

Functional nonsynonymous single nucleotide polymorphisms from the TGF- β protein interaction network

Sevtap Savas, Ian W. Taylor, Jeff L. Wrana and Hilmi Ozelik

Physiol Genomics 29:109-117, 2007. First published Dec 26, 2006;

doi:10.1152/physiolgenomics.00226.2006

You might find this additional information useful...

Supplemental material for this article can be found at:

<http://physiolgenomics.physiology.org/cgi/content/full/00226.2006/DC1>

This article cites 66 articles, 27 of which you can access free at:

<http://physiolgenomics.physiology.org/cgi/content/full/29/2/109#BIBL>

Updated information and services including high-resolution figures, can be found at:

<http://physiolgenomics.physiology.org/cgi/content/full/29/2/109>

Additional material and information about *Physiological Genomics* can be found at:

<http://www.the-aps.org/publications/pg>

This information is current as of May 8, 2007 .

Functional nonsynonymous single nucleotide polymorphisms from the TGF- β protein interaction network

Sevtap Savas,^{1,2,3} Ian W. Taylor,^{4,5} Jeff L. Wrana,^{4,5} and Hilmi Ozcelik^{1,2,3}

¹Fred A. Litwin Centre for Cancer Genetics, Samuel Lunenfeld Research Institute, ²Department of Pathology and Laboratory Medicine, Mount Sinai Hospital; ³Department of Laboratory Medicine and Pathobiology, University of Toronto; ⁴Centre for System Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital; and ⁵Department of Medical Genetics and Microbiology, University of Toronto, Toronto, Ontario, Canada

Submitted 13 October 2006; accepted in final form 22 December 2006

Savas S, Taylor IW, Wrana JL, Ozcelik H. Functional nonsynonymous single nucleotide polymorphisms from the TGF- β protein interaction network. *Physiol Genomics* 29: 109–117, 2007. First published December 26, 2006; doi:10.1152/physiolgenomics.00226.2006.—Protein complexes mediated by protein-protein interactions are essential for many cellular functions. Transforming growth factor (TGF)- β signaling involves a cascade of protein-protein interactions and malfunctioning of this pathway has been implicated in human diseases. Using an *in silico* approach, we analyzed the naturally occurring human genetic variations from the proteins involved in the TGF- β signaling (10 TGF- β proteins and 242 other proteins interacting with them) to identify the ones that have potential biological consequences. All proteins were searched in the dbSNP database for the presence of nonsynonymous single nucleotide polymorphisms (nsSNPs). A total of 118 validated nsSNPs from 63 proteins were retrieved and analyzed in terms of 1) evolutionary conservation status, 2) being located in a functional protein domain or motif, and 3) altering putative protein motif or phosphorylation sites. Our results indicated the presence of 31 nsSNPs that occurred at evolutionarily conserved residues, 37 nsSNPs were located in protein domains, motifs, or repeats, and 46 nsSNPs were predicted to either create or abolish putative protein motifs or phosphorylation sites. We undertook this study to analyze the human genetic variations that can affect the protein function and the TGF- β signaling. The nsSNPs reported in here can be characterized by experimental approaches to elucidate their exact biological roles and whether they are related to human disease.

transforming growth factor- β pathway; protein-protein interactions; evolutionary conservation analysis; protein domains and motifs; phosphorylation sites

TRANSFORMING GROWTH FACTOR BETA (TGF- β) signaling is directly involved in many homeostatic and developmental processes (22), including wound healing, tumorigenesis, cell proliferation and differentiation, neovascularization, and extracellular matrix formation (33). TGF- β signaling starts with the binding of the secreted TGF- β ligands to the transmembrane serine/threonine kinase TGF- β receptors. This binding initiates a cascade of cellular events including the activation of the SMAD proteins, which transduce the TGF- β signals to nucleus for transcriptional regulation of specific genes (12, 33, 61). Deregulation or abnormalities of the TGF- β pathway have been implicated in several conditions including the atherosclerosis (20), Alzheimer's disease (6), autoimmune disease (1),

familial juvenile polyposis (34), pancreatic cancer (29), breast cancer (56), and prostate cancer (3).

Single nucleotide polymorphisms (SNPs) are the most common genetic variation in human (57). They occur, on the average, once every 300–400 base pairs (24). SNP located within the coding or regulatory regions of genes can cause qualitative and quantitative changes in gene expression, RNA splicing, protein translation, or gene function. Nonsynonymous SNPs (nsSNPs) are located in the coding regions, substitute the amino acids, and are the least frequent form of the SNPs probably because of the selective constraints on protein sequences (7, 16). Since nsSNPs change the amino acid sequences, they are also likely to change the structure and the function of the proteins. As nearly 30% of the nsSNPs are predicted to affect the protein function (9, 37, 48), they have been the topic of the studies aiming to identify the disease-susceptibility loci (8).

Pathway-based approaches are beneficial because they help the molecular biologists and epidemiologists to investigate multiple genes by focusing on specific physiological processes and relate them to human disease (62). These kinds of approaches also provide a logical basis for examining the gene-gene and SNP-SNP interactions (epistasis) in human health and disease predisposition (11). In this regard, we and other groups have previously analyzed and evaluated the functionalities of the nsSNPs from a variety of cellular pathways (10, 26, 53, 54). In this study, we aimed to systematically investigate the nsSNPs from the TGF- β signaling protein interaction network. A total of 10 TGF- β superfamily members have been chosen as the TGF- β core proteins, and a TGF- β protein interaction network was constructed. These members of the TGF- β superfamily were chosen based on their occurrence as bait in a high-throughput screen (2), where literature curated interactions were validated by the LUMIER technique and found to be >70% true-positive using conservative cutoffs. The 10 core proteins included the two bone morphogenetic protein receptors (ALK2/ACVR1 and ALK6/BMPRI1B), one transmembrane receptor for TGF- β ligand (TGFBR1), five SMAD members that transduce the TGF- β signals in cells (SMAD1, SMAD2, SMAD3, SMAD4, and SMAD7), and two E3 ligases that facilitate SMAD receptor destruction by ubiquitin-mediated protein degradation (SMURF1 and SMURF2) (33). Our results point to the notion that nsSNPs that can cause functional alterations in the TGF- β interaction network, which can be utilized for further studies to functionally characterize and investigate their relationship to human disease and health.

Article published online before print. See web site for date of publication (<http://physiolgenomics.physiology.org>).

Address for reprint requests and other correspondence: H. Ozcelik, Mount Sinai Hospital Samuel Lunenfeld Research Inst., 60 Murray St. L6–303, Box 29, Toronto ON Canada M5T 3L9 (e-mail: ozcelik@mshri.on.ca).

METHODS

TGF- β pathway interacting proteins. Literature-curated protein-protein interactions (PPIs) were taken from the MedLine Database at the National Center for Biotechnology Information between February and April 2004. Interactions were detected for well-characterized components of the TGF- β superfamily of signal transduction pathway (TGFB1, ALK2/ACVR1, ALK6/BMP1B, SMAD1, SMAD2, SMAD3, SMAD4, SMAD7, SMURF1, and SMURF2). To search for interactions in the MedLine database each molecule and all known synonyms including alternative spellings from human, mouse, *Drosophila melanogaster*, and *Caenorhabditis elegans* were entered in the search field separated by the Boolean operator "OR". The resulting list of abstracts was manually curated by a trained, graduate-level scientist (I. W. Taylor) for interactions of proteins with the component of the TGF- β pathway. In the case of ambiguous language in the abstracts the scientist examined the primary manuscript for direct evidence in the data figures relating to the abstract. Interacting proteins were then mapped to a unique identifier with the UniGene database (65) (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>). When interactions were detected in model organisms, the human UniGene ID was used to describe the homologous human gene. The protein interactions are depicted in the Supplementary Figure (http://www.ozceliklab.com/savas/TGFbeta_PPIs). (The online version of this article contains supplemental material.)

nsSNP. The validated nsSNPs for the TGF- β interaction network proteins were retrieved from the dbSNP build 124 (58) (<http://www.ncbi.nlm.nih.gov/SNP/>). Only the nsSNPs that are found in at least two chromosomes in a sample panel of at least 20 chromosomes were included into this study (validated nsSNPs). Throughout this paper, the SNPs that were reported with lower or more than 5% of the samples are annotated as rare and common SNPs, respectively. In some cases, SNPs were detected with lower and higher than 5% minor allele frequencies in different submissions: for simplicity, we classified these SNPs as common SNPs together with the SNPs with minor allele frequencies of $\geq 5\%$.

Prediction of functional consequences of nsSNPs. A precomputed PolyPhen resource (courtesy of Dr. Shamil Sunyaev) was utilized to retrieve PolyPhen predictions for the nsSNPs (48) (<http://www.bork.embl-heidelberg.de/PolyPhen/>). All predictions were based on protein alignments; the predictions based on fewer than five proteins in the alignment were considered unreliable and thus are annotated as "noninformative" in this study.

Protein domain and motif analysis. The information related to the positions of the protein functional domains and motifs was retrieved from the Swiss-Prot database feature table (5) (<http://www.expasy.org/sprot/sprot-search.html>) and the Interpro database (36) (<http://www.ebi.ac.uk/Interpro>). In those cases when the Interpro entry was not available, we have run the InterproScan program (47) (<http://www.ebi.ac.uk/InterProScan/>) to predict the protein domains and motifs. The Human Protein Reference Database (HPRD, Ref. 43; <http://www.hprd.org>) was also utilized to see whether the nsSNPs were located in protein domains, motifs, or at experimentally verified posttranslational modification sites of the proteins. NetPhos (4) was utilized to predict putative phosphorylation sites (<http://www.cbs.dtu.dk/services/NetPhos/>). ScanSite (40) was utilized to predict the putative protein motifs (a total of 63 motifs including phosphorylation sites, modular signaling, and interaction domains) (http://scansite.mit.edu/motifscan_seq.phtml). ScanSite was run at high-stringency conditions.

RESULTS

A total of 242 proteins interacting with the 10 TGF- β core proteins (TGFB1, ACVR1, BMP1B, SMAD1, SMAD2, SMAD3, SMAD4, SMAD7, SMURF1 and SMURF2) were

identified. The number of proteins interacting with each of these proteins was as follows: TGFB1 ($n = 55$), ACVR1 ($n = 1$), BMP1B ($n = 10$), SMAD1 ($n = 40$), SMAD2 ($n = 41$), SMAD3 ($n = 71$), SMAD4 ($n = 24$), SMAD7 ($n = 7$), SMURF1 ($n = 1$), and SMURF2 ($n = 2$) (Supplementary Figure; http://www.ozceliklab.com/savas/TGFbeta_PPIs). All proteins were searched in the dbSNP database for the presence of validated nsSNPs. As a result, a total of 118 validated nsSNPs from 63 proteins were retrieved at the time of data collection (Table 1). When we categorized the nsSNPs based on their minor allele frequencies (MAFs; see METHODS), the majority of the nsSNPs ($n = 94$, 80.0%) were reported as common variations in the dbSNP database (Table 1).

Informative PolyPhen predictions were available for 70/118 (59.0%) nsSNPs. Among these, 31 (44.3%, 31/70) nsSNPs were predicted to affect the protein function (either possibly or probably damaging predictions; Tables 1 and 2). No functional nsSNP was identified in ACVR1, SMAD7, SMURF1, or SMURF2 nor in the proteins interacting with them ($n = 11$). However, at least one functional nsSNP was identified in proteins interacting with TGFB1 ($n = 5$), BMP1B ($n = 1$), SMAD1 ($n = 2$), SMAD2 ($n = 8$), SMAD3 ($n = 7$), SMAD4 ($n = 1$) (Table 2). Similar to the data from the whole set of nsSNPs (see above), 27/31 (87.1%) of the functional nsSNPs were common nsSNPs with MAFs $\geq 5\%$. Information curated in the Swiss-Prot, Interpro, and the HPRD databases indicated that 37/118 (31.0%) nsSNPs were located in a protein domain, motif, or a repeat sequence (Table 2). Among this group of nsSNPs, PolyPhen predicted 16 to affect protein function.

We also utilized the ScanSite (40) and NetPhos (4) programs to predict impact of the nsSNPs on the putative protein motifs and phosphorylation sites along the TGF- β protein interaction network proteins. As a result, a total of 46 (39%, 46/118) nsSNPs from 34 genes have been predicted to modify putative protein motifs and phosphorylation sites (Table 3). ScanSite predicted that in the case of seven nsSNPs (ALDH3A1-G309E, BRCA2-N289H, FBXO34-G704V, MYC-N11S, NCOA1-P1272S, REV3L-R1892H, and TCF7-G256R), a short protein motif was either created or abolished. NetPhos predicted that 29 nsSNPs altered (either created or abolished) putative kinase recognition motifs and phosphorylation sites. The BRCA2-

Table 1. Summary of the data

	#
Genes screened for nsSNPs	252
Genes with validated nsSNPs	63
Validated nsSNPs	118
Common nsSNPs	94
Rare nsSNPs	24
nsSNPs with informative PolyPhen prediction	70
nsSNPs functional by PolyPhen	31
nsSNPs benign by PolyPhen	39
nsSNPs located in domain/motif/repeat	37
nsSNPs functional by PolyPhen	16
nsSNPs benign by PolyPhen	16
nsSNPs with no informative PolyPhen prediction	5
nsSNPs altering a protein motif/phosphorylation site	46
nsSNPs functional by PolyPhen	15
nsSNPs benign by PolyPhen	14
nsSNPs with no informative PolyPhen prediction	17

nsSNP, nonsynonymous single nucleotide polymorphism.

Table 2. nsSNPs that are either predicted to be functional by PolyPhen or located in a functional protein domain, motif, and/or a repeat

Interacting Core Protein	Gene	SNP ID	Accession #	Variation	Major Allele	mAFs	PolyPhen Prediction	Protein Domain/Motif/Repeat
SMAD2	AGER	rs2070600	NP_001127.1	G82S	G	2	S residue possibly damaging	*Extracellular domain (potential); *Ig-like V-type domain; †(IPR003599) Immunoglobulin subtype domain; †(IPR007110) Immunoglobulin-like domain; †(IPR013151) Immunoglobulin domain; ‡Ig_LIKE domain
TGFBR1	BIRC4	rs5956583	NP_001158.2	Q423P	Q and P	1 and 2	P residue possibly damaging	
BMPR1B	BMP2	rs235768	NP_001191.1	R190S	S and R	2	S residue probably damaging	*Propeptide; †(IPR001111) Transforming growth factor beta NH ₂ -terminal
BMPR1B	BMP2	rs2273073	NP_001191.1	S37A	S	2	A residue benign	*Propeptide; †(IPR001111) Transforming growth factor beta NH ₂ -terminal
BMPR1B	BMP2	rs2273074	NP_001191.1	A106T	A	1	T residue possibly damaging	*Propeptide; †(IPR001111) Transforming growth factor beta NH ₂ -terminal
BMPR1B	BMP4	rs17563	NP_001193.1	V152A	V and A	2	V residue benign	*Propeptide; †(IPR001111) Transforming growth factor beta NH ₂ -terminal
BMPR1B	BMPR2	rs2228545	NP_001195.2	S775N	S	1	N residue NI	*Cytoplasmic domain (potential)
SMAD3	BRCA2	rs766173	NP_000050.1	N289H	N	1 and 2	H residue possibly damaging	
SMAD3	BRCA2	rs1046984	NP_000050.1	S599F	S	2	S residue possibly damaging	
SMAD3	BRCA2	rs1799954	NP_000050.1	R2034C	R	2	C residue possibly damaging	
SMAD3	BRCA2	rs4987047	NP_000050.1	I2944F	I	1 and 2	F residue possibly damaging	
SMAD4	BTRC	rs2270439	NP_003930.1	P556H	P	1	H residue possibly damaging	
TGFBR1	CD44	rs1467558	NP_000601.2	I479T	T	2	I residue benign	*Extracellular domain (potential); stem region
TGFBR1	CD44	rs9666607	NP_000601.2	K417R	R	2	K residue benign	*Extracellular domain (potential); stem region
SMAD2	CNTN2	rs2275697	NP_005067.1	A145T	T and A	2	T residue benign	*Ig-like C2-type 2 domain; †(IPR003599) Immunoglobulin subtype domain; ‡Ig_LIKE domain
SMAD2	CNTN2	rs2305276	NP_005067.1	R657W	R	2	W residue probably damaging	*Fibronectin type-III I domain; †(IPR003961) Fibronectin, type III domain; ‡FN3 domain
SMAD2	CNTN2	rs17416074	NP_005067.1	V1024I	V	2	I residue NI	*Propeptide (potential)
TGFBR1	DUSP13	rs3088142	NP_057448.2	C156Y	Y and C	2	C residue possibly damaging	*Tyrosine-protein phosphatase domain; †(IPR000340) Dual specificity protein phosphatase domain
TGFBR1	DUSP13	rs16931996	NP_057448.2	R190G	G	2	G residue probably damaging	†(IPR000340) Dual specificity protein phosphatase domain
TGFBR1	DUSP13	rs16932004	NP_057448.2	R62Q	R	1	Q residue benign	†(IPR000340) Dual specificity protein phosphatase domain
TGFBR1	ENG	rs1800956	NP_000109.1	D366H	H and D	2	H residue possibly damaging	†(IPR001507) Endoglin/CD105 antigen domain; Zona pellucida domain
SMAD3	ERBB2IP	rs3213837	NP_061165.1	S274L	S	1 and 2	L residue possibly damaging	†(IPR003591) Leucine-rich repeat; †(IPR003591) Leucine-rich repeat, typical subtype
SMAD3	ERBB2IP	rs3805466	NP_061165.1	S1112L	S	1 and 2	L residue possibly damaging	
TGFBR1	FBXO34	rs10138395	NP_060413.2	G704V	G	2	V residue probably damaging	
SMAD2	FLNA	rs17091204	NP_001447.1	S1012L	S	2	L residue possibly damaging	†(IPR001298) Filamin/ABP280 repeat

Table 2. —“continued”

Interacting Core Protein	Gene	SNP ID	Accession #	Variation	Major Allele	mAFs	PolyPhen Prediction	Protein Domain/Motif/Repeat
BMPR1B	GDF5	rs224331	NP_000548.1	S276A	A and S	2	A residue NI	*Propeptide (potential); †(IPR001111) Transforming growth factor beta NH ₂ -terminal domain
TGFBR1	GDF8	rs1805085	NP_005250.1	A55T	A	2	T residue benign	*Propeptide (potential); †(IPR001111) Transforming growth factor beta NH ₂ -terminal domain
TGFBR1	GDF8	rs1805086	NP_005250.1	K153R	K	1 and 2	R residue benign	*Propeptide (potential); †(IPR001111) Transforming growth factor beta NH ₂ -terminal domain
SMAD3	HIF1A	rs11549465	NP_001521.1	P582S	P	2	S residue possibly damaging	*ID region
SMAD2	MYC	rs4645959	NP_002458.1	N11S	N	2	S residue possibly damaging	†(IPR012682) Transcription regulator Myc NH ₂ -terminal domain
SMAD2	MYC	rs4645961	NP_002458.1	V170I	V	1	I residue benign	†(IPR012682) Transcription regulator Myc NH ₂ -terminal domain
SMAD3	NCOAI	rs1804645	NP_003734.3	P1272S	P	1	S residue probably damaging	
SMAD1	NEUROD1	rs1801262	NP_002491.1	T45A	A and T	2	T residue benign	‡CC motif
SMAD2	NFKB1	rs4648072	NP_003989.2	M507V	M	1	V residue possibly damaging	
SMAD2	NFKB1	rs4648099	NP_003989.2	H712Q	H	2	Q residue possibly damaging	*ANK 5 repeat; †(IPR002110) Ankyrin repeat; ‡ANK domain
SMAD3	NGFB	rs6330	NP_002497.1	A35V	V and A	2	A residue benign	*Propeptide
SMAD3	NGFB	rs11466110	NP_002497.1	V72M	V	1	M residue benign	*Propeptide
SMAD1	NOTCH1	rs3125001	NP_060087.2	A863T	T	2	T residue NI	*Extracellular domain (potential); *EGF-like 22 domain; †(IPR000742) EGF-like, type 3 domain; ‡EGFCA domain
SMAD1	NOTCHI	rs11574885	NP_060087.2	Q300R	Q	2	R residue possibly damaging	*Extracellular domain (potential); *EGF-like 8 domain; *calcium-binding domain (potential); †IPR000742 EGF-like, type 3 domain; ‡EGFCA domain
SMAD3	NR3C1	rs6192	NP_000167.1	F65V	F	1	V residue possibly damaging	
SMAD2	NUP214	rs103612	NP_005076.3	P574S	S	2	S residue benign	*11 × 5 AA approximate repeats region
SMAD2	PAXIP1	rs3501	NP_031375.2	M979V	M	2	V residue benign	†(IPR001357) BRCT domain
SMAD3	PEX6	rs2274516	NP_000278.2	V882I	V	1	I residue benign	†(IPR003959) AAA ATPase, central region domain; ‡AAA domain
SMAD3	PPARG	rs1801282	NP_056953.2	P12A	P	1 and 2	A residue possibly damaging	
SMAD1	SMAD4	rs2229083	NP_005350.1	W101G	W	2	G residue probably damaging	*MH1 domain; †(IPR003619) Dwarfing protein, A domain
TGFBR1	SNX1	rs1802376	NP_003090.2	D466N	n/a	2	N residue possibly damaging	
TGFBR1	SNX2	rs1044463	NP_003091.2	A384S	A	2	S residue benign	‡CC motif
SMAD4	TGIF	rs2229333	NP_733796.2	P292L	P	2	L residue probably damaging	
SMAD2	TP53	rs1042522	NP_000537.2	P72R	R and P	2	P residue possibly damaging	*Interacting with PRMT1 protein; *Interacting with WWOX protein

Table 2. —“continued”

Interacting Core Protein	Gene	SNP ID	Accession #	Variation	Major Allele	mAFs	PolyPhen Prediction	Protein Domain/Motif/Repeat
SMAD2	TUBB2A	rs1054331	NP_001060.1	S201C	C	2	C residue probably damaging	†(IPR003008) Tubulin/FtsZ. GTPase domain; ‡tubulin domain
SMAD3	VDR	rs2228570	NP_000367.1	M1T	M	2	T residue probably damaging	
SMAD1	VENTX	rs2270192	NP_055283.1	E101K	E	1	K residue NI	†(IPR001356) Homeobox domain

The protein domain, motif, or repeat information is as in *Swiss-Prot feature table, †Interpro database (in parentheses are the Interpro protein domain and motif identifiers), ‡Human Protein Reference Database (HPRD). mAF, minor allele frequency; NI, noninformative PolyPhen prediction; CC, coiled coil.

S599F, DAB2-T586I, HIF1A-P582S, NEDD9-T577M, REV3L-R1892H, and SNX1-D466N are predicted to alter two motifs and phosphorylation sites (Table 3). PolyPhen predicted 15 of the nsSNPs, which alter protein motif or phosphorylation sites based on NetPhos and ScanSite results, as also affecting the protein function (Table 2).

DISCUSSION

PPIs that form functional protein complexes are essential for a variety of biological processes such as signal transduction, DNA repair, and cell cycle (19, 32, 39, 42, 64). These interactions are mainly facilitated by specific interaction domains and motifs (42). One protein can have multiple interaction domain/binding motifs and thus multiple interaction partners. Not surprisingly, mutations affecting PPIs have been linked to abnormal cellular functions and human diseases (35, 38).

TGF- β is a signaling pathway where PPIs are quite abundant. We had previously suggested that the functional nsSNPs from proteins interacting with each other could affect the function of protein complexes by means of affecting either the function of individual proteins or the dynamics of the PPIs (54). Similarly, in this study, our goal was to analyze the possible consequences of genetic variations on functions of the TGF- β interaction network proteins via an *in silico* approach. Based on the published information, we have retrieved the proteins interacting with a group of TGF- β proteins to construct a TGF- β protein interaction network. This network consisted of a total of 252 proteins, and the highest numbers of protein interactions were with the SMAD3 ($n = 71$), TGFBR1 ($n = 55$), SMAD2 ($n = 41$), SMAD1 ($n = 40$), and SMAD4 ($n = 24$) proteins.

Our results demonstrated that a total of 63 nsSNPs (31 damaging the protein function and/or 32 nsSNPs creating/abolishing short protein motifs and phosphorylation sites) could have biological consequences (Tables 2, 3). A literature search revealed that some of the nsSNPs were already implicated in human diseases. AGER-G82S has been previously implicated in the susceptibility to diabetes-associated microvascular dermatoses (21) and rheumatoid arthritis (18), and the AGER-82S allele was found to increase the inflammatory response compared with AGER-G82 allele (18). HIF1A-P582S has an elevated transcription activity (60) and was associated with the maximal oxygen consumption before and after aerobic exercise training in older humans (46) and Type 2 diabetes (67). PPARG-P12A was associated with Type 2 diabetes in Oji-Cree women (17) and reduced risks of renal cell carcinoma (59) and myocardial infarction (50). TP53-72R induces apop-

tosis better than the proline allele (13) although the proline allele is more efficient in G1 arrest (44). This nsSNP is associated with a variety of cancers, including breast cancer in Jewish women (41) and colorectal cancer (23). NR3C1-N363S (Table 3) was associated with the coronary artery disease (31), obesity (30), and susceptibility to overweight in patients with Type 2 diabetes mellitus (52). Association of AGER, HIF1A, PPARG, and NR3C1 genes with complications related to diabetes is not surprising considering the fact that the biological roles of these genes are consistent with the observed association. For example, AGER, an advanced glycosylation end product-specific receptor, protects tissues from glucose-mediated damage (63); HIF1A induces expression of genes involved in glucose metabolism under hypoxia (28); PPARG modulates a variety of process including insulin sensitivity (27); and NR3C1, the glucocorticoid receptor, plays a role in expression of glucocorticoid responsive genes (51). On the other hand, an nsSNPs in BMP4, BMP4-V152A, was reported to be associated with bone density in postmenopausal women (49), although our analysis suggests that it is functionally benign (Table 2). This nsSNP may represent either a false-negative prediction (see below) or may be in linkage disequilibrium with another genetic factor responsible for the phenotype. Nevertheless, our results suggest that the nsSNPs in Tables 2 and 3 are good candidates for further genetics and molecular analyses to reveal whether they have functional consequences and their effects are linked to human disease.

It has been shown that proteins that interact with multiple partners are more likely to be essential (25), and the biological impact of functional nsSNPs from such proteins is likely to be large and diverse. In our data set, SMAD4 has 24 other proteins interacting with it, and it has one nsSNP that occurred at an evolutionarily conserved residue (SMAD4-W101G; Table 2). Therefore, it is feasible to speculate that this nsSNP could affect the functions of a variety of protein complexes involving SMAD4. In this regard among all, SMAD4-W101G nsSNP can be prioritized for further studies.

Protein domains, motifs, and posttranslational sites such as phosphorylation sites are important for the function, structure, and stability of proteins. Therefore, amino acid substitutions in these regions can change the intrinsic properties of proteins. Previously, the impact of the nsSNPs on short binding motifs and phosphorylation sites has been proposed and analyzed for proteins from other cellular pathways (10, 55). In the present study, we found that 37 nsSNPs were located in a protein domain. These nsSNPs, especially the ones that occur at evolutionarily conserved residues ($n = 16$, Table 2), can alter

Table 3. *nsSNPs altering short putative protein motifs and phosphorylation sites*

Gene ^a	SNP ID ^b	Accession #	mAFs ^c	Variations ^d	Major Variant ^e	ScanSite Results ^f	NetPhos Results ^g
ALDH3A1	rs3744692	NP_000682.3	1	G309E	G	E residue creates casein kinase I motif at T313 GGTEDAATRYIAPTI	
BIRC4	rs5956583	NP_001158.2	1 and 2	Q423P	Q and P		P residue removes S427 in QDESSQTSLS
BMP2	rs235768	NP_001191.1	2	R190S	S and R		S residue creates S190 in FPVTSLLDTS
BMP2	rs2273073	NP_001191.1	2	S37A	S		A residue removes S37 in FAAASSGRP
BRCA2	rs766173	NP_000050.1	1 and 2	N289H	N	H residue removes Lck SH2 motif at Y296 NVLEDEVYETVVDTS	
BRCA2	rs1046984	NP_000050.1	2	S599F	S		S residue creates S599 in HDETSYKGGK and creates T598 in IHDETSYKGG
CNTN2	rs2275697	NP_005067.1	2	A145T	T and A		T residue creates T145 in DPVKTHEGW
DAB2	rs700241	NP_001334.1	1	T586I	T		I residue removes S588 in WSTTSPLGN and removes T586 in NAWSITSPL
DUSP13	rs16932004	NP_057448.2	1	R62Q	R		Q residue removes S65 in ARDKSKLIQ
ELF1	rs1056820	NP_758961.1	2	T343S	S and T		S residue removes T343 in GGATTVLKLP
ENG	rs1800956	NP_000109.1	2	D366H	H and D		H residue creates T369 in DHAMTLVLK
ERBB2IP	rs3213837	NP_061165.1	1 and 2	S274L	S		L residue removes S274 in ETIGSLKNI
ERBB2IP	rs3805466	NP_061165.1	1 and 2	S1112L	S		L residue removes S1112 in REFHSAGRT
FBXO34	rs3742569	NP_060413.2	2	L533P	P and L	P residue creates intersectin SH3A motif at P533 GSAEPFVPPASSVES	
FBXO34	rs10138395	NP_060413.2	2	G704V	G		V residue removes T705 in KAKGTEAEE
GDF5	rs224331	NP_000548.1	2	S276A	A and S		A residue removes S276 in RQPAQLLDV
GDF8	rs1805085	NP_005250.1	2	A55T	A		T residue creates T55 in SRIETIKIQ
HIF1A	rs11549465	NP_001521.1	2	P582S	P		S residue creates S582 in DQLSSLESS and creates S586 in SLESSSASP
HNF4A	rs1800961	NP_000448.3	1 and 2	T139I	T		I residue removes T139 in DRISTRSS
MAP3K1	rs832582	XP_042066.8	2	V1040I	I		I residue creates T1039 in SPECTIHLE
MYC	rs4645959	NP_002458.1	2	N11S	N	S residue creates PDK1 binding motif at S11 NVSFTNRSYDLDYDS	
MYC	rs4645961	NP_002458.1	1	V170I	V		I residue creates S175 in CSTSSLYLQ
NCOA1	rs1804645	NP_003734.3	1	P1272S	P	S residue removes GSK3 kinase motif at S1275 LQQTTPASGYQSPDM	
NEDD9	rs3734401	NP_006394.1	1	T577M	T		M residue removes Y579 in NSTEYPHGG and creates S576 in SIMNSMEYYP
NOTCH1	rs3125001	NP_060087.2	2	A863T	T		T residue removes T866 in QAGQICEVD
NR3C1	rs6192	NP_000167.1	1	F65V	I		V residue creates S69 in VPKGSVSNA
NR3C1	rs6195	NP_000167.1	1	N363S	N		S residue creates S361 in IPVGSESWN

Table 3. —“continued”

Gene ^a	SNP ID ^b	Accession #	mAFs ^c	Variations ^d	Major Variant ^e	ScanSite Results ^f	NetPhos Results ^g
NUP153	rs6906499	NP_005115.2	2	N402K	K		K residue creates S405 in DKKC <u>S</u> TGYE
NUP153	rs12195921	NP_005115.2	1 and 2	I248V	I		V residue creates S252 in VLKTS <u>Q</u> LGD
PIAS3	rs17354559	NP_006090.1	1	S381C	S		C residue removes S381 in EILSS <u>C</u> SDC
PPP1R15A	rs500079	NP_055145.2	2	T597A	A and T		T residue creates T597 in ARRIT <u>Q</u> AQE
PPP1R15A	rs557806	NP_055145.2	2	R251P	P		P residue removes S255 in RSSG <u>S</u> DPRS
PPP1R15A	rs564196	NP_055145.2	2	R31H	R		H residue removes S34 in SRAW <u>S</u> RLRG
REV3L	rs3218599	NP_002903.1	1 and 2	D1734H	D		H residue creates S1735 in QSLH <u>S</u> ANTS
REV3L	rs3218600	NP_002903.1	1	S1142L	S		L residue removes S1142 in EKGTS <u>R</u> KHI
REV3L	rs3218606	NP_002903.1	1	R1892H	R	H residue removes Intersectin SH3A motif at P1890 QNPRPG <u>S</u> PLRSGQGV and removes p85 SH3 mode2 motif at P1890 QNPRPG <u>S</u> PLRSGQGV	
SCAP2	rs17154402	NP_003921.2	2	S253T	T		T residue creates S252 in NPLT <u>S</u> TQPI
SNX1	rs1802376	NP_003090.2	2	D466N	n/a		N residue creates Y463 in RVTQ <u>Y</u> ERNF and removes S471 in FERIS <u>T</u> VVR
SP1	rs3741665	NP_612482.2	1	T737A	T		A residue removes S735 in GSEGS <u>T</u> TAT
STK111P	rs2305053	NP_443134.1	1	R780C	R		C residue removes S784 in RDHGS <u>W</u> SLS
TCF7	rs30489	NP_963964.1	1 and 2	G256R	G	R residue creates calmodulin-dependent Kinase 2 motif at S259 KDGNRQ <u>E</u> LSMSSSS	
TGIF	rs2229333	NP_733796.2	2	P292L	P		L residue removes S291 in PKP <u>S</u> SPGSV
TUBB2A	rs1054331	NP_001060.1	2	S201C	C		C residue removes S201 in DETY <u>S</u> IDNE
VENTX	rs2270192	NP_055283.1	1	E101K	E		K residue creates T99 in RTAF <u>T</u> MKQV
VEPH1	rs1378796	NP_078897.1	2	V263G	V		G residue creates Y264 in EIAG <u>Y</u> EPVA
VEPH1	rs11918974	NP_078897.1	2	S522P	S		P residue removes S522 in SENL <u>S</u> ETVK

^a Gene symbols are as approved by the HUGO Gene Nomenclature Committee (45) <http://www.gene.ucl.ac.uk/nomenclature>. ^bSNP IDs correspond to the dbSNP database SNP identifiers. ^cnsSNPs with mAFs less than and equal/higher than 5% are indicated by 1 and 2, respectively. ^dThe nature and the position of the amino acids specified by the nsSNPs is as reported in the dbSNP database. ^eThe major variants are based on the information under the population diversity section in the dbSNP. ^fScanSite results showing the putative protein motif sequence and the position of the amino acids. ^gThe nine-residue long putative kinase recognition motifs are shown. The putative serine (S), threonine (T), and tyrosine (Y) residues predicted to be phosphorylated by NetPhos are underlined.

the structure/features of protein domains and thus the function of the protein. Literature search and the information of the Swiss-Prot feature table or the HPRD did not reveal any of the nsSNPs in Table 3 to be an experimentally verified phosphorylation sites. However, the sensitivity of the NetPhos predictions is 69–96% and with a false positive rate of 0–26% for tyrosine, 0–11% for serine, and 0–14% for threonine (4). In case of the ScanSite, the accuracy of the program was reported as ~70% (66). Therefore, although some of these predictions are likely to be false positives, a considerable portion of the predictions is supposed to be correct. Thus, these nsSNPs deserve further analysis.

Interestingly, 80% of the all nsSNPs analyzed and 44% of the nsSNPs with a reliable PolyPhen prediction that were

predicted to affect the protein function were presented as common genetic variations (minor allele frequencies $\geq 5\%$). It has been hypothesized that common variations in the contemporary human population contribute to disease susceptibility (8). Thus, it is tempting to say that the common and functional nsSNPs reported in this study can also be utilized for disease association studies to investigate their potential roles in disease proposition.

In conclusion, here we report an analysis of proteins acting with the 10 TGF- β core proteins through protein-protein interactions, and present nsSNPs that might be important for the protein function based on the evolutionary conservation and protein domain and motif analyses. Those nsSNPs that are found to be functional based on evolutionary conservation

analysis and/or altering putative protein binding motifs and phosphorylation sites are good candidates for further functional studies. Given that these nsSNPs are located in candidate proteins, they may also be utilized in disease association studies to test their potential contribution to the altered disease risk. Furthermore, it is possible that in such complexes, some of the nsSNPs could have coevolved together to compensate for the functional impact on each other (10, 11, 14, 15, 54). Thus, further analysis of these nsSNPs can also help with the elucidation of the epistatic relationships among the nsSNPs either on the same protein or individual proteins of protein complexes.

ACKNOWLEDGMENTS

Authors thank Dr. Hamdi Jarjanazi for the automatic retrieval of the SNP information from the dbSNP database and Dr. Shamil Sunyayev for the precomputed PolyPhen resource.

GRANTS

This work was supported by Canadian Breast Cancer Foundation (H. Ozcelik) and by Genome Canada through Ontario Genomics Institute (J. L. Wrana). S. Savas is supported, in part, by a "CIHR Strategic Training Program Grant - The Samuel Lunenfeld Research Institute Training Program: Applying Genomics to Human Health" fellowship.

REFERENCES

- Aoki CA, Borchers AT, Li M, Flavell RA, Bowlus CL, Ansari AA, Gershwin ME. Transforming growth factor beta (TGF-beta) and autoimmunity. *Autoimmun Rev* 4: 450–459, 2005.
- Barrios-Rodiles M, Brown KR, Ozdamar B, Bose R, Liu Z, Donovan RS, Shinjo F, Liu Y, Dembowy J, Taylor IW, Luga V, Przulj N, Robinson M, Suzuki H, Hayashizaki Y, Jurisica I, Wrana JL. High-throughput mapping of a dynamic signaling network in mammalian cells. *Science* 307: 1621–1625, 2005.
- Bello-DeOcampo D, Tindall DJ. TGF-beta/Smad signaling in prostate cancer. *Curr Drug Targets* 4: 197–207, 2003.
- Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol* 294: 1351–1362, 1999.
- Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 31: 365–370, 2003.
- Cacquevel M, Lebourrier N, Cheenne S, Vivien D. Cytokines in neuroinflammation and Alzheimer's disease. *Curr Drug Targets* 5: 529–534, 2004.
- Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Shaw N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22: 231–238, 1999.
- Chakravarti A. Population genetics—making sense out of sequence. *Nat Genet* 21, Suppl 1: 56–60, 1999.
- Chasman D, Adams RM. Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. *J Mol Biol* 307: 683–706, 2001.
- Cho S, Savas S, Ozcelik H. Genetic variation and the mitogen-activated protein kinase (MAPK) signaling pathway. *OMICS* 10: 66–81, 2006.
- Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect in humans. *Hum Mol Genet* 11: 2463–2468, 2002.
- Dibrov A, Kashour T, Amara FM. The role of transforming growth factor beta signaling in messenger RNA stability. *Growth Factors* 24: 1–11, 2006.
- Dumont P, Leu JI, Della Pietra AC 3rd, George DL, Murphy M. The codon 72 polymorphic variants of p53 have markedly different apoptotic potential. *Nat Genet* 33: 357–365, 2003.
- Fay JC, Wu CI. Sequence divergence, functional constraint, and selection in protein evolution. *Annu Rev Genomics Hum Genet* 4: 213–235, 2003.
- Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. Evolutionary rate in the protein interaction network. *Science* 296: 750–752, 2002.
- Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat Genet* 22: 239–247, 1999.
- Hegele RA, Cao H, Harris SB, Zinman B, Hanley AJ, Anderson CM. Peroxisome proliferator-activated receptor-gamma2 P12A and type 2 diabetes in Canadian Oji-Cree. *J Clin Endocrinol Metab* 85: 2014–2019, 2000.
- Hofmann MA, Drury S, Hudson BI, Gleason MR, Qu W, Lu Y, Lalla E, Chitnis S, Monteiro J, Stickland MH, Bucciarelli LG, Moser B, Moxley G, Itescu S, Grant PJ, Gregersen PK, Stern DM, Schmidt AM. RAGE and arthritis: the G82S polymorphism amplifies the inflammatory response. *Genes Immun* 3: 123–135, 2002.
- Houtman JC, Barda-Saad M, Samelson LE. Examining multiprotein signaling complexes from all angles. *FEBS J* 272: 5426–5435, 2005.
- Ishisaki A, Matsuno H. Novel ideas of gene therapy for atherosclerosis: modulation of cellular signal transduction of TGF-beta family. *Curr Pharm Des* 12: 877–886, 2006.
- Kankova K, Zajejsky J, Marova I, Muzik J, Kuhrova V, Blazkova M, Znojil V, Beranek M, Vacha J. Polymorphisms in the RAGE gene influence susceptibility to diabetes-associated microvascular dermatoses in NIDDM. *J Diabetes Complications* 15: 185–192, 2001.
- Kluppel M, Hoodless PA, Wrana JL, Attisano L. Mechanism and function of signaling by the TGF- β superfamily. In: *Protein Kinase Functions: Frontiers in Molecular Biology*, edited by Woodgett J. Oxford, UK: Oxford University Press, 2000.
- Koushik A, Tranah GJ, Ma J, Stampfer MJ, Sesso HD, Fuchs CS, Giovannucci EL, Hunter DJ. p53 Arg72Pro polymorphism and risk of colorectal adenoma and cancer. *Int J Cancer* 119: 1863–1868, 2006.
- Kruglyak L, Nickerson DA. Variation is the spice of life. *Nat Genet* 27: 234–236, 2001.
- Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature* 411: 41–42, 2001.
- Johnson MM, Houck J, Chen C. Screening for deleterious nonsynonymous single-nucleotide polymorphisms in genes involved in steroid hormone metabolism and response. *Cancer Epidemiol Biomarkers Prev* 14: 1326–1329, 2005.
- Kintscher U, Law RE. PPAR γ -mediated insulin sensitization: the importance of fat versus muscle. *Am J Physiol Endocrinol Metab* 288: E287–E291, 2005.
- Lee JW, Bae SH, Jeong JW, Kim SH, Kim KW. Hypoxia-inducible factor (HIF-1)alpha: its protein stability and biological functions. *Exp Mol Med* 36: 1–12, 2004.
- Li M, Becnel LS, Li W, Fisher WE, Chen C, Yao Q. Signal transduction in human pancreatic cancer: roles of transforming growth factor beta, somatostatin receptors, and other signal intermediates. *Arch Immunol Ther Exp (Warsz)* 53: 381–387, 2005.
- Lin RC, Wang XL, Dalziel B, Caterson ID, Morris BJ. Association of obesity, but not diabetes or hypertension, with glucocorticoid receptor N363S variant. *Obes Res* 11: 802–808, 2003.
- Lin RC, Wang XL, Morris BJ. Association of coronary artery disease with glucocorticoid receptor N363S variant. *Hypertension* 41: 404–407, 2003.
- Mayer BJ. Protein-protein interactions in signaling cascades. *Mol Biotechnol* 13: 201–213, 1999.
- Mehra A, Wrana JL. TGF- β and Smad signal transduction pathway. *Biochem Cell Biol* 80: 605–622, 2002.
- Miyaki M, Kuroki T. Role of Smad4 (DPC4) inactivation in human cancer. *Biochem Biophys Res Commun* 306: 799–804, 2003.
- Morrison DK. Mechanisms regulating Raf-1 activity in signal transduction pathways. *Mol Reprod Dev* 42: 507–514, 1995.
- Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bradley P, Bork P, Bucher P, Cerutti L, Copley R, Courcelle E, Das U, Durbin R, Fleischmann W, Gough J, Haft D, Harte N, Hulo N, Kahn D, Kanapin A, Krestyaninova M, Lonsdale D, Lopez R, Letunic I, Madera M, Maslen J, McDowall J, Mitchell A, Nikolskaya AN, Orchard S, Pagni M, Ponting CP, Quevillon E, Selengut J, Sigrist CJ, Silventoinen V, Studholme DJ, Vaughan R, Wu CH. Interpro, progress and status in 2005. *Nucleic Acids Res* 33: D201–D205, 2005.
- Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res* 12: 436–446, 2002.
- Notarangelo LD, Ochs HD. Wiskott-Aldrich syndrome: a model for defective actin reorganization, cell trafficking and synapse formation. *Curr Opin Immunol* 15: 585–591, 2003.

39. Nourry C, Grant SG, Borg JP. PDZ domain proteins: plug and play! *Sci STKE* 179: RE7, 2003.
40. Obenaus JC, Cantley LC, Yaffe MB. Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res* 31: 3635–3641, 2003.
41. Ohayon T, Gershoni-Baruch R, Papa MZ, Distelman Menachem T, Eisenberg Barzilai S, Friedman E. The R72P P53 mutation is associated with familial breast cancer in Jewish women. *Br J Cancer* 92: 1144–1148, 2005.
42. Pawson T, Nash P. Assembly of cell regulatory systems through protein interaction domains. *Science* 300: 445–452, 2003.
43. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, Niranjan V, Muthusamy B, Gandhi TK, Gronborg M, Ibarrola N, Deshpande N, Shanker K, Shivashankar HN, Rashmi BP, Ramya MA, Zhao Z, Chandrika KN, Padma N, Harsha HC, Yatish AJ, Kavitha MP, Menezes M, Choudhury DR, Suresh S, Ghosh N, Saravana R, Chandran S, Krishna S, Joy M, Anand SK, Madavan V, Joseph A, Wong GW, Schiemann WP, Constantinescu SN, Huang L, Khosravi-Far R, Steen H, Tewari M, Ghaffari S, Blobe GC, Dang CV, Garcia JG, Pevsner J, Jensen ON, Roepstorff P, Deshpande KS, Chinnaiyan AM, Hamosh A, Chakravarti A, Pandey A. Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res* 32: D497–D501, 2004.
44. Pim D, Banks L. p53 polymorphic variants at codon 72 exert different effects on cell cycle progression. *Int J Cancer* 108: 196–199, 2004.
45. Povey S, Lovering R, Bruford E, Wright M, Lush M, Wain H. The HUGO Gene Nomenclature Committee (HGNC). *Hum Genet* 109: 678–680, 2001.
46. Prior SJ, Hagberg JM, Phares DA, Brown MD, Fairfull L, Ferrell RE, Roth SM. Sequence variation in hypoxia-inducible factor 1 α (HIF1A): association with maximal oxygen consumption. *Physiol Genomics* 15: 20–26, 2003.
47. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. *Nucleic Acids Res* 33: W116–W120, 2005.
48. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 30: 3894–3900, 2002.
49. Ramesh Babu L, Wilson SG, Dick IM, Islam FM, Devine A, Prince RL. Bone mass effects of a BMP4 gene polymorphism in postmenopausal women. *Bone* 36: 555–561, 2005.
50. Ridker PM, Cook NR, Cheng S, Erlich HA, Lindpaintner K, Plutzky J, Zee RY. Alanine for proline substitution in the peroxisome proliferator-activated receptor gamma-2 (PPARG2) gene and the risk of incident myocardial infarction. *Arterioscler Thromb Vasc Biol* 23: 859–863, 2003.
51. Rosmond R. The glucocorticoid receptor gene and its association to metabolic syndrome. *Obes Res* 10: 1078–1086, 2002.
52. Roussel R, Reis AF, Dubois-Laforgue D, Bellanne-Chantelot C, Timsit J, Velho G. The N363S polymorphism in the glucocorticoid receptor gene is associated with overweight in subjects with type 2 diabetes mellitus. *Clin Endocrinol (Oxf)* 59: 237–241, 2003.
53. Savas S, Kim DY, Ahmad MF, Shariff M, Ozcelik H. Identifying functional genetic variants in DNA repair pathway using protein conservation analysis. *Cancer Epidemiol Biomarkers Prev* 13: 801–807, 2004.
54. Savas S, Ahmad MF, Shariff M, Kim DY, Ozcelik H. Candidate nsSNPs that can affect the functions and interactions of cell cycle proteins. *Proteins* 58: 697–705, 2005.
55. Savas S, Ozcelik H. Phosphorylation states of cell cycle and DNA repair proteins can be altered by the nsSNPs. *BMC Cancer* 5: 107, 2005.
56. Serra R, Crowley MR. Mouse models of transforming growth factor beta impact in breast development and cancer. *Endocr Relat Cancer* 12: 749–760, 2005.
57. Shastry BS. SNP alleles in human disease and evolution. *J Hum Genet* 47: 561–566, 2002.
58. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29: 308–311, 2001.
59. Smith WM, Zhou XP, Kurose K, Gao X, Latif F, Kroll T, Sugano K, Cannistra SA, Clinton SK, Maher ER, Prior TW, Eng C. Opposite association of two PPARG variants with cancer: overrepresentation of H449H in endometrial carcinoma cases and underrepresentation of P12A in renal cell carcinoma cases. *Hum Genet* 109: 146–151, 2001.
60. Tanimoto K, Yoshiga K, Eguchi H, Kaneyasu M, Ukon K, Kumazaki T, Oue N, Yasui W, Imai K, Nakachi K, Poellinger L, Nishiyama M. Hypoxia-inducible factor-1 α polymorphisms associated with enhanced transactivation capacity, implying clinical significance. *Carcinogenesis* 24: 1779–1783, 2003.
61. Ten Dijke P, Hill CS. New insights into TGF-beta-Smad signalling. *Trends Biochem Sci* 29: 265–273, 2004.
62. Thomas DC. The need for a systematic approach to complex pathways in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 14: 557–559, 2005.
63. Vlassara H. The AGE-receptor in the pathogenesis of diabetic complications. *Diabetes Metab Res Rev* 17: 436–443, 2001.
64. Vondriska TM, Pass JM, Ping P. Scaffold proteins and assembly of multiprotein signaling complexes. *J Mol Cell Cardiol* 37: 391–397, 2004.
65. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetverin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Helmberg W, Kapustin Y, Kenton DL, Khovayko O, Lipman DJ, Madden TL, Maglott DR, Ostell J, Pruitt KD, Schuler GD, Schriml LM, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Suzek TO, Tatusov R, Tatusova TA, Wagner L, Yaschenko E. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 34: D173–D180, 2006.
66. Yaffe MB, Leparic GG, Lai J, Obata T, Volinia S, Cantley LC. A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat Biotechnol* 19: 348–353, 2001.
67. Yamada N, Horikawa Y, Oda N, Iizuka K, Shihara N, Kishi S, Takeda J. Genetic variation in the hypoxia-inducible factor-1 α gene is associated with type 2 diabetes in Japanese. *J Clin Endocrinol Metab* 90: 5841–5847, 2005.