



SHORT REPORT

Open Access

Functional genomics annotation of a statistical epistasis network associated with bladder cancer susceptibility

Ting Hu^{1,2}, Qinxin Pan¹, Angeline S Andrew^{2,3}, Jillian M Langer⁴, Michael D Cole⁴, Craig R Tomlinson^{2,4}, Margaret R Karagas^{2,3} and Jason H Moore^{1,2,3*}

* Correspondence:

Jason.H.Moore@Dartmouth.edu
¹Department of Genetics, Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA
²Institute for Quantitative Biomedical Sciences, Dartmouth College, Hanover, NH 03755, USA
Full list of author information is available at the end of the article

Abstract

Background: Several different genetic and environmental factors have been identified as independent risk factors for bladder cancer in population-based studies. Recent studies have turned to understanding the role of gene-gene and gene-environment interactions in determining risk. We previously developed the bioinformatics framework of statistical epistasis networks (SEN) to characterize the global structure of interacting genetic factors associated with a particular disease or clinical outcome. By applying SEN to a population-based study of bladder cancer among Caucasians in New Hampshire, we were able to identify a set of connected genetic factors with strong and significant interaction effects on bladder cancer susceptibility.

Findings: To support our statistical findings using networks, in the present study, we performed pathway enrichment analyses on the set of genes identified using SEN, and found that they are associated with the carcinogen benzo[a]pyrene, a component of tobacco smoke. We further carried out an mRNA expression microarray experiment to validate statistical genetic interactions, and to determine if the set of genes identified in the SEN were differentially expressed in a normal bladder cell line and a bladder cancer cell line in the presence or absence of benzo[a]pyrene. Significant nonrandom sets of genes from the SEN were found to be differentially expressed in response to benzo[a]pyrene in both the normal bladder cells and the bladder cancer cells. In addition, the patterns of gene expression were significantly different between these two cell types.

Conclusions: The enrichment analyses and the gene expression microarray results support the idea that SEN analysis of bladder in population-based studies is able to identify biologically meaningful statistical patterns. These results bring us a step closer to a systems genetic approach to understanding cancer susceptibility that integrates population and laboratory-based studies.

Keywords: Epistasis, Gene-gene interactions, Statistical epistasis networks, Benzo[a]pyrene, Gene-drug association, Bladder cancer

Findings

Introduction

Tobacco consumption has been implicated as the most relevant risk factor for the development of bladder cancer. Among various carcinogens identified in tobacco smoke, the polycyclic aromatic hydrocarbons (PAHs) are well characterized and their risk associations with bladder cancer are well studied [1-4]. In addition to these strong environmental risk factors, recent genome-wide association studies have identified several highly significant genetic factors with small effects [5,6]. Some of these have been shown to modify the effects of smoking on risk of bladder cancer [7].

Most existing genetic association studies have focused on the independent effects of individual genes. That is, they have by design ignored the context of human ecology and the extensive variability in the human genome. As a result, much of the heritability of common human diseases such as bladder cancer remains unexplained. Multiple approaches have been proposed to account for this missing heritability including sequencing to identify causative rare variants and advanced analytical strategies to characterize those genetic effects that are context-dependent [8]. The characterization of gene-gene interaction effects, i.e. *epistasis*, provides a promising research avenue to understand the complex genetic architecture underlying common human diseases in population-based studies [9-11]. Previously, we developed a network-based framework, statistical epistasis networks (SEN), to characterize the global epistatic interaction space of bladder cancer [12,13]. SEN were able to identify a large set of interacting single-nucleotide polymorphisms (SNPs) that had strong associations with bladder cancer. In addition, these SNPs were organized in the form of a connected network structure that was highly significant and nonrandom compared to the null distribution of networks generated using permutation tests.

The goal of the present study was to determine whether the SEN identified for bladder cancer have any biological relevance in the context of the most prevalent risk factor. Because the majority of bladder cancer cases in the United States can be attributed to tobacco exposure, we hypothesized that many of the SEN genes would have a biological role related to biological response to PAH exposure. Here, we applied pathway enrichment analysis of the bladder cancer SEN and found that it is enriched for genes that have strong associations with benzo[*a*]pyrene. We carried out a gene expression microarray experiment in cell lines from normal bladder and bladder cancer cells treated with and without benzo[*a*]pyrene. The goal of this experiment was to provide a functional genomics annotation of the SEN.

Methods and results

In the framework of SEN [12], first we exhaustively measured the non-additive statistical interactions among all pairs of SNPs in a large population-based bladder cancer case-control dataset [14]. The information theoretic measurement *information gain* was used to quantify the epistatic interaction between two SNPs associated with bladder cancer [15-17]. The information gain measure was able to separate the pure synergistic interaction effect from the additive main effects of a pair of genetic attributes, and thus to characterize their statistical epistasis associated with a particular disease status or phenotype. Network properties, including network size, global

network connectivity, and degree distribution, were analyzed to systematically derive a pairwise interaction strength threshold to construct an interaction network that had significantly distinct structures compared to null networks built from permutation testing [12]. This network had a sizable largest connected component including 39 SNPs from 29 cancer susceptibility genes (Table 1), which were absent in any of the 1000 null networks using permuted data. We speculated that these identified 29 genes captured an important aspect of the underlying complex genetic structure of bladder cancer [12].

To further validate the effectiveness of our population-based SEN methodology at capturing meaningful biology in our bladder cancer epistatic interaction network, we performed functional enrichment analyses on those 29 genes using ToppGene [18]. ToppGene detects functional enrichment of a given gene list based on ontologies,

Table 1 The 29 genes identified in our previous genetic interaction study on bladder cancer using SEN

Symbol	Entrez Gene ID	Description
AHRR	57491	Aryl-hydrocarbon receptor repressor
AKR1C3	8644	Aldo-keto reductase family 1, member C3
AXIN2	8313	Axin2
BCL6	604	B-cell CLL/lymphoma 6
BIRC3	330	Baculoviral IAP repeat containing 3
CARD15	64127	Nucleotide-binding oligomerization domain containing 2
CAT	847	Catalase
CCL5	6352	Chemokine (C-C motif) ligand 5
CCNH	902	Cyclin H
FTHFD	10840	Aldehyde dehydrogenase 1 family, member L1
GATA3	2625	GATA binding protein 3
GSTM3	2947	Glutathione S-transferase mu 3 (brain)
HSD17B4	3295	Hydroxysteroid (17-beta) dehydrogenase 4
IGF1R	3480	Insulin-like growth factor 1 receptor
IL1A	3552	Interleukin 1, alpha
INSR	3643	Insulin receptor
MASP1	5648	Mannan-binding lectin serine peptidase 1 (C4/C2 activating component of Ra-reactive factor)
MBD2	8932	Methyl-CpG binding domain protein 2
MMP1	4312	Matrix metalloproteinase 1 (interstitial collagenase)
MYBL2	4605	V-myb myeloblastosis viral oncogene homolog (avian)-like 2
NEDD5	4735	Septin 2
OPRM1	4988	Opioid receptor, mu 1
PARP4	143	Poly (ADP-ribose) polymerase family, member 4
PGR	5241	Progesterone receptor
PIM1	5292	Pim-1 oncogene
RERG	85004	RAS-like, estrogen-regulated, growth inhibitor
TNKS	8658	Tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase
TP53I3	9540	Tumor protein p53 inducible protein 3
XPC	7508	Xeroderma pigmentosum, complementation group C

pathways, phenotypes, drug-gene association, etc. All statistical results were corrected for multiple testing using the Bonferroni method. For our list of 29 genes, the most significant biological process from the Gene Ontology analysis was response to organic cyclic compound ($p = 1.27 \times 10^{-5}$). Figure 1 summarizes all 39 significant biological processes. In addition, the most significant chemical from the drug-gene analysis was benzo[*a*]pyrene ($p = 2.90 \times 10^{-5}$). Figure 2 summarizes all 34 significant drug-gene categories. These enrichment results indicate that our SEN successfully identified interacting genes that are mostly targeted and regulated by the carcinogen benzo[*a*]pyrene.

Next, we performed a gene expression microarray experiment by exposing both normal bladder cells and bladder cancer cells to benzo[*a*]pyrene. Four plates each of the normal urothelial cell line UROtsa and the tumorigenic cell line J82 with and without the treatment of 10 μ M benzo[*a*]pyrene, were cultured. mRNAs were purified and transcription levels were measured using the HumanHT-12 v3 Expression BeadChip (Illumina, Inc., San Diego, CA). There were 47,230 probes on the array of which 29,785 were kept after filtering for detectable probes (detection p -value ≤ 0.05). A quantile normalization method was applied to normalize the data [19]. To test for genes that were differentially expressed in cell lines with and without benzo[*a*]pyrene treatment, we used a

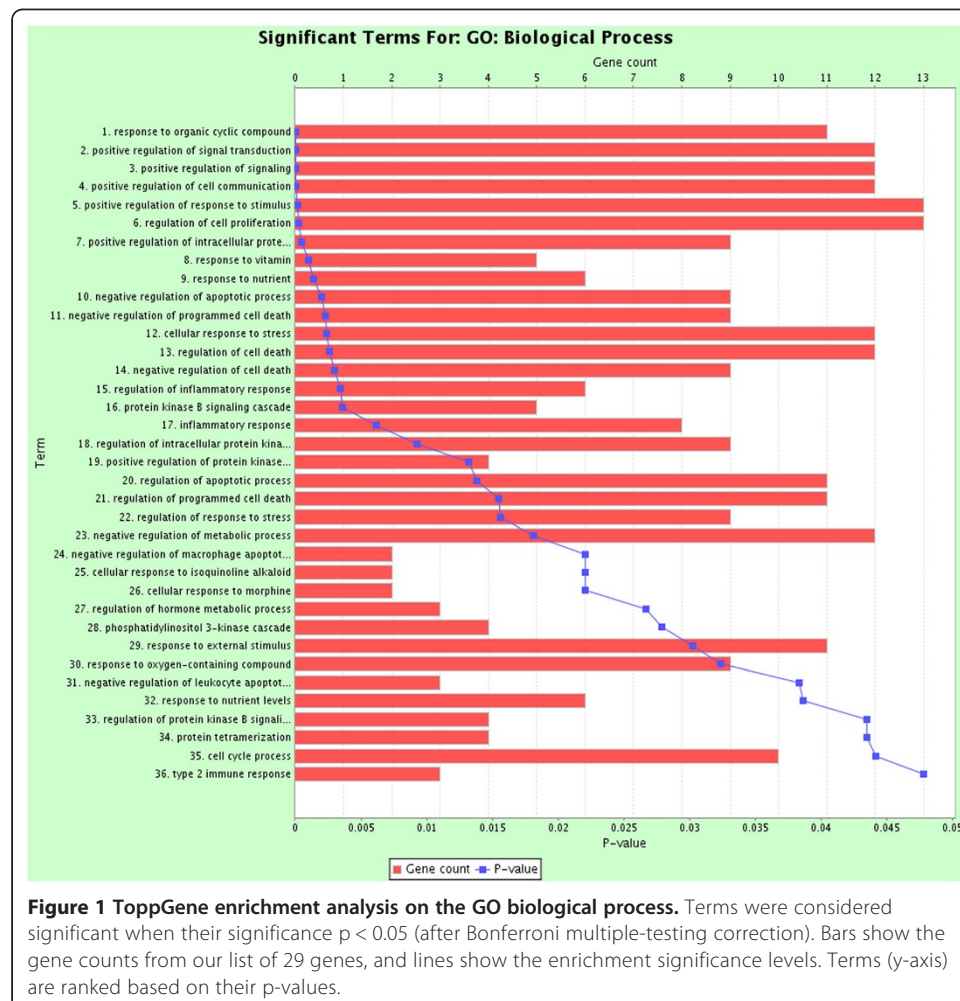
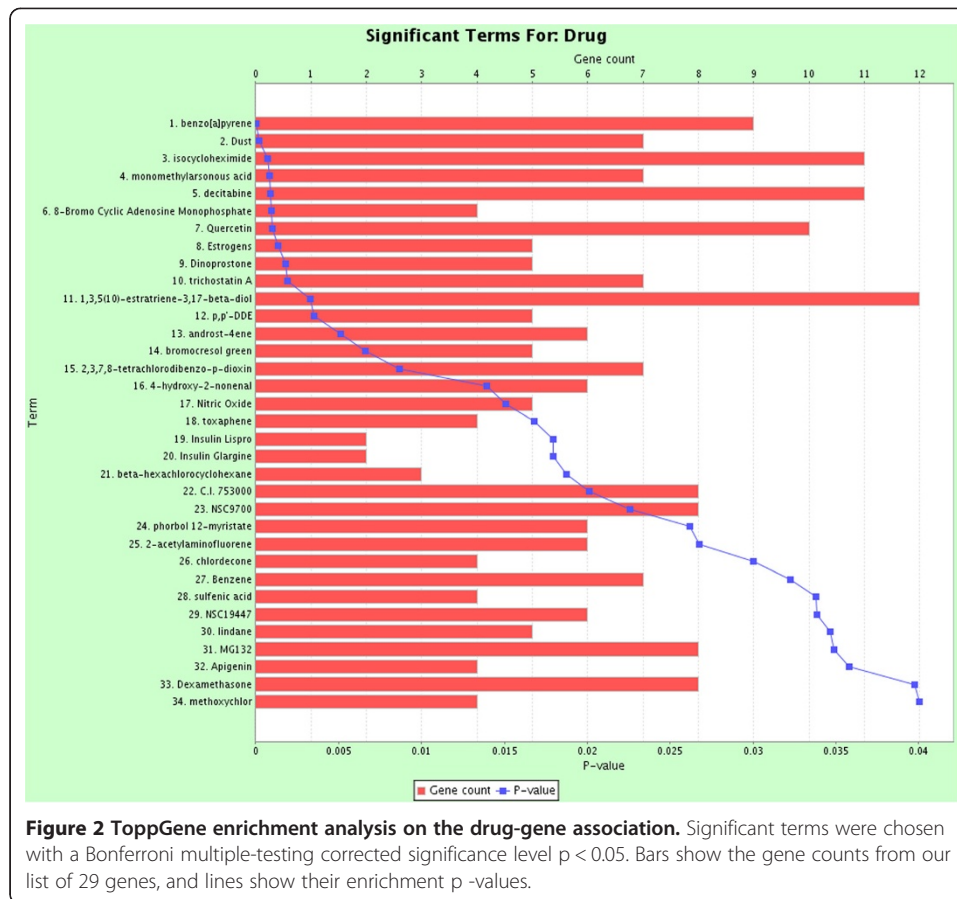


Figure 1 ToppGene enrichment analysis on the GO biological process. Terms were considered significant when their significance $p < 0.05$ (after Bonferroni multiple-testing correction). Bars show the gene counts from our list of 29 genes, and lines show the enrichment significance levels. Terms (y-axis) are ranked based on their p-values.



random variance t -test as described by Wright et al. [20]. Genes with a parametric $p \leq 0.05$ were considered differentially expressed.

Out of the 29 genes identified in our SEN, 26 had detectable probes on the microarray. Their differential expression patterns are shown in Figure 3. Of these, there were 16 genes that had lower expression levels in the J82 untreated bladder cancer cells compared to the UROtsa untreated cells. When comparing J82 cells with and without the treatment, six genes were identified as differentially expressed. Out of these six genes, *BIRC3*, *AKRIC3*, and *BCL6* were up-regulated with the treatment of benzo[*a*]pyrene, whereas *IL1A*, *MBD2*, and *AXIN2* were down-regulated. For the UROtsa cell line with and without the toxicant treatment, six genes were identified as differentially expressed. Among them, *AHRR* and *IL1A* were up-regulated, whereas *NEDD5*, *INSR*, *CAT*, and *BCL6* were down-regulated with the presence of benzo[*a*]pyrene. It is interesting to note that, although *IL1A* was differentially expressed in both cell lines with and without the toxicant treatment, the direction of the change was opposite.

We then annotated the genes with significantly differential expressions, either up-regulated or down-regulated, in our SEN (Figure 4). We first transformed the original SNP-SNP interaction network into a gene-gene interaction network by setting each vertex as a gene and connecting two genes if there existed at least one pair of SNPs, each from one of the two genes, that had a significant and strong pairwise interaction in the original SEN. Genes that were differentially expressed in the normal UROtsa cells with and without the toxicant treatment are shown in blue color in

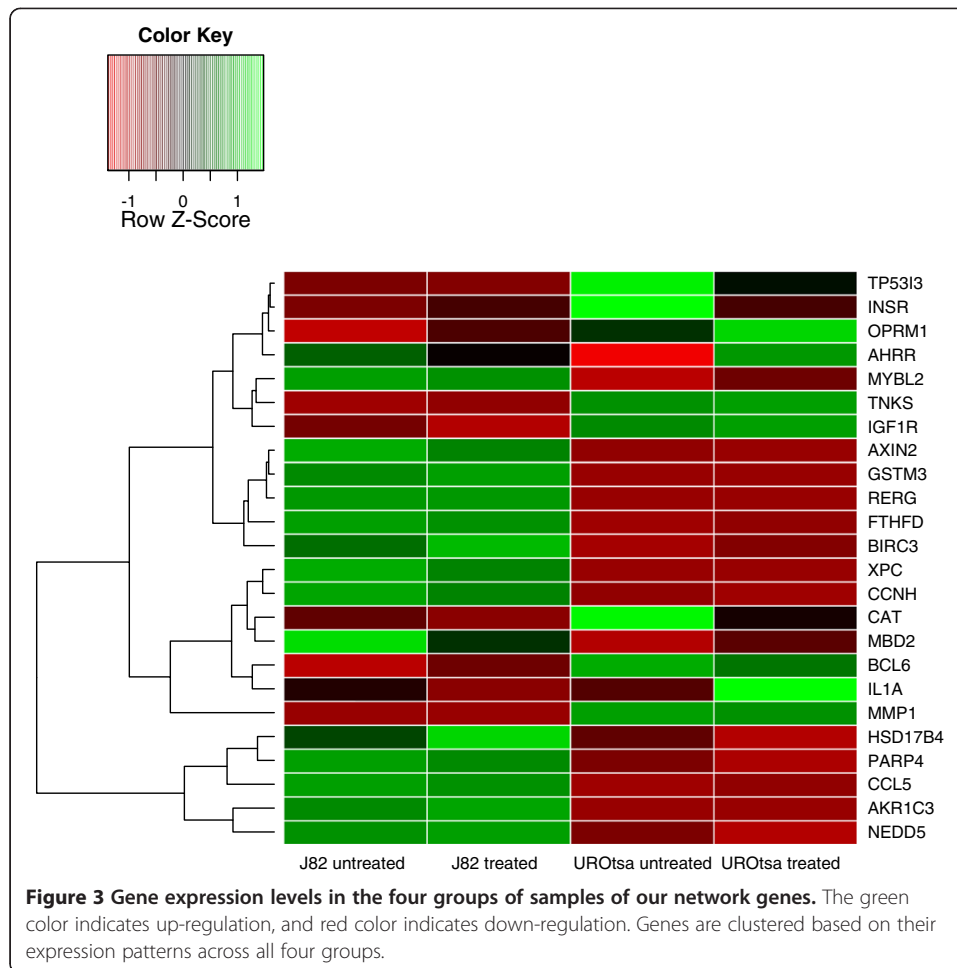
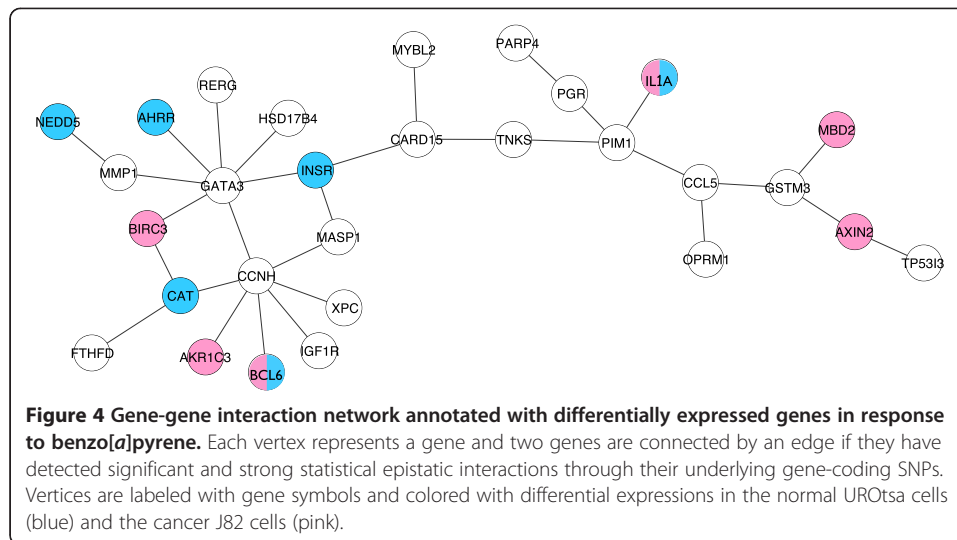


Figure 4. Pink colored vertices represent genes that were differentially expressed in the tumorigenic J82 cells with and without the toxicant treatment. Also note that two genes, *IL1A* and *BCL6*, were identified differentially expressed in both the normal and the cancer cells. The *IL1A* gene encodes for interleukin 1 alpha, a member of the interleukin 1 cytokine family. This cytokine is produced by monocytes and macrophages as a proprotein, which is proteolytically processed and released in response to cell injury and thus induces apoptosis. *IL1A* was up-regulated in the normal UROtsa cell but down-regulated in the tumorigenic J82 cells in response to the treatment of benzo[*a*]pyrene. Since a major function of *IL1A* is to induce apoptosis, a mechanism for programmed killing of potentially harmful cells, the identification of *IL1A* both in the SEN and in the differential gene expression experiment suggests its potential role in interacting with the carcinogen and the development of bladder cancer. Likewise, the Aryl-Hydrocarbon Receptor Repressor (*AHRR*) responded differentially to benzo[*a*]pyrene treatment by cell line. Expression levels are increased by tobacco smoke exposure by a mechanism involving less DNA methylation [21]. We observed increased *AHRR* expression in the normal bladder epithelial cells, as expected, however the opposite relationship in the cancer cell line. To provide further validation for our results, we used the ten differentially expressed SEN genes (Figure 4) for another enrichment



analysis using ToppGene, and a strong drug association with benzo[a]pyrene was detected ($p = 3.733 \times 10^{-3}$).

It is interesting to observe that the normal UROtsa cell and the tumorigenic J82 cells only had two differentially expressed genes in common. The SEN was derived from a population-based case-control study. Genes in the network were those that contained SNPs associated with bladder cancer risk. The fact that their expression levels changed differently in normal and tumorigenic cells supports a biological role for these genes in response to tobacco, that potentially the cells to have greater susceptibility to transformation or carcinogenic progression. The differential responses to benzo[a]pyrene in the normal and cancer cells indicate the possible contribution of interactions between genetic factors and the tobacco carcinogen to bladder cancer risk.

Conclusion

In this study, we used functional genomics data to validate genetic interaction network that was constructed purely based on statistical analysis. We performed an enrichment analysis on a set of genes previously identified by our SEN methodology that had strong and significant epistatic interactions associated with bladder cancer. We tested the hypothesis that these susceptibility genes would be related to tobacco response, since this is the major bladder carcinogen. The enrichment analysis results suggested our network had more genes that respond to a tobacco carcinogen benzo[a]pyrene than would be expected by chance. We then carried out a microarray experiment culturing normal bladder and bladder cancer cells in the presence or absence of benzo[a]pyrene. These results showed that there were more genes in our SEN that respond to benzo[a]pyrene than would be expected by chance given the size of the network. These gene set enrichment and functional genomics annotation analysis results suggest that the SEN methodology may be able to capture biological effects at the heart of the interplay between environmental factors such as smoking and the complex gene-gene interactions that likely influence risk of bladder cancer.

Although not functional proof, this study demonstrates how integrating population-based statistical results with cellular experiments can assist with interpretation of associations to provide a systems genetics approach to understanding cancer susceptibility.

Abbreviations

PAH: Polycyclic aromatic hydrocarbon; SEN: Statistical epistasis networks; SNP: Single-nucleotide polymorphism; GO: Gene Ontology.

Competing interests

The authors declare no competing interests.

Authors' contributions

TH designed the study, performed the analyses, and drafted the manuscript. QP participated in the design of the study, analyzed the microarray experiment results, and participated in drafting the manuscript. ASA and MRK collected the bladder cancer population data and participated in the design of the study. JML, MDC, CRT, and carried out the gene expression experiment and participated in the design of the study. JHM conceived of the study, participated in its design and coordination, and helped drafting the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by NIH grants R01 LM010098, R01 LM009012, R01 AI59694, P20 GM103506 and P20 GM103534.

Author details

¹Department of Genetics, Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA. ²Institute for Quantitative Biomedical Sciences, Dartmouth College, Hanover, NH 03755, USA. ³Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA. ⁴Department of Pharmacology & Toxicology, Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA.

Received: 3 December 2013 Accepted: 5 April 2014

Published: 11 April 2014

References

1. Hoffmann D, Hoffmann I, El-Bayoumy K: **The less harmful cigarette: a controversial issue. A tribute to Ernst L. Wynder.** *Chem Res Toxicol* 2001, **14**(7):767–790.
2. Wolf A, Kutz A, Plottner S, Behm C, Bolt HM, Follmann W, Kuhlmann J: **The effect of benzo[a]pyrene on porcine urinary bladder epithelial cells analyzed for the expression of selected genes and cellular toxicological endpoints.** *Toxicology* 2005, **207**:255–269.
3. Smith CJ, Perfetti TA, Rumble MA, Rodgman A, Doolittle DJ: **"IARC group 2A carcinogens" reported in cigarette mainstream smoke.** *Food Chem Toxicol* 2000, **38**(4):371–383.
4. Smith CJ, Perfetti TA, Mullens MA, Rodgman A, Doolittle DJ: **"IARC group 2B carcinogens" reported in cigarette mainstream smoke.** *Food Chem Toxicol* 2000, **38**(9):825–848.
5. Rothman N, Garcia-Closas M, Chatterjee N, Malats N, Wu X, Figueroa JD, Real FX, Van Den Berg D, Matullo G, Baris D, Thun M, Kiemeneys LA, Vineis P, De Vivo I, Albanes D, Purdue MP, Rafnar T, Hildebrandt MA, Kiltie AE, Cussenot O, Golka K, Kumar R, Taylor JA, Mayordomo JI, Jacobs KB, Kogevinas M, Hutchinson A, Wang Z, Fu YP, Prokunina-Olsson L, et al: **A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci.** *Nat Genet* 2010, **42**:978–984.
6. Garcia-Closas M, Ye Y, Rothman N, Figueroa JD, Malats N, Dinney CP, Chatterjee N, Prokunina-Olsson L, Wang Z, Lin J, Real FX, Jacobs KB, Baris D, Thun M, De Vivo I, Albanes D, Purdue MP, Kogevinas M, Kamat AM, Lerner SP, Grossman HB, Gu J, Pu X, Hutchinson A, Fu YP, Burdett L, Yeager M, Tang W, Tardón A, Serra C, et al: **A genome-wide association study of bladder cancer identifies a new susceptibility locus within SLC14A1, a urea transporter gene on chromosome 18q12.3.** *Hum Mol Genet* 2011, **20**:4282–4289.
7. Garcia-Closas M, Rothman N, Figueroa JD, Prokunina-Olsson L, Han SS, Baris D, Jacobs EJ, Malats N, De Vivo I, Albanes D, Purdue MP, Sharma S, Fu YP, Kogevinas M, Wang Z, Tang W, Tardón A, Serra C, Carrato A, Garcia-Closas R, Lloreta J, Johnson A, Schwenn M, Karagas MR, Schned A, Andriole G Jr, Grubb R 3rd, Black A, Gapstur SM, Thun M, et al: **Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer.** *Canc Res* 2013, **73**:2211–2220.
8. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH: **Missing heritability and strategies for finding the underlying causes of complex disease.** *Nat Rev Genet* 2010, **11**:446–450.
9. Moore JH: **The ubiquitous nature of epistasis in determining susceptibility to common human diseases.** *Hum Hered* 2003, **56**(1–3):73–82.
10. Carlborg O, Haley CS: **Epistasis: too often neglected in complex trait studies?** *Nat Rev Genet* 2004, **5**(8):618–625.
11. Cordell HJ: **Detecting gene-gene interactions that underline human diseases.** *Nat Rev Genet* 2009, **10**(6):392–404.
12. Hu T, Sinnott-Armstrong NA, Kiralis JW, Andrew AS, Karagas MR, Moore JH: **Characterizing genetic interactions in human disease association studies using statistical epistasis networks.** *BMC Bioinforma* 2011, **12**:364.
13. Hu T, Chen Y, Kiralis JW, Moore JH: **ViSEN: methodology and software for visualization of statistical epistasis networks.** *Genet Epidemiol* 2013, **37**(3):283–285.

14. Karagas MR, Tosteson TD, Blum J, Morris JS, Baron JA, Klaue B: **Design of an epidemiologic study of drinking water arsenic exposure and skin and bladder cancer risk in a U.S. population.** *Environ Health Perspect* 1998, **106**(Suppl 4):1047–1050.
15. Moore JH, Gilbert JC, Tsai CT, Chiang FT, Holden T, Barney H, White BC: **A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility.** *J Theor Biol* 2006, **241**(2):252–261.
16. Fan R, Zhong M, Wang S, Zhang Y, Andrew A, Karagas M, Chen H, Amos CI, Xiong M, Moore JH: **Entropy-based information gain approaches to detect and to characterize gene-gene and gene-environment interactions/correlations of complex diseases.** *Genet Epidemiol* 2011, **35**:706–721.
17. Hu T, Chen Y, Kiralis JW, Collins RL, Wejse C, Sirugo G, Williams SM, Moore JH: **An information-gain approach to detecting three-way epistatic interactions in genetic association studies.** *J Am Med Inform Assoc* 2013, **20**:603–636.
18. Chen J, Bardes EE, Aronow BJ, Jegga AG: **ToppGene Suite for gene list enrichment analysis and candidate gene prioritization.** *Nucleic Acids Res* 2009, **37**(suppl 2):W305–W311.
19. Schmid R, Baum P, Itrich C, Fundel-Clemens K, Huber W, Brors B, Eils R, Weith A, Mennerich D, Quast K: **Comparison of normalization methods for illumina BeadChip HumanHT-12 v3.** *BMC Genomics* 2010, **11**:349.
20. Wright GW, Simon RA: **A random variance model for detection of differential gene expression in small microarray experiments.** *Bioinformatics* 2003, **19**:2448–2455.
21. Shenker NS, Ueland PM, Polidoro S, van Veldhoven K, Ricceri F, Brown R, Flanagan JM, Vineis P: **DNA methylation as a long-term biomarker of exposure to tobacco smoke.** *Epidemiology* 2013, **24**(5):712–716.

doi:10.1186/1756-0381-7-5

Cite this article as: Hu et al.: Functional genomics annotation of a statistical epistasis network associated with bladder cancer susceptibility. *BioData Mining* 2014 **7**:5.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

