BMC Cancer

Research article

Genome wide *in silico* SNP-tumor association analysis

Ping Qiu*†, Luquan Wang†, Mitch Kostich†, Wei Ding†, Jason S Simon and Jonathan R Greene

Address: Bioinformatics Group and Discovery Technology Department at Schering-Plough Research Institute, 2015 Galloping Hill Road, Kenilworth, New Jersey 07033, USA

Email: Ping Qiu* - ping.qiu@spcorp.com; Luquan Wang - luquan.wang@spcorp.com; Mitch Kostich - mitchell.kostich@spcorp.com; Wei Ding - wei.ding@spcorp.com; Jason S Simon - jason.simon@spcorp.com; Jonathan R Greene - jonathan.greene@spcorp.com

* Corresponding author   †Equal contributors

Abstract

Background: Carcinogenesis occurs, at least in part, due to the accumulation of mutations in critical genes that control the mechanisms of cell proliferation, differentiation and death. Publicly accessible databases contain millions of expressed sequence tag (EST) and single nucleotide polymorphism (SNP) records, which have the potential to assist in the identification of SNPs overrepresented in tumor tissue.

Methods: An *in silico* SNP-tumor association study was performed utilizing tissue library and SNP information available in NCBI's dbEST (release 092002) and dbSNP (build 106).

Results: A total of 4865 SNPs were identified which were present at higher allele frequencies in tumor compared to normal tissues. A subset of 327 (6.7%) SNPs induce amino acid changes to the protein coding sequences. This approach identified several SNPs which have been previously associated with carcinogenesis, as well as a number of SNPs that now warrant further investigation

Conclusions: This novel *in silico* approach can assist in prioritization of genes and SNPs in the effort to elucidate the genetic mechanisms underlying the development of cancer.

Background

Expressed Sequence Tags (ESTs) are single-pass, partial sequences of cDNA clones derived from a vast number of disease and normal tissues [1]. ESTs have been used extensively for gene discovery and transcript mapping of genes from a wide number of organisms, including human and mouse [1,2]. ESTs have also been used for SNP identification [3-5], gene expression analysis and transcriptome analysis [6,7]. Currently there are more than 4 millions human ESTs in GenBank dbEST database and the number is still growing.

Susceptibility to common, complex diseases is in part genetically determined [8-11], although the genetic contribution might vary greatly depending on the diseases. Single nucleotide polymorphisms (SNPs) are the most common genetic variation in the human genome, and the number of SNPs identified experimentally is growing tremendously. Currently, dbSNP (build 106) contains more than 2.7 million unique SNPs. These data provides a vital resource to study the role of specific sequence alterations on disease susceptibility as well as drug resistance/sensitivity. In recent years SNPs have been favored as more tractable genotypic markers [12]. As genetic markers, SNPs have several advantages over microsatellites sequence

repeats, including abundance (one every 750–1000 bp) [13], stability, and suitability for high throughput analysis. As a consequence, SNPs are being utilized with increasing frequency as markers in human genetic analysis, such as studies of comparative population variation [14-16] and candidate gene association analysis [17-21]. Finally, the combination of SNP analysis with new approaches to investigate profiles of gene expression and proteomics should lead to fundamental insights into the biological importance of common genetic variations in the human genome [22].

Cancer is a polygenic, complex disease caused by the interaction of many genetic and environmental factors [23,24]. Presently, fewer than 10% of tumor cases are attributable to the inheritance of mutations in a single gene, such as BRCA1/2, BRAF and p53. Mutation in any one gene in the polygenic pathway may have a small effect on the risk of developing cancer in a particular individual, but may still make a substantial contribution to cancer incidence within the population if the mutation is present with high frequency [22,23]. Careful study of the huge number of single nucleotide polymorphism (SNPs) will eventually provide new insights into carcinogenic mechanisms [19,25,26]

In this report, we detail a novel approach which utilizes the publicly available dbEST, and dbSNP datasets http://www.ncbi.nlm.nih.gov to identify SNPs located in genes potentially involved in tumor development.

Methods
SNP classification by clustering with dbEST and in silico genotyping
All unique SNP and EST records were obtained from the NCBI databases (dbSNP build 106 :http://www.ncbi.nlm.nih.gov/SNP, dbEST release 092002 http://www.ncbi.nlm.nih.gov/dbEST/). ESTs sequences and their associated tissue library information were extracted and organized in a relational database (Sybase, SQL Server Release 11.0, CA, Sybase Inc.). The EST cDNA libraries were manually curated and cataloged into tumor and non-tumor libraries. A total of 4153 tumor and 2178 non-tumor cDNA libraries were identified.

EST and SNP sequences were clustered using a "common tag" method as previously described [7], SNP sequences that contiged with ESTs were, for purposes of this study, assumed to map to exons, and thus designated coding SNPs (cSNPs). SNP sequences not aligning with ESTs were excluded from further analysis. For each cSNP, sequence alignment was performed against dbEST using BLASTN [27]. In an effort to eliminate false clustering, 95% identity over 50 bp was selected as a minimum homology threshold. The genotype for each EST at the SNP position

was fetched from the BLAST alignment. For those SNP sequences with at least 50 EST hits, ESTs were grouped further by their tissue sources. The cutoff of 50 ESTs was chosen at random, and corresponds to an average representation of 8 distinct tissues. One SNP allele was picked for each tissue if it was present in greater than 80% of ESTs. A tissue was designated heterozygous if both SNP alleles were present in an equal number of ESTs.

SNP distribution analysis in normal vs. tumor cases
The major and minor allele frequencies for each SNP were calculated for tumor and normal tissue. Fisher's exact test was used to test the significance of occurrence of the SNP genotype in both tissue types. Fisher's exact test is calculated using:

$$P = \frac{(R_1!R_2!..R_m!)(C_1!C_2!...C_n!)}{a_{ij}!N!\prod i,j}$$

Let there exist two such variables X and Y, with m and n observed states, respectively. Now form an $n \times m$ matrix in which the entries a_{ij} represent the number of observations in which $x = i$ and $y = j$. Calculate the row and column sums R_i and C_j, respectively, and the total sum of the matrix

$$N = \sum_i Ri = \sum_j Cj$$

All SNP meeting the Fisher's exact test P value < 0.05 significance threshold were further analyzed for amino acid codon conservation.

Codon conservation analysis for cSNP with P < 0.05
SNP sequences were subject to BLAST analysis against Genbank nr database ftp://ftp.ncbi.nih.gov/genbank/ using BLASTX. The top protein hit with percent identity greater than 95% over a 30 amino acids window size was further analyzed to determine whether the SNP resulted in codon change. The codon which contained the IUPAC code of the SNP was replaced with the corresponding nucleotide codes (A,T,C,G) and tested for amino acid codon change.

Results and Discussions
A large number of studies have focused on investigating genetic polymorphisms in individual genes in order to estimate genetic contribution to the development of cancer [28]. Cancer susceptibility SNPs have been identified among genes with known activity in cell cycle maintenance and DNA repair as well as those encoding phase I and phase II enzymes [28]. Recent advancements in large scale SNP genotyping have made genome-wide SNP association analysis possible [29,30]. However, despite large efforts to identify SNPs in genes previously identified as

Table 1: SNPs counts in each analytical step. SNP sequences that overlap with ESTs are referred as (cSNPs).

Total Starting SNPs (dbSNP Build 106)	~2,700,000
SNPs Overlaps with ESTs(cSNPs)	~740,000
cSNPs with over 50 ESTs	~34,000
cSNPs with P < 0.05	4865
cSNPs with P < 0.05 and identifiable AA change	327

candidates for cancer susceptibility, genome wide identification and characterization of SNPs among cancer patients or tumor tissue has not been reported.

This report describes a comprehensive allele frequency analysis of ~2.7 million unique SNPs in tumor vs normal tissues. The goal of the study was to identify SNPs overrepresented in tumor-derived ESTs using dbEST tissue library information. Initially, all SNPs from dbSNP build 106 were downloaded from NCBI. A total of 741,244 (27.5%) SNPs mapped to transcribed regions by clustering to the dbEST database (release 092002) (Table 1). SNPs overlapping with an EST were further subject to allele frequency analysis using dbEST tissue information. Fisher's exact test identified 4865 (0.66%) SNPs with allele frequencies which were significantly different between tumor and normal tissue. A less conservative confidence interval of $P < 0.05$ was used in this study. A Multiple testing correction was not performed as we were more willing to accept false positives than false negatives. Multiple testing correction, including the Bonferroni correction (which assumes independent markers), would markedly overcorrect for the inflated false-positive rate and thereby throw away valid information. This is especially true given the large number of tests involved in this study and the relatively small P values obtained due to the low count of tissues for each SNP.

Table 2 [see Additional file 1, table2.pdf] summarizes those cSNPs identified by the present analysis with allele frequencies significantly different between tumor and normal tissue that result in amino acid change. Many of these genes are known to be involved in tumor development.

HLA/MHC gene SNPs represented approximately 15% (50/327) of cSNPs identified as differing significantly between tumor and normal tissue. It has been previously reported that tumor cells undergo changes in the major histocompatibility complex (MHC) class I locus during tumor development [31,32]. These HLA losses produce tumor cells that are able to escape anti-tumor T cell immune responses. Defects in the antigen processing machinery and in HLA class I antigens in malignant cells may have a significant impact on the clinical course of

malignant diseases and on the outcome of T cell-based immunotherapy [32]. In addition, MHC class I loss or down regulation in cancer cells is a major immune escape route used by a large variety of human tumors to evade anti-tumor immune responses mediated by cytotoxic T lymphocytes. Multiple mechanisms are responsible for such HLA class I alterations. These data suggest that SNPs in HLA might be another important mechanism that causes loss-of-function, affecting the role of HLA in presenting immunogenic peptides to T cells.

Glutathione S-transferases (GST) constitute a large multigene family of phase II enzymes involved in detoxification of potentially genotoxic chemicals. Total or partial deletions or SNPs in alleles encoding GSTM1, GSTM3, GSTPI, GSTT1, GSTZ1 are associated with reduction of enzymatic activity toward several substrates of different GST isoenzymes. In addition, molecular epidemiological studies indicate that a single SNP in glutathione S-transferase appears to be a moderate lung cancer risk factor. However, the risk is higher when interactions with more GST polymorphisms and other risk factors (e.g. cigarette smoking) occur. Individuals with decreased rate of detoxification or with "high risk" glutathione S-transferase genotypes have a slightly higher level of carcinogen-DNA adducts and more cytogenetic damages [28,33]. Blackburn *et al.* have reported that an A/G transition at position 94 of GSTZ1, which reflects a Lys to Glu changes in the encoded peptide, displayed differences in activity towards several substrates [34]. This SNP (rs7975) was also found to display different allele frequencies in normal compared to tumor tissue in this study. The present study also identified a SNP (rs1065411) in GSTM1, causing a Lys to Gln change at reside 173. Based on the association of the SNP in GSTZ1 with increased cancer risk, further analysis of the variability in GSTM1 is warranted.

The protein kinase PITSLRE is part of the large family of p34[cdc2] related kinases whose functions appear to be linked to the control of cell division and possibly programmed cell death [35]. Evidence also suggests that one or more PITSLRE kinase isoforms may be tumor suppressor genes [36]. It has been suggested that one PITSLRE isoform p110 protein kinase are cleaved *in vivo* by multiple caspases during Fas-mediated cell death at several sites

within the amino-terminal domain and the caspase cleavage of this protein is affected by the phosphorylation [37]. This study identified one SNP (rs1059828) in PITSLRE kinase (amino acid 401 on CDC2L1, NP_277021.1 and amino acid 396 on CDC2L2, NP_284922.1) which yields an amino acid alteration of Ser->Leu with significantly different allele frequencies in normal compared to tumor tissues. Feng et al [38] discovered a similar mutation (C/T at nucleotide location 97 of exon 7, Ser-Leu) on PITSLRE CDC2L1 in the melanoma cell line UACC903. While their exact role remains to be tested, the potential of these two independently identified mutations to induce phosphorylation site changes on PILSLRE kinase, suggest importance in tumor development.

Finally, mitochondria have been reported to play a key role in various apoptotic processes including cell death induced by cytotoxic agents [39,40]. Mitochondria undergoing permeability transition release apoptogenic proteins such as cytochrome *c* and apoptosis-inducing factor from the mitochondrial intermembrane space into the cytosol, where they can activate caspases and endonucleases [39,40]. This analysis has identified several mitochondrial genes including dUTP Pyrophosphatase and ATP synthase with significantly difference SNP allele frequencies. While their role in apoptosis remains to be determined, the large number of SNPs in mitochondrial genes revealed by this analysis suggests that such mutations may contribute to the tumor development.

The approach described here has several limitations. Due to the continually evolving nature of the human protein catalog, SNPs located in previously unannotated coding regions were not included in this analysis. A complete list of all significant SNPs is available [see Additional file 2, all_snp.xls], allowing the analysis to be repeated as the protein catalog is updated. *In silico* analyses are also limited by the quantity and quality of the data present in databases used in the analysis. The data present in dbEST is not well annotated with regard to the precise origin of the source tissue used in cDNA library construction. It is possible, for example, that EST data from multiple tissues sourced from the same donor were used in the present analysis. This lack of diversity could artificially bias the significance of any particular allelic imbalance observed. The homogeneity of the tissue characterized as tumor-derived is another potential source of error. Analysis of actual tumor tissue might contain a large portion of normal tissue into which the tumor infiltrate. For those reasons, the number of ESTs which contain a particular SNP and the diversity of source tissues that contain those SNPs will affect the quality of the analysis. In addition, limited representation of low-abundance transcripts in dbEST likely has introduces a bias towards SNPs present in genes which display widespread tissue distribution, or are

present in tissue types overrepresented in the database. SNPs present in genes expressed at low levels are under represented from this analysis as they were likely to fall short of the protocol thresholds. Another drawback is that somatic mutations might be excluded from the list since the majority of dbSNP entries represent chromosomal mutations and therefore primarily represent inherited polymorphisms. Somatic mutations that cause cancer in some genes (i.e: BRAF) [19] might not be detected if the same mutation is not stably inherited. Lastly, bias was introduced by the using EST tissue library information to assess allele frequencies. This limited the present analysis to SNPs present in known or predicted amino acid coding sequences, excluding those common, functional intronic or promoter region SNPs which may result in splicing or expression changes.

Large scale genotyping of samples from patients will lead to important breakthroughs in understanding mechanism of gene-environment and gene-gene interactions in common polygenic cancers. Effort has been initiated in large sequencing laboratories to carry out comprehensive SNP analysis in all disease candidate genes. However, this is a labor intensive, lengthy and very costly effort. The in silico analysis described here provide a quick and economic approach to screen through a large number of identified SNPs in the human genome to pinpoint possible cancer susceptibility genes, utilizing the rich tissue and library information present in the public dbEST database. Nevertheless, positive associations of SNPs with cancers reported here are very preliminary and are subject to interpretation and careful experimental validation. Only the combined consideration of studies in different populations produce similar results will result in the belief that a SNP is indeed a cancer risk factor.

Although we do not validate all the tumor related genes identified in this report, the approach taken here identified numerous hits in DNA repair genes, genes encoding phase I and phase II enzymes and other tumor related genes, some of which are already under scrutiny by the cancer research community. A couple of the SNPs revealed in this analysis have been suggested to have roles in tumor development in previously published studies [33]. Complementary to any other disease gene and SNP association study, this approach can help to prioritize the genes that need to be validated and further help to elucidate the genetic contribution to the development of cancer. This method can also help to identify new genes or SNPs that might be crucial to tumor development. Additional genome wide screens through cancer cell DNA for somatic mutations ultimately will provide a more complete picture of the number and patterns of mutations underlying human oncogenesis.

Authors' contributions

PQ, LW, WD, MK carried out the data analysis. PQ, JS drafted the manuscript. PQ, LW, WD, MK, JS, JG participated in the design of study. All authors read and approved the final manuscript.

Additional material

Additional File 1

SNPs with significantly different allele frequency in normal vs tumor tissues which result in codon change. P value < 0.05.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2407-4-4-S1.pdf]

Additional File 2

Complete list of SNPs. All SNPs with significantly different allele frequency in normal vs tumor tissue. P value < 0.05.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2407-4-4-S2.xls]

Acknowledgements

References

1. Schuler G: **Pieces of the puzzle: expressed sequence tags and the catalog of human genes.** *J Mol Med* 1997, **75**:694-698.
2. Kawai J, Shinagawa A, Shibata K, Yoshino M, Itoh M, Ishii Y, Arakawa T, Hara A, Fukunishi Y, Konno H, Adachi J, Fukuda S, Aizawa K, Izawa M, Nishi K, Kiyosawa H, Kondo S, Yamanaka I, Saito T, Okazaki Y, Gojobori T, Bono H, Kasukawa T, Saito R, Kadota K, Matsuda H, Ashburner M, Batalov S, Casavant T, Fleischmann W, Gaasterland T, Gissi C, King B, Kochiwa H, Kuehl P, Lewis S, Matsuo Y, Nikaido I, Pesole G, Quackenbush J, Schriml LM, Staubli F, Suzuki R, Tomita M, Wagner L, Washio T, Sakai K, Okido T, Furuno M, Aono H, Baldarelli R, Barsh G, Blake J, Boffelli D, Bojunga N, Carninci P, de Bonaldo MF, Brownstein MJ, Bult C, Fletcher C, Fujita M, Gariboldi M, Gustincich S, Hill D, Hofmann H, Hume DA, Kamiya M, Lee NH, Lyons P, Marchionni L, Mashima J, Mazzarelli J, Mombaerts P, Nordone P, Ring B, Ringwald M, Rodriguez I, Sakamoto N, Sasaki H, Sato K, Schonbach C, Seya T, Shibata Y, Storch KF, Suzuki H, Toyo-oka K, Wang KH, Weitz C, Whittaker C, Wilming L, Wynshaw-Boris A, Yoshida K, Hasegawa Y, Kawaji H, Kohtsuki S, Hayashizaki Y, RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium: **Functional annotation of a full-length mouse cDNA collection.** *Nature* 2001, **409**:685-690.
3. Irizarry K, Kustanovich V, Li C, Brown N, Nelson S, Wong W, Lee CJ: **Genome-wide analysis of single-nucleotide polymorphisms in human expressed sequences.** *Nat Genet* 2000, **26**:233-236.
4. Schmid KJ, Sorensen TR, Stracke R, Torjek O, Altmann T, Mitchell-Olds T, Weisshaar B: **Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in Arabidopsis thaliana.** *Genome Res* 2003, **13**:1250-1257.
5. Barker G, Batley J, O' Sullivan H, Edwards KJ, Edwards D: **Redundancy based detection of sequence polymorphisms in expressed sequence tag data using autoSNP.** *Bioinformatics* 2003, **19**:421-422.
6. Qiu P, Benbow L, Liu S, Greene JR, Wang L: **Analysis of a human brain transcriptome map.** *BMC Genomics* 2002, **3**:10.
7. Benbow L, Wang L, Laverty M, Liu S, Qiu P, Bond RW, Gustafson E, Hedrick JA, Kostich M, Greene JR, Wang L: **A reference database for tumor-related genes co-expressed with interleukin-8 using genome-scale in silico analysis.** *BMC Genomics* 2002, **3**:29.
8. Lander ES: **The new genomics: Global views of biology.** *Science* 1996, **274**:536-539.
9. Collins FS, Guyer MS, Chakravarti A: **Variations on a theme: Cataloging human DNA sequence variation.** *Science* 1997, **278**:1580-1581.
10. Chakravarti A: **Population genetics – Making sense out of sequence.** *Nat Genet* 1999, **21**:56-60.
11. Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, Rainbow DB, Hunter KM, Smith AN, Di Genova G, Herr MH, Dahlman I, Payne F, Smyth D, Lowe C, Twells RC, Howlett S, Healy B, Nutland S, Rance HE, Everett V, Smink LJ, Lam AC, Cordell HJ, Walker NM, Bordin C, Hulme J, Motzo C, Cucca F, Hess JF, Metzker ML, Rogers J, Gregory S, Allahabadia A, Nithiyananthan R, Tuomilehto-Wolf E, Tuomilehto J, Bingley P, Gillespie KM, Undlien DE, Ronningen KS, Guja C, Ionescu-Tirgoviste C, Savage DA, Maxwell AP, Carson DJ, Patterson CC, Franklyn JA, Clayton DG, Peterson LB, Wicker LS, Todd JA, Gough SC: **Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease.** *Nature* 2003, **423**:506-511.
12. Chanock S: **Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease.** *Dis Markers* 2001, **17**:89-98.
13. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J: **Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome.** *Science* 1998, **280**:1077-1082.
14. Holden C: **Race and medicine.** *Science* 2003, **302**:594-596.
15. Goddard KA, Hopkins PJ, Hall JM, Witte JS: **Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations.** *Am J Hum Genet* 2000, **66**:216-234.
16. Ober C, Leavitt SA, Tsalenko A, Howard TD, Hoki DM, Daniel R, Newman DL, Wu X, Parry R, Lester LA: **Variation in the interleukin 4-receptor gene confers susceptibility to asthma and atopy in ethnically diverse populations.** *Am J Hum Genet* 2000, **66**:517-526.
17. Lai E, Riley J, Purvis I, Roses A: **A 4-Mb high-density single nucleotide polymorphism-based map around human APOE.** *Genomics* 1998, **54**:31-38.
18. Martin ER, Gilbert JR, Lai EH, Riley J, Rogala AR, Slotterbeck BD, Sipe CA, Grubber JM, Warren LL, Conneally PM: **Analysis of association at single nucleotide polymorphisms in the APOE region.** *Genomics* 2000, **63**:7-12.
19. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Mould C, Parker A, Stevens C, Watt S, Hooper S, Wilson R, Jayatilake H, Gusterson BA, Cooper C, Shipley J, Hargrave D, Pritchard-Jones K, Maitland N, Chenevix-Trench G, Riggins GJ, Bigner DD, Palmieri G, Cossu A, Flanagan A, Nicholson A, Ho JW, Leung SY, Yuen ST, Weber BL, Seigler HF, Darrow TL, Paterson H, Marais R, Marshall CJ, Wooster R, Stratton MR, Futreal PA: **Mutations of the BRAF gene in human cancer.** *Nature* 2002, **417**:949-954.
20. Long AD, Langley CH: **The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits.** *Genome Res* 1999, **9**:720-731.
21. Risch N, Merikangas K: **The future of genetic studies of complex human diseases.** *Science* 1996, **273**:1516-1517.
22. Taylor JG, Choi EH, Foster CB, Chanock SJ: **Using genetic variation to study human disease.** *Trends Mol Med* 2001, **7**:507-12.
23. Hanahan D, Weinberg RA: **The hallmarks of cancer.** *Cell* 2000, **100**:57-70.
24. Hemminki K, Mutanen P: **Genetic epidemiology of multistage carcinogenesis.** *Mutat Res* 2001, **473**:11-21.
25. Ameyaw MM, Tayeb M, Thornton N, Folayan G, Tariq M, Mobarek A, Evans DA, Ofori-Adjei D, McLead HL: **Ethnic variation in the HER-2 codon 655 genetic polymorphism previously associated with breast cancer.** *J Hum Genet* 2002, **47**:172-175.
26. Mimori K, Inoue H, Shiraishi T, Ueo H, Mafune K, Tanaka Y, Mori M: **A Single-Nucleotide Polymorphism of SMARCB1 in Human Breast Cancers.** *Genomics* 2002, **80**:254-258.
27. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.

28. Reszka E, Wasowicz W: **Significance of genetic polymorphisms in glutathione S-transferase multigene family and lung cancer risk.** *Int J Occup Med Environ Health* 2001, **14:**99-113.

29. Hoque MO, Lee CC, Cairns P, Schoenberg M, Sidransky D: **Genome-wide genetic characterization of bladder cancer: a comparison of high-density single-nucleotide polymorphism arrays and PCR-based microsatellite analysis.** *Cancer Res* 2003, **63:**2216-2222.

30. Dumur CI, Dechsukhum C, Ware JL, Cofield SS, Best AM, Wilkinson DS, Garrett CT, Ferreira-Gonzalez A: **Genome-wide detection of LOH in prostate cancer using human SNP microarray technology.** *Genomics* 2003, **81:**260-269.

31. Ferrone S, Marincola FM: **Loss of HLA class I antigens by melanoma cells: molecular mechanisms, functional significance and clinical relevance.** *Immunol Today* 1995, **16:**487-494.

32. Garrido F, Ruiz-Cabello F, Cabrera T, Perez-Villar JJ, Lopez-Botet M, Duggan-Keen M, Stern PL: **Implications for immunosurveillance of altered HLA class I phenotypes in human tumours.** *Immunol Today* 1997, **18:**89-95.

33. Stucker I, Hirvonen A, de Waziers I, Cabelguenne A, Mitrunen K, Cenee S, Koum-Besson E, Hemon D, Beaune P, Loriot MA: **Genetic polymorphisms of glutathione S-transferases as modulators of lung cancer susceptibility.** *Carcinogenesis* 2002, **23:**1475-1481.

34. Blackburn AC, Tzeng HF, Anders MW, Board PG: **Discovery of a functional polymorphism in human glutathione transferase zeta by expressed sequence tag database analysis.** *Pharmacogenetics* 2000, **10:**49-57.

35. Loyer P, Trembley JH, Lahti JM, Kidd VJ: **The RNP protein, RNPS1, associates with specific isoforms of the p34cdc2-related PITSLRE protein kinase in vivo.** *J Cell Sci* 1998, **111:**1495-1506.

36. Beyaert R, Kidd VJ, Cornelis S, Van de Craen M, Denecker G, Lahti JM, Gururajan R, Vandenabeele P, Fiers W: **Cleavage of PITSLRE kinases by ICE/CASP-1 and CPP32/CASP-3 during apoptosis induced by tumor necrosis factor.** *J Biol Chem* 1997, **272:**11694-11697.

37. Tang D, Gururajan R, Kidd V: **Phosphorylation of PITSLRE p110 Isoforms Accompanies Their Processing by Caspases during Fas-mediated.** *Cell Death J Biol Chem* 1998, **273:**16601-16607.

38. Feng Y, Shi J, Goldstein AM, Tucker MA, Nelson MA: **Analysis of mutations and identification of several polymorphisms in the putative promoter region of the P34CDC2-related CDC2L1 gene located at 1P36 in melanoma cell lines and melanoma families.** *Int J Cancer* 2002, **99:**834-838.

39. Beltinger C, Fulda S, Kammertoens T, Uckert W, Debatin K: **Mitochondrial Amplification of Death Signals Determines Thymidine Kinase/Ganciclovir-triggered Activation of Apoptosis.** *Cancer Res* 2000, **60:**3212-3217.

40. Sanchez-Alcazar JA, Khodjakov A, Schneider E: **Anticancer drugs induce increased mitochondrial cytochrome c expression that precedes cell death.** *Cancer Res* 2001, **61:**1038-1044.

Pre-publication history

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

http://www.biomedcentral.com/1471-2407/4/4/prepub