BMC Cancer

Research article

Phosphorylation states of cell cycle and DNA repair proteins can be altered by the nsSNPs

Sevtap Savas[1,2,3] and Hilmi Ozcelik*[1,2,3]

Address: [1]Fred A. Litwin Centre for Cancer Genetics, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, M5G 1X5, ON, Canada, [2]Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, M5G 1X5, ON, Canada and [3]Department of Laboratory Medicine and Pathobiology, University of Toronto, M5G 1L5, Toronto, ON, Canada

Email: Sevtap Savas - savas@mshri.on.ca; Hilmi Ozcelik* - ozcelik@mshri.on.ca

* Corresponding author

Abstract

Background: Phosphorylation is a reversible post-translational modification that affects the intrinsic properties of proteins, such as structure and function. Non-synonymous single nucleotide polymorphisms (nsSNPs) result in the substitution of the encoded amino acids and thus are likely to alter the phosphorylation motifs in the proteins.

Methods: In this study, we used the web-based NetPhos tool to predict candidate nsSNPs that either introduce or remove putative phosphorylation sites in proteins that act in DNA repair and cell cycle pathways.

Results: Our results demonstrated that a total of 15 nsSNPs (16.9%) were likely to alter the putative phosphorylation patterns of 14 proteins. Three of these SNPs (CDKN1A-S31R, OGG1-S326C, and XRCC3-T241M) have already found to be associated with altered cancer risk. We believe that this set of nsSNPs constitutes an excellent resource for further molecular and genetic analyses.

Conclusion: The novel systematic approach used in this study will accelerate the understanding of how naturally occurring human SNPs may alter protein function through the modification of phosphorylation mechanisms and contribute to disease susceptibility.

Background

Phosphorylation is a common, reversible post-translational modification that occurs at serine (S), threonine (T), and tyrosine (Y) residues in proteins [1]. Overall, phosphorylation can alter the structure, function, interaction, stability, and the sub-cellular location of the proteins [2-4], and therefore play an indispensable role in regulation of the cellular processes such as signal transduction, gene expression, cytoskeletal regulation, apoptosis, homeostasis, cell cycle, and DNA damage recognition and repair [5-11]. The phosphorylation state of a protein is

determined by the opposing actions of kinases and phosphatases [12]. Proteins may contain multiple phosphorylation sites, which may be targeted by different kinases/phosphatases [2]. The activity of kinases and phosphatases at different times and/or upon different stimuli provides a means of powerful control over the protein phosphorylation state and thus the biological processes the protein is involved in.

In the post-genomic era, there is an expanding interest in identification of the single nucleotide polymorphisms

(SNPs) that might affect the protein function and thus contribute to the disease susceptibility. The non-synonymous SNPs (nsSNPs) substitute encoded amino acids in proteins, and therefore are good candidates as disease-modifiers. A variety of approaches have been developed and applied, based on criteria such as the evolutionary conservation status or structural parameters, to characterize and select the nsSNPs that are most likely to have functional consequences [13-19].

In this report, we predicted the potential effect of a set of nsSNPs [20,21] in altering the phosphorylation status of DNA repair and cell cycle proteins using the NetPhos tool [22], which is an artificial neural network method that predicts the phosphorylation sites with a sensitivity of 69–96%. DNA repair and cell cycle pathways interact during the cell growth and division to maintain the genomic stability of dividing cells. Abnormalities in the DNA repair and/or the cell cycle pathways can lead to abnormal cell growth/division or cellular death [23], and are implicated in many human diseases, including cancer [24-30]. Functional significance of many phosphorylated residues of several DNA repair and cell cycle proteins has already been evaluated. For example, phosphorylation of STATα residue S727 is required for its maximal transcriptional activation [31] and enhances its binding to the BRCA1 protein [32]. Similarly, phosphorylation of S383 and S387 are required for the FANCG function during mitosis [33]. Likewise, mutations of the phosphorylated residues Ser366 and Thr387 of p53 affect its transactivation function [34]. To our knowledge, although SNPs of DNA repair and cell cycle proteins have already been shown to contribute to cancer risk [35-37], the potential role of nsSNPs in alteration of phosphorylation patterns of proteins has not been evaluated before. Therefore, the novel approach described in this study will accelerate the formation of a bridge between variations in DNA repair/cell cycle function and predisposition to disease.

Methods
The nsSNPs extracted from public SNP databases were previously reported [20,21], however, only the nsSNPs that were found in ≥2 chromosomes in a sample panel of ≥46 chromosomes were included into that manuscript. A total of 89 nsSNPs from 47 genes involved in DNA repair and cell cycle constituted the final data set. The NetPhos [22] algorithm was utilized to predict putative phosphorylation sites for both the wild type and the variant protein sequences. Only the predictions that remove or create a site at either the SNP location or at kinase recognition motifs are included into this manuscript. Please note that the BRCA1 and NFKB1 proteins were initially identified as cell cycle protein interacting proteins [21]. However, in this manuscript, we classified the BRCA1 as a DNA repair and the NFKB1 as a cell cycle protein. The mouse ortho-

logues were retrieved from the LocusLink resource of NCBI [38] and aligned with the human proteins using the ClustalW program [39] to identify the corresponding mouse residue.

Results and discussion
We utilized the NetPhos algorithm to predict putative phosphorylation sites along the DNA repair and cell cycle proteins, and studied whether 89 naturally occurring nsSNPs (64 from 28 DNA repair and 25 from 19 cell cycle genes) might alter the phosphorylation patterns in these proteins. The sensitivity of NetPhos prediction has been reported to be 69–96% with a false-positive prediction rate of 0–26% for Y, 0–11% for S, and 0–14% for T [22]. The results obtained using the NetPhos software are shown in Table I, and are summarized in Table II. Our results have shown that 16.9% (15/89) of the nsSNPs studied are likely to abolish or create 17 putative phosphorylation sites in 44.0% (14/32) of the proteins. As summarized in Table II, five nsSNPs (ERCC5-S311C, OGG1-S326C, XRCC3-T241M, CCND3-S259A, and CDKN1A-S31R) were predicted to abolish putative phosphorylation sites, whereas four nsSNPs were predicted to create putative phosphorylation sites in the proteins (ERCC2-H201Y, ERCC4-P379S, LIG4-P231S, and XRCC1-P309S). These nsSNPs resulted in the addition or removal of a S, T or Y residue at the predicted phosphorylation site.

The kinase recognition/interaction motif involves 7–12 amino acids around the phosphorylated residue [40], and the physicochemical characteristics of these amino acids determine the specificity of the protein kinases [41,42]. Thus, the amino acid substitutions within the kinase recognition motifs are likely to influence the substrate recognition and the subsequent phosphorylation by kinases. Accordingly, we have identified six nsSNPs (Table I, II) located within the phosphorylation motif of six proteins (within 4 amino acids on either side of the putative phosphorylated residue based on NetPhos outputs) that abolished eight putative phosphorylation sites (BRCA1-P871L at S868, BRCA1-S1040N at S1041, ERCC5-S311C at S310, IGHMBP2-T671A at S672, WRN-S1079L at S1083 and at S1084, CCNI-V207I at S208, and NFKB1-H712Q at T716). Interestingly, NetPhos predicts two overlapping phosphorylation motifs for the ERCC5-S311C nsSNP (S311 SLPS<u>S</u>SKMH and S310 ESLP<u>S</u>SKM), which are both completely abolished by the substitution of the serine residue (position 311) with a cysteine (Table I). Similarly, the WRN-S1079L nsSNP was also predicted to remove 2 putative overlapping phosphorylation motifs (S1083 SKTV<u>S</u>SGTK and S1084 KTVS<u>S</u>GTKE) simultaneously.

Table 1: nsSNPs that abolish or create putative phosphorylated residues in DNA repair and cell cycle proteins. Only the NetPhos [22] predictions that remove or create a site at either the SNP location or at kinase recognition motifs are shown. The nsSNPs that create or abolish putative phosphorylation sites at the nsSNP position are shown in bold. Under the wild type and variant columns are the NetPhos outputs with the location of the amino acid, the phosphorylation motif (the putative phosphorylated residue is underlined), the score, and the residue being phosphorylated. 1 and 2 under the frequency column represents the nsSNP minor allele frequencies <5% and ≥5%, respectively [20-21]. Please note that the BRCA1 and NFKB1 proteins were initially identified as cell cycle protein interacting proteins [21]. However, in this manuscript, we classified the BRCA1 as a DNA repair and the NFKB1 as a cell cycle protein. §The putative phosphorylation sites that are also predicted in mouse proteins.

Pathway	Gene	Accession #	SNP ID	nsSNP	Freq.	Wild Type	Variant
DNA repair	BRCA1	NM_007294.1	SNP000007492 rs799917	P871L	2	868 SKRQSFAPF 0.599 *S*	-
	BRCA1	NM_007294.1	rs4986852	S1040N	1	§1041 KEASSSNIN 0.557 *S*	-
	ERCC2	NM_000400.1	SNP000000054 rs1799792	H201Y	1	-	**201 YSILYANVV 0.745 *Y***
	ERCC4	NM_005236.1	SNP000000067	P379S	1 and 2	-	**379 LESNSKWEA 0.507 *S***
	ERCC5	NM_000123.1	SNP001026027 rs2307491	S311C	1	a) **311 SLPSSSKMH 0.990 *S*** b) §310 ESLPSSSKM 0.645 *S*	-
	IGHMBP2	NM_002180.1	SNP000012785 rs622082	T671A	2	§672 GPATSRTRG 0.634 *S*	-
	LIG4	NM_002312.2	rs3093765	P231S	1	-	**231 QLHDSSVGL 0.562 *S***
	OGG1	NM_002542.4	SNP000064679	S326C	2	**326 DLRQSRHAQ 0.990 *S***	-
	WRN	NM_000553.2	SNP001026663 rs3087414	S1079L	1	a) §1083 SKTVSSGTK 0.790 *S* b) 1084 KTVSSGTKE 0.829 *S*	-
	XRCC1	NM_006297.1	SNP000064196 rs25491	P309S	1	-	**309 EPRRSRAGP 0.996 *S***
	XRCC3	NM_005432.2	SNP000000060	T241M	2	§**241 SLGATLREL 0.849 *T***	-
Cell Cycle	CCND3	NM_001760.2	rs1051130	S259A	2	**259 LREASQTSS 0.982 *S***	-
	CCNI	NM_006835.2	rs4252903	V207I	1	§208 LAMVSLEME 0.664 *S*	-
	CDKN1A	NM_000389.2	SNP000003435 rs1801270 GAI870831 GAI1503061	S31R	2	**31 SEQLSRDCD 0.924 *S***	-
	NFKB1	NM_003998.2	rs4648099	H712Q	1	716 HVDSTTYDG 0.595 *T*	-

Table 2: Distribution of the nsSNPs predicted to alter the phosphorylation sites.

nsSNP	DNA repair	Cell cycle	Total
Abolished ≥1 putative phosphorylated residue (S, T or Y) at the nsSNP location	3	2	5
Abolished ≥1 putative phosphorylated residue by changing the kinase recognition motif	5	2	7
Created ≥1 putative phosphorylated residue (S, T or Y) at the nsSNP location	4	0	4
Created ≥1 putative phosphorylated residue by changing the kinase recognition motif	0	0	0

The Swiss-Prot [43], HPRD [44], PhosphoBase [45], and Phospho.ELM [46] databases and the existing literature did not reveal any experimentally verified phosphorylation at the predicted sites. Analysis of the mouse orthologues showed that the corresponding amino acids at the BRCA1-S1041, CCNI-S208, ERCC5-S310, IGHMBP2-S672, WRN-S1083 and XRCC3-T241 residues were also predicted to be phosphorylated, suggesting that these motifs/sites might have been evolutionarily conserved between two species. On the other hand, the remaining phosphorylation sites, which are not detected in mouse proteins, may represent the newly evolved phosphorylation motifs in human. However, considering the false-

positive rate of NetPhos as well as the possibility that the negative selection acting on the nsSNP sites can result in higher false-positive rates, we cannot totally rule out that all predictions in Table 2 are false. Yet these predictions are still of a great value and suggest possible phosphorylation sites that can be experimentally evaluated. In future, when sufficient molecular data regarding the phosphorylation status of orthologous proteins is available, more systematic analyses can be performed to maximize the accuracy of phosphorylation predictions.

We have also performed an extensive literature review to investigate the role of the reported nsSNPs (minor allele

Table 3: Common nsSNPs with a possible role in cancer predisposition. Only the information derived from the studies on the protein function as well as the studies with a suggestion of disease-association have been included. 1 and 2 under the frequency column represents the nsSNP with minor allele frequencies <5% and ≥5%, respectively [20-21].

Pathway	Gene	nsSNP	Possible effect on phosphorylation	Frequency	Functional analysis	Cancer risk association
DNA repair	BRCA1	P871L	Abolishes at S869	2	-	-
	ERCC4	P379S	Creates at S379	1 and 2	-	-
	IGHMBP2	T671A	Abolishes at S672	2	-	-
	OGG1	S326C	Abolishes at S326	2	Yamane *et al.* [55]	Sugimura *et al.* [52]; Xing *et al.* [53]; Elahi *et al.* [54]
	XRCC3	T241M	Abolishes and T241	2	Yoshihara *et al.* [51]	Winsey *et al.* [49]; Kuschel *et al.* [47]; Shen *et al.* [50]; Figueiredo *et al.* [48]
Cell cycle	CCND3	S259A	Abolishes at S359	2	-	-
	CDKN1A	S31R	Abolishes at S31	2	Chedid *et al.* [59]	Wu *et al.* [58]; Roh *et al.* [56]; Tsai *et al.* [57]

frequencies ≥5%) in human cancer predisposition (Table III). Supporting our hypothesis, three SNPs (CDKN1A-S31R, OGG1-S326C, and XRCC3-T241M) have already found to be associated with altered cancer risk. XRCC3-T241M nsSNP was reported to be associated with increased breast cancer [47,48] and melanoma risk [49], and was also found to be protective against bladder cancer in heavy smokers [50]. XRCC3 is a key DNA repair protein involved in base excision repair [29] and is involved in repairing the alterations caused by many DNA damaging agents. Recently, the XRCC3-M241 variant has been associated with increased risk of incidence of tetraploid cells, frequently observed in cancers, through affecting the function of the XRCC3- and Rad52-associated RPA protein [51]. Similarly, the OGG1-S326C SNP was found to be associated with increased lung [52], orolaryngeal and esophageal cancer risk [53,54]. OGG1 is a DNA repair protein that is protective against the mutations induced by the 8-hydroxyguanine. Yamane *et al.*, [55] suggested that OGG1-C326, when compared to OGG1-S326, was associated with a lower repair capacity for 8-hydroxyguanine induced mutations in human cells. In the case of CDKN1A-S31R, the CDKN1A-S31 was suggested to be associated with increased endometrial cancer [56] whereas CDKN1A-R31 was associated with increased primary open-angle glaucoma [57] and esophageal cancer risk [58]. The CDKN1A-R31 form of the protein was not significantly different than the CDKN1A-S31 form in terms of its ability to suppress colony formation [59]. However, it is not clear whether this result would suggest that the CDKN1A-R31 would be functionally equivalent to the wild type allele in other diverse cellular mecha-

nisms that the CDKN1A protein is involved in, such as apoptosis, cell migration, and senescence [60,61].

In addition to the SNPs already implicated in cancer risk, we identified one relatively common nsSNP potentially altering the phosphorylation pattern of a major breast and ovarian cancer susceptibility gene, BRCA1. The BRCA1-P871L SNP was not found to be associated with either breast [62] or ovarian cancer risk [63], however, further analyses is required to see whether this nsSNP or the other nsSNPs in Table III play a role in susceptibility to other cancer types.

How can we explain that commonly occurring nsSNPs (minor allele frequencies ≥5%) are likely to affect the phosphorylation and thus the function of the proteins? If the phosphorylation site is necessary for the function of the protein and the protein is necessary for the fitness of the organism (indispensable/essential protein), then we would expect such nsSNPs (deleterious alleles) to be either removed from the population or be kept at low allele frequency by means of the purifying selection. Thus, in this case, one can conclude that the common nsSNPs presented in this report can be falsely predicted as removing/creating putative phosphorylation sites by NetPhos program. However, the allele frequencies of the deleterious alleles from proteins that are essential for fitness get higher than expected when the nsSNPs are a) created by hot-spot mutation mechanism(s), b) subject to balancing selection, too [64]. Alternatively, even though the nsSNPs (and the abolished/created phosphorylation sites) have important impact on the protein function, the protein and/or the altered protein function may not affect the fit-

ness, which can also explain the lack of purifying selection against such nsSNPs and their relatively high minor allele frequencies. Besides, the biological consequences of altered protein function may only be exerted under certain environmental conditions.

Conclusion

Here we report a set of nsSNPs in DNA repair and cell cycle genes that are predicted to alter the phosphorylation motifs of the encoded proteins, with possible consequences on protein function, structure, interaction, and stability. If the nsSNPs with a ≥5% minor allele frequency listed in Table III do indeed alter the phosphorylation state of the corresponding proteins, they then represent important candidates for disease susceptibility studies, especially relating to cancer risk. We conclude with the suggestion that our approach and the resulting data indicate a novel mechanism of SNP action: alteration of the functional characteristics of the proteins through phosphorylation may significantly contribute to our understanding of the molecular basis of complex diseases, such as cancer. This study is unique in the sense that it systematically links the possible post-translational modification functional effects of SNPs to disease (cancer) predisposition.

List of abbreviations

SNP: single nucleotide polymorphism; nsSNP: non-synonymous SNP.

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

SS participated in design of the study, collected and analyzed the data and prepared the draft of the manuscript. HO participated in the design and coordination of the study, and helped to draft the manuscript. Both authors read and approved the final manuscript.

Acknowledgements

References

1. Johnson LN, O'Reilly M: **Control by phosphorylation.** *Curr Opin Struct Biol* 1996, **6(6):**762-769.
2. Cohen P: **The regulation of protein function by multisite phosphorylation – a 25 year update.** *Trends Biochem Sci* 2000, **25(12):**596-601.
3. Pawson T: **Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems.** *Cell* 2004, **116(2):**191-203.
4. Zhang HG, Wang J, Yang X, Hsu HC, Mountz JD: **Regulation of apoptosis proteins in cancer cells by ubiquitin.** *Oncogene* 2004, **23(11):**2009-2015.
5. Szumiel I: **Monitoring and signaling of radiation-induced damage in mammalian cells.** *Radiat Res* 1998, **150(5 Suppl):**S92-101.
6. Graves JD, Krebs EG: **Protein phosphorylation and signal transduction.** *Pharmacol Ther* 1999, **82(2–3):**111-121.
7. Hunter T: **Signaling – 2000 and beyond.** *Cell* 2000, **100(1):**113-127.
8. Henneke G, Koundrioukoff S, Hubscher U: **Multiple roles for kinases in DNA replication.** *EMBO Rep* 2003, **4(3):**252-256.
9. McGowan CH: **Regulation of the eukaryotic cell cycle.** *Prog Cell Cycle Res* 2003, **5:**1-4.
10. Vermeulen K, Van Bockstaele DR, Berneman ZN: **The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer.** *Cell Prolif* 2003, **36(3):**131-149.
11. Westermann S, Weber K: **Post-translational modifications regulate the microtubule function.** *Nat Rev Mol Cell Biol* 2003, **4(12):**938-947.
12. Lester LB, Scott JD: **Anchoring and scaffold proteins for kinases and phosphatases.** *Recent Prog Horm Res* 1997, **52:**409-429.
13. Chasman D, Adams RM: **Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation.** *J Mol Biol* 2001, **307:**683-706.
14. Ng PC, Henikoff S: **Predicting deleterious amino acid substitutions.** *Genome Res* 2001, **11:**863-874.
15. Ng PC, Henikoff S: **Accounting for human polymorphisms predicted to affect protein function.** *Genome Res* 2002, **12:**436-446.
16. Wang Z, Moult J: **SNPs, protein structure, and disease.** *Hum Mutat* 2001, **17:**263-270.
17. Ramensky V, Bork P, Sunyaev S: **Human non-synonymous SNPs: server and survey.** *Nucleic Acids Res* 2002, **30:**3894-3900.
18. Stitziel NO, Tseng YY, Pervouchine D, Goddeau D, Kasif S, Liang J: **Structural location of disease-associated single-nucleotide polymorphisms.** *J Mol Biol* 2003, **327(5):**1021-1030.
19. Clifford RJ, Edmonson MN, Nguyen C, Buetow KH: **Large-scale analysis of non-synonymous coding region single nucleotide polymorphisms.** *Bioinformatics* 2004, **20(7):**1006-1014.
20. Savas S, Kim DY, Ahmad MF, Shariff M, Ozcelik H: **Identifying functional genetic variants in DNA repair pathways using protein conservation analysis.** *Cancer Epidemiol Biomarkers Prev* 2004, **13(5):**801-807.
21. Savas S, Ahmad MF, Shariff M, Kim DY, Ozcelik H: **Candidate nsSNPs that can affect the functions and the interactions of cell cycle proteins.** *Proteins* 2005, **58(3):**697-705.
22. Blom N, Gammeltoft S, Brunak S: **Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites.** *J Mol Biol* 1999, **294(5):**1351-1362.
23. Norbury CJ, Zhivotovsky B: **DNA damage-induced apoptosis.** *Oncogene* 2004, **23(16):**2797-2808.
24. Latif C, Harvey SH, O'Connell MJ: **Ensuring the stability of the genome: DNA damage checkpoints.** *Scientific World Journal* 2001, **1:**684-702.
25. Malumbres M, Barbacid M: **To cycle or not to cycle: a critical decision in cancer.** *Nat Rev Cancer* 2001, **1(3):**222-231.
26. Duker NJ: **Chromosome breakage syndromes and cancer.** *Am J Med Genet* 2002, **115(3):**125-129.
27. Thompson LH, Schild D: **Recombinational DNA repair and human disease.** *Mutat Res* 2002, **509(1–2):**49-78.
28. Sarasin A: **An overview of the mechanisms of mutagenesis and carcinogenesis.** *Mutat Res* 2003, **544(2–3):**99-106.
29. Thacker J, Zdzienicka MZ: **The mammalian XRCC genes: their roles in DNA repair and genetic stability.** *DNA Repair (Amst)* 2003, **2(6):**655-672.
30. Motoyama N, Naka K: **DNA damage tumor suppressor genes and genomic instability.** *Curr Opin Genet Dev* 2004, **14(1):**11-16.
31. Zhu X, Wen Z, Xu LZ, Darnell JE Jr: **Stat1 serine phosphorylation occurs independently of tyrosine phosphorylation and requires an activated Jak2 kinase.** *Mol Cell Biol* 1997:6618-6623.
32. Ouchi T, Lee SW, Ouchi M, Aaronson SA, Horvath CM: **Collaboration of signal transducer and activator of transcription 1 (STAT1) and BRCA1 in differential regulation of IFN-**

gamma target genes. *Proc Natl Acad Sci U S A* 2000, **97(10)**:5208-5213.

33. Mi J, Qiao F, Wilson JB, High AA, Schroeder MJ, Stukenberg PT, Moss A, Shabanowitz J, Hunt DF, Jones NJ, Kupfer GM: **FANCG is phosphorylated at serines 383 and 387 during mitosis.** *Mol Cell Biol* 2004, **24(19)**:8576-8585.

34. Ou YH, Chung PH, Sun TP, Shieh SY: **p53 C-Terminal Phosphorylation by CHK1 and CHK2 Participates in the Regulation of DNA-Damage-induced C-Terminal Acetylation.** *Mol Biol Cell* 2005, **16(4)**:1684-1695.

35. Goode EL, Ulrich CM, Potter JD: **Polymorphisms in DNA repair genes and associations with cancer risk.** *Cancer Epidemiol Biomarkers Prev* 2002, **11(12)**:1513-1530.

36. Koike H, Suzuki K, Satoh T, Ohtake N, Takei T, Nakata S, Yamanaka H: **Cyclin D1 gene polymorphism and familial prostate cancer: the AA genotype of A870G polymorphism is associated with prostate cancer risk in men aged 70 years or older and metastatic stage.** *Anticancer Res* 2003, **23(6D)**:4947-4951.

37. Chang BL, Zheng SL, Isaacs SD, Wiley KE, Turner A, Li G, Walsh PC, Meyers DA, Isaacs WB, Xu J: **A polymorphism in the CDKN1B gene is associated with increased risk of hereditary prostate cancer.** *Cancer Res* 2004, **64(6)**:1997-1999.

38. Wheeler DL, Church DM, Edgar R, Federhen S, Helmberg W, Madden TL, Pontius JU, Schuler GD, Schriml LM, Sequeira E, Suzek TO, Tatusova TA, Wagner L: **Database resources of the National Center for Biotechnology Information: update.** *Nucleic Acids Res* 2004:D35-40.

39. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD: **Multiple sequence alignment with the Clustal series of programs.** *Nucleic Acids Res* 2003, **31(13)**:3497-3500.

40. Songyang Z, Blechner S, Hoagland N, Hoekstra MF, Piwnica-Worms H, Cantley LC: **Use of an oriented peptide library to determine the optimal substrates of protein kinases.** *Curr Biol* 1994, **4(11)**:973-982.

41. Kreegipuu A, Blom N, Brunak S, Jarv J: **Statistical analysis of protein kinase specificity determinants.** *FEBS Lett* 1998, **430(1–2)**:45-50.

42. Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, Brunak S: **Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence.** *Proteomics* 2004, **4(6)**:1633-1649.

43. O'Donovan C, Martin MJ, Gattiker A, Gasteiger E, Bairoch A, Apweiler R: **High-quality protein knowledge resource: SWISS-PROT and TrEMBL.** *Brief Bioinform* 2002, **3**:275-284.

44. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, Niranjan V, Muthusamy B, Gandhi T, 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: **Development of human protein reference database as an initial platform for approaching systems biology in humans.** *Genome Res* 2003, **13(10)**:2363-2371.

45. Kreegipuu A, Blom N, Brunak S: **PhosphoBase, a database of phosphorylation sites: release 2.0.** *Nucleic Acids Res* 1999, **27(1)**:237-239.

46. Diella F, Cameron S, Gemund C, Linding R, Via A, Kuster B, Sicheritz-Ponten T, Blom N, Gibson TJ: **Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins.** *BMC Bioinformatics* 2004, **5(1)**:79.

47. Kuschel B, Auranen A, McBride S, Novik KL, Antoniou A, Lipscombe JM, Day NE, Easton DF, Ponder BA, Pharoah PD, Dunning A: **Variants in DNA double-strand break repair genes and breast cancer susceptibility.** *Hum Mol Genet* 2002, **11(12)**:1399-1407.

48. Figueiredo JC, Knight JA, Briollais L, Andrulis IL, Ozcelik H: **Polymorphisms XRCC1-R399Q and XRCC3-T241M and the risk of breast cancer at the Ontario site of the Breast Cancer Family Registry.** *Cancer Epidemiol Biomarkers Prev* 2004, **13(4)**:583-591.

49. Winsey SL, Haldar NA, Marsh HP, Bunce M, Marshall SE, Harris AL, Wojnarowska F, Welsh KI: **A variant within the DNA repair gene XRCC3 is associated with the development of melanoma skin cancer.** *Cancer Res* 2000, **60(20)**:5612-5616.

50. Shen M, Hung RJ, Brennan P, Malaveille C, Donato F, Placidi D, Carta A, Hautefeuille A, Boffetta P, Porru S: **Polymorphisms of the DNA repair genes XRCC1, XRCC3, XPD, interaction with environmental exposures, and bladder cancer risk in a case-control study in northern Italy.** *Cancer Epidemiol Biomarkers Prev* 2003, **12(11 Pt 1)**:1234-1240.

51. Yoshihara T, Ishida M, Kinomura A, Katsura M, Tsuruga T, Tashiro S, Asahara T, Miyagawa K: **XRCC3 deficiency results in a defect in recombination and increased endoreduplication in human cells.** *EMBO J* 2004, **23(3)**:670-680.

52. Sugimura H, Kohno T, Wakai K, Nagura K, Genka K, Igarashi H, Morris BJ, Baba S, Ohno Y, Gao C, Li Z, Wang J, Takezaki T, Tajima K, Varga T, Sawaguchi T, Lum JK, Martinson JJ, Tsugane S, Iwamasa T, Shinmura K, Yokota J: **hOGG1 Ser326Cys polymorphism and lung cancer susceptibility.** *Cancer Epidemiol Biomarkers Prev* 1999, **8(8)**:669-674.

53. Xing DY, Tan W, Song N, Lin DX: **Ser326Cys polymorphism in hOGG1 gene and risk of esophageal cancer in a Chinese population.** *Int J Cancer* 2001, **95(3)**:140-143.

54. Elahi A, Zheng Z, Park J, Eyring K, McCaffrey T, Lazarus P: **The human OGG1 DNA repair enzyme and its association with orolaryngeal cancer risk.** *Carcinogenesis* 2002, **23(7)**:1229-1234.

55. Yamane A, Kohno T, Ito K, Sunaga N, Aoki K, Yoshimura K, Murakami H, Nojima Y, Yokota J: **Differential ability of polymorphic OGG1 proteins to suppress mutagenesis induced by 8-hydroxyguanine in human cell in vivo.** *Carcinogenesis* 2004, **25(9)**:1689-1694.

56. Roh JW, Kim JW, Park NH, Song YS, Park IA, Park SY, Kang SB, Lee HP: **p53 and p21 genetic polymorphisms and susceptibility to endometrial cancer.** *Gynecol Oncol* 2004, **93(2)**:499-505.

57. Tsai FJ, Lin HJ, Chen WC, Tsai CH, Tsai SW: **A codon 31ser-arg polymorphism of the WAF-1/CIP-1/p21/tumour suppressor gene in Chinese primary open-angle glaucoma.** *Acta Ophthalmol Scand* 2004, **82(1)**:76-80.

58. Wu MT, Wu DC, Hsu HK, Kao EL, Yang CH, Lee JM: **Association between p21 codon 31 polymorphism and esophageal cancer risk in a Taiwanese population.** *Cancer Lett* 2003, **201(2)**:175-180.

59. Chedid M, Michieli P, Lengel C, Huppi K, Givol D: **A single nucleotide substitution at codon 31 (Ser/Arg) defines a polymorphism in a highly conserved region of the p53-inducible gene WAF1/CIP1.** *Oncogene* 1994, **9(10)**:3021-3024.

60. Roninson IB: **Oncogenic functions of tumour suppressor p21(Waf1/Cip1/Sdi1): association with cell senescence and tumour-promoting activities of stromal fibroblasts.** *Cancer Lett* 2002, **179(1)**:1-14.

61. Coqueret O: **New roles for p21 and p27 cell-cycle inhibitors: a function for each cell compartment?** *Trends Cell Biol* 2003, **13(2)**:65-70.

62. Menzel HJ, Sarmanova J, Soucek P, Berberich R, Grunewald K, Haun M, Kraft HG: **Association of NQO1 polymorphism with spontaneous breast cancer in two independent populations.** *Br J Cancer* 2004, **90(10)**:1989-1994.

63. Wenham RM, Schildkraut JM, McLean K, Calingaert B, Bentley RC, Marks J, Berchuck A: **Polymorphisms in BRCA1 and BRCA2 and risk of epithelial ovarian cancer.** *Clin Cancer Res* 2003, **9(12)**:4396-4403.

64. Fay JC, Wu CI: **Sequence divergence, functional constraint, and selection in protein evolution.** *Annu Rev Genomics Hum Genet* 2003, **4**:213-235.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2407/5/107/prepub