverse). However, none of these isoforms were changed following 5FU treatment (data not shown).

Promoter assay analysis

A 226 bp region encompassing the intronic p53 binding site within *PRKAG2* was amplified using hotstart PCR with forward primer 5′-TAGGAGACCTGGGGGGACTTT-3′ and reverse primer 5′-CAGGCATCTCGAAGAGATCA-3′ and 50 ng of genomic DNAs isolated from the individuals carrying either the wild-type (WT) G or mutant (MUT) A allele. The PCR conditions were; 94°C for 15 min, followed by 35 cycles of denaturation at 94°C for 45 s, annealing 55°C for 45 s, and extension at 72°C for 45 s. The resultant PCR products of 226 bp were purified from agarose gels and cloned using TOPO-TA cloning system (Invitrogen, Calsbad, CA). The genotypes of the cloned

DNA fragments were confirmed by DNA sequencing. Subsequently, the DNA fragments were subcloned into the upstream of TATA-luciferase (fire-fly) containing pGL4 vector (Promega) using Kpn I and Xho I restriction enzymes (New England Biolabs).

Reporter assay analysis was performed by using both HCT116 wild type and null for p53 cells (provided by Dr Bert Vogelstein's lab at the Johns Hopkins School of Medicine) that were maintained in DMEM containing 10% fetal bovine serum. 5×10^4 cells were plated in triplicate in 24-well plates and transfected next day with 250 ng of either parent TATA-luc, WT-TATA-luc or MUT-TATA-luc plasmid DNAs under serum free conditions using 1 μg per well of Lipofectamine 2000 (Invitrogen, Calsbad, CA). 2.5 ng of pRL-CMV vector containing renilla luciferase was co-transfected in each well to normalize transfection efficiency across wells. After 8 h the cells were recovered for 3 h in serum containing medium, following which the cells were treated for 12 h with 375 μM 5-Fluorouracil or DMSO. The cells were lysed in passive lysis buffer and promoter assays were carried out as per manufacturer's instructions using Promega Dual-luciferase assay system. The values obtained for each construct were normalized as fold-change to that of the activity of parental TATA-luc vector in HCT116 WT cells (designated as 1).

Statistical analysis

Hardy–Weinberg Equilibrium (HWE) test was performed in the five control samples separately, and no evidence for deviation from HWE was found. Association analysis was performed using the χ^2 test under a recessive model of inheritance. For the joint association analyses of the combined samples, the Mantel–Haenszel method for meta-analysis was used by assuming fixed effect. All statistical analyses were performed by using the StataSE8 system.

Results

Identification of p53 binding site SNPs

Of 542 high confidence p53 binding sites identified in HCT116 cell line by our genome-wide ChIP-PET mapping analysis (Wei et al. 2006), we selected 235 sites for SNP mining where an unequivocal p53 consensus binding motif sequence (5′-RRRCWWGYYYRRRCWWGYYY-3′) can be found (see Supplementary Fig. 1 for the position weighted matrix presentation of the consensus sequence). The sequences of the 235 binding sites were blasted against the dbSNP database (version 115), and 14 SNPs were identified to be directly located within the binding motifs. Of the 14 SNPs, 12 SNPs were successfully genotyped in

76 anonymous germ-line DNA samples in Caucasian population, and six SNPs were confirmed to be polymorphic with a minor allele frequency (MAF) above 1%.

Of the six confirmed p53 binding motif germ-line polymorphisms, rs1860746 was found to be located within the consensus motif sequence of an p53 binding site in the third intron of the PRKAG2 gene where high p53 protein occupancy was observed (Wei et al. 2006). rs1860746 (a G/T substitution) is located at one of the highly conserved bases of p53 motif sequence, and its minor allele T causes a mismatch to the p53 consensus motif sequence: 5′-RRRCWWGYYYRRRCWW[G/T]YYY-3′. According to the data from the HapMap project, the SNP is common in African and Caucasian populations (MAF = 20%, confirmed by our genotyping analysis), but extremely rare in Asian populations (Chinese and Japanese) (MAF = 1%).

PRKAG2 encodes the gamma two noncatalytic subunit of the AMPK protein complex, a central sensor of energy stress. The known involvement of AMPK and p53 in cancer development and its interesting frequency pattern in different populations encouraged us to characterize the molecular function of this germ-line p53 binding motif polymorphism in cancer development.

Functional characterization of the binding site and its germ-line polymorphism (rs1860746)

To characterize the molecular function of this p53 binding site and its polymorphism, we chose lymphoblastoid cell lines (LCLs) as in vitro system because LCLs have a normal diploid genome and a large collection of cell lines where cells carrying different genotypes of germ-line SNPs

are available for functional analysis. Further, western blot analysis showed that the p53 protein in LCLs could be induced in a time-dependent fashion by 5FU treatment (Supplementary Fig. 2).

First, we performed the ChIP analysis in eight LCL cell lines: three homozygous for the mutant T allele; two homozygous for the wild-type G allele, and three heterozygous. A significant enrichment of the binding site sequence was observed at the baseline and further augmented after 5FU treatment (for 10 h) in the five cell lines carrying either one or two copies of the wild-type G allele (12-fold enrichment on average), whereas the three cell lines carrying two copies of the mutant T allele showed little enrichment of binding sequence (Fig. 1a) (twofold enrichment on average). In addition, we performed real-time PCR analysis to directly measure the relative abundance of the wild-type (G allele) and mutant (T allele) motif sequences in the ChIP pull-down DNAs from the three heterozygous cell lines (after 5FU treatment for 6 or 32 h) and found fivefold to tenfold enrichment of the wild-type G allele over the mutant T allele in the ChIP pull-down DNAs (Fig. 1b). The enrichment of the wild-type G over mutant T allele could also be observed at the baseline, although the enrichment was less prominent. The ChIP analyses clearly showed that the p53 protein has a higher binding affinity to the wild-type G allele than to the mutant T allele. As a control, a similar enrichment of p53 binding at the p21 (CDKN1A, a well-characterized p53 target gene) promoter at the baseline (about 200-fold) as well as after 5FU treatment for 6 (about 300-fold higher) and 10 h (about 500-fold higher) was observed in the cell lines carrying either G or T allele at rs1860746 (Supplement Fig. 3).
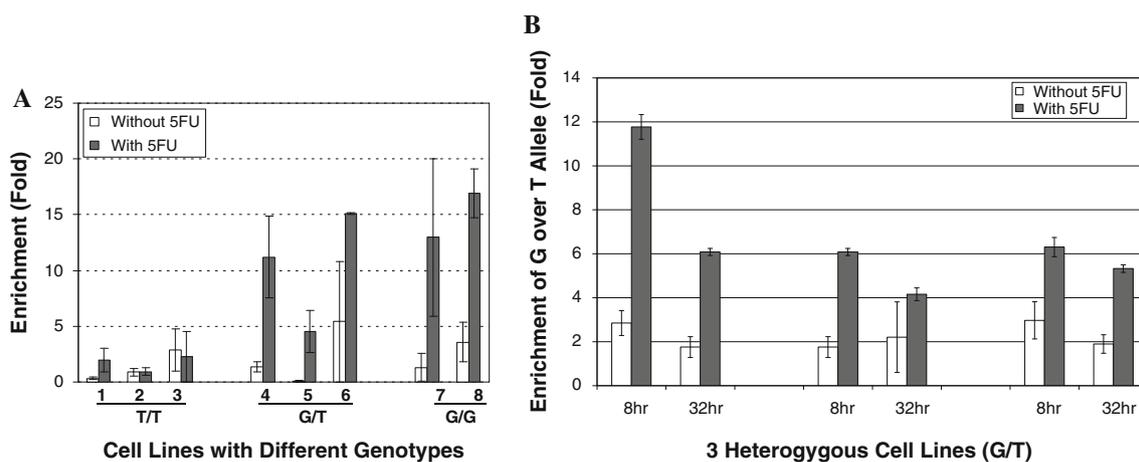

Fig. 1 The results from ChIP and real-time PCR analyses, showing that the wild-type allele (G) is associated with stronger p53 binding activity than the mutant allele (T) in LCLs. **a** The differential enrichment of the binding site sequence at the baseline and after 5FU treatment in the cell lines carrying either only wild-type allele (G/G) (two cell lines), or mutant (T/T) allele (three cell lines), or both alleles (G/T) (three cell lines). **b** The enrichment of the wild-type G allele over the mutant T allele in the ChIP pull-down DNAs from the three heterozygous cell lines (G/T) after 5FU treatment for 8 and 32 h

Subsequently, we measured the transcription regulation activities of the wild-type and mutant binding site sequences through a reporter assay analysis. Both wild-type and mutant binding site sequences were cloned into a TATA-luciferase reporter vector and then transfected into HCT116 cells with either wild-type p53 protein or with the p53 disrupted by homologous recombination (p53 null). In the p53 wild-type HCT116 cells, the presence of the wild-type binding site sequence strongly induced the expression of the reporter gene (20-fold induction), and the induction is augmented by the activation of p53 by 5FU treatment (about 30-fold induction) (Fig. 2). In the p53 null HCT116 cells, this induction was largely abolished. In both p53 wild-type and null HCT116 cells, the mutant binding site sequence (T allele) showed a minimal induction of the report gene expression. Together with the results of ChIP analysis, our results demonstrate that this is a functional p53 binding site whose binding and regulatory activity can be disrupted by the SNP identified.

In six different lymphoblastoid cell lines (three with the TT genotype, and three with the GG genotype), the PRKAG2 gene, however, did not change in expression after p53 induction with 5FU regardless of the genotype at the rs1860746 locus (Supplementary Fig. 4, Panel A). Moreover, no protein product of the PRKAG2 gene could be detected in the cell lines (data not shown). In contrast, p21 (CDKN1A) showed increased expression at each time point following 5FU treatment, demonstrating that p53 was functionally induced in these cell lines (Supplementary Fig. 4, Panel B). We surmise that either the binding site at the rs1860746 locus does not regulate the PPKAG2 gene transcription or that this regulation is silent in lymphoblastoid cell lines. Furthermore, after treatment with 5FU, no induction of PRKAG2 by p53 was observed in the HCT116 cell line that harbors wild-type p53 and is responsive to p53 action (Tan et al. 2005) (data not shown).
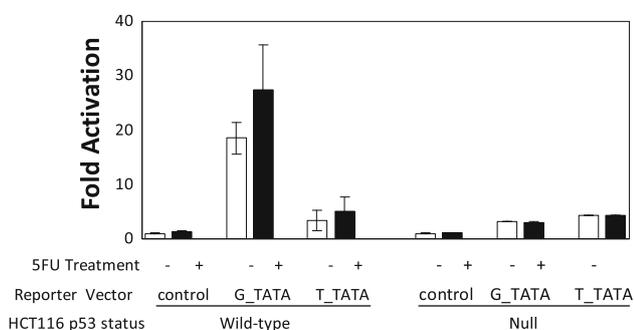
Fig. 2 Functional analysis of the binding site sequence (226 bp fragment) and its polymorphism (rs1860746) by reporter gene assay in wild-type and p53-null HCT116 cells with or without 5FU treatment. *Control* TATA-luciferase pGL4 vector; *G_TATA* TATA-luciferase pGL4 vector with a insert of the 226 bp binding site sequence of G allele; *T_TATA* TATA-luciferase pGL4 vector with a insert of the 226 bp binding site sequence of T allele

Taken together, our results show that though the binding site at the rs1860746 locus binds p53 and can be used as a p53 responsive enhancer, it does not regulate PRKAG2.

Association analysis of rs1860746 in breast cancer

Given that p53 has been implicated in cancer development (Vousden and Lu 2002; Shaw 2006), we hypothesize that the polymorphism at rs1860746 may have an impact on cancer susceptibility. To test this hypothesis, we genotyped the SNP in the sample consisting of 1,290 breast cancer patients, and 1,483 healthy controls from Sweden and 2,222 breast cancer patients and 1,256 healthy controls from Finland. Given that only the homozygous TT genotype showed the aberrant p53 binding in our in vitro functional analyses, the association was tested under a recessive model in the combined 3,512 breast cancer patients and 2,739 controls from both the Sweden and Finland. Significant association of the TT genotype with breast cancer susceptibility was found (OR = 1.34 (95%CI = 1.01, 1.77), $P = 0.043$). Each of the Swedish and Finnish samples showed a trend for association, but neither achieved statistical significance (OR = 1.45 for Swedish and 1.25 for Finnish sample) (Fig. 3), due to the rarity of the TT homozygous genotype (<5%) in population. Furthermore, we performed a subgroup analysis by stratifying the breast cancer patients according to their menopausal status, family history and ER status and found a stronger association in the premenopausal patients as compared to post menopausal individuals (OR = 1.66 vs. 1.30), the patients with family history vs. sporadic cases (OR = 1.48 vs. 1.25) or ER negative tumors as compared to ER positive tumors (OR = 1.48 vs. 1.26) (Table 2).

To further validate the association, we then genotyped rs1860746 in another three breast cancer case–control samples of European origin (ABCFS, GENICA, and KCONFAB), consisting of additional 2,615 cases and 2,458 controls. The joint analysis of all the four samples with ER status information (consisting of 4,190 cases and 5,197 controls) showed the significant association with ER negative breast cancer (OR = 1.47 (95%CI = 1.02–2.12), $P = 0.038$) (Fig. 3), while consistent association was observed across all the four samples. In contrast, no consistent association was observed across all the independent samples for the overall breast cancer risk as well as other patient subgroups (Fig. 3, Supplement Fig. 5).

Discussion

This study presents one of the few efforts where p53-related regulatory variants were investigated molecularly and genetically (Pietsch et al. 2006). In addition to the T/G
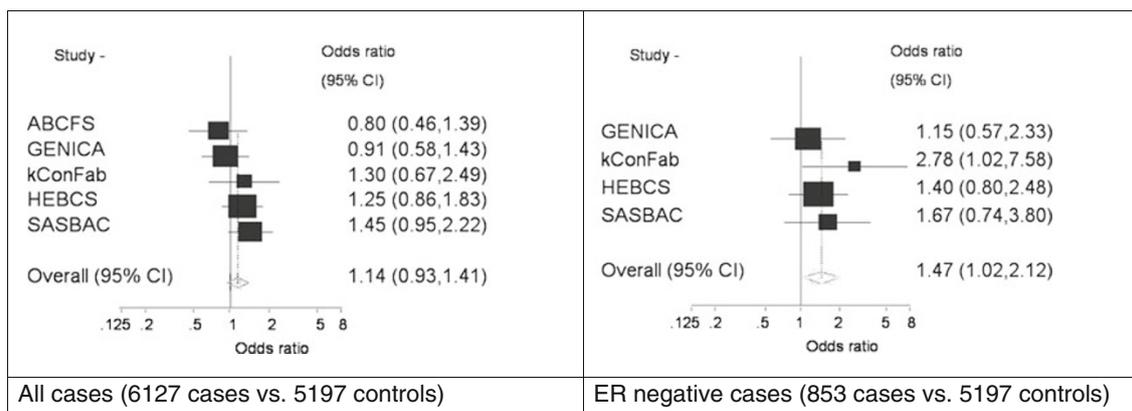
Fig. 3 Forest plots for odds ratio estimates of rs1860746 under a recessive model of inheritance by study in overall breast cancer and ER negative patient subgroup of breast cancer. The size of the box is inversely proportional of the standard error for the log odds ratio estimate

polymorphism within the intronic promoter of MDM2 (Bond et al. 2004), Mendendez et al. (2007) also identified a C/T polymorphism within the proximal promoter region of the flt-1 gene, where the minor allele of T created a half-binding site for p53 and brought the system under the control of p53 network. A more recent effort by the same group has further demonstrated that the presence of this polymorphism also created a partial responsiveness to estrogen receptor upstream of the previously identified binding half-site for p53. This results in synergistic simulation of transcription at this flt-promoter site through the combined action of p53 and ER (Menendez et al. 2007). The importance of these p53-related regulatory variants in disease development, however, has not been demonstrated.

We sought to identify potential regulatory SNPs by a "forward" genetic strategy that first assesses all DNA binding sites of p53 in a genome wide manner, mines polymorphisms within the binding sites, interrogates the functional impact of these SNPs on the primary property of p53 occupancy and transcriptional regulation, and lastly investigates their association with disease susceptibility. As a result of the initial attempt, we found that the homozygous state of the minor allele (TT) of one such binding site variant, rs1860746, showed significantly lower p53 occupancy after cellular induction with a genotoxic agent. Given the known function of p53 as an important tumor suppressor gene, and the placement of the binding site variant within another cancer related gene PRKAG2 (encoding AMPKγ) (Inoki et al. 2003; Shaw et al. 2004; Jones et al. 2005; Laderoute et al. 2006), we sought to determine the association of this SNP with cancer susceptibility in breast cancer. Our results show a modest effect of the homozygous TT state on breast cancer susceptibility which is significant only in ER negative breast cancers after examining over 5,000 cases and controls. Such modest effects are intriguing but not definitive and therefore will require larger studies to validate, especially

since the frequency of the effective homozygous state is low (<5%). This is similar to the results of SNP rs3020314 which tags a region of ESR1 intron 4 (the estrogen receptor gene) where after analysis in 55,000 cases and controls showed an OR effect of 1.05 confined to women bearing ER positive tumors (Dunning et al. 2009). The greater association of the homozygous TT genotype of rs1860746 with susceptibility for ER negative cancers is consistent with the molecular understanding of breast cancer biology since ER negative tumors have greater aneuploidy and are associated with aberrations of p53 itself (Miller et al. 2005). Carriers of germ-line p53 mutations in the families affected with Li-Fraumeni syndromes (LFS) are at risk for early-onset breast cancer (Olivier et al. 2003), and our findings further suggest that genetic disturbances in downstream transcriptional regulation by p53 may also have an effect on breast cancer risk. We also want to point out that BRCA mutation carriers have been excluded from the familial patients from the KConFab and HEBC studies. In familial patients, BRCA mutations, especially BRCA1 mutations, would be a strong confounding risk factor for ER negative breast cancer, if not excluded. Therefore, by excluding BRCA carriers, our association results are expected to be independent from the confounding effect of BRCA mutations. Further studies will be needed to investigate the risk effect of the polymorphism in BRCA mutation carriers.

Despite the definitive binding of p53 at the rs1860746 locus, the transcriptional analysis of the PRKAG2 gene, however, did not show direct regulation of this gene by p53 either in lymphoblastoid cell lines or in HCT116 colon carcinoma cell line. This discrepancy can be due to different scenarios. First, the p53 binding site at the rs1860746 locus may regulate another gene in the vicinity in a p53 dependent manner through distant regulatory control which we have recently demonstrated with the estrogen receptor (Fullwood et al. 2009). Second, as the

most conservative explanation, rs1860746 may be in linkage disequilibrium with another causal mutation within the PRKAG2 gene or another gene in the vicinity and that the differential p53 binding was a fortuitous but irrelevant observation. It should also be noted that though the effect size of the rs1860746 SNP in breast cancer is small, it may be greater in other cancer types that are more p53 driven, as suggested by our subgroup analysis in breast cancer.

Our study also raises an interesting facet about potential causal regulatory SNPs with low effect size. We could see a signal in our genetic analysis only using a recessive model which would have been missed by standard GWAS analyses where either allele or trend association are usually tested. The functional assignment of a regulatory SNP allows for the selection of the appropriate analytical approach. Moreover, it raises yet another tier of genetic polymorphisms that may contribute to disease susceptibility: one of low effect size and recessive in nature. As proof-of-principal, our study has highlighted that combining the genome-wide discovery of transcription regulatory elements (such as transcription factor binding sites) with the forward genetic analysis in both model and human systems can greatly advance our understanding on the molecular and physiological functions of regulatory genetic variation. We further posit that intersect between the new genome-wide knowledge of various regulatory sequences and the rapidly accumulated disease association data on germ-line polymorphisms will bring new insights to the role of genetic variants in regulatory variation in human populations.

References

Ahmed S, Thomas G et al (2009) Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet 41(5): 585–590

Bond GL, Hu W et al (2004) A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. Cell 119(5): 591–602

Bourque G, Leong B et al (2008) Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res 11:1752–1762

Cawley S, Bekiranov S et al (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 116(4): 499–509

Dunning AM, Healey CS et al (2009) Association of ESR1 gene tagging SNPs with breast cancer risk. Hum Mol Genet 18(6):1131–1139

Easton DF, Pooley KA et al (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447(7148):1087–1093

Fullwood MJ, Liu MH et al (2009). An oestrogen-receptor-α-bound human chromatin interactome. Nature (in press)

Hu Z, Jin G et al (2007) MDM2 promoter polymorphism SNP309 contributes to tumor susceptibility: evidence from 21 case-control studies. Cancer Epidemiol Biomarkers Prev 16(12): 2717–2723

Inoki K, Zhu T et al (2003) TSC2 mediates cellular energy response to control cell growth and survival. Cell 115(5):577–590

Jones RG, Plas DR et al (2005) AMP-activated protein kinase induces a p53-dependent metabolic checkpoint. Mol Cell 18(3):283–293

King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. Science 188(4184):107–116

Knight JC (2005) Regulatory polymorphisms underlying complex disease traits. J Mol Med 83(2):97–109

Laderoute KR, Amin K et al (2006) 5′-AMP-activated protein kinase (AMPK) is induced by low-oxygen and glucose deprivation conditions found in solid-tumor microenvironments. Mol Cell Biol 26(14):5336–5347

Menendez D, Inga A et al (2007) A single-nucleotide polymorphism in a half-binding site creates p53 and estrogen receptor control of

vascular endothelial growth factor receptor 1. Mol Cell Biol 27(7):2590–2600

Miller LD, Smeds J et al (2005) An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci USA 102(38):13550–13555

Olivier M, Goldgar DE et al (2003) Li-Fraumeni and related syndromes: correlation between tumor type, family structure, and TP53 genotype. Cancer Res 63(20):6643–6650

Pastinen T, Hudson TJ (2004) Cis-acting regulatory variation in the human genome. Science 306(5696):647–650

Pietsch EC, Humbey O et al (2006) Polymorphisms in the p53 pathway. Oncogene 25(11):1602–1611

Shaw RJ (2006) Glucose metabolism and cancer. Curr Opin Cell Biol 18(6):598–608

Shaw RJ, Kosmatka M et al (2004) The tumor suppressor LKB1 kinase directly activates AMP-activated kinase and regulates apoptosis in response to energy stress. Proc Natl Acad Sci USA 101(10):3329–3335

Sladek R, Hudson TJ (2006) Elucidating cis- and trans-regulatory variation using genetical genomics. Trends Genet 22(5):245–250

Tan J, Zhuang L et al (2005) Pharmacologic modulation of glycogen synthase kinase-3beta promotes p53-dependent apoptosis through a direct Bax-mediated mitochondrial pathway in colorectal cancer cells. Cancer Res 65(19):9012–9020

Vousden KH, Lu X (2002) Live or let die: the cell's response to p53. Nat Rev Cancer 2(8):594–604

Wei CL, Wu Q et al (2006) A global map of p53 transcription-factor binding sites in the human genome. Cell 124(1):207–219

Weinmann AS, Farnham PJ (2002) Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. Methods 26(1):37–47

Wells J, Farnham PJ (2002) Characterizing transcription factor binding sites using formaldehyde crosslinking and immunoprecipitation. Methods 26(1):48–56

Wray GA (2007) The evolutionary significance of cis-regulatory mutations. Nat Rev Genet 8(3):206–216