Published in final edited form as:

Hum Genet. 2014 December ; 133(12): 1467–1476. doi:10.1007/s00439-014-1483-8.

Linking polymorphic p53 response elements with gene expression in airway epithelial cells of smokers and cancer risk

Xuting Wang¹, Gary S. Pittman¹, Omari J. Bandele¹, Jason J. Bischof¹, Gang Liu², John Brothers II², Avrum Spira², and Douglas A. Bell^{1,*}

¹Environmental Genomics Section, Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709

²Section of Computational Biomedicine, Department of Medicine, Boston University, Boston, MA 02118

Abstract

Chronic cigarette smoking exposes airway epithelial cells to thousands of carcinogens, oxidants and DNA damaging agents, creating a field of molecular injury in the airway and altering gene expression. Studies of cytologically normal bronchial epithelial cells from smokers have identified transcription-based biomarkers that may prove useful in early diagnosis of lung cancer, including a number of p53-regulated genes. The ability of p53 to regulate transcription is critical for tumor suppression, and this suggests that single nucleotide polymorphisms (SNPs) in functional p53 binding sites (p53 response elements, or p53REs) that affect gene expression could influence susceptibility to cancer. To connect p53RE SNP genotype with gene expression and cancer risk, we identified a set of 204 SNPs in putative p53REs, and performed *cis* eQTL (expression quantitative trait loci) analysis, assessing associations between genotypes and mRNA levels of adjacent genes in bronchial epithelial cells obtained from 44 cigarette smokers. To further test and validate these genotype-expression associations, we searched published eQTL studies from independent populations and determined that 53% (39/74) of the bronchial epithelial eQTLs were observed in at least one of other studies. SNPs in p53REs were also evaluated for effects on p53-DNA binding using a quantitative *in vitro* protein-DNA binding assay. Last, based on linkage disequilibrium, we found 6 p53RE SNPs associated with gene expression were identified as cancer risk SNPs by either genome-wide association studies (GWAS) or candidate gene studies. We provide an approach for identifying and evaluating potentially functional SNPs that may modulate the airway gene expression response to smoking and may influence susceptibility to cancers.

*To whom correspondence should be addressed: Environmental Genomics Section, Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Mail Drop: C3-03, 111 TW Alexander Drive, PO Box 12233, Research Triangle Park, NC 27709. Tel: 919-541-7686, Fax: 919-541-7593, BELL1@niehs.nih.gov.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Author Contributions

Conceived and designed the experiments: AS DAB. Performed the experiments: XW GSP OJB JB JBII GL. Analyzed the data: XW DAB. Wrote the paper: XW DAB.

Keywords

SNP; polymorphism; p53 response element; smoking; gene expression; eQTL; cancer risk

Introduction

Cigarette smoking contributes to millions of preventable deaths each year (http://www.wpro.who.int/media_centre/fact_sheets/fs_20060530.htm) and is the most important cause of lung cancer, cancer-related death, and chronic obstructive lung disease (Parkin et al. 2005). Despite the strong correlation between smoking and disease, only a modest fraction (10–15%) of smokers develop lung cancer (Shields 1999). Recent studies have demonstrated that cigarette exposure damages epithelial cells in the bronchial airway, creating a molecular field of injury that may represent an early stage of carcinogenesis (Franklin et al. 1997; Slaughter et al. 1953; Spira et al. 2004; Steiling et al. 2008). Gene expression profiling of cytologically normal bronchial epithelial cell brushings obtained from nonsmokers, healthy smokers and smokers with and without lung cancer, has revealed a number of differentially expressed genes and gene pathways among these groups (Spira et al. 2004; Spira et al. 2007). Prominent features of the injured bronchial cells include altered gene expression for oxidative stress, xenobiotic and DNA damage response genes (Spira et al. 2004), presumably due to the presence of oxidants and DNA damaging agents in cigarette smoke. Variability in the airway's transcriptome response to damage from tobacco smoke among smokers appears to be associated with risk of lung cancer, and these airway gene expression profiles may potentially serve as an early diagnostic biomarker for disease (Spira et al. 2007; Steiling et al. 2008).

The observations that individuals differ in the response to smoking suggests a possible role for genetic variability in susceptibility to development of lung cancer. Genome-wide association studies have identified many single nucleotide polymorphisms (SNPs) associated with lung cancer risk but the mechanisms linking SNP genotype to biological function are largely unknown. We and others (Dunham et al. 2012; Schadt et al. 2008; Tomso et al. 2005) have hypothesized that these SNPs impact the regulation of cancer genes by affecting transcription factor binding. A number of studies have examined the relationship between SNPs and *in vitro* gene expression in cell-lines (Cheung et al. 2005; Morley et al. 2004; Stranger et al. 2005; Stranger et al. 2007; Wang et al. 2007) in order to identify expression quantitative trait loci (eQTLs). Using a bioinformatics-directed approach related to the present work, Wang *et al.* (Wang et al. 2007) examined SNPs occurring in binding sites of the transcription factor NRF2 (*NFE2L2*), predicted the functional impact, and compared genotype with *in vitro* expression of HapMap lymphoblastoid cell lines. Until recently there were relatively few eQTL studies comparing *in vivo* expression with genotype in primary human tissues (Schadt et al. 2008; Wang et al. 2010) but now it is possible to compare candidate eQTLs across numerous studies.

The p53 pathway is central in the response to DNA damage, and p53 is frequently mutated in lung cancer. The p53 protein is activated following DNA damage and binds response elements (REs) found in the regulatory regions of p53 target genes. Depending on the

exposure and the extent of damage, the subsequent transactivation or repression of target genes results in divergent outcomes including cell cycle arrest or apoptosis (Riley et al. 2008). We hypothesized that cigarette smoke-induced gene transcription in the airway may be modulated by genetic variation, SNPs in p53 binding sites, which affect p53 binding and may be associated with cancer susceptibility. To test this hypothesis, we carried out a pilot study using the approach shown in Figure 1, identifying p53RE SNPs, and assessing the association of SNP genotypes with gene expression profiles in discovery set of samples from cytologically normal airway epithelial cells obtained at bronchoscopy from smokers. SNPs were evaluated for functional effects on p53 binding to DNA using an *in vitro* microsphere-based protein binding assay. We then asked if SNPs' effect on expression identified in this study were reproducible in independent genome-wide eQTL studies or if these SNPs were associated with cancer risk, either in GWAS or candidate gene studies. Our findings suggest this is a useful approach for identifying and evaluating potentially functional SNPs that may modulate the airway gene expression response to smoking and might influence risk for developing tobacco-related lung cancer.

Materials and Methods

Study population and airway epithelial cell collection

Previously we recruited a group of current and former smokers who were undergoing bronchoscopy as a diagnostic study for clinical suspicion of lung cancer at four institutions: Boston University Medical Center, Boston Veterans Administration, Lahey Clinic and St. James's Hospital (Spira et al. 2007). The study was approved by the Institutional Review Boards of all medical centers, and all participants provided written informed consent. These subjects were included in a previous study of gene expression (Spira et al. 2007; Wang et al. 2010) and 45% of the smokers were later diagnosed with lung cancer. As described in Wang et al. (2010), age varied considerably among the patients in this study and we identified 12 probe-sets (representing 11 genes) with expression strongly correlated with age at FDR 0.1 level; these probes were excluded from further analysis. For this study normal bronchial airway epithelial cells were obtained from the uninvolved mainstem bronchus with an endoscopic cytobrush (Cellebrity Endoscopic Cytobrush, Boston Scientific). RNA was extracted, and its integrity and epithelial cell content was confirmed as described previously (Spira et al. 2007).

Microarray gene expression data acquisition and preprocessing

Approximately eight micrograms of total RNA was processed, labeled and hybridized to Affymetrix HG-U133A GeneChips containing 22,283 probe-sets as described previously (Spira et al. 2007). The MAS 5.0 algorithm was used for background adjustment, normalization, and probe-level summarization of microarray data. All microarray data have been submitted to the Gene Expression Omnibus (GEO) under accession number GSE19027.

Selection of putative p53 response element SNPs

Putative p53 response element SNPs were selected using a published approach developed by our group (Bandelet et al. 2011; Tomso et al. 2005; Wang et al. 2007). Briefly, we

constructed a position weight matrix (PWM) for the p53 binding sites based on the collection of experimentally validated binding sites. Then, we predicted p53REs among all SNP sequences in dbSNP (build 126) using the PWM and mapped putative p53RE SNPs to genes within a 100kb window of the transcription start site (TSS) and the transcription end site (TES). We selected 307 of these putative p53RE SNPs for genotyping.

SNP genotyping

Utilizing Illumina's Assay Design Tool, we designed Illumina GoldenGate assays for a total of 307 putative functional p53RE SNPs. DNA samples derived from whole blood were genotyped according to the manufacturer's protocol using the Illumina BeadStation 500G GoldenGate genotyping platform. Genotypes were assigned using Beadstudio v3.0 genotyping software with GenTrain scores greater than 0.25. Among successfully genotyped SNPs, 204 had minor allele frequency > 0.05 and met the Hardy-Weinberg equilibrium test $p > 0.001$ (are listed in Supplemental Table 1).

Association analysis and permutation testing

A linear regression model was used to evaluate the association between $\log_2$ -transformed expression values of a probe-set and genotypes of a SNP that was within 100kb from a gene start or end, as previously described (Cheung et al. 2005; Stranger et al. 2005). To assess the significance of associations, we performed 10,000 permutations of each expression value relative to the genotypes. Permutation procedures provide a computationally intensive approach to generating significance levels empirically, and has been used for studies with small sample sizes (Cheung et al. 2005; Morley et al. 2004; Stranger et al. 2005; Stranger et al. 2007; Wang et al. 2010). To perform permutation (described in reference (Stranger et al. 2005)), genotype labels were swapped randomly (assuming that individuals are interchangeable under the null) to provide a new datasets to be sampled under the null hypothesis. An association was considered nominally significant if the p-value from the analysis of the observed data was lower than the threshold of the 0.05 tail of the distribution of the minimal p-values from 10,000 permutations. All association tests and permutations were performed using the PLINK software v1.0.3 whole genome association analysis toolset (Purcell et al. 2007), available at <http://pngu.mgh.harvard.edu/purcell/plink/>. We consider the eQTLs identified from these tests as "nominally" significant, and these candidates (Table 1) were then assessed by cross validation in independent eQTL studies listed in Supplemental Table S4.

Cross validation of eQTLs in public eQTL datasets

For each SNP-gene pair identified in this study, we examined if it was also significant in five published eQTL studies (Supplementary Table S4)(Grundberg et al. 2012; Liang et al. 2013; Veyrieras et al. 2008; Westra et al. 2013; Zeller et al. 2010), which have listed *cis* SNP-gene associations in lymphoblastoid cells, peripheral blood cells, adipose, and skin obtained from ~800 to 5000 healthy Caucasians. We limited our search to their most significant eQTLs (at FDR 0.05) that reside within the whole coding region plus 100kb upstream the transcriptional start site (TSS) and 100kb downstream from the transcriptional end site (TES).

Linking eQTL SNPs with cancer risk

To determine whether our candidate eQTL SNPs were among previously reported cancer-associated SNPs, we first curated cancer-associated SNPs listed in the Catalog of Published Genome-Wide Association Studies (GWAS catalog, <http://www.genome.gov/gwastudies>) and the Cancer Genome-wide Association and Meta Analyses Database (Cancer GAMAdb, <http://www.hugenavigator.net/CancerGEMKB>) (Yu et al. 2008); then, using the linkage disequilibrium (LD) data from the 1000 Genomes Project (Abecasis et al. 2012), we also identified all SNPs (proxies) that were in strong LD ($r^2 \geq 0.8$) with cancer SNPs; last we matched our candidate SNPs with cancer-associated SNPs and their proxies.

Microsphere Assay for Protein-DNA binding (MAPD)

To characterize the impact of selected SNPs on p53 binding, we utilized a fluorescent microsphere binding assay to evaluate allele-specific p53-DNA interactions (Bandelet et al. 2011; Noureddine et al. 2009). This microsphere-based flow cytometric assay utilizes fluorescent-dyed beads. Each bead type has a distinct spectral signature and is coated with oligonucleotide sequences containing the p53RE sequences to be tested (Supplementary Table S3). The DNA coated microspheres were hybridized in triplicate with nuclear extracts containing activated p53, and repeated three times. The p53-DNA binding of each allele was quantified using the BioRad Bioplex flow cytometer equipped with a dual laser. It detects both the bead identity and a fluorescent reporter dye attached to the antibody against the target of interest (p53) that is bound to p53RE oligonucleotide coated to the bead. The fluorescence signal intensity reflects relative strength of the p53-DNA interactions. Error bars represent the standard deviation, significance was assessed by t-test.

Results

Association of p53 RE SNP genotypes with gene expression in bronchial epithelial cells from cigarette smokers

Our focus in this project was to examine if genotypes of a set of polymorphisms in bioinformatically identified p53 binding sites were associated with mRNA levels in a discovery set of bronchial epithelial cell samples collected by bronchoscopy from smokers with suspicion of cancer (Spira et al. 2007). We then aimed to validate any tentatively positive associations in a p53 binding assay and also by comparison with independent eQTL studies (Figure 1).

Using a position weight matrix-based bioinformatics approach to identify polymorphic p53 response elements (Bandelet et al. 2011), we selected 307 SNPs in known or putative p53 response elements but only 204 p53RE SNPs had minor allele frequencies large enough (≥ 0.05) to examine among subjects in this pilot study (Supplemental Table S1). Normalized $\log_2$ -transformed gene expression values were evaluated against SNP genotypes by linear regression in all smokers ($n=44$), and permutation based resampling test was applied as previously described by Stranger *et al.* (Stranger et al. 2005). We observed 74 associations ($p^{\text{perm}} < 0.05$) between SNP genotypes of 60 p53RE SNPs and mRNA levels of 58 nearby genes (68 probe-sets), listed in Supplemental Table S2.

Because the small number of samples limited statistical power of our test, we further assessed whether the associations observed here could be validated by replication in independent genome-wide eQTL studies. To do this we queried five eQTL datasets (Grundberg et al. 2012; Liang et al. 2013; Veyrieras et al. 2008; Westra et al. 2013; Zeller et al. 2010)(Supplementary Table S4) that report the top 5% (by FDR) *cis* SNP-gene associations in lymphoid cell lines (LCLs), peripheral blood cells, monocytes, as well as adipose, and skin tissue that were each obtained from populations ranging from 400 to 5000 healthy Caucasians. We found that 39 of 74 (53%) SNP-gene pairs that were nominally significant in our airway epithelial cell study were replicated in at least one of the queried tissues or cells, 14 (19%) SNP-gene pairs were identified as significant eQTLs across 3 or more studies and remarkably, one association (rs10797056, in the upstream of S100A13) was significant in all queried tissues (Table 1). To test if our bioinformatics selection of p53RE SNPs was enriching for reproducible eQTLs, we compared the reproducible rate from our study (53%, 39 of 74) to that among all eQTL studies (14%, 152,048 of 1,082,839 reported in up to 6500 European ancestry individuals). We found a 6.82-fold enrichment ($p = 5.16 \times 10^{-15}$ by Fisher's exact test).

Differential binding of p53 to p53RE SNPs associated with gene expression

We selected 25 putative polymorphic p53REs for testing in an *in vitro* p53 binding assay (Noureddine et al. 2009) to evaluate the impact of the SNP on p53 binding. Oligonucleotides containing the p53RE sequence (supplementary Table S3) were attached to fluorescent microspheres, incubated with human cell nuclear extracts containing activated p53 and a fluorescent p53 antibody and the complexes were read on a flow cytometer (Bandelet et al. 2011; Noureddine et al. 2009). This approach allows the multiplex assessment of relative p53 binding, and the results of three independent trials (in triplicate) are shown in Figure 2A. Two types of information are displayed by the bars, the relative magnitude of p53 binding to the putative p53 target sequence, and the differential impact that the SNP has on p53 binding. Of these putative p53REs, three (rs3737717, rs347519, and rs2286748) showed strong binding to activated p53, with an additional eight showing moderate binding (rs1704743, rs28371212, rs2290847, rs1042546, rs8006147, and rs292512). The rest showed weak binding (e.g. less than 10% of p21 element). Several SNP alleles resulted in significant differential binding, including rs7240884, which produced 84% reduction in binding. The putative p53RE SNP rs347519 resulted in a 51% reduced binding.

Overall, with this small sample of putative p53REs, we found that differential binding was correlated to the difference of PWM scores between predicted strong and weak alleles ($r^2 = 0.362$, $p = 0.0015$, Figure 2B), suggesting PWM score is a relatively good predictor of p53 binding, consistent with a previous study (Bandelet et al. 2011). We also examined whether or not the differential binding was consistent with the difference of gene expression between predicted strong and weak alleles. Among the eQTL SNPs tested in the binding assay, 92% (12/13) displayed a directional change in gene expression consistent with PWM score, and for each of the six reproducible eQTL SNPs, we found that the stronger binding allele corresponded to the higher mRNA level. These analyses provide evidence that SNPs affecting p53 binding also have the potential to influence gene transcription.

SNPs associated with gene expression were also associated with cancer susceptibility

Given the recent increase in our knowledge of genetic variation important in human cancer through genome-wide association studies (GWAS) and candidate gene studies, we sought to determine whether the discovered p53RE eQTL SNPs were among previously reported cancer-associated SNPs. Thus, we looked at the overlap with cancer-associated SNPs listed in the Catalog of Published Genome-Wide Association Studies (GWAS catalog, <http://www.genome.gov/gwastudies>) (Hindorff et al. 2009)) and the Cancer Genome-wide Association and Meta Analyses Database (Cancer GAMAdb, <http://www.hugenavigator.net/CancerGEMKB>) (Yu et al. 2008), including 977 SNPs from GWAS and 309 SNPs from candidate gene studies as of 11/26/2013. We found a p53RE SNP (rs11062386, an eQTL of FKBP4) associated with lung cancer risk, and another p53RE SNP (rs10993994, an eQTL of MSMB) associated with prostate cancer risk.

Because commercial genotyping arrays used in GWAS contain a sample of <10% of the common SNPs in the genome, most SNPs tested in this study were not directly analyzed in any population study of cancer risk. To identify indirect associations, we used linkage disequilibrium (LD) data from the 1000 Genomes Project (Abecasis et al. 2012), to determine if any of our SNPs were in strong LD ($r^2 \geq 0.8$) with cancer risk SNPs. We further found 5 p53RE SNPs that were in strong LD with 8 cancer risk SNPs (Table 3). Of note, rs10993994, a putative p53RE SNP associated with prostate cancer risk, was also in strong LD with other 4 SNPs associated with prostate cancer risk. To test if our eQTL SNPs were more likely to be cancer risk SNPs than non-eQTL SNPs, we tested the proportion of each and found that 10.0% (6/60) of eQTL SNPs were also cancer risk SNPs whereas only 1.6% (4/247) of non-eQTL SNPs were cancer risk SNPs. There was a significant difference in the ratio of cancer risk SNPs between these two groups (enrichment of 6.7-fold, $p = 0.0049$, by Fisher's exact test). More detailed functional analysis, as well as population studies, will be needed to establish if these connections between putatively functional p53 pathway SNPs and cancer risk are useful in predictive models.

Discussion

This analysis, while limited in scope and statistical power, highlights an approach for linking genetic variants in p53 or other transcription factor binding sites with cancer risk through the regulation of gene expression. A central role for p53 in lung tumors is well established (Herbst et al. 2008; Toyooka et al. 2003), but there have been relatively few *in vivo* studies of p53 and p53-regulated gene expression in normal human tissues chronically exposed to cigarette smoke (Lampe et al. 2004). Early involvement of p53 in lung tumorigenesis has been strongly suggested by studies that have detected mutant p53 DNA sequences in normal tissues and in tissues at the margins of tumors (Ahrendt et al. 1999; Franklin et al. 1997; Harden et al. 2004). The implication of such events could be altered regulation of p53 pathway genes and susceptibility to transformation. A previous, related study (Spira et al. 2007) indicated that gene expression signatures in cytologically normal large-airway epithelial cells can serve as a useful biomarker for early diagnosis of lung cancer, and these signatures may be due to an acquired airway-wide response to cigarette exposure, sometimes termed a "field effect" (Steiling et al. 2008). We initiated this study with an *a priori*

hypothesis that the p53-mediated airway response to cigarette smoke might be modulated by genetic variation in the p53 pathway, in particular SNPs within p53 binding motifs. Recent reports (Nikulenkov et al, 2012; Zeron-Medina, et al. 2013) suggest there may be thousands of possible p53 binding sites (p53REs) in human genome, and in this work we explored an approach that combines bioinformatics, genetics, and population studies for identifying functional polymorphic p53REs.

Using a position weight matrix model to identify potential p53RE SNPs (Bandelet et al. 2011), we scanned dbSNP build 126 and picked 204 p53RE SNPs (MAF $\geq$ 0.05) to evaluate as eQTLs in bronchial epithelial cells. The procedure yielded 60 *cis*-acting, putative p53RE eQTL SNPs associated with 58 genes. We tested a set of 25 p53RE SNPs to determine if they disrupted binding of p53 to DNA and found that 11 elements bound p53 to a substantial degree and 9 displayed significant differential binding. Notably, this functional assay provides evidence that SNPs affecting p53 binding also have the potential to influence gene transcription.

This pilot study had the goal of testing a bioinformatics-based method for picking candidate eQTLs in the p53 pathway using bronchial epithelial cells sampled from a small group of individuals being examined for risk of lung cancer. Because the detected bronchial epithelial cell eQTL associations in these groups have limited statistical power, these findings as they pertain to lung cells or lung cancer, need to be verified in larger studies. Thus, while similar studies of lung tissue were unavailable for comparison, we endeavored to validate our findings by querying several independent eQTL databases (Grundberg et al. 2012; Liang et al. 2013; Veyrieras et al. 2008) and found that 39 SNP-gene pairs identified here (53%) were also observed in at least one of the queried tissues, and 14 were significant in 3 or more queried tissues. This is a highly significant enrichment over the reproducible rate among all eQTL databases (14%).

Conceptually, it seems that an observation of consistency among SNP functional effects should strengthen biological plausibility of the finding. For example, a SNP that affected p53 binding could reduce transcription of a target gene, resulting in lower expression in tissues of cancer cases, and this might also be reflected in a higher genotype frequency among cancer cases. One disappointing outcome of the present study was that among the bioinformatically chosen p53RE SNPs examined, we were unable to observe consistency across all of the outcomes evaluated. Strategy for selection of SNPs could be improved by using p53 chromatin immunoprecipitation sequencing (ChIP-Seq) as a functional screen to identify p53 binding sites and then searching for SNPs within these p53 occupied genomic locations (Zeron-Medina et al. 2013).

Despite these limitations, it is clear that SNPs that impact p53 binding and p53-induced gene expression could potentially affect cancer risk, thus we searched public available SNP databases on cancer risk to determine if any of the SNPs studied had been reported as cancer risk SNPs or if they were in linkage disequilibrium with cancer risk SNPs. We identified that 10 p53 pathway SNPs were associated with both gene expression in bronchial epithelial cells of smokers and risk to various cancers, including two SNPs associated with lung cancer risk. It strongly suggests that this pathway-directed, eQTL approach has a promise for

identifying disease-associated SNPs. Our analysis reveals the potential for genetic modulation of the p53-mediated exposure response pathway in bronchial epithelial cells, and future work will examine this possibility in larger studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded in part by the Intramural Research Program (projects: Z01-ES100475 and Z01 ES046008) of the National Institute of Environmental Health Sciences, National Institutes of Health and grants to Avrum Spira from the National Institute of Health (U01ES016035, R01CA124640). We are grateful for comments from Drs. Daniel Menendez and Jack Taylor, NIEHS.

References

- Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491:56–65. [pii]. 10.1038/nature11632 [PubMed: 23128226]
- Ahrendt SA, Chow JT, Xu LH, Yang SC, Eisenberger CF, Esteller M, Herman JG, Wu L, Decker PA, Jen J, Sidransky D. Molecular detection of tumor cells in bronchoalveolar lavage fluid from patients with early stage lung cancer. *J Natl Cancer Inst*. 1999; 91:332–9. [PubMed: 10050866]
- Bandelet OJ, Wang X, Campbell MR, Pittman GS, Bell DA. Human single-nucleotide polymorphisms alter p53 sequence-specific binding at gene regulatory elements. *Nucleic Acids Res*. 2011; 39:178–89. 10.1093/nar/gkq764 [PubMed: 20817676]
- Bergamaschi D, Gasco M, Hiller L, Sullivan A, Syed N, Trigiante G, Yulug I, Merlano M, Numico G, Comino A, Attard M, Reelfs O, Gusterson B, Bell AK, Heath V, Tavassoli M, Farrell PJ, Smith P, Lu X, Crook T. p53 polymorphism influences response in cancer chemotherapy via modulation of p73-dependent apoptosis. *Cancer Cell*. 2003; 3:387–402. [PubMed: 12726864]
- Bergamaschi D, Samuels Y, Sullivan A, Zvelebil M, Breysens H, Bisso A, Del Sal G, Syed N, Smith P, Gasco M, Crook T, Lu X. iASPP preferentially binds p53 proline-rich region and modulates apoptotic function of codon 72 polymorphic p53. *Nature Genetics*. 2006; 38:1133–1141. [PubMed: 16964264]
- Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. Mapping determinants of human gene expression by regional and genome-wide association. *Nature*. 2005; 437:1365–9. [PubMed: 16251966]
- Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis C, Doyle F, Epstein CB, Frietze S, Harrow J, Kaul R, Khatun J, Lajoie BR, Landt SG, Lee BK, Pauli F, Rosenbloom KR, Sabo P, Safi A, Sanyal A, Shores N, Simon JM, Song L, Trinklein ND, Altshuler RC, Birney E, Brown JB, Cheng C, Djebali S, Dong XJ, Dunham I, Ernst J, Furey TS, Gerstein M, Giardine B, Greven M, Hardison RC, Harris RS, Herrero J, Hoffman MM, Iyer S, Kellis M, Khatun J, Kheradpour P, Kundaje A, Lassmann T, Li QH, Lin X, Marinov GK, Merkel A, Mortazavi A, Parker SCJ, Reddy TE, Rozowsky J, Schlesinger F, Thurman RE, Wang J, Ward LD, Whitfield TW, Wilder SP, Wu W, Xi HLS, Yip KY, Zhuang JL, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M, Pazin MJ, Lowdon RF, Dillon LAL, Adams LB, Kelly CJ, Zhang J, Wexler JR, Green ED, Good PJ, Feingold EA, Bernstein BE, Birney E, Crawford GE, Dekker J, Elnitski L, Farnham PJ, Gerstein M, Giddings MC, Gingeras TR, Green ED, Guigo R, Hardison RC, Hubbard TJ, Kellis M, Kent WJ, Lieb JD, Margulies EH, Myers RM, Snyder M, Stamatoyannopoulos JA, Tenenbaum SA, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57–74. 10.1038/Nature11247 [PubMed: 22955616]
- Franklin WA, Gazdar AF, Haney J, Wistuba IL, La Rosa FG, Kennedy T, Ritchey DM, Miller YE. Widely dispersed p53 mutation in respiratory epithelium. A novel mechanism for field carcinogenesis. *J Clin Invest*. 1997; 100:2133–7. [PubMed: 9329980]

- Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A, Nisbett J, Sekowska M, Wilk A, Shin SY, Glass D, Travers M, Min JL, Ring S, Ho K, Thorleifsson G, Kong A, Thorsteindottir U, Ainali C, Dimas AS, Hassanali N, Ingle C, Knowles D, Krestyaninova M, Lowe CE, Di Meglio P, Montgomery SB, Parts L, Potter S, Surdulescu G, Tsaprouni L, Tsoka S, Bataille V, Durbin R, Nestle FO, O'Rahilly S, Soranzo N, Lindgren CM, Zondervan KT, Ahmadi KR, Schadt EE, Stefansson K, Smith GD, McCarthy MI, Deloukas P, Dermitzakis ET, Spector TD. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet.* 2012; 44:1084–9. ng.2394 [pii]. 10.1038/ng.2394 [PubMed: 22941192]
- Harden SV, Thomas DC, Benoit N, Minhas K, Westra WH, Califano JA, Koch W, Sidransky D. Real-time gap ligase chain reaction: a rapid semiquantitative assay for detecting p53 mutation at low levels in surgical margins and lymph nodes from resected lung and head and neck tumors. *Clin Cancer Res.* 2004; 10:2379–85. [PubMed: 15073114]
- Herbst RS, Heymach JV, Lippman SM. Lung cancer. *N Engl J Med.* 2008; 359:1367–80. [PubMed: 18815398]
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A.* 2009; 106:9362–7. 10.1073/pnas.0903103106 [PubMed: 19474294]
- Lampe JW, Stepaniants SB, Mao M, Radich JP, Dai H, Linsley PS, Friend SH, Potter JD. Signatures of environmental exposures using peripheral leukocyte gene expression: tobacco smoke. *Cancer Epidemiol Biomarkers Prev.* 2004; 13:445–53. [PubMed: 15006922]
- Liang L, Morar N, Dixon AL, Lathrop GM, Abecasis GR, Moffatt MF, Cookson WO. A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Res.* 2013; 23:716–26. gr.142521.112 [pii]. 10.1101/gr.142521.112 [PubMed: 23345460]
- Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. Genetic analysis of genome-wide variation in human gene expression. *Nature.* 2004; 430:743–7. [PubMed: 15269782]
- Nikulenkov F, Spinnler C, Li H, Tonelli C, Shi Y, Turunen M, Kivioja T, Ignatiev I, Kel A, Taipale J, Selivanova G. Insights into p53 transcriptional function via genome-wide chromatin occupancy and gene expression analysis. *Cell Death Differ.* 2012; 19:1992–2002. 10.1038/cdd.2012.89 [PubMed: 22790872]
- Noureddine MA, Menendez D, Campbell MR, Bandele OJ, Horvath MM, Wang X, Pittman GS, Chorley BN, Resnick MA, Bell DA. Probing the functional impact of sequence variation on p53-DNA interactions using a novel microsphere assay for protein-DNA binding with human cell extracts. *PLoS Genet.* 2009; 5:e1000462. [PubMed: 19424414]
- Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin.* 2005; 55:74–108. [PubMed: 15761078]
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81:559–75. [PubMed: 17701901]
- Riley T, Sontag E, Chen P, Levine A. Transcriptional control of human p53-regulated genes. *Nat Rev Mol Cell Biol.* 2008; 9:402–12. [PubMed: 18431400]
- Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY, Kasarskis A, Zhang B, Wang S, Suver C, Zhu J, Millstein J, Sieberts S, Lamb J, GuhaThakurta D, Derry J, Storey JD, Avila-Campillo I, Kruger MJ, Johnson JM, Rohl CA, van Nas A, Mehrabian M, Drake TA, Lusk AJ, Smith RC, Guengerich FP, Strom SC, Schuetz E, Rushmore TH, Ulrich R. Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* 2008; 6:e107. [PubMed: 18462017]
- Shields PG. Molecular epidemiology of lung cancer. *Ann Oncol.* 1999; 10(Suppl 5):S7–11. [PubMed: 10582132]
- Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer.* 1953; 6:963–8. [PubMed: 13094644]
- Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, Palma J, Brody JS. Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci U S A.* 2004; 101:10143–8. [PubMed: 15210990]

- Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, Gilman S, Dumas YM, Calner P, Sebastiani P, Sridhar S, Beamis J, Lamb C, Anderson T, Gerry N, Keane J, Lenburg ME, Brody JS. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med.* 2007; 13:361–6. [PubMed: 17334370]
- Steiling K, Ryan J, Brody JS, Spira A. The field of tissue injury in the lung and airway. *Cancer Prev Res (Phila Pa).* 2008; 1:396–403.
- Stranger BE, Forrest MS, Clark AG, Minichiello MJ, Deutsch S, Lyle R, Hunt S, Kahl B, Antonarakis SE, Tavare S, Deloukas P, Dermitzakis ET. Genome-wide associations of gene expression variation in humans. *PLoS Genet.* 2005; 1:e78. [PubMed: 16362079]
- Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, Ingle CE, Dunning M, Flicek P, Koller D, Montgomery S, Tavare S, Deloukas P, Dermitzakis ET. Population genomics of human gene expression. *Nat Genet.* 2007; 39:1217–24. [PubMed: 17873874]
- Tomso DJ, Inga A, Menendez D, Pittman GS, Campbell MR, Storici F, Bell DA, Resnick MA. Functionally distinct polymorphic sequences in the human genome that are targets for p53 transactivation. *Proc Natl Acad Sci U S A.* 2005; 102:6431–6. [PubMed: 15843459]
- Toyooka S, Tsuda T, Gazdar AF. The TP53 gene, tobacco exposure, and lung cancer. *Hum Mutat.* 2003; 21:229–39. [PubMed: 12619108]
- Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, Pritchard JK. High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* 2008; 4:e1000214.10.1371/journal.pgen.1000214 [PubMed: 18846210]
- Wang X, Chorley BN, Pittman GS, Kleeberger SR, Brothers J 2nd, Liu G, Spira A, Bell DA. Genetic variation and antioxidant response gene expression in the bronchial airway epithelium of smokers at risk for lung cancer. *PLoS One.* 2010; 5:e11934.10.1371/journal.pone.0011934 [PubMed: 20689807]
- Wang X, Tomso DJ, Chorley BN, Cho HY, Cheung VG, Kleeberger SR, Bell DA. Identification of polymorphic antioxidant response elements in the human genome. *Hum Mol Genet.* 2007; 16:1188–200. [PubMed: 17409198]
- Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, Zhernakova A, Zhernakova DV, Veldink JH, Van den Berg LH, Karjalainen J, Withoff S, Uitterlinden AG, Hofman A, Rivadeneira F, t Hoen PA, Reinmaa E, Fischer K, Nelis M, Milani L, Melzer D, Ferrucci L, Singleton AB, Hernandez DG, Nalls MA, Homuth G, Nauck M, Radke D, Volker U, Perola M, Salomaa V, Brody J, Suchy-Dicey A, Gharib SA, Enquobahrie DA, Lumley T, Montgomery GW, Makino S, Prokisch H, Herder C, Roden M, Grallert H, Meitinger T, Strauch K, Li Y, Jansen RC, Visscher PM, Knight JC, Psaty BM, Ripatti S, Teumer A, Frayling TM, Metspalu A, van Meurs JB, Franke L. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet.* 2013; 45:1238–43.10.1038/ng.2756 [PubMed: 24013639]
- Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. *Nat Genet.* 2008; 40:124–5.10.1038/ng0208-124 [PubMed: 18227866]
- Zeller T, Wild P, Szymczak S, Rotival M, Schillert A, Castagne R, Maouche S, Germain M, Lackner K, Rossmann H, Eleftheriadis M, Sinning CR, Schnabel RB, Lubos E, Mennerich D, Rust W, Perret C, Proust C, Nicaud V, Loscalzo J, Hubner N, Tregouet D, Munzel T, Ziegler A, Tired L, Blankenberg S, Cambien F. Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One.* 2010; 5:e10693.10.1371/journal.pone.0010693 [PubMed: 20502693]
- Zeron-Medina J, Wang X, Repapi E, Campbell MR, Su D, Castro-Giner F, Davies B, Peterse EF, Sacilotto N, Walker GJ, Terzian T, Tomlinson IP, Box NF, Meinshausen N, De Val S, Bell DA, Bond GL. A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection. *Cell.* 2013; 155:410–22.10.1016/j.cell.2013.09.017 [PubMed: 24120139]

Figure 1.

Schematic overview: (A) Putative p53RE SNPs were identified using bioinformatics methods and SNP genotyping arrays were created; (B) SNP genotyping in 44 smokers; (C) Association analysis of SNP genotype and gene expression carried out in bronchial epithelial cells from smokers at risk for lung cancer; (D) Cross validation against eQTL databases; (E) *In vitro* test of the impact of SNP on p53 binding; (F). Linking eQTL SNPs with cancer risk SNPs discovered by GWAS or candidate gene studies.

Fig. 2a

Fig. 2b

Figure 2.

Figure 2A. Binding of activated p53 to putative p53 response element target sequences using MAPD method. Relative magnitude of p53 binding to the target sequence (y-axis) is

displayed. The impact that the SNP has on p53 binding is demonstrated by the difference in the paired bars (mean $\pm$ standard deviation). Significance assessed by t-test.

Figure 2B. Plot of the differential p53 binding and the difference of PWM scores between predicted strong and weak alleles.

Table 1

Cross validation of SNP-gene expression associations. Association was assessed by linear regression between log₂-transformed expression values of a probe-set and genotypes from all smokers (n=44). The P-values were calculated from a permutation analysis (distribution of the minimal p-values from 10,000 permutations). We compared our eQTLs to the top 5% FDR of eQTLs from independent studies (listed in supplementary Table S4) and the reported p-values are listed.

rs	Gene	BEC	LCL	PBC	Adipose	Skin	Monocyte
rs2301559	AKAP9	0.0011	9.8E-38				
rs1391221	C11orf73	0.0053		1.7E-06			
rs12364970	CD59	0.0053			7.1E-05		
rs831629	CD59	0.0084	2.9E-20		1.5E-09	5.4E-09	
rs6107581	CDS2	0.0199			7.0E-04		
rs12105811	CFLAR	0.0065	1.5E-09				
rs686402	COQ9	0.0082		2.6E-10			
rs2302459	CRK	0.0309		1.2E-08			
rs2472649	CXCL5	0.0370		5.4E-90			
rs6504187	DDX42	1.00E-04	4.2E-12	1.5E-14	5.5E-08		
rs7223966	DDX42	0.0113	1.4E-10	2.6E-09	3.3E-10	5.5E-04	
rs1648282	DUOX1	0.0115	2.4E-04				
rs12364970	FBXO3	0.0055		2.1E-07			
rs831626	FBXO3	0.0139		4.1E-06			
rs11157930	HAUS4	0.0067	3.9E-25				
rs14391	INPP5K	0.0045		4.6E-05			
rs12453963	JUP	0.0344		1.5E-05			
rs1407885	LAPTM5	0.0500		1.6E-23			
rs17705633	LOC100506469	0.0098					
rs2290847	LRBA	0.0023	1.0E-19	3.3E-09			
rs2256974	LST1	0.0400	2.5E-36	1.1E-06	2.6E-13	6.8E-18	
rs6922865	MAPK14	0.0063	1.8E-13	1.1E-19			8.6E-12
rs10993994	MSMB	0.0203					
rs10098474	MSRA	0.0335	5.0E-08				
rs3132424	PPFIA1	0.0285	9.8E-10	2.6E-63			

rs	Gene	BEC	LCL	PBC	Adipose	Skin	Monocyte
rs10797056	S100A13	0.0113	4.8E-44	6.6E-30	3.8E-56	2.2E-28	1.1E-69
rs884171	SAE1	0.0237		3.1E-19			
rs2073083	SAMM50	0.0035	6.0E-09	9.8E-198	6.1E-08	1.0E-04	
rs276685	SENP6	0.0430	5.8E-09				
rs337253	SLC30A5	0.0456	2.9E-08	2.4E-13	1.7E-07		
rs2054393	SLC39A8	0.0053		7.7E-06			4.2E-05
rs2297216	SMU1	0.0442	3.4E-08				
rs7741729	STK38	0.0408		2.2E-07			
rs12151019	TMEM147	0.0490		7.3E-07			
rs11658168	TNFSF12	0.0311		1.7E-29			5.2E-10
rs12195078	UBE2J1	0.0487	1.7E-12	1.9E-46			5.5E-06
rs7744352	UBE2J1	0.0010	6.7E-10	1.1E-12	1.6E-07		
rs1358851	UBE2J1	0.0132		4.2E-23			
rs729437	ZFYVE21	0.0490	1.2E-11				

BEC =bronchial epithelial cells (this study); LCL= lymphoblastoid cell lines; PBC=peripheral blood cell.

Table 2

Evaluation of p53RE SNPs for differential binding to p53 by *in vitro* binding assay

SNP	Associated Gene	Sequence	Predicted Strong/weak Allele	Minor Allele Frequency (MAF)	Binding Strength % of p21	MAPD p53 Binding Difference
p53 consensus RRRWWGYYY...RRRCWWGYYY						
rs9724673	ACBD3	CAACRTGGTAcAGTCATGCTC	A/G	A (0.38)	5.7	-9.0
rs2286748	TRIM8	TGGCAGGTCAGAMCTTGCCC	A/C	A (0.26)	24.0 ^S	-71.2 ^{**}
rs2254343	C19orf54	CSTCATGTCTcCGGCCCTGGAG	G/C	G (0.24)	5.9	-4.8
rs831629	CD59	CCCYAAGCTTTGGCTCTCTT	C/T	T (0.40)	4.7	-33.0
rs12105811	CFLAR	GATCCTGCTTGGTCRTGTCC	A/G	A (0.17)	7.0	-42.2 [*]
rs3132424	CTTN	GGCTGGCCTGGCCTKGCCT	T/G	G (0.35)	5.7	4.4
rs1704743	DDX18	CAGCATGAATaAGACYTGTTC	T/C	C (0.06)	11.2 ^M	-31.1
rs7223966	DDX42	ACACTRGCTaAGCCTGCTG	A/G	A (0.26)	4.9	-42.9
rs1057556	EEF1D	KITGAAGCTCAGACTTGGTC	G/T	T (0.13)	4.0	-13.8
rs2274227	FAM20B	TGACAAAGTTGtTGKCAAGGTT	G/T	T (0.42)	3.5	-8.6
rs3737717	GLTPD1	GAAATGYTTGCACATGTGT	C/T	C (0.06)	55.6 ^S	-18.6 ^{**}
rs28371212	HLA-DQB1	GAGCTARCCTACACATGTCA	G/A	A (0.25)	14.5 ^M	-45.6 [*]
rs1407885	LAPTM5	TAACTTGYYCgGCCCAAGCTC	C/T	T (0.37)	8.7	-4.98
rs2290847	LRBA	AAACAGGCTTaTAGCAYGTCT	C/T	T (0.41)	17.9 ^M	-77.0 ^{**}
rs1042546	MARCKSL1	GCACATGCTCTGTRCATGTGT	G/A	G (0.07)	13.0 ^M	62.2
rs2235930	MFAP2	AGAYTTGTAGcGAGCAAGCAG	C/T	C (0.47)	6.1	-66.9 ^{**}
rs7037941	MLLT3	CKTCTAGCTCATGCTTGTCC	G/T	T (0.46)	3.5	-15.7
rs9380152	NEDD9	GGGCAAGACTgAGACTYGGTG	C/T	T (0.17)	8.2	-1.9
rs8006147	PGF	CTTAATGCCCCAGCRTGTCT	A/G	G (0.46)	13.0 ^M	-53.0 [*]
rs347519	KCNN4/PLAUR	GGACAGGCATGGACAyGCAC	T/C	C (0.23)	28.7 ^S	-50.6 ^{**}
rs7240884	PMAIP1	CAACATRTTTGATCTTGTAG	G/A	A (0.39)	14.3 ^M	-83.6 ^{***}
rs618369	PTPRF	GGGCAGGTGTGGGGCTKAC	T/G	G (0.31)	6.3	8.9
rs762670	SBF1	TCACAAGCCTGGTCTTGCgY	C/T	T (0.20)	11.2 ^M	-15.7
rs292512	TMEM140	AAGCTTGTCaAGACTAGTYT	T/C	T (0.24)	13.7 ^M	18.9

SNP	Associated Gene	Sequence	Predicted Strong/weak Allele	Minor Allele Frequency (MAF)	Binding Strength % of p21	MAPD p53 Binding Difference
rs7524763	WDTC1	GAGCMTGCTcACACCAGCCC	A/C	A (0.34)	9.4	-34.5

^{\$}R= A/G, Y= T/C, W= A/T K = T/G, M =A/C; lower case indicates spacer nucleotides between p53 RE half-sites;

M Moderate binding;

S Strong binding.

* p< 0.05,

** p<0.01

*** p< 0.001

Table 3

eQTL SNPs associated with cancer risk in GWAS or candidate gene studies.

eQTL SNP	gene	eQTL	risk SNP	LD with risk SNP	PUBMED ID	cancer type	OR[95%CI]	study type	p-value
rs11062386	FKBP4	yes	yes		19836008	lung cancer	1.10[1.06–1.15]	GWAS	4.14E-06
rs10993994	MSMB	yes	yes		18264097	prostate cancer	1.25[1.17–1.34]	GWAS	9.00E-29
rs10993994	MSMB	yes		rs2611512	18264097	prostate cancer		GWAS	2.70E-12
rs10993994	MSMB	yes		rs3123078	19767753	prostate cancer		GWAS	1.00E-19
rs10993994	MSMB	yes		rs7920517	18264097	prostate cancer		GWAS	1.30E-19
rs10993994	MSMB	yes		rs11006207	18264097	prostate cancer		GWAS	3.00E-14
rs2301559	AKAP9	yes		rs6964587	18334708	breast cancer	1.1[1.04–1.17]	Candidate	0.001
rs729437	ZFYVE21	yes		rs1799794	19116388	breast cancer	1.09[1.00–1.19]	Candidate	0.05
rs2072195	PES1			rs2283873	22544364	Wilms tumor	2.05[1.57–2.69]	GWAS	5.00E-12
rs12602273	MPDU1			rs12947788	21448430	esophageal cancer	0.73 [0.57–0.94]	Candidate	0.014